

Ausgewählte Schranken der Standardnormalverteilung und der χ^2 -Verteilung (1 FG)
für die einseitige und für die zweiseitige Fragestellung

P	z		χ^2 für einen Freiheitsgrad	
	einseitig	zweiseitig	einseitig	zweiseitig
0,001	3,090	3,291	9,550	10,828
0,01	2,326	2,576	5,412	6,635
0,05	1,645	1,960	2,706	3,841
0,10	1,282	1,645	1,642	2,706
0,20	0,842	1,282	0,708	1,642
0,50	0	0,674	0	0,455

Das griechische Alphabet

Griechischer Buchstabe	Name des Buchstabens	Griechischer Buchstabe	Name des Buchstabens
<i>A</i> α	Alpha	<i>N</i> ν	Ny
<i>B</i> β	Beta	<i>Ξ</i> ξ	Xi
<i>Γ</i> γ	Gamma	<i>O</i> o	Omikron
<i>Δ</i> δ	Delta	<i>Π</i> π	Pi
<i>E</i> ε	Epsilon	<i>P</i> ρ	Rho
<i>Z</i> ζ	Zeta	<i>Σ</i> $\sigma \varsigma$	Sigma
<i>H</i> η	Eta	<i>T</i> τ	Tau
<i>Θ</i> ϑ	Theta	<i>Υ</i> υ	Ypsilon
<i>I</i> ι	Jota	<i>Φ</i> φ	Phi
<i>K</i> κ	Kappa	<i>X</i> χ	Chi
<i>Λ</i> λ	Lambda	<i>Ψ</i> ψ	Psi
<i>M</i> μ	My	<i>Ω</i> ω	Omega

Lothar Sachs
Jürgen Hedderich

Angewandte Statistik

Methodensammlung mit R

Zwölfte, vollständig neu bearbeitete Auflage
mit 142 Abbildungen
und 180 Tabellen

Professor Dr. rer. nat. Lothar Sachs
Seebrooksberg 5
24147 Klausdorf
Deutschland

Dipl. Inform. Jürgen Hedderich
Bimöhler Straße 16
24623 Großenaspe
Deutschland
hedderich@medinfo.uni-kiel.de

Von der 1. bis zur 3. Auflage als
„Statistische Auswertungsmethoden“
1968, 1969 und 1972 erschienen

ISBN-10 3-540-32160-8 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-32160-6 Springer Berlin Heidelberg New York
ISBN 3-540-40555-0 11. Auflage Springer Berlin Heidelberg New York

Bibliografische Information Der Deutschen Bibliothek
Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie;
detaillierte bibliografische Daten sind im Internet über <<http://dnb.ddb.de>> abrufbar.

Dieses Werk ist urheberrechtlich geschützt. Die dadurch begründeten Rechte, insbesondere die der Übersetzung, des Nachdrucks, des Vortrags, der Entnahme von Abbildungen und Tabellen, der Funksendung, der Mikroverfilmung oder der Vervielfältigung auf anderen Wegen und der Speicherung in Datenverarbeitungsanlagen, bleiben, auch bei nur auszugsweiser Verwertung, vorbehalten. Eine Vervielfältigung dieses Werkes oder von Teilen dieses Werkes ist auch im Einzelfall nur in den Grenzen der gesetzlichen Bestimmungen des Urheberrechtsgesetzes der Bundesrepublik Deutschland vom 9. September 1965 in der jeweils geltenden Fassung zulässig. Sie ist grundsätzlich vergütungspflichtig. Zuwiderhandlungen unterliegen den Strafbestimmungen des Urheberrechtsgesetzes.

Springer ist ein Unternehmen von Springer Science+Business Media
springer.de

© Springer-Verlag Berlin Heidelberg 1974, 1978, 1992, 1997, 1999, 2002, 2004, 2006
Printed in Germany

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

Umschlaggestaltung: Erich Kirchner, Heidelberg

SPIN 10984253 42/3153-5 4 3 2 1 0 – Gedruckt auf säurefreiem Papier

Vorwort zur zwölften Auflage

„Viele Forscher machen sich bei der Behandlung der statistischen Beobachtungen die Sache zu leicht. Allerdings kann man in den letzten Jahren einen gewissen Fortschritt wahrnehmen, viele statistische Arbeiten der Neuzeit lassen aber noch viel zu wünschen übrig. Es ist zwar von demjenigen, welcher nur einen gelegentlichen Gebrauch von statistischen Untersuchungen macht, nicht zu erwarten, dass er die Methoden der mathematischen Statistik vollständig beherrscht; jedenfalls kann aber ein jeder ohne Schwierigkeit einen Überblick über viele der wichtigsten Elementargrundsätze gewinnen und dadurch einer Menge von Fehlern und Fehlschlüssen entgehen.“ (Harald Westergaard (1901))

Diese Erkenntnis liegt zwar mehr als 100 Jahre zurück, lange bevor die Grundlagen der Wahrscheinlichkeitsrechnung (A.N. Kolmogoroff) und der modernen Statistik (R.A. Fisher oder J. Neyman und E.S. Pearson) gelegt wurden, ist aber auch heute noch uneingeschränkt gültig. Unter Angewandter Statistik verstehen die Autoren zugleich den Methodenkörper anwendbarer mathematischer Verfahren und die Anwendung dieses Methodenkörpers auf gemessene und/oder gezählte Beobachtungen. Der Schwerpunkt des Buches liegt daher auf Prinzipien der statistischen Denkansätze und auf der Darstellung der Voraussetzungen, die erfüllt sein müssen, bevor man eine bestimmte Formel oder einen bestimmten Test anwenden darf. Berücksichtigt werden insbesondere die Analyse von Stichproben kleiner Umfänge und verteilungsunabhängige Methoden. Angesprochen werden in diesem Lehr- und Nachschlagebuch Nichtmathematiker, insbesondere Praktiker in Technik und Wissenschaft, Ingenieure, Mediziner sowie Studierende und Wissenschaftler dieser und anderer Bereiche. Dem an der praktischen statistischen Arbeit interessierten Mathematiker gibt es einen Überblick.

Für die neue Auflage der „Angewandten Statistik“ war eine Überarbeitung des vor 40 Jahren konzipierten Werkes nicht mehr ausreichend. Schon die letzten Auflagen boten kaum Gelegenheit, die Gliederung und den Inhalt grundlegend zu modifizieren oder zu ergänzen. So konnten nur einige ältere oder überholte Verfahren entfernt werden, um Platz für einiges Neue zu schaffen. Die vorliegende 12. Auflage ist somit ein neues Buch, das der neue Autor (Dipl. Inform. J. Hedderich) in enger Zusammenarbeit mit dem Namengeber (Prof. Dr. rer. nat. L. Sachs) konzipiert und realisiert hat, wobei größere Teile der 11. Auflage übernommen und in einen neuen Kontext gestellt worden sind. Die neue Gliederung in acht Kapiteln erleichtert einerseits den Einstieg in und das Auffinden von statistischen Verfahren. Andererseits wird diese Gliederung auch zukünftigen Auflagen gerecht, wenn es um Neuerungen und Ergänzungen hinsichtlich der statistischen Methodik geht.

Das **1. Kapitel** gibt eine Einführung in die statistische Arbeitsweise bei wissenschaftlichen Fragestellungen. Es verdeutlicht, dass statistische Methoden Kern wissenschaftlicher Erkenntnisprozesse sind. Grundlagen aus der Mathematik, von den Grundrechenarten bis zum Funktionsbegriff und der Kombinatorik, sind im **2. Kapitel** zusammengefasst. Dieses Kapitel wird ergänzt durch eine kurze Einführung in die Matrixalgebra, die hilfreich für ein besseres Verständnis der Verfahren zur Modellbildung im achten Kapitel ist.

Verfahren der deskriptiven Statistik, konsequent gegliedert nach dem Skalenniveau der zu beschreibenden Merkmale, sind im **3. Kapitel** zusammengefasst. Methoden zur Exploration von Daten, insbesondere auch die Erfassung von Abhängigkeiten und Zusammenhängen in den Beob-

achtungen, ermöglichen den Einstieg in eine weiterführende Analyse und Bewertung der Daten.

Der Begriff der Wahrscheinlichkeit, insbesondere im Hinblick auf ein Verständnis von Voraussetzungen und Konsequenzen der Unabhängigkeit von Ereignissen wird ausführlich im **4. Kapitel** mit zahlreichen Beispielen eingeführt. Die Ausführungen zum diagnostischen Test stehen dabei eher beispielhaft für die in der Regel auf bedingten Wahrscheinlichkeiten basierende Terminologie und Argumentationsweise statistischer Verfahren.

Von zentraler Bedeutung bei der Auswahl und Anwendung statistischer Methoden ist nach Ansicht der Autoren der Begriff der Zufallsvariablen, eine Modellvorstellung, die erst eine formale Übertragung der „realen“ Beobachtungen in die Sprache und die numerischen Analyseverfahren der Mathematik ermöglicht. Daher sind im **5. Kapitel** die wichtigsten Verteilungsmodelle zusammengefasst, um neue Modelle ergänzt (z.B. die negative Binomialverteilung und die Weibullverteilung) und mit zahlreichen Beispielen versehen worden. Neu ist hier eine einheitliche Notation zu den Quantilen (kritischen Schranken) spezieller Verteilungen, die für Leser der vorangegangenen Auflagen verwirrend sein könnte. Dabei wird nun einheitlich das obere Quantil einer Verteilung, z.B. 0,95 für „0,05; einseitig“ und 0,975 für „0,05; zweiseitig“, verwendet.

Die neue Auflage der Angewandten Statistik versucht, möglichst klar die Methodenansätze für das „Schätzen“ von Parametern (**6. Kapitel**) und für das „Testen“ von Hypothesen (**7. Kapitel**) zu trennen. Eine eindeutige und stringent eingeführte Notation soll hier einerseits die Brücke zur vertiefenden Lektüre der Spezialliteratur der (theoretischen) Statistik bilden, andererseits sollen Gemeinsamkeiten und Parallelen der verschiedenen Ansätze deutlich werden. Dabei wurden ältere Verfahren präzisiert und neue Verfahren mit zahlreichen Beispielen aufgenommen, z.B. das Bootstrapping, Randomisierungsverfahren und das Prüfen von Äquivalenzaussagen. Weitere Ergänzungen betreffen die Verfahren zur Fallzahlbestimmung (Powerberechnung), die mit dem Programm **R** flexibel eingesetzt werden können. Der Abschnitt zur Analyse von Häufigkeiten wurde um eine ausführliche Darstellung des Kappa-Koeffizienten ergänzt.

Völlig neu ist das **8. Kapitel**. Die Autoren sind überzeugt, dass Methoden zur Bildung und Bewertung von statistischen Modellen heute als zentraler Bestandteil der Angewandten Statistik anzusehen sind. Somit werden die multiple lineare Regression, die logistische Regression, loglineare Modelle und letztlich auch die Analyse von Ereigniszeiten (Überleben) mit Beispielen eingeführt und diskutiert. Diese Verfahren können nicht so elementar und ausführlich dargestellt werden wie die Methoden in den vorangehenden Kapiteln. Dazu gibt es umfangreiche spezielle und vertiefende Monographien. Im Rahmen dieser kurzen Einführung soll zumindest das Verständnis für Verfahren der Modellbildung gefördert und die weitverbreitete Zurückhaltung bei der Anwendung und Interpretation im Rahmen explorativer Datenanalysen abgebaut werden.

Das Verständnis für statistische Methoden erschließt sich letztlich auch aus der selbständigen Analyse (eigener) Daten nach festen Anleitungen und Formeln. Dafür wurden früher Rechenblätter entworfen, mit denen schrittweise durch elementare Berechnungen Ergebnisse hergeleitet und geprüft werden konnten. Ein frühes Hilfsmittel war dabei sicher der Taschenrechner, mit dem diese Arbeit sicherer und schneller zu bewerkstelligen war. Seit den 70iger Jahren des vergangenen Jahrhunderts ist die Entwicklung von kommerziellen Statistik-Programmpaketen, genannt seien hier nur SPSS und SAS, weit voran geschritten. Diese stellen „vorkonfektionierte“ Lösungen bereit, die von dem Anwender häufig nur schwer nachzuvollziehen sind. Mit dem kostenlosen Programm **R** steht ein Werkzeug zur Verfügung, mit dem einerseits elementare Berechnungen einfach durchgeführt werden können, andererseits auch komplexe statistische Verfahren und Modelle aus festen Paketen genutzt werden können. Daher wurden viele Beispiele in dieser Auflage mit **R** berechnet und zahlreiche erklärende Abbildungen mit **R** neu erstellt. Die dafür verwendeten Befehle sind im

Internet auf der Produktseite des Buches (**Download**) beim Springer-Verlag abrufbar und können parallel zur Lektüre des Buches modifiziert und ergänzt werden. Einen Einstieg in die Verwendung von **R** bietet das **9. Kapitel**. Wichtige Befehle sind in einer Übersicht (**Lesezeichen**) am Ende des Buches zusammengefasst. Die Autoren sind überzeugt, dass sich hieraus ein besseres Verständnis der statistischen Methodik ohne die häufig abschreckende Rechenarbeit entwickeln und die Statistik mehr Freunde finden kann.

Um die 12. Auflage zu entlasten, ist auf Teile des Textes und auf die Übernahme der sehr ausführlichen Bibliographie älterer Auflagen verzichtet worden. Die neue Bibliographie und das neue Sachverzeichnis sind an den Schwerpunkten der neuen Auflage orientiert und müssen sich unter der geänderten Ausrichtung erst entwickeln.

Unser Dank gilt den Kolleginnen am Institut für Medizinische Informatik und Statistik der Christian-Albrechts-Universität Kiel (Direktor Prof. Dr. rer. nat. M. Krawczak), Frau Dr. A. Caliebe und Frau Dipl. Math. U. Schulz, für zahlreiche Anregungen und die kritische Durchsicht von Teilen des Manuskripts. Herrn Dipl. Inform. O. Junge danken wir für die Hilfestellung bei technischen Problemen mit $\text{\LaTeX}$, die insbesondere durch den Übergang von der 11. zur 12. Auflage aufgetreten sind. Unser Dank gilt auch den Damen und Herren der Kieler Universitätsbibliothek, vor allen Dingen Herrn Dr. J. Aschenbach.

Am Schluss ist es uns eine angenehme Pflicht, zahlreichen Lesern früherer Auflagen zu danken, die durch ihre kritischen Anmerkungen manches Versehen auszumerzen halfen. Den Damen und Herren des Springer Verlages, insbesondere Herrn C. Heine, Frau L. Braun und Frau R. Milewski danken wir für die angenehme Zusammenarbeit. Trotz einer sorgfältigen Bearbeitung von Texten, Formeln und Beispielen lassen sich Fehler und Unklarheiten nicht ausschließen. Wir bitten den Leser, uns diese mitzuteilen (schriftlich an die Adresse der Autoren oder auch per E-mail an j.hedderich@t-online.de). Auch für Verbesserungsvorschläge sind wir dankbar. Hoffentlich wenige Korrekturen werden aktuell über die Produktseite des Buches beim Springer-Verlag (**Errata**) im Internet angegeben.

Kiel, März 2006

J. Hedderich

Lothar Sachs

Übersetzungen älterer Auflagen liegen vor:

- ins Russische (1976): ohne ISBN Nummer, der vergleichbare sowjetische Code 3[(10805* – 146)/(008(01) – 76)][115 – 76], *BTOROI INDEKS-10803, CTATISTIKA, MOSKBA;
- ins Spanische (1978): ISBN 84-335-6412-9, Editorial Labor, S.A., Barcelona;
- ins Amerikanische (1984): ISBN 0-387-90976-1, Springer, New York.

Vorwort zur zehnten Auflage**Ziele älterer Auflagen, die auch für diese Neubearbeitung gelten**

Das Buch wendet sich an Interessierte, die ich im Einzelnen in meinen Vorworten zur 1., 7. bis 9. Auflage (vgl. S. VI–XI) charakterisiert habe und die „etwas“ suchen, das dem LERNEN dient, die Grundlagen vermittelnd, einführend und vertiefend, auch anhand vieler durchgerechneter Beispiele, dem ANWENDEN mit zahlreichen Planungs- und Auswertungsempfehlungen aus der Praxis und dem NACHSCHLAGEN, um einen Überblick über ein weitgefasstes Methodenspektrum zu gewinnen. Allen drei Zielen dient neben den Querverweisen und den weiterführenden Literatur-Hinweisen insbesondere das zum Nachschlagen und Wiederfinden durchstrukturierte sehr ausführliche Sachverzeichnis.

Kurz nach der 9. folgt jetzt die neu gesetzte und damit lesbarere 10. Auflage, weitreichend überarbeitet und aktualisiert. Im Text wurden Unstimmigkeiten und Druckfehler beseitigt, Präzisierungen vorgenommen, zahlreiche Ergänzungen und Hinweise sowie weitere Web-Sites aufgenommen. Manche Anregungen kamen von ehemaligen Teilnehmern an meinen Oberseminaren, einige aufgrund von Leserbriefen, herzlichen Dank! Auch das Sachverzeichnis und die Literatur habe ich auf den neuesten Stand gebracht, wobei dem Leser, der sich intensiver mit der Statistik beschäftigen möchte, auf S. 690 ein eleganter Weg aufgezeigt wird. Andere folgen hier weiter unten sowie auf S. XXXVI. Herrn Prof. Dr. Carsten Stick, Direktor des Instituts für Medizinische Klimatologie der Universität Kiel, danke ich für eine Liste häufiger Fehler in Dissertationsschriften (vgl. S. XXXVII).

Mein Dank gilt auch wieder den Damen und Herren der Kieler Universitätsbibliothek, vor allem Herrn Dr. Jürgen Aschenbach. Den Damen und Herren des Springer-Verlages danke ich für die ausgezeichnete Zusammenarbeit. Für Leserzuschriften bin ich weiterhin dankbar, insbesondere für jeden Verbesserungsvorschlag.

Klausdorf, im Januar 2002

Lothar Sachs

Vorwort zur achten Auflage

Auch die 8., völlig neu bearbeitete und erweiterte Auflage dient zum Lernen, Anwenden und Nachschlagen für anwendungsorientierte Leser mit unterschiedlichen Vorkenntnissen und breit gestreuten Interessen. Es ist ein ausführlich gefasstes Lehrbuch und Nachschlagewerk, das dem Anfänger anhand zahlreicher Arbeitshilfen und vertiefender Wiederholungen, unterschiedlich akzentuiert, den Einstieg in die Anwendung statistischer Methoden ermöglicht und ihn unterstützt. Dem Fortgeschrittenen bietet es eine Fülle von Hinweisen und Berechnungsmethoden zu weiteren wichtigen, speziellen Verfahren der Statistik. Hierzu dienen auch die wesentlich erweiterten drei Verzeichnisse: das Literaturverzeichnis, das Namenverzeichnis und das Sachverzeichnis. Es ergänzt daher auch jedes Statistik-Software-Handbuch. Angesprochen werden in erster Linie Studenten und Praktiker aus den Bereichen der Naturwissenschaften, der Medizin und der Technik. Es eignet sich aber auch für Interessierte und Wissenschaftler anderer Disziplinen, die sich um Erkenntnisgewinnung durch statistische Ansätze bemühen und die hier Hinweise und Details zur Planung und Auswertung von Untersuchungen erhalten. Die Neubearbeitung habe ich zunächst auf Formulierungs-, Formel- und Druckfehler durchgesehen, wobei mir aufmerksame Leser Hinweise gegeben haben, für die ich herzlich danke. Weiter habe ich Anfragen von Lesern, Fachkollegen und Teilnehmern an meinen Oberseminaren berücksichtigt, denen ich ebenfalls herzlich danke. Da

jetzt auf den Informationsstatistik-Ansatz nach Woolf und Kullback verzichtet werden kann, waren die Seiten 456/465 und 608/611 wieder frei verfügbar. Außerdem ist ein kleiner Anhang hinzugekommen. Generell habe ich zahlreiche Textstellen neu formuliert, Aussagen präzisiert und vieles ergänzt: Anwendungsschwerpunkte, Methoden, Formeln, Tabellen, Übersichten, Beispiele, Kommentare, Querverweise sowie Warnungen und Empfehlungen für die praktische Arbeit. Wichtige Abschnitte habe ich auch in dieser Auflage weitgehend „autark“ belassen und eine Wiederholung nicht gescheut. Bevor ein bestimmtes Verfahren angewandt wird, ist ein Blick auf zugehörige Hinweise und Querverweise unerlässlich. Bewusst einfach gehaltene Beispiele bieten sich an, sie zur Übung in gering modifizierter Form durchzurechnen, etwa indem ein Messwert variiert wird, so dass sich das erwartete Resultat abschätzen lässt. Die zahlreichen Ergänzungen hat zwar die Informationsdichte erhöht, die Seitenzahl des Textes konnte jedoch konstant bleiben. Manches Interessante ist jetzt als Kleingedrucktes etwas stiefmütterlich behandelt worden. Deutlich erweitert und vertieft habe ich das zum Nachschlagen und Wiederfinden besonders wichtige strukturierte Sachverzeichnis mit Übersichtscharakter sowie die nicht nur für den Praktiker unentbehrlichen Literaturangaben. Erstaunlich schnell gelangt man hier in unwegsames Gelände, was auch für andere reizvolle Fachgebiete gilt, sobald man ausgetretene Pfade verlässt. Den Damen und Herren des Springer-Verlages danke ich herzlich für die ausgezeichnete Zusammenarbeit. Für Leserzuschriften bin ich dankbar, insbesondere für jeden Verbesserungsvorschlag.

Klausdorf, im Herbst 1996

Lothar Sachs

Vorwort zur siebenten Auflage

Auch die 7., völlig neu bearbeitete Auflage mit wesentlich mehr mathematisch-statistischen Tabellen, Übersichten, Formeln und vollständig durchgerechneten Zahlenbeispielen dient zum LERNEN, daher die für das Selbststudium unerlässlichen vertiefenden Wiederholungen mit bewusst unterschiedlicher Akzentsetzung, zum ANWENDEN statistischer Verfahren in der praktischen Arbeit, daher der Handbuch-Charakter, und zum NACHSCHLAGEN, um genau das aufzuspueren, was dem Suchenden weiterhilft. Aus diesen Gründen war ein völlig neu bearbeitetes ausführliches Literaturverzeichnis notwendig. Hierzu dienen neben den 94 meist neuen Übersichten vier völlig neu bearbeitete ausführliche Verzeichnisse: das Inhaltsverzeichnis (20 Seiten), das Literaturverzeichnis (51 S.), das Namenverzeichnis (14 S.) und das Sachverzeichnis (79 S.).

Statistische Programmpakete sind weit verbreitet. So konnte manches weggelassen werden. Dafür habe ich mehr zur Planung einer Untersuchung ausgeführt, Zusammenhänge und Verweise stärker aktualisiert, die Zahl der Hinweise, Übersichten, Tabellen, Formeln und insbesondere der Beispiele deutlich vermehrt sowie zahlreiche Gebiete ausführlicher behandelt (z.B. die Kombinatorik) und neue Methoden (z.B. den Jonckheere Test) aufgenommen. Auf das rapide anwachsende und interessante Gebiet der multivariaten Statistik, das die im Buch behandelten Themen wesentlich ergänzt, habe ich an einigen Stellen hingewiesen und weiterführende Monographien genannt. Da sich Wahrscheinlichkeitsrechnung und Kombinatorik mit interessanten Beispielen schmücken lassen, die weiterführende Ansätze enthalten, sind diese Beispiele im ersten Kapitel von B1 bis B172 durchnummeriert worden, so dass sich in späteren Kapiteln leicht auf sie zurückkommen lässt. Auch einige Bemerkungen zu Simulationen sind mit anderen Hinweisen in das 1. Kapitel integriert worden. Kapitel 2 enthält jetzt allgemein interessierende Bemerkungen zu epidemiologischen und ähnlichen Studien sowie drei vielseitig verwendbare geschlossene Folgetestpläne. Die restlichen fünf Kapitel sind ebenfalls neu bearbeitet worden. Details bietet das völlig neu und sehr ausführlich angelegte Inhaltsverzeichnis, das durch die Übersichten ergänzt wird. Teilweise gestaffelte schlagwortartige Untertitel zu den einzelnen Abschnitten erleichtern die Übersicht; das Thema selbst wird im Untertitel nur selten gegliedert oder noch einmal genannt.

Wiederholungen waren u.a. dort nicht zu vermeiden, wo wichtige Abschnitte weitgehend „autark“ sein sollten; zusätzliche Querverweise sollte der Leser beachten, bevor ein bestimmtes Verfahren angewandt wird. Viele Beispiele sind bewußt einfach gehalten. Sie sollten zur Übung in gering modifizierter Form durchgerechnet werden, etwa einen Messwert variieren, so dass sich das erwartete Resultat abschätzen lässt.

Wer tiefer in die statistische Methodik eindringen möchte, wird den im Literaturverzeichnis angeführten Arbeiten wesentlich mehr entnehmen als die knappen Hinweise im Text ahnen lassen. Erstaunlich schnell gelangt man hier in unwegsames Gelände, was auch für andere reizvolle Fachgebiete gilt, sobald man die ausgetretenen Pfade verlässt.

Der Biometric Society danke ich für die Erlaubnis aus der Arbeit von J.K. Haseman: Exact sample sizes for use with the Fisher-Irwin Test for 2×2 tables. *Biometrics* **34** (1978), 106–109 Tables 1 + 2, pages 107 und 108 übernehmen zu dürfen. Mein Dank gilt auch wieder den Damen und Herren der Kieler Universitätsbibliothek, insbesondere Frau Dr. Gudrun Otto und Herrn Dr. Jürgen Aschenbach.

In einem losen Zusammenhang mit dieser Neubearbeitung steht mein Oberseminar, das von der Abteilung, insbesondere von ihrem Direktor, Herrn Prof. Dr.-Ing. K. Sauter, stets nachhaltig gefördert worden ist. Herrn Prof. Sauter sowie Frau Katrin Anger und Frau Petra Neumann, die meine Kartei geführt und Entwürfe für das Oberseminar geschrieben haben, sei herzlich gedankt. Den Damen und Herren des Springer-Verlages danke ich für die ausgezeichnete Zusammenarbeit. Für Leserzuschriften bin ich dankbar, insbesondere für jeden Verbesserungsvorschlag.

Klausdorf, im Januar 1992

Lothar Sachs

Vorwort zur ersten Auflage

„Das kann kein Zufall sein“, sagte sich im Jahre 1710 der Arzt der Königin Anne, John Arbuthnot (1667–1735), Wissenschaftler und Satiriker (er erfand „John Bull“), Freund und Mitarbeiter von Jonathan Swift, Alexander Pope und John Gay, außerordentlich geschätzt von Dr. Samuel Johnson, als er in den Geburtsregistern von 82 Jahrgängen (1629–1710) ausnahmslos die Knabengeburten häufiger vertreten fand als die Mädchengeburten. Dieser Stichprobenumfang bot ihm eine ausreichende Sicherheit für seinen Schluss. Er konnte hinter die Zahl der Knabengeburten jedesmal ein Pluszeichen setzen (größer als die Anzahl der Mädchengeburten), und schuf so den Vorzeichen-test. Bei großen Stichproben genügt Zweidrittelmehrheit des einen Vorzeichens. Bei kleinen Stichproben ist eine 4/5- oder sogar eine 9/10-Mehrheit für den Nachweis eines verlässlichen Stichprobenunterschiedes notwendig.

Charakteristisch für unsere Zeit ist die stürmische Entwicklung von Wahrscheinlichkeitsrechnung, mathematischer Statistik und ihrer Anwendungen in Wissenschaft, Technik, Wirtschaft und Politik.

Dieses Buch ist auf Anregung von Herrn Prof. Dr. H.-J. Staemmler, jetzt Chefarzt der Städtischen Frauenklinik in Ludwigshafen am Rhein, geschrieben worden. Ihm bin ich für die geleistete vielfältige Unterstützung zu großem Dank verpflichtet!

Bei der Beschaffung von Literatur waren mir Herr Prof. Dr. W. Wetzel, Direktor des Seminars für Statistik der Universität Kiel, jetzt Direktor des Institutes für angewandte Statistik der F.U. Berlin, Frau Brunhilde Memmer, Bibliothek des Wirtschaftswissenschaftlichen Seminars der Universität Kiel, Herr Priv. Doz. Dr. E. Weber, Landwirtschaftliche Fakultät der Universität Kiel, Variationsstatistik, sowie die Herren Dr. J. Neumann und Dr. M. Reichel von der hiesigen Universitäts-Bibliothek behilflich. Nicht unerwähnt lassen möchte ich die wertvolle Mitarbeit bei der Abfassung des Manuskriptes, insbesondere durch Frau W. Schröder, Kiel, durch Fräulein Christa

Diercks, Kiel, und durch den medizinisch-technischen Assistenten Herrn F. Niklewicz, Kiel, dem ich die Anfertigung der graphischen Darstellungen verdanke.

Herrn Prof. Dr. S. Koller, Direktor des Institutes für Medizinische Statistik und Dokumentation der Universität Mainz und besonders Herrn Prof. Dr. E. Walter, Direktor des Institutes für Medizinische Statistik und Dokumentation der Universität Freiburg i. Br. verdanke ich viele wertvolle Anregungen.

Beim Lesen der Korrekturen haben mich die Herren Dipl. Math. J. Schimmler und Oberstudienrat Dr. K. Fuchs unterstützt. Ihnen sei herzlich gedankt!

Weiter danke ich den zahlreichen Autoren, Herausgebern und Verlagen, die den Abdruck der Tafeln und Abbildungen ohne Vorbehalt gestattet haben.

Zu Dank verpflichtet bin ich insbesondere dem literarischen Vollstrecker des verstorbenen Sir Ronald A. Fisher, F.R.S., Cambridge, Herrn Prof. Frank Yates, Rothamsted und den Herren der Oliver und Boyd Ltd., Edinburgh, für die Erlaubnis, Tafel II 1, Tafel III, Tafel IV, Tafel V und Tafel VII 1 ihres Buches „Statistical Tables for Biological, Agricultural and Medical Research“ zu reproduzieren; Herrn Prof. O.L. Davies, Alderley Park, und den Herren des Verlages von Oliver und Boyd Ltd., Edinburgh, für die Erlaubnis, einen Teil der Tafel *H* aus dem Buch „The Design and Analysis of Industrial Experiments“ von O.L. Davies übernehmen zu dürfen; den Herren des Verlages C. Griffin and Co. Ltd., London, sowie ihren Autoren, den Herren Prof. M.G. Kendall und Prof. M.H. Quenouille, für die Erlaubnis, aus dem Buch von Kendall und Stuart „The Advanced Theory of Statistics“, Vol. II, die Tafeln 4a und 4b, aus dem Büchlein von Quenouille „Rapid Statistical Calculations“, die Abbildungen auf den Seiten 28 und 29 sowie Tafel 6 reproduzieren zu dürfen; den Herren Prof. E.S. Pearson und H.O. Hartley, Herausgeber der „Biometrika Tables for Statisticians“, Vol. 1, 2nd ed., Cambridge 1958, für die Erlaubnis, Kurzfassungen der Tafeln 18, 24 und 31 übernehmen zu dürfen. Mein Dank gilt weiter Mrs. Marjorie Mitchell, der McGrawHill Book Company, New York, und Herrn Prof. W.J. Dixon für die Erlaubnis, aus dem Buch von W.J. Dixon und F.J. Massey Jr.: „Introduction to Statistical Analysis“ Tafel A-12 c und Tafel A-29 reproduzieren zu dürfen (Copyright vom 13. April 1965, 1. März 1966 und 21. April 1966) sowie Herrn Prof. C. Eisenhart für die Genehmigung, aus „Techniques of Statistical Analysis“, herausgegeben von C. Eisenhart, M.W. Hastay und W.A. Wallis, die Tafel der Toleranzfaktoren für die Normalverteilung entnehmen zu dürfen. Herrn Prof. F. Wilcoxon, Lederle Laboratories, a Division of American Cyanamid Company, Pearl River, danke ich für die Erlaubnis, aus „Some Rapid Approximate Statistical Procedures“ von F. Wilcoxon und Roberta A. Wilcox, die Tafeln 2, 3 und 5 zu reproduzieren. Herrn Prof. W. Wetzel, Berlin-Dahlem, und den Herren des de Gruyter-Verlages, Berlin W 35, danke ich für die Erlaubnis, aus den Elementaren Statistischen Tabellen von W. Wetzel die Tafel auf S. 31 übernehmen zu dürfen. Besonderen Dank schulde ich Herrn Prof. Dr. K. Diem, Redaktion des Documenta Geigy, Basel, für die freundliche Überlassung einer verbesserten Tafel der oberen Signifikanzschranken des studentisierten Extrembereiches, die für die 7. Auflage der „Wissenschaftlichen Tabellen“ vorgesehen ist.

Den Herren des Springer-Verlages danke ich für die sehr erfreuliche Zusammenarbeit.

Kiel, November 1967

Lothar Sachs

Inhaltsverzeichnis

1	Einführung	1
1.1	Definition und Aufgaben der Statistik	1
1.2	Wissenschaftliche Arbeitstechnik	3
1.2.1	Daten und Modelle	3
1.2.2	Kreisprozesse	4
1.2.3	Modelle in der Statistik	6
1.3	Statistik und wissenschaftliche Methode	7
1.3.1	Wiederholbare Erfahrungen	7
1.3.2	Deskriptive Statistik	8
1.3.3	Explorativer Ansatz	9
1.3.4	Konfirmativer Ansatz	10
1.3.5	Merkmale, Grundgesamtheit, Stichprobe	11
1.3.6	Stichproben	12
1.3.7	Zufallsstichproben	13
1.4	Datenanalyse	14
1.4.1	Klassierung von Merkmalen	14
1.4.2	Skalierung von Variablen	15
1.4.3	Daten	17
2	Grundlagen aus der Mathematik	20
2.1	Logische und relationale Operatoren	20
2.2	Mengen	21
2.2.1	Begriffsbildung	21
2.2.2	Mengenoperationen	22
2.3	(Grund-) Rechenarten	23
2.3.1	Summen und Produkte	24
2.3.2	Potenzen und Wurzeln	29
2.3.3	Logarithmen	30
2.3.4	Rundungen	32
2.3.5	Rechnen mit fehlerbehafteten Zahlen	33
2.4	Einführung in die Matrixalgebra	34
2.4.1	Definition und Schreibweise	34
2.4.2	Matrixoperationen	35
2.4.3	Determinanten	39
2.4.4	Die Inverse Matrix	39
2.4.5	Lineare Abhängigkeit, Rang einer Matrix	40
2.4.6	Lineare Gleichungssysteme	41

2.4.7	Eigenwerte und Eigenvektoren	41
2.5	Funktionen	42
2.5.1	Lineare Funktionen	43
2.5.2	Nichtlineare Funktionen	44
2.5.3	Periodische Funktionen	45
2.5.4	Exponentialfunktion und logarithmische Funktion	46
2.5.5	Flächen unter einer Funktion - Integrale	46
2.6	Kombinatorik	47
2.6.1	Permutationen	47
2.6.2	Kombinationen - der Binomialkoeffizient	49
2.6.3	Kombinationen mit Wiederholungen und mit Berücksichtigung der Anordnung	50
2.6.4	Zerlegung einer Menge	52
2.6.5	Das Pascalsche Dreieck	52
2.6.6	Der Multinomialkoeffizient	54
3	Deskriptive Statistik	55
3.1	Häufigkeiten	56
3.1.1	Absolute und relative Häufigkeiten	56
3.1.2	Sinnvolle Quotienten: Verhältniszahlen	57
3.1.3	Prozentwerte	59
3.1.4	Torten- und Balkendiagramme	59
3.1.5	Tabellen	60
3.1.6	Bedingte Häufigkeiten	62
3.2	Beschreibung von Ordinaldaten	62
3.2.1	Medianwert und andere Quartile	63
3.2.2	Quantile	64
3.2.3	Streuung ordinal skalierten Daten	65
3.2.4	Punktdiagramm und Box-Plot	66
3.2.5	Korrelationskoeffizient nach Kendall	67
3.3	Beschreibung von metrischen Daten	68
3.3.1	Arithmetischer Mittelwert	68
3.3.2	Standardabweichung, Varianz	70
3.3.3	Variationskoeffizient	72
3.3.4	Der $(\bar{x} \pm s)$ -Bereich	73
3.3.5	Klassierte Messwerte; Berechnung des Mittelwertes und der Standardabweichung	74
3.3.6	Das gewogene arithmetische Mittel, die gewogene Varianz und das gewichtete arithmetische Mittel	74
3.3.7	Geometrischer Mittelwert	76
3.3.8	Harmonischer Mittelwert	78
3.4	Häufigkeitsverteilung	80
3.4.1	Histogramm	80
3.4.2	Stamm-Blatt Darstellung	83
3.5	Konzentration; Gini Index	83
3.6	Maßzahlen für den Zusammenhang metrischer Daten	85
3.6.1	Punktwolken	85
3.6.2	Die empirische Kovarianz	85
3.6.3	Der empirische Korrelationskoeffizient	87
3.6.4	Der Rangkorrelationskoeffizient	88
3.6.5	Typisierung korrelativer Zusammenhänge	90
3.6.6	Die lineare Regression	91

3.6.7	Spezielle Schätzungen der Regressionsgeraden	93
3.6.8	Robuste lineare Regression	97
3.7	Nichtlineare Regression	99
3.7.1	Einige linearisierende Transformationen	104
4	Wahrscheinlichkeiten	108
4.1	Zufallsexperiment, Ereignis	109
4.2	Begriff der Wahrscheinlichkeit	111
4.2.1	Definition nach Laplace	111
4.2.2	Axiome nach Kolmogoroff	113
4.3	Bedingte Wahrscheinlichkeiten, stochastische Unabhängigkeit	116
4.3.1	Bedingte Wahrscheinlichkeit	116
4.3.2	Stochastische Unabhängigkeit	120
4.4	Bayessches Theorem	128
4.4.1	Bayessches Theorem und Pfadregel	129
4.4.2	Acht Beispiele zum Bayesschen Theorem	130
4.5	Der diagnostische Test	132
4.5.1	ROC - Analyse	136
4.5.2	Der Likelihoodquotient	137
4.6	Maßzahlen in der Epidemiologie	139
4.6.1	Prävalenz und Inzidenz	139
4.6.2	Standardisierungen	141
5	Zufallsvariablen, Verteilungen	144
5.1	Die Zufallsvariable	144
5.1.1	Wahrscheinlichkeitsfunktion, Wahrscheinlichkeitsdichte und Verteilungsfunktion	145
5.2	Maßzahlen zur Kennzeichnung der Verteilung	150
5.2.1	Erwartungswert	151
5.2.2	Varianz	152
5.2.3	Momente: Schiefe und Exzess	154
5.3	Diskrete Verteilungen	161
5.3.1	Das Urnenmodell	161
5.3.2	Gleichverteilung	163
5.3.3	Binomialverteilung	164
5.3.4	Poisson-Verteilung	173
5.3.5	Negative Binomial-Verteilung	181
5.3.6	Hypergeometrische Verteilung	186
5.4	Stetige Verteilungen	190
5.4.1	Gleichverteilung	190
5.4.2	Normalverteilung	191
5.4.3	Lognormalverteilung	204
5.4.4	Exponentialverteilung	207
5.4.5	Weibull-Verteilung	209
5.5	Testverteilungen	210
5.5.1	Student-Verteilung (t)	211
5.5.2	Chiquadrat-Verteilung (χ^2)	215
5.5.3	Fisher-Verteilung (F)	216
5.5.4	Verteilungen wichtiger Stichprobenfunktionen aus normalverteilten Grundgesamtheiten	222
5.6	Verteilung zweidimensionaler Zufallsvariablen	224
5.6.1	Modellbildung	224

5.6.2	Randverteilungen und Unabhängigkeit	226
5.6.3	Korrelationskoeffizient	230
5.6.4	Zweidimensionale Normalverteilung	231
5.6.5	Multinomialverteilung (Polynomialverteilung)	233
6	Schätzen	235
6.1	Zufallsstichproben und Zufallszahlen	235
6.1.1	Spezielle Stichprobenverfahren	238
6.2	Das Schätzen von Parametern	239
6.2.1	Vorbemerkungen	240
6.2.2	Wünschenswerte Eigenschaften von Schätzfunktionen	241
6.2.3	Gesetz der großen Zahlen	243
6.2.4	Der mittlere quadratische Fehler	243
6.3	Schätzverfahren für Maßzahlen einer Verteilung	245
6.3.1	Momentenmethode	245
6.3.2	Schätzung nach der größten Erwartung (MLE)	246
6.3.3	Kleinsten Fehler (OLS)	251
6.4	Konfidenzintervalle	252
6.5	Konfidenzintervall für einen Anteilswert aus einer dichotomen Grundgesamtheit (π)	254
6.5.1	Approximation durch die Normalverteilung	256
6.5.2	Sonderfälle mit $\hat{p} = 0$ bzw. $\hat{p} = 1$	258
6.5.3	Schnellschätzung der Vertrauensgrenzen anhand einer beobachteten relativen Häufigkeit nach Clopper und Pearson	259
6.5.4	Angenähertes 95%-Konfidenzintervall für $\pi_1 - \pi_2$ (n_1 und n_2 groß)	261
6.5.5	Schätzung des Mindestumfangs einer Stichprobe bei ausgezählten Werten	262
6.6	Konfidenzintervalle für μ bei Normalverteilung	263
6.6.1	Vertrauenswahrscheinlichkeit und Irrtumswahrscheinlichkeit	263
6.6.2	Konfidenzintervall für den Erwartungswert μ	265
6.6.3	Konfidenzintervall für die Differenz $\mu_1 - \mu_2$	267
6.6.4	Das Konfidenzintervall für den Erwartungswert μ_d der Paardifferenzen	269
6.6.5	Konfidenzintervall für das Verhältnis μ_1/μ_2	269
6.6.6	Mindestzahl von Beobachtungen zur Schätzung eines Mittelwertes	271
6.7	Konfidenzintervall für die mittlere absolute Abweichung	271
6.8	Konfidenzintervall für den Median	273
6.8.1	Angenäherte verteilungsunabhängige Konfidenzintervalle für beliebige Quantile	274
6.9	Konfidenzintervalle nach dem Bootstrap-Verfahren	275
6.10	Konfidenzintervall für σ^2 bzw. σ	278
6.10.1	Konfidenzintervall für den Variationskoeffizienten γ	279
6.10.2	Konfidenzintervall für den Quotienten zweier Varianzen σ_1^2/σ_2^2	279
6.10.3	Mindestzahl von Beobachtungen zur Schätzung einer Standardabweichung	280
6.11	Konfidenzintervall für den Erwartungswert λ einer Poisson-Verteilung	280
6.12	Weibull-Verteilung	284
6.12.1	Bestimmung der Parameter	284
6.12.2	Das Konfidenzintervall für die Weibull-Gerade	285
6.13	Konfidenzintervalle für die Parameter einer linearen Regression	286
6.13.1	Die Schätzung einiger Standardabweichungen	286
6.13.2	Konfidenzintervalle für den Regressionskoeffizienten, für den Achsenabschnitt und für die Restvarianz	291
6.13.3	Konfidenzintervalle und Prädiktionsintervalle für die Regressionsgerade	292
6.13.4	Inverse Prädiktion aus einer linearen Regression	296

6.13.5	Das Konfidenzintervall für den Korrelationskoeffizienten ρ	297
6.14	Toleranzgrenzen	299
6.14.1	Verteilungsunabhängige Toleranzgrenzen	301
6.15	Übereinstimmung von Messwerten nach Bland-Altman	302
7	Hypothesentest	305
7.1	Der statistische Test	305
7.1.1	Entscheidungsprinzipien	306
7.1.2	Statistische Hypothesen und Testentscheidungen	307
7.1.3	Statistischer Test - Schritt für Schritt	310
7.1.4	Powerfunktion und Operationscharakteristik	314
7.1.5	Die Operationscharakteristik	318
7.1.6	Die Formulierung von Hypothesen	321
7.1.7	Der P-Wert nach R.A. Fisher	322
7.1.8	Äquivalenztests	324
7.1.9	Verteilungsunabhängige Verfahren	325
7.2	Tests der Verteilung (goodness of fit)	327
7.2.1	Der Quotient R/s	327
7.2.2	Überprüfung des 3. und 4. Momentes	328
7.2.3	Das Wahrscheinlichkeitsnetz, QQ-Plot	330
7.2.4	Der Chiquadrat-Anpassungstest	333
7.2.5	Kolmogoroff-Smirnoff-Anpassungstest	337
7.2.6	Shapiro-Wilk Test	341
7.2.7	Anderson-Darling Test	342
7.2.8	Ausreißerproblem	343
7.3	Einstichprobenverfahren	347
7.3.1	Hypothesen zu Wahrscheinlichkeiten	347
7.3.2	Hypothesen zu Erwartungswerten, die sich auf einen empirischen Mittelwert beziehen	352
7.3.3	Einstichproben-Median-Test	358
7.3.4	Vergleich einer empirischen Varianz mit ihrem Parameter	359
7.3.5	Prüfung der Zufällsmäßigkeit einer Folge von Alternativdaten oder von Messwerten	360
7.3.6	Prüfung der Erwartungswerte von Poisson-Verteilungen	366
7.4	Zweistichprobenverfahren	367
7.4.1	Vergleich zweier Varianzen (F-Test)	367
7.4.2	Rangdispersionstest von Siegel und Tukey	371
7.4.3	Ansari-Bradley-Test	375
7.4.4	t-Test für unabhängige Stichproben	377
7.4.5	t-Test für Paardifferenzen	387
7.4.6	Wilcoxon Rangsummentest für zwei unabhängige Stichproben	391
7.4.7	Wilcoxon-Paardifferenzentest	400
7.4.8	Vergleich zweier unabhängiger Stichproben nach Kolmogoroff und Smirnoff	405
7.4.9	Cramér-von Mises Test	408
7.4.10	Einige weitere verteilungsunabhängige Verfahren für den Vergleich unabhängiger Stichproben	410
7.4.11	Zweistichproben-test auf Äquivalenz	414
7.5	Mehrstichprobenverfahren, varianzanalytische Methoden	418
7.5.1	Prüfung der Gleichheit mehrerer Varianzen	418
7.5.2	Einfache Varianzanalyse	423
7.5.3	Multiple Vergleiche, Multiples Testproblem	428

7.5.4	H-Test von Kruskal und Wallis	442
7.5.5	Varianzanalyse für Messwiederholungen (Blockvarianzanalyse)	454
7.5.6	Friedman-Test	456
7.5.7	Zweifache Varianzanalyse	465
7.5.8	Prinzipien der Versuchsplanung	470
7.6	Die Analyse von Häufigkeiten	477
7.6.1	Vergleich zweier relativer Häufigkeiten	477
7.6.2	Die Analyse von Vierfeldertafeln	479
7.6.3	Odds Ratio und relatives Risiko	487
7.6.4	Exakter Fisher-Test	496
7.6.5	Der von McNemar modifizierte Vorzeichentest	497
7.6.6	Test nach Mantel-Haenszel	503
7.6.7	Der $k \cdot 2$ -Felder- χ^2 -Test nach Brandt und Snedecor	507
7.6.8	Cochran-Armitage Test auf linearen Trend	516
7.6.9	Die Analyse von Zweiwegtafeln des Typs $r \cdot c$	519
7.6.10	Bowker-Test auf Symmetrie in quadratischen Mehrfeldertafeln	535
7.6.11	Cohen's Kappa-Koeffizient κ	537
7.7	Hypothesentests zur Korrelation und Regression	543
7.7.1	Prüfung des Vorhandenseins einer Korrelation	544
7.7.2	z -Transformation nach R.A. Fisher	548
7.7.3	Weitere Anwendungen der z -Transformation	549
7.7.4	Der Vergleich mehrerer Korrelationskoeffizienten	551
7.7.5	Prüfung der Linearität einer Regression	552
7.7.6	Prüfung der Regressionsparameter	553
7.7.7	Prüfung des Rang-Korrelationskoeffizienten ρ_S	557
8	Statistische Modellbildung	560
8.1	Einführung	560
8.2	Regressionsmodelle	562
8.2.1	Die einfache lineare Regression	562
8.2.2	Die multiple lineare Regression	566
8.2.3	Verfahren der Variablenauswahl	573
8.2.4	Nominalskalierte Einflussgrößen	576
8.3	Varianzanalyse im linearen Modell	577
8.3.1	Einfaktorielle Varianzanalyse	577
8.3.2	Zweifaktorielle Varianzanalyse	581
8.4	Logistische Regression	585
8.4.1	Hypothesentest im logistischen Regressionsmodell	589
8.4.2	Multiple logistische Regression	591
8.4.3	Interpretation der Regressionskoeffizienten (odds)	594
8.4.4	Variablenauswahl im Rahmen der Modellbildung	595
8.4.5	Residuenanalyse	597
8.5	Log-lineare Modelle	598
8.5.1	Kontingenztafeln	598
8.5.2	Log-lineares Modell am Beispiel von 2 Faktoren	602
8.5.3	Drei-dimensionale Kontingenztafeln	604
8.6	Analyse von Überlebenszeiten	609
8.6.1	Kaplan-Meier Schätzung der Überlebensfunktion	611
8.6.2	Der Logrank-Test	616
8.6.3	Parametrische Modelle für Überlebenszeiten	618
8.6.4	Das Cox-Regressionsmodell	621

9	Einführung in R	633
9.1	Das Konsolfenster	633
9.2	Objekte in R	636
9.3	Hilfestellung in R	637
9.4	Erzeugen von Daten in R mittels Funktionen	638
9.5	Dateneingabe: „Daten in Rahmen“ (data.frame)	639
9.6	Auswahl und Sortierung von Daten	640
9.7	Ablaufsteuerung: logische Bedingungen und Funktionen in R	641
9.8	Einige mathematische und statistische Funktionen	643
9.8.1	Formulierung von Modellgleichungen	645
9.9	Einfache graphische Funktionen und Hilfsmittel	646
10	Übungsaufgaben zu ausgewählten Themen	650
	Lösungen der Übungsaufgaben	657
	Literaturverzeichnis	667
	Namensverzeichnis	680
	Sachverzeichnis	684

Einführung

- Definition und Aufgaben der Statistik
- Wissenschaftliche Arbeitstechnik
- Statistik und wissenschaftliche Methode
- Datenanalyse

1.1 Definition und Aufgaben der Statistik

Statistik ist die Lehre von der Variabilität / Streuung in den Beobachtungen.

Statistik ist die Kunst, Daten zu gewinnen, darzustellen, zu analysieren und zu interpretieren, um zu neuem Wissen zu gelangen.

Jeder von uns hat es erlebt, dass er wie der eingebildete Kranke und der eingebildete Gesunde echte Zusammenhänge oder echte Unterschiede nicht erkennt bzw. dass er nicht existente Unterschiede oder Zusammenhänge zu erkennen glaubt.

Im Alltag erfassen wir einen Zusammenhang oder einen Unterschied mit Hilfe von Sachkenntnis und nach dem sogenannten ersten „Eindruck“. Der Wissenschaftler, der gewisse neue Erscheinungen, Abhängigkeiten, Trends, Effekte vieler Art entdeckt und darauf eine Arbeitshypothese gründet, sichert diese ab gegen die Hypothese: die festgestellten Effekte sind allein durch den Zufall bedingt.

Die Frage, ob beobachtete Erscheinungen nur als Zufallsergebnisse gelten können oder typisch sind, beantwortet die Beurteilende Statistik. Mit Hilfe statistischer Verfahren lassen sich Fragen beantworten und Behauptungen überprüfen. Beispielsweise: Wie viele Personen sollte man vor einer Wahl befragen, um ein ungefähres Bild vom Wahlergebnis zu erhalten? Hat der zweistündige Schulsport in der Woche einen Trainingseffekt auf Herz und Kreislauf? Welche von mehreren Zahnpasten ist für die Kariesprophylaxe zu empfehlen? Wie hängt die Stahlqualität von der Zusammensetzung des Stahles ab? Die neue Verkäuferin hat den Tagesumsatz um DM 1000 erhöht. Die für eine bestimmte Krankheit charakteristische Überlebensrate (60%) wird durch Heilmittel A auf 90% erhöht. Die Kunstdünger K1, K2 und K3 zeigen bei Hafer keine unterschiedliche Wirkung. Zur Beantwortung dieser und anderer Fragen und Behauptungen benötigt man Daten (auf die wir in Abschnitt [1.4.3] näher eingehen werden). **Daten sind wichtig, um Annahmen zu bewerten und neues Wissen zu entdecken.**

Statistische Methoden befassen sich mit Daten aus unserer Umwelt, mit ihrer Gewinnung und Aufbereitung: Beschreibung, Auswertung und Beurteilung; das Ziel ist die Vorbereitung von Entscheidungen. Als Vorläufer der Statistik gelten (1) von Herrschern benötigte Daten über die Bevölkerung wie die Zahl wehrfähiger Männer und (2) durch den Spieltrieb angeregte Überlegungen über Wettchancen beim Würfelspiel. „Statistik“ war im 18. Jahrhundert die „Lehre von der Zustandsbeschreibung der Staaten“, wobei auch Daten über Bevölkerung, Heer und Gewerbe gesammelt wurden. Hieraus entwickelte sich die „**Beschreibende Statistik**“ mit der

Aufgabe, Zustände und Vorgänge zu beschreiben; hierzu dienen Tabellen, graphische Darstellungen, Verhältniszahlen, Indexzahlen und typische Kenngrößen, wie Lagemaße (z. B. arithmetischer Mittelwert) und Streuungsmaße (z. B. Varianz oder Standardabweichung).

Die „**Beurteilende Statistik**“ schließt anhand geeigneter Daten auf allgemeine Gesetzmäßigkeiten, die über den Beobachtungsraum hinaus gültig sind. Sie entwickelte sich aus der „Politischen Arithmetik“, die sich hauptsächlich mit Tauf-, Heirats- und Sterberegistern beschäftigte, um Geschlechtsverhältnis, Fruchtbarkeit, Altersaufbau und Sterblichkeit der Bevölkerung abzuschätzen. Die Beurteilende Statistik basiert auf der **Wahrscheinlichkeitsrechnung**, die mathematische Methoden zur Erfassung zufallsbedingter oder **stochastischer Experimente** beschreibt. Beispiele für stochastische Experimente oder Zufallsexperimente sind: das Werfen eines Würfels, Glücksspiele und Lotterien aller Art, das Geschlecht eines Neugeborenen, Tagestemperaturen, Ernteerträge, die Brenndauer einer Glühlampe, die Zeigerstellung eines Messinstruments bei einem Versuch, kurz jede Beobachtung und jeder Versuch, bei denen die Ergebnisse durch Zufallsschwankungen oder Messfehler beeinflusst sind. Fast stets interessieren hierbei weniger die Beobachtungen oder Messergebnisse selbst, sondern die übergeordnete Gesamtheit, der die Beobachtungen oder Messergebnisse entstammen. Beispielsweise die Wahrscheinlichkeit, mit einem intakten Würfel eine 4 zu werfen, oder der Anteil der Zwillingsgeburten in Deutschland. Bei vielen, **Wiederholbare Erfahrungen** betreffende Fragestellungen wird man nicht die zu untersuchende Menge aller möglichen Erfahrungen oder Beobachtungen, die so genannte Grundgesamtheit, vollständig erfassen können, sondern nur einen geeignet auszuwählenden Teil. Um einen Wein zu beurteilen, entnimmt der Kellermeister einem großen Fass mit dem Stechheber eine kleine Probe.

Diese **Stichprobe** gibt dann Aufschluss über die Häufigkeit und Zusammensetzung der interessierenden Merkmale der zu beurteilenden **Grundgesamtheit**, die man aus finanziellen, zeitlichen oder prinzipiellen Gründen nicht als Ganzes untersuchen kann. Vorausgesetzt wird das Vorliegen von **Zufallsstichproben**, bei denen jedes Element der Grundgesamtheit die gleiche Chance hat, ausgewählt zu werden. Enthält die Grundgesamtheit unterschiedliche Teilgesamtheiten, dann wird man geschichtete Zufallsstichproben wählen. Sinnvolle und repräsentative Teilmenge einer Tortensendung ist weder der Tortenboden, noch die Füllung, noch die Garnierung, sondern allenfalls ein Stück Torte. Besser noch sind mehreren Torten entnommene Proben von Boden, Füllung und Garnierung.

Zufallsstichproben gewinnt man im Zahlenlotto mit Hilfe einer mechanischen Vorrichtung. Im allgemeinen bedient man sich zur Gewinnung von Zufallsstichproben einer Tabelle von Zufallszahlen: Die Elemente werden nummeriert, ein Element gilt als ausgewählt, sobald seine Nummer in der Tabelle erscheint. Nach einem Zufallsverfahren entnommene Stichproben haben den Vorzug, daß die aus ihnen ermittelten statistischen Kenngrößen gegenüber denen der Grundgesamtheit im allgemeinen nur die unvermeidlichen **Zufallsfehler** [symmetrisch und meist klein] aufweisen, die, da sie das Resultat nicht verzerren – bei mehrfachen Wiederholungen gleichen sich zufällige Fehler im Mittel aus – abgeschätzt werden können, während bei den Verfahren ohne Zufallsauswahl noch so genannte methodische oder **systematische Fehler** [vermeidbar!] hinzukommen können, über deren Größe sich in der Regel keine Angaben machen lassen. Insbesondere die Abschätzung des Zufallsfehlers und die Prüfung, ob beobachtete Erscheinungen auch für die Grundgesamtheiten charakteristisch sind oder lediglich als Zufallsergebnisse gelten können, die so genannte **Prüfung von Hypothesen** über die Grundgesamtheit oder über Prozesse stehen im Vordergrund.

Bei der Übertragung eines Problems in **statistisch prüfbare Hypothesen** sollte auf die Auswahl und Definition geeigneter problemnaher und aussagekräftiger, möglichst messbarer Merkmale, auf die Präzisierung und Konstanz der Untersuchungsbedingungen sowie auf die Verwendung kostenoptimaler Stichproben- bzw. Versuchspläne Wert gelegt werden. Wir konzentrieren unser Augenmerk auf uns wesentlich erscheinende Teile des Sachverhalts und versuchen, diese stark vereinfachte Nachbildung als Modell zu formulieren, wobei einige Annahmen notwendig sind.

[Ohne Annahmen kommt man aus, wenn lediglich eine Beschreibung geplant ist, oder wenn nur Hypothesen gewonnen aber nicht geprüft werden sollen.]

1.2 Wissenschaftliche Arbeitstechnik

- Daten und Modelle
- Kreisprozesse
- Modelle in der Statistik

1.2.1 Daten und Modelle

Der Wissenschaftler stellt Fragen und bemüht sich, sie zu beantworten. Hierbei helfen statistische Methoden, indem sie Planung und Auswertung wissenschaftlicher Studien präzisieren. Dies erfordert Sachkenntnis, Umsicht und Grundkenntnisse der Sprache der Statistik. Insbesondere ist zu kontrollieren, ob die Voraussetzungen der hierbei genutzten statistischen Modelle seitens der gewonnenen Daten erfüllt sind.

Statistische Modelle sind Annahmen über Entstehung und Struktur der zu analysierenden Daten in der Sprache des Statistikers.

Wichtig ist der Vergleich der beobachteten Datenstruktur mit der im Modell formalisierten Datenstruktur, formuliert aufgrund des **Vorwissens** des Untersuchers über den **die Daten erzeugenden Prozess** (1) und die zu prüfenden Hypothesen (2), die durch Ziel und Zweck der Untersuchung bestimmt sind.

Meist wird ein Standardmodell genutzt, seine Brauchbarkeit überprüft und gegebenenfalls ein anderes statistisches Modell gewählt, das den Daten angemessener ist und die gewünschten bzw. modifizierten Hypothesen effizienter zu prüfen gestattet.

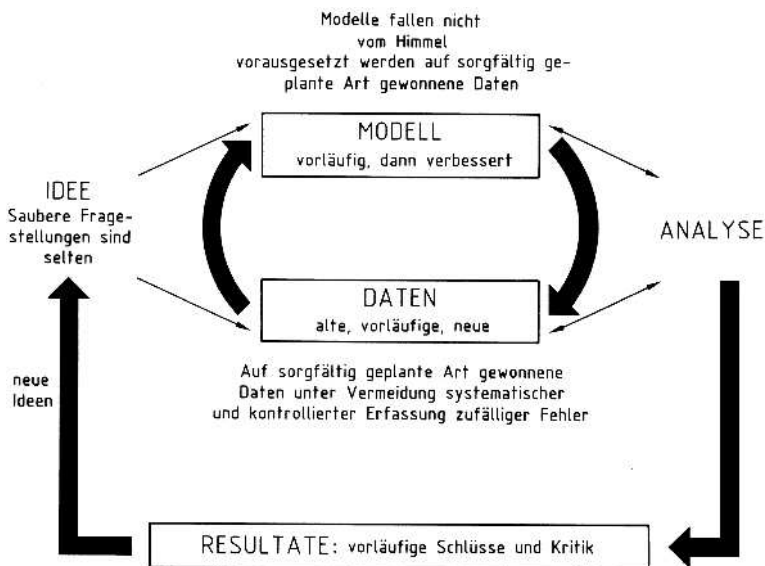


Abb. 1.1. Kreisprozesse in der wissenschaftlichen Arbeit

Modelle sind wichtige Entscheidungshilfen. Modelle beschreiben und erklären, außerdem ermöglichen sie Voraussagen. Die wissenschaftliche Arbeitsweise ist eine Strategie, die darauf abzielt, allgemeine Gesetzmäßigkeiten zu finden und sie mit Hilfe prüfbarer und ablehnbarer (falsifizierbarer) Aussagen zu einer logisch-mathematisch strukturierten Theorie zu entwickeln. Hierbei resultiert eine angenäherte Beschreibung der erfassbaren Wirklichkeit. Diese angenäherte Beschreibung ist revidierbar und komplettierbar. Typisch für die wissenschaftliche Methodik ist der **Kreisprozess** oder **Iterationszyklus**:

Mutmaßungen (Ideen) → Plan → Beobachtungen → Analyse → Ergebnisse → Neue Mutmaßungen (Neue Ideen) → ...; hierbei werden Widersprüche und Unverträglichkeiten ausgeschaltet sowie die Modelle und Theorien verbessert. Die bessere Theorie ist die, die uns erlaubt, mehr zu erklären und bessere Voraussagen zu machen.

1.2.2 Kreisprozesse

Die Wissenschaft ist ein Kreisprozess, ein Prozess von Wechselwirkungen zwischen (Erwartung und Erfahrung) Theorienbildung UND empirischer Forschung; dieser Prozess unterliegt der Selbstkorrektur.

Für uns ist wichtig: Aufgrund der problemspezifischen Fragestellung werden **Annahmen** gemacht hinsichtlich der Struktur des zugrunde liegenden Modells und des entsprechenden statistischen Modells. Nach Prüfung der Verträglichkeit von Beobachtungen und statistischem Modell werden Kenngrößen zur statistischen Beschreibung einer Grundgesamtheit, so genannte **Parameter**, – feste Zahlen, die Modelleigenschaften beschreiben – , geschätzt und Hypothesen über die Parameter geprüft. In beiden Fällen resultieren **Wahrscheinlichkeitsaussagen**. Aufgabe der Statistik ist es somit, der Fragestellung und den Daten angemessene statistische Modelle zu finden und zu schaffen und durch sie die in den Daten steckende wesentliche Information herauszuschälen, d. h. die **Statistik liefert Modelle** für die Informationsreduktion, um Zusammenhänge zu erkunden und spezielle Fragen zu beantworten.

Diese und andere Verfahren bilden den Kern einer auf die kritische Gewinnung und Beurteilung von Messwerten und Häufigkeiten ausgerichteten Datenanalyse, wie sie für viele Bereiche in Technik, Wirtschaft, Politik und Wissenschaft notwendig ist. **Datenanalyse** ist die systematische Suche nach aufschlussreichen Informationen über Erscheinungen, Strukturen und Vorgänge anhand von Daten und graphischen, mathematischen sowie insbesondere statistischen Verfahren ohne oder mit Wahrscheinlichkeitskonzept. Hierbei geht es weniger darum, Daten zu Wahrscheinlichkeiten zu „vermahlen“ und statistisch signifikante Befunde zu erzielen, die ja bedeutungslos oder unwichtig sein können. Nicht die statistische Signifikanz, sondern die **praktische Relevanz** zählt. Eine Bewertung von Befunden hängt von vielen Faktoren ab, etwa von der fachspezifischen Bedeutung, von der Verträglichkeit mit anderen Resultaten oder von den Voraussagen, die sie ermöglichen. Diese Evidenz kann kaum statistisch bewertet werden.

Daten haben viele Wirkungen auf uns, die über eine Entscheidung hinausgehen. Sie geben uns Verständnis, Einsicht, Anregungen und überraschende Ideen, um neue Aktivitäten zu planen.

Planen heißt überlegen, wie, mit welchen Ressourcen und in welchem Zeitraum ein angestrebtes Ziel erreicht werden kann. Dabei sollte man Alternativen und Konsequenzen aufzeigen und damit künftige Entscheidungen rationalisieren, möglichst flexibel und unter vorausschauender Begegnung möglicher zusätzlicher Schwierigkeiten. Unvorhersehbare Umstände können zur Revision des Gesamtplans führen. Übersicht 1 gibt Details, ergänzt durch Sachs [Sac06].

Übersicht 1. Erfahrungsbedingte Hypothesen und theoriegeleitete Erfahrungen ergänzen sich

Bemerkungen zur Behandlung wissenschaftlicher Probleme

1. **Formulierung der Fragestellung**, der Wünsche und Ziele: Häufig ist es zweckmäßig, das gesamte Problem in Teilprobleme zu zerlegen und einige Fragen zu stellen:
 - a) Anlass und Zweck der Studie? Nahziel(e) und Fernziel(e)?
 - b) Skizzierung der Ausgangssituation anhand von Standardfragen: was? wie? wo? wann? wie viel? was ist unbekannt? was wird vorausgesetzt?
 - c) Problemtyp: Schätzungen? Standardisierungen? Vergleiche? Aufsuchen von Optimalbedingungen? Bedeutsamkeit von Änderungen? Zusammenhänge zwischen Variablen?
 - d) Angestrebter Gültigkeitsbereich und erforderliche Genauigkeit der Aussagen?
 - e) Konsequenzen sowie Interessenten der möglichen Resultate?
2. **Prüfung aller Informationsquellen**: Hauptsächlich Erkundigungen und Literatur-Recherchen und Suche im **Internet**: was ist mit welchen Methoden bereits erkundet worden? Sind diese Befunde zuverlässig [begründete Annahmen oder Tatsachen („woher wissen Sie das?)]? Welche Alternativen existieren?
3. **Wahl der Strategie**:
 - a) **Entwicklung des problemspezifischen Modells**. Anzahl der zu berücksichtigenden Variablen. Einführung vereinfachender Annahmen. Prüfung, ob eine Möglichkeit besteht, das Problem durch Transformation weiter zu vereinfachen, z. B. Untersuchungen an Zellkulturen oder an isolierten Organen anstatt am Menschen.
 - b) **Entwicklung der Untersuchungstechnik**. Die Methode sollte problemnahe Messwerte (bzw. Häufigkeiten) liefern, gewonnen ohne systematische Fehler!
 - c) **Entwicklung des statistischen Modells**. Plan der statistischen Analyse. Klare Formulierung: des Modells, der Voraussetzungen des Modells, der Parameter und Konfidenzintervalle, der Hypothesenpaare sowie weiterer Details, etwa Art der Randomisierung.
4. **Prüfung der Strategie**: Anhand von Probe-Erhebungen und Vorversuchen. Überprüfung der Untersuchungstechnik und der Verträglichkeit der Beobachtungswerte mit dem statistischen Modell.
5. **Festlegung und Realisierung der Strategie**: Aufgrund jetzt vorliegender Erfahrungen.
 - a) **Endgültige Festlegung aller wesentlichen Punkte**, z. B. der Untersuchungsmethode, der Versuchsobjekte, der Merkmalsträger, der Merkmale und Einflussgrößen, der Kontrollen, der Bezugsbasis; Berücksichtigung des Nulleffektes, Ausschaltung der unkontrollierbaren Variablen; Stichprobenumfang bzw. Zahl der Wiederholungen, Berücksichtigung des Aufwandes an Arbeitskräften, Geräten, Material, Zeit u. a.; Umfang des gesamten Programmes; endgültige Formulierung des Modells der statistischen Analyse; Vorbereitung und Kontrolle der Datenerfassung, Strukturierung der geplanten Tabellen und Formulierung der zu prüfenden Hypothesen mit Vorgabe des Signifikanzniveaus.
 - b) **Durchführung der Untersuchung**, möglichst ohne Modifikation. Datenanalyse, Angabe von Konfidenzintervallen und Prüfung weniger Hypothesen.
6. **Entscheidungen und Schlussfolgerungen**:
 - a) **Ergebnis**: Kontrolle der Berechnungen. Darlegung der Resultate (Konfidenzintervalle!) in Form von Tabellen und/oder graphischen Darstellungen.
 - b) **Interpretation**: Hinweise auf Plausibilität, praktische Bedeutung, Überprüfbarkeit und Gültigkeitsbereich der Untersuchungen. Unter Berücksichtigung der vereinfachenden Annahmen wird das Ergebnis der Hypothesenprüfung kritisch gewürdigt und, wenn möglich und sinnvoll, mit den Befunden anderer Autoren verglichen. Ist eine Wiederholung der Untersuchung mit weniger vereinfachenden Annahmen, mit verbesserten Modellen, neuer Untersuchungstechnik usw. erforderlich? Ergeben sich neue, aus den Daten gewonnene Hypothesen, die durch unabhängige neue Untersuchungen überprüft werden müssen?
 - c) **Bericht**: Beschreibung wesentlicher Details der gesamten Untersuchung, einschließlich der negativen Befunde und wünschenswerter neuer Ansätze.

1.2.3 Modelle in der Statistik

Ein Modell, etwa eine Landkarte oder ein Globus, ist eine vereinfachte Nachbildung eines Sachverhaltes. Es dient zur **Erklärung und Voraussage**. Modellvorstellungen sind unerlässlich, wenn Untersuchungen geplant werden: es beginnt mit theoretischen Überlegungen zur Identifizierung und Definition des Problems: Jede Anwendung statistischer Methoden setzt ein Modell voraus, es sei denn man begnügt sich mit einer einfachen Beschreibung von Daten anhand von Maßzahlen.

Ein statistisches Modell ist der **mathematische Ausdruck für eine durch Randbedingungen eingeschränkte Wirklichkeit**; formal erfasst und analysiert wird die Struktur eines Systems oder Prozesses. Bestimmte Merkmale der zu modellierenden Realität werden als wesentlich aufgefasst und im Modell angemessen nachgebildet. Nach der empirischen Überprüfung anhand von Experimenten, Beobachtungen oder Erhebungen wird das Modell korrigiert und verfeinert, bis die Modell-Rechnungen die Wirklichkeit hinreichend gut beschreiben. Der Einfluss als unwesentlich aufgefasster Merkmale, die im Modell unberücksichtigt bleiben, ist die Ursache für die Abweichungen des Modells von der Realität. Diese **Abweichungen oder Residuen** sind natürlich umso kleiner, je detaillierter und angemessener ein Modell ist. Die Residuen dürfen keine Struktur aufweisen; sie müssen zufällig verteilt sein (vgl. im Kapitel [8] zur Modellbildung). Modelle sollten einfach und gut interpretierbar sein, sowie eine überzeugende Antwort auf die zugrundeliegende Fragestellung ermöglichen.

Statistische Methoden geben eine unvollständige aber aufschlussreiche Beschreibung von Phänomenen, die zu kompliziert sind, um vollständig durch ein Modell erfasst zu werden. Die Wahl eines Modells hängt ab von dem zu modellierenden Objekt oder Prozess und von der Aufgabenstellung und dem Ziel der Untersuchung. Bei der Wahl des Modells wird man bestrebt sein, alle wesentlichen Umstände zu berücksichtigen, damit die aufgrund dieses Modells erzielten Ergebnisse der Wirklichkeit entsprechen, wobei, falls möglich, ein eher einfaches Modell zu bevorzugen ist.

Ein **statistisches** oder **stochastisches Modell** ist ein mathematisches Modell, das neben strukturgebenden Konstanten **Zufallsvariable** (ausführlich im Kapitel [5] zu Zufallsvariablen) enthält, um Erscheinungen zu beschreiben, in denen der Zufall eine wesentliche Rolle spielt. Gedanklich gehen wir hierbei von **Zufallsexperimenten** aus. Die Konstanten heißen **Parameter**; sie charakterisieren als Kennzahlen einer Grundgesamtheit, etwa einer normalverteilten Grundgesamtheit, das Modell, die den Zufallsvariablen zugrundeliegende **Wahrscheinlichkeitsverteilung**: das ist die Gesetzmäßigkeit, nach der die betrachtete Zufallsvariable ihre Werte annimmt.

Modelle sind um so flexibler, je mehr Parameter sie haben. Einen Parameter enthält z. B. die Poisson-Verteilung, zwei Parameter, Erwartungswert und Standardabweichung, charakterisieren eine Normalverteilung. Parameter sind meist unbekannt. Man schätzt sie aus den **Beobachtungen**, die als **Realisierungen von Zufallsvariablen** angesehen werden (die einer konkreten Zufallsstichprobe entstammen). Mehrere Parameter zugleich aus dem vorliegenden Datenkörper richtig zu schätzen, ist jedoch schwierig. Aus diesem Grunde beschränkt man sich oft lieber auf einfache, übersichtliche Modelle, auch wenn man weiß, dass es sich um eine Approximation handelt. In manchen Fällen lässt sich durch **Transformation** der Realisierungen von Zufallsvariablen eine der bekannten Verteilungen annähern, z. B. eine Normalverteilung. Dann ist es möglich, die für dieses Modell entwickelten Standardverfahren der Beurteilenden Statistik auf die vorliegenden Beobachtungen anzuwenden. Dieses erfordert:

1. Umsicht und Beherrschung des Fachgebiets,
2. Vertrautheit mit der statistischen Terminologie, mit den wichtigen Modellen und Methoden, einschließlich ihrer Voraussetzungen sowie
3. eine gründliche Überprüfung, ob im Anwendungsfalle die gewonnenen Daten diese Voraussetzungen, auch bei voraussetzungsarmen sogenannten verteilungsunabhängigen statistischen Verfahren, erfüllen (zumindest teilweise) bzw.
4. inwieweit Abweichungen noch toleriert werden dürfen und wie sich diese auf die Resultate auswirken werden.

1.3 Statistik und wissenschaftliche Methode

- Wiederholbare Erfahrungen
- Deskriptive Statistik
- Explorativer Ansatz
- Konfirmativer Ansatz
- Merkmale, Grundgesamtheit
- Stichproben
- Zufallsstichproben

1.3.1 Wiederholbare Erfahrungen

Die **Wissenschaft** lehrt uns:

- wie etwas erkannt wurde,
- was, genau, bisher bekannt ist und
- was noch unbekannt ist sowie
- den Umgang mit Empirie, Unsicherheit und Wahrheit.

Den Gegenstand empirischer Wissenschaften bilden nicht einmalige isolierte, ein einzelnes Individuum oder Element betreffende Ereignisse oder Merkmale, sondern **wiederholbare Erfahrungen**, eine Gesamtheit von – als gleichartig betrachteter – Erfahrungen, über die Aussagen gefordert werden.

Als Semmelweis im Jahre 1847 in der Geburtshilfe-Klinik in Wien gegen den Widerstand seiner Kollegen hygienische Maßnahmen durchsetzte, wusste er nichts über die bakteriologischen Erreger des Kindbettfiebers. Auch konnte er den Erfolg seines Experimentes nicht direkt beweisen, denn auch nach der Einführung der Hygiene starben noch Frauen in seiner Klinik am Kindbettfieber. Die Müttersterblichkeit aber war von 10,7% (1840–1846) über 5,2% (1847) auf 1,3% (1848) zurückgegangen, und da Semmelweis diese Prozentsätze an einer großen Zahl von Wöchnerinnen (21 120; 3375; 3556) errechnet hatte, ergab sich die Schlussfolgerung, die Hygiene beizubehalten.

Statistische Methoden sind überall da erforderlich, wo Ergebnisse nicht beliebig oft und exakt reproduzierbar sind. Die Ursachen dieser Nichtreproduzierbarkeit liegen in **unkontrollierten und unkontrollierbaren Einflüssen**, in der Ungleichartigkeit der Versuchsobjekte, der Variabilität des Beobachtungsmaterials und in den Versuchs- und Beobachtungsbedingungen. Diese Ursachen führen in den Beobachtungsreihen zu der **Streuung** quantitativ erfasster Merkmale. Da infolge dieser Streuung ein gefundener Einzelwert – die Variabilität einzelner Merkmale ist bei naturwissenschaftlichen Untersuchungen meist kleiner als bei sozialwissenschaftlichen – kaum exakt

reproduzierbar sein wird, müssen sichere und eindeutige Schlussfolgerungen zurückgestellt werden. Die Streuung führt damit zu einer **Ungewissheit**, die häufig nur **Entscheidungen** ermöglicht. Dieses ist der Ansatzpunkt einer modernen Definition der Statistik als Entscheidungshilfe, die auf Abraham Wald (1902–1950) zurückgeht: **Statistik ist eine Zusammenfassung von Methoden, die uns erlauben, vernünftige optimale Entscheidungen im Falle von Ungewissheit zu treffen.**

Die **Beschreibende (Deskriptive) Statistik** befasst sich mit der Untersuchung und Beschreibung **möglichst der ganzen Grundgesamtheit**. Sie ist einfach und verständlich; graphische Methoden, die auch gut zur Darstellung der Resultate dienen, zeigen Unerwartetes deutlich. Außerdem ist sie unerlässlich, wenn für die Daten (noch) kein Modell vorliegt.

Die **Beurteilende (Schließende) Statistik** untersucht demgegenüber nur einen Teil, der für die Grundgesamtheit, deren Eigenschaften uns interessieren, charakteristisch oder repräsentativ sein soll. Es wird also von einem Teil der Beobachtungen auf die Grundgesamtheit aller geschlossen ("schließende" Statistik). Entscheidend ist hierbei, dass der zu prüfende Teil der Grundgesamtheit – die **Stichprobe** – zufällig, sagen wir nach einem Lotterieverfahren, ausgewählt wird. Wir bezeichnen eine Stichprobenentnahme als zufällig, wenn jede mögliche Kombination von Stichprobenelementen der Grundgesamtheit dieselbe Chance der Entnahme besitzt. **Zufallsstichproben** sind wichtig, da nur sie Rückschlüsse auf die Grundgesamtheit zulassen. Totalerhebungen sind häufig kaum oder nur mit großem Kosten- und Zeitaufwand möglich!

1.3.2 Deskriptive Statistik

Die wissenschaftliche Arbeitsweise ist eine Strategie, die darauf abzielt, **allgemeine Gesetzmäßigkeiten** zu finden und sie zu einer möglichst logisch-mathematisch strukturierten Theorie zu entwickeln. Hierbei resultiert eine angenäherte Beschreibung der Wirklichkeit, eine Rekonstruktion der erfassbaren Wirklichkeit. Diese Approximation ist **revidierbar und komplettierbar**. Typisch für die Wissenschaft ist daher ein Iterationszyklus (Abbildung 1.1) der Art: Ideen, Beobachtungen, Ergebnisse, neue Ideen. Die Ideen sind Bausteine für Modelle und Theorien. Durch die Iterationen werden Unverträglichkeiten und Widersprüche eliminiert und die Modelle und Theorien verbessert. Hierfür müssen Beobachtungen gemacht und Daten gewonnen werden, die dann analysiert werden, um das Ausgangskonzept zu modifizieren und zu präzisieren.

Dass zu viele Daten nicht angemessen analysiert werden, hat meist mehrere Ursachen:

1. Die Fakten sind komplizierter als ursprünglich erwartet.
2. Mit zunehmender Anhäufung der Daten legt sich die ursprüngliche Begeisterung.
3. Man strebt nach immer neueren und besseren Daten und schiebt so die Analyse vor sich her.

Für medizinische Daten kommt neben der **biologischen Variabilität** und ihrer Problematik noch hinzu, dass fast stets viele Variablen eine Rolle spielen, mehr als in Physik und Chemie. Von diesen Variablen werden in der Regel die üblichen Voraussetzungen statistischer Verfahren kaum erfüllt. Daher spielen gerade hier **datenanalytische Konzepte** wie z. B. graphische Darstellungen eine große Rolle.

Ein wesentlicher Teil der Statistik ist die **Datenbeschreibung** einschließlich einer systematischen Suche nach aufschlussreichen Informationen über die Struktur eines Datenkörpers. **Strukturen in den Daten** und bedeutsame Abweichungen von diesen Strukturen sollen aufgedeckt werden. Die Bewertung derartiger Befunde hängt von mehreren Faktoren ab, etwa von ihrer Repräsentativität, von der medizinischen Bedeutung, von der Verträglichkeit mit anderen Resultaten oder von den Voraussagen, die sie ermöglichen. Diese Evidenz gilt es, angemessen abzuschätzen. Daten haben zudem viele Wirkungen auf uns, die über eine Entscheidung hinausgehen. Sie geben uns Verständnis, Einsicht, Anregungen und überraschende Ideen.

1.3.2.1 Dimensionalität

Daten sind stets **mehrdimensional** oder **multivariat**, wenn die Bedingungen beobachtet und protokolliert werden, unter denen sie entstehen. Wie groß soll diese Liste aussagekräftiger Einflussgrößen und damit die Dimensionalität p sein? Um dies zu entscheiden, bedarf es der Kombination von Vorinformation und experimenteller Einsicht. Bei der Verringerung oder Reduktion der Dimensionalität multivariater Daten muss ein Optimum in Bezug auf Einfachheit, Klarheit und Detaillierungsgrad angestrebt werden. Ist der Verdichtungsgrad zu niedrig, so lassen sich die Daten nicht überschauen, ist er zu hoch, so ist die Aussage dürftig. Welche Variablen sollten vernachlässigt werden? Welche Variablen können zu einem neuen Maß mit stabilen statistischen Eigenschaften zusammengefasst werden? Zur Übersicht und zur Beantwortung mancher Frage dienen hier graphische Darstellungen. Zu viele gewonnene Daten werden eher oberflächlich ausgewertet und, wenn überhaupt, unübersichtlich dargestellt. Wenigen Daten misstraut der Leser, viele Daten überblättert er. Es ist keineswegs trivial, die geeignete Informationsdichte für Daten und Resultate zu finden, zumal sie auch vom mutmaßlichen Leser und seinen Kenntnissen abhängt. Besonders instruktiv sind Tabellen mit 3 x 4 oder höchstens 4 x 5 Fächern.

1.3.2.2 Data Editing

Nach der Datengewinnung erfolgt die Entfernung oder Modifizierung von Daten, die mit der Masse der Daten nicht harmonisieren. Hierfür gibt es statistische Verfahren, ohne dass zu subjektiv entschieden wird. Entsprechende Verfahren und Regeln sind vor der Datengewinnung festzulegen. Dieses **Data Editing** (Data Cleaning) ist ein Teil der statistischen Analyse, der besonders viel Umsicht erfordert, da sonst wertvolle Evidenz verschwindet und die Möglichkeit, zu Wahrscheinlichkeitsaussagen zu gelangen, eingeschränkt wird. Vor dem „Data Editing“ wird man nach der Untersuchungseinheit, nach ihrer Anzahl, ihrer möglichen Unabhängigkeit und ihrer Anordnung/Gruppierung fragen, dann nach den untersuchten Merkmalen. Weitere Fragen:

- Traten bei der Datengewinnung Besonderheiten auf?
- Gab es systematische Unterschiede zwischen Maschinen bzw. Beobachtern?
- Lassen sich fehlende Beobachtungen ersetzen?
- Gibt es Beobachtungen, die den erwarteten Variationsbereich deutlich sprengen und als mutmaßliche Ausreißer zu behandeln sind?
- Welchem Verteilungstyp folgen die Daten?
- Ist ihre Variabilität auffallend?
- Was folgt aus den Punktwolken untereinander korrelierender Variablen? Mitunter lassen sich anhand der Punktwolken deutlich abweichende Beobachtungen erkennen.

1.3.3 Explorativer Ansatz

Liegen Daten vor, die nicht auf sorgfältig geplante Art gewonnen worden sind, etwa „Nichtzufallsstichproben“ oder irgendwelche Teilgesamtheiten, und ist es noch nicht möglich, präzise Fragen zu stellen, dann werden diese Daten - deren Struktur und Auffälligkeiten interessieren - anhand von Methoden der Beschreibenden Statistik und der (erkundenden) **Explorativen Datenanalyse (EDA)** untersucht, ohne dass ein Modell vorausgesetzt wird.

Die Explorative Datenanalyse dient

1. der Darstellung von Daten,
2. dem Auffinden von Strukturen,
3. dem Erkennen von Besonderheiten und
4. der Suche nach neuen Möglichkeiten.

Aufgabe der EDA ist das Aufspüren und die zusammenfassende Darstellung unbekannter Strukturen in meist umfangreichen Stichproben auch mehrdimensionaler Daten bei (1) wenig klarer Fragestellung, (2) fehlender (unbekannter) Grundgesamtheit, (3) kaum geplanter Datenerhebung, (4) Unklarheit über die Auswahl geeigneter Modelle (da viele gleich gute Modelle in Frage kommen) und bei (5) quantitativ nicht fassbarer Aussagegenauigkeit.

Während die EDA, **datengesteuert**, hilft, durch Modellsuche und Gewinnung neuer Hypothesen neue Strukturen und neue Hypothesen aufzuspüren, bemüht sich die Beurteilende Statistik, **hypothesengesteuert**, darum, falsche Aussagen zu verhindern; und zwar aufgrund von Zufallsstichproben und mitunter auch im Anschluss an Methoden der Beschreibenden Statistik bzw. der Explorativen Datenanalyse. Explorative Verfahren geben durch die Suche nach Auffälligkeiten Anstöße zur Bildung von Hypothesen und Modellen und helfen bei der Präzisierung der Fragestellung etwa im Sinne eines Konfidenzbereichs. In der explorativen Phase errechnete P-Werte - hier wird die Beurteilende Statistik explorativ eingesetzt und interpretiert - können als Plausibilitätsmaß interpretiert werden. Die dort gefundenen Modelle und Hypothesen bedürfen einer Überprüfung bzw. Bestätigung durch die Beurteilende Statistik (**Konfirmative Datenanalyse**), im allgemeinen anhand neuer Daten. Im Gegensatz zur Beurteilenden Statistik baut die EDA nicht auf einem vorformulierten Wahrscheinlichkeitsmodell auf: es werden keine Annahmen gemacht und keine Hypothesen geprüft. Die EDA beginnt vielmehr mit dem Studium der Daten, mit ihrer Darstellung und Zusammenfassung, um neue Einsichten zu gewinnen. Werden dabei nichttriviale Strukturen gefunden, so kann man versuchen, diese durch ein statistisches Modell zu beschreiben. Dabei sollten jedoch stark einschränkende Modellannahmen vermieden werden. Deshalb ist die Anwendung „**resistenter**“ **Schätzverfahren** geboten; das sind Verfahren, die unempfindlich sind gegenüber schlechten (ausreißerverdächtigen) Daten. Die Daten können dann mit dem geschätzten Modell bereinigt werden. In einer verfeinerten Analyse kann anschließend in den Residuen wiederum nach Strukturen gesucht werden, die dann in Modellen berücksichtigt werden können. Dieser Prozess lässt sich iterativ fortsetzen, bis die Residuen zufällig verteilt sind. Dabei werden auf jeder Stufe tabellarische, graphische und andere Darstellungen als Hilfsmittel herangezogen. Diese Vorgehensweise erfordert eine gute Kenntnis sowohl des Sachproblems als auch der Art und Weise, wie die Daten zustandegekommen sind.

1.3.4 Konfirmativer Ansatz

Auf der Beschreibenden Statistik aufbauend, spielt die **Beurteilende, schließende, mathematische, wertende, induktive oder analytische Statistik** (statistical inference) die entscheidende Rolle. Sie ermöglicht den Schluss von der Stichprobe auf die zugehörige Grundgesamtheit (z. B. die Schätzung des Wahlresultates anhand bekannter Einzelergebnisse ausgewählter Wahlkreise), auf **allgemeine Gesetzmäßigkeiten**, die über den Beobachtungsbereich hinaus gültig sind. In allen empirischen Wissenschaften ermöglicht sie durch Gegenüberstellung empirischer Befunde mit Ergebnissen, die man aus wahrscheinlichkeitstheoretischen Modellen – Idealisierungen spezieller experimenteller Situationen – herleitet, die Beurteilung empirischer Daten und die Überprüfung wissenschaftlicher Hypothesen und Theorien; wobei allerdings nur Wahrscheinlichkeitsaussagen möglich sind, die dann dem Praktiker unentbehrliche Informationen als Grundlage für seine Entscheidungen bieten.

In der **Schätztheorie** ist eine Entscheidung darüber zu treffen, wie man anhand einer Stichprobe möglichst viel über die charakteristischen Kennwerte der zugehörigen Grundgesamtheit erfährt. In der **Testtheorie** handelt es sich darum, zu entscheiden, ob die Stichprobe aus einer bestimmten (vorgegebenen) Grundgesamtheit entnommen wurde.

Die moderne Statistik ist interessiert an der problemgerechten und am Modell orientierten Planung, Durchführung und Auswertung von Experimenten und Erhebungen: Ein **Experiment** ist eine geplante und kontrollierte Einwirkung eines Untersuchers auf Objekte – eine **Erhebung** ist eine

geplante und kontrollierte Erfassung eines Zustandes oder Vorgangs an Objekten einer Gesamtheit. Hier: vorliegende Merkmalsstruktur, meist von Individuen (Einflussgrößen kaum variierbar) – dort: Zuordnung der Objekte oder Individuen zu den experimentellen Bedingungen (Einflussgrößen variierbar). Entscheidend für die Versuchsplanung ist die Frage, für welche Grundgesamtheit die Ergebnisse repräsentativ sein sollen.

Aufgabe und Ziel der Beurteilenden Statistik - auffassbar als **mathematische Theorie wiederholbarer Ereignisse** - ist die Entwicklung mathematischer und vor allem stochastischer Verfahren sowie ihre Anwendung auf Probleme in praktisch allen Gebieten, in denen **zufallsabhängige Phänomene** zu modellieren und dadurch Strukturen und deren Variabilität zu erkunden sind. Modelliert werden bestehende und mögliche reale Strukturen, die als Vorbilder für Planung und Entwicklung oder als Hypothesen zur Erklärung realer Phänomene dienen können. So erhält man unvollständige aber aufschlussreiche Beschreibungen von Phänomenen, die zu kompliziert sind, als dass sie vollständig durch ein Modell erfasst werden. Wesentliche Inhalte der Beurteilenden Statistik sind Prinzipien der Versuchsplanung und der Planung und Analyse von Erhebungen, Wahrscheinlichkeitsrechnung, Hypothesenprüfung und Zusammenhangsanalysen. Im Vordergrund steht die Entwicklung und Anpassung spezieller Verfahren, die den jeweiligen Besonderheiten und Fragestellungen gerecht werden und die es ermöglichen, **zu Entscheidungen und Schlussfolgerungen zu gelangen und deren Unsicherheit abzuschätzen**. Vorausgesetzt wird eine sorgfältige Planung, die es gestattet, hierfür aussagekräftige Daten zu gewinnen und diese dann angemessen auszuwerten, so dass sie ihre Herkunft offenbaren, eine Abschätzung ihrer Unsicherheit möglich wird und auch die angestrebte Verallgemeinerung möglich wird. Damit wird dem Praktiker geholfen, Fragen aus den meisten Fachbereichen zu beantworten. Das sind selten Fragen, die sich, aufgrund wohlbekannter und intensiv genutzter statistischer Modelle, leicht beantworten lassen. Häufig bereitet die gezielte Beantwortung der sachlogisch formulierten Frage einige Mühe - mitunter ist sie nicht möglich. Die Antwort, die die Statistik bietet, wird auch immer von einer Schätzung der Ungenauigkeit dieser Antwort begleitet sein müssen.

Übersicht 2. Beurteilende Statistik

Die Beurteilende (Schließende) Statistik

- (1) geht über die Beschreibende Statistik hinaus, indem sie insbesondere bei Erhebungen nach einem Zufallsverfahren gewonnene Stichproben, **Zufallsstichproben**, bzw. bei Experimenten **randomisierte Beobachtungen** voraussetzt;
- (2) nutzt auf der Grundlage der **Wahrscheinlichkeitsrechnung** vielfältige Methoden für die Modellierung und Beschreibung stochastischer Erscheinungen und Gesetzmäßigkeiten;
- (3) ermöglicht anhand der aus Experimenten und Erhebungen gewonnenen Daten allgemeingültige **Aussagen über die den Daten zugrundeliegenden Grundgesamtheiten, Zusammenhänge und Prozesse**.

Sorgfältig gewonnenen Daten wird unterstellt, sie entstammen einer bestimmten Grundgesamtheit, über die Aussagen erwünscht sind. Die Daten dienen dann dazu, **Annahmen oder Hypothesen über diese Grundgesamtheit zu prüfen**, d. h. gültige von ungültigen Hypothesen zu unterscheiden. Die Methoden der Beurteilenden oder Schließenden Statistik dienen zum Rückschluss von den Daten auf die Grundgesamtheit, zur Beurteilung der Hypothesen über die Grundgesamtheit, anhand von Vertrauensbereichen und statistischen Tests, indem

GÜLTIGE HYPOTHESEN ANGEREICHERT WERDEN.

1.3.5 Merkmale, Grundgesamtheit, Stichprobe

In der Statistik bezeichnet man die betrachteten Charakteristika der **Untersuchungsobjekte** als **Merkmale**. Diese treten an den Untersuchungsobjekten (**Fälle**, **Einheiten** oder **Merkmals-**

trägern), die ein oder mehrere Merkmale aufweisen, in verschiedenen Ausprägungen auf. Das Auffinden aussagekräftiger Merkmale ist eine wichtige Teilaufgabe der Statistik. Je nachdem wie die **Merkmalsausprägungen** beschrieben werden, unterscheidet man durch Zählen (Kinderzahl) oder Messen (Körpermaße) erfasste **quantitative Merkmale** von den **qualitativen Merkmalen**, wie z. B. Geschlecht, Beruf, Familienstand sowie ordinale Merkmale, die sich nach der Intensität der Merkmalsausprägung in eine Rangfolge mit numerisch nicht definierbaren Intervallen bringen lassen (Hühnerhof-Hackordnung, Schulnoten).

Die Menge aller möglichen Einheiten, welche der statistischen Betrachtung zugrunde liegen, nennen wir **Grundgesamtheit**. Man unterscheidet zwei Arten von Grundgesamtheiten: Einmal eine endliche Grundgesamtheit existierender Objekte wie sie für eine Erhebung typisch ist, zum anderen eine beliebig große Grundgesamtheit hypothetischer Objekte, wie sie für Experimente typisch sind; hier wird durch Wiederholung der Messung unter gleichen bis ähnlichen Bedingungen eine Grundgesamtheit von Messwerten geschaffen, die als Realisierungen von Zufallsvariablen mit bestimmter Verteilung aufgefasst werden. Insbesondere bei Erhebungen ist jede Grundgesamtheit **sachlich, räumlich und zeitlich** zu definieren. Zusätzlich wichtig sind Angaben über den Anlass der Untersuchung und über die Auswahl der Grundgesamtheit und der Stichprobe, eine präzise Formulierung der Fragestellungen sowie Definitionen der Untersuchungs- bzw. Beobachtungseinheiten und der Merkmale. Wichtig ist auch eine Liste von Merkmalen, die stets vorhanden sind (Positivkatalog) und seltener, eventuell zusätzlich, eine andere mit stets fehlenden Merkmalen (Negativkatalog). Einige weiterführende Fragen :

1. Anlass und Zweck der Studie?
2. Was ist bereits bekannt, was ist strittig, was ist unbekannt?
3. Inwieweit ist das thematische und methodische Umfeld der Studie bekannt?
4. Welche Konsequenzen ergeben sich aus den möglichen Resultaten?
5. Wen interessiert das Resultat?

1.3.5.1 Offene und geschlossene Grundgesamtheiten

Ein anderes Hemmnis, zu Zufallsstichproben zu gelangen, die diesen Namen voll verdienen, besteht darin, dass wir eben keine Urne mit definiertem Inhalt vor uns haben, sondern eine nicht abgegrenzte OFFENE Grundgesamtheit in der Zeit. In der Praxis wird eine abgeschlossene Grundgesamtheit mit festen Parametern postuliert; auf diese fiktive Grundgesamtheit - man hofft, sie sei repräsentativ für die offene Grundgesamtheit - beziehen sich dann alle Schlüsse, die die Zufallsstichprobe(n) gestatten, wobei angenommen wird, mögliche **Selektionseffekte** seien zu vernachlässigen. Erinnerung sei, dass - etwas vereinfacht dargestellt - bei einem Experiment der „Urneninhalt“ überprüfbar geschaffen wird, stochastischen Prinzipien unterworfen ist und nur Stichproben möglich sind, um eine „Warum“-Frage zu beantworten. Demgegenüber interessiert bei einer Erhebung die Beantwortung einer „Wie viel“-Frage: historisch einmalig liegt eine Grundgesamtheit vor, die direkt oder anhand einer Stichprobe zunächst deskriptiv analysiert wird. Der Unterschied zwischen Experiment und Erhebung verschwindet, wenn wir die jeweilige Grundgesamtheit, sei sie endlich oder unendlich, durch ihre Verteilungsfunktion $F(x)$ charakterisieren.

1.3.6 Stichproben

Aufschluss über das Vorkommen und die Verteilung der interessierenden Merkmale in der Grundgesamtheit, die man häufig aus finanziellen, zeitlichen oder prinzipiellen Gründen nicht als Ganzes untersuchen kann, erhält man anhand einer **Stichprobe**, wie sie z. B. die Kostprobe, der Probe-Einstich in den Hochofen und die Entnahme einer Blutprobe durch eine eingestochene Hohnadel darstellen. **Stichprobenverfahren** sind gegenüber Vollerhebungen:

- billiger,
- schneller durchführbar, ohne dass sich Umstände und Rahmenbedingungen ändern;
- häufig ist wegen der kleineren Zahl die Erhebung und die Aufbereitung sorgfältiger,
- die Resultate sind schneller verfügbar (aktueller);
- in speziellen Situationen sind Stichprobenerhebungen unerlässlich etwa bei Ernteschätzungen oder bei komplizierten Untersuchungen, die z.B. Interviewer voraussetzen; zur Kontrolle der Elemente einer laufenden Produktion sowie bei Lebensdauer-Härte-Tests von Maschinen und bei der zerstörenden Prüfung, etwa von Munition und Maschinen.

Vollerhebungen sind wichtig zur: (1) Planung von Stichproben-Erhebungen, (2) regionalen Detaillierung und zur Erfassung von Minderheiten sowie zur (3) Aktualisierung früherer Vollerhebungen.

Zur Erhebung typischer Einzelfälle: Es gibt Fälle, in denen weder die repräsentative Erhebung noch die Vollerhebung möglich ist. Dann wird man bewusst nach objektiv-sachlogischen Kriterien ausgewählte typische Einzelfälle erheben, um ungefähre Aufschlüsse über die unbekannte Grundgesamtheit zu erhalten. Außerdem bedient man sich der Erhebung typischer Einzelfälle zur Planung umfangreicher Erhebungen sowie nach umfangreichen Erhebungen zur Vertiefung spezieller bisher gewonnener Erkenntnisse. Verallgemeinerungsfähige Resultate sind so aber kaum zu erzielen, wohl aber erste Hinweise.

1.3.7 Zufallsstichproben

Wenn es nicht sinnvoll ist, alle Einheiten einer Grundgesamtheit zu untersuchen, wählt man aus der Grundgesamtheit Einheiten aus, und zwar so, dass alle Einheiten voneinander unabhängig und mit der gleichen Wahrscheinlichkeit in die so genannte **Zufallsstichprobe** gelangen können. Das Auswahlverfahren für die Stichprobeneinheiten muss unabhängig von dem zu untersuchenden Merkmal sein. Man benutzt z. B. eine Tafel von Zufallszahlen: Die Einheiten der Grundgesamtheit werden nummeriert; mit Hilfe der Zufallszahlen werden die in die Zufallsstichprobe aufzunehmenden Einheiten festgelegt. Die Zufallsstichprobe der Merkmalswerte steht jetzt stellvertretend für die Grundgesamtheit der Merkmalswerte aller Untersuchungsobjekte, über die Aussagen gewünscht werden. Eine stark untergliederte oder umfangreiche heterogene Grundgesamtheit wird durch eine kleine Stichprobe kaum hinreichend repräsentiert. Bei kleinen Stichproben wird es häufig sinnvoll sein, weitgehend informell zu befragen, damit auch Unerwartetes erfasst werden kann. Auch bei gesicherter Repräsentativität der kleinen Stichprobe ist die Verallgemeinerungsfähigkeit der Resultate zu prüfen.

Nach einem Zufallsverfahren entnommene Stichproben haben den Vorzug, dass die aus ihnen ermittelten statistischen Kenngrößen, zum Beispiel der Mittelwert, gegenüber den entsprechenden „wahren“ Werten der Grundgesamtheit nur die unvermeidlichen **zufälligen Abweichungen** aufweisen; da sie das Resultat nicht verzerren - bei mehrfachen Wiederholungen gleichen sich zufällige Abweichungen im Mittel aus - können zufällige Abweichungen geschätzt werden. Bei den Verfahren ohne Zufallsauswahl (d.h. Selektion) können noch so genannte methodische oder **systematische Fehler** hinzukommen, über deren Größe sich in der Regel kaum Angaben machen lassen. Enthält die Stichprobe systematische Fehler, dann ist sie nicht repräsentativ für die Grundgesamtheit.

Einige Beispiele: Obst-Prachtexemplare im Schaufenster brauchen nicht der durchschnittlichen Qualität des für den Verkauf bestimmten Obstes zu entsprechen. Nicht- oder Falschbeantwortung von Interviewerfragen oder Fragebogen, etwa unrichtige Altersangaben aus Nachlässigkeit (Bevorzugung der mit 0 oder 5 endenden Lebensjahre), Eitelkeit oder Betrug; Falschbeantwortung zur Täuschung, um der befragenden Stelle gefällig zu sein oder um sie zu ärgern oder auch aus



Unsicherheit über den möglichen Zweck der Erhebung. Beim Experimentieren: Messungen mit falsch geeichten Maßen oder Messgeräten wie Waage, Uhr, Tacho, letzteres mit proportionalem systematischen Fehler. Methodische Fehler dieser Art sind leider allgegenwärtig und nur durch Sachkenntnis auf dem speziellen Arbeitsgebiet zu vermeiden.

In der Statistik interessieren die in repräsentativen Stichproben auftretenden **zufälligen Abweichungen**, ihre Abschätzung und die Prüfung, ob in den Stichproben beobachtete Erscheinungen auch für die Grundgesamtheiten charakteristisch sind oder lediglich als Zufallsergebnisse gelten können (Prüfung von Hypothesen über die Grundgesamtheit). Aus den Stichproben-Kennwerten wird auf die Kennwerte der Grundgesamtheit geschlossen. Die Kennwerte der Grundgesamtheit werden Parameter genannt und meist mit griechischen Buchstaben geschrieben. So wird von dem Stichprobenmittelwert $\bar{x}$ (lies: x quer) auf den Mittelwert der Grundgesamtheit, auf den Parameter μ (mü) geschlossen. Je mehr Einheiten n die Zufallsstichprobe umfasst, desto größer ist die Genauigkeit dieses Schlusses. Für den Schluss auf die Grundgesamtheit dienen Konfidenzintervalle (Vertrauensbereiche) und statistische Tests.

1.4 Datenanalyse

- Klassierung von Merkmalen
- Skalierung von Variablen
- Daten

1.4.1 Klassierung von Merkmalen

Merkmale, über deren Verteilung in einer Grundgesamtheit Aussagen zu machen sind, die das Ziel der Untersuchung sind, nennt man **Zielgrößen**. Andere Merkmale, die an den Merkmalsträgern auftreten und die in einem funktionalen Zusammenhang zur Zielgröße stehen, oder Merkmalsausprägungen, die man als Behandlungen den Merkmalsträgern zuordnen kann und die die Ausprägungen der Zielgrößen beeinflussen, nennt man **Einflussgrößen**. Nicht quantifizierbare und damit nicht kontrollierbare und im Modell nicht berücksichtigte Einflussgrößen nennt man **Störgrößen**; sie sind verantwortlich für die zufallsabhängige Streuung der Untersuchungsergebnisse.

Wird für einen bestimmten Wagentyp die Abhängigkeit der Länge des Bremsweges von der Geschwindigkeit untersucht, so ist diese eine Einflussgröße. Die Bremsweglänge ist die Zielgröße. Weitere Einflussgrößen können wegen der unerwünschten Variabilität, die sie auf die Zielgröße ausüben (zum Beispiel Reifentyp, Straßenbelag, Nässe) als Störgrößen miterfasst werden. Als **Identifikationsgrößen** kommen schließlich Nummer des Wagens bzw. Nummer der Mannschaft (Wagen + Fahrer) infrage. Je nach der Fragestellung wird man bestimmte Einflussgrößen als Störgrößen auffassen. Liegen Zufallsstichproben vor, so spricht man besser von **Variablen** als von Größen und Merkmalen.

Man unterscheidet Variablen, die Einfluss ausüben und Variablen, die beeinflusst werden und präzisiert die interessierenden Variablen:

- Definition, Bezeichnung, Kurzbezeichnung, Skalentyp, Messgenauigkeit und angemessene Stellenzahl, Verschlüsselung nicht-numerischer Daten. Maßeinheiten, Standards, interne und externe Kontrollen, Plausibilität, Bezugsrahmen und Vergleichbarkeit der Daten.
- Rangordnung sowie Abhängigkeiten. Welche Variablen sind besonders wichtig, gleichberechtigt, gemeinsam zu betrachten, zu einer neuen Variable zusammenzufassen oder voneinander abhängig? Welche Strukturen sind zwischen Variablen erkennbar? Welche unabhängigen Variablen erklären einige oder sämtliche abhängigen Variablen?

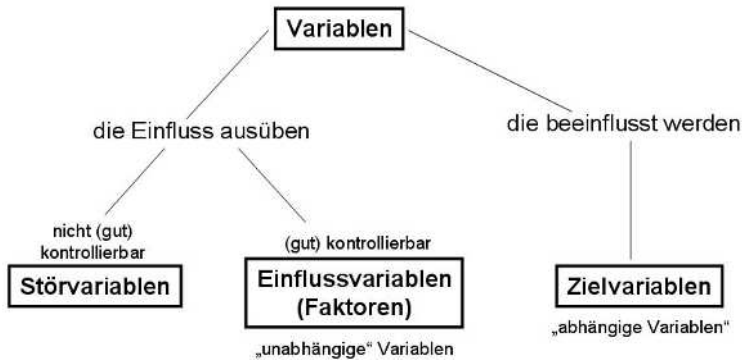


Abb. 1.2. Klassierung von Variablen nach der Problemstellung

1.4.2 Skalierung von Variablen

Unter **Skalierung** versteht man die vollständige Klassifizierung einer Gesamtheit von Objekten nach gemeinsamen Merkmalen durch Zuordnung von Zeichen zu den Objekten. Die mehr oder weniger geordnete Menge der möglichen Zeichen heißt **Skala**. Jeder Merkmalsausprägung entspricht ein Skalenwert. Die Skala ist dabei quasi ein Modell des Merkmals, das die Relation zwischen den Objekten bezüglich der Ausprägungen des betrachteten Merkmals adäquat widerspiegeln und formalen Operationen, vor allem mathematisch-statistischen Verfahren, zugänglich machen soll. Eine Skala muss disjunkt und erschöpfend sein.

Liegen zwei Merkmalsausprägungen vor, so können diese (1) gleich oder ungleich sein (**Nominalskala**), (2) der Größe nach geordnet werden (**Rang- oder Ordinalskala**), (3) der Abstand zwischen beiden kann gemessen werden (**Intervall- und Verhältnis-Skala**, letztere mit echtem Nullpunkt - siehe Abbildung 1.3 und Übersicht 3).

Messen ist die Bestimmung einer Merkmalsausprägung an einem Merkmalsträger, wobei der Merkmalsausprägung eine Zahl zugeordnet wird. Dieses gilt für Merkmalsausprägungen, die intervall- oder verhältnisskaliert sind. Auf dem Niveau der Nominal- und Ordinalskalen spricht man nicht von Messungen sondern von der **Zuordnung einer Qualität** ohne bzw. mit natürlicher Reihenfolge: so kann ein Käfer durch die Merkmalsausprägungen „aus Taiwan“, „von blauer Farbe“ und „linsengroß“ allein durch die Benennungen charakterisiert werden. Die Skala schulischer Zensuren ist eine Rangskala, deren Aussagekraft auf die Gruppe der beurteilten Schüler beschränkt ist. Vergleichende Aussagen zwischen mehreren Schulklassen sind nicht zulässig; selbst dann kaum, wenn die Zensuren vom selben Lehrer vergeben worden sind.

Intervall- und Verhältnisskala bilden die so genannte **„metrische Skala“**. Nominal- und Ordinalskala heißen dagegen auch **„nichtmetrisch“**. Nichtmetrische Merkmale nennt man „Gruppierungsmerkmale“, die sich aus metrischen Merkmalen auch durch die Bildung von Klassen gewinnen lassen.

Alternativmerkmale, dichotome Merkmale (zweiwertige oder binäre (0/1) Merkmale) und geeignet transformierte Ränge von ordinalen Merkmalen lassen sich praktisch als metrische Merkmale behandeln.

Nach ihrer Berufszugehörigkeit gefragte Versuchspersonen lassen sich in keine eindeutige und objektive Reihenfolge bringen. Klassifizierungen dieser Art – wir sprechen von der **Nominalskala** – liegen bei der Aufstellung von Geschlechts-, Berufs-, Sprach- und Nationalitätengruppen vor. Häufig bietet sich eine zweckbestimmte Ordnung an: Beispielsweise, wenn die Untersuchungsobjekte nach dem Alter oder nach einem anderen Merkmal in eine objektive Reihenfolge gebracht werden, wobei jedoch die Abstände auf der **Rangskala oder Ordinalskala** keine echten „Real-

abstände" darstellen: So kann bei einer nach dem Alter orientierten Rangskala auf einen Zwanzigjährigen ein Dreißigjähriger und dann weiter ein Zweiunddreißigjähriger folgen.

Sind aufeinanderfolgende Intervalle konstant, es sei an die konventionelle Temperaturmessung nach Celsius gedacht, so ermöglicht die **Intervallskala** noch keinen sinnvollen Vergleich: Es ist unkorrekt, zu behaupten, dass zehn Grad Celsius doppelt so warm seien wie fünf Grad Celsius.

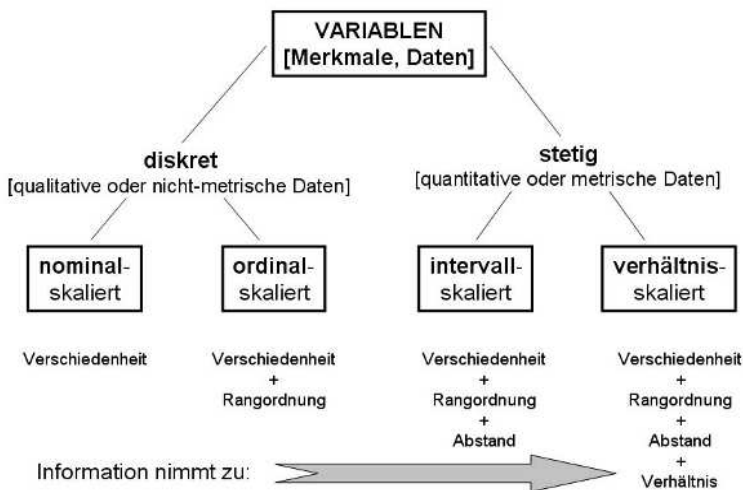


Abb. 1.3. Skalentypen zur Messbarkeit von Beobachtungen

Erst eine Intervallskala mit **absolutem Nullpunkt** lässt sinnvolle Vergleiche zu. Merkmale, für die ein solcher Nullpunkt angegeben werden kann, sind etwa Temperaturmessung in Kelvin, Länge, Gewicht und Zeit. Skalen dieser Art sind die leistungsfähigsten, sie werden als **Verhältnisskalen** bezeichnet. Während sich eine Verhältnisskala durch Multiplikation mit einer positiven Konstanten in eine andere überführen lässt, beispielsweise 1 US-Mile = 1,609347 mal 1 Kilometer, d. h. $y = ax$ – wobei das Verhältnis zweier numerischer Beobachtungen bei der Multiplikation mit einer Konstanten erhalten bleibt – ändert es sich bei Einheiten der Intervallskala [z. B. Umrechnung von x Grad Celsius in y Grad Fahrenheit: $y = ax + b$ mit $a = 9/5$ und $b = 32$; übrigens $-40^\circ\text{C} = -40^\circ\text{F}$; zu Kelvin: $0^\circ\text{C} = 273,15\text{ K}$ und $0\text{ K} = -273,15^\circ\text{C}$].

Weitere Beispiele sind:

1. **Nominalskala:** Autonummern und Postleitzahlen (willkürliche Numerierung) sowie Familienstand, Farbangaben und Berufsbezeichnungen bzw. Häufigkeitsdaten.
2. **Rangskala:** Schulnoten u. a. Daten, die (wie Bonituren zugeteilt werden und) eine Rangordnung ausdrücken.
3. **Intervallskala** (Nullpunkt konventionell festgelegt; Intervalle mit empirischem Sinn; direkte Quotientenbildung verboten): Kalenderdatum, Intelligenzquotient, Temperaturmessung in Grad Celsius oder Fahrenheit.
4. **Verhältnisskala** (mit wahrem Nullpunkt): Temperaturmessung in Kelvin, die Einheiten der Physik wie m, kg, s.

Hinweis: In dieser Auflistung von 1. bis 4. nehmen Informationsgehalt und Empfindlichkeit gegenüber Messfehlern zu.

1.4.2.1 Zulässige Skalentransformationen

Das Skalenniveau entscheidet darüber wie die Zuordnung der Merkmalsausprägungen zu den Zahlen erfolgen soll, welche **Transformationen** der Zahlen erlaubt sind und letztlich welche **Operationen** mit den Zahlen sinnvoll (bedeutungsvoll) sind. Je nach Skalenniveau können den Beobachtungen mehr oder weniger beliebige Zahlen zugeordnet werden bzw. die zugeordneten Zahlen transformiert werden. Wichtig ist, dass nach einer Transformation die relevanten Beziehungen / Verhältnisse in den Beobachtungen erhalten bleiben.

- Bei der Nominalskala muss bei einer Transformation lediglich die Information über Gleichheit und Verschiedenheit erhalten bleiben. Es können also den Objekten beliebige Zahlen zugeordnet werden, solange alle Objekte derselben Klasse dieselben Zahlen, und alle Objekte einer anderen Klasse eine andere Zahl erhalten (**eindeutige Zuordnung**).
- Bei der Ordinalskala muss die Information über die Ordnung erhalten bleiben. Somit können den Objekten beliebige Zahlen zugeordnet werden, solange Objekte mit einer größeren Merkmalsausprägung auch eine größere Zahl erhalten (**streng monotone Transformationen**).
- Bei der Intervallskala muss auch die Information über Abstände zwischen Merkmalsausprägungen erhalten bleiben. Bei Transformationen müssen also den Objekten Zahlen so zugeordnet werden, dass die Zahlen die Information über die Abstände zwischen den Merkmalsausprägungen korrekt wiedergeben (**lineare Transformationen** der Form $y = a + bx$).
- Bei der Verhältnisskala muss zusätzlich die Information über die Verhältnisse der Merkmalsausprägungen erhalten bleiben. Nur die Multiplikation mit einer Konstanten ist erlaubt: „**Einheitentransformation**“ (Streckung / Stauchung) $y = bx$ ($b > 1$ / $b < 1$), z.B. Länge in Metern = Länge in Fuss mal 0,3048.

1.4.3 Daten

Aufzeichnungen wie Messergebnisse oder Rechenresultate sowie persönliche Angaben (z. B. Geschlecht, Name, Größe) bezeichnet man als „**Daten**“; es sind Folgen von Zeichen für einen bestimmten Zweck. In der Statistik gelten Daten als realisierte Merkmalsausprägungen. Durch **Anwendung von Interpretationsvorschriften** auf Daten - etwa: die Messergebnisse betreffen ..., die Berechnungen beziehen sich auf ..., die untersuchten Personen bilden ..., erhält man aus den Daten **Informationen**. Entsprechend der Ausführungen aus dem vorangehenden Abschnitt unterscheidet man auch Daten nach zulässigen Operationen (vgl. Übersicht 3).

Mit den durch Zahlen verschlüsselten (codierten) Informationen (**Verschlüsselungen**) - es liegen **Nominaldaten** vor - kann nicht sinnvoll gerechnet werden. Neben dem Prüfen auf Gleichheit kann unter bestimmten Aspekten auch ein Sortieren, z.B. nach der Häufigkeit oder nach inhaltlichen Überlegungen heraus, sinnvoll sein.

Ordinaldaten, wie z.B. Schulnoten, entstehen aus eindeutig definierten Reihenfolgen. Für diese sind die Grundrechenarten nicht definiert. Möglichkeiten der Beschreibung und Bewertung von Ordinaldaten ergeben sich aus der natürlichen Anordnung der Beobachtungen, zum Beispiel hinsichtlich Minimum, Maximum, Medianwert und ausgezeichneten Quantilen.

Metrische Daten entstehen aus Messungen oder Zählungen. Auf der Grundlage einer eindeutigen Definition (Vorgehen bzw. Verfahren) und Skalierung (Einheit) kann mit metrischen Daten gerechnet werden. Insbesondere können Maßzahlen wie Mittelwerte oder Standardabweichungen bestimmt werden.

Übersicht 3. Abgrenzung zu den vier Skalentypen

Skalentyp	Informationswert	Hinweise	Beispiele
Nominalskala	$A = B; A \neq B$ Gleichheit und Ungleichheit von Objekten	Klassifizierung / Etikettierung qualitativer Merkmalsausprägungen: mit Namen versehene ungeordnete Werte: reelle Zahlen sind als Schlüsselziffern (Codierungen) möglich	Geschlecht, Vorname, Beruf, Nationalität, Parteizugehörigkeit, Postleitzahl
Ordinalskala	$A = B; A \neq B$ zusätzlich $A < B < C$ Rangordnung	Geordnete Werte: Merkmalsausprägungen sind anordbar , d.h. mit Größer- oder Kleiner-Relation. Rangreihe: reelle Zahlen sind als Rangordnungsziffern möglich	Schulnoten, Betriebsklima - Ränge: im Sport, beim Militär, Güteklassen bei landwirtschaftlichen Erzeugnissen
Intervallskala	$A = B; A \neq B$ zusätzlich sind die Differenzen $d = A - B$ vergleichbar	Skala mit konstanten Abständen und willkürlichem Nullpunkt: Subtraktion und Addition reeller Zahlen sind zugelassen	Temperatur, gemessen in °Celsius und °Fahrenheit, Kalenderdatum, Punktwerte im Intelligenztest, Höhenangaben
Verhältnisskala	$A = B; A \neq B$ zusätzlich sind die Differenzen $d = A - B$ und die Quotienten $c = A : B$ sind vergleichbar	Skala mit echtem Nullpunkt, d.h. exakt vergleichbare Messverhältnisse: Division und Multiplikation reeller Zahlen sind zugelassen, d.h. Prozentangaben sind möglich	Länge, Fläche, Volumen, Gewicht, Alter, Temperatur gemessen in Kelvin, Stromstärke, Kosten, Gewinn

Kalenderdaten dienen z.B. zur Berechnung des Alters, einer Verweildauer im Krankenhaus oder der Überlebenszeit. Mitunter sind Überlebenszeiten von Patienten durch konkurrierende Ereignisse, etwa tödliche Verkehrsunfälle, nicht beobachtbar; oder sie sind nicht beobachtbar, weil die Studie vorzeitig abgeschlossen wird.

1.4.3.1 Datengewinnung und Datenstruktur

Entsprechend ihrer Gewinnung unterscheidet man:

1. Daten, die bereits vorliegen und die weder experimentell unter Randomisierung entstanden noch als Zufallsstichprobe einer Erhebung aufzufassen sind.
2. Daten, gewonnen anhand eines Experimentes mit zufälliger Zuordnung der Untersuchungseinheiten zu den Bedingungen (Randomisierung).
3. Daten, gewonnen aufgrund einer Erhebung als Zufallsstichprobe aus einer definierten Grundgesamtheit.

Nach (2) oder (3) gewonnene Daten lassen sich als **Realisierungen von Zufallsvariablen** auffassen.

Bei der **Datengewinnung** ist zu beachten: 1) die richtigen Merkmale bzw. Variablen auswählen, 2) sie standardisiert und exakt messen und 3) hinreichend viele Daten gewinnen. Wichtig sind

hierbei und danach **Prüfungen** auf Vollzähligkeit, inhaltliche Vollständigkeit, formale Richtigkeit und logische Richtigkeit, letztere auch anhand bereits vorliegender Daten anderer Untersucher. Kombiniert man Daten, so sind **Fehler** der Datengewinnung, der Datenkombination (Datenagglomeration) und deren Interpretation zu unterscheiden.

Daten werden als **Datenmatrix (Tabelle)** übersichtlich dargestellt: ihre n Zeilen repräsentieren die Merkmalsträger (Fälle) - Objekte wie Untersuchungseinheiten oder Personen -, ihre p Spalten die Merkmale oder Variablen. Die Zellen dieser $n \times p$ - Tabelle enthalten die in der Untersuchung gewonnenen Merkmalsausprägungen. Eine andere Bezeichnung hierzu sind **empirischen Daten**, d.h. diejenigen Werte, die in einem vorliegenden (konkreten) Fall von den Variablen angenommen werden, die Realisierungen dieser Variablen.



	age	sex	height	weight	bmp	fev1	rv	frc
1	7	0	109	13.1	68	32	258	183
2	7	1	112	12.9	65	19	449	245
3	8	0	124	14.1	64	22	441	268
4	8	1	125	16.2	67	41	234	146
5	8	0	127	21.5	93	52	202	131
6	9	0	130	17.5	68	44	308	155
7	11	1	139	30.7	89	28	305	179
8	12	1	150	28.4	69	18	369	198
9	12	0	146	25.1	67	24	312	194
10	13	1	155	31.5	68	23	413	225
11	13	0	156	39.9	89	39	206	142
12	14	1	153	42.1	90	26	253	191

Abb. 1.4. Datentabelle - Beispiel in R: `data(cystfibr)` aus `library(ISwR)` mit dem Befehl `edit(cystfibr)`

Beispiel: Die Abbildung 1.4 zeigt eine Datentabelle zu einer Untersuchung der **Lungenfunktion** bei zystischer Fibrose (aus Dalgaard [Dal05]). Neben dem Alter (Jahre), dem Geschlecht (kodiert: 0-männlich und 1-weiblich), der Größe (cm), dem Gewicht (kg) und dem Bodymaß (% von der Norm) werden in diesem Beispiel die Ergebnisse der Lungenfunktionsprüfung, z.B. das „forcierte Ausatmungsvolumen“ (fev1), das „Residualvolumen“ (rv) und die „Residualkapazität“ (frc) tabelliert.

Für metrische Daten ist festzulegen, wie viele **Kommastellen** (Genauigkeit) erforderlich sind (z. B. für das Gewicht eine).

Rangdaten werden durch Rangziffern oder Rangzahlen in der Datentabelle dokumentiert (z.B. 1 - 6).

Nominaldaten müssen verschlüsselt erfasst werden. So benutzt man zur **Verschlüsselung des Berufes** einen dreistelligen Schlüssel, etwa Damenschneider(in) = 402 und Buchbinder(in) = 708. Ist der Beruf unbekannt oder listenmäßig nicht erfasst, dann muss auch für diesen Sachverhalt ein fester Schlüssel verwendet werden, z. B. 999. Bei der Vergabe von Schlüsselziffern sind sämtliche möglichen Merkmalsausprägungen zu erfassen (man denke auch an „Sonstiges“ z. B. für Berufe, die wir nicht kennen) und Überschneidungen auszuschließen.

Für die folgenden speziellen Sachverhalte sind **eindeutige Regelungen** (Kodierungen) festzulegen:

- für den betreffenden Fall nicht zutreffend (z. B. Beruf für ein Kind oder Anzahl der Schwangerschaften für einen Mann),
- nicht untersucht (fehlende Angabe),
- nicht feststellbar, obwohl untersucht,
- unklar, ob untersucht oder negativ (o.B. ohne Befund).



Grundlagen aus der Mathematik

- Logische und relationale Operatoren
- Mengen
- (Grund-)Rechenarten
- Einführung in die Matrixalgebra
- Funktionen
- Kombinatorik

In diesem Kapitel werden einige **elementare mathematische Kenntnisse** wiederholt. Sie bilden mit einigen Ausnahmen (insbesondere hinsichtlich einer kurzen Einführung in den Umgang mit Matrizen) einen Teil des für die mittlere Reife geforderten Wissens. Diese Kenntnisse reichen vollauf für das Verständnis der in den weiteren Kapiteln behandelten Probleme.

2.1 Logische und relationale Operatoren

Die Sprache der Mathematik verwendet Symbole, z. B. Buchstaben oder andere Zeichen, um bestimmte Sachverhalte präzise und kurz darzustellen. Zahlen werden in der Algebra im allgemeinen mit kleinen lateinischen Buchstaben ($a, b, c, d, \dots$) oder, wenn sehr viele unterschieden werden sollen, mit $a_1, a_2, a_3, \dots, a_n$ bezeichnet.

Für die Darstellung bestimmter Sachverhalte in den Definitionen, Gesetzmäßigkeiten und Beispielen werden Symbole zu den Beziehungen (**relationale Operatoren, Relationen**) zwischen zwei Zahlen verwendet, die in Tabelle 2.1 zusammengestellt sind.

Tabelle 2.1. Einige mathematische Relationen

Beziehung	Bedeutung	Beispiel
$a = b$	a ist gleich b	$8 = 12 - 4$
$a < b$	a ist kleiner als b	$4 < 5$
$a > b$	a ist größer als b	$6 > 5; -5 > -6$
$a \leq b$	a ist gleich oder kleiner als b	Verdienst a beträgt höchstens b Euro
$a \geq b$	a ist gleich oder größer als b	Verdienst a beträgt mindestens b Euro
$a \simeq b$	a ist angenähert, nahezu	$109,8 \simeq 110$
$a \approx b$	gleich, ungefähr gleich b	$109,8 \approx 110$
$a \neq b$	a ist nicht gleich b	$4 \neq 6$

Für „x ist größer als a und kleiner oder gleich b“ schreibt man: $a < x \leq b$

Für „x ist wesentlich größer als a“ schreibt man: $x \gg a$

Aus der Ungleichung $a > b$ folgt $-a < -b$ und (für $b > 0$) $1/a < 1/b$.

Logische Operatoren werden verwendet, um Zusammenhänge und logische Aussagen kurz und prägnant in mathematischen Darstellungen und Abhandlungen wiedergeben zu können. Eine Zusammenstellung der wichtigsten logischen Symbole enthält Tabelle 2.2.

Tabelle 2.2. Einige wichtige logische Operatoren

Operator	Bedeutung
A, B, C	Aussagen können ausführlich formuliert und symbolisch abgekürzt werden
$\top$	zeigt an, dass eine Aussage „ wahr “ ist (vgl. engl. True)
$\perp$	zeigt an, dass eine Aussage „ falsch “ ist
$\wedge$	Verknüpfung von Aussagen mit „ und “ - Konjunktion (geklammert)
$\vee$	Verknüpfung von Aussagen mit „ oder “ - Disjunktion (getrennt)
$\neg$	verneinen einer Aussage mit „ nicht “ - Negation
$\Rightarrow$	Folgerung einer neuen Aussage - „ Implikation “
$\Leftrightarrow$	Gleichheit zweier Aussagen - „ Äquivalenz “
$\forall$	steht für „ alle “
$\exists$	steht für „ es gibt ein “

2.2 Mengen

- Begriffsbildung
- Mengenoperationen

2.2.1 Begriffsbildung

Das gedankliche Modell einer **Menge** ist eines der wichtigsten Voraussetzungen zum Verständnis der '**neuen**' **Mathematik**, die sich im frühen 20. Jahrhundert entwickelte. Es ermöglicht einerseits eine neue Sicht auf die Eigenschaften der Zahlen (Zahlenmengen wie natürliche Zahlen, rationale Zahlen, reelle Zahlen) sowie Zuordnungen (Funktionen) zwischen Zahlenmengen. Andererseits eröffnet dieses Modell neue Möglichkeiten in der Darstellung und Behandlung komplexer und abstrakter Zusammenhänge, zum Beispiel in der Wahrscheinlichkeitsrechnung (Ereignismengen). Der Mathematiker **Georg Cantor** (1845-1918) definierte eine Menge als eine Zusammenfassung von **wohlunterscheidbaren Objekten (Elementen)** zu einem Ganzen. Die Unterscheidbarkeit garantiert, dass für jedes Element zu entscheiden ist, ob es zu einer Menge gehört oder nicht. In einer Menge gibt es somit keine identischen Elemente. Die **Mächtigkeit** einer Menge $n = |A|$ bezeichnet die Zahl der Elemente in der Menge. Diese ist nicht notwendig endlich. Beispiele zu Mengen sind:

$$\begin{aligned}
 A &= \{a, b, c, d, e, f\} \\
 N &= \{1, 2, 3, 4, 5, 6, \dots\} \\
 G &= \{x \mid x \in N, x \text{ ist durch } 2 \text{ teilbar}\}
 \end{aligned}
 \tag{2.1}$$

Der griechische Buchstabe **Epsilon** $\in$ wird verwendet, um die Zugehörigkeit von einzelnen Elementen zu einer Menge zu kennzeichnen: $5 \in N$ oder $d \in A$ aber $5 \notin A$.

Eine Menge **C ist in A enthalten** ($C \subseteq A$), wenn jedes Element von C auch in A enthalten ist.

$$\begin{aligned}
 A = B &\Leftrightarrow \forall x \in A \Rightarrow x \in B \quad \text{und} \quad \forall y \in B \Rightarrow y \in A \\
 C \subseteq A &\Leftrightarrow \forall x \in C \Rightarrow x \in A
 \end{aligned}
 \tag{2.2}$$

Damit ist jede Menge in sich selbst enthalten. **D ist eine echte Teilmenge von A** ($D \subset A$, vgl. Abb. 2.1), wenn es Elemente in A gibt, die nicht zu D gehören. Die Analogie zu den Relationen zwischen Zahlen hinsichtlich kleiner ($<$) bzw. kleiner oder gleich ($\leq$) ist offensichtlich. Zwei

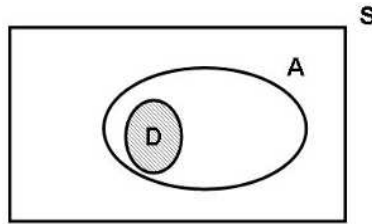


Abb. 2.1. Venn-Diagramm zu Darstellung von Mengen ($D \subset A$)

Mengen **A und B sind gleich** ($A = B$), wenn jedes Element von A in B enthalten ist und umgekehrt.

Ein Sonderfall, der in dieser Terminologie nicht leicht einsehbar ist, ist **die leere Menge** ($\emptyset$ oder $\{\}$), eine Menge, die keine Elemente enthält. Die leere Menge ist insbesondere eine Teilmenge jeder anderen Menge!

Ein weiterer Sonderfall ist die dem sicheren Ereignis entsprechende **Gesamtmenge** (S). Sie setzt den Rahmen für alle weiteren Betrachtungen, d.h. jede Menge wird als eine Teilmenge der Gesamtmenge gesehen.

Die Menge aller Teilmengen einer Menge A wird **Potenzmenge** genannt. Die **Mächtigkeit der Potenzmenge** P_A einer Menge A mit n Elementen ist $|P_A| = 2^n$. Sei zu Beispiel $A = \{a, b, c\}$, dann ist $P_A = \{\{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}, \{\}\}$. Diese hat die Mächtigkeit $2^3 = 8$; dabei ist nicht zu vergessen, dass jede Menge in sich selbst als Teilmenge aufgefasst werden kann und auch die leere Menge Teilmenge einer jeden anderen Menge ist.

2.2.2 Mengenoperationen

Für das Rechnen mit Mengen werden Mengenoperationen definiert.

Die **Vereinigung zweier Mengen A und B** enthält die Elemente, die **in A oder in B** oder in beiden Mengen enthalten sind. Dabei ist das mathematische "oder" gemeint, das in diesem Sinn nicht ausschließlich zu verwenden ist.

$$C = A \cup B = \{x | x \in A \text{ oder } x \in B\} \quad (2.3)$$

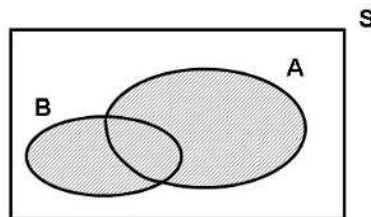


Abb. 2.2. Venn-Diagramm zur Vereinigungsmenge zweier Mengen ($A \cup B$)

Der Operator $\cup$ erinnert dabei an ein U (für das Wort Union, Vereinigung). Zu beachten ist insbesondere, dass

$$A \cup A = A \quad \text{und} \quad B \subset A \Rightarrow B \cup A = A \quad (2.4)$$

Die **Schnittmenge zweier Mengen A und B** enthält die Elemente, die **in A und in B** enthalten sind. Das Wort "und" wird hier im Sinn von „sowohl als auch“ verwendet.

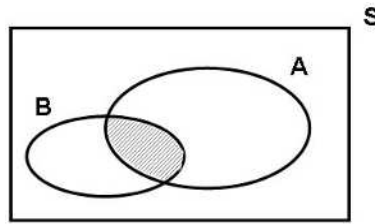


Abb. 2.3. Venn-Diagramm zur Schnittmenge zweier Mengen ($A \cap B$)

$$D = A \cap B = \{x | x \in A \text{ und } x \in B\} \quad (2.5)$$

Der Operator $\cap$ wurde als Gegensatz zum $\cup$ bei der Vereinigung gewählt. Haben die beiden Mengen A und B keine gemeinsamen Elemente, dann ist die Schnittmenge leer. A und B sind dann **disjunkt** genannt.

$$A \cap B = \emptyset \Rightarrow A \text{ und } B \text{ disjunkt} \quad (2.6)$$

Zu beachten ist insbesondere, dass

$$A \cap A = A \quad \text{und} \quad B \subset A \Rightarrow B \cap A = B \quad .$$

Die **Komplementärmenge** oder **das Komplement einer Menge A** enthält alle Elemente, die **nicht zu A** gehören. Der Bezug zur **Gesamtmenge S** ist somit zum Verständnis einer Komplementärmenge notwendig.

$$\bar{A} = \{x | x \notin A\} \quad (2.7)$$

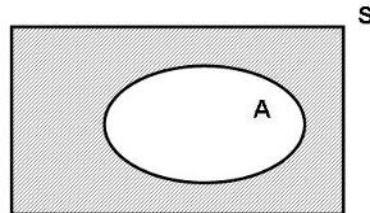


Abb. 2.4. Venn-Diagramm zum Komplement einer Menge ($\bar{A}$)

Die Vereinigungsmenge einer Menge A mit dem Komplement von $\bar{A}$ ergibt somit stets die Gesamtmenge S. Die Komplementärmenge der Gesamtmenge ist die leere Menge und umgekehrt.

$$A \cup \bar{A} = S \quad \text{und} \quad \bar{\bar{S}} = \{\}$$

Die **Restmenge** oder **logische Differenz** zwischen zwei Mengen A und B besteht aus allen Elementen, die zu A, aber nicht zu B gehören (Differenzmenge $A \setminus B$).

$$A \setminus B = \{x | x \in A \wedge x \notin B\} = A \cap \bar{B} \quad (2.8)$$

2.3 (Grund-) Rechenarten

- Summen und Produkte
- Potenzen und Wurzeln
- Logarithmen
- Rundungen
- Rechnen mit fehlerbehafteten Zahlen

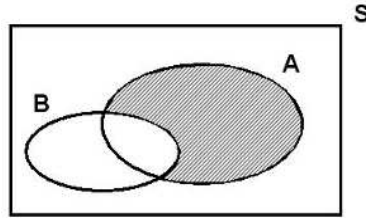


Abb. 2.5. Venn-Diagramm zur Mengendifferenz / Restmenge ($A \setminus B$)

Auf **René Descartes** (1596–1650) geht ein Teil der mathematischen Symbolik zurück: die Zeichen $+$, $-$ und $\sqrt{}$, die Potenzschreibweise sowie die Verwendung der Buchstaben $a, b, c, \dots$ als bekannte und $\dots x, y, z$ als unbekannte Größen. Von **Gottfried Wilhelm Leibniz** (1646-1716) stammen: Multiplikationspunkt, Divisionspunkte, das Integralzeichen $\int$ und die Begriffe Indices ($x_i; i = 1, \dots, n$), Konstante, Variable, Parameter sowie der Funktionsbegriff. Im 18. Jahrhundert führte **Leonhard Euler** (1707-1783) das Funktionssymbol $f(x)$ ein, den Buchstaben e für die Basis der natürlichen Logarithmen, das Symbol π für die Kreiszahl und das Summenzeichen $\sum$.

2.3.1 Summen und Produkte

Die Beherrschung der 4 Grundrechenarten: **Addition, Subtraktion, Multiplikation und Division** wird vorausgesetzt. Trotzdem seien die folgenden Vereinbarungen getroffen.

Eine Rechenoperation ist eine Vorschrift, aus zwei Zahlen eindeutig eine neue Zahl, z. B. die Summe oder die Differenz, zu bilden.

Addition: Summand + Summand = Ausgerechnete Summe [$5 + 8 = 13$]

Subtraktion: Minuend – Subtrahend = Ausgerechnete Differenz [$13 - 8 = 5$]

2.3.1.1 Das Summenzeichen

Soll die **Summe** der Zahlen $x_1, x_2, \dots, x_n$ gebildet werden, so wird für diese Operation das folgende Symbol $\sum_{i=1}^n$ eingeführt. $\sum$ ist der große griechische Buchstabe **Sigma**, das Zeichen für „Summe von“. Gelesen wird diese Operation: die Summe aller Zahlen x_i von $i = 1$ bis $i = n$. Der Index der ersten zu addierenden Größe wird dabei unter das Summenzeichen gesetzt, der Index der letzten Größe darüber. Allgemein wird die Summation vom Index 1 bis zum Index n geführt. Für die Summe von x_1 bis x_n sind also folgende Schreibweisen gleichwertig:

$$x_1 + x_2 + x_3 + \dots + x_n = \sum_{i=1}^{i=n} x_i = \sum_{i=1}^n x_i = \sum_i x_i = \sum x$$

$$\sum_{i=3}^5 x_i = x_3 + x_4 + x_5, \quad \text{d.h. } 5 - 3 + 1 = 3$$

Summanden

(2.9)



Im Folgenden werden Beispiele mit dem Programm **R** eingefügt. Sie sollen einerseits die behandelten mathematischen und statistischen Zusammenhänge verdeutlichen und zu eigenen Berechnungen anregen. Andererseits führen die Beispiele schrittweise in die vielfältigen Möglichkeiten dieses Programmes ein. Hilfreich zum Verständnis ist die Lektüre des Kapitels [9] (Einführung in R). Die Beispiele sind einheitlich vom laufenden Text des Buches hervorgehoben.

```

> 12 + 32                                # Addition
[1] 44
> 43 - 15                                # Subtraktion
[1] 28
> Zahlen <- c(2, 5, 7, 8, 9, 6)          # Werte in einem Vektor
> sum(Zahlen)                             # Summe
[1] 37

```

Hinweis zu der **Sprache der Statistik**: Merkmalswerte x_i sind Ausprägungen des Merkmals X beim i -ten Merkmalsträger ($i = 1, 2, \dots, n$) der Stichprobe des **Umfangs** n . Für die Summe der n Merkmalswerte x_i schreibt man $x_1 + x_2 + \dots + x_n = \sum_{i=1}^n x_i = \sum_{i=1}^n x_i$ oder einfach $\sum x$. Jede aus Merkmals- oder Beobachtungswerten x_i berechnete summarische Größe ist eine statistische Maßzahl, eine Stichprobenfunktion, auch „Statistik“ genannt: basiert sie auf einer Zufallsstichprobe, dann schätzt sie einen Parameter.



Um Ausdrücke mit Summen, wie zum Beispiel $\sum_{i=1}^n (3 + 2x_i + x_i^2) = 3n + 2 \sum_{i=1}^n x_i + \sum_{i=1}^n x_i^2$ berechnen zu können, helfen die folgenden Regeln:

$$\begin{aligned}
 \sum_{i=1}^n (x_i + y_i) &= (x_1 + y_1) + (x_2 + y_2) + \dots \\
 &= (x_1 + x_2 + \dots) + (y_1 + y_2 + \dots) \\
 &= \sum_{i=1}^n x_i + \sum_{i=1}^n y_i \\
 \sum_{i=1}^n kx_i &= kx_1 + kx_2 + \dots = k \sum_{i=1}^n x_i \quad \text{insb.} \quad \sum_{i=1}^n k = nk \\
 \sum_{i=1}^n (k + x_i) &= (k + x_1) + (k + x_2) + \dots = nk + \sum_{i=1}^n x_i \\
 \sum_{i=m}^n k &= (n - m + 1)k, \quad (m < n)
 \end{aligned} \tag{2.10}$$

Sind a und b reelle Zahlen so gilt:

$$\sum_{i=1}^n (ax_i - b)^2 = a^2 \sum_{i=1}^n x_i^2 - 2ab \sum_{i=1}^n x_i + nb^2 \tag{2.11}$$

Beispiel:

$$a = 2, b = 3, x_i = 4 \text{ und } 5, \text{ d. h. } n = 2$$



$$(2 \cdot 4 - 3)^2 + (2 \cdot 5 - 3)^2 = 4(16 + 25) - 2 \cdot 2 \cdot 3(4 + 5) + 2 \cdot 9 \quad 25 + 49 = 74 = 164 - 108 + 18$$

$$\left(\sum x_i \right)^2 = \sum x_i^2 + 2 \sum_{i < j} x_i x_j \quad \text{mit} \quad 1 \leq i < j \leq n \tag{2.12}$$

Beispiel:

$$x_i = 1, 2, 3; \quad (1 + 2 + 3)^2 = 36 = (1^2 + 2^2 + 3^2) + 2(1 \cdot 2 + 1 \cdot 3 + 2 \cdot 3)$$



Übersicht 4. Verbindungen der vier Grundrechenarten

Rechnen heißt, aus 2 oder mehreren Zahlen eine neue zu finden. Jedes der vier üblichen Rechenzeichen (+; −; ·; :) stellt eine Rechenvorschrift dar:

+	plus, Additionszeichen
−	minus, Subtraktionszeichen
·	mal, Multiplikationszeichen
:	geteilt durch, Divisionszeichen

Das Ergebnis jeder Rechnung sollte zu Beginn der Rechnung geschätzt, danach zweimal gerechnet und **anhand einer Probe kontrolliert** werden. Beispielsweise ist $4,8 + 16,1$ etwa gleich 21, genau 20,9; Probe $20,9 - 4,8 = 16,1$ oder $15,6:3$ ist etwa gleich 5, genau 5,2; Probe $5,2 \cdot 3 = 15,6$.

Für die **Reihenfolge der vier Grundrechenarten** gelten zwei Regeln:

1. Punktrechnung (*Multiplikation und Division*) geht vor *Strichrechnung* (*Addition und Subtraktion*).

$$\begin{aligned}\text{Beispiele: } 2 + 3 \cdot 8 &= 2 + 24 = 26 \\ 6 \cdot 2 + 8 : 4 &= 12 + 2 = 14\end{aligned}$$

Die positiven Zahlen (+1, +2, +3, +...), die Null und die negativen Zahlen (−1, −2, −3, −...) bilden die *ganzen Zahlen*, einen Zahlenbereich, in dem jede Subtraktionsaufgabe eine Lösung hat (z. B.: $8 - 12 = -4$). Bei der Punktrechnung sind folgende etwas salopp formulierte *Vorzeichenregeln* (Diophantos von Alexandria, um 250 n. Chr.) zu beachten:

$+$	$+$	$+$	Gleiche Vorzeichen
$+$	$:$	$+$	ergeben plus
$-$	$+$	$-$	$(-8) : (-2) = +4 = 4$
$-$	$:$	$-$	$(-8) : (+2) = -4$
$-$	$+$	$-$	$(-8) : (-2) = +4 = 4$
$-$	$:$	$+$	$(-8) : (+2) = -4$
			Rechenzeichen
			Vorzeichen

Der Wert einer reellen Zahl a , unabhängig von ihrem Vorzeichen, wird ihr *absoluter Betrag* genannt und $|a|$ geschrieben, z. B. $|-4| = | +4| = 4$.

2. Was in der Klammer steht, wird zuerst berechnet. Stecken mehrere Klammern ineinander, so ist mit der innersten Klammer zu beginnen. Vor einer Klammer verzichtet man im allgemeinen auf das Multiplikationszeichen, z. B.:

$$4(3 + 9) = 4(12) = 4 \cdot 12 = 48; \quad 9 - (7 - 3) = \left\{ \begin{array}{l} 9 - 7 + 3 \\ 9 - 4 \end{array} \right\} = 5$$

Die Division wird häufig als Bruch dargestellt, z. B.:

$$\begin{aligned}\frac{3}{4} &= 3/4 = 3 : 4 = 0.75 & \frac{1}{a} + \frac{1}{b} &= \frac{a+b}{a \cdot b} \\ 4[12 - (8 \cdot 2 + 18)] &= 4[12 - (16 + 18)] = 4(12 - 34) = 4(-22) = -88 \\ 12 \left[\frac{(9-3)}{2} - 1 \right] &= 12 \left[\frac{6}{2} - 1 \right] = 12(3 - 1) = 12(2) = 24\end{aligned}$$

Große Zahlen anschaulich gemacht :

$$\begin{array}{l} 1 \text{ Jahr} = \\ 8760 \\ \text{Stunden} \end{array} \quad \left\{ \begin{array}{l} 10^3 = \\ 10^6 = \\ 10^9 = \end{array} \right. \left. \begin{array}{l} 1000 \text{ Sekunden} \\ 1 \text{ Millionen Sekunden} \\ 1 \text{ Milliarde Sekunden} \end{array} \right\} \text{ sind rund } \left\{ \begin{array}{l} 17 \text{ Minuten} \\ 11 \frac{1}{2} \text{ Tage} \\ 32 \text{ Jahre} \end{array} \right.$$

[vgl. Billion: 10^{12} und Trillion: 10^{18}]

2.3.1.2 Spezielle Summen

Einige Sonderfälle sind beim Umgang mit Summen in der Statistik von besonderer Bedeutung.

1. Summe der ersten n **natürlichen Zahlen**:

$$\sum_{i=1}^n i = \frac{n(n+1)}{2} \quad (2.13)$$

2. Summe der ersten n **ungeraden Zahlen**:

$$\sum_{i=1}^n (2i-1) = n^2$$

3. Summe der ersten n **geraden Zahlen**:

$$\sum_{i=1}^n 2i = n(n+1)$$

4. Summe der **Quadrate** der ersten n natürlichen Zahlen:

$$\sum_{i=1}^n i^2 = \frac{n(n+1)(2n+1)}{6}$$

5. Summe der **dritten Potenzen** der ersten n natürlichen Zahlen:

$$\sum_{i=1}^n i^3 = \frac{1}{4}n^4 + \frac{1}{2}n^3 + \frac{1}{4}n^2 = \left[\frac{n(n+1)}{2} \right]^2$$

6. Summe der **vierten Potenzen** der ersten n natürlichen Zahlen:

$$\sum_{i=1}^n i^4 = \frac{1}{5}n^5 + \frac{1}{2}n^4 + \frac{1}{3}n^3 - \frac{1}{30}n$$

7. Summe der **fünften Potenzen** der ersten n natürlichen Zahlen:

$$\sum_{i=1}^n i^5 = \frac{1}{6}n^6 + \frac{1}{2}n^5 + \frac{5}{12}n^4 - \frac{1}{12}n^2$$

```
> 1:20
[1] 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
> sum(1:20)
[1] 210
```

Summen zu Rangzahlen

Summen über **unendlich viele Summanden** können feste, endliche Werte annehmen. Beispiele hierzu sind :

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n \frac{1}{i} \approx 0,5772156649 + \ln(n+1)$$

Eulersche In n ist der natürliche
Konstante Logarithmus von n

Das Symbol „lim“ steht dabei für den „Grenzwert“ und das Symbol „∞“ steht für „unendlich“, z.B. in $\lim_{i \rightarrow \infty} \frac{1}{i} = 0$.

$$\sum_{i=1}^{\infty} \frac{1}{i^2} \approx \pi^2/6 \left[\begin{array}{l} \text{Euler} \\ 1736 \end{array} \right] \text{ Die entsprechende Summe für } 1/i^3 \text{ ist unbekannt, für } 1/i^4 \text{ lautet sie } \pi^4/90.$$

$$\boxed{\sum_{n=1}^{\infty} \frac{1}{n} = +\infty \quad \sum_1^{\infty} \frac{n-1}{n!} = 2 \quad \sum_1^{\infty} \frac{1}{4n^2-1} = 1} \quad (2.14)$$

[zur Berechnung von $n!$ (n-Fakultät) vgl. (2.16)]

2.3.1.3 Multiplikation und Division; Fakultät

Multiplikation: Faktor $\times$ Faktor = Ausgerechnetes Produkt [$2 \times 3 = 6$]

Division: Dividend / Divisor = Ausgerechneter Quotient [$6/3 = 2$] (Divisor $\neq 0$)

Das Produkt zweier Zahlen wird nur selten durch das Zeichen $\times$ zwischen den beiden Faktoren charakterisiert, da eine Verwechslung mit dem Buchstaben x möglich ist; im allgemeinen deuten wir die Multiplikation durch einen hochgestellten Punkt an oder setzen die Faktoren ohne jedes Zeichen direkt nebeneinander, beispielsweise $5 \cdot 6$ oder pq . Die Aufgabe $1,23 \cdot 4,56$ schreibt man in den USA $1.23 \cdot 4.56$ oder $(1.23)(4.56)$, in England und Kanada $1 \cdot 23 \cdot 4 \cdot 56$ oder $1 \cdot 23 \times 4 \cdot 56$. Ein Komma wird in diesen Ländern zur übersichtlicheren Darstellung großer Zahlen verwendet (z. B. $5,837 \cdot 43$ bzw. $5,837.43$ anstatt $5837,43$). Das **Produktzeichen** $\prod$ ist wie folgt definiert:

$$\boxed{\prod_{i=1}^n x_i = x_1 \cdot x_2 \cdot \dots \cdot x_n} \quad (2.15)$$

und wird gelesen „**Produkt über** i “.

Speziell für das Produkt über die ersten n natürlichen Zahlen wird ein neues Symbol, gelesen **n-Fakultät**, eingeführt:

$$\boxed{n! = \prod_{n=1}^n n = n \cdot (n-1) \cdot \dots \cdot 3 \cdot 2 \cdot 1} \quad (2.16)$$

```
> 4 * 17                                # Multiplikation
[1] 68
> 56 / 8                                # Division
[1] 7
> Zahlen <- c(2, 3, 4, 5)
> prod(Zahlen)                           # Produkt
[1] 120
> 1:10
[1] 1 2 3 4 5 6 7 8 9 10
> prod(1:10)                             # Fakultät
[1] 3628800
```

2.3.2 Potenzen und Wurzeln

Potenzrechnung (Potenzieren): Ein Produkt gleicher Faktoren a ist eine Potenz a^n ; gesprochen: „ a hoch n “ oder „ n -te Potenz von a “. Hierbei ist a die Basis und n der Exponent der Potenz ($a^1 = a$).

$$\text{Basis}^{\text{Exponent}} = \text{Potenzwert} \qquad 2 \cdot 2 \cdot 2 = 2^3 = 8$$

```
> 2^4
[1] 16
> 12^4
[1] 20736
```

Potenzieren

Die zweiten Potenzen a^2 werden **Quadratzahlen** genannt, denn a^2 gibt den Flächeninhalt eines Quadrats mit der Seite a an, daher liest man a^2 auch „ a Quadrat“. Die dritten Potenzen werden **Kubikzahlen** genannt; a^3 gibt den Rauminhalt eines Würfels mit der Kante a an.

Eine besondere Bedeutung haben die **Zehnerpotenzen**. Man benutzt sie bei Überschlagsrechnungen, um sich einen Überblick über die Größenordnung zu verschaffen, sowie um sehr große und sehr kleine Zahlen abgekürzt und übersichtlich zu schreiben:

$$1\,000 = 10 \cdot 10 \cdot 10 = 10^3 \quad \text{oder} \quad 1\,000\,000 = 10^6$$

$$(10^3 - 10^2) \text{ ist nicht } 10^1 \text{ sondern } 900 = 0,9 \cdot 10^3.$$

Einige Vorsilben zu Zehnerpotenzen

10^{-k}	k	10^{+k}
Milli	3	Kilo
Mikro	6	Mega
Nano	9	Giga

Zunächst einige **Potenzgesetze** mit Beispielen (m und n seien **natürliche Zahlen**):

$$a^m \cdot a^n = a^{m+n} \qquad \rightarrow 2^4 \cdot 2^3 = 2^{4+3} = 2^7 = 128$$

$$a^m : a^n = a^{m-n} \qquad \rightarrow 2^4 : 2^3 = 2^{4-3} = 2^1 = 2$$

$$a^n \cdot b^n = (ab)^n \qquad \rightarrow 6^2 \cdot 3^2 = 6 \cdot 6 \cdot 3 \cdot 3 = (6 \cdot 3)^2 = 18^2 = 324$$

$$a^m : b^m = \left(\frac{a}{b}\right)^m \qquad \rightarrow 7^4 : 5^4 = \left(\frac{7}{5}\right)^4 = 1,4^4 = 3,8416$$

$$(a^m)^n = a^{m \cdot n} = (a^n)^m \rightarrow (5^2)^3 = 5^2 \cdot 5^2 \cdot 5^2 = 5^{2 \cdot 3} = 5^6 = 15625$$

$$a^{-n} = \frac{1}{a^n} \qquad \rightarrow 10^{-3} = \frac{1}{10^3} = \frac{1}{1000} = 0,001; 2^{-6} = \frac{1}{2^6} = \frac{1}{64} = 15,625 \cdot 10^{-3}$$

$$a^0 = 1 \quad \text{für} \quad a \neq 0 \quad \rightarrow \frac{a^5}{a^5} = a^{5-5} = a^0 = 1 \quad (\text{vgl. auch: } 0^a = 0 \text{ für } a > 0)$$

Diese Gesetzmäßigkeiten gelten auch, wenn m und n keine ganzen Zahlen sind; das heißt, wenn $a \neq 0$, gelten die angegebenen Potenzgesetze auch für gebrochene Exponenten ($m = p/q, n = r/s$).

Wurzelrechnung (Radizieren): Statt $a^{(1/n)}$ schreibt man auch $\sqrt[n]{a^1} = \sqrt[n]{a}$ und liest n -te Wurzel aus a . Für $n = 2$ (Quadratwurzel) schreibt man kurz $\sqrt{a}$. $\sqrt[n]{a}$ ist die Zahl, die, in die n -te Potenz erhoben, den Radikanden a ergibt: $[\sqrt[n]{a}]^n = a$. Folgende Bezeichnung ist üblich:

$$\text{Wurzelexponent} \sqrt{\text{Radikand}} = \text{Wurzelwert} \qquad \sqrt[2]{25} = \sqrt{25} = 5, \quad \text{denn} \quad 5^2 = 25$$

Einige Formeln und Beispiele für das Rechnen mit Wurzeln:

$\sqrt[n]{a} \cdot \sqrt[n]{b} = \sqrt[n]{ab}$	$\frac{\sqrt[n]{a}}{\sqrt[n]{b}} = \sqrt[n]{\frac{a}{b}}$	$a^{m/n} = \sqrt[n]{a^m}$	$[\sqrt[n]{a}]^m = \sqrt[n]{a^m}$	$\sqrt[n]{\sqrt[n]{a}} = \sqrt[n \cdot n]{a}$

$$\sqrt{50} = \sqrt{25 \cdot 2} = 5\sqrt{2}, \quad \sqrt{50}/\sqrt{2} = \sqrt{50/2} = \sqrt{25} = 5, \quad \sqrt[4]{3^{12}} = 3^{12/4} = 3^3 = 27$$

$$(\sqrt[3]{8})^2 = 4 = \sqrt[3]{8^2}; \quad \sqrt[3]{\sqrt{64}} = 2 = \sqrt[3 \cdot 2]{64}; \quad \sqrt[3]{35^5} = 35^{5/3} = 374,5.$$

```

> sqrt(3)                                     # Radizieren in R
[1] 1.732051
> sqrt(234)
[1] 15.29706
> 35^(5/3)
[1] 374.4956
```

2.3.3 Logarithmen

Das Rechnen mit Logarithmen (Logarithmieren): Logarithmen sind Exponenten. Wenn a eine positive Zahl ist, und y eine beliebige Zahl (> 0), dann gibt es eine eindeutig bestimmte Zahl x , so dass $a^x = y$ ist. Diese Zahl x heißt **Logarithmus** von y zur Basis a , geschrieben:

$$x = {}^a \log y \quad \text{oder} \quad x = \log_a y \quad \text{Mit } a^0 = 1 \text{ gilt } \log_a 1 = 0.$$

Die Zahl y heißt **Numerus** des Logarithmus zur Basis a . Meist werden Logarithmen zur Basis 10 verwendet, geschrieben ${}^{10} \log x$, $\log_{10} x$ oder einfach $\lg x$. Andere Logarithmensysteme werden weiter unten erwähnt. Nehmen wir $a = 10$ und $y = 3$, dann ergibt sich mit den Logarithmen zur Basis 10 (**Briggssche**, dekadische oder Zehnerlogarithmen) $x = 0,4771$ und $10^{0,4771} = 3$. Weitere Beispiele mit vierstelligen Logarithmen:

$5 = 10^{0,6990}$	oder	$\lg 5 = 0,6990$
$1 = 10^0$	oder	$\lg 1 = 0$
$10 = 10^1$	oder	$\lg 10 = 1$
$1000 = 10^3$	oder	$\lg 1000 = 3$
$0,01 = 10^{-2}$	oder	$\lg 0,01 = -2$

Da Logarithmen Exponenten sind, gelten also die Potenzgesetze, z. B.:

$$2 \cdot 4 = 10^{0,3010} \cdot 10^{0,6021} = 10^{0,3010+0,6021} = 10^{0,9031} = 8.$$

Die Multiplikation von Zahlen wird zurückgeführt auf die Addition der Logarithmen der Zahlen. Entsprechend gilt: Division wird zu Subtraktion, Potenzieren wird zu Multiplikation, Radizieren wird zu Division – allgemein:

$\left. \begin{array}{l} 1. \lg(ab) = \lg a + \lg b \\ 2. \lg \frac{a}{b} = \lg a - \lg b \end{array} \right\} (a > 0, b > 0)$	$\left. \begin{array}{l} 3. \lg a^n = n \lg a \\ 4. \lg \sqrt[n]{a} = \lg a^{1/n} = \frac{1}{n} \lg a \end{array} \right\} \begin{array}{l} (a > 0, \\ n = \text{Dezi-} \\ \text{malzahl}) \end{array}$

$$\lg \left[\frac{1}{c} \right] = \lg 1 - \lg c = 0 - \lg c = -\lg c$$

$$= \lg c^{-1} = (-1) \lg c = -\lg c$$

[vgl. $\lg 630 = \lg(10 \cdot 63) = \lg 10 + \lg 63 = 1 + 1,7993 = 2,7993$;
 $\lg 6,30 = \lg(63/10) = \lg 63 - \lg 10 = 1,7993 - 1 = 0,7993$]

(1) **Effekt:** multiplikativ $\longrightarrow$ additiv

$$y = ab \quad \lg y = \lg a + \lg b$$

(2) **Symmetrie:** nein $\longrightarrow$ ja

$$a/b \neq b/a \quad \lg(a/b) = -\lg(b/a)$$

Die sogenannten **natürlichen Logarithmen (ln)** haben als Basis die Konstante

$$e \approx 2,718281828459 \dots$$

Die Umrechnungsformeln lauten mit gerundeten Werten (vgl. auch Tabelle 2.3):

$\ln x = \ln 10 \cdot \lg x \simeq 2,302585 \cdot \lg x$ $\lg x = \lg e \cdot \ln x \simeq 0,4342945 \ln x$	(vgl. $\ln 1 = 0, \ln e = 1$, $\ln 10^k \simeq k \cdot 2,302585$)

Anstelle „ $\ln x$ “ findet man auch „ $^e \log x$ “ und „ $\log_e x$ “ [vgl. auch $\ln e^x = x, e^{\ln x} = x$ und insbesondere $a^x = e^{x \cdot \ln a} (a > 0)$].

Den **Logarithmus zur Basis 2**, Logarithmus dualis, ld (bzw. mit lb [binär, aus zwei Einheiten bestehend] bezeichnet), erhält man nach:

$\text{ld } x = \frac{\lg x}{\lg 2} \simeq 3,321928 \lg x$ $\text{ld } x = \frac{\ln x}{\ln 2} \simeq 1,442695 \cdot \ln x$	(z.B. $\text{ld } 5 = 2,322$ $= 3,322 \cdot 0,699$ $= 1,1443 \cdot 1,609$)	Sonderfälle: $\log_a a = 1$ $\log_a 1 = 0$ $\log_a(a^x) = x$

Tabelle 2.3. Wichtige Konstanten

Wichtige Konstanten	
π	3,141592654
$(2\pi)^{-1/2}$	0,398942280
Eulersche Zahl e	2,718281828
$\lg e$	0,434294482
$\ln 10$	2,302585093
Eulersche Konstante	0,577215665

```

> pi                      # Die Zahl pi = 3.141593
[1] 3.141593
> exp(1)                  # Die Zahl e = 2.718282
[1] 2.718282
> log(12, base = exp(1))  # Logarithmus zur Basis e
[1] 2.484907
> log10(16)               # Logarithmus zur Basis 20
[1] 1.20412
> log2(20)                # Logarithmus zur Basis 2
[1] 4.321928

```

2.3.4 Rundungen

Sollen die Werte 14,6, 13,8, 19,3, 83,5 und 14,5 auf die jeweils nächste ganze Zahl gerundet werden, so bereitet dies bei den ersten drei Werten keine Schwierigkeit; sie werden zu 15, 14 und 19. Bei den folgenden Werten kämen die Zahlen 83 und 84 bzw. 14 und 15 in Betracht. Es hat sich als zweckmäßig erwiesen, jeweils zu der nächsten geraden Zahl auf- oder abzurunden, so dass 83,5 in 84 und 14,5 in 14 übergeht. Die Null wird hierbei als gerade Zahl gewertet. Je mehr Werte auf diese Weise gerundet und zur Summe zusammengefasst werden, um so schneller gleichen sich die Rundungsfehler aus. Man runde: (1) ab bei Endziffern bis 4, (2) auf bei Endziffern über 5, (3) zur nächsten geraden Zahl bei der Endziffer = 5: $2,45 \rightarrow 2,4$ oder, falls nötig $2,4^+$ bzw. $3,55 \rightarrow 3,6$ oder $3,6^-$.

Übersicht 5. „Altväterliches“ Rezept für jede **Rechnung von Hand**

1. **Rechenschema:** Aufeinanderfolgende Rechenschritte in allen Einzelheiten festlegen. Eine umfangreiche Berechnung sollte so gut durchdacht und vorbereitet sein, dass ihre Durchführung angelernten Hilfskräften überlassen werden kann. Übersichtliche Rechenschemata, die die gesamte Zahlenrechnung enthalten und nach denen die Rechnung planmäßig-schematisch abläuft, helfen auch Fehler zu vermeiden.
2. Bogen einseitig beschreiben; Ziffern deutlich schreiben; breite Randspalte für Nebenrechnungen frei lassen; Übertragungen vermeiden; falsche Zahlen durchstreichen, die richtigen darüberschreiben.
3. **Überschlagsrechnungen** zur Vermeidung von Kommafehlern einschalten; Kontrolle der Rechnung!

Jeder Rechenoperation hat eine Überschlagsrechnung voranzugehen oder zu folgen, wobei zumindest die Kommastellung im Ergebnis sicher entschieden wird. Hierbei ist die Schreibweise mit Zehnerpotenzen zu empfehlen:

$$\frac{0,00904}{0,167} = \frac{9,04 \cdot 10^{-3}}{1,67 \cdot 10^{-1}} \simeq 5 \cdot 10^{-2}, \text{ auf 3 Stellen genau: } 5,413 \cdot 10^{-2}.$$

4. Wenn möglich, sollte die Aufgabe **zur besseren Kontrolle noch nach einer anderen Methode** gelöst werden. Mitunter ist es besser, wenn 2 Mitarbeiter die Berechnungen unabhängig voneinander ausführen und ihre Resultate vergleichen.
5. Je nach den zur Verfügung stehenden **Rechenhilfsmitteln** sind diese Empfehlungen und die im Buch angeführten Rechenkontrollen zu modifizieren und durch optimalere zu ersetzen.

Rundungsfehler bzw. sichere Ergebnisse auf Rechenanlagen behandeln ausführlich z.B. Meis [Mei87] und Rump [Rum86].

Wichtig ist auch der Begriff der **signifikanten Ziffern**. Unter den signifikanten Ziffern einer Zahl versteht man die Ziffernfolge der Zahl ohne Berücksichtigung des evtl. vorhandenen Kommas und bei Zahlen kleiner als 1 ohne die Null vor dem Komma und ohne die dann noch folgenden Nullen. Tabelle 2.4 vergleicht drei gerundete Resultate, die Anzahl der signifikanten Ziffern und die hiermit zum Ausdruck gebrachte Genauigkeit: die im Ergebnis mit einbegriffenen Genauigkeitsgrenzen sowie ihren maximalen Rundungsfehler.

Werden zwei Zahlen, jede mit x genauen oder signifikanten Ziffern multipliziert, dann sind höchstens $(x - 1)$ Ziffern des Produktes als verlässlich anzusehen. Für die Division gilt Entsprechendes.

Tabelle 2.4. Beispiele zu Rundungen und dem daraus resultierenden Fehler

Resultat (R)	Anzahl signifikanter Ziffern	Grenzwerte des Fehlerbereiches	Differenz (D)	Größter Fehler (%) $= \frac{0,5 \cdot D }{R} \cdot 100$
4	1	3,5 - 4,5	1	12,5
4,4	2	4,35 - 4,45	0,1	1,14
4,44	3	4,435 - 4,445	0,01	0,113

Beispiel: Berechne die Fläche eines Rechtecks aus den gemessenen Seitenlängen 38,22 cm und 16,49 cm. Die Antwort als $38,22 \cdot 16,49 = 630,2478 \text{ cm}^2$ zu formulieren wäre falsch, da die Fläche jeden Wert zwischen $38,216 \cdot 16,486 = 630,02898$ und $38,224 \cdot 16,494 = 630,46666$ annehmen kann. Dieses Gebiet wird charakterisiert durch $630,2 \text{ cm}^2 \pm 0,3 \text{ cm}^2$. Der Wert kann nur durch drei signifikante Ziffern dargestellt werden (630 cm^2).

Resultate sind auf eine sinnvolle Zahl signifikanter Ziffern zu runden, die selten die der Originaldaten (die am wenigsten genaue oder kleinste Zahl) übersteigt, was dann methodisch bedingt bzw. besonders zu begründen ist.

Rundungen können in **R** einfach mit den Funktionen **ceiling()**, **floor()**, **trunc()** und **round()** durchgeführt werden.

```
> ceiling(2.34)           # kleinste ganze Zahl nicht < x
[1] 3
> floor(1.56)             # groesste ganze Zahl nicht > x
[1] 1
> trunc(-2.765)           # ganzzahliger Anteil
[1] -2
> round(1.2654, digits = 2) # Rundung
[1] 1.27
```

2.3.5 Rechnen mit fehlerbehafteten Zahlen

Werden fehlerbehaftete Zahlen durch Rechenoperationen verbunden, dann lässt sich die so genannte **Fehlerfortpflanzung** abschätzen. Hierzu können zwei parallele Rechnungen durchgeführt werden, einmal mit den Fehlerschranken, die im Endergebnis zu einem Minimum führen, und ein zweites Mal mit den Fehlerschranken, die im Ergebnis zu einem Maximum führen.

Beispiel:

30 ± 3
Bereich: von 27 bis 33

20 ± 1
Bereich: von 19 bis 21

- Addition:** Die wahre Summe beider Zahlen liegt zwischen $27 + 19 = 46$ und $33 + 21 = 54$. Der relative Fehler der Summe beträgt $(54 - 46)/(54 + 46) = 8/100 = 0,08$; er liegt in den Grenzen von $\pm 8\%$.
- Subtraktion:** Die wahre Differenz liegt zwischen $27 - 21 = 6$ und $33 - 19 = 14$ (Subtraktion „überkreuz“, d. h. der obere Grenzwert einer Zahl wird von dem unteren Grenzwert der anderen Zahl abgezogen, der untere Grenzwert einer Zahl wird von dem oberen der anderen Zahl abgezogen). Der relative Fehler der Differenz beträgt: $(14 - 6)/(14 + 6) = 8/20 = 0,40, \pm 40\%$!
- Multiplikation:** Das wahre Produkt liegt in den Grenzen von $27 \cdot 19 = 513$ bis $33 \cdot 21 = 693$. Der relative Fehler des Produktes beträgt

$$\frac{513 - 30 \cdot 20}{30 \cdot 20} = \frac{513 - 600}{600} = \frac{-87}{600} = -0,145 \quad -14,5\% \text{ bzw.}$$

$$\frac{693 - 30 \cdot 20}{30 \cdot 20} = \frac{693 - 600}{600} = \frac{93}{600} = 0,155 \quad +15,5\%$$

4. **Division:** Der wahre Quotient liegt zwischen $27/21 = 1,286$ und $33/19 = 1,737$ (Division „überkreuz“). Den relativen Fehler des Quotienten erhält man zu

$$\frac{1,286 - 30/20}{30/20} = \frac{0,214}{1,500} = -0,143 \quad -14,3\% \text{ bzw.}$$

$$\frac{1,737 - 30/20}{30/20} = \frac{0,237}{1,500} = 0,158 \quad +15,8\%$$

Von allen vier Grundrechenoperationen ist die Subtraktion bei fehlerbehafteten Zahlen besonders gefährlich, der Endfehler liegt mitunter wesentlich höher als bei den anderen Rechenoperationen.



Ein grundsätzlich vermeidbarer Fehler, der ein Resultat bedingt, das vom wahren Befund oder Wert systematisch [und nicht-zufällig] abweicht, wird **systematischer Fehler** oder **Bias** genannt, z.B durch ein falsch eingestelltes Messinstrument. Er bedroht alle Stufen einer wissenschaftlichen Untersuchung, etwa die Auswahl der Untersuchungseinheiten, die Durchführung der Experimente oder Beobachtungen, die Datenanalyse bis zur Interpretation der Resultate.

2.4 Einführung in die Matrixalgebra

- Definition und Schreibweise
- Matrixoperationen
- Determinanten
- Die Inverse Matrix
- Lineare Abhängigkeit, Rang einer Matrix
- Lineare Gleichungssysteme
- Eigenwerte und Eigenvektoren

Eine **Matrix** ist eine feste Struktur von Zahlen, die in Zeilen und Spalten angeordnet sind. Die große Bedeutung des Begriffes in der Statistik liegt darin, dass Beobachtungen im Rahmen von Experimenten und Erhebungen „natürlich“ in dieser Form vorliegen (vgl. auch Abbildung 1.4 im 1. Kapitel). Die Zeilen sind die Fälle (Merkmalsträger, Beobachtungseinheiten) und die Spalten sind die Beobachtungen (Merkmale, Variablen). Die Einführung in die Matrixalgebra geht über das Niveau der vorangehenden Abschnitte hinaus und ist für dieses Buch ausschließlich zum Verständnis des Kapitels zur Modellbildung (Beschreibung und Bewertung **multipler Einflussgrößen**) notwendig. In den anderen Kapiteln wird traditionell die algebraische Darstellung verwendet. Für den Einstieg kann dieses Kapitel zunächst übersprungen und bei Bedarf zu einem späteren Zeitpunkt nachgelesen werden.

2.4.1 Definition und Schreibweise

Eine **Matrix A** ist ein rechteckiges Schema von (reellen) Zahlen a_{ij} ; die Zahlen in diesem Schema heißen Elemente von **A**. Matrizen werden in der Regel mit großen lateinischen Buchstaben bezeichnet.

$$\mathbf{A}_{(n \times m)} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} = (a_{ij}) \quad (2.17)$$

Der erste Index (i) heißt **Zeilenindex**, der zweite Index (j) heißt **Spaltenindex**. Die **Dimension** oder auch Ordnung einer Matrix wird durch die Anzahl der Zeilen und Spalten ($n \times m$) bestimmt. Eine Matrix ($\mathbf{A}$) wird transponiert, indem Zeilen und Spalten vertauscht werden. Dieses entspricht einer Spiegelung der Matrixelemente an der **Hauptdiagonalen** (a_{ii}) der Matrix. Für die **Transponierte** einer Matrix $\mathbf{A}$ schreibt man $\mathbf{A}^T$ oder auch $\mathbf{A}'$.

$$\mathbf{A}'_{(m \times n)} = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & & \vdots \\ a_{1m} & a_{2m} & \cdots & a_{nm} \end{pmatrix} = (a_{ji}) \quad (2.18)$$

Für die Definition einer Matrix und für das Transponieren stehen in dem Programm **R** spezielle Funktionen zur Verfügung:

```
> A <- matrix( c(1, 2, 3, 6, 5, 4), nrow=2, ncol=3, byrow=TRUE)
> A.trans <- t(A); A; A.trans      # Transponieren einer Matrix
      [,1] [,2] [,3]
[1,]    1    2    3
[2,]    6    5    4
      [,1] [,2]
[1,]    1    6
[2,]    2    5
[3,]    3    4
```

Insbesondere gilt $(\mathbf{A}')' = \mathbf{A}$.

Eine Matrix heißt **symmetrisch**, wenn: $\mathbf{A}' = \mathbf{A}$ (2.19)

Eine Matrix mit n Zeilen und nur einer Spalte heißt auch **Spaltenvektor**. Entsprechend wird eine Matrix mit einer Zeile und m Spalten auch **Zeilenvektor** genannt.

$$\text{Spaltenvektor: } \mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \quad \text{Zeilenvektor: } \mathbf{a}' = (a_1, a_2, \dots, a_m) \quad (2.20)$$

Zwei Matrizen $\mathbf{A}$ und $\mathbf{B}$ sind gleich, wenn sie in allen Elementen übereinstimmen:

$$a_{i,j} = b_{i,j} \Leftrightarrow \mathbf{A}_{(n \times m)} = \mathbf{B}_{(n \times m)}$$

Einige wichtige Sonderfälle und die dazugehörige Notation sind in Tabelle 2.5 zusammengestellt.

2.4.2 Matrixoperationen

2.4.2.1 Matrixaddition und -subtraktion

Zwei Matrizen $\mathbf{A}$ und $\mathbf{B}$ werden **elementweise addiert (subtrahiert)**. Dazu müssen sie in der Anzahl der Zeilen und Spalten übereinstimmen.

Tabelle 2.5. Einige Definitionen und Notationen zur Matrixalgebra

Bezeichnung	Definition	Notation
Nullvektor	$(0, 0, \dots, 0)'$	$\mathbf{o}'$
Einsvektor	$(1, 1, \dots, 1)'$	$\mathbf{1}'$
Einheitsvektor	$(0, \dots, 0, 1, 0, \dots, 0)'$	$\mathbf{i}'_i$
Nullmatrix	$a_{ij} = 0 \quad \forall i, j$	$\mathbf{0}$
Einsmatrix	$a_{ij} = 1 \quad \forall i, j$	$\mathbf{J}$
Einheitsmatrix	$a_{ii} = 1; \quad a_{ij} = 0 \quad \forall i \neq j$	$\mathbf{I}$
quadratische Matrix	$n = m$	$\mathbf{A}_{(n \times n)}$
symmetrische Matrix	$a_{ij} = a_{ji}$	

$$c_{ij} = a_{ij} \pm b_{ij} \quad \forall i, j \quad \Leftrightarrow \quad C_{(n \times m)} = \mathbf{A}_{(n \times m)} \pm \mathbf{B}_{(n \times m)} \quad (2.21)$$

Im folgenden Beispiel werden in **R** zwei einfache Matrizen **A** und **B** definiert und anschließend wird die Summe **A + B = C** gebildet:

```
> A <- matrix( c(1, 2, 3, 6, 5, 4), nrow=2, ncol=3, byrow=TRUE)
> B <- matrix(c(4, 5, 6, 9, 8, 7), nrow=2, ncol=3, byrow=TRUE)
> C <- A + B; A; B; C           # Addition zweier Matrizen
      [,1] [,2] [,3]
[1,]    1    2    3
[2,]    6    5    4
      [,1] [,2] [,3]
[1,]    4    5    6
[2,]    9    8    7
      [,1] [,2] [,3]
[1,]    5    7    9
[2,]   15   13   11
```

Für die Matrixaddition und -subtraktion sind folgende Regeln zu beachten. Das neutrale Element der Matrixaddition ist die **Nullmatrix**.

$$\begin{aligned} \mathbf{A} + \mathbf{B} &= \mathbf{B} + \mathbf{A} \\ (\mathbf{A} + \mathbf{B}) + \mathbf{C} &= \mathbf{A} + (\mathbf{B} + \mathbf{C}) \\ (\mathbf{A} \pm \mathbf{B})' &= \mathbf{A}' \pm \mathbf{B}' \\ \mathbf{A} \pm \mathbf{0} &= \mathbf{A} \end{aligned} \quad (2.22)$$

2.4.2.2 Matrixmultiplikation

Einzelne reelle Zahlen, die als Faktoren in der Matrixmultiplikation auftreten, werden **Skalare** genannt. Eine Matrix **A** wird mit einem Skalar *c* multipliziert, indem jedes Element der Matrix mit dem Skalar multipliziert wird.

$$c\mathbf{A} = (ca_{ij}) \quad \forall i, j \quad (2.23)$$

```
> A <- matrix( c(1, 2, 3, 6, 5, 4), nrow=2, ncol=3, byrow=TRUE);
> A; 2 * A           # Multiplikation mit einem Skalar
      [,1] [,2] [,3]
[1,]    1    2    3
[2,]    6    5    4
      [,1] [,2] [,3]
```

[1,]	2	4	6
[2,]	12	10	8

Dabei gelten die folgenden Regeln (c und d sind Skalare):

$$\begin{aligned}
 c\mathbf{A} &= \mathbf{A}c \\
 (cd)\mathbf{A} &= c(d\mathbf{A}) \\
 (c \pm d)\mathbf{A} &= c\mathbf{A} \pm d\mathbf{A} \\
 c(\mathbf{A} \pm \mathbf{B}) &= c\mathbf{A} \pm c\mathbf{B}
 \end{aligned}
 \tag{2.24}$$

Bei der **Multiplikation zweier Matrizen** $\mathbf{A}$ und $\mathbf{B}$ werden die Elemente über die Zeilen von $\mathbf{A}$ und den Spalten von $\mathbf{B}$ miteinander multipliziert und addiert. Somit muss die Zahl der Spalten in $\mathbf{A}$ mit der Zahl der Zeilen in $\mathbf{B}$ übereinstimmen:

$$\mathbf{C}_{(n \times m)} = \mathbf{A}_{(n \times k)} \mathbf{B}_{(k \times m)} \tag{2.25}$$

$$c_{ij} = \sum_{l=1}^k a_{il} b_{lj} \quad (i = 1, \dots, n \quad \text{und} \quad j = 1, \dots, m)$$

```

> A <- matrix( c(1, 2, 3, 6, 5, 4), nrow=2, ncol=3, byrow=TRUE);
> B <- matrix(c(4, 5, 6, 9, 8, 7), nrow=3, ncol=2, byrow=TRUE);
> C <- A %*% B; A; B; C           # Multiplikation (%*%) zweier Matrizen
      [,1] [,2] [,3]
[1,]    1    2    3
[2,]    6    5    4
      [,1] [,2]
[1,]    4    5
[2,]    6    9
[3,]    8    7
      [,1] [,2]
[1,]   40   44
[2,]   86  103

```

Die Matrixmultiplikation ist generell nicht kommutativ ($\mathbf{AB} \neq \mathbf{BA}$). Für Produkte von Matrizen gelten unter Beachtung der entsprechenden Dimensionen folgende Regeln:

$$\begin{aligned}
 (\mathbf{AB})\mathbf{C} &= \mathbf{A}(\mathbf{BC}) \\
 \mathbf{A}(\mathbf{B} + \mathbf{C}) &= \mathbf{AB} + \mathbf{AC} \\
 (\mathbf{B} + \mathbf{C})\mathbf{A} &= \mathbf{BA} + \mathbf{CA}
 \end{aligned}
 \tag{2.26}$$

Das **neutrale Element** der Matrixmultiplikation ist die Einheitsmatrix $\mathbf{I}$:

$$\mathbf{IA} = \mathbf{AI} = \mathbf{A} \tag{2.27}$$

Die **Multiplikation mit der Nullmatrix** $\mathbf{0}$ liefert stets die Nullmatrix:

$$\mathbf{0A} = \mathbf{A0} = \mathbf{0} \tag{2.28}$$

Für das **Transponieren eines Matrixproduktes** gilt:

$$\begin{aligned}
 (\mathbf{AB})' &= \mathbf{B}'\mathbf{A}' \\
 (\mathbf{ABC})' &= \mathbf{C}'\mathbf{B}'\mathbf{A}'
 \end{aligned}
 \tag{2.29}$$

Ein Sonderfall bei der Matrixmultiplikation ist das Produkt von Vektoren mit gleicher Anzahl von Elementen (**Skalarprodukt**). Dabei wird entsprechend elementweise multipliziert und die Summe über die Produkte gebildet:

$$\mathbf{a}'\mathbf{b} = (a_1, a_2, \dots, a_n) \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = \sum_{i=1}^n a_i b_i \quad (2.30)$$

```
> a <- c(1, 2, 3)
> b <- c(4, 5, 6)
> c <- t(a) %*% b; a; b; c           # Skalarprodukt zweier Vektoren
[1] 1 2 3
[1] 4 5 6
     [,1]
[1,] 32
```

Das Skalarprodukt hat die folgenden Eigenschaften:

$$\begin{aligned} \mathbf{a}'\mathbf{b} &= \mathbf{b}'\mathbf{a} \\ \mathbf{a}'\mathbf{1} &= \sum_{i=1}^n a_i \\ \mathbf{a}'\mathbf{a} &= \sum_{i=1}^n a_i^2 \end{aligned} \quad (2.31)$$

Die **Länge** oder **Norm eines Vektors** wird wie folgt definiert:

$$\|\mathbf{a}\| := \sqrt{\mathbf{a}'\mathbf{a}} = \sqrt{\sum a_i^2} \quad (2.32)$$

```
> a <- c(1, 2, 3, 4, 5, 6)
> a.trans <- t(a)
> a.norm <- sqrt(a.trans %*% a)      # Norm eines Vektors
> a; a.norm
[1] 1 2 3 4 5 6
     [,1]
[1,] 9.539392
```

Ein Vektor **a** heißt **normiert**, wenn er die Länge 1 hat, d.h. $\|\mathbf{a}\| = 1$ gilt. Ein Vektor kann somit durch die Multiplikation mit dem Kehrwert seiner Norm normiert werden.

$$\mathbf{a}_n = \frac{1}{\|\mathbf{a}\|} \mathbf{a} \quad (2.33)$$

Zwei Vektoren, die vom Nullvektor verschieden sind, heißen **orthogonal**, wenn ihr Skalarprodukt gleich dem Nullvektor ist:

$$\mathbf{a}'\mathbf{b} = \sum a_i b_i = 0 \quad (2.34)$$

Zwei normierte Vektoren, die orthogonal zueinander sind, heißen **orthonormal**.

2.4.3 Determinanten

Für das Rechnen mit Matrizen, insbesondere die Bestimmung der Inversen (vgl. nächsten Abschnitt) und des Ranges ist die Berechnung von Determinanten erforderlich. Die **Determinante** einer quadratischen Matrix $\mathbf{A}$ ist ein Skalar, welcher wie folgt (hier für die Entwicklung nach der i -ten Spalte für beliebiges i) rekursiv definiert wird:

$$|\mathbf{A}| = a \quad \text{wenn} \quad \mathbf{A}_{(1 \times 1)} = a \quad (2.35)$$

$$|\mathbf{A}| = \sum_{j=1}^n (-1)^{i+j} a_{ij} |\mathbf{A}_{ij}|$$

Dabei ist $|\mathbf{A}_{ij}|$ die Determinante jener $(n-1) \times (n-1)$ Matrix $\mathbf{A}_{ij}$, die man aus $\mathbf{A}$ durch Streichen der i -ten Zeile und j -Spalte erhält.

```
> A <- matrix(c(3, 1, 2, 4, 5, 6, 9, 7, 8), nrow=3, ncol=3, byrow=TRUE)
> A.det <- det(A); A; A.det      # Bestimmung der Determinante
      [,1] [,2] [,3]
[1,]    3    1    2
[2,]    4    5    6
[3,]    9    7    8
[1] -18
```

Die Bestimmung von Determinanten ist für die Berechnungen in den folgenden Abschnitten notwendig.

2.4.4 Die Inverse Matrix

Eine quadratische Matrix A heißt **invertierbar**, wenn es eine Matrix A^{-1} (die **Inverse** oder **Kehrmatrix**) gibt, für die gilt:

$$\mathbf{A} \mathbf{A}^{-1} = \mathbf{A}^{-1} \mathbf{A} = \mathbf{I} \quad (2.36)$$

Eine quadratische Matrix ist genau dann invertierbar, wenn ihre Determinante von Null verschieden ist ($|\mathbf{A}| \neq 0$). In diesem Fall ist die Matrix numerisch nicht singular (lineare Unabhängigkeit der Vektoren in $\mathbf{A}$). Eine Möglichkeit der Berechnung erfolgt über die Determinante nach:

$$\mathbf{A}^{-1} = (a_{ij}^{(-1)}) \quad \text{mit} \quad a_{ij}^{(-1)} = \frac{(-1)^{j+i} |\mathbf{A}_{ij}|}{|\mathbf{A}|}$$

In **R** wird hierzu in dem Paket **library(MASS)** (Venables [VB02]) die so genannte verallgemeinerte „Moore-Penrose“-Inverse mit der Funktion **ginv()** eingeführt:

```
> A <- matrix(c(3, 1, 2, 4, 5, 6, 9, 7, 8), nrow=3, ncol=3, byrow=TRUE)
> A.inv <- ginv(A)      # Berechnung der inversen Matrix
> A; round(A.inv, 2); round(A %*% A.inv, 2)
      [,1] [,2] [,3]
[1,]    3    1    2
[2,]    4    5    6
[3,]    9    7    8
      [,1] [,2] [,3]
[1,]  0.11 -0.33  0.22
[2,] -1.22 -0.33  0.56
[3,]  0.94  0.67 -0.61
      [,1] [,2] [,3]
[1,]    1    0    0
[2,]    0    1    0
[3,]    0    0    1
```

Für das Rechnen mit der Inversen einer Matrix gelten folgende Regeln:

$$\begin{aligned}
 (\mathbf{A}^{-1})^{-1} &= \mathbf{A} \\
 (\mathbf{AB})^{-1} &= \mathbf{B}^{-1} \mathbf{A}^{-1} \\
 (\mathbf{ABC})^{-1} &= \mathbf{C}^{-1} \mathbf{B}^{-1} \mathbf{A}^{-1} \\
 (\mathbf{A}')^{-1} &= (\mathbf{A}^{-1})' \\
 (c\mathbf{A})^{-1} &= \frac{1}{c} \mathbf{A}^{-1} \quad \forall c \neq 0 \\
 \mathbf{I}^{-1} &= \mathbf{I}
 \end{aligned} \tag{2.37}$$

2.4.5 Lineare Abhängigkeit, Rang einer Matrix

Eine **Linearkombination** von Vektoren besteht aus einer Summe von Vektoren $\mathbf{a}_j$, die mit Skalaren x_j multipliziert werden:

$$\mathbf{b} := x_1 \mathbf{a}_1 + x_2 \mathbf{a}_2 + \dots + x_m \mathbf{a}_m = \sum_{j=1}^m x_j \mathbf{a}_j \tag{2.38}$$

Die Vektoren in einer Linearkombination $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_m$ (keine Nullvektoren) heißen **linear unabhängig**, wenn es reelle Zahlen x_j gibt, die nicht alle Null sind, so dass gilt:

$$\sum_{j=1}^m x_j \mathbf{a}_j = \mathbf{0} \tag{2.39}$$

Anderenfalls heißen die Vektoren linear abhängig.

Die Anzahl linear unabhängiger Spaltenvektoren (Zeilenvektoren) in einer Matrix $\mathbf{A}$ bestimmt den **Rang** (rg) der Matrix.

$$\text{rg}(\mathbf{A}_{(n \times m)}) \leq \min(n, m) \tag{2.40}$$

$$\text{rg}(\mathbf{A}_{(n \times m)}) = \min(n, m) \Rightarrow \mathbf{A} \text{ hat vollen Rang}$$

Eine **quadratische Matrix** mit **vollem Rang** heißt **regulär**, anderenfalls **singulär**:

$$\text{rg}(A_{n \times n}) = n \Leftrightarrow |A| \neq 0 \Leftrightarrow A^{-1} \text{ existiert!} \tag{2.41}$$

Für das Rechnen mit Rängen gelten die folgenden Regeln:

$$\begin{aligned}
 \text{rg}(\mathbf{A}) &= \text{rg}(\mathbf{A}') \\
 \text{rg}(\mathbf{AB}) &\leq \min\{\text{rg}(\mathbf{A}), \text{rg}(\mathbf{B})\} \\
 \text{rg}(\mathbf{A}'\mathbf{A}) &= \text{rg}(\mathbf{AA}') = \text{rg}(\mathbf{A})
 \end{aligned} \tag{2.42}$$

2.4.6 Lineare Gleichungssysteme

Lineare Gleichungssysteme bestehen aus einer Anzahl von **Linearkombinationen** der Form:

$$\begin{array}{ccccccc} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m & = & b_1 \\ \vdots & & \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nm}x_m & = & b_n \end{array} \quad (2.43)$$

Ein Vektor $\mathbf{x}$, der diese Bedingungen erfüllt, heißt **Lösung des Gleichungssystems** $\mathbf{Ax} = \mathbf{b}$; dabei bezeichnet $\mathbf{A}$ die **Koeffizientenmatrix**. Das Gleichungssystem ist **homogen**, wenn $\mathbf{b} = \mathbf{0}$, anderenfalls nennt man es inhomogen.

Ein lineares Gleichungssystem besitzt genau dann eine Lösung, wenn

$$rg(\mathbf{A}, \mathbf{b}) = rg(\mathbf{A}) \quad (2.44)$$

Hinsichtlich der Lösung eines linearen Gleichungssystems unterscheidet man insbesondere die folgenden Sonderfälle:

1. Die Koeffizientenmatrix $\mathbf{A}$ ist quadratisch und hat vollen Rang $rg(\mathbf{A}_{m,m}) = m$:

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b} \quad \text{ist eine eindeutige Lösung} \quad (2.45)$$

2. Die Koeffizientenmatrix $\mathbf{A}$ hat vollen Spaltenrang $rg(\mathbf{A}_{n,m}) = m < n$:

$$\mathbf{x} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{b} \quad \text{ist die OLS - Lösung} \quad (2.46)$$

OLS steht dabei für „ordinary least squares“ - kleinste Fehlerquadrate. Diese Bezeichnung geht auf Gauss zurück; der Term $\mathbf{A}'\mathbf{A}$, der bei der Lösung linearer Gleichungssysteme in der **multivariaten Statistik** eine zentrale Rolle spielt, wird **Gauss-Transformation** genannt.

```
> A <- matrix(c(3, 1, 2, 4, 5, 6, 9, 7, 8), nrow=3, ncol=3, byrow=TRUE)
      [,1] [,2] [,3]
[1,]    3    1    2
[2,]    4    5    6
[3,]    9    7    8
> b <- c(2,4,8)
[1] 2 4 8
> x <- solve(A, b); round(x, 2)
[1] 0.67 0.67 -0.33
> A %o% x
      [,1]
[1,]    2
[2,]    4
[3,]    8
```

2.4.7 Eigenwerte und Eigenvektoren

Gibt es für eine quadratische Matrix $\mathbf{A}_{(n \times n)}$ einen Vektor $\mathbf{x}$ und einen Skalar λ (Lambda), so dass gilt:

$$\mathbf{Ax} = \lambda\mathbf{x}, \quad (2.47)$$

dann heißt λ **Eigenwert** und $\mathbf{x}$ **Eigenvektor** der Matrix $\mathbf{A}$.

Die Eigenwerte λ_i ($i = 1, \dots, n$) von $\mathbf{A}$ ergeben sich aus der Lösung des **charakteristischen Polynoms** $|\mathbf{A} - \lambda\mathbf{I}| = 0$

Für jeden Eigenwert λ_i kann der zugehörige Eigenvektor $\mathbf{x}_i$ durch die Lösung der **charakteristischen Gleichung** $(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}$ bestimmt werden.


```

> A <- matrix(c(3, 1, 2, 4), nrow = 2, ncol=2, byrow=TRUE);A
     [,1] [,2]
[1,]    3    1
[2,]    2    4
> l <- eigen(A)$values; round(l, 2)      # Eigenwerte in A
[1] 5 2
> x <- eigen(A)$vectors; round(x, 2)    # Eigenvektoren in A
     [,1] [,2]
[1,] -0.45 -0.71
[2,] -0.89  0.71

```

Für eine reguläre, reellwertige und symmetrische Matrix $A_{(n \times n)}$ gilt:

- Alle Eigenwerte von A sind reelle Zahlen.
- Die zu verschiedenen Eigenwerten gehörenden Eigenvektoren sind orthogonal.
- Zu A gehört eine **orthogonale Matrix** P mit der eine so genannte **Diagonalisierung** durchgeführt werden kann:

$$P'AP = \Delta \quad \text{bzw.} \quad A = P\Delta P'$$

Dabei ist Δ (Delta) eine **Diagonalmatrix**, deren Diagonalelementen gerade die Eigenwerte von A sind. Die Spalten der Matrix P bestehen aus paarweise orthonormalen Eigenvektoren von A , d.h. $PP' = I$.

Von besonderer Bedeutung in der multivariaten Statistik, insbesondere in der Faktorenanalyse, sind **quadratische Formen** einer symmetrischen Matrix $A_{(n \times n)}$:

$$Q = x'Ax = \sum_{i=1}^n a_{ii}^2 + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n a_{ij}x_i x_j \quad (2.48)$$

- Eine Matrix A ist **positiv definit** ($Q > 0 \quad \forall x \neq 0$), wenn alle Eigenwerte größer sind als Null: $\lambda_i > 0 \quad \forall i$.
- Eine Matrix A ist **positiv semidefinit**, wenn $\lambda_i \geq 0$ und mindestens ein $\lambda_i = 0$.
- Eine Matrix A ist **negativ definit** ($Q < 0 \quad \forall x \neq 0$), wenn alle Eigenwerte kleiner sind als Null: $\lambda_i < 0 \quad \forall i$.
- Eine Matrix A ist **negativ semidefinit**, wenn $\lambda_i \leq 0$ und mindestens ein $\lambda_i = 0$.

Mit Hilfe der orthogonalen Matrix P lässt sich eine so genannte **orthogonale Transformation** einer Matrix $A_{(n \times n)}$ definieren:

$$y := P'x \quad \Rightarrow \quad x = Py \quad (2.49)$$

$$\Rightarrow x'Ax = (Py)'APy = y'P'APy = y'\Lambda y = \sum_{i=1}^n \lambda_i y_i^2$$

2.5 Funktionen

- Lineare Funktionen
- Nichtlineare Funktionen
- Periodische Funktionen
- Exponentialfunktion und logarithmische Funktion
- Lineare Abhängigkeit, Rang einer Matrix
- Flächen unter einer Funktion - Integrale

Eine Funktion ist eine **Zuordnungsvorschrift**: Wie jedem Sitzplatz in einem Theater bei jeder Vorstellung eine bestimmte Eintrittskarte zugeordnet ist, so ordnet eine Funktion jedem Element einer Menge (Wertemenge, Wertebereich) ein bestimmtes Element einer anderen Menge (Bildmenge) zu. Dabei handelt es sich in der Regel jeweils um die Menge der reellen Zahlen ($\mathbb{R}$). Im einfachsten Fall ist jedem Wert der **unabhängigen Variablen** x ein bestimmter Wert der **abhängigen Variablen** y zugeordnet: $y = f(x)$ (sprich: y gleich f von x) (das Funktionssymbol $f(x)$ hat Euler eingeführt); die unabhängige Variable x heißt Argument. So ist z. B. für die Funktion $y = x^3$ dem Argument $x = 2$ der Funktionswert $y = 2^3 = 8$ zugeordnet. Für die graphische Darstellung derartiger Zuordnungen wird das Kartesische Koordinatensystem verwendet, indem zwei Koordinaten (senkrecht zueinander), die **y-Koordinate (Ordinate)** und die **x-Koordinate (Abszisse)** eine Fläche aufspannen, in der Wertepaare (x, y) durch Punkte dargestellt werden können. Das Bild der Funktion im Koordinatensystem wird auch als **Graph der Funktion** bezeichnet.

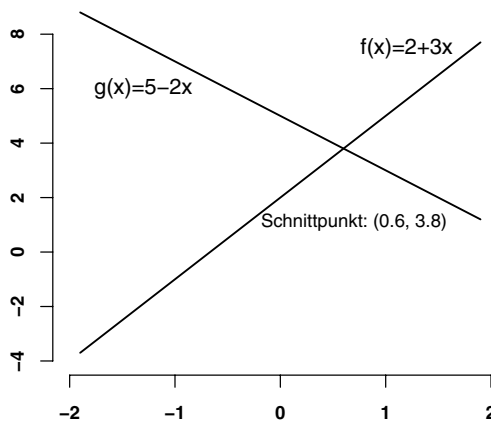


Abb. 2.6. Graphische Darstellung der linearen Funktionen $y = f(x) = 2 + 3x$ und $y = g(x) = 5 - 2x$

2.5.1 Lineare Funktionen

Eine lineare Funktion wird im einfachsten Fall durch die folgende Funktionsgleichung beschrieben:

$$\boxed{y = a + bx} \quad (2.50)$$

Der Graph der Funktion ist eine Gerade, die die Ordinate bei dem Wert a schneidet (**Achsenabschnitt**) und die die **Steigung** b aufweist. Ist $b < 0$, dann fällt die Gerade, ist $b > 0$, dann steigt die Gerade. Für $b = 0$ liegt die Gerade parallel zur Abszisse. Insbesondere ist $b = \tan(\alpha)$, wobei α den Winkel angibt, unter dem die Gerade die Abszisse schneidet.

Die Koordinaten für den **Schnittpunkt zweier Geraden** (x_S, y_S) , bzw. für den Schnittpunkt mit der Abszisse bestimmt man aus der Lösung der entsprechenden linearen Gleichungen.

Beispiel: Darstellung der linearen Funktionen $f(x) = 2 + 3x$ und $g(x) = 5 - 2x$ im kartesischen Koordinatensystem (Abbildung 2.6):

$$\begin{aligned} \rightarrow (2 + 3x) &= (5 - 2x) \rightarrow (2 + 3x) - (5 - 2x) = 0 \rightarrow -3 + 5x = 0 \rightarrow x = 3/5 = 0,6 \\ \rightarrow x_S &= 0,6 \quad \text{und} \quad y_S = 5 - 2x = 5 - 2 \cdot 0,6 = 3,8. \end{aligned}$$

Anmerkung zur Darstellung: Kurven werden im Folgenden in **R** mit der Funktion **plot()** gezeichnet. Die Koordinatenachsen liegen grundsätzlich am linken (y-Achse) und am unteren Rand (x-Achse) der Abbildung, ohne dass sich die beiden Achsen schneiden.

In einer **multiplen linearen Funktion** wird der Zusammenhang zwischen einer abhängigen Variablen y und mehreren unabhängigen Variablen x_i ($i = 1, \dots, n$) beschrieben. Neben einem konstanten Term a wird für jede unabhängige Variable ein entsprechender Koeffizient eingeführt:

$$y = a + b_1x_1 + b_2x_2 + \dots, b_nx_n$$

2.5.2 Nichtlineare Funktionen

2.5.2.1 Polynomfunktionen

Die allgemeine Form einer Polynomfunktion ist

$$y = a + b_1x + b_2x^2 + \dots + b_nx^n \quad (2.51)$$

n kennzeichnet darin den **Grad der Funktion** (des Polynoms). Speziell für $n = 2$ führt dieser Ansatz auf **quadratischen Funktionen**, deren Graphen durch **Parabeln** im Koordinatensystem dargestellt werden:

$$y = a + b_1x + b_2x^2$$

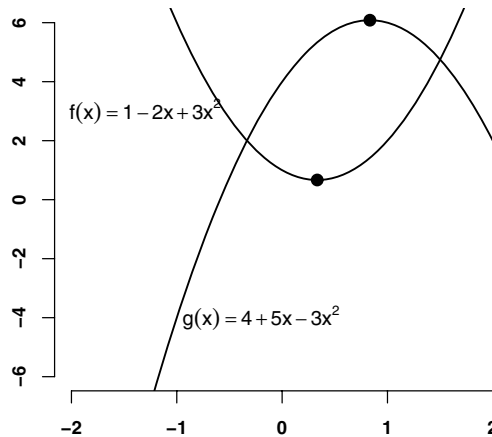


Abb. 2.7. Graphische Darstellung der quadratischen Funktionen $y = f(x) = 1 - 2x + 3x^2$ und $y = g(x) = 4 + 5x - 3x^2$

Der **Scheitelpunkt** einer Parabel wird durch den maximalen (minimalen) Funktionswert bestimmt. Er kann durch Umformung der Funktionsgleichung in die so genannte **Scheitelform** bestimmt werden.

$$y = b_2(x - s_x)^2 + s_y$$

$$\text{mit } s_x = -\frac{b_1}{2b_2} \quad \text{und} \quad s_y = a - \frac{b_1^2}{4b_2}$$

Ein weitere Möglichkeit zur Bestimmung des Scheitelpunktes erfolgt über die 1. Ableitung mit $f'(x) = 0$ (Bestimmung von Maximum bzw. Minimum).

2.5.3 Periodische Funktionen

Periodische (**trigonometrische**) Funktionen können am **Einheitskreis** (Radius = 1) definiert werden oder sie ergeben sich (für spitze Winkel $< 90^\circ$) aus den Seitenverhältnissen an einem rechtwinkligen Dreieck. Die abhängige Variable x wird dabei entweder als **Winkel** (0° bis 360°) oder als reelle Zahl im **Bogenmaß** (0 bis $2\pi \equiv 360^\circ$) angegeben ($B = (W/180)\pi$).

Sinus: $y = \sin(x) = \frac{\overline{BC}}{\overline{OB}} = \frac{\text{Gegenkathete}}{\text{Hypotenuse}}$	(2.52)
Kosinus: $y = \cos(x) = \frac{\overline{AB}}{\overline{OB}} = \frac{\text{Ankathete}}{\text{Hypotenuse}}$	
Tangens: $y = \tan(x) = \frac{\overline{CD}}{\overline{AD}} = \frac{\text{Gegenkathete}}{\text{Ankathete}}$	
Kotangens: $y = \cotg(x) = \frac{\overline{AE}}{\overline{DE}} = \frac{\text{Ankathete}}{\text{Gegenkathete}}$	

Der **Einheitskreis** (s. Abbildung 2.8) wird im kartesischen Koordinatensystem durch die Funktion $x^2 + y^2 = 1$ dargestellt. Bildlich entsteht er durch einen „Zeiger“ der Länge 1, der sich entgegengesetzt zum Uhrzeigersinn dreht. Betrachtet werden entsprechend x -Werte im Bereich $[-1, 0, -1, 0, +1]$ (dieses entspricht den Zeigerwinkeln α im Bereich $[0, 90^\circ, 180^\circ, 270^\circ, 360^\circ]$). Die Funktionswerte zu den **trigonometrischen Funktionen** können dann durch entsprechende Seitenverhältnisse bzw. Strecken nach (2.52) berechnet werden. Die in der Definition verwendeten Strecken sind in der Abbildung am Einheitskreis gekennzeichnet.

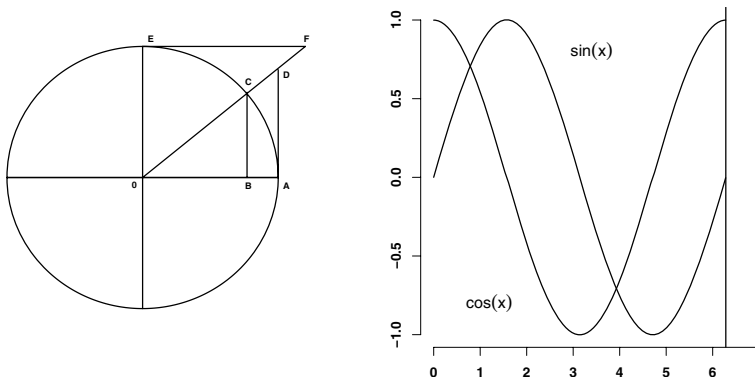


Abb. 2.8. Graphische Darstellung zum Einheitskreis und zu den periodischen Funktionen $y = f(x) = \sin(x)$ und $y = g(x) = \cos(x)$ im Intervall $[0, 2\pi]$

Die wichtigsten Formeln zu den Funktionen in der **Trigonometrie** sind in einer Übersicht (2.53) zusammengestellt:

$\sin^2(x) + \cos^2(x) = 1$	$\frac{\sin(x)}{\cos(x)} = \tan(x)$	$\sin(x) \cdot \cos(x) = 1$	(2.53)
$\frac{\cos(x)}{\sin(x)} = \cotg(x)$	$\tan(x) \cdot \cotg(x) = 1$		

2.5.4 Exponentialfunktion und logarithmische Funktion

Die **Exponentialfunktion** $y = a^x$ (für $a = e = 2,718282$ die natürliche Exponentialfunktion) nimmt nur positive reelle Werte an. Für $a > 1$ wächst sie monoton von 0 bis ∞ ; für $0 < a < 1$ nimmt sie monoton von ∞ bis 0 ab.

Die **logarithmische Funktion** $y = \log_a x$ $a > 0$ ist die Umkehrfunktion zur Exponentialfunktion; für $a = e$ natürliche logarithmische Funktion. Ihr Graph ist das Spiegelbild an der Winkelhalbierenden im kartesischen Koordinatensystem.

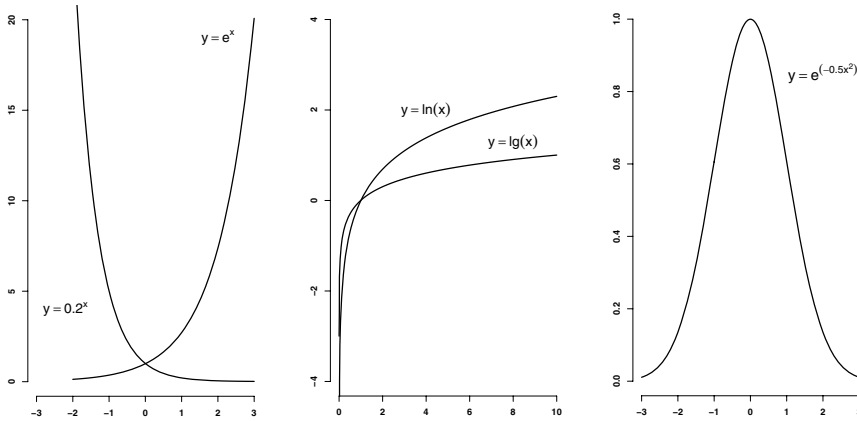


Abb. 2.9. Graphische Darstellung der Exponentialfunktionen $y = e^x$ und $y = (\frac{1}{5})^x$, der Logarithmusfunktionen $y = \ln(x)$ und $y = \log_{10}(x)$ und der speziellen Funktion $y = e^{-0.5x^2}$

Von besonderer Bedeutung in der Statistik ist die Funktion $y = e^{-\frac{1}{2}x^2}$. Sie wird als Grundlage für die **Dichtefunktion der Standardnormalverteilung** einer stetigen Zufallsvariable genutzt. Ihr Verlauf ist in Abbildung 2.9 dargestellt.

2.5.5 Flächen unter einer Funktion - Integrale

Der Fläche, die innerhalb bestimmter Grenzen (in einem festen **Intervall** $[a, b]$) durch die x-Achse und den Graphen einer Funktion $f(x)$ begrenzt wird (kurz die Fläche unter der Funktion), ist in zahlreichen Situationen von großer Bedeutung, so zum Beispiel in der Pharmakokinetik als **AUC (area under curve)** zur Kennzeichnung von aufgenommenen / ausgeschiedenen Mengen einer Substanz in Abhängigkeit von der Zeit. In der Statistik kommt der Fläche unter der Funktion eine zentrale Bedeutung zu, zunächst in dem Begriff der **Wahrscheinlichkeitsdichte**, und darauf aufbauend in dem Modell der **Verteilungsfunktion**, mit der wesentliche Eigenschaften von (speziell stetigen) Zufallsvariablen beschrieben werden.

Für die Bestimmung des Flächeninhalts wird das Intervall $[a, b]$ in n gleiche Teile zerlegt. Die Teilpunkte sind dann bestimmt durch:

$$x_k = a + k \cdot \Delta x \quad \text{mit} \quad k = 0, 1, 2, \dots, n \quad \text{und} \quad \Delta x := \frac{b - a}{n}$$

Zwischen den Teilpunkten kann die Fläche unter der Kurve durch die Summe einzelner Rechtecke (genauer noch durch Trapeze, vgl. Abbildung 2.10) näherungsweise gekennzeichnet werden. Dabei ist die Annäherung um so besser, je schmalere die Breite der Rechtecke gewählt wird. Allgemein gilt:

$$F = \int_a^b f(x)dx = \lim_{n \rightarrow \infty} \sum_{k=1}^n f(x_k) \cdot \Delta x \quad (2.54)$$

Den linken Teil der Formel 2.54 nennt man **das bestimmte Integral der Funktion $f(x)$** im Intervall $[a, b]$. Die praktische Berechnung von Integralen basiert auf der sogenannten **Stammfunktion** $F(x)$ mit $F'(x) = f(x)$, die auf die Mathematiker Newton (1643-1727) und Leibniz (1646-1716) zurückgeht:

$$\int_a^b f(x)dx = F(b) - F(a) = F(x) \Big|_a^b \quad (2.55)$$

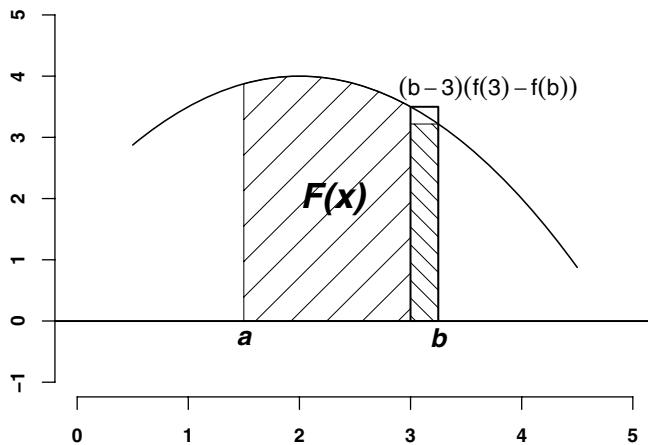


Abb. 2.10. Fläche unter der Kurve im Intervall $[a, b]$

2.6 Kombinatorik

- Permutationen
- Kombinationen - der Binomialkoeffizient
- Kombinationen mit Wiederholung
- Kombinationen mit Berücksichtigung der Anordnung
- Zerlegung einer Menge
- Das Pascalsche Dreieck
- Der Multinomialkoeffizient

2.6.1 Permutationen

Jede Reihenfolge, in der eine Menge von n verschiedenen Elementen angeordnet ist, bezeichnet man als **Permutation**. Das Programm **R** stellt in dem Paket **library(combinat)** [Cha02] spezielle Funktionen zum Permutieren und Kombinieren von Objekten zur Verfügung.

```
> library(combinat)
> x <- c("a", "b", "c")
> permn(x)
```

```
[[1]] [1] "a" "b" "c"
[[2]] [1] "a" "c" "b"
[[3]] [1] "c" "a" "b"
[[4]] [1] "c" "b" "a"
[[5]] [1] "b" "c" "a"
[[6]] [1] "b" "a" "c"
```

Insgesamt gibt es [vgl. (2.16 auf Seite 28)]

$$n(n-1) \cdot \dots \cdot 1 = n! \quad (\text{gelesen: } n\text{-Fakultät}) \quad (2.56)$$

verschiedene Permutationen. Für die Auswahl des 1. Elements gibt es nämlich n Möglichkeiten, für die Auswahl des nächsten nur noch $n-1$, da ein Element schon ausgewählt wurde. Diese Anzahlen müssen miteinander multipliziert werden, um die Gesamtzahl der möglichen Permutationen zu erhalten. Folgende Sonderfälle sind besonders zu beachten:

$$1! = 1 \quad \text{und auch} \quad 0! = 1$$

Eine Erklärung für $0! = 1$ liefert $(k+1)! = (k+1)k!$ und $k! = \frac{(k+1)!}{k+1}$.

Es gibt $n!$ Möglichkeiten, um n unterschiedliche Objekte in einer Reihe anzuordnen.



Beispiel: Acht unterschiedliche Bücher lassen sich auf $8! = 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 40320$ verschiedenen Arten nebeneinander ins Regal stellen.

```
> n <- 8
> prod(1:n)           # Produkt der Zahlen von 1 bis n
[1] 40320
```

Einen Sonderfall stellen **kreisförmige Permutationen** dar:

n unterschiedliche Objekte lassen sich auf $(n-1)!$ Arten kreisförmig anordnen.



Beispiel: Wie viele ungleiche Halsketten lassen sich allein durch unterschiedliche Anordnungen von 8 farblich unterschiedlichen Steinen herstellen? $(8-1)!/2 [= 2520]$; denn die Kette kann umgedreht werden.

In **Stammbäumen** lassen sich zum Beispiel drei Objekte A , B und C hinsichtlich ihrer Ähnlichkeit oder Verwandtschaft auf drei Arten anordnen:

$$A B C \quad A C B \quad B C A.$$

Für $n \geq 2$ Objekte (z. B. Pflanzen- oder Tiergattungen) gibt es

$$N = \frac{(2n-2)!}{2^{n-1}(n-1)!} \quad (2.57)$$

Stammbäume. Für $n = 20$ erhält man bereits

$$N = \frac{38!}{2^{19} \cdot 19!} \approx \frac{5,2302 \cdot 10^{44}}{5,2429 \cdot 10^5 \cdot 1,2165 \cdot 10^{17}} \approx 8,20 \cdot 10^{21}.$$

```
> n <- 20
> prod(1:(2*n-2)) / (2^(n-1)*prod(1:(n-1)))
[1] 8.200795e+21
```

k-Permutationen: Will man nicht alle Elemente anordnen, sondern nur einen Teil, also k Elemente aus n Elementen auswählen und diese anordnen, so hat man nur das Produkt über die ersten k dieser n Faktoren zu bilden, also

$$n(n-1) \dots (n-k+1) = \frac{n!}{(n-k)!} \quad (2.58)$$

Man bezeichnet diese Anordnungen als k -Permutationen, abgekürzt durch $(n)_k$.

Beispiel: Wenn aus 30 Vereinsmitgliedern ein Vorsitzender, ein Schriftführer und ein Kassenwart gewählt werden sollen, so gibt es hierfür rein theoretisch $30!/(30-3)! = 30!/27! = 30 \cdot 29 \cdot 28 = 24360$ Möglichkeiten.



2.6.2 Kombinationen - der Binomialkoeffizient

Häufig interessiert die **Zahl der Teilmengen mit k Elementen aus einer Menge von n Elementen**, die Anzahl „ k -elementiger“ Teilmengen. Für die Anzahl der Möglichkeiten, k Elemente in einer bestimmten Reihenfolge auszuwählen, haben wir $n!/(n-k)!$ erhalten. Hierbei wird aber jede Reihenfolge der k Elemente als eigene Möglichkeit angesehen. Wir haben also die Gesamtzahl durch die Anzahl der Möglichkeiten, k Elemente anzuordnen, zu dividieren, d. h. durch $k!$. Somit ergibt sich als Zahl möglicher Teilmengen mit k Elementen aus einer Menge von n Elementen, d. h. als Zahl k -elementiger Teilmengen von n Elementen (oder je k gleichzeitig):

$$\frac{n!}{(n-k)!k!} = \binom{n}{k} \quad (\text{gelesen: } n \text{ über } k) \quad (2.59)$$

Diese Größe heißt **Binomialkoeffizient** oder Euler-Symbol (nach Leonhard Euler: 1707–1783). Für das Rechnen mit Binomialkoeffizienten beachte man insbesondere

$$\binom{n}{k} = \binom{n}{n-k} = \begin{cases} \frac{n!}{(n-k)!k!} & \text{für } k \leq n \\ 0 & \text{für } k > n \end{cases} \quad (2.60)$$

und die Sonderfälle:

$$\binom{n}{0} = 1 = \binom{n}{n} \quad \binom{n}{1} = n = \binom{n}{n-1} \quad \binom{0}{0} = 1$$

Häufig findet man auch die Schreibweise:

$${}_nC_x = \binom{n}{x} = \frac{n!}{(n-x)!x!} = \binom{n}{n-x}$$

mit $0 \leq x \leq n$ ganzzahlig

Beispiel: $\binom{9}{7}$ wird als $\binom{9}{2} = \frac{9 \cdot 8}{2 \cdot 1} = 36$ berechnet, und nicht als

$$\frac{9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 \cdot 2 \cdot 1} = 36.$$



Die Berechnung der Binomialkoeffizienten im Programm **R** erfolgt über die Funktion **”choose”**:

```
> n <- 9
> k <- 7
> choose(n, k)           # Binomialkoeffizienten in R
[1] 36
```

2.6.2.1 Rechnen mit dem Binomialkoeffizienten - weitere Formeln

$$\binom{n+1}{x+1} = \frac{n+1}{x+1} \binom{n}{x} \quad \left| \quad \binom{n-1}{x-1} = \frac{x}{n} \binom{n}{x} \right| \quad \binom{n}{x-1} = \frac{x}{n-x-1} \binom{n}{x}$$

$$\text{sowie} \quad \binom{n+1}{x} = \frac{n+1}{n-x+1} \binom{n}{x}$$

$$\binom{n+1}{k} = \binom{n}{k} + \binom{n}{k-1} \quad \binom{n}{k+1} = \frac{n-k}{k+1} \binom{n}{k}$$

Binomialkoeffizienten lassen sich auch rekursiv berechnen (**Rekursionsformel**):

$$\begin{aligned} \binom{n+1}{k+1} &= \binom{n}{k} + \binom{n}{k+1} \\ &= \binom{n}{k} + \binom{n-1}{k} + \cdots + \binom{k}{k} \\ &= \sum_{i=0}^{n-k} \binom{n-i}{k} = \sum_{i=k}^n \binom{i}{k} \end{aligned} \quad (2.61)$$

Die **Reduktionsformel** für Binomialkoeffizienten lautet:

$$\binom{n-1}{k-1} = \frac{k}{n} \binom{n}{k} = \binom{n}{k} - \binom{n-1}{k} \quad (2.62)$$

2.6.3 Kombinationen mit Wiederholungen und mit Berücksichtigung der Anordnung

Eine Auswahl von k Elementen aus einer Menge von n Elementen ($n \geq k$) heißt eine Kombination von n Elementen zur k -ten Klasse, oder einfacher, eine Kombination k -ter Ordnung. Je nachdem, ob die gegebenen Elemente evtl. gleich oder alle verschieden sind, spricht man von Kombinationen mit oder ohne Wiederholung. Sollen zwei Kombinationen, die zwar genau dieselben k Elemente, aber in verschiedener Anordnung enthalten, als verschieden gelten, so spricht man von Kombinationen mit Berücksichtigung der Anordnung, andernfalls von Kombinationen ohne Berücksichtigung der Anordnung. Danach können wir 4 Modelle unterscheiden. Die Anzahl der Kombinationen k -ter Ordnung (je k zugleich) von n Elementen

I **ohne Wiederholung** und **ohne Berücksichtigung der Anordnung** ist durch den Binomialkoeffizienten gegeben.

$$\binom{n}{k} = \frac{n!}{(n-k)!k!} = \frac{n(n-1) \cdot \dots \cdot (n-k+1)}{k!} \quad (2.63)$$

II **ohne Wiederholung**, aber **mit Berücksichtigung der Anordnung** ist gleich

$$\binom{n}{k} k! = \frac{n!}{(n-k)!} = n(n-1) \cdot \dots \cdot (n-k+1) \quad (2.64)$$

Beispiel: Bewerben sich n Personen um k unterschiedliche Preise (1., 2., ..., k . Preis), dann kann die Preisverteilung auf $\binom{n}{k} k!$ unterschiedliche Arten folgen: 10 Rennfahrer kämpfen um 3 Medaillen (G,S,B); dann kann die Preisverteilung auf $\binom{10}{3} 3! = 120 \cdot 6 = 720$ unterschiedliche Arten erfolgen. Für die Goldmedaille gibt es 10 Möglichkeiten, für die silberne kommen 9 Rennfahrer in Frage und für die bronzene 8, d. h. insgesamt $10 \cdot 9 \cdot 8 = 720$ unterschiedliche Arten der Preisverteilung.



```
> choose(10, 3) * prod(1:3)
[1] 720
```

III **mit Wiederholung**, aber **ohne Berücksichtigung der Anordnung** ist gleich

$$\binom{n+k-1}{k} = \frac{(n+k-1)!}{k!(n-1)!} \quad (2.65)$$

Beispiel: Sind von fünf verschiedenen Bonbonsorten jeweils 12 Bonbons in eine Tüte abzufüllen, so gibt es hierfür $\binom{5+12-1}{12} = \binom{16}{12} = 1820$ Möglichkeiten.



```
> choose(5+12-1, 12)
[1] 1820
```

IV **mit Wiederholung** und **mit Berücksichtigung der Anordnung** ist gleich:

$$n^k \quad (2.66)$$

Beispiel: Wie viele „Wörter“ aus maximal drei Buchstaben lassen sich aus den 26 Buchstaben des Alphabets bilden, wenn Wiederholungen zugelassen werden? Zunächst lassen sich 26 „Wörter“ aus einem Buchstaben bilden, dann 26^2 aus zwei Buchstaben und 26^3 aus drei Buchstaben, insgesamt somit $26 + 26^2 + 26^3 = 18\,278$.



```
> 26^3 + 26^2 + 26
[1] 18278
```

2.6.4 Zerlegung einer Menge

Beispielsweise gibt es für die dreielementige Menge $\{A, B, C\}$

$$(\bar{A}, \bar{B}, \bar{C}) + (A) + (B) + (C) + (A, B) + (A, C) + (B, C) + (A, B, C)$$

$$2^3 = 8 = \binom{3}{0} + \binom{3}{1} + \binom{3}{2} + \binom{3}{3} = 1 + 3 + 3 + 1 = 8 \quad \text{mögliche Teilmengen.}$$

Allgemein kann eine Menge in Teilmengen mit $0, 1, 2, \dots, n$ Elementen zerlegen, deren Anzahl 2^n beträgt:

$$\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \dots + \binom{n}{n-1} + \binom{n}{n} = \sum_{k=0}^n \binom{n}{k} = 2^n \quad (2.67)$$

Die Zahl der Auswahlmöglichkeiten von k aus n Elementen beträgt für $k \geq 1$ (dabei verzichtet man speziell auf das erste Glied der Reihe, die Betrachtung der leeren Menge):

$$\binom{n}{1} + \binom{n}{2} + \dots + \binom{n}{n} = \sum_{k=1}^n \binom{n}{k} = 2^n - 1 \quad (2.68)$$



Beispiel: Von den ursprünglichen 32 Zähnen bilden die beim älteren Menschen vorhandenen Zähne eine Teilmenge; insgesamt gibt es $2^{32} = 4,3 \cdot 10^9$ verschiedene Teilmengen. Ohne die leere Menge $\binom{n}{0} = 1$ ist dann die Zahl der Teilmengen einer n -elementigen Menge zu jeweils k Elementen, $k \leq n$, gleich $2^n - 1$.



Beispiel: Ausstattungs-Varianten beim Autokauf mit $k = 1, 2, \dots, n$ verschiedenen Komponenten: Drei beliebig miteinander kombinierbare Zusatzausrüstungen (z. B. Antiblockiersystem, Stahlkurbeldach und schwenkbare Scheinwerfer) führen z. B. zu

$$\sum_{k=1}^n \binom{3}{k} = \binom{3}{1} + \binom{3}{2} + \binom{3}{3} = 3 + 3 + 1 = 7$$

Varianten. Häufig liegt die Zahl der Varianten (V) unter diesem Maximum, da gewisse Einzelvarianten nicht in Frage kommen.

Für 4 Karosserie-, 5 Motor-, 2 Getriebe-Varianten und 5 unterschiedliche Ausstattungspakete ergeben sich $4 \cdot 5 \cdot 2 \cdot 5 = 200$ Varianten; 10 Zusatzausstattungen bringen allein

$$\sum_{k=1}^{10} \binom{10}{k} = 10 + 45 + 120 + 210 + 252 + 210 + 120 + 45 + 10 + 1 = 1023 \text{ Varianten.}$$

```
> sum(choose(10, 1:10))
[1] 1023
```

Mit den 200 Varianten sind das (ohne Farbvarianten) schon $200 \cdot 1023 = 204\,600$ Varianten.

2.6.5 Das Pascalsche Dreieck

Die Binomialkoeffizienten ergeben sich elegant aus dem unten aufgeschriebenen Pascalschen Dreieck (Pascal 1623–1662): Werden zwei nebeneinander stehende Zahlen des Dreiecks addiert,

so erhält man die darunter auf Lücke stehende Zahl. Die Gesetzmäßigkeit des Pascalschen Dreiecks lautet:

$$\boxed{\binom{n}{x} + \binom{n}{x+1} = \binom{n+1}{x+1}} \quad (2.69)$$

Die Besonderheit des Dreiecks besteht darin, dass es unmittelbar die Anzahl der Kombinationen mit einem festen n zeilenweise übersichtlich wiedergibt:

	Zeile n	Binomialkoeffizienten
1	0	$\binom{0}{0}$
1 1	1	$\binom{1}{0} \quad \binom{1}{1}$
1 2 1	2	$\binom{2}{0} \quad \binom{2}{1} \quad \binom{2}{2}$
1 3 3 1	3	$\binom{3}{0} \quad \binom{3}{1} \quad \binom{3}{2} \quad \binom{3}{3}$
1 4 6 4 1	4	$\binom{4}{0} \quad \binom{4}{1} \quad \binom{4}{2} \quad \binom{4}{3} \quad \binom{4}{4}$
1 5 10 10 5 1	5	$\binom{5}{0} \quad \binom{5}{1} \quad \binom{5}{2} \quad \binom{5}{3} \quad \binom{5}{4} \quad \binom{5}{5}$

Abb. 2.11. Das Pascalsche Dreieck von $n = 1, \dots, 5$

Die Entwicklung der Binomialkoeffizienten nach dem Pascalschen Dreieck aus Abbildung 2.11 wird auch bei der Auflösung binomischer Formeln deutlich, z.B. für Zeile $n = 5$:

$$(a+b)^5 = \sum_{i=0}^5 \binom{5}{i} a^{(5-i)} b^i = a^5 b^0 + 5a^4 b^1 + 10a^3 b^2 + 10a^2 b^3 + 5a^1 b^4 + a^0 b^5$$

Einige Identitäten zum Pascalschen Dreieck

1. Jede Zeile ist symmetrisch, d. h.

$$\binom{n}{x} = \binom{n}{n-x}$$

2. Jeder Binomialkoeffizient ist gleich der Summe der beiden über ihm stehenden, d. h.

$$\binom{n}{x} = \binom{n-1}{x-1} + \binom{n-1}{x}$$

3. Für die Zeilensumme der Binomialkoeffizienten gilt:

$$\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \dots + \binom{n}{n} = 2^n$$

4. Für die Zeilensumme der quadrierten Binomialkoeffizienten gilt:

$$\binom{n}{0}^2 + \binom{n}{1}^2 + \binom{n}{2}^2 + \dots + \binom{n}{n}^2 = \binom{2n}{n}$$

5. Für alle Zeilen ist die alternierende Summe gleich Null:

$$\binom{n}{0} - \binom{n}{1} + \binom{n}{2} - \binom{n}{3} + \dots + (-1)^n \binom{n}{n} = 0$$

Sind n gleiche Objekte in k unterschiedliche Fächer zu legen, kein Fach darf leer bleiben, so gibt es hierfür $\binom{n-1}{k-1}$ mit $n \geq k$ unterschiedliche Arten.



Beispiel: $n = 4, k = 3$, d. h. $\binom{4-1}{3-1} = \binom{3}{2} = \binom{3}{1} = 3$.

2.6.6 Der Multinomialkoeffizient

Wenn n Elemente in k Gruppen angeordnet werden, so dass $n_1 + n_2 + \dots + n_k = n$, wobei $n_1, n_2, \dots, n_k$ die Anzahl der Elemente pro Gruppe bezeichnet, dann gibt es

$$\frac{n!}{n_1! \cdot n_2! \cdot \dots \cdot n_k!} \quad (2.70)$$

unterschiedliche Möglichkeiten, die n Elemente in diese k Gruppen zu gruppieren (Multinomialkoeffizient).



Beispiel: Zehn Studenten sollen in zwei Gruppen zu je 5 Basketballspielern eingeteilt werden. Wie viele unterschiedliche Teams können gebildet werden?

$$\frac{10!}{5! \cdot 5!} = \frac{3\,628\,800}{120 \cdot 120} = 252$$



Beispiel: Ein Satz von 52 Spielkarten soll so unter 4 Spielern verteilt werden, dass jeder 13 Karten enthält. Wie viele unterschiedliche Möglichkeiten gibt es für die Verteilung der Karten?

$$\frac{52!}{13! \cdot 13! \cdot 13! \cdot 13!} = \frac{8,0658 \cdot 10^{67}}{(6,2270 \cdot 10^9)^4} \simeq 5,36 \cdot 10^{28}.$$

```
> karten <- 52                                # Anzahl der Karten
> spieler <- 4                                # Anzahl der Spieler
> k.spiel <- karten/spieler                    # Anzahl Karten pro Spieler
> prod(1:karten)/(prod(1:k.spiel)^spieler)
[1] 5.364474e+28
```


Deskriptive Statistik

- Häufigkeiten
- Ordinaldaten
- Metrische Daten
- Häufigkeitsverteilung
- Konzentration; Gini-Index
- Maßzahlen für den Zusammenhang
- Nichtlineare Regression

Die Verfahren der deskriptiven Statistik können grundsätzlich nach vier Gesichtspunkten eingeteilt werden. Maßzahlen, die

1. eine **zentrale Tendenz** (Lage) von Beobachtungen / Messungen ausdrücken,
2. die eine **Streuung** oder Variabilität in den Beobachtungen / Messungen erfassen,
3. die die **Form** bzw. das Profil der (Häufigkeits-) Verteilung beschreiben und
4. die weitere spezielle Aspekte, z.B. den **Zusammenhang** oder eine **Assoziation** zwischen zwei Beobachtungsreihen, untersuchen.

Übersicht 6. Maßzahlen der deskriptiven Statistik nach den vier Skalentypen

Skalentyp	Lagemaße	Streuungsmaße	Assoziationsmaße
Nominalskala	absolute, relative und prozentuale Häufigkeiten (H), Dichtemittel (D)	Gini-Simpson-Index (V_G)	„bedingte“ Häufigkeiten in Tabellen, Kontingenzkoeffizient (K), Chancenverhältnis oder Odds Ratio (OR)
Ordinalskala	H , D , Medianwert ($\tilde{x}$), Perzentile, speziell Quartile: Q_1 und Q_2	Spannweite (Range R), Interquartilbereich (IQR), Median-Deviation ($\tilde{D}$)	Rangkorrelationskoeffizient nach Kendall (r_τ) oder nach Spearman (r_s)
Intervallskala	H , D , $\tilde{x}$, Perzentile, arithmetischer Mittelwert ($\bar{x}$)	R , IQR , $\tilde{D}$, Standardabweichung (s) und Varianz (s^2)	lineare Abhängigkeit - Korrelationskoeffizient (r) und Regressionskoeffizienten
Verhältnisskala	H , D , $\tilde{x}$, Perzentile, $\bar{x}$, geometrischer Mittelwert ($\bar{x}_G$), harmonischer Mittelwert ($\bar{x}_H$)	R , IQR , $\tilde{D}$, s , s^2 , Variationskoeffizient (V)	r_τ ; r_s ; r und Regressionskoeffizienten

3.1 Häufigkeiten

- Absolute und relative Häufigkeiten
- Sinnvolle Quotienten: Verhältniszahlen
- Prozentwerte
- Torten- und Balkendiagramme
- Tabellen
- Bedingte Häufigkeiten

3.1.1 Absolute und relative Häufigkeiten

Die Beschreibung nominal skalierten Daten erfolgt (univariat) stets über Häufigkeiten, d. h. zu einem Merkmal wird die **Häufigkeit des Auftretens** der einzelnen Merkmalsausprägungen in einer Stichprobe bzw. Gesamtheit von n -Fällen bestimmt. Für ein Merkmal mit k Ausprägungen gilt:

$$x_1 + x_2 + x_3 + \dots + x_k = \sum_{i=1}^k x_i = n \quad (3.1)$$

Die x_i mit $i = 1, \dots, k$ heißen **absolute Häufigkeiten**. Für das Verständnis und die Interpretation von Häufigkeiten ist die Angabe der Gesamtzahl n (Anzahl der Fälle) zwingend erforderlich. Ein Beispiel zur Darstellung von Häufigkeiten ist in Tabelle 3.1 dargestellt. Der **Modalwert (Dichtemittel, D)** kennzeichnet die Merkmalsausprägung, die am häufigsten auftritt.

Tabelle 3.1. Beispieldaten zur Häufigkeit der Blutgruppen des AB0-Systems in einer Untersuchung von $n=155$ Personen

Blutgruppe	A	B	AB	0
Anzahl (gesamt $n=155$)	69	17	7	62
relative Häufigkeit	0,45	0,11	0,05	0,40
prozentuale Häufigkeit	44,5%	11,0%	4,5%	40,0%

Für den Vergleich von Häufigkeiten innerhalb von Gesamtheiten oder Stichproben mit unterschiedlicher Anzahl von Fällen ist die **Normierung** durch die jeweilige Anzahl zu **relativen Häufigkeiten** sinnvoll. Die Summe der relativen Häufigkeiten ist nach der Formel 3.2 stets 1.

$$\frac{x_1}{n} + \frac{x_2}{n} + \frac{x_3}{n} + \dots + \frac{x_k}{n} = \sum_{i=1}^k \frac{x_i}{n} = 1 \quad (3.2)$$

Im Programm **R** werden die Häufigkeiten in Vektoren gespeichert; die Bezeichnungen zu den Merkmalsausprägungen können durch die Funktion **names()** ergänzt werden.

```
> absolut <- c(69, 17, 7, 62) # Blutgruppen A, B, AB, 0
> names(absolut) <- c("A", "B", "AB", "0"); absolut
  A  B AB 0
69 17  7 62
> anzahl <- sum(absolut); anzahl
[1] 155
> relativ <- absolut / anzahl; round(relativ, 2)
  A  B AB 0
0.45 0.11 0.05 0.40
```



```
> prozent <- relativ * 100; round(prozent, 1)
  A    B    AB    0
44.5 11.0  4.5 40.0
```

Eine Maßzahl für die **Variabilität**, hier besser für die **Dispersion** oder **Diversität**, der beobachteten Häufigkeiten in den einzelnen Merkmalsausprägungen ist der **Gini-Simpson-Index** [Sim49]:

$$V_G = \sum_{i=1}^k \frac{x_i}{n} \left(1 - \frac{x_i}{n}\right) = 1 - \sum_{i=1}^k \left(\frac{x_i}{n}\right)^2 \quad (3.3)$$

Der Gini-Simpson-Index nimmt den Wert 0 an, wenn alle Beobachtungen in eine Kategorie fallen (maximale Konzentration). Die größte Variabilität in den beobachteten Häufigkeiten eines Merkmals mit k Merkmalsausprägungen ist durch $\frac{(k-1)}{k}$ gegeben.

```
> Gini <- sum(relativ*(1-relativ)); Gini
[1] 0.6277627
```

3.1.2 Sinnvolle Quotienten: Verhältniszahlen

Verhältniszahlen sind Quotienten zweier Zahlen, zwischen denen sachlogisch eine Beziehung besteht oder zu vermuten ist. Eine typische Verhältniszahl ist die **relative Häufigkeit** bzw. die prozentuale Häufigkeit. So betrug der Anteil der **Knabengeburten** an der Gesamtzahl der Mädchen- und Knabengeburten in der Bundesrepublik Deutschland (BRD) in den Jahren 1965, 1969 und 1978 0,514 oder 51,4%. In den letzten 20 Jahren wurden bei uns jeweils 5% bis 6% mehr Knaben als Mädchen geboren. Dieser Knabenüberschuss wird nach etwa 55 Jahren ausgeglichen; danach folgt ein Frauenüberschuss.

Der Anteil der Knabengeburten ist eine typische Gliederungszahl. Sie entsteht, wenn das zahlenmäßige Verhältnis einer Teilmenge zur zugehörigen Gesamtmenge ausgedrückt wird. Gliederungszahlen dienen zur Darstellung der inneren Struktur einer Gesamtmenge und zum Vergleich mehrerer Gesamtmengen. Erwähnt sei z. B. die Aufgliederung der Erwerbstätigen mehrerer Länder nach Wirtschaftsbereichen, wobei deren Vergleichbarkeit gegeben ist bzw. durch Korrekturen ermöglicht wird. Man unterscheidet mindestens drei Arten von Verhältniszahlen:

1. **Gliederungszahlen** („Anteil“ für eine endliche Grundgesamtheit) die das zahlenmäßige Verhältnis einer Teilmenge zur zugehörigen Gesamtmenge ausdrücken; z. B. den Anteil der Lebendgeborenen an der Gesamtzahl der Geburten (Lebend- und Totgeborene):

$$\frac{\text{Zahl der Lebendgeborenen}}{\text{Gesamtzahl der Geburten}}$$

2. **Beziehungszahlen** („Dichte“; mit Maßeinheit und umkehrbar) die das zahlenmäßige Verhältnis zweier verschiedenartiger Mengen, die logisch miteinander verknüpft sind, ausdrücken; z. B. die Zahl der Lebendgeborenen zur Zahl der Frauen im gebärfähigen Alter (15–45 J.):

$$\frac{\text{Zahl der Lebendgeburten}}{\text{Zahl der Frauen im gebärfähigen Alter (15–45 J.)}}$$

Bei Beziehungszahlen bezieht man Ereignisse direkt auf den zugehörigen Bestand („**Verursachungszahlen**“) wie im Beispiel oder nicht, etwa km/Std., PKW/Einwohner, Schüler/Lehrer, Niederschlagsmenge/Quadratmeter („**Entsprechungszahlen**“). Den richtigen (Bezugs-) Nenner zu finden, ist mitunter schwierig. Wichtige Beziehungszahlen werden wir später z. B. als arithmetisches Mittel und als Varianz kennenlernen.

Bei Verursachungszahlen bezieht man Ereignismassen auf die zugehörigen Bestandsmassen: z. B. Zahl der Eheschließungen pro 1000 Einwohner (BRD, 1950: 10,7; 1960: 9,4; 1970: 7,3; 1980: 5,9; 1988: 6,5; 1990: 6,5; 1993: 5,5; 1995: 5,3; 1997: 5,2) oder die Zahl der Lebendgeborenen pro 1000 Einwohner (BRD, 1950: 16,2; 1960: 17,4; 1970: 13,4; 1980: 10,1; 1988: 11,0; D, 1990: 11,4; 1993: 9,8; 1995: 9,4; 1997: 9,9. Quelle: Stat. Jahrbuch 1999 für die Bundesrepublik Deutschland. Hrsg.: Stat. Bundesamt, Wiesbaden).

3. **Messzahlen** ("Teil zum anderen Teil"; um Vergleiche zu ermöglichen) die das zahlenmäßige Verhältnis einer Menge zu einer gleichartigen nebengeordneten Menge - beide sind logisch miteinander verknüpft - ausdrücken; z. B. das Verhältnis der Lebendgeborenen (Stat. Jb. 1999):

$$\frac{\text{Zahl der Knaben Geburten}}{\text{Zahl der Mädchen Geburten}} \quad (\text{BRD, 1973, 1974, 1976, 1980: 1,054; 1986: 1,054; 1987: 1,062; 1988: 1,058; 1989: 1,051; D, 1990: 1,057; 1991: 1,055; 1992: 1,052; 1993: 1,056; 1995: 1,054; 1996: 1,058; 1997: 1,055}).$$

Wird eine Reihe von Werten auf eine gemeinsame Basis bezogen, so heißen diese Verhältniszahlen **Messzahlen**, etwa die prozentuale monatliche Abweichung der Arbeitslosenziffer vom Jahresdurchschnitt. Man setzt auch den Umsatz für ein bestimmtes Jahr gleich 100 und berechnet für die Umsatzzahlen der folgenden Jahre entsprechende Messzahlen. Ein aus Messzahlen derselben Berichtsperiode oder Basisperiode gebildeter Mittelwert heißt **Indexzahl** oder **Index**.

Tabelle 3.2. Verschiedene Arten von Verhältniszahlen

Relativzahlen	Verhältnis	Beispiel
Gliederungszahlen	Unterordnung: Menge und Teilmenge	Altersaufgliederung der Bevölkerung
Beziehungszahlen	Nebenordnung: verschiedenartige Mengen	Bevölkerungszahl je km^2 der Fläche eines Landes
Messzahlen	Nebenordnung: gleichartige Mengen	Vergleich der Betriebsangehörigen zweier Werke eines Industriezweiges an einem Stichtag



Verhältniszahlen haben den Nachteil, dass die tatsächlichen (absoluten) Zahlen nicht erkennbar sind. Dafür sind sie diesen gegenüber jedoch leichter erfassbar und besser vergleichbar.

Hinweis: Bestandsmassen und Ereignismassen

Bestimmte interessierende Sachverhalte, z. B. Daten der Wirtschaft, werden nicht einmalig, sondern periodisch (etwa täglich, monatlich, jährlich, . . .) erhoben. Die Ergebnisse dieser Erhebungen in der Zeit bilden eine **Zeitreihe**. Man unterscheidet Zeitreihen von gemessenen Bestandsmassen und von gezählten Ereignismassen.

Bestandsmassen wie Bevölkerung, Lagerbestand und Hotel-Belegung sind durch eine mittlere Verweildauer charakterisiert. Der Bestand ändert sich durch Zu- und Abgänge. Eine Bestandsmasse, gewonnen durch Erhebung eines Bestandes am Stichtag, bezieht sich stets auf diesen Zeitpunkt, z. B. Lagerbestand am 1. jeden Monats. Andere Beispiele sind Bevölkerung, Spareinlagenbestand, Betriebe, Anbaufläche.

Ereignismassen sind Ereignisse, die zu bestimmten Zeitpunkten auftreten: Zu- und Abgänge (Ereignismassen werden daher auch **Bewegungsmassen** genannt) innerhalb eines Zeitraumes: Geburten, Todesfälle; Empfangenes, Ausgegebenes; Anmeldung neuer Gäste, Abmeldung von Gästen. Eine Ereignis- (oder Bewegungsmasse, gewonnen durch laufende Registrierung isolierter Ereignisse in der Zeit, bezieht sich dagegen auf einen Zeitraum, z. B. monatliche Geburtenzahlen einer

Klinik und jährliche Steuereinnahmen eines Landes. Den im vorangehenden Absatz genannten Bestandsmassen entsprechen die folgenden Ereignismassen: Geburten und Todesfälle, Ein- und Auszahlungen, Lohnsummen, Ernteertrag.

3.1.3 Prozentwerte

Die Quotienten aus Häufigkeiten, z. B. $2/5$, $12/30$ und $80/200$ liefern jeweils die Dezimalzahl 0,4, die nur dann, wenn der Nenner (die Basis) größer als 100 ist, als **Prozentzahl** 40% geschrieben werden darf und deren statistische Aussagekraft mit zunehmendem Nenner (z. B. $400/1000$) ansteigt.

Bedeutet x die Anzahl der Treffer unter n Beobachtungen der Stichprobe, dann ist $p = x/n$ die relative Häufigkeit. Die **prozentuale Häufigkeit** der Treffer in der Stichprobe ist

$$p = \frac{x}{n} \cdot 100\% \quad \text{mit} \quad n \geq 100 \quad (3.4)$$

Die Summe der prozentualen Häufigkeiten ergibt stets 100.

Für $n < 80$ gibt man „ x von n “ oder x/n nur als relative Häufigkeit an, jedoch keine Prozentzahlen.

Für $80 \lesssim n < 150$ sind Prozentzahlen ohne Kommastelle zu schreiben [für $80 \leq n < 100$ schreibt man, falls aus Vergleichsgründen erwünscht, z. B. $29/80 = 0,3625$ als „(36%)“], erst ab etwa $n = 2000$ mit zwei Stellen nach dem Komma. Beispiel: $p = 33/189 = 0,17460$ wird als Prozentzahl 17,5% angegeben.

Beachtet sei der Unterschied zwischen „**Prozente**“ und „**Prozentpunkte**“, der Differenz zweier Prozente: nimmt z. B. etwas von 70% auf 77% zu, so wächst es um 10% bzw. um sieben Prozentpunkte. Steigt dagegen etwas von 143% auf 144% an, d.h. akkurat um einen Prozentpunkt, so erhöht es sich nur um $1/143$, d.h. $100/143\% = 0,7\%$. Nur für die Zunahme von 100% auf 101% erhöht sich etwas um genau $1/100$, d.h. um $100/100\% = 1\%$. Je nachdem, welche Tendenz dem Leser suggeriert werden soll, wird mit Prozentpunkten oder mit Erhöhung um ...% „argumentiert“.

3.1.4 Torten- und Balkendiagramme

Graphische Darstellungen im weiteren Sinne sind besonders aus der Werbung bekannt. Man unterscheidet Linien-, Stab-, Flächen- und Körperdiagramme.

Das **Stab-** oder **Säulendiagramm** (Balkendiagramm) ist durch konstante Breite ausgezeichnet. Beim **Flächendiagramm** bevorzugt man Quadrate (Fläche = a^2) oder Kreise (Fläche = πr^2) - **Tortendiagramm** -, beim Körperdiagramm Würfel (Inhalt = a^3). Da Flächen und Körper in ihren Größenverhältnissen leicht falsch beurteilt werden können, sind Linien- und Stabdiagramme allen anderen Diagrammen an Klarheit überlegen. Ergänzen sich unterschiedliche Elemente einer Häufigkeitsverteilung zu 100%, dann vermittelt das 100%-Stab- bzw. **Rechteckdiagramm** eine gute Übersicht.

Für die graphische Darstellung von Häufigkeiten stehen im Programm **R** unter anderem die Funktionen **pie()** und **barplot()** zur Verfügung, z.B. Abbildung 3.1 mit den Blutgruppendaten aus Tabelle 3.1.

```
> pie(absolut, labels=c("A", "B", "AB", "0"),
+     col=c("white", "grey80", "black", "grey60"))
> barplot(absolut, names.arg=c("A", "B", "AB", "0"),
+         density=c(5, 10, 15, 20), angle=c(45, 135, 45, 135), col="black",
+         ylim=c(0, 70))
```

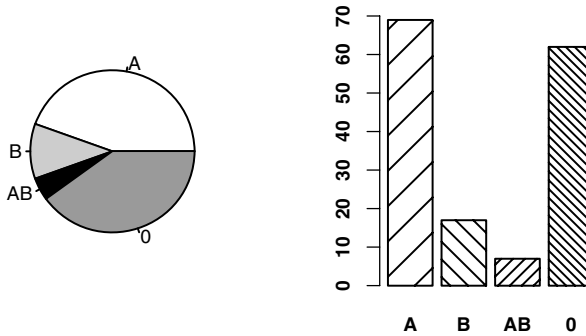


Abb. 3.1. Häufigkeiten der Blutgruppen im AB0-System bei $n=155$ Personen (Kreis- und Balkendiagramm)

Hierbei – wie bei allen Prozentangaben – muss die 100% entsprechende Anzahl der Gesamt- Stichprobenelemente, der Umfang der Stichprobe, im Diagramm selbst oder in der Legende vermerkt werden.

3.1.5 Tabellen

Die gemeinsame Häufigkeitsverteilung von zwei, an n Fällen beobachteten, Merkmalen wird in einer Tabelle (Schema in Tabelle 3.3) dargestellt. Dabei wird zwischen dem **Zeilenmerkmal** (r Ausprägungen) und dem **Spaltenmerkmal** (c Ausprägungen) unterschieden. Die beobachteten Häufigkeiten werden durch die Randsummen (**Zeilen-** und **Spaltensummen**) ergänzt.

Tabelle 3.3. Schema für die zweifache Klassifikation: Eine der beiden Merkmalsreihen dieser $r \cdot c$ -Tafel ist auch als Stichprobenreihe auffassbar (r -rows, c -columns)

1. Merkmal (r Zeilen)	2. Merkmal (c Spalten)	1	2	-	j	-	c	Zeilensumme
1		n_{11}	n_{12}	-	n_{1j}	-	n_{1c}	$n_{1.}$
2		n_{21}	n_{22}	-	n_{2j}	-	n_{2c}	$n_{2.}$
-		-	-	-	-	-	-	-
i		n_{i1}	n_{i2}	-	n_{ij}	-	n_{ic}	$n_{i.}$
-		-	-	-	-	-	-	-
r		n_{r1}	n_{r2}	-	n_{rj}	-	n_{rc}	$n_{r.}$
Spaltensumme		$n_{.1}$	$n_{.2}$	-	$n_{.j}$	-	$n_{.c}$	$n_{..} = n$ Ecksumme



Beispiel: Wird neben der Blutgruppe auch das Geschlecht der Personen erfasst, dann kann die gemeinsame Häufigkeitsverteilung der beiden Merkmale durch die Tabelle 3.4 dargestellt werden: Tabellen werden im Programm **R** in der **Struktur einer Matrix** gespeichert. Die Bezeichnungen zu den Zeilen- und Spaltenmerkmalen können durch einige spezielle Funktionen hinzugefügt werden. In dem Beispiel wird auch die Berechnung der relativen Häufigkeiten auf die Zeilen- und Spaltensummen (vgl. bedingte Häufigkeiten) mit den Funktionen **margin.table()** und **prop.table()** gezeigt.

Tabelle 3.4. Beispiel zur Häufigkeitsverteilung für die Blutgruppe nach dem Geschlecht für $n = 155$ Personen

Geschlecht / Blutgruppe	A	B	AB	0	Summe
männlich	30	10	5	40	85
weiblich	39	7	2	22	70
Summe	69	17	7	62	155

```

> absolut <- matrix(c(30, 10, 5, 40, 39, 7, 2, 22), nrow=2, byrow=T)
> colnames(absolut) <- c("A", "B", "AB", "0")
> rownames(absolut) <- c("maennlich", "weiblich")
> names(dimnames(absolut)) <- c("Geschlecht", "Blutgruppe"); absolut
      Blutgruppe
Geschlecht A B AB 0
maennlich 30 10 5 40
weiblich  39  7 2 22
> margin.table(absolut, 1)
Geschlecht maennlich weiblich
      85      70
> margin.table(absolut, 2)
Blutgruppe
A B AB 0
69 17 7 62
> round(prop.table(absolut, 1), 3)
      Blutgruppe
Geschlecht A B AB 0
maennlich 0.353 0.118 0.059 0.471
weiblich  0.557 0.100 0.029 0.314
> round(prop.table(absolut, 2), 3)
      Blutgruppe
Geschlecht A B AB 0
maennlich 0.435 0.588 0.714 0.645
weiblich  0.565 0.412 0.286 0.355

```

Die Häufigkeiten aus Tabellen werden in den entsprechenden Formeln doppelt indiziert, d. h. man unterscheidet zwischen einem **Zeilen-** und einem **Spaltenindex**. Für die Summen in einer Tabelle folgt daraus die allgemeine Darstellung entsprechend Tabelle 3.3:

$$\begin{aligned}
 \sum_{i=1}^r n_{ij} &= n_{.j} \\
 \sum_{j=1}^c n_{ij} &= n_{i.} \\
 \sum_{i=1}^r \sum_{j=1}^c n_{ij} &= n_{..} = n
 \end{aligned} \tag{3.5}$$

Mit der **Punktnotation** im Index wird jeweils der Index gekennzeichnet, über den die Summe gebildet wird. Die Summe über beide Indices (zweifach summiert, **Doppelsumme**) führt dann auf die Gesamtsumme.



3.1.5.1 Rechteckdiagramm und Mosaikplot

Für die graphische Darstellung von Häufigkeiten aus Tabellen werden Rechteckdiagramme oder Mosaikplots verwendet.

In einem **Rechteckdiagramm** werden Rechtecke gestapelt dargestellt, den Häufigkeiten eines Merkmals in den einzelnen Kategorien des anderen Merkmals entsprechend. Das Prinzip der **Mosaikplots** ist ähnlich, allerdings wird die Aufteilung der Rechteckflächen mit Bezug auf die entsprechenden Zeilen- bzw. Spaltensummen vorgenommen (vgl. Abbildung 3.2).

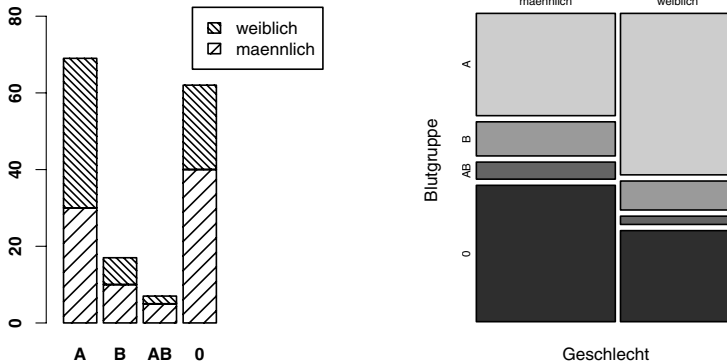


Abb. 3.2. Häufigkeiten der Blutgruppen im ABO-System nach dem Geschlecht für $n=155$ Personen (Rechteckdiagramm und Mosaikplot)

3.1.6 Bedingte Häufigkeiten

Ein Zusammenhang bzw. eine Abhängigkeit zwischen zwei nominal-skalierten Merkmalen zeigt sich im Rahmen einer deskriptiven Statistik in den „**bedingten**“ Häufigkeiten. Dazu werden die relativen Häufigkeiten auf die Zeilen- oder Spaltensummen bezogen bestimmt und miteinander verglichen. Stimmen z.B. die zeilenweise berechneten relativen Häufigkeiten mit dem Anteil in der jeweiligen Merkmalskategorie überein, dann sind die beiden Merkmale **unabhängig** voneinander.

$$\begin{aligned} \frac{n_{ij}}{n_{i.}} &\approx \frac{n_{.j}}{n_{..}} && \text{für alle Zeilen} && i = 1, \dots, r \\ \frac{n_{ij}}{n_{.j}} &\approx \frac{n_{i.}}{n_{..}} && \text{für alle Spalten} && j = 1, \dots, c \end{aligned} \quad (3.6)$$

Der Sachverhalt der Unabhängigkeit lässt sich somit sehr leicht in der entsprechenden Darstellung im Mosaikplot nachvollziehen, in welchem sich die Flächenanteile in diesem Fall nicht 'wesentlich' überdecken sollten. Die Quantifizierung und Bewertung von Abhängigkeiten in nominal skalierten Merkmalen wird in Kapitel [7.6] näher ausgeführt.

3.2 Beschreibung von Ordinaldaten

- Medianwert und andere Quartile
- Quantile
- Streuung ordinal skalierten Daten
- Punktdiagramm und Box-Plot
- Korrelationskoeffizient nach Kendall

Die Ausprägungen ordinal skalierten Merkmale unterliegen einer natürlichen **Anordnung**. Damit können die Beobachtungen **sortiert** werden, jede Beobachtung nimmt im Vergleich zu den anderen Beobachtungen einen **festen Platz (Rang)** ein.

Beispiel: Die Schmerzintensität vor / nach Gabe eines Analgetikums wird auf einer **Visuellen Analogskala (VAS)** erfasst. Dazu markiert der Patient den Befund auf einer Strecke zwischen 0 cm



(schmerzfrei) und 10 cm (unerträgliche Schmerzen). Die Schmerzen sind durch dieses Vorgehen zwar nicht (metrisch) messbar, allerdings kann eine ordinale Bewertung in den Kategorien 0 bis 10 vorgenommen werden. Bei $n=13$ Patienten zeigt sich folgendes Ergebnis.

Tabelle 3.5. Schmerzzintensität nach der visuellen Analogskala bei $n=13$ Patienten vor und nach Gabe eines Analgetikums

Patient	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII
vor Therapie	3	4	6	4	8	9	2	7	10	7	5	6	5
nach Therapie	4	4	1	5	3	3	1	3	4	5	6	9	1

Die Beobachtungen x_i , ($i : 1, \dots, n$) werden der Größe nach aufsteigend geordnet (sortiert) und dann mit in Klammern gesetzten Indizes versehen (um sie von den die ursprüngliche Reihenfolge wiedergebenden Indizes zu unterscheiden): $x_{(1)}$ ist somit die kleinste Beobachtung, $x_{(n)}$ die größte. Der Wert $x_{(i)}$ heißt **i-ter Anordnungswert**.

Sind alle Werte unterschiedlich, bezeichnet man den Index als **Rang**. Treten dem Wert nach gleiche Beobachtungen auf, so spricht man von **Bindungen** (ties). In diesem Fall ordnet man den Beobachtungen, die dem Wert nach gleich sind, den Durchschnitt der hierfür vorgesehenen Rangzahlen zu (Mittelrangmethode): z. B. für 5,3 5,3 5,3 anstatt der Ränge 1, 2, 3 jeweils den mittleren Rang $(1 + 2 + 3)/3 = 2$, d. h. 2, 2, 2.

Anordnungswerte sind wichtig zur Schätzung von Quantilen und zur Schätzung verteilungs-unabhängiger Toleranzintervalle; außerdem sind sie wichtig für multiple Vergleiche und Auswahlverfahren, für nichtparametrische Tests (Rangtestverfahren), für die Entdeckung von Ausreißern und für die Voraussage extremer Ereignisse wie z. B. Erdbeben.

Für das Sortieren und für die Vergabe der Rangzahlen stehen im Programm **R** die Funktionen **sort()** und **rank()** zur Verfügung.

```
> vor; sort(vor)
[1] 3 4 6 4 8 9 2 7 10 7 5 6 5 10 8
[1] 2 3 4 4 5 5 6 6 7 7 8 8 9 10 10
>
#####
> vor; rank(vor)
[1] 3 4 6 4 8 9 2 7 10 7 5 6 5 10 8
[1] 2.0 3.5 7.5 3.5 11.5 13.0 1.0 9.5 14.5 9.5 5.5 7.5 5.5 14.5 11.5
```

3.2.1 Medianwert und andere Quartile

Maßzahlen zur Kennzeichnung der Verteilung eines ordinal-skalierten Merkmals werden aus der Anordnung der Beobachtungen, der sogenannten **Rangliste**, abgeleitet.

$$x_{(1)} \leq x_{(2)} \leq x_{(3)} \dots \leq x_{(n-1)} \leq x_{(n)} \quad (3.7)$$

$$x_{(1)} = \text{Minimum}$$

$$x_{(n)} = \text{Maximum}$$

Neben der kleinsten (**Minimum**) und der größten (**Maximum**) Beobachtung ist die Mitte der Beobachtungen von zentraler Bedeutung in der Statistik. Der **Medianwert** gibt denjenigen beobachteten Wert an, der die Anzahl der Beobachtungen in zwei gleich große Hälften teilt, so dass jeder Teil 50% der Verteilung enthält. Das Symbol für den Medianwert ist $\tilde{x}$ (gelesen: x Schlange). Umfasst die Stichprobe eine ungerade Anzahl von Werten, so ist der Medianwert der „mittlere“, der nach der Größe geordneten Werte, ist n gerade, dann gibt es zwei mittlere Werte $\tilde{x}_1$ und $\tilde{x}_2$: der Medianwert (oder besser Pseudomedianwert) wird dann als $\tilde{x} = \frac{1}{2}(\tilde{x}_1 + \tilde{x}_2)$ ermittelt.

Es existieren 3 Werte, die eine Häufigkeitsverteilung in 4 gleiche Teile zerlegen. Der zentrale Wert ist wiederum der Medianwert $\tilde{x} = Q_2$, die anderen beiden bezeichnet man als unteres (erstes) und oberes (drittes) Quartil, d. h. das erste **Quartil** Q_1 ist die Maßzahl, die am Ende des ersten Viertels in der nach der Größe geordneten Reihe der Messwerte steht; Q_3 ist die Maßzahl am Ende des dritten Viertels der Reihe.

Die Berechnung der Quartile bei n aufsteigend geordneten Beobachtungen erfolgt nach der Regel:

$$\begin{array}{lll} Q_1 = x_{(k)} & \text{mit} & k = \lfloor (n+1) \cdot 0,25 \rfloor \\ Q_2 = \tilde{x} = x_{(l)} & \text{mit} & l = \lfloor (n+1) \cdot 0,50 \rfloor \\ Q_3 = x_{(m)} & \text{mit} & m = \lfloor (n+1) \cdot 0,75 \rfloor \end{array} \quad (3.8)$$

In der Formel 3.8 bezeichnet der Operator $\lfloor \cdot \rfloor$ die größte ganze Zahl, die nicht größer ist als das Ergebnis der Multiplikation (ganzzahliger Anteil). Die Quartile Q_1 und Q_3 werden im allgemeinen erst bei $n > 12$ (besser: $n > 20$) angegeben.

Die Bestimmung der Quartile kann in **R** mit der Funktion **floor()** (für den ganzzahligen Anteil) nach dem Sortieren direkt erfolgen. Andererseits stehen aber auch spezielle Funktionen **median()** und **quantile()** zur Verfügung. Abweichend von Formel 3.8 werden die Quartile in **R** allerdings durch eine Interpolation ermittelt.

```
> vor <- c(3, 4, 6, 4, 8, 9, 2, 7, 10, 7, 5, 6, 5)
> vsort <- sort(vor); n <- length(vsort) # Quartile
> Q1 <- vsort[floor((n+1)*0.25)]; Q1
[1] 4
> Q2 <- vsort[floor((n+1)*0.50)]; Q2
[1] 6
> Q3 <- vsort[floor((n+1)*0.75)]; Q3
[1] 7
>
> median(vor);
[1] 6
> quantile(vor, c(0.25, 0.50, 0.75))
25% 50% 75%
 4   6   7
```

Klassierte Beobachtungen: Liegt eine in Klassen eingeteilte Reihe von Einzelwerten vor, dann schätzt man den Medianwert durch lineare Interpolation nach

$$\tilde{x} = \tilde{U} + b \left(\frac{n/2 - (\sum f)_{\tilde{U}}}{f_{\text{Median}}} \right) \quad (3.9)$$

$\tilde{U}$ = untere Klassengrenze der Medianwertklasse; b = Klassenbreite; n = Anzahl der Werte; $(\sum f)_{\tilde{U}}$ = Summe der Häufigkeitswerte aller Klassen unterhalb der Medianwertklasse; f_{Median} = Anzahl der Werte in der Medianwertklasse.

3.2.2 Quantile

Ein Quantil x_α (auch **Fraktil** genannt) ist ein Lokalisationsmaß, das durch den Anteil α der Beobachtungen definiert ist, die unterhalb dieses Wertes liegen. Diese Verteilungsanteile werden bei metrischen Daten in der empirischen Verteilungsfunktion zusammengefasst. Der Median ist durch $\tilde{x}_{0,5}$ definiert. Sonderfälle der Quantile ergeben sich für $\alpha = 0,5, 0,25$ und $0,75$ (**Quartile**), für $\alpha = k/10$ ($k = 1, 2, \dots, 9$) (**Dezile**), $\alpha = k/100$ ($k = 1, 2, \dots, 99$) (**Perzentile**). Die Berechnung der Quantile erfolgt nach:

$$x_\alpha = \begin{cases} x_{(k)}: & k = \lceil n \cdot \alpha \rceil & \text{wenn } n \cdot \alpha \text{ nicht ganzzahlig ist} \\ \frac{1}{2}(x_{(k)} + x_{(k+1)}): & k = n \cdot \alpha & \text{sonst} \end{cases} \quad (3.10)$$

In der Formel 3.10 bezeichnet der Operator $\lceil \cdot \rceil$ die kleinste ganze Zahl, die nicht kleiner ist als das Ergebnis aus der Multiplikation (nächste ganze Zahl).

Bei gruppierten (klassierten) Beobachtungen werden die Quantile nach 3.9 berechnet, indem $n/2$ ersetzt wird durch $in/4$ ($i=1, 2, 3$; Quartile), $jn/10$ ($j = 1, 2, \dots, 9$; Dezentile), $kn/100$ ($k = 1, 2, \dots, 99$; Perzentile) sowie Medianwert und Medianwertklasse durch das gewünschte Quantil und seine Klasse.

3.2.3 Streuung ordinal skalierten Daten

Für die Kennzeichnung der Variabilität in den Beobachtungen zu ordinal-skalierten Merkmalen gibt es eine Anzahl spezieller Maßzahlen. Die **Spannweite (Range R)** wird aus der Differenz von Maximum und Minimum bestimmt und umfasst alle Beobachtungen. Der **Interquartilbereich (IQR oder I_{50})** wird aus der Differenz des 3. und 1. Quartils bestimmt und umfasst damit die zentralen 50% der Beobachtungen, d. h. 25% der Beobachtungen liegen unterhalb und 25% der Beobachtungen liegen oberhalb des Interquartilbereichs.

$$\begin{aligned} x_{(1)} &\leq x_{(2)} \leq x_{(3)} \dots \leq x_{(n-1)} \leq x_{(n)} \\ R &= x_{(n)} - x_{(1)} \\ I_{50} = IQR &= x_{0,75} - x_{0,25} = Q_3 - Q_1 \end{aligned} \quad (3.11)$$

Die **mittlere absolute Abweichung vom Medianwert** und die **Median-Deviation** (Median Absolute Deviation, **MAD** oder knapp **$\tilde{D}$**) sind weitere hervorragende Streuungsmaße für ordinale Beobachtungen, insbesondere letztere sollte stets angegeben werden. Die mittlere absolute Abweichung vom Medianwert wird berechnet nach:

$$MA_{\tilde{x}} = \frac{1}{n} \sum_{i=1}^n |x_i - \tilde{x}| = \frac{1}{n} \sum_{j=1}^k f_j |x_j - \tilde{x}| \quad (3.12)$$

Die rechte Seite in (3.12) betrifft die Berechnung bei klassierten Beobachtungen in k Klassen mit den Klassenmitten x_j und den absoluten Besetzungszahlen f_j .

Beziehung zwischen Medianwert und den absoluten Abweichungen: **Die Summe der absoluten Abweichungen nimmt für den Medianwert ein Minimum an.**



Die **Median-Deviation** wird berechnet nach:

$$\tilde{D} = \text{Medianwert}\{|x_i - \tilde{x}|\} = \text{Medianwert}\{f_j |x_j - \tilde{x}|\} \quad (3.13)$$

Die Berechnung der mittleren absoluten Abweichung kann in **R** elementar formuliert werden. Für die Berechnung der Median-Deviation **$\tilde{D}$** steht die Funktion **mad()** zur Verfügung. Allerdings muss entsprechend der Definition ein Skalierungsfaktor **const=1** gesetzt werden, da sonst standardmäßig **const=1,4826** angenommen wird, um eine konsistente, robuste Schätzung für die Standardabweichung σ nach dem Modell einer Normalverteilung ($N(\mu, \sigma)$) zu erhalten.

```
> MA <- mean(abs(vor - median(vor))); MA
[1] 1.846154
>
> D <- mad(vor, const=1); D
[1] 2
```

3.2.4 Punktdiagramm und Box-Plot

Die graphische Darstellung der beobachteten Werte eines mindestens ordinalskalierten Merkmals kann durch Dot-Plots oder Box-Plots erfolgen. In einem **Punktdiagramm (Dot-Plot)** werden die einzelnen Beobachtungen als Punkte zu y-Werten in einem Koordinatensystem aufgetragen. Die x-Achse dient der Unterteilung nach möglichen Untergruppen / Vergleichsgruppen oder nach wiederholten Untersuchungen / Messwiederholungen. Gleiche Werte in einer Gruppe sollten nebeneinander dargestellt werden (vgl. Abbildung 3.4), um Häufungen (Bindungen) in den Beobachtungen erkennen zu können.

In einem **Box-Plot** werden die im vorangehenden Abschnitt abgeleiteten Maßzahlen in einem Koordinatensystem dargestellt.

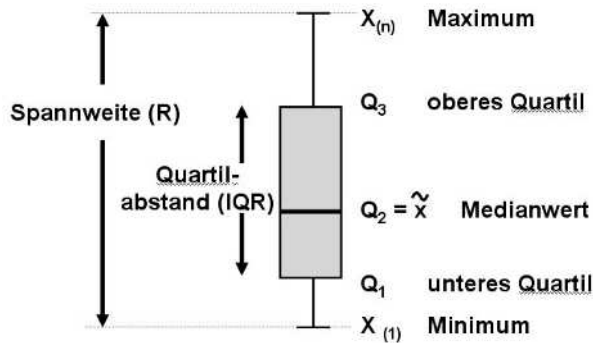


Abb. 3.3. Statistische Maßzahlen in der Box-Plot Darstellung

In dem Programm **R** stehen für die graphische Darstellung nach diesen Ansätzen die Funktionen **stripchart()** und **boxplot()** zur Verfügung. Gleiche Werte werden dabei im Punktdiagramm (stripchart) zufällig auf gleicher Höhe nebeneinander („jitter“) angeordnet. Für das Beispiel in diesem Abschnitt folgt (vgl. Abbildung 3.4):

```
> stripchart(list(vor, nach), method="jitter", jitter=0.1,
+   vertical=TRUE, group.names=c("vor", "nach"),
+   xlim=c(0.5, 2.5), ylim=c(0, 10), pch=16, cex=1.3)
>
> boxplot(vor, nach, range = 1.5, names=c("vor", "nach"),
+   ylim=c(0, 10), col=8)
```

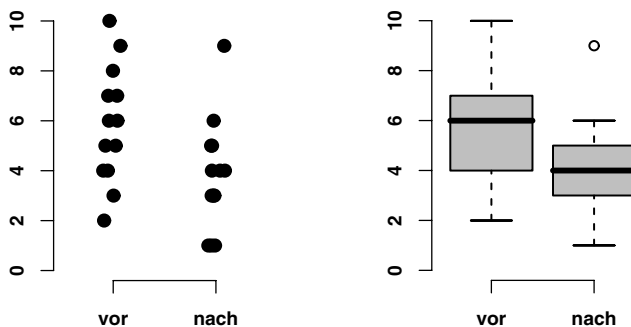


Abb. 3.4. Dot-Plot und Box-Plot Darstellung zu der Schmerzzintensität (VAS) vor und nach Gabe eines Analgetikums bei n=13 Patienten

Mit dem Parameter **range=1.5** kann in dem Box-Plot ein Bereich der **unauffälligen Streuung** gekennzeichnet werden. Dieser wird zum Beispiel entsprechend der Erfahrungen aus der Explorativen Datenanalyse (EDA nach Tukey) auf die **1,5-fache** (oder auch 3-fache) **Quartildistanz** begrenzt. Beobachtungen außerhalb dieser Grenzen gelten als Ausreißer (Extremwerte). Die Modifikation der Box-Plots in diesem Sinn wird auch häufig **Box-Whisker-Plot** genannt.

3.2.5 Korrelationskoeffizient nach Kendall

Mit der Rangkorrelation von Kendall soll der Zusammenhang zweier abhängiger Rangreihen beschrieben und bewertet werden. Der Korrelationskoeffizient r_τ (gelesen r-tau) basiert auf den ordinalen Informationen, d.h. er vergleicht die relative Lage der Werte zueinander über so genannte **Inversionen** und **Proversionen**.

R_x	1	2	3	4	5	6	7
R_y	2	3	1	4	6	5	7

Inversion: $R_{y_i} > R_{y_j}$ für $R_{x_i} < R_{x_j}$

Proversion: $R_{y_i} < R_{y_j}$ für $R_{x_i} < R_{x_j}$

Die Ordnung in den Beobachtungen wird durch x-Werte vorgegeben. Die entsprechenden y-Werte folgen dieser „natürlichen“ Ordnung (Proversion) oder entgegengesetzt in „unnatürlicher“ Ordnung (Inversion). Ein idealer Zusammenhang drückt sich somit durch 100% Proversionen ($r_\tau=1$) oder durch 100% Inversionen ($r_\tau=-1$) aus. Der Korrelationskoeffizient nach Kendall wird nach (3.14) berechnet.

$$r_\tau = 1 - \frac{4 \cdot (\text{Zahl der Inversionen})}{n(n-1)} \quad (3.14)$$

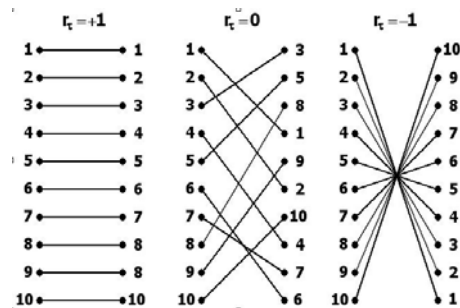


Abb. 3.5. Inversionen in gepaarten Rangfolgen zur Berechnung des Korrelationskoeffizienten nach Kendall

Beispiel: 10 verschiedene **Weinsorten** (a bis j) werden von zwei Juroren bewertet.

Wein	d	c	e	j	g	a	h	i	f	b
Richter A	1	2	3	4	5	6	7	8	9	10
Richter B	2	1	5	3	4	6	7	9	8	10

Anzahl der Inversionen = 1 + 0 + 2 + 0 + 0 + 0 + 0 + 1 + 0 + 0 = 4

$$r_\tau = 1 - \frac{4 \cdot 4}{10 \cdot 9} = 1 - 0,178 = 0,822$$



Die folgende elementare Berechnung von r_τ mit **R** zeigt einige Besonderheiten hinsichtlich der Programmierung. Eine spezielle Funktion steht mit **cor.test()** zur Verfügung.

```
> x <- c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10)
> y <- c(2, 1, 5, 3, 4, 6, 7, 9, 8, 10)
> n <- length(x); inv <- 0; prov <- 0
> for (i in 1:n) {
+   for (j in i:n) {
+     if (x[i]<x[j] & y[i]>y[j]) inv <- inv + 1
+     if (x[i]<x[j] & y[i]<y[j]) prov <- prov + 1
+   }
+ }
> r.tau <- 1 - 4*inv/(n*(n-1)); r.tau
[1] 0.8222222
```

3.3 Beschreibung von metrischen Daten

- Arithmetischer Mittelwert
- Standardabweichung, Varianz
- Variationskoeffizient
- Der $(\bar{x} \pm s)$ -Bereich
- Klassierte Messwerte
- Das gewogene arithmetische Mittel, die gewogene Varianz und das gewichtete arithmetische Mittel
- Geometrischer Mittelwert
- Harmonischer Mittelwert

Die Beschreibung metrischer Daten kann grundsätzlich auch nach den Maßzahlen und Verfahren erfolgen, die im vorangehenden Abschnitt für ordinal-skalierte Beobachtungen näher beschrieben sind. Ergänzend ergeben sich aber aus der Messbarkeit weitere Möglichkeiten, die mehr Informationen aus den Beobachtungen nutzen.

3.3.1 Arithmetischer Mittelwert

Das arithmetische Mittel $\bar{x}$ (gelesen: x quer) ist gleich der Summe aller Beobachtungen, geteilt durch die Anzahl dieser Beobachtungen. Es dient zur repräsentativ-nivellierenden Mittelung (Informationsverdichtung) für mehrere bis zahlreiche nicht zu heterogene Beobachtungen:

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \dots + x_n) = \frac{1}{n} \sum_{i=1}^n x_i \quad (3.15)$$



Beispiel: Der Body-Mass-Index ist eine Kennzahl, die aus der Körpergröße (m) und dem Gewicht (kg) berechnet wird: $BMI = \text{Gewicht}/(\text{Groesse})^2$. Für 13 Personen wurden die folgenden Werte ermittelt.

Hinweis: Die Bewertung des Body-Mass-Index erfolgt nach Kuczmarski und Mitarbeitern [KCFT97] in 4 Kategorien:

Kategorie	BMI-Bereich
Untergewicht	$BMI < 20$
Normalgewicht	$20 \leq BMI < 25$
Übergewicht	$25 \leq BMI < 30$
Fettleibigkeit	$BMI \geq 30$

Tabelle 3.6. Bestimmung des Body-Mass-Index bei 13 Personen

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}
28.2	23.9	20.3	26.7	25.6	32.5	23.5	19.7	27.8	26.7	20.7	28.4	33.3

Die Summe über alle Werte ist 337,3. Daraus folgt für den arithmetischen Mittelwert $\bar{x} = 337,3/13 = 25,95$. Die Berechnung in **R** erfolgt direkt nach der Definition (3.15) oder mittels der Funktion **mean()**.

```
> bmi <- c(28.2, 23.9, 20.3, 26.7, 25.6, 32.5, 23.5, 19.7, 27.8, 26.7, 20.7, 28.4, 33.3)
> n <- length(bmi)
> Summe <- sum(bmi); Summe
[1] 337.3
> Summe/n
[1] 25.94615
> mean(bmi)
[1] 25.94615
```

arithmetisches Mittel

Die Summe der Abweichungen der einzelnen Werte vom arithmetischen Mittelwert ist gleich Null, das heißt die Summe der Abweichungen der Werte, die kleiner sind als der Mittelwert (mit negativem Vorzeichen), ist im Betrag nach gleich der Summe der Abweichungen der Werte, die größer sind als der Mittelwert (mit positivem Vorzeichen):



$$\left| \sum_{x_i \leq \bar{x}} (x_i - \bar{x}) \right| = \left| \sum_{x_i \geq \bar{x}} (x_i - \bar{x}) \right|$$

Eine anschauliche Erklärung der Eigenschaften eines arithmetischen Mittelwertes liefern auch zwei Phänomene aus der Physik:

- Das **Prinzip der kommunizierenden Glasröhren**: nach dem Öffnen der Verbindungshähne gleichen sich die unterschiedlichen Höhen der Wassersäulen zum gemeinsamen Mittelwert aus.
- Denkt man sich Daten auf der reellen Zahlengeraden durch Punkte mit gleicher Masse ersetzt, so ist das arithmetische Mittel die Stelle auf der Geraden, die dem **Schwerpunkt aller Massenpunkte** entspricht.

Der Medianwert ist als Lagemaß zu bevorzugen beim Vorliegen:

- ordinalskalierte Beobachtungen (Rangdaten),
- nur weniger Meßwerte,
- asymmetrischer Verteilungen,
- von Verteilungen mit offenen Endklassen und
- bei Verdacht auf Ausreißer.

Beachte: Für $\bar{x}$ gilt: $\sum_i (x_i - \bar{x}) = 0$ und $\sum_i (x_i - \bar{x})^2 \leq \sum_i (x_i - x)^2$ für jedes x ; für den Medianwert $\tilde{x}$ gilt dagegen $\sum_i |x_i - \tilde{x}| \leq \sum_i |x_i - x|$ für jedes x ; d.h. $\sum_i (x_i - \bar{x})^2$ und $\sum_i |x_i - \tilde{x}|$ sind jeweils minimal! **Der Mittelwert minimiert die Summe der Abweichungsquadrate** (demgegenüber minimiert der Medianwert die absolut genommene Summe der Abweichungen).



Diese Eigenschaft des Mittelwertes hat zur Folge, dass Extremwerte (Ausreißer) einen großen Einfluss auf die Bewertung der zentralen Lage der Messwerte ausüben - man sagt, **der Mittelwert ist ein empfindliches Lagemaß**.

Robuste Mittelwerte sind neben dem Medianwert das **α -gestutzte Mittel** (Trimmen) und das **α -Winsorisierte Mittel** (nach C.P. Winsor), wobei $\alpha = 0,05$, $\alpha = 0,1$ oder $\alpha = 0,2$ gewählt wird. Bei stärkerem Verdacht auf Ausreißer wird der erste Ansatz bevorzugt.

1. **Stutzen:** Das 10%-gestutzte arithmetische Mittel ($\alpha = 0,1$) erhält man, indem man zunächst die Daten ansteigend anordnet, anschließend auf jeder „Seite“ 10% der Daten verwirft und dann aus den restlichen Daten das arithmetische Mittel bildet (in den folgenden Formeln Formeln bezeichnet der Operator $\lfloor \cdot \rfloor$ den „ganzzahligen Anteil“ aus dem Produkt $\alpha \cdot n$):

$$\begin{aligned} x_{(1)} &\leq x_{(2)} \leq x_{(3)} \leq \dots \leq x_{(n)} \\ \bar{x}_t &= \frac{x_{(g+1)} + \dots + x_{(n-g)}}{n-2g} = \frac{1}{n-2g} \sum_{i=g+1}^{n-g} x_{(i)} \\ \text{mit } g &= \lfloor \alpha \cdot n \rfloor \quad \text{für } 0 \leq \alpha \leq 0,5 \end{aligned} \quad (3.16)$$

2. **Winsorisieren:** Das 10%-Winsorisierte arithmetische Mittel ergibt sich, indem man zunächst die Daten ansteigend sortiert, anschließend auf jeder „Seite“ 10% der Daten (jeden dieser als extrem beurteilten Werte) durch den nächstgelegenen Wert der restlichen Daten ersetzt und dann aus sämtlichen Daten das arithmetische Mittel bildet.

$$\begin{aligned} x_{(1)} &\leq x_{(2)} \leq x_{(3)} \leq \dots \leq x_{(n)} \\ \bar{x}_w &= \frac{1}{n} (gx_{(g+1)} + \sum_{i=g+1}^{n-g} x_{(i)} + gx_{(n-g)}) \\ \text{mit } g &= \lfloor \alpha \cdot n \rfloor \quad \text{für } 0 \leq \alpha \leq 0,5 \end{aligned} \quad (3.17)$$

In **R** kann das gestutzte arithmetische Mittel durch einen optionalen Parameter in der Funktion **mean()** berechnet werden. Dazu folgt ein Beispiel mit modifizierten BMI-Werten:

```
> bmi <- c(22.2, 23.9, 20.3, 26.7, 25.6, 22.5, 23.5, 24.7, 27.8, 26.7, 20.7, 26.4, 40.3)
> sort(bmi)
[1] 20.3 20.7 22.2 22.5 23.5 23.9 24.7 25.6 26.4 26.7 26.7 27.8 40.3
> mean(bmi)
[1] 25.48462
> mean(bmi, trim=0.1)
[1] 24.60909
```

3.3.2 Standardabweichung, Varianz

Die Streuung metrischer Daten wird durch eine gemittelte Abweichung vom Mittelwert gekennzeichnet. Dabei ist die Summe der Abweichungen als Maß ungeeignet (da sie stets gleich Null ist). Stattdessen werden die quadrierten Abweichungen betrachtet.

Die Standardabweichung (s) ist *praktisch* gleich der positiven Quadratwurzel aus dem Mittelwert der quadrierten Abweichungen vom Mittelwert; in der Statistik ist der Mittelwert $\bar{x}$ stets zusammen mit n und s anzugeben!

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \quad (3.18)$$

Der Ausdruck „*praktisch*“ bezieht sich hierbei auf die Tatsache, dass in der Wurzel der Nenner nicht n , wie es einem Mittelwert entspräche, steht, sondern die um 1 verminderte Zahl der Werte. Eine nähere Begründung hierfür wird im Abschnitt Schätzen (Erwartungstreue Schätzung) gegeben. Gelegentlich wird auch zwischen der **Standardabweichung aus einer Stichprobe** und

der **Standardabweichung in der Gesamtheit** unterschieden, bei der dann durch die Anzahl n dividiert wird. Das Quadrat der Standardabweichung wird als **Varianz** (s^2) bezeichnet. Die Berechnung der Standardabweichung in **R** erfolgt direkt nach der Formel oder mit der Funktion **sd()**.

```
> bmi <- c(28.2, 23.9, 20.3, 26.7, 25.6, 32.5, 23.5, 19.7, 27.8, 26.7, 20.7, 28.4, 33.3)
> m <- mean(bmi)
> saq <- (bmi - m)^2 # Abweichungsquadrate
> sqrt(sum(saq)/(n-1)) # Standardabweichung
[1] 4.295466
> sd(bmi)
[1] 4.295466
```

Praktisch erfolgt die Berechnung der Standardabweichung meist nach:

$$s = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}} \quad s = \sqrt{\frac{n \sum x^2 - (\sum x)^2}{n(n-1)}}$$

Beispiel: Berechne $\bar{x}$ und s aus den Werten: 27, 22, 24 und 26 ($n = 4$).

$$\bar{x} = \frac{\sum x}{n} = \frac{99}{4} = 24,75$$

$$s = \sqrt{\frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n-1}} = \sqrt{\frac{2465 - \frac{99^2}{4}}{4-1}} = \sqrt{4,917} = 2,22 \quad \text{bzw.}$$

$$s = \sqrt{\frac{n \sum x^2 - (\sum x)^2}{n(n-1)}} = \sqrt{\frac{4 \cdot 2465 - 99^2}{4(4-1)}} = \sqrt{4,917} = 2,22$$

Die Varianz s^2 besitzt als **Dimension** das Quadrat der Dimension der einzelnen Beobachtungen x_i (z.B. cm^2 bei Längenmessungen in cm); deshalb wird s bevorzugt. Die Varianz ist **rund doppelt so genau** anzugeben wie die Standardabweichung, etwa: $s^2 = 1,44$ und $s = 1,2$.

Die Streuungsmaße s^2 und s bleiben **unverändert**, wenn die Beobachtungen x_i um einen konstanten Wert c verkleinert oder vergrößert werden. Somit ist die Varianz **lageunabhängig** (lokalisationsinvariant oder translationsinvariant).

Beispiel: Gegeben sei eine Verteilung der Körpergröße von n Personen. Die Form, d. h. die Varianz, bliebe unverändert, würde man die Personen entweder in einen Graben stellen, aus dem der Kleinste noch herausragt oder auf eine Mauer stellen, wobei Grabentiefe bzw. Mauerhöhe konstant seien. Dementsprechend kann man bei der Berechnung einer Varianz alle Werte x um einen bestimmten Wert a vermindern bzw. vermehren. So würde man zur Berechnung der Varianz der Körpergröße Erwachsener z. B. von den um 100 cm verminderten Messwerten ausgehen.

Die Varianz lässt sich auch nach (3.19) berechnen:

$$s^2 = \frac{1}{2n(n-1)} \sum_i \sum_j (x_i - x_j)^2 = \frac{\sum_{i < j} (x_i - x_j)^2}{n-1} \quad (3.19)$$

$$\text{z. B. } x_i : 1, 2, 3; n = 3; \bar{x} = 2; s^2 = \frac{1^2 + 2^2 + 1^2}{3(3-1)} = 1$$



In dieser Darstellung wird die Eigenschaft der **Varianz als Maßzahl für die Variabilität** in den Beobachtungen besonders deutlich.

Ein anderes Streuungsmaß ist $s_{\text{Gini}} = \{2/[n(n-1)]\} \sum_i \sum_j |x_i - x_j|$, in dem die Abweichungsquadrate durch die absoluten Abweichungen ersetzt sind.

Soll ein **zusätzlicher Wert** x_z bei der Berechnung von Mittelwert und Varianz berücksichtigt werden, dann gilt für die jetzt $n + 1$ Beobachtungen: $\bar{x}_{n+1} = (x_z + n\bar{x})/(n + 1) = \bar{x} + (x_z - \bar{x})/(n + 1)$ und $s_{n+1}^2 = (n + 1)(\bar{x}_{n+1} - \bar{x})^2 + (n - 1)s^2/n$.

3.3.2.1 Arithmetische Mittelwerte und Standardabweichungen vergleichbar gemacht

Zum vorläufigen Vergleich insbesondere von Messreihen bei zumindest angenähert symmetrisch-eingipfliger Verteilung (vgl. Häufigkeitsverteilung) verwendet man die Umrechnung der Rohwerte x_i in **Standardwerte** y_i innerhalb jeder Messreihe nach:

$$y_i = 100 \frac{x_i - \bar{x}}{s} + 500$$

Durch diese Umrechnung (lineare Transformation) ergibt sich insgesamt ein Mittelwert von 500 und eine Standardabweichung von 100; individuelle Mittelwerte liegen dann darunter oder darüber. Mitunter wird man anstatt des Faktors 100 und anstatt der die Nullpunktverschiebung bedingenden Größe 500 auch andere Werte bevorzugen.

Beispiel:

Stichprobe j	Beobachtungen x_{ij}	Anzahl n_j	Mittelwert $\bar{x}_j$	Standardabweichung s_j
1	40 50 72	3	54,0	16,37
2	30 60 80 90 100	5	72,0	27,75
3	40 50 60 70	4	55,0	12,91
1+2+3	30 + 40 + ... + 90 + 100	12	61,833	21,328

Die Umrechnung nach obiger Transformation führt auf:

Stichprobe j	Beobachtungen y_{ij}	$\bar{y}_j$	s_j
1	397,63 444,52 547,67	463,27	76,76
2	350,74 491,40 585,18 632,07 678,95	547,67	130,11
3	397,63 444,52 491,40 538,29	467,96	60,53
1+2+3	350,74+397,63+397,63...+678,95	500,00	100,00

Diese spezielle Form einer standardisierenden Transformation ist von zentraler Bedeutung bei der Behandlung von normalverteilten Zufallsvariablen!

3.3.3 Variationskoeffizient

Zum Vergleich der Variabilität von Verteilungen, der relativen Standardabweichung von Merkmalen mit unterschiedlichen Mittelwerten, dient der von K. Pearson (1875-1936) eingeführte Variationskoeffizient V . Dieser kennzeichnet die Standardabweichung in Einheiten des arithmetischen Mittels.

$$V = \frac{s}{\bar{x}} \quad \text{alle } x > 0 \quad (3.20)$$

V setzt eine Verhältnisskala voraus und bleibt als **relatives Streuungsmaß** konstant, wenn jedes x_i mit einer Konstanten multipliziert wird (**Skaleninvarianz**). V dient auch als **Konzentrationsmaß**: ein im Verhältnis zu $\bar{x}$ großes s entspricht einer schwachen Konzentration.

Der Variationskoeffizient ist ein relatives dimensionsloses Streuungsmaß mit dem Mittelwert als Einheit. Da sein Maximum $\sqrt{n}$ beträgt, gibt man auch gern den in Prozent ausgedrückten **relativen Variationskoeffizienten** V_r an, der Werte zwischen 0% und 100% annehmen kann:

$$V_r[\%] = \frac{s/\bar{x}}{\sqrt{n}} 100 \quad \text{alle } x > 0 \quad \text{und} \quad \frac{s}{\bar{x}} \leq \sqrt{n} \quad (3.21)$$

Beispiel: Für $n = 50$, $s = 4$ und $\bar{x} = 20$ erhält man

$$V = \frac{4}{20} = 0,20 \quad \text{und} \quad V_r = \frac{4/20}{\sqrt{50}} 100 = 2,8\% \quad \text{oder} \quad V_r = 0,03.$$



3.3.4 Der $(\bar{x} \pm s)$ -Bereich

Die graphische Darstellung von Mittelwerten und Standardabweichungen aus Messwerten erfolgt häufig in einem sogenannten „**Fehlerbalkendiagramm**“ (engl. error bar chart, vgl. Abbildung 3.6). Der Name ist missverständlich, da die Variabilität von Messwerten dargestellt werden soll und „Fehler“ nur eine mögliche Ursache für die Streuung in den Beobachtungen darstellen.

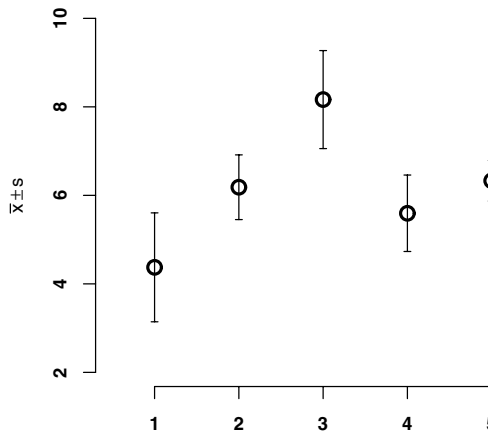


Abb. 3.6. „Fehlerbalkendiagramm“ für den Vergleich der Mittelwerte in 5 Gruppen aus je 10 Messungen mit unterschiedlicher Streuung

Im „Fehlerbalkendiagramm“ wird die Lage des Mittelwertes durch einen Punkt (oder einen Balken) markiert und die Streuung der Werte durch vertikale Linien entsprechend der berechneten Standardabweichung nach oben bzw. unten (in y-Richtung) eingezeichnet. Damit können insbesondere Mittelwerte aus verschiedenen Stichproben oder zu unterschiedlichen Zeitpunkten sehr übersichtlich vergleichend gegenübergestellt werden. Informativer, insbesondere bei kleinen Fallzahlen, ist jedoch die Darstellung im Box-Plot bzw. Punktdiagramm.

3.3.5 Klassierte Messwerte; Berechnung des Mittelwertes und der Standardabweichung

Berechnung von Mittelwert ($\bar{x}$) und Varianz (s^2) aus n Beobachtungen (unklassiert bzw. klassiert: n_j sind die Besetzungszahlen pro Klasse mit $n = \sum_{j=1}^n n_j$; x_j sind die entsprechenden Klassenmitten):

$$\text{unklassiert: } \bar{x} = \frac{\sum x}{n}; \quad s^2 = \frac{n \sum x^2 - (\sum x)^2}{n(n-1)} \quad (3.22)$$

$$\text{klassiert: } \bar{x} = \frac{\sum n_j x_j}{n}; \quad s^2 = \frac{n \sum n_j x_j^2 - (\sum n_j x_j)^2}{n(n-1)} \quad (3.23)$$

Werden Daten klassiert, so wird der „Rechenaufwand“ vereinfacht, da man alle Werte einer Klasse in der Mitte der Klasse vereinigt und die Streuung innerhalb der Klassen vernachlässigt. **Mittelwert und Varianz sowie andere Maße sollten aus den unklassierten Daten berechnet werden**, da sich sonst fehlerhafte Werte ergeben, die auch durch entsprechende Korrekturmaßnahmen (Sheppard-Korrektur) nicht vollständig auszugleichen sind.

3.3.6 Das gewogene arithmetische Mittel, die gewogene Varianz und das gewichtete arithmetische Mittel

Sollen mehrere Messreihen oder Stichproben mit den Umfängen $n_1, n_2, \dots, n_k$, den Mittelwerten $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k$ und den Varianzen $s_1^2, s_2^2, \dots, s_k^2$ zu einer gemeinsamen Messreihe vereinigt werden, die den Umfang $n = n_1 + n_2 + \dots + n_k$ hat, dann ist das **gewogene arithmetische Mittel** $\bar{x}_{\text{gew}}$

$$\bar{x}_{\text{gew}} = \frac{n_1 \cdot \bar{x}_1 + n_2 \cdot \bar{x}_2 + \dots + n_k \cdot \bar{x}_k}{n} \quad (3.24)$$

auch **Gesamtmittel** $\bar{x}$ genannt und die Standardabweichung s_{in} innerhalb der Messreihen

$$s_{\text{in}} = \sqrt{\frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1) + \dots + s_k^2(n_k - 1)}{n - k}} \quad (3.25)$$



Beispiel: $n_1 = 8, \bar{x}_1 = 9, (s_1 = 2) \quad s_1^2 = 4$

$$n_2 = 10, \bar{x}_2 = 7, (s_2 = 1) \quad s_2^2 = 1 \quad \bar{x} = \frac{8 \cdot 9 + 10 \cdot 7 + 6 \cdot 8}{24} = 7,917$$

$$n_3 = 6, \bar{x}_3 = 8, (s_3 = 2) \quad s_3^2 = 4$$

$$s_{\text{in}} = \sqrt{\frac{4(8-1) + 1(10-1) + 4(6-1)}{24-3}} = 1,648$$

Die **gewogene Varianz** der x -Werte in der Gesamtmessreihe wird nach

$$s_{\text{gew}}^2 = \frac{1}{n-1} \left[\sum_i (n_i - 1) s_i^2 + \sum_i n_i (\bar{x}_i - \bar{x})^2 \right] \quad (3.26)$$

berechnet, d. h.

$$s_{\text{gew}}^2 = (1/23)[(7 \cdot 4 + 9 \cdot 1 + 5 \cdot 4) + (8 \cdot 1,083^2 + 10 \cdot 0,917^2 + 6 \cdot 0,083^2)] = 3,254.$$

Der **gewichtete arithmetische Mittelwert**: Einzelmessungen ungleicher Genauigkeit lassen sich durch **unterschiedliche Gewichte** w_i (1, 2, 3 bzw. mit $0 \leq w_i \leq 1$ und $\sum w_i = 1$) kennzeichnen. Das gewichtete arithmetische Mittel erhält man nach $\bar{x} = (\sum w_i x_i) / \sum w_i$ bzw. zweckmäßiger durch Wahl eines günstigen Hilfswertes a , man geht dann von den Abweichungen $z_i = x_i - a$ aus:

Beispiel: Drei unterschiedlich gewichtete Messwerte



x_i	w_i	$x_i - a = z_i$ ($a = 137,8$)	$w_i z_i$
138,2	1	0,4	0,4
137,9	2	0,1	0,2
137,8	1	0,0	0,0

$$\sum w_i = 4 \quad \sum w_i z_i = 0,6$$

$$\bar{x} = a + \frac{\sum w_i z_i}{\sum w_i}$$

$$\bar{x} = 137,8 + \frac{0,6}{4} = 137,95$$

Sind Objekte hinsichtlich unterschiedlich zu bewertender Merkmale zu vergleichen, so benutzt man gern **bewertende „Gewichte“** w_i . Etwa für den Vergleich von 3 Schülern hinsichtlich ihrer Leistungen, ihrer Benotungen ($b_i : 1, \dots, 5$) in 5 Fächern, denen eine unterschiedliche Bedeutung w_i so zugeordnet wird, dass die Summe der Gewichte w_i der 5 Fächer gleich 1 ist (vgl. Tabelle 3.7).

Beispiel: Rangfolge dreier Schüler



Tabelle 3.7. Leistungsvergleich von drei Schülern aufgrund der Produktsummen ihrer gewichteten Benotungen

Nr.	Fach	w_i	Schüler A		Schüler B		Schüler C	
			b_i	$b_i w_i$	b_i	$b_i w_i$	b_i	$b_i w_i$
1	Deutsch	0,3	2	0,6	4	1,2	3	0,9
2	Mathem.	0,3	3	0,9	2	0,6	1	0,3
3	Physik	0,2	3	0,6	2	0,4	2	0,4
4	Chemie	0,1	2	0,2	2	0,2	2	0,2
5	Englisch	0,1	1	0,1	3	0,3	4	0,4
Summe		1	–	2,4	–	2,7	–	2,2

Aus den Produktsummen $\sum b_i w_i$ folgt, dass C der Beste ist, gefolgt von A und B. Entsprechend geht man vor, wenn n Objekte aufgrund ihrer Merkmalsausprägungen – es liegen p unterschiedlich zu bewertende Merkmale vor – in eine Rangfolge zu bringen sind.

3.3.7 Geometrischer Mittelwert

Sind die Merkmalsausprägungen **relative Änderungen** (z. B. Wachstum, Zuwachsraten, Produktionssteigerungen), so wird das geometrische Mittel $\bar{x}_G$ berechnet; ebenfalls, wenn **mittlere Arbeits- und Wartezeiten** erwünscht sind.

Liegen die positiven Werte $x_1, x_2, \dots, x_n$ vor, dann heißt die n -te Wurzel aus dem Produkt aller dieser Werte das geometrische Mittel $\bar{x}_G$

$$\bar{x}_G = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdot \dots \cdot x_n} \quad \text{mit } x_i > 0 \quad (3.27)$$

Die Berechnung erfolgt auf logarithmischem Wege nach

$$\lg \bar{x}_G = \frac{1}{n} (\lg x_1 + \lg x_2 + \lg x_3 + \dots + \lg x_n) = \frac{1}{n} \sum_{i=1}^n \lg x_i \quad (3.28)$$

Der Logarithmus des geometrischen Mittels ist also gleich dem arithmetischen Mittel der Logarithmen. Sollen mehrere, sagen wir k geometrische Mittel, die aus den Reihen mit den Umfängen $n_1, n_2, \dots, n_k$ ermittelt wurden, zu einem Gesamtmittel vereinigt werden, so wird ein **gewogenes geometrisches Mittel** gebildet

$$\lg \bar{x}_G = \frac{n_1 \cdot \lg \bar{x}_{G1} + n_2 \cdot \lg \bar{x}_{G2} + \dots + n_k \cdot \lg \bar{x}_{Gk}}{n_1 + n_2 + \dots + n_k} \quad (3.29)$$

Das geometrische Mittel ist vor allem dann anzuwenden, wenn ein Durchschnitt von **Verhältniszahlen** berechnet werden soll, wobei die Veränderungen in jeweils gleichen zeitlichen Abständen angegeben sind (vgl. Beispiel 1). Es wird verwendet, wenn sich eine Variable in der Zeit in einem einigermaßen konstanten Verhältnis ändert. Das ist der Fall bei **Wachstumserscheinungen** mannigfaltiger Art. Die durchschnittliche Zunahme der Bevölkerung in der Zeit, der Patientenzahl oder Unterhaltskosten einer Klinik sind bekannte Beispiele. Ob eine sich im konstanten Verhältnis ändernde Geschwindigkeit vorliegt, lässt sich überschlagsmäßig beurteilen, indem man die Daten auf einfachem Logarithmenpapier (Ordinate: logarithmisch geteilt, für das Merkmal; Abszisse linear geteilt, für die Zeit) notiert. Es müsste sich wenigstens angenähert eine Gerade ergeben. $\bar{x}_G$ ist dann der **Mittelwert der Zuwachsraten** (vgl. Beispiele 2 und 3)



Beispiel 1: Ein Angestellter erhält in drei aufeinanderfolgenden Jahren **Gehaltserhöhungen** von 2,5%, 10% und 22%. Der Prozentsatz ist jeweils auf das Gehalt des Vorjahres bezogen. Gefragt ist nach der durchschnittlichen Gehaltserhöhung.

Das geometrische Mittel von 1,025, 1,10 und 1,22 ist zu ermitteln:

$$\begin{array}{l} \lg 1,025 = 0,010724 \\ \lg 1,10 = 0,041393 \\ \lg 1,22 = 0,086360 \\ \hline \sum \lg x_i = 0,138477 \end{array} \quad \begin{array}{l} \frac{1}{3} \cdot \sum \lg x_i = 0,046159 = \lg \bar{x}_G \\ \bar{x}_G = 1,112 \end{array}$$

Im Durchschnitt ist somit das Gehalt um 11,2% gestiegen [vgl. mit dem arithmetischen Mittelwert $\bar{x} = (2,5 + 10 + 22)/3 = 11,5$].

```
> gehalt <- c(1.025, 1.10, 1.22) # Gehaltserhoenungen
> lg.gehalt <- log10(gehalt)
> 10^mean(lg.gehalt) # mittlere Gehaltserhoenung
[1] 1.112138
```



Beispiel 2: In einer bestimmten Kultur erhöhte sich in drei Tagen die Zahl der **Bakterien** pro Einheit von 100 auf 500. Gefragt ist nach der durchschnittlichen täglichen Zunahme, ausgedrückt in Prozenten.

Diese Größe bezeichnen wir mit x , dann beträgt die Zahl der Bakterien nach dem

1. Tag: $100 + 100x = 100(1 + x)$
2. Tag: $100(1 + x) + 100(1 + x)x = 100(1 + x)^2$
3. Tag: $100(1 + x)^2 + 100(1 + x)^2x = 100(1 + x)^3$

Dieser letzte Ausdruck muss gleich 500 sein, d. h.

$$100(1 + x)^3 = 500, \quad (1 + x^3) = 5, \quad 1 + x = \sqrt[3]{5}$$

Mit Hilfe von Logarithmen finden wir $\sqrt[3]{5} = 1,710$, d. h. $x = 0,710 = 71,0\%$.

Allgemein: Beginnen wir mit einer Menge M , die sich mit konstanter Zuwachsrate r in der Zeiteinheit vermehrt, dann erhalten wir nach n Zeiteinheiten den Betrag

$$B = M(1 + r)^n \quad (3.30)$$

Beispiel 3: Eine Summe sei in $n = 4$ Jahren von 4 Millionen **Euro** (M) auf 5 Millionen Euro (B) angewachsen. Gefragt ist nach der durchschnittlichen jährlichen Zuwachsrate.

Wenn ein Anfangskapital von M (Euro) nach n Jahren auf B (Euro) angewachsen ist, dann ist das geometrische Mittel r der Zuwachsraten für die n Jahre gegeben durch

$$r = \left[\sqrt[n]{B/M} \right] - 1 \quad (3.31)$$

Wir erhalten $r = \sqrt[4]{5 \cdot 10^6 / 4 \cdot 10^6} - 1 = \sqrt[4]{5/4} - 1$ und setzen $\sqrt[4]{5/4} = x$, dann ist $\lg x = (1/4) \lg(5/4) = (1/4)(\lg 5 - \lg 4) = 0,0217$; damit ergibt sich $x = 1,052$ und $r = 1,052 - 1 = 0,052$. Die durchschnittliche Zuwachsrate beträgt 5,2% jährlich.

Beispiel 4: Eine Aktie, für 100 Euro gekauft, steigt im ersten Jahr auf 150 Euro, fällt aber im zweiten Jahr auf 90 Euro. Die mittlere Wachstumsrate ist dann $-5,13\%$.

$$10 \xrightarrow{+} 150 \xrightarrow{-} 90, \text{ d. h. } (150/100) = 1,5 \quad \text{und} \quad (90/150) = -0,6$$

$$\sqrt{1,5 \cdot 0,6} = 0,9487; \quad 0,9487 - 1 = -0,0513 \quad \text{oder} \quad -5,13\%$$

[vgl. $100 - 5,13 = 94,87$; $5,13\%$ von $94,87$ ist gleich $4,87$ und $94,87 - 4,87 = 90$].

Exponentielles Wachstum: Bedeutet d die Verdoppelungszeit, r die relative Zuwachsrate pro Jahr und beschreibt die Funktion $y = ke^{rt}$ mit der Konstanten k (und wegen $\ln e = 1$ schreibbar als $\ln y = \ln k + rt$) das Wachstum, dann ist $d = (\ln 2)/r = 0,693/r$. So ergibt sich für eine jährliche Zuwachsrate von 0,07 oder 7% eine Verdoppelungszeit von $0,693/0,07 \approx 10$ Jahren. Die **kritische Zeit** t_{cr} in Jahren, die eine Größe Q benötigt, um von ihrem gegenwärtigen Wert Q_0 zu einem kritischen Wert oder Grenzwert Q_{cr} anzuwachsen, ist unter der Annahme, dass ein exponentielles Wachstum mit konstanter Zuwachsrate r in % pro Jahr vorliegt, $t_{cr} = (230/r) \lg(Q_{cr}/Q_0)$; z. B. $Q_{cr}/Q_0 = 25$; $r = 7\%$; $t_{cr} = (230/7) \lg 25 = 32,8571 \cdot 1,3979 = 45,9$ oder 46 Jahre.

Für **relative Änderungen** benutze man die Maßzahl $100 \ln(y/x)$.



3.3.8 Harmonischer Mittelwert

Als Mittel von Leistungslimits bei freiem Zeitverbrauch und als mittlere Überlebenszeit wird das harmonische Mittel $\bar{x}_H$ berechnet, ebenfalls z. B. als Mittel der durchschnittlichen Bevölkerungsdichte mehrerer Länder.

Liegen die positiven (oder negativen) Werte $x_1, x_2, \dots, x_n$ vor, dann heißt **der reziproke Wert des arithmetischen Mittels** aller reziproken Werte das harmonische Mittel $\bar{x}_H$

$$\bar{x}_H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}} = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}} \quad \text{mit } x_i \neq 0 \quad (3.32)$$

Bei praktischen Anwendungen ist es vielfach notwendig, den Einzelwerten x_i Gewichte w_i zuzuordnen und daraus ein **gewichtetes harmonisches Mittel** (vgl. Beispiele 3 und 4) zu berechnen:

$$\bar{x}_H = \frac{w_1 + w_2 + \dots + w_n}{\frac{w_1}{x_1} + \frac{w_2}{x_2} + \dots + \frac{w_n}{x_n}} = \frac{\sum_{i=1}^n w_i}{\sum_{i=1}^n \left(\frac{w_i}{x_i} \right)} \quad (3.33)$$

Das **gewogene harmonische Mittel** ist

$$\bar{x}_H = \frac{n_1 + n_2 + \dots + n_k}{\frac{n_1}{\bar{x}_{H_1}} + \frac{n_2}{\bar{x}_{H_2}} + \dots + \frac{n_k}{\bar{x}_{H_k}}} = \frac{\sum_{i=1}^k n_i}{\sum_{i=1}^k \frac{n_i}{\bar{x}_{H_i}}} \quad (3.34)$$

Das harmonische Mittel wird dann benötigt, wenn Beobachtungen das, was wir mit dem arithmetischen Mittel ausdrücken wollen, **im umgekehrten Verhältnis** angeben, wenn die Beobachtungen gewissermaßen eine Reziprozität enthalten, etwa Angaben wie Stunden pro Kilometer (anstatt km/Std). Es wird weiter gebraucht, wenn aus verschiedenen Geschwindigkeiten für Teilstrecken die mittlere Geschwindigkeit berechnet werden soll (Beispiel 2) oder wenn aus verschiedenen Dichten von Gasen, Flüssigkeiten, Teilchen usw. in einzelnen Teilräumen die mittlere Dichte zu ermitteln ist. Als mittlere Überlebenszeit wird es auch benutzt.



Beispiel 1: In drei verschiedenen Läden wird ein bestimmter Gegenstand zu den folgenden Preisen verkauft: 10 Stück für EURO 1,-, 5 Stück für EURO 1,- und 8 Stück für EURO 1,-. Gefragt ist nach der Durchschnittszahl – wie viel **Stück pro EURO**.

$$\bar{x}_H = \frac{3}{\frac{1}{10} + \frac{1}{5} + \frac{1}{8}} = \frac{3}{\frac{17}{40}} = \frac{120}{17} = 7,06 \simeq 7,1$$

Kontrolle: 1 Stck. = EURO 0,100
 1 Stck. = EURO 0,200
 1 Stck. = EURO 0,125
 3 Stck. = EURO 0,425

das heißt

$$1 \text{ Stck.} = \text{EURO } \frac{0,425}{3} = 0,1417$$

und $\frac{1,0000}{0,1417} = 7,06$
 stimmt mit 7,1 Stück pro EURO überein.

```

> stueck <- c(10, 5, 8) # Kosten / Stueckzahl
> rez.stueck <- 1/stueck; n <- length(stueck)
> n / sum(rez.stueck) # mittlere Stueckzahl
[1] 7.058824

```

Beispiel 2: Das klassische Beispiel für das harmonische Mittel ist eine Bestimmung des Geschwindigkeitsdurchschnitts. Es fährt jemand von A nach B mit einer Durchschnittsgeschwindigkeit von 30 km/Std. Für den Rückweg von B nach A benutzt er dieselbe Straße mit einer Durchschnittsgeschwindigkeit von 60 km/Std. Gefragt ist nach der Durchschnittsgeschwindigkeit für die Gesamtfahrt (D_G)

$$D_G = 2 \left/ \left(\frac{1}{30} + \frac{1}{60} \right) \right. = 40 \text{ km/Std.}$$

Hinweis: Angenommen, die Entfernung AB betrage 60 km, dann braucht man für die Reise von A nach B $\frac{60 \text{ km}}{30 \text{ km/Std.}} = 2$ Stunden, für die Reise von B nach A $\frac{60 \text{ km}}{60 \text{ km/Std.}} = 1$ Stunde, d. h. $D_G = \frac{\text{Gesamtstrecke}}{\text{Gesamtzeit}} = \frac{120 \text{ km}}{3 \text{ Std.}} = 40 \text{ km/Std.}$

Die Geschwindigkeit ist ein sog. **intensives** Merkmal. Demgegenüber ist bei einem sog. **extensiven** Merkmal (z. B. Entfernung) die Bildung eines Durchschnittes nicht erlaubt, wohl aber die Summenbildung.

Beispiel 3: Bei einem bestimmten Arbeitsgang sind für $n = 5$ Arbeiter die sogenannten Stückzeiten in Minuten je Stück festgestellt worden. Die durchschnittliche Stückzeit der Gruppe von fünf Arbeitern soll unter der Annahme berechnet werden, dass vier Arbeiter 8 Stunden arbeiten und der fünfte Arbeiter 4 Stunden arbeitet:

Arbeitsaufwand w_i (in Minuten)	Stückzeit x_i (in Minuten/Stück)	Fertigung w_i/x_i (in Stück)
480	0,8	480/0,8 = 600
480	1,0	480/1,0 = 480
480	1,2	480/1,2 = 400
480	1,2	480/1,2 = 400
240	1,5	240/1,5 = 150
$\sum w_i = 2160$		$\sum (w_i/x_i) = 2040$

$$\bar{x}_H = \frac{\sum w_i}{\sum (w_i/x_i)} = \frac{2160}{2040} = 1,059$$

Die durchschnittliche Stückzeit beträgt somit 1,06 Minuten/Stück.

Beispiel 4: Aus den Daten der folgenden Tabelle ist die durchschnittliche Bevölkerungsdichte für das Gesamtgebiet beider Länder zu berechnen.

Land	Bevölkerungsdichte x_i (Einwohner/qkm)	Wohnbevölkerung N_i (Mill.)	w_i
1	250	60	0,75
2	300	20	0,25

Man erhält:

$$\bar{x}_H = \frac{\frac{60 + 20}{\frac{60}{250} + \frac{20}{300}}}{\frac{60}{250} + \frac{20}{300}} = 261 \text{ Einw./qkm bzw. mit den Gewichten}$$

$$w_i [0 \leq w_i \leq 1 \quad \text{und} \quad \sum_{i=1}^n w_i = 1]$$

$$\bar{x}_{H,g} = \frac{1}{\sum_{i=1}^n \frac{w_i}{x_i}} = \frac{1}{\frac{0,75}{250} + \frac{0,25}{300}} = 261 \text{ Einwohner/qkm}$$

Für **klassierte Beobachtungen** (k Klassen) mit den Klassenmitten x_i und den Häufigkeiten f_i (wobei $\sum_{i=1}^k f_i = n$) gilt:

$$\bar{x}_G = \sqrt[n]{x_1^{f_1} \cdot x_2^{f_2} \cdot \dots \cdot x_k^{f_k}} \quad \text{oder} \quad \lg \bar{x}_G = \left[\frac{1}{n} \right] \sum_{i=1}^k f_i \lg x_i \quad \text{mit} \quad x_i > 0 \quad (3.35)$$

$$\frac{1}{\bar{x}_H} = \left[\frac{1}{n} \right] \sum_{i=1}^k f_i / x_i \quad \text{oder} \quad \bar{x}_H = n / \sum_{i=1}^k f_i / x_i \quad \text{mit} \quad x_i \neq 0 \quad (3.36)$$

Zwischen den **drei Mittelwerten** besteht nach A.-L. Cauchy (1789–1857) die folgende Beziehung

$$\bar{x}_H \leq \bar{x}_G \leq \bar{x} \quad (3.37)$$

wobei die Gleichheitszeichen für gleiche Stichprobenwerte gelten. Für zwei Werte gilt

$$\bar{x} / \bar{x}_G = \bar{x}_G / \bar{x}_H \quad \text{oder} \quad \bar{x} \bar{x}_H = \bar{x}_G^2 \quad (3.38)$$

Hinweis: Das **quadratische Mittel** $\left[\sqrt{\frac{1}{n} \sum x_i^2} \right]$, es ist $\geq \bar{x}$, wird selten verwendet.

3.4 Häufigkeitsverteilung

- Histogramm
- Stamm-Blatt Darstellung

Die Zuordnung von Häufigkeiten zu den Merkmalsausprägungen heißt Häufigkeitsverteilung. Die Maßzahlen in den vorangehenden Abschnitten kennzeichnen ausgewählte Eigenschaften (speziell die mittlere Lage und die Variation betreffend) einer Häufigkeitsverteilung von Messwerten. Insbesondere hinsichtlich der Form der Verteilung, unimodal (eingipflig) oder multimodal, symmetrisch oder asymmetrisch (links-steil, rechts-steil), sagen die Maßzahlen nichts aus. Spezielle graphische Darstellungsformen wie **Histogramm**, Häufigkeitspolygon und **empirische Verteilungsfunktion** müssen (bei ausreichendem Stichprobenumfang) genutzt werden, um Fehlinterpretationen bei der Beurteilung von Häufigkeitsverteilungen zu vermeiden.

3.4.1 Histogramm

Um eine Vorstellung von der Form einer Stichprobenverteilung zu erhalten und die Kenngrößen leichter berechnen und besser interpretieren zu können, werden die n Beobachtungen ($n > 50$) in k Klassen gruppiert. Man bildet, entsprechend dem Umfang n der Stichprobe, etwa 7 bis 20 Klassen mit gleicher Klassenbreite b . Die Anzahl der Beobachtungen, die in die j -te Klasse fallen, wird mit n_j bezeichnet ($j = 1, \dots, k$); n_j wird **absolute Häufigkeit** oder **Besetzungszahl** der

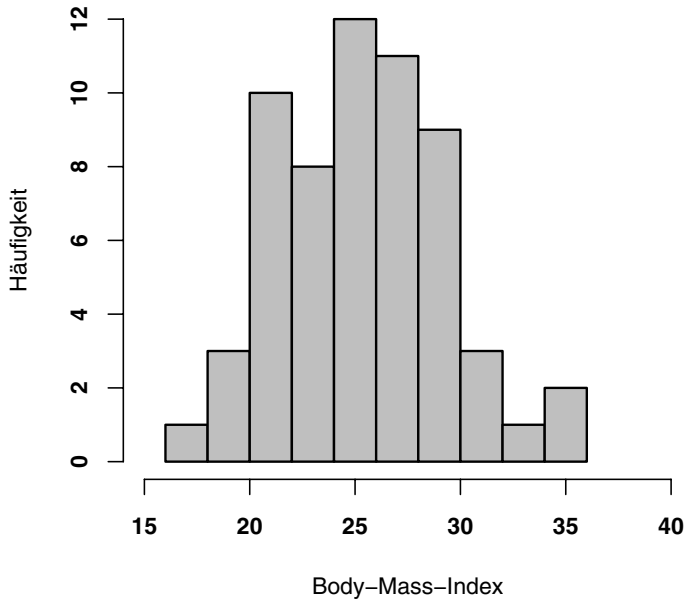


Abb. 3.7. Histogramm zur Verteilung des Body-Mass-Index, gemessen bei $n=60$ Personen

Klasse j genannt. Die relativen Besetzungszahlen $h_j = n_j/n$ können auch in Prozent $100n_j/n\%$ ausgedrückt werden.

Durch schrittweises Aufsummieren ergeben sich die **absoluten Summenhäufigkeiten** (aufsummierte Besetzungszahlen) $B_j = n_1 + n_2 + \dots + n_j$; aus diesen Werten werden nach $H_j = 100B_j/n\%$ auch die prozentualen Summenhäufigkeiten berechnet.

In einem Histogramm werden in einem Koordinatensystem die Häufigkeiten durch Rechtecke der Höhe n_j und der Breite b über den Klassenmitten auf der x-Achse aufgetragen. Im Programm **R** wird das Histogramm mit der Funktion **hist()** erstellt.

Beispiel: Bei $n = 60$ Personen wurde der Body-Mass-Index (aus dem Verhältnis von Körpergewicht (kg) und Körpergröße² (m²)) bestimmt. 

```
> bmi <- c(20.8, 29.7, 27.6, 28.6, 20.7, 21.0, 23.1, 21.9, 24.8, 25.3, 27.1, 23.3,
+ 19.5, 25.2, 25.8, 21.6, 28.7, 30.6, 23.3, 26.6, 35.3, 17.0, 22.6, 25.9,
+ 29.0, 23.7, 21.7, 26.5, 18.5, 24.5, 29.0, 23.2, 27.9, 18.8, 27.1, 21.5,
+ 26.5, 20.3, 25.5, 32.0, 26.7, 34.9, 24.6, 25.6, 26.7, 22.1, 28.8, 28.1,
+ 28.8, 32.2, 30.3, 24.9, 28.0, 21.1, 22.0, 25.5, 24.0, 26.6, 24.7, 28.8)
> hist(bmi, breaks=c(16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36), col="grey",
+      xlim=c(15, 40), xlab="Body-Mass-Index", ylab="Häufigkeit", main=" ")
```

Der kleinste Wert ist 17,0, der größte Wert ist 35,3. Für die Darstellung der Häufigkeitsverteilung im Histogramm ist somit die Wahl einer Klassenbreite $b = 2$ sinnvoll (10 Klassen), wobei man die Klassengrenzen möglichst ganzzahlig festlegt (vgl. Tabelle 3.8).

Die Rechteckflächen im Histogramm sind proportional zu den Anteilen an der Häufigkeitsverteilung. Im Beispiel haben 12 (20%) Personen einen BMI-Wert, der größer ist als 24 und kleiner oder gleich 26. Andererseits haben 45 (75%) Personen einen BMI-Wert, der kleiner oder gleich 28 ist. Dieser Wert entspricht somit dem 3. Quartil.

Tabelle 3.8. Häufigkeitsverteilung zu klassierten Messwerten (Body-Mass-Index [kg/m^2]) mit als Prozentzahlen angegebenen Häufigkeitssummen

j	Klassen- grenzen	Mitte	Häufigkeit (absolut) n_j	kum. Häufig- keit B_j	rel. Häufig- keit h_j	kum. relative Häufigkeit H_j
1	$16 < x \leq 18$	17	1	1	1,7	1,7
2	$18 < x \leq 20$	19	3	4	5,0	6,7
3	$20 < x \leq 22$	21	10	14	16,7	23,3
4	$22 < x \leq 24$	23	8	22	13,3	36,7
5	$24 < x \leq 26$	25	12	34	20,0	56,7
6	$26 < x \leq 28$	27	11	45	18,3	75,0
7	$28 < x \leq 30$	29	9	54	15,0	90,0
8	$30 < x \leq 32$	31	3	57	5,0	95,0
9	$32 < x \leq 34$	33	1	58	1,7	96,7
10	$34 < x \leq 36$	35	2	60	3,3	100,0

Die letzte Spalte H_j in Tabelle 3.8 kann verwendet werden, um näherungsweise die **empirische Verteilungsfunktion** (kumulierte relative Häufigkeiten, relative Summenhäufigkeiten) darzustellen (linke Seite in Abbildung 3.8).

Das Programm **R** stellt für die Darstellung der empirischen Verteilungsfunktion aus Messwerten die Funktion **ecdf()** zur Verfügung (rechte Seite in Abbildung 3.8). Dabei werden die kumulierten Häufigkeiten für jeden einzelnen Messwert berücksichtigt, so dass die entsprechenden Quantile, z.B. für 80% (0,80) der BMI-Wert 28,5, direkt abgelesen werden können. Die empirische Verteilungsfunktion für eine eingipflige symmetrische Häufigkeitsverteilung zeigt, vom Wert 0 auf den Wert 1 ansteigend, typischerweise einen **sigmoiden (s-förmigen) Verlauf**, wie im Beispiel punktsymmetrisch zum Mittelwert (50% entsprechend 0,50). Asymmetrische oder mehrgipflige Häufigkeitsverteilungen sind in der Regel sehr einfach am Verlauf der empirischen Verteilungsfunktion zu erkennen.

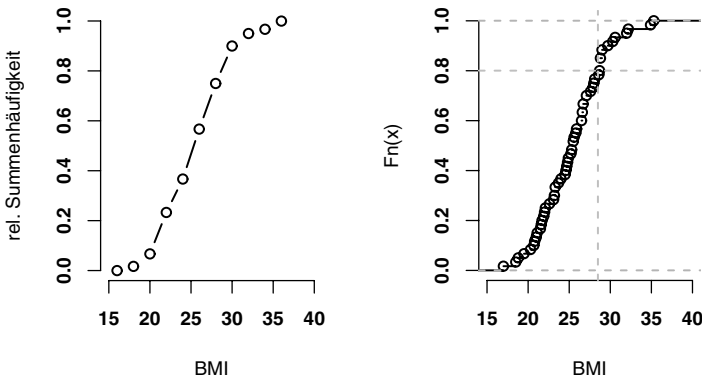


Abb. 3.8. Empirische Verteilungsfunktion zum Body-Mass-Index, gemessen bei $n=60$ Personen: links nach den Klassen aus der Histogrammdarstellung; rechts auf der Basis der kumulierten Einzelwerte



Das Histogramm und die empirische Verteilungsfunktion bilden die Grundlage zum anschaulichen Verständnis der **Wahrscheinlichkeitsfunktion** (Wahrscheinlichkeitsdichte) und der **Verteilungsfunktion** von Zufallsvariablen in der Statistik (Kapitel [5]).

3.4.2 Stamm-Blatt Darstellung

Eine wichtige und aufschlussreiche Alternative zum Histogramm ist die so genannte Stamm-Blatt-Darstellung (stem-and-leaf-Plot). Dabei werden die Messwerte direkt der Größe nach wiedergegeben, wobei die vorderen Dezimalstellen den Stamm, und die hinteren Dezimalstellen die Blätter darstellen. Im folgenden Beispiel werden Stamm und Blätter durch den Dezimalpunkt (Komma) getrennt.

Stamm-und-Blatt-Darstellung für die Verteilung:			Stamm Blatt	Schema des Vergleichs zweier Verteilungen
14,3	14,7	15,0	18 3	Blatt St. Blatt
15,2	15,6	15,6	17 69	· · · · · · · · · · ·
15,7	16,0	16,5	16 05	· · · · · · · · · · ·
17,6	17,9	18,3	15 02667	· · · · · · · · · · ·
			14 37	· · · · · · · · · · ·

Back-to-back stem-and-leaf plot

Das Programm **R** stellt zur Stamm-Blatt-Darstellung die Funktion **stem()** zur Verfügung:

```
> stem(bmi)

The decimal point is at the |

16 | 0
18 | 585
20 | 378015679
22 | 01612337
24 | 0567892355689
26 | 5566771169
28 | 0167888007
30 | 36
32 | 02
34 | 93
```

3.5 Konzentration; Gini Index

Das Streuungskonzept fragt danach, wie sich die einzelnen Beobachtungswerte um einen zentralen Wert herum verteilen. Das **Konzentrationskonzept** fragt, ob sich große Merkmalsbeträge geballt auf nur wenige Merkmalsträger konzentrieren. Die Verteilung der Merkmalsbeträge auf die Merkmalsträger kann so sein, dass sich **bei wenigen** ein Großteil des Gesamtbetrages konzentriert. Die Verteilung könnte aber auch so sein, dass der gesamte Merkmalsbetrag **gleichmäßig** auf alle Merkmalsträger aufgeteilt ist (Gleichverteilung = keine Konzentration). Bei wenigen kann in diesem Zusammenhang bedeuten: - absolut wenige: auf wenige Merkmalsträger konzentriert sich der Merkmalsbetrag - relativ wenige: auf einen geringen Anteil von Merkmalsträgern fällt ein hoher Anteil des Merkmalsbetrages.

Die gebräuchlichste Art der Darstellung der relativen Konzentration eines Merkmales ist die so genannte **Lorenzkurve** (nach M.O. Lorenz). Auf der Abszisse (u-Achse) erscheinen für die Gruppe von Untersuchungseinheiten deren kumulierte relative Häufigkeiten. Auf der Ordinate (v-Achse) erscheinen die zugehörigen kumulierten Anteile an der Gesamtsumme.

$$u = \frac{m}{n} \quad (3.39)$$

$$v = \frac{\sum_{i=1}^m x_i}{\sum_{i=1}^n x_i}$$

Je näher die empirisch ermittelte Konzentrationskurve an der Diagonalen liegt, um so geringer ist die relative Konzentration. Je stärker die Lorenzkurve **durchhängt**, je größer ist auch die Konzentration. Die Lorenzkurve ist eine Gerade mit der Steigung 1 (Diagonale), wenn auf p-Prozent der Fälle (Merkmalsträger) auch p-Prozent der entsprechenden Merkmalssumme entfallen.

Der Gini-Index misst die relative Konzentration; er beträgt das zweifache der Fläche zwischen der Diagonalen des Einheitsquadrates und der Lorenzkurve. Sehr unterschiedliche Kurven führen möglicherweise zu dem selben Gini-Koeffizienten. Daher sind beide zur Bewertung der Konzentration in den Beobachtungen zu berücksichtigen.

Beispiel: 6 Unternehmen teilen sich den Marktanteil an einem bestimmten Produkt wie folgt auf:

Firma	1	2	3	4	5	6	Σ
Anteil	2	8	10	15	20	45	100

```
> gini <- function(x, y) {
+   area <- 0
+   # Berechnung der Fläche – Trapezregel
+   for (i in 2:n+1) area <- area + 0.5*((x[i]-x[i-1])*(y[i]+y[i-1]))
+   gini <- 1 - 2*area; round(gini, 3) # Gini-Index
+ }
> b <- c(2, 8, 10, 15, 20, 45); n <- length(b)
> u <- c(0, (1:n)/n); v <- c(0, (cumsum(b) / sum(b)))
> gini(u,v)
[1] 0.43
```

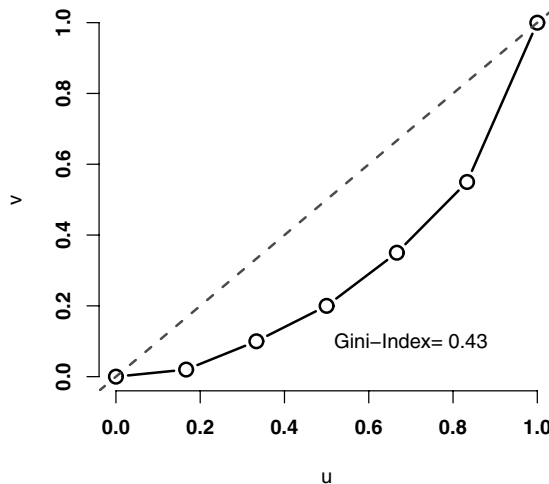


Abb. 3.9. Lorenz-Kurve zu den Marktanteilen von 6 Unternehmen

3.6 Maßzahlen für den Zusammenhang metrischer Daten

- Punktwolken
- Die empirische Kovarianz
- Der empirische Korrelationskoeffizient
- Der Rangkorrelationskoeffizient
- Typisierung korrelativer Zusammenhänge
- Die lineare Regression
- Spezielle Schätzungen der Regressionsgeraden
- Robuste lineare Regression

In vielen Situationen ist es wünschenswert, etwas über die Abhängigkeit zwischen zwei metrischen Merkmalen eines Individuums, Materials, Produktes oder Prozesses zu erfahren. In einigen Fällen mag es auf Grund theoretischer Überlegungen sicher sein, dass zwei Merkmale, X und Y , miteinander zusammenhängen. Das Problem besteht dann darin, Art und Grad des Zusammenhangs zu ermitteln.

3.6.1 Punktwolken

Zunächst wird man die **Wertepaare** (x_i, y_i) in ein *Koordinatensystem* eintragen. Hierdurch erhält man eine Vorstellung über **Streuung und Form der Punktwolke**.

Beispiel: Die mittlere Körpergröße (cm) der Kinder aus Kalama (ein ägyptisches Dorf, das an einer Studie zur Ernährungssituation in den Entwicklungsländern teilnahm) soll dem Alter (in Monaten) in einer Punktwolke gegenübergestellt werden.



Tabelle 3.9. Alter und Körpergröße bei Kindern aus Kalama

Gruppe	1	2	3	4	5	6	7	8	9	10	11	12
Alter x_i	18	19	20	21	22	23	24	25	26	27	28	29
Größe y_i	76.1	77.0	78.1	78.2	78.8	79.7	79.9	81.1	81.1	81.8	82.8	83.5

Für die Erstellung von Punktwolken in **R** ist die allgemeine Plotfunktion **plot()** gut geeignet. Das Ergebnis ist in Abbildung 3.10 wiedergegeben.

```
> x <- seq(18, 29, by=1)
> y <- c(76.1, 77.0, 78.1, 78.2, 78.8, 79.7, 79.9, 81.1, 81.2, 81.8, 82.8, 83.5)
> plot(x, y, pch=16, cex=1.5, xlab="Alter (Monate)", ylab="Größe (cm)",
+      xlim=c(17, 30), ylim=c(75, 85))
```

Liegen die Punkte „nahezu“ auf einer Geraden, dann lässt sich nach Augenmaß, mehr oder weniger sicher, eine **Ausgleichsgerade** einzeichnen. Mit ihrer Hilfe lässt sich ablesen, welcher y -Wert zu einem vorgegebenen x -Wert erwartet werden kann. Lässt die beobachtete Punktwolke einen linearen Anstieg (Abfall) erkennen, dann kann rechnerisch eine geeignete **Regressionsgerade** bestimmt und die Stärke eines linearen Zusammenhangs mit dem **Korrelationskoeffizienten** beurteilt werden. Die Punktwolke gibt direkt weiterhin Aufschluss über die beurteilbaren Wertebereiche (**Definitionsbereich**) der x - und y -Werte und die **Streuung in den Wertepaaren**. Insbesondere können auch **Ausreißer** recht einfach erkannt werden.

3.6.2 Die empirische Kovarianz

Ein Maß für den wechselseitigen linearen Zusammenhang zweier Messwertreihen ist die empirische Kovarianz [mittleres Abweichungsprodukt (3.40)]. Sie basiert auf der Überlegung, dass die

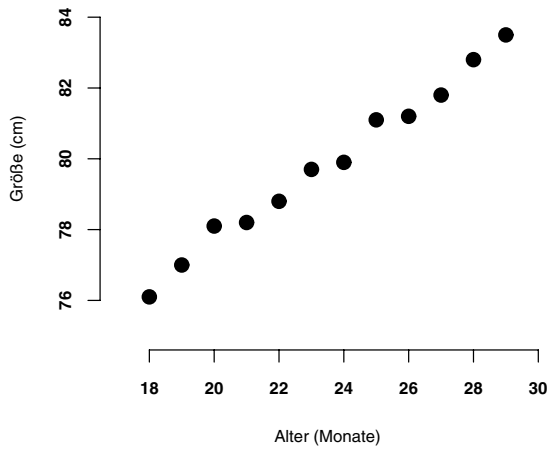


Abb. 3.10. Punktwolke zu Alter und Körpergröße der Kinder aus Kalama

Streuung der beobachteten Punktwolke durch die **Summe der Rechtecke** $\sum (x_i - \bar{x})(y_i - \bar{y})$ beschrieben werden kann. Dabei wird der „zentral“ gelegene Punkt $(\bar{x}, \bar{y})$ als **Schwerpunkt der Punktwolke** bezeichnet.

$$s_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n - 1} \quad (3.40)$$

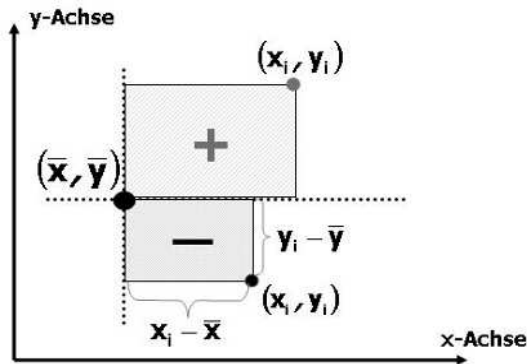


Abb. 3.11. Kennzeichnung der Streuung in bivariaten Messungen

Das Vorzeichen der empirischen Kovarianz resultiert aus der Orientierung am Schwerpunkt der Punktwolke, d.h. inwieweit positive oder negative **Abweichungsprodukte** (Rechtecke) überwiegen. Die empirische Kovarianz ist 0, wenn beide Anteile gleich groß sind. In diesem Fall bezeichnet man die beiden Messwertreihen als **statistisch unabhängig**. Andererseits ist die empirische Kovarianz **nicht normiert**. Sie ist mit Dimensionsangaben zu versehen und ihr maximaler Wert hängt von der Streuung in den x - und y -Werten ab.

3.6.3 Der empirische Korrelationskoeffizient

Der empirische Korrelationskoeffizient als **Maßzahl für die Stärke eines linearen Zusammenhangs** wird mittels einer Normierung der empirischen Kovarianz durch das Produkt der Standardabweichungen berechnet.

$$r = \frac{s_{xy}}{s_x s_y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (3.41)$$

Die Berechnung des empirischen Korrelationskoeffizienten erfolgt praktisch nach (3.42) durch Umformung von (3.41):

$$r = \frac{\sum xy - \frac{1}{n}(\sum x)(\sum y)}{\sqrt{[\sum x^2 - \frac{1}{n}(\sum x)^2][\sum y^2 - \frac{1}{n}(\sum y)^2]}} \quad (3.42)$$

Beispiel: Für die Rechnung nach dieser Formel ist ein Tabellenschema hilfreich, in dem die entsprechenden Summen bestimmt werden können. Einige „künstliche“ Beispieldaten sind in Tabelle 3.10 aufgeführt.



Tabelle 3.10. Beispieldaten zur Korrelation und Regression

x	y	x^2	y^2	xy
13	12	169	144	156
17	17	289	289	289
10	11	100	121	110
17	13	289	169	221
20	16	400	256	320
11	14	121	196	154
15	15	225	225	225
103	98	1593	1400	1475

$$r = \frac{1475 - \frac{1}{7}103 \cdot 98}{\sqrt{[1593 - \frac{1}{7}103^2][1400 - \frac{1}{7}98^2]}} = 0,7087357 = 0,709$$

Die Berechnung des Korrelationskoeffizienten in **R** erfolgt durch die Funktion **cor()**:

```
> x <- c(13, 17, 10, 17, 20, 11, 15)
> y <- c(12, 17, 11, 13, 16, 14, 15)
>
> cov(x, y)                                # Berechnung der Kovarianz
[1] 5.5
> cor(x, y)                                # Berechnung des Korrelationskoeffizienten
[1] 0.7087357
```

Der Korrelationskoeffizient r ist ein Maß für die **Stärke eines linearen Zusammenhangs** zwischen zwei Messwertreihen:

- r ist normiert, d.h. $-1 \leq r \leq +1$
- ein positiver Zusammenhang liegt vor, wenn $r > 0$.
- ein negativer Zusammenhang liegt vor, wenn $r < 0$.
- kein linearer Zusammenhang liegt vor, wenn $r = 0$.
- ein $r = 1$, ($r = -1$) bedeutet, dass alle Punkte auf einer Geraden mit positiver (negativer) Steigung liegen.

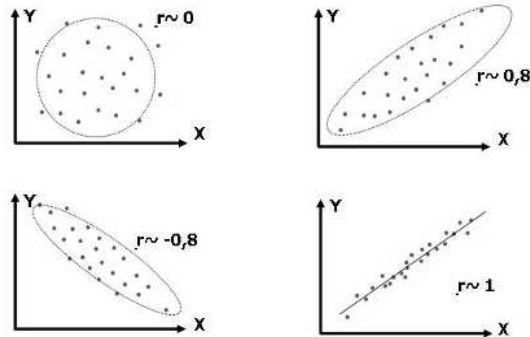


Abb. 3.12. Punktwolken zu unterschiedlichen Korrelationskoeffizienten (beachte: links unten liegt eine „negative“ Korrelation vor)



Die Berechnung des Korrelationskoeffizienten ist unsinnig, wenn aufgrund der Punktwolke kein linearer Zusammenhang (Punktwolke) erkennbar ist! Die Höhe des Korrelationskoeffizienten hängt empfindlich von einzelnen Extremwerten (Ausreißern) und von der Häufigkeitsverteilung der x - und y -Werte ab. Ein robustes Maß für die Stärke des Zusammenhangs ist der Rangkorrelationskoeffizient nach Spearman.

3.6.4 Der Rangkorrelationskoeffizient

Zur Berechnung des Rangkorrelationskoeffizienten transformiert man beide Reihen durch Zuordnung von Rangzahlen in Rangreihen, bildet die Differenzen D der n **Rangpaare**, quadriert und summiert sie zu $\sum D^2$ und setzt diesen Wert in folgende Formel ein.

$$r_S = 1 - \frac{6 \sum D^2}{n(n^2 - 1)} \quad (3.43)$$

Bei gleichen Werten, man spricht von sogenannten **Bindungen** (ties), werden mittlere Rangplätze zugeordnet; in einer der beiden Reihen sollten **höchstens etwa 1/5 der Beobachtungen ranggleich** sein. Sind viele Bindungen vorhanden, so berechnet man r_S , indem man den Korrelationskoeffizient r aus den Rängen berechnet oder eine Korrektur für die Bindungen einführt.

Wenn zwei Rangordnungen gleich sind, werden die Differenzen Null, d. h. $r_S = 1$. Wenn eine Rangordnung die Umkehrung der anderen ist, also vollständige Diskrepanz besteht, erhält man $r_S = -1$. Der Rangkorrelationskoeffizient kann somit in gleicher Weise interpretiert werden wie der Korrelationskoeffizient ($-1 \leq r_S \leq +1$).



Beispiel: Der Zusammenhang zwischen mathematischer und altsprachlicher Begabung von 8 Schülern (S) anhand von Schulnoten in Latein (L) und in Mathematik (M) soll mit dem Rangkorrelationskoeffizient nach Spearman (mit und ohne Berücksichtigung der Bindungen) bewertet werden:

S	d	b	g	a	f	e	h	c
L	1	2	2	2	3	3	4	4
M	2	4	1	3	4	3	4	3
R_L	1	3	3	3	5,5	5,5	7,5	7,5
R_M	2	7	1	4	7	4	7	4
D	-1	-4	2	-1	-1,5	1,5	0,5	3,5
D^2	1	16	4	1	2,25	2,25	0,25	12,25

$$\sum D^2 = 39 \quad r_S = 1 - \frac{6 \cdot 39}{8(8^2 - 1)} = 0,536$$

$$T_L = \frac{1}{2}[(3^3 - 3) + (2^3 - 2) + (2^3 - 2)] = 18$$

$$T_M = \frac{1}{2}[(3^3 - 3) + (3^3 - 3)] = 24$$

$$r_{S,B} = 1 - \frac{6 \cdot 39}{(8^3 - 8) - (18 + 24)} = 0,4935$$

[zur Berechnung des Rangkorrelationskoeffizienten vgl. auch 3.44]

Ohne Berücksichtigung der Bindungen wird die Korrelation überschätzt ($0,536 > 0,494$). Die elementare Berechnung des Rangkorrelationskoeffizienten in **R** soll mit der Funktion **rank()** gezeigt werden.

```
> L <- c(1, 2, 2, 2, 3, 3, 4, 4)
> M <- c(2, 4, 1, 3, 4, 3, 4, 3)
> r.L <- rank(L); r.L          # Rangzahlen zu x
[1] 1.0 3.0 3.0 3.0 5.5 5.5 7.5 7.5
> r.M <- rank(M); r.M          # Rangzahlen zu y
[1] 2 7 1 4 7 4 7 4
>
> D <- r.L - r.M; n <- length(D)
> 1 - 6*sum(D^2)/(n*(n^2-1))    # Rangkorrelationskoeffizient (Spearman)
[1] 0.5357143
> cor(r.L, r.M)                # Korrelationskoeffizient aus Rangzahlen
[1] 0.4935481
```

3.6.4.1 Spearmansche Rangkorrelation bei Bindungen

Nur wenn **Bindungen** (gleiche Werte) gehäuft auftreten, lohnt sich die folgende Korrekturformel (vgl. Text unter (3.43):

$$r_{S,B} = 1 - \frac{6 \sum D^2}{(n^3 - n) - (T_{x'} + T_{y'})} \quad (3.44)$$

$$T_{x'} = \frac{1}{2} \cdot \sum (t_{x'}^3 - t_{x'}) ; \quad T_{y'} = \frac{1}{2} \cdot \sum (t_{y'}^3 - t_{y'})$$

mit $t_{x'}$ (der Strich am x deutet an, dass wir uns auf Ranggrößen beziehen) gleich der Anzahl der Bindungen in aufeinanderfolgenden Gruppen (gleicher Ranggrößen) der x' -Reihe, $t_{y'}$ gleich der Anzahl der Bindungen in aufeinanderfolgenden Gruppen (gleicher Ranggrößen) der y' -Reihe: Man zählt also an der ersten Gruppe, wie oft derselbe Wert erscheint, setzt diese Häufigkeit in die dritte Potenz und subtrahiert hiervon die Häufigkeit. Entsprechend verfährt man mit allen Gruppen und bildet schließlich die Summen $T_{x'}$ und $T_{y'}$; demonstriert am obigen Schulnoten-Beispiel.

3.6.5 Typisierung korrelativer Zusammenhänge

Man spricht von stochastischer Abhängigkeit oder von einem **stochastischen Zusammenhang**, wenn die Nullhypothese, es bestehe stochastische Unabhängigkeit, widerlegt wird. Die sachliche Deutung gefundener statistischer Zusammenhänge und ihre Prüfung auf mögliche kausale Zusammenhänge liegt außerhalb der statistischen Methodenlehre.

„**Kausalität**“ in der Statistik lässt sich auffassen als stochastischer Zusammenhang, der durch Störgrößen nicht zu erklären ist und weitgehend vom Fachwissen und nicht von der Statistik begründet wird. Kausale Schlüsse aus statistischer Sicht setzen voraus, dass wiederholbare randomisierte Experimente vorliegen: die Untersuchungseinheiten sind vor den unterschiedlichen Behandlungen als gleichwertig und die Behandlungsergebnisse als möglicherweise kausal unterschiedlich aufzufassen. Echte Kausaleffekte sind nicht beobachtbar, da jeder Untersuchungseinheit nur eine Behandlung zugeordnet werden kann. Werden kausale Schlüsse bei Erhebungen angestrebt, so sollten es Längsschnittstudien vom Typ einer Kohortenstudie sein.

Erscheint die stochastische Abhängigkeit gesichert, dann ist zu bedenken, dass die Existenz eines funktionalen Zusammenhangs – beispielsweise die Zunahme der Störche und der Neugeborenen während eines gewissen Zeitraumes in Schweden – nichts aussagt über den kausalen Zusammenhang. So kann zwischen der Dosis eines Arzneimittels und der Letalität einer Krankheit eine ausgesprochen positive Korrelation bestehen, da bei sehr ernsten Erkrankungen die Letalität nicht wegen der größeren Dosis des Medikaments, sondern trotz derselben erhöht ist. Eine Korrelation kann durch direkte kausale Zusammenhänge zwischen X und Y , durch eine gemeinsame Abhängigkeit von dritten Größen oder durch Heterogenität des Materials oder rein formal bedingt sein.

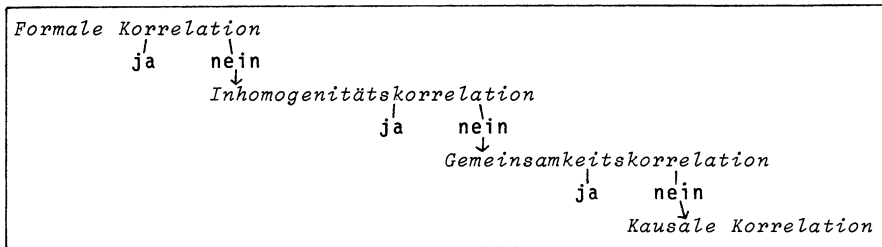
Kausale Korrelationen existieren z. B. zwischen Begabung und Leistung, zwischen Dosis und Wirkung von Heilmitteln, zwischen Arbeitszeit und Preis von Produkten. Beispiele für eine **Gemeinsamkeitskorrelation** sind der Zusammenhang zwischen Körpermaßen, etwa zwischen der Länge des rechten und linken Armes oder zwischen Körperlängen und Körpergewicht sowie die Korrelation zwischen Zeitreihen: Die Abnahme der Zahl der Storchennester in Ostpreußen und die Abnahme der Zahl der Geburten: Basis war die zunehmende Industrialisierung (Scheinkorrelation, „spurious correlation“).

Bei der **Inhomogenitätskorrelation** besteht das Material aus verschiedenen Teilmassen, die in verschiedenen Bereichen des Koordinatensystems liegen. Unterscheidet man die Teilmassen nicht, so wird durch die Lageunterschiede der Punktwolken ein Korrelationseffekt erzielt, der die Korrelationsverhältnisse innerhalb der Teilmassen völlig verändern kann. Besonders eindrucksvoll ist folgendes Beispiel: Der Hämoglobingehalt des Blutes und die Oberflächengröße der Blutkörperchen zeigen weder bei Neugeborenen noch bei Männern noch bei Frauen eine Korrelation. Die Werte sind $-0,06$ bzw. $-0,03$ bzw. $+0,07$. Würde man das Material zusammenfassen, so erhielte man für das Gesamtmaterial einen Korrelationskoeffizienten von $+0,75$.

Auch die positive Korrelation zwischen Schuhgröße und Einkommen gehört hierher. Sie kam zustande, wenn man Männer und Frauen gemeinsam betrachtete. Da Frauen durchschnittlich kleinere Schuhe und niedrigere Einkommen hatten, kam bei gemeinsamer Betrachtung aller Daten eine deutliche Korrelation zustande, obwohl sie weder bei den Männern noch bei den Frauen auftrat.

Werden zwei unabhängige Variablen X und Y durch die Variable Z dividiert, so wird hiermit zwischen X/Z und Y/Z eine sog. **Indexkorrelation** erzeugt (vgl. McNemar [McN69]). Sind beispielsweise X und Y sich zu 100% ergänzende Prozentsätze, so muss zwangsläufig eine negative Korrelation zwischen ihnen auftreten, Eiweiß und Fettanteile in Nahrungsmitteln usw. Der Ausdruck „**Schein-Korrelation**“ ist für diese Zusammenhänge üblich, er ist jedoch besser zu vermeiden, da ja auch eine Schein-Korrelation zwischen zwei Prozentzahlen nicht Schein, sondern für die betrachteten Variablen Tatsache ist. Neben dieser **formalen Korrelation** gibt es, wie oben

angedeutet worden ist, noch eine Reihe weiterer nichtkausaler Korrelationen. In einer Deutungsanalyse von Korrelationen im praktischen Anwendungsfall gibt Koller [Kol63] Richtlinien, die es gestatten, durch Ausschließung anderer Möglichkeiten (vgl. die Selektionskorrelation) echte oder besser kausale Korrelationen zu erkennen. Danach kann man zur Deutung einer Korrelation so vorgehen, dass man prüft, ob eine formale Korrelation vorliegt. Kann dies verneint werden, so wird nach folgendem Schema weiter geprüft:



Die Anerkennung einer kausalen Korrelation erfolgt also durch Ausschließen der anderen Möglichkeiten. Wegen der möglichen Überschneidung der Typen lässt sich das Schema in der Praxis nicht immer so streng und ausschließend anwenden, wie es im Modell dargestellt wird. Häufig wird man auch nicht bis zum Typ der kausalen Korrelation vordringen, sondern bereits vorher stehen bleiben und diesen Typ für den jeweiligen Fall nicht widerlegen können. Die Höhe des Korrelationskoeffizienten wird dabei nur selten eine Rolle spielen.

Inhomogenitätskorrelation und Gemeinsamkeitskorrelation sind als Schein-Korrelationen (vgl. auch Aitchison [Ait87]) durch Drittvariablen bedingt; diese können natürlich auch eine Korrelation zwischen X und Y verdecken. Mehr hierüber sowie die Beziehungen zur **kollektiven Korrelation**, auch **Gruppenfehlschluss** genannt, ist Tiede [Tie87] zu entnehmen, der auf die so genannte **Kovarianz-Zerlegung** näher eingeht.

3.6.6 Die lineare Regression

Ziel der Regressionsanalyse ist es, anhand einer empirischen Funktion $\bar{y}_i(x_i)$, der graphischen Darstellung der bedingten Mittelwerte $\bar{y}_i(x_i)$ als Funktion von x_i eine *funktionale Beziehung* zwischen den Größen Y und X zu finden, die es gestattet, aus vorgegebenen bzw. zu beliebigen Werten der unabhängigen Einflussgröße X die jeweils abhängige Zielgröße Y zu schätzen. Man spricht von der **Regression von Y auf X**.

Die Bezeichnung geht historisch auf Untersuchungen von Francis Galton zurück, der den Zusammenhang der Körpergröße von Eltern und Kindern im Rahmen der Vererbung untersuchte und dem Phänomen einer „regression to mediocrity“ (regression to the mean) nachging.

Man sollte beachten, dass der Variationsbereich der x -Werte hinreichend groß ist und dass die ersten Differenzen $(y_{i+1} - y_i) = d_i$ und die zweiten Differenzen $(d_{i+1} - d_i)$ keinen Trend aufweisen. Liegen nur Datenpaare (x_i, y_i) vor, so ist diese Beziehung $y_i(x_i)$, d. h. $y_i(x_i)$ als Funktion von x_i im einfachsten Fall die Gleichung der geraden Linie (vgl. auch lineare Funktionen im Abschnitt [2.5.1] auf Seite 43).



Die Kennzahlen der Regressionsgeraden sind, wenn $y = a + bx$ die allgemeine Gleichung der Geraden darstellt, a und b : a stellt den Abschnitt auf der y -Achse dar, der von ihrem Nullpunkt 0 (Ursprung, lat. origo) gerechnet, durch die Regressionsgerade abgetrennt wird, a wird als **Achsenabschnitt** (auf der Ordinate) bezeichnet (**intercept**); b gibt die mittlere Änderung im Merkmal Y an, wenn zu Beobachtungseinheiten übergegangen wird, die im Merkmal X eine Einheit größer sind und heißt die Richtungskonstante, die **Steigung** (**slope**) oder der **Regressionskoeffizient**.

$$y = a + bx = \bar{y} - b\bar{x} + bx = \bar{y} + b(x - \bar{x}) \quad \text{bzw.} \quad y - \bar{y} = b(x - \bar{x}) \quad (3.45)$$

Aufgabe der Regressionsrechnung ist es, die beiden Größen a und b so zu bestimmen, dass die Summe der **Abweichungsquadrate** zwischen den beobachteten y -Werten und den aus der Regressionsgleichung geschätzten Werten minimal ist. Die Abweichungen werden **Residuen** genannt und mit ε (griech. epsilon) gekennzeichnet.

$$\begin{aligned} y_i &= a + bx_i + \varepsilon_i = \hat{y}_i + \varepsilon_i \\ \varepsilon_i &= y_i - \hat{y}_i \\ \sum_{i=1}^n \varepsilon_i^2 &\rightarrow \min \end{aligned}$$

Die Schätzung erfolgt durch die **Methode der kleinsten Quadrate (ordinary least-squares method OLS)** nach Carl Friedrich Gauss (1777 - 1855) und kann als Lösung der entsprechenden Normalgleichungen direkt angegeben werden:

$$\begin{aligned} b &= \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} = \frac{s_{xy}}{s_x^2} \\ a &= \bar{y} - b\bar{x} \end{aligned} \quad (3.46)$$

Für die Berechnung ist somit die Bestimmung der entsprechenden Mittelwerte, Varianzen und der Kovarianz erforderlich. Dazu können die Formeln und Rechenhilfen aus den vorangehenden Abschnitten verwendet werden. In **R** wird die Berechnung durch eine spezielle Funktion **lm()** (**linear model**) geleistet, die im Kapitel Modellbildung [8] ausführlich dargestellt wird.



Beispiel: Im Tierversuch wurde die Entstehung von Tumoren der Lunge (prozentualer Anteil) unter erhöhten Expositionen von Asbestfasern (50 bis 3000 Fasern/ml) in der Atemluft untersucht.

Tabelle 3.11. Entwicklung von Tumoren der Lunge in Abhängigkeit von der Höhe der Exposition mit Asbestfasern

Asbest (Fasern/ml)	50	400	500	900	1100	1600	1800	2000	3000
Tumorentstehung (%)	2	6	5	10	26	42	37	28	50

```
> asbest <- c(50, 400, 500, 900, 1100, 1600, 1800, 2000, 3000)
> lungca <- c(2, 6, 5, 10, 26, 42, 37, 28, 50)
> lm(lungca ~ asbest)
```

```
Call: lm(formula = lungca ~ asbest)
Coefficients: (Intercept)          asbest
            0.54047         0.01772
```

Das Ergebnis der Rechnung kann durch den Graphen der entsprechenden linearen Funktion $\hat{y} = 0,54 + 0,018x$ in die beobachtete Punktwolke eingetragen werden. Die Güte der Anpassung an die beobachteten Werte wird graphisch durch die Darstellung der Residuen bewertet, die markante Modellabweichungen (Abbildung 3.13 rechts) andeuten.

Rechnerisch wird die Stärke einer linearen Abhängigkeit zwischen zwei Merkmalen (x , y) durch den Korrelationskoeffizienten r von Pearson, beschrieben. Das Quadrat r^2 des Korrelationskoeffizienten wird auch als **Bestimmtheitsmaß (B)** bezeichnet. Damit wird der Anteil der Streuung in der y -Variablen beschrieben, der durch die lineare Regression aus der x -Variablen erklärt werden kann.

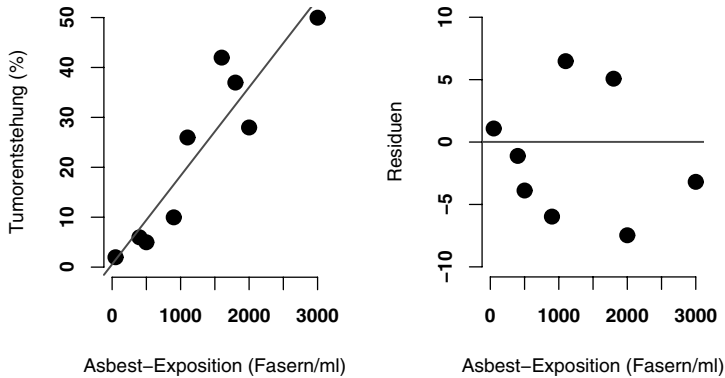


Abb. 3.13. Entwicklung von Tumoren der Lunge in Abhängigkeit von der Höhe der Exposition mit Asbestfasern

$$\left(\frac{s_{xy}}{s_x} \right)^2 = b \cdot s_{xy} = r^2 s_y^2$$

In der Regel wird die Zuordnung der Beobachtungen in abhängige Variable Y und unabhängige Variable X aus der Fragestellung / Problemstellung eindeutig festgelegt sein. Allgemein ist der dargestellte Lösungsweg jedoch auch mit einer Vertauschung der Variablen möglich, man spricht dann von einer Regression von X auf Y , die zu einer anderen Regressionsgeraden führt. Die beiden Regressionsgeraden schneiden sich im Schwerpunkt der Punktwolke $(\bar{x}, \bar{y})$, wobei für den kleineren der beiden Schnittwinkel gilt:

$$\tan(\alpha) = \frac{1 - r^2}{r^2} \frac{s_{xy}}{s_x^2 + s_y^2} = \frac{1 - r^2}{r} \frac{s_x s_y}{s_x^2 + s_y^2}$$

Die beiden Regressionsgeraden fallen zusammen, wenn $|r| = 1$, und somit gilt $\tan(\alpha) = 0$ und $\alpha = 0$. Der Zusammenhang zwischen Korrelation und Regression wird in Abbildung 3.14 zusammenfassend dargestellt. Der absolute Wert des Korrelationskoeffizienten kann als Maß für den Winkel zwischen den beiden Regressionslinien aufgefasst werden. Für $r = 0$ mit $\alpha = 90^\circ$ sind beide Regressionsgeraden orthogonal.

3.6.7 Spezielle Schätzungen der Regressionsgeraden

Im folgenden werden für den Fall, dass nicht nur die Variable Y , sondern auch die Variable X fehlerbehaftet ist (vgl. Tukey [Tuk51], Acton [Act59], Madansky [Mad59], Carlson u. Mitarb. [CSW66]), Schnellschätzungen der Regressionsgeraden nach Bartlett und Kerrich angegeben.

3.6.7.1 Das Bartlett-Verfahren

Nach Bartlett [Bar49] ermittelt man die Steigung der Geraden $y = a + bx$, indem man die n Punkte in der x -Richtung in drei sich nicht überlappende, möglichst gleich große Gruppen teilt, wobei die erste und dritte Gruppe genau k Punkte enthalten und k möglichst nahe an $n/3$ herankommen sollte. Dann ist der Regressionskoeffizient

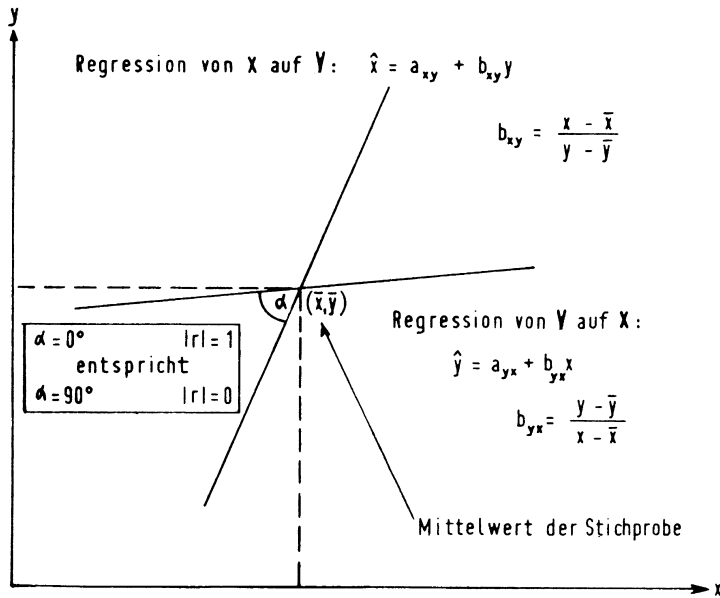


Abb. 3.14. Zusammenhang zwischen Korrelation und Regression

$$\hat{b} = \frac{\bar{y}_3 - \bar{y}_1}{\bar{x}_3 - \bar{x}_1} \quad (3.47)$$

mit $\bar{y}_3$ = Mittelwert y der dritten Gruppe; $\bar{y}_1$ = Mittelwert y der ersten Gruppe; $\bar{x}_3$ = Mittelwert x der dritten Gruppe; $\bar{x}_1$ = Mittelwert x der ersten Gruppe. Der Achsenabschnitt errechnet sich dann nach

$$\hat{a} = \bar{y} - \hat{b}\bar{x} \quad (3.48)$$

wobei $\bar{x}$ und $\bar{y}$ die Mittelwerte aller n Punkte darstellen.

Wenn der Abstand aufeinanderfolgender x -Werte konstant gehalten wird, hat diese Methode eine überraschend hohe Wirksamkeit. Wendy Gibson und Jowett [GJ57] erwähnen in einer interessanten Studie, dass das Verhältnis der drei Gruppen zueinander etwa 1 : 2 : 1 betragen sollte. Doch ist der Unterschied zum Gruppenverhältnis 1 : 1 : 1 nicht sehr kritisch: Bei U -förmigen und rechteckigen Verteilungen ist dieses Verhältnis optimal, während das 1 : 2 : 1-Verhältnis bei J -förmigen und schiefen Verteilungen so wie beim Vorliegen einer Normalverteilung zu bevorzugen ist.

Zur Kontrolle kann die Schnellschätzung $\hat{b} \simeq \sum y / \sum x$ benutzt werden. Geht die Gerade nicht durch den Nullpunkt, so lassen sich anhand der oberen 30% und anhand der unteren 30% der Werte die Kennzahlen a und b abschätzen (Cureton [Cur66]):

$$\hat{b} \simeq \frac{\sum y_{\text{ob.}} - \sum y_{\text{unt.}}}{\sum x_{\text{ob.}} - \sum x_{\text{unt.}}} \quad \hat{a} \simeq \sum y_{\text{unt.}} - \hat{b} \sum x_{\text{unt.}} \quad (3.49)$$

Beispiel: Schätzung der Regressionsgeraden, wenn beide Variablen (X , Y) Messfehler aufweisen.



Stichprobe (Nr.)	Merkmal I (X)	Merkmal II (Y)
1	38,2	54,1
2	43,3	62,0
3	47,1	64,5
4	47,9	66,6
5	55,6	75,7
6	64,0	83,3
7	72,8	91,8
8	78,9	100,6
9	100,7	13,4
10	116,3	138,3

Die angepasste Gerade geht durch den Punkt $(\bar{x}, \bar{y})$ mit den Werten $\bar{x} = 66,48$ und $\bar{y} = 86,03$. Den Regressionskoeffizienten schätzen wir anhand der Mittelwerte des ersten und letzten Drittels beider Reihen nach (3.47):

$$\hat{b} = (\bar{y}_3 - \bar{y}_1)/(\bar{x}_3 - \bar{x}_1) = (120,767 - 60,200)/(98,633 - 42,867) = 1,0861.$$

Den Achsenabschnitt erhalten wir nach (3.48) über die Gesamtmittelwerte zu $\hat{a} = \bar{y} - \hat{b}\bar{x} = 86,03 - 1,0861 \cdot 66,48 = 13,826$. Die angepasste Regressionsgerade lautet somit $\hat{y} = 13,826 + 1,0861x$.

3.6.7.2 Das Kerrich-Verfahren

Wenn beide Variablen fehlerbehaftet sind, ausschließlich positive Werte x_i und y_i auftreten und die Punktwolke sich einer durch den Koordinatenursprung gehenden Geraden ($y = bx$) anschmiegt, kann man zur Schätzung von b folgendes elegante Verfahren (Kerrich [Ker66]) benutzen:

Man bildet für die n unabhängigen Datenpaare (x_i, y_i) die Differenzen $d_i = \lg y_i - \lg x_i$, ihren Mittelwert $\bar{d}$ und die Standardabweichung

$$s_d = \sqrt{\sum (d_i - \bar{d})^2 / (n - 1)} \quad (3.50)$$

Da jeder Quotient y_i/x_i eine Schätzung von b darstellt, ist jedes d_i eine Schätzung von $\lg b$. Ein brauchbarer Schätzwert von $\lg b$ ist $\bar{d}$, und zwar insbesondere dann, wenn die Werte x_i und y_i kleine Variationskoeffizienten aufweisen. Vorausgesetzt, $\lg y_i$ und $\lg x_i$ sind wenigstens angenähert normalverteilt.

Hinweis: Ein 95%-Konfidenzintervall (vgl Kapitel [6]) für β erhält man über

$$\lg b \pm s_d t_{n-1; 0,95} / \sqrt{n} \quad (3.51)$$

Beispiel: Gegeben sind $n = 16$ Datenpaare (angepasste Gerade geht durch den Nullpunkt!) mit $\bar{d} = 9,55911 - 10 = \lg b$ und $s_d = 0,00555$; d. h. $t_{15; 0,95} = 2,131$ und $s_d \cdot t_{n-1; 0,95} / \sqrt{n} = 0,00555 \cdot 2,131 / \sqrt{16} = 0,00296$. Das 95%-Konfidenzintervall für $\lg \beta$ lautet $9,55911 - 10 \pm 0,00296$; d. h. $\hat{b} = 0,362$ und $0,359 \leq \beta \leq 0,365$.



3.6.7.3 Orthogonale kleinste Quadrate

Sind beide Variablen (X , Y) fehlerbehaftet, dann ist es sinnvoll, die Gerade zu bestimmen, die die senkrecht auf ihr stehenden (orthogonalen) Abweichungen minimiert; diese Gerade liegt dann natürlich zwischen $\hat{y}$ und $\hat{x}$. Man schätzt sie anhand von (3.52).

$$\begin{aligned}
 b &= \frac{-(Q_x - Q_y) + \sqrt{(Q_x - Q_y)^2 + 4Q_{xy}}}{2Q_{xy}} \\
 Q_x &= \sum (x - \bar{x})^2 \\
 Q_y &= \sum (y - \bar{y})^2 \\
 Q_{xy} &= \sum (x - \bar{x})(y - \bar{y}) \\
 a &= \bar{y} - b\bar{x}
 \end{aligned} \tag{3.52}$$

Näheres ist z. B. Casella und Berger [CB02] zu entnehmen.



Beispiel: Mit den Daten aus Tabelle 3.10 auf Seite 87 können die Werte nach (3.52) in **R** einfach bestimmt und das Ergebnis im Koordinatensystem dargestellt werden.

```

> x <- c(13, 17, 10, 17, 20, 11, 15)
> y <- c(12, 17, 11, 13, 16, 14, 15)
>
> Q.x <- sum((x - mean(x))^2); Q.y <- sum((y - mean(y))^2)
> Q.xy <- sum((x - mean(x))*(y - mean(y)))
>
> b <- (-(Q.x - Q.y) + sqrt((Q.x - Q.y)^2 + 4 * Q.xy^2)) / (2 * Q.xy); b
[1] 0.5004332
> a <- mean(y) - b * mean(x); a
[1] 6.636483

```

Das Ergebnis dieser „orthogonalen“ Schätzung $\hat{y}_{ortho} = 6,637 + 0,5004x$ ist in Abbildung 3.15 dargestellt.

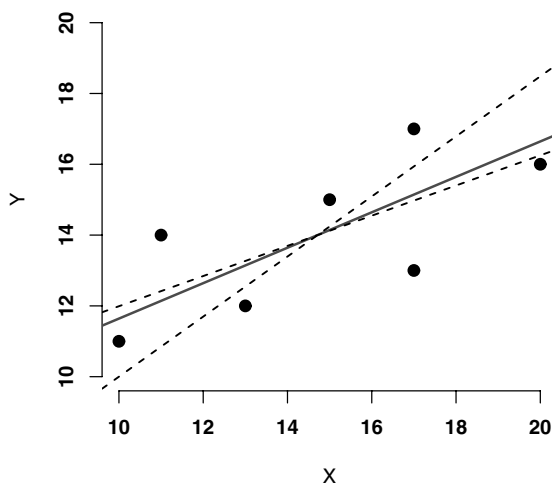


Abb. 3.15. Regression nach der Methode der orthogonal kleinsten Quadrate

3.6.8 Robuste lineare Regression

Die OLS-Schätzung aus dem vorangehenden Abschnitt wird erheblich durch extreme Beobachtungen (Ausreißer) beeinflusst, das bedeutet sie sind empfindlich gegenüber „ungewöhnlichen“ / „atypischen“ Einzelwerten. Eine robuste Schätzung sollte effizient und unverzerrt (ohne systematischen Fehler) sein und Abweichungen von der Modellannahme, hier bezüglich der Linearität, sollten das Verfahren nicht grundsätzlich in Frage stellen. Der Ansatz robuster Schätzverfahren beruht auf einer verallgemeinerten Betrachtung der Residuen mittels der **Zielfunktion** $\rho(\varepsilon)$:

$$\sum_{i=1}^n \rho(\varepsilon_i) \rightarrow \min$$

$$\rho_{OLS}(\varepsilon_i) = (y_i - \hat{y}_i)^2 \quad \text{für die OLS-Schätzung}$$

(3.53)

Bei der **LAD-Methode (least absolute deviation)** werden die Schätzwerte für a und b so bestimmt, dass die Summe der Beträge der Abweichungen minimal wird:

$$\rho_{LAD}(\varepsilon_i) = |y_i - \hat{y}_i| \quad \text{für die LAD-Schätzung}$$

(3.54)

Der Einfluss hinsichtlich der resultierenden Abweichungen (Fehler), den einzelne Beobachtungen auf die Schätzung der Regressionsparameter a und b haben, wird durch die 1. Ableitung der Funktion ρ (nach dem Fehler), die so genannte **Einflussfunktion** ρ' , bemessen:

$$\sum_{i=1}^n \rho(\varepsilon_i) = \sum_{i=1}^n (y_i - \hat{y}_i)$$

Einflussfunktion:
$$\sum_{i=1}^n \rho'(\varepsilon)x_i = 0$$

Für die Lösung dieses Problems werden iterative Schätzverfahren eingesetzt. In **R** steht für die LAD-Schätzung die Funktion **rq()** in der **library(quantreg)** [Koe05] (quantile regression) zur Verfügung, deren Verwendung an einem Beispiel gezeigt werden soll.

Beispiel: Der mittlere Preis (hier in cent per pound), den Fischer bei der Anlandung verschiedener Fischarten und Meeresfrüchte in den Jahren 1970 und 1980 in einer westenglischen Hafenstadt erhielten, zeigt erhebliche Schwankungen. Wie kann trotz der Ausreißer und Extremwerte auf einen „linearen Trend“ geschlossen werden?



Tabelle 3.12. Mittlere Preise bei der Anlandung von Meeresfrüchten in den Jahren 1970 und 1980

Sorte	Preis 1970	Preis 1980
COD (Kabeljau)	13.1	27.3
FLOUNDER (Flunder)	15.3	42.4
HADDOCK (Schellfisch)	25.8	38.7
OCEAN PERCH (Barsch)	4.9	123.1
SALMON CHINOOK (Lachs)	55.4	166.3
TUNA, ALBACORE (Thunfisch)	26.7	80.1
CLAMS, SOFT-SHELLED (Venusmuscheln)	47.5	150.7
CLAMS, BLUE HARD-SHELLED (Venusmuscheln)	6.6	20.3
LOBSTERS, AMERICAN (Hummer)	94.7	159.7
OYSTERS, EASTERN (Austern)	61.1	131.3
SEA SCALLOPS (Kammmuscheln)	135.6	404.2
SHRIMP (Garnelen)	47.6	149.0

$$\rho'_{OLS}(\varepsilon) = 2\varepsilon \quad \text{proportional zur Abweichung}$$

$$\rho'_{LAD}(\varepsilon) = \begin{cases} +1 & \text{für } \varepsilon_i > 0 \\ 0 & \text{für } \varepsilon_i = 0 \\ -1 & \text{für } \varepsilon_i < 0 \end{cases} \quad (3.55)$$

Die LAD-Schätzung ist sehr robust gegen Ausreißer und Extremwerte, allerdings wenig effizient im Vergleich zur OLS-Schätzung hinsichtlich der Stichprobenvarianz. Einen Kompromiss bieten die sogenannten **M-Schätzer**. Sie verfolgen ebenfalls den Ansatz, dass die 1. Ableitung der Zielfunktion durch geeignete Schranken begrenzt wird. Ein Beispiel ist die **M-Schätzung nach Huber**:

$$\rho'_{HUBER}(\varepsilon) = \begin{cases} +k & \text{für } \varepsilon_i > k \\ \varepsilon & \text{für } -k \leq \varepsilon_i \leq k \\ -k & \text{für } \varepsilon_i < -k \end{cases} \quad (3.56)$$

Die Konstante k , mit der hier die Begrenzung erfolgt, orientiert sich an der Streuung in den beobachteten Werten der Y-Variablen und wird mit Bezug auf die Median-Deviation in dem Ansatz nach Huber mit $k = 1,345$ festgelegt. Für die M-Schätzung nach Huber steht im Programm **R** die Funktion **rlm()** in der **library(MASS)** (robust fitting of linear model) [VB02] zur Verfügung. Ein Vergleich der verschiedenen Ansätze der Schätzung der Parameter a und b in einer linearen Regression wird an den Daten der Preise für Meeresfrüchte gezeigt.

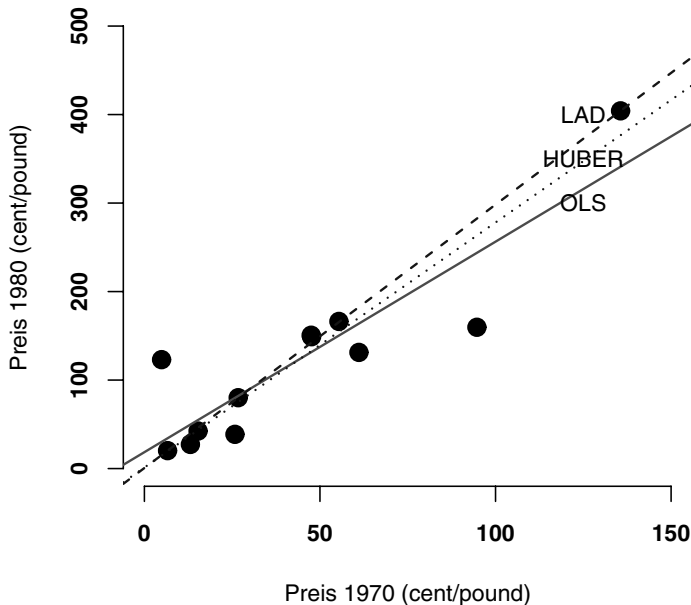


Abb. 3.16. Entwicklung der Preise für Fische und Meeresfrüchte bei der Anlandung in den Jahren 1970 und 1980; Schätzung der linearen Regression mit den Verfahren OLS, LAD und HUBER

```
> library(quantreg)
> library(MASS)
> p.1970 <- c(13.1,15.3,25.8, 4.9, 55.4,26.7, 47.5, 6.6, 94.7, 61.1,135.6, 47.6)
> p.1980 <- c(27.3,42.4,38.7,123.1,166.3,80.1,150.7,20.3,159.7,131.3,404.2,149.0)
> ols.regr <- lm(p.1980 ~ p.1970); ols.regr
```

```

Call: lm(formula = p.1980 ~ p.1970)

Coefficients:
(Intercept)      p.1970
      18.525         2.378

> lav.regr <- rq(p.1980 ~ p.1970, tau=0.5); lav.regr
Call: rq(formula = p.1980 ~ p.1970, tau = 0.5)

Coefficients:
(Intercept)      p.1970
      0.6374656      2.9761249

> huber.regr <- rlm(p.1980 ~ p.1970); huber.regr
Call: rlm(formula = p.1980 ~ p.1970)
Converged in 12 iterations

Coefficients:
(Intercept)      p.1970
      1.186329      2.768727

```

Die OLS-Schätzung ist empfindlich gegenüber Ausreißern. Im Beispiel ist das insbesondere der Preis für den Barsch, der im OLS-Ansatz in der Regression auf einen „mittleren Preisanstieg“ des 2,4fachen führt. Die Schätzung nach der LAD-Methode führt auf das 3fache, nach der Methode von Huber auf das 2,8fache.

3.7 Nichtlineare Regression

In vielen Fällen zeigt eine graphische Darstellung, dass die interessierende Beziehung nicht durch eine Regressionsgerade beschrieben werden kann. In den Abbildungen 3.17 - 3.20 sind einige nichtlineare Funktionen dargestellt. Häufig entspricht eine **Gleichung zweiten Grades** (quadratische Gleichung) ausreichend genau den tatsächlichen Verhältnissen. Wir bedienen uns im folgenden wieder der Methode der kleinsten Quadrate.

Die **allgemeine Gleichung zweiten Grades** lautet:

$$y = a + bx + cx^2 \quad (3.57)$$

Die Konstanten a , b und c für die gesuchte Funktion zweiten Grades gewinnt man aus folgenden **Normalgleichungen**:

$$\begin{array}{l} \text{I } an + b \sum x + c \sum x^2 = \sum y \\ \text{II } a \sum x + b \sum x^2 + c \sum x^3 = \sum xy \\ \text{III } a \sum x^2 + b \sum x^3 + c \sum x^4 = \sum x^2 y \end{array} \quad (3.58)$$

Beispiel: Die Bestimmung der Konstanten einer Gleichung zweiten Grades anhand der Normalgleichungen soll für die Daten aus folgender Tabelle erfolgen:



x	y	xy	x^2	x^2y	x^3	x^4
1	4	4	1	4	1	1
2	1	2	4	4	8	16
3	3	9	9	27	27	81
4	5	20	16	80	64	256
5	6	30	25	150	125	625
15	19	65	55	265	225	979

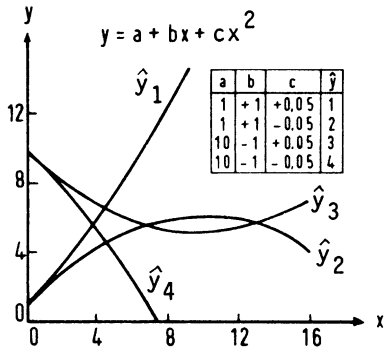


Abb. 3.17. Kurvenformen I

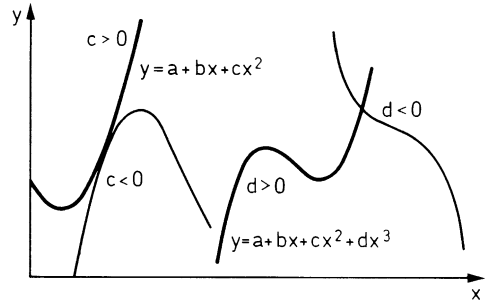


Abb. 3.19. Kurvenformen III

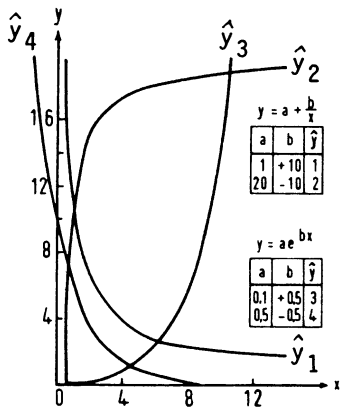


Abb. 3.18. Kurvenformen II

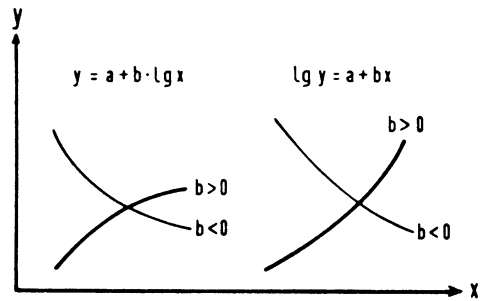


Abb. 3.20. Kurvenformen IV

Diese Werte werden in die Normalgleichungen eingesetzt:

I $5a + 15b + 55c = 19$

II $15a + 55b + 225c = 65$

III $55a + 225b + 979c = 265$

Aus I und II sowie aus III wird zuerst die Unbekannte a beseitigt:

$$\begin{array}{rcl} 5a + 15b + 55c = 19 & \cdot 3 & \\ 15a + 55b + 225c = 65 & & \end{array}$$

$$\begin{array}{rcl} 15a + 45b + 165c = 57 & & \\ 15a + 55b + 225c = 65 & & \end{array}$$

IV $10b + 60c = 8$

$$\begin{array}{rcl} 15a + 55b + 225c = 65 & \cdot 11 & \\ 55a + 225b + 979c = 265 & \cdot 3 & \end{array}$$

$$\begin{array}{rcl} 165a + 605b + 2475c = 715 & & \\ 165a + 675b + 2937c = 795 & & \end{array}$$

V $70b + 462c = 80$

Aus IV und V beseitigen wir b und erhalten c :

$$\begin{array}{rcl}
 70b + 462c & = & 80 \\
 10b + 60c & = & 8 \quad \cdot 7 \\
 \hline
 70b + 462c & = & 80 \\
 70b + 420c & = & 56 \\
 \hline
 42c & = & 24, \text{ d. h. } c = \frac{24}{42} = \frac{12}{21} = \frac{4}{7} (= 0,571)
 \end{array}$$

Durch Einsetzen von c in IV erhalten wir b :

$$\begin{array}{rcl}
 10b + 60c & = & 8 \\
 10b + \frac{60 \cdot 4}{7} & = & 8 \\
 70b + 240 & = & 56 \text{ und } b = \frac{56 - 240}{70} = -\frac{184}{70} = -\frac{92}{35} (= -2,629)
 \end{array}$$

Durch Einsetzen von b und c in I erhalten wir a :

$$\begin{array}{rcl}
 5a + 15 \cdot \left(-\frac{92}{35}\right) + 55 \cdot \frac{4}{7} & = & 19 \\
 5a - \frac{15 \cdot 92}{35} + \frac{55 \cdot 4 \cdot 5}{7 \cdot 5} & = & 19 \\
 35 \cdot 5a - 15 \cdot 92 + 55 \cdot 20 & = & 19 \cdot 35 \\
 175a - 1380 + 1100 & = & 665 \\
 175a - 280 & = & 665 \text{ und } a = \frac{945}{175} = \frac{189}{35} (= 5,400)
 \end{array}$$

Kontrolle: Einsetzen der Werte in die Normalgleichung I:

$$5 \cdot 5,400 - 15 \cdot 2,629 + 55 \cdot 0,571 = 27,000 - 39,435 + 31,405 = 18,970 \simeq 19,0$$

Die Gleichung zweiten Grades lautet:

$$\hat{y} = \frac{189}{35} - \frac{92}{35}x + \frac{4}{7}x^2 \simeq 5,400 - 2,629x + 0,5714x^2$$

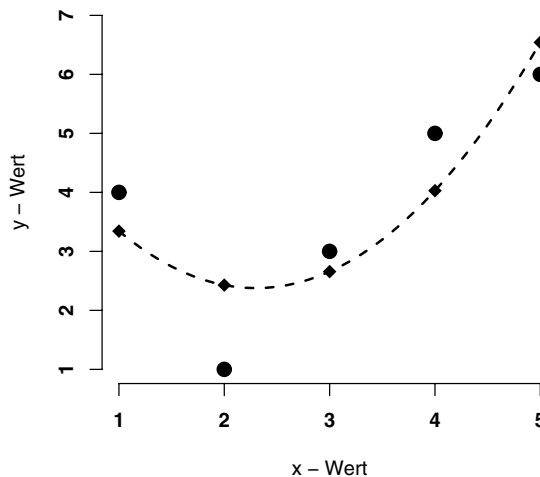


Abb. 3.21. Beispiel zur nichtlinearen Regression mit einer Gleichung zweiten Grades

In **R** erfolgt die Berechnung der Konstanten (Regressionskoeffizienten) durch die Funktion **nls()**. Neben der Spezifikation der entsprechenden Funktion müssen für die Koeffizienten (geeignete) Ausgangswerte angegeben werden. Das Ergebnis ist in Abbildung 3.21 dargestellt.

```
> x <- c(1,2,3,4,5)
> y <- c(4,1,3,5,6)
> nls( y ~ a + b*x + c*x^2, start = list(a = 1, b = 1, c = 1))
Nonlinear regression model
model: y ~ a + b * x + c * x^2
data: parent.frame()
      a      b      c
5.4000000 -2.6285714 0.5714286
residual sum-of-squares: 3.828571
```

Die **Güte der Anpassung** wird über die Abweichungen $y - \hat{y}$, Residuen genannt, beurteilt. Diese sind in dem Beispiel recht hoch, insbesondere für den Punkt ($x = 2$; $y = 1$).

x	y	$\hat{y}$	$y - \hat{y}$	$(y - \hat{y})^2$
1	4	3,343	0,657	0,432
2	1	2,429	-1,429	2,042
3	3	2,657	0,343	0,118
4	5	4,029	0,971	0,943
5	6	6,543	-0,543	0,295
	19	19,00	-0,001	3,83

In **R** können die berechneten Koeffizienten und nach der Gleichung „geschätzte“ y -Werte über spezielle Funktionen **coef()** und **predict()** angezeigt werden. Abbildung 3.21 zeigt die beobachteten und geschätzten Werte zu dem Beispiel.

```
> mod <- nls( y ~ a + b*x + c*x^2, start = list(a = 1, b = 1, c = 1))
> formula(mod); coef(mod)
y ~ a + b * x + c * x^2
      a      b      c
5.4000000 -2.6285714 0.5714286
> predict(mod, x)
[1] 3.342857 2.428571 2.657143 4.028571 6.542857
```

Drei Bemerkungen zum Modell der nichtlinearen Regression:

1. Stimmt das Modell, dann sind für jedes Regressionsmodell die **Residuen** $y - \hat{y}$ als beobachtete zufällige Fehler aufzufassen. Aufschluss hierüber verschafft die graphische Darstellung:
 - als Histogramm,
 - $(y_i - \hat{y}_i)$ (Ordinate) gegen i („horizontales Band“),
 - $(y_i - \hat{y}_i)$ gegen $\hat{y}_i$,
 - $(y_i - \hat{y}_i)$ gegen x_i („horizontales Band“) und
 - $(y_i - \hat{y}_i)$ gegen eine möglicherweise wichtige Variable, die bisher unberücksichtigt geblieben ist.
2. Als **nichtlineares Bestimmtheitsmaß** ($\hat{B}_{nl} = r_{nl}^2$) gilt allgemein $\hat{B}_{nl} = 1 - (A/Q_y)$ mit $A = \sum (y - \hat{y})^2$; für (3.57) elegant:

$$A = \sum y^2 - a \sum y - b \sum xy - c \sum x^2 y$$

$$Q_y = \sum y^2 - \frac{1}{n} (\sum y)^2$$

d. h. für unser Beispiel: $A = 87 - (189/35)19 + (92/35)65 - (4/7)265 = 87 - 102,6000 + 170,8571 - 151,4286 = 3,8285$; $Q_y = 87 - (19)^2/5 = 14,8000$; $\hat{B}_{nl} = 1 - (3,8285/14,8000)$

= 0,7413 und der nichtlineare Korrelationskoeffizient $r_{nl} = \sqrt{0,7413} = 0,8610$.

3. Zur Zusammenfassung kann man für (3.57) als **durchschnittliche Änderungsrate** die Steigung $b + 2cx$ der Kurve an der Stelle $(x_1 + x_n)/2$ angeben.

Ist der Zusammenhang durch eine **Exponentialfunktion** des Typs

$$y = ab^x \quad (3.59)$$

darstellbar, dann ergibt sich, wenn beide Seiten der Gleichung logarithmiert werden:

$$\lg y = \lg a + x \cdot \lg b \quad (3.60)$$

Die entsprechenden Normalgleichungen lauten:

$$\begin{array}{l} \text{I } n \cdot \lg a + (\sum x) \cdot \lg b = \sum \lg y \\ \text{II } (\sum x) \cdot \lg a + (\sum x^2) \cdot \lg b = \sum (x \cdot \lg y) \end{array} \quad (3.61)$$

Beispiel: Die Konstanten einer Exponentialfunktion sollen aus den Normalgleichungen für folgende Beispieldaten bestimmt werden.



x	y	$\lg y$	$x \lg y$	x^2
1	3	0,4771	0,4771	1
2	7	0,8451	1,6902	4
3	12	1,0892	3,2376	9
4	26	1,4150	5,6600	16
5	51	1,7076	8,5380	25
15	99	5,5240	19,6029	55

Die Summen werden in die Gleichung eingesetzt:

$$\begin{array}{l} \text{I } 5 \lg a + 15 \lg b = 5,5240 \quad \cdot 3 \\ \text{II } 15 \lg a + 55 \lg b = 19,6029 \end{array}$$

$$15 \lg a + 45 \lg b = 16,5720$$

$$15 \lg a + 55 \lg b = 19,6029$$

$$10 \lg b = 3,0309$$

$\lg b = 0,30309$, in I eingesetzt:

$$5 \lg a + 15 \cdot 0,30309 = 5,5240$$

$$5 \lg a + 4,54635 = 5,5240$$

$$5 \lg a = 0,9776$$

$$\lg a = 0,19554$$

Die entsprechenden entlogarithmierten Werte sind

$$a = 1,569 \quad b = 2,009 .$$

Die den obigen Werten angepasste Exponentialgleichung zur Schätzung von y aus x lautet somit $\hat{y} = 1,569 \cdot 2,009^x$. Die Lösung in **R** mit der Funktion **nls()** ist:

```

> x <- c(1,2,3,4,5)
> y <- c(3,7,12,26,51)
> nls( y ~ a*b^x, start = list(a = 1, b = 1))
Nonlinear regression model
  model: y ~ a * b^x
  data: parent.frame()
      a      b
1.602022 1.998596
residual sum-of-squares: 1.225082

```

3.7.1 Einige linearisierende Transformationen

Wenn die Form einer nicht-linearen Abhängigkeit zwischen zwei Variablen bekannt ist, dann ist es manchmal möglich, durch Transformation einer oder beider Variablen eine lineare Beziehung, eine gerade Linie zu erhalten.

Die soeben besprochene Gleichung $y = ab^x$ (3.59) hatten wir logarithmiert $\lg y = \lg a + x \lg b$; dies ist die Gleichung einer Geraden mit $\lg a =$ Achsenabschnitt (auf der Ordinate) und $\lg b =$ Regressionskoeffizient.

Rechnet man nicht mit den Normalgleichungen, dann sind die einzelnen Schritte:

1. Man transformiert alle y -Werte in $\lg y$ -Werte und rechnet mit den Logarithmen der beobachteten y -Werte ($y^* = \lg y$).
2. Die Regressionsgerade $\hat{y}^* = a^* + b^* x$ wird wie üblich geschätzt.
3. Durch Entlogarithmierung von $a^* = \lg a$, $b^* = \lg b$ erhält man die gesuchten Konstanten a und b der Originalgleichung $y = ab^x$.

Die Tabelle 3.13 (Natrella [Nat63]) zeigt einige Beziehungen zwischen x und y , die sich leicht linearisieren lassen: Die Tabelle weist auf die notwendigen Transformationen hin und gibt die Formeln für den Übergang von den Kennzahlen der geraden Linie zu den Konstanten der ursprünglichen Beziehung.

Diese linearisierenden Transformationen können auch dazu benutzt werden, um rein empirisch die Form eines Zusammenhangs zu ermitteln. Wir lesen die Tabelle 3.13 jetzt von den transformierten Werten zur Art des Zusammenhangs:

1. Trage y gegen $1/x$ in ein normales Koordinatensystem ein. Liegen die Punkte auf einer geraden Linie, so gilt die Beziehung $y = a + \frac{b}{x}$.
2. Trage $\frac{1}{y}$ gegen x in ein normales Koordinatensystem ein. Liegen die Punkte auf einer geraden Linie, so gilt die Beziehung $y = a/(b + x)$.
3. Trage in ein einfach logarithmisches Papier (halblogarithmisches Netz, Exponentialpapier) y (logarithmische Teilung) gegen x (arithmetische Teilung) ein. Liegen die Punkte auf einer geraden Linie, so gilt die Beziehung

$$y = ab^x \quad \text{oder} \quad y = ae^{bx}.$$

4. Trage in ein doppelt logarithmisches Papier (logarithmisches Netz, Potenzpapier) y gegen x ein. Liegen die Punkte auf einer geraden Linie, so gilt die Beziehung

$$y = ax^b.$$

Papiere, deren Koordinatenraster nicht wie bei gewöhnlichem Millimeterpapier gleichförmig (äquidistant, linear) ist, sondern bei denen die Koordinatenachsen beliebige Funktionsskalen tragen, nennt man **Funktionspapiere**. Neben dem Exponential- und dem Potenzpapier gibt es noch

Tabelle 3.13. Einige linearisierende Transformationen; verändert und erweitert nach Natrella, M. G.: Experimental Statistics, National Bureau of Standards Handbook 91, US. Government Printing Office, Washington 1963

Besteht eine Beziehung der Form	Trage die transformierten Variablen in das Koordinatensystem ein		Ermittle aus a' und b' die Konstanten a und b	
	$y' =$	$x' =$	$a' =$	$b' =$
$y = a + \frac{b}{x}$	y	$\frac{1}{x}$	a	b
$y = \frac{a}{b + x}$	$\frac{1}{y}$	x	$\frac{b}{a}$	$\frac{1}{a}$
$y = \frac{ax}{b + x}$	$\frac{1}{y}$	$\frac{1}{x}$	$\frac{1}{a}$	$\frac{b}{a}$
$y = \frac{x}{a + bx}$	$\frac{x}{y}$	x	a	b
$y = ab^x$	$\lg y$	x	$\lg a$	$\lg b$
$y = ax^b$	$\lg y$	$\lg x$	$\lg a$	b
$y = ae^{bx}$	$\ln y$	x	$\ln a$	b
$y = \frac{b}{ae^x}$	$\ln y$	$\frac{1}{x}$	$\ln a$	b
$y = a + bx^n \quad n \text{ bekannt}$	y	x^n	a	b
und schätze $\hat{y}' = a' + b'x'$				

andere wichtige Papiere, die komplizierte nichtlineare Funktionen linearisieren. Erwähnt sei das Sinuspapier, bei dem eine Achse gleichförmig, die andere nach einer Sinusleiter geteilt ist und in dem man Funktionen der Art

$ax + b \sin y + c = 0$

 durch die gerade Linie

$ax' + by' + c = 0$

darstellen kann ($x' = x \cdot e_x, y' = (\sin y)e_y$ mit $e_x = e_y = 1$).

Exponentialpapiere sind wichtig für das Studium radioaktiver und chemischer Zerfallsprozesse wie für die Analyse des Längenwachstums vieler Lebewesen. In der theoretischen Biologie und in der Physik spielen Potenzgesetze und damit auch Potenzpapiere eine gewisse Rolle.

In **R** können Probleme zu häufig verwendeten nichtlinearen Funktionen, z. B. für asymptotische Regression, exponentielle Modelle, Compartment-Modelle 1. Ordnung, logistische Modelle, Wachstumsmodelle (Gompertz), das Michaelis-Menten Modell und das Weibull Modell, mit speziellen Funktionen bearbeitet werden.

Beispiel: (Michaelis-Menten Gleichung) Das Modell beschreibt die Enzymkinetik (Geschwindigkeit einer enzymatischen Reaktion) in Abhängigkeit von der Konzentration (X) (Abbildung 3.22).



$$V = \frac{V_{max} \cdot X}{K_m + X}$$

(3.62)

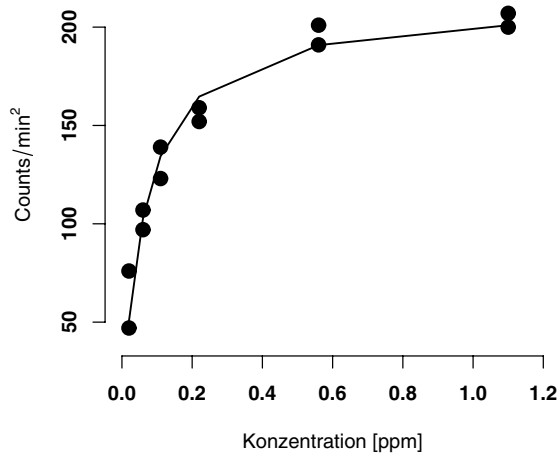


Abb. 3.22. Nichtlineare Regression am Beispiel der Michaelis-Menten Gleichung (3.62)

In dem Beispiel „**data(Puromycin)**” in **R** ist dies hinsichtlich der Aktivität (Anzahl von Counts einer radioaktiven Substanz) als Funktion der Substratkonzentration (*ppm*, parts per million) die initiale Rate („velocity”) der Reaktion (*Counts/min*²). Das Modell kann durch die Funktion **SSmicmen()** gekennzeichnet werden.

```
> conc <- c(0.02, 0.02, 0.06, 0.06, 0.11, 0.11, 0.22, 0.22, 0.56, 0.56, 1.10, 1.10)
> rate <- c(76, 47, 97, 107, 123, 139, 159, 152, 191, 201, 207, 200)
> nls(rate ~ SSmicmen(conc, Vm, K))
Nonlinear regression model
model: rate ~ SSmicmen(conc, Vm, K)
data: parent.frame()
      Vm      K
212.68370749 0.06412123
residual sum-of-squares: 1195.449
```

Aus den Daten wird für die Michaelis-Menten Gleichung eine maximale Geschwindigkeit $V_{max} = 212,68$ und die Konzentration für eine Halbsättigung ($V_{max}/2$) mit $K_m = 0,064$ bestimmt.

Linearisierung gekrümmter Punktwolken

Zur Linearisierung **kurvilinearer Zusammenhänge** wählt man für y (oder x) eine geeignete Transformation. Abbildung 3.23 (links) zeigt, dass dann, wenn y gegen x aufgetragen, eine Gerade resultiert, sich für x gegen y^3 aufwärts, für x gegen $(-1/y^2)$ abwärts gekrümmte Kurven ergeben. Erhält man also für eine Punktwolke x gegen y einen aufwärts gekrümmten Zusammenhang, so wird man, dem Krümmungsgrad entsprechend, die geeignete Transformation, z. B. $(-1/y)$ wählen, um eine Gerade zu erhalten; man hätte auch, anstatt y durch $(-1/y)$ zu ersetzen, x durch x^3 ersetzen können (vgl. Abbildung 3.23, rechts), um dann die Gerade $\hat{y} = a + bz$ mit $z = x^3$ zu erhalten.

So resultiert z. B. für Punktwolken (y gegen x) des Typs  durch das Auftragen von $\log y$ oder $\sqrt{y}$ gegen x (bzw. von y gegen x^2) häufig eine Gerade.

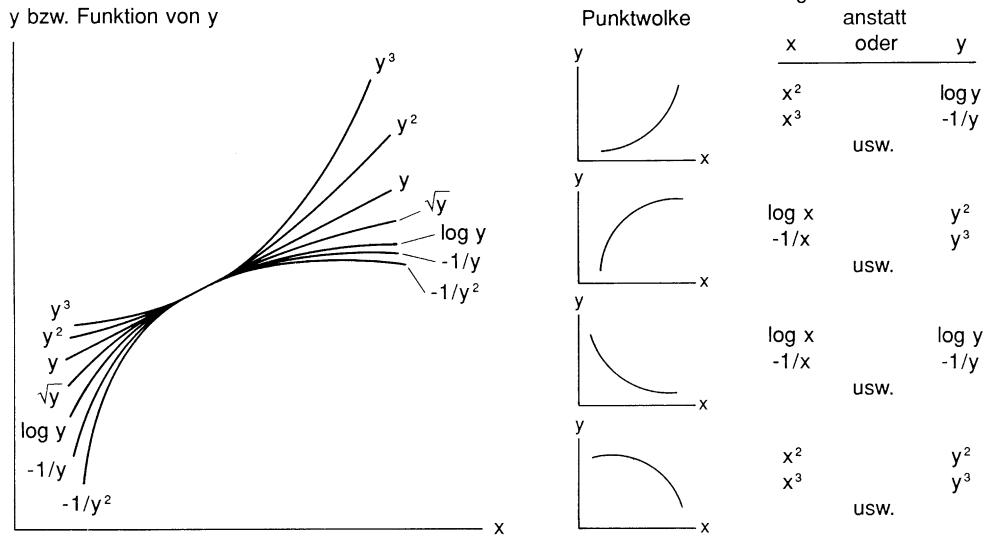


Abb. 3.23. Transformation der Werte y oder x zur Linearisierung von Punktwolken; rechts Begründungen der vier Krümmungsfälle mit den jeweiligen x - oder/und y -Transformationen

Wahrscheinlichkeiten

- Zufallsexperiment
- Begriff der Wahrscheinlichkeit
- Bedingte Wahrscheinlichkeit, stochastische Unabhängigkeit
- Bayessches Theorem
- Der diagnostische Test
- Maßzahlen in der Epidemiologie

Fünf Jahreszahlen zur Geschichte der Wahrscheinlichkeitsrechnung und der Statistik

1654 Der Chevalier de Méré fragt Blaise Pascal (1623–1662), warum es vorteilhaft sei, beim Würfelspiel auf das Erscheinen der Sechs in 4 Würfeln, aber nicht vorteilhaft sei, beim Spiel mit zwei Würfeln auf das Erscheinen der Doppelsechs in 24 Würfeln zu wetten. Hierüber korrespondiert Pascal mit Pierre de Fermat (1601–1665): Die beiden Wahrscheinlichkeiten sind 0,518 und 0,491. Die Frage nach den Aussagen, die aufgrund der erhaltenen Spielausgänge über die zugrundeliegenden Wahrscheinlichkeitsgesetze zu treffen sind, d. h. die Frage nach der Wahrscheinlichkeit für die Richtigkeit von Modellen oder Hypothesen untersucht Thomas Bayes (1702–1761).

1713/18 erscheinen die Lehrbücher der Wahrscheinlichkeitsrechnung von Jakob Bernoulli (1654–1705; *Ars Conjectandi*, opus posthumum, 1713) mit dem Begriff, Stochastik, der Binomialverteilung und dem Gesetz der großen Zahlen und Abraham de Moivre (1667–1754; *The Doctrine of Chances*, 1718) mit dem Begriff der stochastischen Unabhängigkeit und im Jahre **1733** mit dem Grenzübergang von der Binomial- zur **Normalverteilung**.

1812 Pierre Simon de Laplace (1749–1827): *Théorie Analytique des Probabilités*, die erste zusammenfassende Übersicht über die Wahrscheinlichkeitsrechnung.

1901 Gründung der auch heute noch richtungsweisenden Zeitschrift *Biometrika* als Kristallisationspunkt der angelsächsischen Schule der Statistik durch Karl Pearson (1857–1936), der mit Ronald Aylmer Fisher (1890–1962), dem u. a. Versuchsplanung und Varianzanalyse zu verdanken sind (1935 erscheint *The Design of Experiments*), die Mehrzahl der biometrischen Methoden entwickelt, die Jerzy Neyman (1894–1981) und Egon S. Pearson [(1895–1980) in den dreißiger Jahren durch den Vertrauensbereich (confidence interval) und die allgemeine Testtheorie erweitern. Nach der Axiomatisierung der Wahrscheinlichkeitsrechnung (1933) baut Andrej Nikolajewitsch Kolmogoroff (1903–1987) die von russischen Mathematikern geschaffene Theorie der stochastischen Prozesse aus.

1950 erscheint *Statistical Decision Functions* von Abraham Wald (1902–1950), in der die während des Zweiten Weltkrieges entwickelte, als stochastischer Prozess auffassbare Sequentialanalyse als Spezialfall der statistischen Entscheidungstheorie enthalten ist, die Richtlinien für das Verhalten in ungewissen Situationen liefert: Statistische Schlussweisen werden als Entscheidungsprobleme aufgefasst.

S.E. Fienberg [Fie92] gibt einen sehr instruktiven Überblick über die Geschichte der Statistik, indem er sieben einschlägige Bücher bespricht, die in den Jahren 1986/90 erschienen sind; zur Vertiefung nennt er weitere 50 Biographien, Monographien, Übersichten und Nachrufe.

Die Unsicherheit von Entscheidungen lässt sich durch die *Wahrscheinlichkeitstheorie* quantitativ erfassen. Anders ausgedrückt: Wahrscheinlichkeitstheoretische Begriffe gestatten die Gewinnung optimaler Entscheidungsverfahren. Wir haben uns daher zunächst dem Begriff Wahrscheinlichkeit zuzuwenden.

4.1 Zufallsexperiment, Ereignis

Ein Experiment mit nicht vorhersagbarem Ergebnis wird Zufallsexperiment genannt. Denken wir an den Münzwurf: hierzu gehört eine Vorschrift (1), diese ist beliebig oft wiederholbar (2), genau 2 Ergebnisse sind möglich [allgemein: „mindestens 2 . . . “] (3) sowie das Ergebnis ist nicht vorhersagbar (4).

Die Menge der möglichen Ergebnisse eines Zufallsexperimentes bilden den sog. **Ereignisraum** oder Stichprobenraum Ω (gr. Omega).

Besonders beachtet sei, dass der Ablauf eines Zufallsexperimentes im Gegensatz zu seinem Resultat streng geregelt ist.



Beim Werfen eines Würfels besteht der Ereignisraum aus 6 Punkten, die wir von 1 bis 6 nummerieren. In diesem Beispiel ist der Ereignisraum also endlich; aber schon wenn man beim Mensch-ärgere-Dich-nicht-Spiel keine Figur mehr im Feld hat und so lange würfeln muss, bis eine 6 auftritt, liegt ein Ereignisraum mit unendlich vielen Ereignissen vor, wenn man nämlich als Ereignisse die Anzahl der Würfe zählt, die bis zum Auftreten einer 6 gewürfelt werden müssen. Dann sind alle positiven ganzen Zahlen als Ereignis möglich.

Wird ein stetiges Merkmal betrachtet, wie die Körpergröße oder die Schlafdauer, dann können wir uns die Ereignisse (Messergebnisse) als Punkte auf der reellen Zahlenachse vorstellen. Der Ereignisraum umfasst dann z. B. alle Punkte eines Intervalls. Eine Teilmenge des Ereignisraumes heißt Ereignis und wird mit großen lateinischen Buchstaben, meist **E** oder **A**, bezeichnet. Betont sei, dass auch der gesamte Ereignisraum Ω als Ereignis aufgefasst wird. Dieses Ereignis heißt **das sichere Ereignis S**. Im Würfelbeispiel kann es interpretiert werden als $S = \{1, 2, 3, 4, 5, 6\}$, das Ereignis, irgendeine Augenzahl zu werfen. Einelementige Teilmengen wie die eben genannten sechs Ereignisse heißen **Elementarereignisse**. Seien E_1 und E_2 Ereignisse, dann interessiert man sich oft dafür, ob eine Messung in E_1 oder in E_2 liegt, es dürfen auch beide Fälle auftreten. Dieses Ereignis ist durch diejenige Teilmenge $E_1 \cup E_2$ des Ereignisraumes charakterisiert, die dadurch entsteht, dass man die Punkte, die in E_1 oder in E_2 oder in beiden liegen, zusammenlegt. Die „Oder-Verknüpfung“, die logische Summe $E_1 \cup E_2$, gelesen: „ E_1 vereinigt mit E_2 “. Diese Verknüpfung wird als **„Vereinigung“ (union)** bezeichnet. Sie besteht im Eintreffen von mindestens einem der beiden Ereignisse E_1 und E_2 . Das Symbol $\cup$ stammt von „Union“, Vereinigung (vgl. auch Kapitel [2.2], Seite 21).

Beispiel: $E_1 = \{2, 4\}$ $E_2 = \{1, 2\}$ $E_1 \cup E_2 = \{1, 2, 4\}$.

Diese Menge charakterisiert das **Ereignis: E_1 oder E_2 oder beide**.



Ganz entsprechend fragt man danach, ob eine Messung in E_1 und E_2 liegt. Dieses Ereignis ist durch diejenigen Punkte des Ereignisraumes charakterisiert, die sowohl in E_1 als auch in E_2 liegen. Diese Menge wird mit $E_1 \cap E_2$ bezeichnet: Die „Sowohl-als-auch-Verknüpfung“; das logische Produkt $E_1 \cap E_2$, gelesen: „ E_1 geschnitten mit E_2 “. Diese Verknüpfung wird als **„Durchschnitt“ [intersection]** bezeichnet. Sie besteht im Eintreffen sowohl des Ereignisses E_1 als auch des Ereignisses E_2 .

Beispiel: $E_1 \cap E_2 = \{2, 4\} \cap \{1, 2\} = \{2\}$.



Tritt der Fall auf, dass E_1 und E_2 keinen Punkt gemeinsam haben, dann sagt man, dass die Ereignisse E_1 und E_2 sich gegenseitig ausschließen. Die Operation $E_1 \cap E_2$ liefert die sog. „leere

Menge“, die keinen Punkt enthält. Der leeren Menge $\emptyset$ entspricht **das unmögliche Ereignis**. Da in der leeren Menge gar kein möglicher Messwert liegt, kann keine Messung nach $\emptyset$ fallen. Wenn E ein Ereignis ist, gibt es ein Ereignis $\bar{E}$, das aus denjenigen Punkten des Stichprobenraumes besteht, die nicht in E liegen. $\bar{E}$ (lies: „nicht E “) heißt das zu E entgegengesetzte, **komplementäre Ereignis** bezüglich S oder das logische Komplement. Sei z. B. E das Ereignis, mit einem Würfel eine gerade Zahl zu werfen, dann ist $E = \{2, 4, 6\}$ und $\bar{E} = \{1, 3, 5\}$.

$$E \cup \bar{E} = S \quad (\text{Sicheres Ereignis}) \quad (4.1)$$

$$E \cap \bar{E} = \emptyset \quad (\text{Unmögliches Ereignis}) \quad (4.2)$$

Die Diagramme in Abbildung 4.1 veranschaulichen die wichtigsten Zusammenhänge (vgl. auch Kapitel [2.2]):

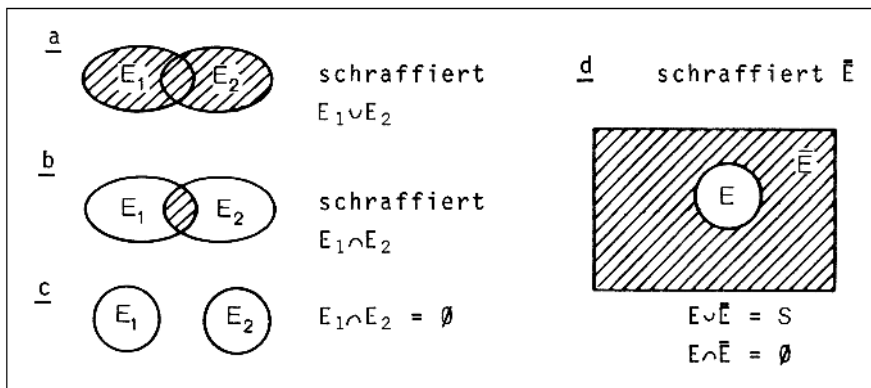


Abb. 4.1. Venn-Diagramme zur Darstellung wichtiger Beziehungen zwischen Ereignissen

Das **Komplement** (complement) zum Ereignis E bezüglich des sicheren Ereignisses S , also $\bar{E}$ besteht aus allen Ereignissen in S , die in E nicht enthalten sind. Anstatt $\bar{E}$ schreibt man auch E^c . Die **leere Menge** $\emptyset$ ist die Teilmenge jeder Menge. Das Komplement der leeren Menge ist der Ereignisraum Ω (d. h. $\bar{\emptyset} = \Omega$). Damit ist das $\emptyset$ entsprechende unmögliche Ereignis Komplement zum sicheren Ereignis.

Weitere Verknüpfungen

- 1)

$A \cup S = S \quad A \cap S = A$ $A \cup \emptyset = A \quad A \cap \emptyset = \emptyset$

Das „ $\cup$ “ entspricht einem „+“, das „ $\cap$ “ einem „ $\cdot$ “ oder „ $\times$ “.

- 2) Für „ B “ ist Teilmenge von A bzw. „ B “ ist in A enthalten schreibt man „ $B \subset A$ “.

Beachtet sei: Für $B \subset A$ gilt: (1) $B \cup A = A$.

(2) $B \cap A = B$.

Ein Ereignis $A \neq \emptyset$ heißt ein **Elementarereignis**, wenn es kein $B \neq \emptyset$ mit $B \subset A$ gibt.

- 3) Der Durchschnitt $A \cap \bar{B}$, A aber nicht B , heißt **Differenz** $A - B$ oder $A \setminus B$ zweier Ereignisse, auch Differenzmenge oder **Restmenge** genannt.

- 4)

$\overline{A \cup B} = \bar{A} \cap \bar{B}$ $\overline{A \cap B} = \bar{A} \cup \bar{B}$

De Morgan-Gesetze: Das Komplement der Vereinigung (des Durchschnitts) zweier Mengen ist gleich dem Durchschnitt (der Vereinigung) der Komplemente der einzelnen Mengen.

Übersicht 7. Die wichtigsten Verknüpfungen zwischen Ereignissen

Nr.	Von den Ereignissen A und B treten ein	Formale Beschreibung
1	beide	$A \cap B$
2	mindestens eines von beiden (A oder B oder beide)	$A \cup B$
3	genau eines von beiden (entweder A oder B, d. h. A und Nicht-B oder Nicht-A und B)	$(A \cap \bar{B}) \cup (\bar{A} \cap B)$
4	höchstens eines von beiden (keines oder A oder B)	$\overline{A \cap B} = \bar{A} \cup \bar{B}$
5	beide nicht (weder A noch B)	$\overline{A \cup B} = \bar{A} \cap \bar{B}$

4.2 Begriff der Wahrscheinlichkeit

- Definition nach Laplace
- Axiome nach Kolmogoroff

Ein **Ereignis** ist eine Teilmenge der Menge aller möglichen Resultate eines Zufallsexperimentes, die als Ereignisraum oder Stichprobenraum bezeichnet wird. Den Ereignissen werden Wahrscheinlichkeiten zugeordnet, die den Axiomen zu genügen haben.

**4.2.1 Definition nach Laplace**

Im täglichen Leben kennen wir verschiedene Arten von Aussagen, in denen das Wort „wahrscheinlich“ (Bedeutungsbereich: vermutlich bis todsicher) auftritt:

1. Die Wahrscheinlichkeit, mit einem intakten Würfel eine „4“ zu werfen, ist $1/6$.
2. Die Wahrscheinlichkeit für das Auftreten einer Zwillingsgeburt ist $1/86$.

Beide Sätze stehen zu dem **Begriff der relativen Häufigkeit** in einer engen Beziehung. Beim Würfeln nehmen wir an, dass im Mittel jede Seite gleich häufig auftritt, so dass wir erwarten, dass bei häufigen Wiederholungen die relative Häufigkeit, mit der eine 4 auftritt, gegen $1/6$ streben wird. Der 2. Satz ist aus einer relativen Häufigkeit entstanden. Man hat in den letzten Jahren beobachtet, dass die relative Häufigkeit der Zwillingsgeburten $1:86$ beträgt, so dass man annehmen kann, dass eine zukünftige Geburt mit der durch diese relative Häufigkeit der früheren Geburten gegebenen Wahrscheinlichkeit eine Zwillingsgeburt sein wird. Wir wollen im folgenden nur Wahrscheinlichkeiten betrachten, die sich als relative Häufigkeiten interpretieren lassen. Bei häufigen Wiederholungen zeigen diese relativen Häufigkeiten im allgemeinen eine auffallende Stabilität. Der Anteil der Linkshänder liegt für eine umfangreiche Bevölkerung bei etwa 8 von 100 Personen. Neuerdings werden auch deutlich höhere Werte angegeben.

Historische Grundlage dieses Wahrscheinlichkeitsbegriffes ist das bekannte dimensionslose Verhältnis, die Formel:

$$\frac{\text{Anzahl der günstigen Fälle}}{\text{Anzahl der möglichen Fälle}} = \frac{g}{m} = P \quad (4.3)$$

Jede Wahrscheinlichkeit (**probability**, von lat. probare [beglaubigen], kurz „ P “) ist damit eine Zahl zwischen Null und Eins:

$$0 \leq P \leq 1 \quad (4.4)$$

Ein unmögliches Ereignis hat die Wahrscheinlichkeit Null [„0%“], ein sicheres Ereignis die Wahrscheinlichkeit Eins [„100%“]. Günstige Fälle in (4.3) wären hierbei z. B.: (1) das Werfen einer „4“, (2) das Auftreten einer Zwillinggeburt bzw. (3) Linkshändigkeit bei einer Person – mögliche Fälle wären dann (1) das Werfen überhaupt einer Augenzahl (1 bis 6), (2) beliebige Geburten (Einzelkind, Zwillinge, . . .) bzw. (3) beliebige Personen.

Diese Definition der Wahrscheinlichkeit geht auf Jakob Bernoulli (1654–1705) und Pierre Simon de Laplace (1749–1827) zurück. Es wird stillschweigend vorausgesetzt, dass alle möglichen Fälle wie beim Würfelspiel **gleich-wahrscheinlich** sind.

Die Wahrscheinlichkeit, mit einem „idealen“ einwandfrei symmetrischen unverfälschten Würfel eine 4 zu werfen, beträgt $1/6$, da alle sechs Seiten die gleiche Chance haben aufzuliegen. Man erkennt den sechs Flächen eines symmetrischen Würfels gleiche Wahrscheinlichkeiten zu.

Die Definition der Wahrscheinlichkeit nach Bernoulli und de Laplace hat natürlich nur dann einen Sinn, wenn alle möglichen Fälle gleich wahrscheinlich, statistisch symmetrisch sind. Sie trifft nur für die üblichen Glücksspielgeräte (Münze, Würfel, Spielkarten und Roulette) zu. Bei ihnen liegt eine physikalische Symmetrie vor, die den Schluss auf die statistische Symmetrie zulässt. Die statistische Symmetrie ist aber für diese Wahrscheinlichkeitsdefinition unbedingt erforderlich. Es handelt sich hierbei um eine a-priori-Wahrscheinlichkeit, die auch mathematische Wahrscheinlichkeit genannt werden kann.

4.2.1.1 Relative Häufigkeit und Wahrscheinlichkeit

Für einen **unsymmetrischen Würfel** ist die Voraussetzung der physikalischen Symmetrie nicht mehr erfüllt und ein Schluss auf statistische Symmetrie nicht mehr möglich. Ein Wahrscheinlichkeitsverhältnis lässt sich nicht angeben. Hier hilft nur der Versuch mit einer großen Anzahl von Würfeln. Man erhält in diesem Fall unter Zuhilfenahme der **Erfahrung aus den Versuchen** die Wahrscheinlichkeit a posteriori oder die statistische Wahrscheinlichkeit: die relative Häufigkeit in einer sehr großen Versuchsserie. So ergebe sich für unseren selbstgebastelten Würfel (sgW) die Wahrscheinlichkeit, eine 4 zu werfen, als $P(4|\text{sgW}) = \frac{\cdot}{\cdot} = 0,154$ [und nicht $0,166 \dots = 1/6$]. Die Unterscheidung von mathematischer und statistischer Wahrscheinlichkeit betrifft lediglich die Art der Gewinnung des Wahrscheinlichkeitswertes.



Wahrscheinlichkeiten werden im allgemeinen als relative Häufigkeiten bestimmt und interpretiert.

Der Unterschied zwischen einer relativen Häufigkeit und einer Wahrscheinlichkeit besteht darin, dass sich eine relative Häufigkeit auf den erfassten Zustand einer Stichprobe in der mehr oder weniger unmittelbaren Vergangenheit bezieht, eine Wahrscheinlichkeit bezieht sich jedoch auf ein zukünftiges Ereignis.



Jede Wahrscheinlichkeit ist als **Erwartungswert** einer relativen Häufigkeit deren mathematische Abstraktion: aus Erfahrung wird Erwartung.

4.2.1.2 Chancen und Risiken

Wahrscheinlichkeiten werden häufig auch als **Wettchancen (odds)** durch ein Verhältnis angegeben, z. B. als 9 zu 12; diesem Verhältnis entspricht die Wahrscheinlichkeit $P = 9/(9+12) = 0,429$ (subjektive Wahrscheinlichkeit). Beiden Angaben gemeinsam ist der Zähler. Der Nenner ist unterschiedlich: Odds zählen hier nur die im Zähler nicht berücksichtigten Ereignisse, Wahrscheinlichkeiten zählen sämtliche Ereignisse. Der Wert $P = 0,429$ approximiert die Wahrscheinlichkeit, von 12 Fechtgängen drei aufeinanderfolgende zu gewinnen ($P = 1815/4096 = 0,443$; Hamlet: V, 2).

Wahrscheinlichkeiten und Odds

$$\text{Odds} = \frac{P}{1-P} \quad P = \frac{\text{Odds}}{\text{Odds} + 1}$$

Wahrscheinlichkeiten um $P \lesssim 0,1$ unterscheiden sich wenig von den entsprechenden Odds. Man rechnet meist weniger mit den Odds als mit deren Logarithmen, die symmetrisch um den Wert Null liegen (Wahrscheinlichkeiten liegen symmetrisch um $1/2$).

P	Odds	lg Odds
0	0	$-\infty$
0,01	0,01	-2
0,1	0,1111	-0,954
0,429	0,75	-0,125
0,5	1	0
0,9	9	0,954
0,99	99	2
1	∞	∞

4.2.2 Axiome nach Kolmogoroff

Ein Ereignis ist eine Teilmenge der Menge aller möglichen Resultate eines Zufallsexperimentes, die als Ereignisraum oder Stichprobenraum bezeichnet wird. Den Ereignissen werden Wahrscheinlichkeiten zugeordnet, die den folgenden Axiomen (Kolmogoroff 1903-1987) zu genügen haben.

I Jedem Ereignis ist eine Wahrscheinlichkeit, eine Zahl zwischen Null und Eins zugeordnet (**Nichtnegativitätsaxiom**):

$$0 \leq P(E) \leq 1 \quad (4.5)$$

II Das sichere Ereignis hat die Wahrscheinlichkeit Eins (**Normierungsaxiom**):

$$P(S) = 1 \quad (4.6)$$

Axiom II lässt sich auch schreiben: $\sum_{\text{alle } i} P(E_i) = 1$

III Die Wahrscheinlichkeit dafür, dass von mehreren, paarweise einander sich ausschließenden Ereignissen ($E_i \cap E_j = \emptyset$ für $i \neq j$; d. h. je zwei verschiedene Ereignisse schließen sich aus) eines eintritt, ist gleich der Summe der Wahrscheinlichkeiten der Ereignisse (**Additivitätsaxiom**):

$$P(E_1 \cup E_2 \cup \dots) = P(E_1) + P(E_2) + \dots \quad (4.7)$$

[Einfache Fassung: Für zwei Ereignisse mit $E_1 \cap E_2 = \emptyset$ gilt: $P(E_1 \cup E_2) = P(E_1) + P(E_2)$.]

Folgerungen aus den Axiomen

Das nicht realisierbare Ereignis ist charakterisiert durch die leere Menge $\emptyset$; für sie gilt $P(\emptyset) = 0$. Aus $P(E) = 0$ folgt jedoch nicht $E = \emptyset$. Entsprechend gilt zwar $P(S) = 1$, man darf jedoch

aus $P(E) = 1$ nicht auf $E = S$ schließen; denn $P(E) = 1$ gilt mitunter auch für Teilmengen E von S , wie wir bei der Definition der bedingten Wahrscheinlichkeit mit der auf einen neuen Stichprobenraum eingeschränkten Menge sehen werden.

Gilt $\begin{cases} P(E) = 0 \\ P(E) = 1 \end{cases}$, so heißt das Ereignis E $\begin{cases} \text{fast unmöglich} \\ \text{fast sicher} \end{cases}$

- Speziell für eine feste Zahl n einander ausschließender Ereignisse gilt nach dem **dritten Axiom**:

$$\boxed{P(E_1 \cup E_2 \cup \dots \cup E_n) = P(E_1) + P(E_2) + \dots + P(E_n)} \quad (4.8)$$

mit $E_i \cap E_j = \emptyset$ für alle $i \neq j$

- Für beliebige Ereignisse gilt die **Ungleichung nach Bonferroni**:

$$\boxed{P(E_1 \cup E_2 \cup \dots \cup E_n) \leq P(E_1) + P(E_2) + \dots + P(E_n)} \quad (4.9)$$

$$\begin{aligned} P(A \cup B) &= 1 - P(\overline{A \cup B}) = 1 - P(\bar{A} \cap \bar{B}) \\ \text{bzw.} \quad &= 1 - [P(\bar{A}) + P(\bar{B}) - P(\bar{A} \cup \bar{B})] \\ &\text{folgt aus (4.12)} \end{aligned}$$

$$\boxed{P(A \cup B) \geq 1 - [P(\bar{A}) + P(\bar{B})]} \quad (4.10)$$

- Die Wahrscheinlichkeiten für ein Ereignis E und für das Komplement $\bar{E}$ addieren sich zu 1:

$$1 = P(S) = P(E \cup \bar{E}) = P(E) + P(\bar{E}), \text{ d. h. } \boxed{P(E) = 1 - P(\bar{E})} \quad (4.11)$$

Beispiele zum Additivitätsaxiom:

- (1) Die Wahrscheinlichkeit, mit einem regelmäßigen Würfel eine 3 oder eine 4 zu werfen, beträgt: $1/6 + 1/6 = 1/3$. Bei einer größeren Serie von Würfeln ist also in 33% der Fälle mit einem Aufliegen einer 3 oder einer 4 zu rechnen.
- (2) Hat man bei einem bestimmten Pferderennen Wetten auf 3 Pferde abgeschlossen, dann ist die Wahrscheinlichkeit, zu gewinnen, gleich der Summe der Gewinnchancen für die drei Pferde (totes Rennen ausgeschlossen).

4.2.2.1 Additionssatz

Die Wahrscheinlichkeit, dass von zwei Ereignissen E_1 und E_2 , die sich nicht ausschließen [d. h. im Gegensatz zum Additivitätsaxiom (4.7)], mindestens eines eintritt, ist gegeben durch

$$\boxed{P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)} \quad (4.12)$$

Dieser **Additionssatz** lässt sich einfach nachvollziehen, wenn man eine Zerlegung der Ereignisse E_1 und E_2 in disjunkte Teilereignisse vornimmt (vgl. Abbildung 4.2):

$$\begin{aligned} E_1 &= (E_1 \cap \bar{E}_2) \cup (E_1 \cap E_2) \\ E_2 &= (\bar{E}_1 \cap E_2) \cup (E_1 \cap E_2) \end{aligned}$$

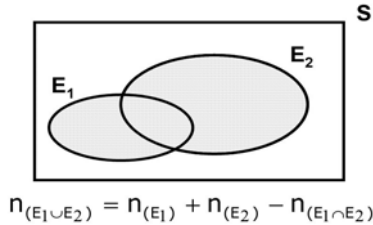


Abb. 4.2. Venn-Diagramm zum Additionssatz für „vereinbare“ Ereignisse

Um also die „Sowohl-als-auch-Wahrscheinlichkeit“ $P(E_1 \cap E_2)$ nicht doppelt zu berücksichtigen, muss diese von der Summe der Einzelwahrscheinlichkeiten subtrahiert werden.

Beispiele zum Additionssatz:

- (1) Entnimmt man einem Spiel von 52 Karten eine Karte und fragt nach der Wahrscheinlichkeit, dass diese ein As oder ein Karo ist – beide schließen sich nicht aus –, so ergibt sich, dass die Wahrscheinlichkeit für das Ziehen eines Asses $P(E_1) = 4/52$, für das Ziehen eines Karos $P(E_2) = 13/52$ und für das Ziehen eines Karo-Asses $P(E_1 \cap E_2) = 1/52$ beträgt: $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2) = 4/52 + 13/52 - 1/52 = 16/52 = 0,308$.
- (2) Die Wahrscheinlichkeit, daß es regnen wird, sei $P(E_1) = 0,70$, daß es schneien wird, sei $P(E_2) = 0,35$ und die, daß beide Ereignisse zugleich eintreten $P(E_1 \cap E_2) = 0,15$. Dann beträgt die Wahrscheinlichkeit für Regen, Schnee oder beides $P(E_1 \cup E_2) = P(E_1 \text{ oder } E_2 \text{ oder beide}) = 0,70 + 0,35 - 0,15 = 0,90$.
- (3) Bei einer Vorsorgeuntersuchung waren 15% der untersuchten Personen herzkrank und 10% lungenkrank; 80% hatten keine der beiden Krankheiten. Wie hoch war der Anteil der untersuchten Personen, die sowohl herz- als auch lungenkrank waren? Bekannt sind somit $P(H) = 0,15$, $P(L) = 0,1$ und $P(\overline{H \cup L}) = 0,8$. Die Wahrscheinlichkeit $P(H \cup L)$, eine der beiden Krankheiten oder auch beide (d. h. mindestens eine) zu haben, ist dann $P(H \cup L) = 1 - P(\overline{H \cup L}) = 1 - 0,8 = 0,2$; d. h. $P(H \cap L) = 0,15 + 0,1 - 0,2 = 0,05$.
- (4) In einem Patientengut betrage der Anteil der Kranken, die (a) herzkrank sind, 60%, (b) lungenkrank sind 50%; (c) beide Krankheiten haben 30% der Patienten. Wie groß ist der Anteil der Kranken, die genau eine der beiden Krankheiten (die also eine von beiden, aber nicht beide gemeinsam) haben. Bekannt sind somit $P(H) = 0,6$, $P(L) = 0,5$ und $P(H \cap L) = 0,3$. Gefragt wird nach der Differenz [„H“ oder „L“ oder „beide“] minus [„beide“], d. h. $P(H \cup L) - P(H \cap L) = P(H) + P(L) - P(H \cap L) - P(H \cap L) = P(H) + P(L) - 2P(H \cap L) = 0,6 + 0,5 - 2 \cdot 0,3 = 0,5$.

Eine **Verallgemeinerung des Additionssatzes** für drei einander nicht ausschließende Ereignisse führt zu

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) - P(A \cap B) \\ &\quad - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C) \end{aligned} \quad (4.13)$$

und die **Verallgemeinerung auf k Ereignisse**, d.h. die Wahrscheinlichkeit dafür, dass mindestens eines der Ereignisse E_i mit $i = 1, \dots, k$ eintritt, ist

$$P(E_1 \cup E_2 \cup \dots \cup E_k) = 1 - \{[1 - P(E_1)] \cdot [1 - P(E_2)] \cdot \dots \cdot [1 - P(E_k)]\} \quad (4.14)$$

Gilt zusätzlich $P(E_i) = p$, d.h. alle Ereignisse sind gleichwahrscheinlich, dann vereinfacht sich diese Formel zu:

$$P(E_1 \cup E_2 \cup \dots \cup E_k) = 1 - (1 - p)^k \quad (4.15)$$



Beispiel: Angenommen, Objekte werden nacheinander durch drei unabhängig voneinander wirk-same Instanzen kontrolliert, die (a) mit den Wahrscheinlichkeiten 90%, 95% und 99% bzw. (b) mit der jeweils konstanten Wahrscheinlichkeit von 95% die fehlerhaften Objekte ausschalten. Welche Methode ist wirksamer?

Für die Wahrscheinlichkeit, dass alle fehlerhaften Objekte ausgeschaltet werden, ergibt sich für (a)

$$P = 0,90 + 0,95 + 0,99 - (0,90 \cdot 0,95) - (0,90 \cdot 0,99) - (0,95 \cdot 0,99) + (0,90 \cdot 0,95 \cdot 0,99) = 0,99995 \quad \text{bzw. nach (4.14)}$$

$$P = 1 - [(1 - 0,90) \cdot (1 - 0,95) \cdot (1 - 0,99)] \\ = 1 - [0,10 \cdot 0,05 \cdot 0,01] = 0,99995.$$

Für (b) ergibt sich: $P = 1 - (1 - 0,95)^3 = 0,999875$ eine formal etwas kleinere Wahrscheinlichkeit.

4.3 Bedingte Wahrscheinlichkeiten, stochastische Unabhängigkeit



Beispiel: In zwei Werken werden Glühlampen hergestellt, und zwar 70% und 30% der Gesamtproduktion. Durchschnittlich weisen von je 100 Lampen des ersten Werkes 83 und von 100 Lampen des zweiten Werkes nur 63 die normgerechten Brennstunden auf. Im Mittel werden von je 100 Glühlampen, die an die Verbraucher gelangen, $77 (= 0,83 \cdot 70 + 0,63 \cdot 30)$ normgerecht sein, d. h. die Wahrscheinlichkeit, eine Normallampe zu kaufen, wird gleich 0,77 sein. Angenommen, wir hätten erfahren, dass die Glühlampen eines bestimmten Geschäftes alle im ersten Werk hergestellt wurden, dann wird die Wahrscheinlichkeit, eine normgerechte Lampe zu kaufen $83/100 = 0,83$ betragen. Die unbedingte Wahrscheinlichkeit des Kaufs einer Normallampe beträgt 0,77; die **bedingte Wahrscheinlichkeit** – Bedingung: im ersten Werk produziert – beträgt 0,83. Das lässt sich schreiben: $P(\text{Kauf einer Normallampe}) = 0,77$ sowie $P(\text{Kauf einer Normallampe} \mid \text{im ersten Werk produziert}) = 0,83$.

4.3.1 Bedingte Wahrscheinlichkeit

Da man von der Wahrscheinlichkeit irgendeines Ereignisses nur unter genau bestimmten Voraussetzungen sprechen kann, ist jede Wahrscheinlichkeit eine bedingte Wahrscheinlichkeit. Eine unbedingte Wahrscheinlichkeit kann im eigentlichen Sinne des Wortes nicht existieren.

Unter der bedingten Wahrscheinlichkeit des Ereignisses E_2 , d. h. unter der Bedingung oder Voraussetzung, dass das Ereignis E_1 schon eingetreten ist (geschrieben $P(E_2|E_1)$), verstehen wir die Wahrscheinlichkeit

$$P(E_2|E_1) = \frac{P(E_1 \cap E_2)}{P(E_1)} \quad (4.16)$$

die natürlich nur für $P(E_1) \neq 0$ definiert ist; analog gilt

$$P(E_1|E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)} \quad (4.17)$$

für $P(E_2) \neq 0$.

4.3.1.1 Multiplikationssatz

Aus der Definition der Bedingten Wahrscheinlichkeit ergibt sich der Multiplikationssatz der Wahrscheinlichkeitsrechnung für das gleichzeitige Eintreffen zweier beliebiger (unabhängiger oder abhängiger) Ereignisse E_1 und E_2 :

$$P(E_1 \cap E_2) = P(E_1) \cdot P(E_2|E_1) = P(E_2) \cdot P(E_1|E_2) = P(E_2 \cap E_1) \quad (4.18)$$

$$P(E_1 \cap E_2 \cap E_3) = P(E_1)P(E_2|E_1)P(E_3|E_1 \cap E_2)$$

Der **Multiplikationssatz für beliebige Ereignisse** $E_1, E_2, \dots, E_n$ (Pfadregel) lautet:

$$P(E_1 \cap E_2 \cap \dots \cap E_n) = P(E_1) \cdot P(E_2|E_1)P(E_3|E_1 \cap E_2) \cdot \dots \cdot P(E_n|E_1 \cap E_2 \cap \dots \cap E_{n-1}) \quad (4.19)$$

Einige Spezialfälle bedingter Wahrscheinlichkeiten mit Bezug zum sicheren Ereignis S :

$$(1) P(B|S) = P(B \cap S)/P(S) = P(B)/P(S) = P(B)$$

$$(2) P(S|B) = P(S \cap B)/P(B) = P(B)/P(B) = 1$$

$$(3) P(B|B) = P(B \cap B)/P(B) = P(B)/P(B) = 1$$

Weitere Regeln für den Umgang mit bedingten Wahrscheinlichkeiten sind:

$$\begin{aligned} P(B|A) + P(\bar{B}|A) &= \frac{P(B \cap A)}{P(A)} + \frac{P(\bar{B} \cap A)}{P(A)} \\ &= \frac{P(B \cap A) + P(\bar{B} \cap A)}{P(A)} = \frac{P(A)}{P(A)} = 1 \end{aligned}$$

$$P(C|A \cap B) = \frac{P(A \cap B \cap C)}{P(A \cap B)}$$

$$\begin{aligned} P(C|B \cap A)P(B|A)P(A) &= \frac{P(A \cap B \cap C)}{P(A \cap B)} \cdot \frac{P(A \cap B)}{P(A)} \cdot P(A) \\ &= P(A \cap B \cap C) \end{aligned}$$

Beispiele zum Multiplikationssatz:



- (1) Bei einer Wahlbeteiligung von 60% entscheiden sich 70% der Wählenden für eine bestimmte Alternative. Wie hoch ist der Anteil der Wahlberechtigten für diese Alternative?

$$P(B \cap A) = P(A) \cdot P(B|A) = 0,6 \cdot 0,7 = 0,42.$$

- (2) Von 10 gleichgroßen Äpfeln weisen 4 braune Flecken auf, 6 seien einwandfrei. Wie groß ist die Wahrscheinlichkeit, rein zufällig drei nicht einwandfreie Äpfel auszuwählen? Dass der erste entnommene Apfel braune Flecken aufweise, sei das Ereignis A ; entsprechend bezeichnen wir die anderen beiden Ereignisse mit B und C . Erfragt ist

$$P(A \cap B \cap C) = P(A) \cdot P(B|A) \cdot P(C|A \cap B) = \frac{4}{10} \cdot \frac{3}{9} \cdot \frac{2}{8} = \frac{4}{10} \cdot \frac{1}{3} \cdot \frac{1}{4} = \frac{1}{30} \approx 0,033.$$

- (3) Aus zwei Fußballmannschaften, K und B , werden nacheinander 3 Spieler „zufällig und ohne Zurücklegen“ aus dem Spiel genommen. Wie groß ist die Wahrscheinlichkeit, dass hierbei

3 Spieler der Mannschaft K betroffen sind. E_i sei das Ereignis, dass der i -te aus dem Spiel genommene Spieler zu K gehört, d. h.

$$P(E_1 \cap E_2 \cap E_3) = P(E_1) \cdot P(E_2|E_1) \cdot P(E_3|E_1 \cap E_2) = \frac{11}{22} \cdot \frac{10}{21} \cdot \frac{9}{20} = \frac{9}{84} = 0,1071.$$

- (4) Eierkuchen: Eine Eierschachtel enthalte 6 Eier, von denen 2 schlecht seien. Wie groß ist die Wahrscheinlichkeit, daß ein Eierkuchen aus drei Eiern, die der Schachtel entnommen werden, einwandfrei ist? Hierzu müssen alle 3 einwandfrei sein. Nacheinander gezogen seien die Wahrscheinlichkeiten, daß das erste, zweite bzw. dritte Ei einwandfrei ist $4/6$, $3/5$ bzw. $2/4$, d. h.

$$P = \frac{4}{6} \cdot \frac{3}{5} \cdot \frac{2}{4} = \frac{1}{5} = 0,2.$$

4.3.1.2 Risikomaße [wird durch Abschnitt 7.6.3 ergänzt]

Personen, die einem bestimmten Risikofaktor ausgesetzt sind, heißen **Exponierte**; diejenigen, bei denen dieser Risikofaktor nicht vorliegt, heißen **Nicht-Exponierte** (Kontrollpersonen). Bezeichnen wir die Wahrscheinlichkeit, bei Vorliegen des Risikofaktors R an der Krankheit K zu erkranken, mit $P(K|R)$, dann ergibt sich die folgende Tabelle 4.1, an der die Wahrscheinlichkeiten $P(\bar{K}|R)$ und $P(K|\bar{R})$ irritieren; die erstere ist bei kleinen Erkrankungswahrscheinlichkeiten ebenso wie $P(\bar{K}|\bar{R})$ fast Eins, die letztere fast Null.

Tabelle 4.1. Bedingte Wahrscheinlichkeiten zum relativen Risiko

	K	$\bar{K}$	
R	$P(K R)$	$P(\bar{K} R)$	$P(R)$
$\bar{R}$	$P(K \bar{R})$	$P(\bar{K} \bar{R})$	$P(\bar{R})$
	$P(K)$	$P(\bar{K})$	1

Mit einer bestimmten Krankheit K , einem Risikofaktor R und den Wahrscheinlichkeiten aus Tabelle 4.1 bezeichnet man die die folgenden Parameter

Delta: $\delta = P(K|R) - P(K|\bar{R})$ als **zuschreibbares Risiko**.

(4.20)

Psi: $\psi = \frac{P(K|R)}{P(K|\bar{R})}$ als **relatives Risiko**.

(4.21)

Omega: $\omega = \frac{P(K|R)}{P(\bar{K}|\bar{R})} \bigg/ \frac{P(K|\bar{R})}{P(\bar{K}|\bar{R})} = \frac{P(K|R) \cdot P(\bar{K}|\bar{R})}{P(\bar{K}|R) \cdot P(K|\bar{R})} = \frac{P(K|R)}{P(K|\bar{R})} \cdot \frac{P(\bar{K}|\bar{R})}{P(\bar{K}|R)}$
 als **Chancen-Verhältnis** (Odds Ratio)

(4.22)

Ein Faktor gilt als Risikofaktor nachgewiesen, wenn $\delta > 0$, $\psi > 1$ und $\omega > 1$.

4.3.1.3 Wahrscheinlichkeiten aus einer Sterbetafel

Eine Sterbetafel beschreibt die Verteilung der Lebensdauer von meist 10^5 Personen. Hierzu werden die Sterbewahrscheinlichkeiten der Toten des entsprechenden Geburtsjahrgangs während eines Jahres zugrunde gelegt. Es wird angenommen, dass die Daten eines Jahres für das ganze Leben der Personen gelten.

Tabelle 4.2. Aus dem Statistischen Jahrbuch 1992 der Bundesrepublik Deutschland auf S. 81 entnommene Sterbetafel-Daten (Absterbeordnung) [in stark gekürzter Form] für 1986/88; von 100 000 Lebendgeborenen erreichen das Alter x in vollendeten Jahren [in Klammern daneben Werte aus dem Stat. Jb. 2002 für 1997/99, S. 72]

x	Überlebende im Alter x			
	männlich		weiblich	
0	100000	(100000)	100000	(100000)
10	98764	(99293)	99073	(99439)
20	98284	(98870)	98843	(99238)
50	92471	(93557)	95842	(96609)
75	51273	(57185)	71775	(75865)
90	6406	(9903)	16903	(22359)

Beispiel: Aus der Tabelle 4.2 lassen sich z. B. folgende Wahrscheinlichkeiten entnehmen:

(1) Die Wahrscheinlichkeit für einen Mann, höchstens 49 Jahre alt zu werden:

$$P(X < 50|M) = \frac{100\,000 - 92\,471}{100\,000} = 0,0753$$

(2) Die Wahrscheinlichkeit für eine Frau, mindestens 90 Jahre alt zu werden:

$$P(X \geq 90|F) = \frac{16\,903}{100\,000} = 0,169.$$

(3) Die Wahrscheinlichkeit für einen 75-jährigen, noch mindestens 15 Jahre zu leben:

$$P(X \geq 15|x = 75|M) = \frac{6\,406}{51\,273} = 0,125.$$

4.3.1.4 Baumdiagramm und Pfadregeln

Ein Baumdiagramm besteht aus Pfaden (Teilstrecken) mit zugehörigen Wahrscheinlichkeiten und aus Ereignissen und Wahlmöglichkeiten an den Knoten (Verzweigungspunkten), die durch 3 Regeln verknüpft werden.

- (1) **Multiplikationspfadregel:** Bei einem mehrstufigen Zufallsversuch erhält man die Wahrscheinlichkeiten der einzelnen Ereignisse, indem man alle Wahrscheinlichkeiten längs des Pfades multipliziert und die Pfadwahrscheinlichkeit erhält.
- (2) **Additionspfadregel:** Gehören zu einem Ereignis eines Zufallsexperimentes mehrere Pfade, so erhält man die Wahrscheinlichkeit des Ereignisses durch Addition der zugehörigen einzelnen Pfadwahrscheinlichkeiten.
- (3) **Totalwahrscheinlichkeitsregel:** Die Summe der Teilwahrscheinlichkeiten an den Endknoten ist gleich Eins.

Beispiel: Mit den Daten aus Tabelle 4.2 lässt sich ein Baumdiagramm-Beispiel erstellen: Die Wahrscheinlichkeit, dass ein 50-jähriger ein Alter von 75 Jahren erreicht, ist durch $51\,273/92\,471 = 0,55448$ gegeben, für eine 50-Jährige beträgt sie dagegen $71\,775/95\,842 = 0,74889$. Wählt man jetzt rein zufällig einen 50-Jährigen und eine 50-Jährige aus und fragt nach (a) der Wahrscheinlichkeit, dass beide ein Alter von 75 Jahren erreichen, bzw. (b), dass wenigstens eine Person dieses Alter erreicht, so lassen sich beide Fragen anhand des Baumdiagramms (vgl. Abbildung 4.3 und die Details zu $P = 0,8881$ im nächsten Kapitel) beantworten.

Man erhält wegen der stochastischen Unabhängigkeit beider Ereignisse:

$$(a) \quad P(M \cap F) = P(M) \cdot P(F) = 0,5545 \cdot 0,7489 = 0,4153.$$



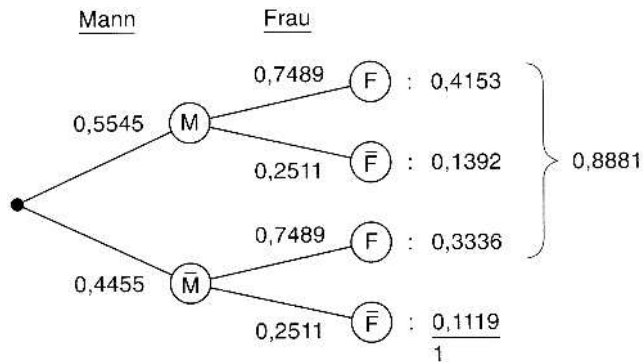


Abb. 4.3. Baumdiagramm für ein zweistufiges Zufallsexperiment zur Stochastischen Unabhängigkeit

- (b) $P(M \cap F) + P(M \cap \bar{F}) + P(\bar{M} \cap F) = 0,4153 + 0,1392 + 0,3336 = 0,8881$ bzw.
 $P = 1 - P(\bar{M} \cap \bar{F}) = 1 - 0,4455 \cdot 0,2511 = 1 - 0,1119 = 0,8881$ oder
 $P(M) + P(\bar{M} \cap F) = 0,5545 + 0,4455 \cdot 0,7489 = 0,8881$.

4.3.2 Stochastische Unabhängigkeit



Beispiel: Zwei Würfel, die in 2 getrennten Räumen geworfen werden, führen zu unabhängigen Resultaten. **Unabhängigkeit** von Ereignissen bedeutet, dass sie sich nicht gegenseitig beeinflussen oder gemeinsam von anderen Ereignissen beeinflusst werden.

Nehmen wir an, wir werfen mit einem einwandfreien Würfel mehrere Sechsen hintereinander, dann sinkt die Chance, weitere Sechsen zu werfen, nicht im geringsten! Sie bleibt für jeden Wurf konstant ($1/6$). Die Ergebnisse späterer Würfe müssen auf keinen Fall die der vorangegangenen ausgleichen. Vorausgesetzt wird natürlich ein regelmäßiger Spielwürfel und die Unabhängigkeit der einzelnen Würfe, d. h. kein vorheriger Wurf beeinflusst den nächsten; der Würfel wird beispielsweise durch den letzten Wurf nicht deformiert.

Zwei Ereignisse E_1 und E_2 nennt man **stochastisch unabhängig** („stochastisch“ bedeutet: mit Zufallsexperimenten und Wahrscheinlichkeiten zusammenhängend), wenn

$$P(E_2|E_1) = P(E_2) \quad \text{mit} \quad P(E_1) > 0 \quad (4.23)$$

Es gilt dann auch: $P(E_1|E_2) = P(E_1) \quad \text{mit} \quad P(E_2) > 0$.

Außerdem gilt

$$P(E_2|E_1) = P(E_2|\bar{E}_1) \quad \text{und} \quad P(E_1|E_2) = P(E_1|\bar{E}_2)$$

Mathematiker bevorzugen anstelle von (4.23) die **Produktdefinition** der Stochastischen Unabhängigkeit

$$P(E_1 \cap E_2) = P(E_1) \cdot P(E_2) \quad (4.24)$$

Insbesondere müssen hier die Fälle $P(E_1) = 0$ und $P(E_2) = 0$ nicht explizit ausgeschlossen werden.



Beispiel: Die Produktdefinition der Stochastischen Unabhängigkeit, (4.24), lässt sich anhand eines Zweiwürfelexperimentes anschaulich machen. Wirft man zwei einwandfreie (unterscheidbare) Würfel, so ergibt sich für jede Kombination

(1,1) (1,2) (1,3) ... (1,6)
 (2,1) (2,6)
 (3,1) (3,6)

 (6,1) (6,6)]

Die Wahrscheinlichkeit für die Augenzahlsumme 7 ergibt sich aus den Ereignissen $1 + 6, 6 + 1, 2 + 5, 5 + 2, 3 + 4, 4 + 3$ (6 von 36 Möglichkeiten) als $P(X = 7) = 6/36 = 1/6$ [bei nicht unterscheidbaren Würfeln [d. h. z. B. $(1,2) = (2,1)$] gibt es nur 21 Ereignisse].

die Wahrscheinlichkeit $\frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36}$:

Zwei Ereignisse A und B nennt man **voneinander stochastisch unabhängig**, wenn die Wahrscheinlichkeit für ihr gleichzeitiges Eintreten gleich dem Produkt ihrer Wahrscheinlichkeiten ist:

$$P(A \cap B) = P(A) \cdot P(B)$$

Symmetrie der Stochastischen Unabhängigkeit: A ist von B genau dann stochastisch unabhängig, wenn auch B von A stochastisch unabhängig ist.

Stochastische Abhängigkeit und kausale Abhängigkeit sind zweierlei: erstere ist symmetrisch („voneinander abhängig“), dagegen hat die kausale Abhängigkeit stets eine Richtung. Aus der kausalen folgt notwendigerweise die stochastische Abhängigkeit, die damit Hinweise geben kann.



Die Kennzeichnung der Eigenschaften eines Zufallsexperimentes lässt sich nun hinsichtlich des „nicht vorhersagbaren“ Ausgangs präzisieren: die Menge aller möglichen Ergebnisse ist bekannt, welches Ergebnis im konkreten Fall auftreten wird, kann nicht vorhergesagt werden, da es von den vorhergehenden Ergebnissen stochastisch unabhängig ist.

Beispiel: Für die Ereignisse A und B gelten die Wahrscheinlichkeiten $P(A) = 0,8, P(B) = 0,7$ und $P(A \cap B) = 0,6$. (1) Wir berechnen $P(A|B), P(B|A)$, wenden (2) den Multiplikationssatz an und prüfen (3) beide Ereignisse auf stochastische Unabhängigkeit:



$$(1) \quad P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0,6}{0,7} = \frac{6}{7}$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{0,6}{0,8} = \frac{6}{8}$$

$$(2) \quad P(A \cap B) = P(A) \cdot P(B|A) = P(B) \cdot P(A|B)$$

$$0,6 = \frac{6}{10} = \frac{8}{10} \cdot \frac{6}{8} = \frac{7}{10} \cdot \frac{6}{7}$$

$$(3) \quad P(A \cap B) = P(A) \cdot P(B)$$

$$\frac{6}{10} \neq \frac{8}{10} \cdot \frac{7}{10} = \frac{56}{100} \quad \text{bzw.}$$

$$P(A|B) = P(A) \quad \text{oder} \quad P(B|A) = P(B)$$

$$\frac{6}{7} \neq \frac{8}{10} \quad \frac{6}{8} \neq \frac{7}{10}$$

Beachte: Für $P(A) = a$ und $P(B) = b$

gilt $P(B|A) \geq (a + b - 1)/a$.

$$P(A) = 0,8, \quad P(B|A) > (0,8 + 0,7 - 1)/0,8 = 0,625$$

$$P(B) = 0,7 \quad [\text{vgl. } P(B|A) = 6/8 = 0,75]$$

Die Ereignisse A und B sind somit stochastisch abhängig. Das zeigt auch die folgende Wahrscheinlichkeitstabelle, deren „**Diagonalprodukte**“ (odds ratio) ungleich sind.

	B	$\bar{B}$	Σ	
A	0,60	0,20	0,8	Diagonalprodukte: $0,60 \cdot 0,10 \neq 0,20 \cdot 0,10$ bzw. $\frac{0,6 \cdot 0,1}{0,2 \cdot 0,1} \neq 1$
$\bar{A}$	0,10	0,10	0,2	
Σ	0,7	0,3	1	

Für $P(A \cap B) = P(A) \cdot P(B) = 0,8 \cdot 0,7 = 0,56$ sind die Ereignisse A und B stochastisch unabhängig.

	B	$\bar{B}$	Σ	
A	0,56	0,24	0,8	Diagonalprodukte: $0,56 \cdot 0,06 = 0,24 \cdot 0,14$ bzw. $\frac{0,56 \cdot 0,06}{0,24 \cdot 0,14} = 1$
$\bar{A}$	0,14	0,06	0,2	
Σ	0,7	0,3	1	

4.3.2.1 Stochastische Unabhängigkeit für drei und mehr Ereignisse

Für die stochastische Unabhängigkeit von drei Ereignissen lautet (4.24) dementsprechend

$$P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C) \quad (4.25)$$

Bei **vollständiger stochastischer Unabhängigkeit** kommen allerdings noch drei paarweise Gleichungen hinzu:

$$\begin{aligned} P(A \cap B) &= P(A) \cdot P(B) \\ P(A \cap C) &= P(A) \cdot P(C) \\ P(B \cap C) &= P(B) \cdot P(C) \end{aligned} \quad (4.26)$$

n Ereignisse $E_1, E_2, \dots, E_n$ mit $n \geq 2$ heißen stochastisch unabhängig, wenn die Wahrscheinlichkeit für alle möglichen Durchschnitte gleich dem Produkt der Einzelwahrscheinlichkeiten ist, d. h., wenn für alle Kombinationen von 2 oder mehr Ereignissen die Produktregeln erfüllt sind. Sonst heißen sie stochastisch abhängig. Um die **stochastische Unabhängigkeit für n Ereignisse** zu definieren, sind somit

$$2^n - n - 1 \quad \text{Gleichungen notwendig.}$$

Für das aus n voneinander stochastisch unabhängigen Experimenten mit den Ereignissen E_i ($i = 1, 2, \dots, n$) zusammengesetzte Ereignis gilt (4.27), vorausgesetzt $P(E_i) = P(E_i|E_j \dots E_q)$ für alle $i, j, \dots, q$ mit $i \neq j \neq \dots \neq q$.

$$P(E_1 \cap E_2 \cap \dots \cap E_n) = P(E_1) \cdot P(E_2) \cdot \dots \cdot P(E_n) \quad (4.27)$$



Beispiel: Eine Mutter von vier Jungen glaubt, dass ihr fünftes Kind ein Mädchen sein wird. Denn die Wahrscheinlichkeit, nacheinander 5 Knabengeburt zu erleben, ist [vgl. (4.27)] mit $(1/2)^5 = 1/2^5 = 1/32 = 0,03125$ sehr gering. Die bedingte Wahrscheinlichkeit für die Geburt eines Knabens bzw. eines Mädchens beim Vorliegen von 4 Knaben ist jedoch ebenso groß wie die unbedingte Wahrscheinlichkeit und damit unverändert $1/2$ und nicht $1/32$.



Beispiel: Unabhängige Ereignisse, die täglich mit der Wahrscheinlichkeit 0,99 bzw. 0,9999 eintreten, also als **praktisch „sicher“** gelten können, sind an allen Tagen eines Jahres mit der Wahrscheinlichkeit $0,99^{365} = 0,0255$ bzw. $0,9999^{365} = 0,9642$ zu erwarten: diese Werte 2,6% und 96,4% differieren beträchtlich.

4.3.2.2 Unvereinbarkeit und stochastische Unabhängigkeit

Für die stochastische Unabhängigkeit zweier Ereignisse A und B gilt: $P(A \cap B) = P(A) \cdot P(B)$.
Für die Unvereinbarkeit zweier Ereignisse A und B gilt: $A \cap B = \emptyset$, hieraus folgt $P(A \cap B) = 0$.

Die Unvereinbarkeit zweier Ereignisse schließt die stochastische Unabhängigkeit dieser Ereignisse aus: A und B sind unvereinbar bedeutet ja, dass A und B nicht gemeinsam auftreten können. Sie beeinflussen sich also gegenseitig. Tritt z. B. A ein, so tritt B nicht ein.

(1) Für die Ereignisse A und B mit $P(A) > 0$ und $P(B) > 0$ gilt:

Wenn A und B $\left\{ \begin{array}{l} \text{stoch. unabh.} \\ \text{unvereinbar} \end{array} \right\}$ sind, dann können sie nicht $\left\{ \begin{array}{l} \text{unvereinbar} \\ \text{stoch. unabh.} \end{array} \right\}$ sein.

(2) Sind A und B unvereinbar und stochastisch unabhängig, dann ist $P(A) = 0$ oder $P(B) = 0$.

4.3.2.3 Ungleichungen nach Bonferroni

Mit $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ und $P(A \cup B) \leq 1$ erhält man

$$\begin{aligned} P(A) + P(B) - P(A \cap B) &\leq 1 \quad (-1) \\ -P(A) - P(B) + P(A \cap B) &\geq -1 \\ P(A \cap B) &\geq P(A) + P(B) - 1 \end{aligned}$$

$$\boxed{P(A \cap B) \geq P(A) + P(B) - 1} \quad (4.28)$$

einen Spezialfall der Bonferroni-Ungleichung, die untere Grenze für die Wahrscheinlichkeit gemeinsamer Ereignisse ausgedrückt in Wahrscheinlichkeiten der Ereignisse. Angenommen, zwei Ereignisse A und B mit $P(A) = P(B) = 0,9$ treten gemeinsam auf, dann ist die untere Grenze für diese Wahrscheinlichkeit

$$P(A \cap B) \geq P(A) + P(B) - 1 = 0,9 + 0,9 - 1 = 0,8.$$

Für kleine Wahrscheinlichkeiten wird diese Bonferroni-Grenze negativ und damit wertlos. Die allgemeine Version der Bonferroni-Ungleichung lautet:

$$P(E_1 \cap E_2 \cap \dots \cap E_n) \geq 1 - \sum_{i=1}^n P(\bar{E}_i) \text{ bzw. } P(E_1 \cap E_2 \cap \dots \cap E_n) \geq \sum_{i=1}^n P(E_i) - (n-1)$$

Sind die E_i voneinander stochastisch unabhängig, so gilt das Gleichheitszeichen und der Term $(n-1)$ entfällt.

Die Ungleichungen von Bonferroni:

$$\boxed{P\left(\bigcup_{i=1}^n E_i\right) \leq \sum_{i=1}^n P(E_i) \quad \text{und} \quad P\left(\bigcap_{i=1}^n E_i\right) \geq 1 - \sum_{i=1}^n P(\bar{E}_i)} \quad (4.29)$$

4.3.2.4 Korrelation

Für $P(B|A) > P(B)$ sagt man: B und A sind positiv korreliert
 Für $P(B|A) < P(B)$ sagt man: B und A sind negativ korreliert

$$P(A \cap B) = P(A) \cdot P(B|A) = P(B) \cdot P(A|B) \quad (4.30)$$

Obere und untere Grenzen für $P(A \cap B)$ in den Spezialfällen I und II sind:

I Positive Korrelation

$$\boxed{P(A) \cdot P(B) \leq P(A \cap B) \leq \min[P(A), P(B)]} \quad (4.31)$$

II Negative Korrelation

$$\boxed{\max[0, P(A) + P(B) - 1] \leq P(A \cap B) \leq P(A) \cdot P(B)} \quad (4.32)$$

Als Korrelationskoeffizient dient

$$\text{Korr}(A, B) = [P(A \cap B) - P(A) \cdot P(B)] / \sqrt{P(A) \cdot P(\bar{A}) \cdot P(B) \cdot P(\bar{B})}$$

4.3.2.5 Dreizehn Beispiele zur stochastischen Unabhängigkeit



- (1) Wie gross ist die Wahrscheinlichkeit, mit drei regelmäßigen Würfeln zugleich dreimal die Sechs zu werfen? $P = \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{216}$. Das ist auch die Wahrscheinlichkeit, mit einem Würfel dreimal hintereinander die Sechs zu werfen. In einer langen Versuchsreihe werden im Durchschnitt nur einmal unter 216 Würfeln alle drei Würfel gleichzeitig eine Sechs zeigen.
- (2) Angenommen, zwei Lotterien versehen 10% bzw. 20% ihrer Lose mit Gewinnen. Mit welcher Wahrscheinlichkeit gewinnt man in beiden zugleich?
 $P = 0,10 \cdot 0,20 = 0,02 = 2\%$.
- (3) Für zwei stochastisch unabhängige Ereignisse A und B gelte $P(B|A) = 0,4$ und $P(B \cap A) = 0,2$. Wie groß sind die Wahrscheinlichkeiten $P(A)$ und $P(B)$?
 $P(B|A) = P(B) = 0,4$ und $P(B \cap A) = P(B) \cdot P(A) = 0,2$,
d. h. $P(A) = 0,2/0,4 = 0,5$.
- (4) Ein regelmäßiger Würfel wird viermal nacheinander geworfen. Wie groß ist die Wahrscheinlichkeit, mindestens eine Sechs zu erzielen? Ersetzt man „mindestens eine Sechs“ durch seine Negation „keine Sechs“, dann erhält man: die Wahrscheinlichkeit, mit einem Wurf keine Sechs zu werfen, ist $5/6$, mit 4 Würfeln beträgt sie $(5/6)^4$. Die Wahrscheinlichkeit, mit 4 Würfeln mindestens eine Sechs zu erhalten, ist $1 - (5/6)^4 = 0,518$, also etwas größer als $1/2$. Das verspricht Vorteile, wenn man mit Geduld, Kapital und gutem Würfel auf das Erscheinen einer Sechs in 4 Würfeln wettet. Entsprechend kann man für den Fall des Werfens zweier Würfel fragen, bei wie vielen Würfeln es sich lohne, auf das Erscheinen einer Doppelsechs zu wetten.

Die Wahrscheinlichkeit, in einem Spiel keine Doppelsechs zu erhalten, beträgt $35/36$, da 36 gleich wahrscheinliche Fälle $1 - 1, 1 - 2, \dots, 6 - 6$ vorhanden sind. Die Wahrscheinlichkeit, in n Würfeln mindestens eine Doppelsechs zu erhalten, ist dann wieder gegeben durch $P = 1 - (35/36)^n$. P soll $> 0,5$ sein, das heißt $(35/36)^n < 0,5$ und somit $n \lg(35/36) < \lg 0,5$ und hieraus $n > 24,6$. Wir setzen $n \lg(35/36) = \lg 0,5$ und erhalten

$$n = \frac{\lg 0,5}{\lg(35/36)} = \frac{0,6990 - 1}{\lg 35 - \lg 36} = \frac{9,6990 - 10}{1,5441 - 1,5563} = \frac{-0,3010}{-0,0122} = 24,6.$$

Man wird also auf das Erscheinen einer Doppelsechs in mindestens 25 Würfeln wetten; die Wahrscheinlichkeit, eine Doppelsechs zu werfen, ist dann größer als 50%.

- (5) Der **Chevalier de Méré** (1607–1684) erwarb eine größere Geldsumme mit dem Abschluss der Wette: bei viermaligem Würfeln wenigstens eine Sechs zu erhalten und verlor sie durch den Abschluss der folgenden: bei 24maligem Wurf mit zwei Würfeln mindestens eine Doppelsechs zu bekommen: $1 - (35/36)^{24} = 0,491 < 0,5 < 0,518 = 1 - (5/6)^4$.

Die Wahrscheinlichkeit, mit zwei Würfeln in n Würfen wenigstens eine Doppelsechs zu werfen, beträgt: $P(\geq 1 \text{ Doppelsechs}) = 1 - [1 - (\frac{1}{6})^2]^n$

$$\text{mit } \begin{cases} P = 0,4914 \\ P = 0,5055 \end{cases} \text{ für } \begin{cases} n = 24 \\ n = 25 \end{cases}$$

Der Briefwechsel zwischen **Pierre de Fermat** (1601–1665) und **Blaise Pascal** (1623–1662), der vom Chevalier de Méré um die Lösung der oben erwähnten Probleme gebeten worden war, begründete im Jahre 1654 die Wahrscheinlichkeitsrechnung, die später durch **Jakob Bernoulli** (1654–1705) zu einer mathematischen Theorie der Wahrscheinlichkeit ausgebaut worden ist.

- (6) Drei Geschütze mögen unabhängig voneinander auf dasselbe Flugzeug schießen. Jedes Geschütz habe die Wahrscheinlichkeit $1/10$, unter den gegebenen Umständen zu treffen. Wie groß ist die Wahrscheinlichkeit, dass das Flugzeug getroffen wird? Erfragt wird die Wahrscheinlichkeit, mindestens einen Treffer zu erzielen. Die Wahrscheinlichkeit, dass kein Flugzeug getroffen wird, beträgt $(9/10)^3$. Für die Wahrscheinlichkeit, dass mindestens ein Treffer erfolgt, ergibt sich dann

$$P = 1 - (9/10)^3 = 1 - \frac{729}{1000} = \frac{271}{1000} = 27,1\%$$

$$(\text{vgl. } P = 1 - [9/10]^{28} = 94,8\% \quad \text{oder} \quad P = 1 - [1/2]^4 = 93,7\%)$$

Trefferwahrscheinlichkeiten: Die Wahrscheinlichkeit P für wenigstens einen Erfolg (Treffer) in n unabhängigen Versuchen jeweils mit der Erfolgswahrscheinlichkeit p ist gegeben durch:

$$P = 1 - (1 - p)^n$$

$$\text{Beispiel: } \begin{matrix} p = 0,05 \\ n = 20 \end{matrix} P = 1 - (1 - 0,05)^{20} = 1 - 0,3585 = 0,642$$

In der Tabelle 4.3 sind einige Trefferwahrscheinlichkeiten aufgeführt:

Tabelle 4.3. Ausgewählte Trefferwahrscheinlichkeiten

p	0,01						0,02						0,05			
n	1	5	10	15	30	50	2	5	10	15	30	50	2	5	10	15
P	0,010	0,049	0,096	0,140	0,260	0,395	0,040	0,096	0,183	0,261	0,455	0,636	0,098	0,226	0,401	0,537
p	0,10				0,20				0,30		0,50		0,75		0,90	
n	2	5	10	15	5	10	15	30	5	10	5	10	2	5	2	3
P	0,190	0,410	0,651	0,794	0,672	0,893	0,965	0,999	0,832	0,972	0,969	0,999	0,937	0,999	0,990	0,999

Die relative Häufigkeit einer Krankheit K in der Bevölkerung betrage $p = 0,01$. Die Wahrscheinlichkeit, in Zufallsstichproben des Umfangs $n = 30$ wenigstens ein an K erkranktes Individuum zu finden, ist nach Tabelle 4.3 $P = 0,26$.

- (7) Es werden 24 Personen nach einem Zufallsverfahren ausgewählt. Wie gross ist die Wahrscheinlichkeit, dass mindestens 2 Personen am selben Tage Geburtstag haben? Sie beträgt $P = 0,538$. Angenommen, das Jahr habe 365 als Geburtstag gleichwahrscheinliche Tage.

Uns interessiere das Ereignis $\overline{E}$, „keine 2 (von insgesamt n) Personen haben denselben Geburtstag“. Für $\overline{E}$ gibt es dann 365^n mögliche und $365 \cdot 364 \cdot 363 \cdot \dots \cdot (365 - n + 1)$ günstige Fälle, d. h. die Wahrscheinlichkeit, dass in einer Gruppe von 24 Personen wenigstens 2 Personen am selben Tage Geburtstag haben, beträgt

$$P = P(E) = 1 - P(\overline{E}) = 1 - \frac{365 \cdot 364 \cdot \dots \cdot 342}{365^{24}} = 0,5383 \quad .$$

Zur schnellen Approximation dient

$$P(E) \geq 1 - e^{-k} \text{ mit } k = \frac{n(n-1)}{2 \cdot 365} \quad (4.33)$$

$$\text{z. B. } n = 24, \quad k = \frac{24 \cdot 23}{2 \cdot 365} = 0,7562; \quad 1/e^{0,7562} = 0,4694; \quad P(E) \geq 0,5306 \quad .$$

Mit anderen Worten, eine Wette, daß von 24 Personen mindestens 2 am selben Tag Geburtstag feiern, würde sich bei einer größeren Serie gleichartiger Wetten lohnen, da von 100 Wetten nur 46 verloren gingen, aber 54 gewonnen würden. Hierbei haben wir den 29. Februar ignoriert; außerdem ist unberücksichtigt geblieben, dass sich die Geburten in bestimmten Monaten häufen. Ersteres verringert die Wahrscheinlichkeit, letzteres erhöht sie: wäre jeder am 1.1. geboren, so resultierte $P(E) = 1$.

- (8) Eine Urne enthalte 15 rote und 5 schwarze Kugeln. E_1 bedeute Ziehen einer roten, E_2 Ziehen einer schwarzen Kugel. Wie groß ist die Wahrscheinlichkeit, in zwei aufeinanderfolgenden Ziehungen zuerst eine rote und dann eine schwarze Kugel zu erhalten?

Die Wahrscheinlichkeit, eine rote Kugel zu ziehen, ist $P(E_1) = 15/20 = 3/4$. Ohne die Kugel zurückzulegen, wird wieder gezogen. Die Wahrscheinlichkeit, eine schwarze Kugel zu ziehen, wenn rot gezogen war, ist $P(E_2|E_1) = 5/19 \simeq 0,26$. Die Wahrscheinlichkeit, in zwei Ziehungen ohne Zurücklegen eine rote und eine schwarze Kugel zu ziehen, ist $P(E_1) \cdot P(E_2|E_1) = 3/4 \cdot 5/19 = 15/76 \approx 0,20$.

- (9) Zehn Prozent einer Bevölkerung seien in einem gegebenen Zeitraum im Durchschnitt von einer Krankheit befallen ($P(E_1) = 0,10$). Von diesen Erkrankten mögen in der Regel 8% sterben ($P(E_2|E_1) = 0,08$). Dann ist die Wahrscheinlichkeit für dieses Ereignis $P = 0,08$ eine bedingte Wahrscheinlichkeit (Bedingung: Erkrankung). Die Wahrscheinlichkeit dafür, dass eine Person der betrachteten Bevölkerung in einem gegebenen Zeitabschnitt erkrankt und an dieser Krankheit stirbt, ist dann $P(E_1 \cap E_2) = P(E_1) \cdot P(E_2|E_1) = 0,1 \cdot 0,08 = 0,008 = 0,8\%$. Der Mediziner würde in diesem Falle sagen: Die **Morbidität** der Krankheit ist 10%, die **Letalität** 8% und die **Mortalität** 0,8%; es ist also Mortalität = Morbidität · Letalität (Maßzahlen der Epidemiologie und die Standardisierungen IDR und SMR werden im Abschnitt Epidemiologie [4.6] vorgestellt).

Von einer anderen Krankheit mögen 20% infiziert sein (E_1), davon mögen in einem bestimmten Zeitraum beispielsweise 30% erkranken (E_2), von denen schließlich 5% sterben (E_3). Dann ist die Mortalität gegeben durch $P(E_1 \cap E_2 \cap E_3) = P(E_1) \cdot P(E_2|E_1) \cdot P(E_3|E_2) = 0,20 \cdot 0,30 \cdot 0,05 = 0,003 = 0,3\%$. Aus **klinischen Statistiken** lassen sich ohne Bezug auf die Bevölkerung keine Aussagen über Morbiditätsverhältnisse gewinnen (auch nicht über ihre Altersabstufung), da das Einzugsgebiet der Klinik, die Personenzahl, die auch von diesem Ereignis hätte betroffen sein können – Personen unter Risiko – meist unbekannt ist.

- (10) Zwei voneinander unabhängige Operationen (A, B) werden jeweils mit den Wahrscheinlichkeiten 0,9 bzw. 0,8 überlebt; und zwar beide mit der Wahrscheinlichkeit $P(A \cap B) =$

$P(A) \cdot P(B) = 0,9 \cdot 0,8 = 0,72$. Mindestens eine Operation wird dann mit $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0,9 + 0,8 - 0,72 = 0,98$ überlebt und genau eine mit $P(A \cup B) - P(A \cap B) = P(A) + P(B) - 2P(A \cap B) = 0,9 + 0,8 - 2 \cdot 0,72 = 0,26$.

- (11) Zwei Ereignisse A und B seien stochastisch unabhängig voneinander. Sie treten mit den Wahrscheinlichkeiten $P(A) = 0,6$ und $P(B) = 0,8$ auf. Wie groß ist die Wahrscheinlichkeit, dass mindestens eines dieser beiden Ereignisse eintritt, d. h. dass entweder nur A oder nur B oder beide Ereignisse eintreten? Zwei Lösungsmöglichkeiten bieten sich an:

$$(1) P(A \cup B) = P(A) + P(B) - P(A \cap B) = P(A) + P(B) - P(A)P(B) \\ = 0,6 + 0,8 - 0,6 \cdot 0,8 = 0,92 ;$$

$$(2) P(A \cup B) = 1 - P(\text{„nicht beide“}) = 1 - P(\bar{A} \cap \bar{B}) \\ = 1 - P(\bar{A})P(\bar{B}) = 1 - 0,4 \cdot 0,2 = 0,92.$$

- (12) Zwei Kinder versuchen unabhängig voneinander, eine bestimmte Aufgabe zu lösen, jeweils mit einer Wahrscheinlichkeit von 0,5. Die Wahrscheinlichkeit dafür, dass wenigstens eines der beiden Kinder die Aufgabe löst, ist mit den Ereignissen

A: Kind I gelingt die Lösung, B: Kind II gelingt die Lösung,

C: wenigstens eines der beiden Kinder löst die Aufgabe und $C = A \cup B$

$P(C) = P(A \cup B) = P(A) + P(B) - P(A \cap B)$, für stochastisch unabhängige Ereignisse A und B gilt $P(A \cap B) = P(A) \cdot P(B)$, d.h. $P(C) = P(A) + P(B) - P(A) \cdot P(B) = 0,5 + 0,5 - 0,5 \cdot 0,5 = 0,75$.

- (13) In einer Gemeinde werden drei Brotsorten, A, B, C verzehrt, den Wahrscheinlichkeiten

$$P(A) = 0,50 ; \quad P(B) = 0,40 ; \quad P(C) = 0,30 ; \\ P(A \cap B) = 0,20 ; \quad P(A \cap C) = 0,15 ; \quad P(B \cap C) = 0,10 ; \\ P(A \cap B \cap C) = 0,05$$

entsprechend.

Die Wahrscheinlichkeiten, dass ein Bewohner dieser Gemeinde die Brotsorte(n):

1. A oder B oder C verzehrt,
2. keine dieser Brotsorten verzehrt,
3. nur A verzehrt,
4. weder B noch C verzehrt,
5. B und C nur gemeinsam verzehrt,
6. höchstens zwei der Brotsorten verzehrt, sind dann:

$$(1) P_1 = P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) \\ - P(B \cap C) + P(A \cap B \cap C) \\ = 0,50 + 0,40 + 0,30 - 0,20 - 0,15 - 0,10 + 0,05 = 0,80$$

$$(2) P_2 = 1 - P(A \cup B \cup C) = 1 - 0,80 = 0,20$$

$$(3) P_3 = P(A) - P(A \cap B) - P(A \cap C) + P(A \cap B \cap C) \\ = 0,50 - 0,20 - 0,15 + 0,05 = 0,60$$

$$(4) P_4 = 1 - P(B \cup C) = 1 - P(B) - P(C) + P(B \cap C) = 1 - 0,40 - 0,30 + 0,10 = 0,40$$

$$(5) P_5 = P(B \cap C) - P(A \cap B \cap C) = 0,10 - 0,05 = 0,05$$

$$(6) P_6 = 1 - P(A \cap B \cap C) = 1 - 0,05 = 0,95.$$

4.4 Bayessches Theorem

- Pfadregeln und das Bayessche Theorem
- Beispiele zum Bayesschen Theorem

Bezeichnen wir mit K eine spezielle Krankheit, mit F eine Frau und mit M einen Mann, dann ist die Erkrankungswahrscheinlichkeit für einen aus der Bevölkerung zufällig ausgewählten Erwachsenen:

$$P(K) = P(F) \cdot P(K|F) + P(M) \cdot P(K|M) \quad (4.34)$$

Es ist danach davon auszugehen, dass in der Bevölkerung der Anteil der Männer und Frauen nicht gleich ist und dass auch die Wahrscheinlichkeit für das Vorliegen der Krankheit bei Männern und Frauen unterschiedlich ist. Die Frage, mit welcher Wahrscheinlichkeit auf das Vorliegen der Krankheit geschlossen werden kann, wenn das Geschlecht bekannt ist, kann mit dem Bayesschen Theorem beantwortet werden.

Die totale Wahrscheinlichkeit:

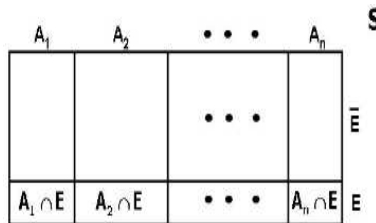


Abb. 4.4. Zerlegung der einer Ergebnismenge S in disjunkte Teilmengen E und $\bar{E}$ bzw. A_i ($i = 1, \dots, n$)

- Die A_i bilden eine disjunkte Zerlegung von S (vgl. Abb. 4.4); man spricht von einer **totalen Ereignisdisjunktion** der A_i .
- Die $A_i \cap E$ bilden eine disjunkte Zerlegung von E .
- Dann gilt $P(E) = \sum_i P(A_i \cap E)$ und unter Anwendung des Multiplikationssatzes erhält man die **totale Wahrscheinlichkeit des Ereignisses E**

$$P(E) = \sum_i P(A_i \cap E) = \sum_i P(A_i) \cdot P(E|A_i) \quad (4.35)$$

Angenommen, das Ereignis E kann nur unter n sich gegenseitig ausschließenden Umständen A_i eintreten, dann zeigt (4.35), wie sich die Totale Wahrscheinlichkeit des Ereignisses E als Summe von Produkten aus den Wahrscheinlichkeiten der verschiedenen Umstände und den entsprechenden bedingten Wahrscheinlichkeiten ergibt.

Uns interessiere der Umstand oder das Ereignis A_k , das in Verbindung mit dem Ereignis E auf-trete. Dann gilt die bedingte Wahrscheinlichkeit

$$P(A_k|E) = \frac{P(A_k \cap E)}{P(E)}$$

Wenden wir auf den Zähler den Multiplikationssatz an und auf den Nenner die „Totale Wahr-scheinlichkeit“, dann erhalten wir das **Bayessche Theorem**:

$$P(A_k|E) = \frac{P(A_k) \cdot P(E|A_k)}{\sum_{i=1}^n P(A_i) \cdot P(E|A_i)} \quad (4.36)$$

Die Ereignisse $A_1, A_2, \dots, A_n$ schließen sich gegenseitig aus, indem sie eine vollständige Zerlegung von S darstellen. Tritt nun eines dieser Ereignisse, sagen wir A_k , in Verbindung mit dem Ereignis E auf, das ebenfalls zu S gehört, wobei $P(E) > 0$ gilt, dann kann nach dem Bayesschen Theorem die Wahrscheinlichkeit $P(A_k|E)$ (nach 4.36) berechnet werden.

Mit $P(E \cap A) = P(E) \cdot P(A|E) = P(A) \cdot P(E|A)$ erhält man

$$P(A|E) = P(E|A) \frac{P(A)}{P(E)} = \frac{P(A) \cdot P(E|A)}{P(E)} \quad (4.37)$$

die so genannte **spezielle Bayes-Formel**; (4.36) heißt dagegen auch allgemeine Bayes-Formel.

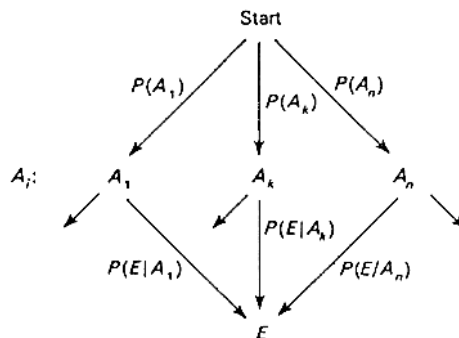
Der Schluss nach Bayes von einem Symptom S auf die Krankheit K (einfachster Fall: 2 Ereignisse K und $\bar{K}$), gegeben: $P(K)$, $P(S|K)$ und $P(S|\bar{K})$:

$$P(K|S) = \frac{P(K) \cdot P(S|K)}{P(K) \cdot P(S|K) + P(\bar{K}) \cdot P(S|\bar{K})} \quad (4.38)$$

Die Wahrscheinlichkeit $P(K)$ wird als „**a-priori**“ **Wahrscheinlichkeit** für K bezeichnet. $P(K|S)$ als die „**a-posteriori**“ **Wahrscheinlichkeit** für K [vgl. auch den Abschnitt zum diagnostischen Test]. Im Rückblick wird anhand von Wahrscheinlichkeiten eine bestimmte Situation eingeschätzt. Aus der Wahrscheinlichkeit an K zu erkranken und den Wahrscheinlichkeiten, das betreffende Symptom aufzuweisen – und zwar im Krankheitsfall und („lätigerweise“) auch im Nicht-Krankheitsfall – wird die Wahrscheinlichkeit berechnet, dass beim Vorliegen eines Symptoms auch die Krankheit vorhanden ist, wird vom Symptom auf die Krankheit geschlossen (wird aus Erfahrung gelernt).

4.4.1 Bayessches Theorem und Pfadregel

- (1) Die Wahrscheinlichkeit eines Pfades ergibt sich als Produkt der Einzelwahrscheinlichkeiten längs des Pfades.
- (2) Die Wahrscheinlichkeit eines Ereignisses ergibt sich als Summe der Wahrscheinlichkeiten aller Pfade, die zu dem Ereignis führen [(4.39)].



I. Die Wahrscheinlichkeit, vom Start nach E zu gelangen, ist

$$P(E) = \sum_{i=1}^n P(A_i) \cdot P(E|A_i) \quad (4.39)$$

II. Angenommen, ich gelange über A_k nach E , dann ist diese Wahrscheinlichkeit

$$P(A_k|E) = \frac{P(A_k) \cdot P(E|A_k)}{\sum_{i=1}^n P(A_i) \cdot P(E|A_i)} \quad (4.40)$$

4.4.2 Acht Beispiele zum Bayesschen Theorem



- (1) **Brückenbau nach erfolgreicher Bewerbung:** Drei Personen bewerben sich um ein öffentliches Amt. Eine Meinungsumfrage ergebe die individuellen Wahlchancen 0,25, 0,35 und 0,40. Die Chancen, dass die Drei nach ihrer Wahl einen Brückenbau durchsetzen, betragen 0,60, 0,90 und 0,80.

Wie groß ist die Wahrscheinlichkeit, dass die Bevölkerung nach der Wahl die Brücke erhält?

Die Wahlchancen sind: $P(A_1) = 0,25$

$$P(A_2) = 0,35$$

$$P(A_3) = 0,40$$

Die Chancen für den Brückenbau sind: $P(B|A_1) = 0,60$

$$P(B|A_2) = 0,90 \text{ und}$$

$$P(B|A_3) = 0,80;$$

dann ist die totale Wahrscheinlichkeit für den Brückenbau

$$P(B) = \sum_{i=1}^3 P(A_i)P(B|A_i) = 0,25 \cdot 0,60 + 0,35 \cdot 0,90 + 0,40 \cdot 0,80 = 0,785.$$

- (2) **Karotten aus drei Gärtnereien:** Ein Gemüsehändler erhalte Karotten aus drei Gärtnereien: 50% stamme aus A_1 , 30% aus A_2 und 20% aus A_3 . Der Händler weiß, daß A_1 1 % Ausschuß liefert, A_2 3% und A_3 4%. Wie viel Prozent Ausschuss sind zu erwarten?

Mit $P(A_1) = 0,5$, $P(A_2) = 0,3$ und $P(A_3) = 0,2$ und $P(E|A_1) = 0,01$, $P(E|A_2) = 0,03$ sowie $P(E|A_3) = 0,04$ ergibt sich insgesamt $P(E) = 0,5 \cdot 0,01 + 0,3 \cdot 0,03 + 0,2 \cdot 0,04 = 0,022$ ein Ausschussanteil von 2,2%.

- (3) **Krankheit K:** In einer bestimmten Grundgesamtheit mit 60% Frauen (F) und 40% Männern ($\bar{F}$) leiden 10% der Frauen und 3% der Männer an der Krankheit K . Wie groß ist die Wahrscheinlichkeit dafür, dass (a) eine zufällig ausgewählte Person an K leidet, (b) diese Person eine Frau bzw. (c) ein Mann ist?

$$P(F) = 0,6 \quad P(\bar{F}) = 0,4 \quad P(K|F) = 0,1 \quad P(K|\bar{F}) = 0,03$$

$$(a) \quad P(K) = P(F) \cdot P(K|F) + P(\bar{F}) \cdot P(K|\bar{F}) \\ = 0,6 \cdot 0,1 + 0,4 \cdot 0,03 = 0,06 + 0,012 = 0,072,$$

$$(b) \quad P(F|K) = \frac{P(F) \cdot P(K|F)}{P(K)} = \frac{0,06}{0,072} = 0,833 = 1 - P(\bar{F}|K)$$

$$(c) \quad P(\bar{F}|K) = \frac{P(\bar{F}) \cdot P(K|\bar{F})}{P(K)} = \frac{0,012}{0,072} = 0,167 = 1 - P(F|K)$$

- (4) **Terrorismus im Flugverkehr:** Auf dem Flughafen werden alle Passagiere vorsorglich kontrolliert. Ein Terrorist werde mit $P(F|T) = 0,98$ festgenommen, ein Nicht-Terrorist mit $P(F|\bar{T}) = 0,001$. Jeder hunderttausendste Flugpassagier sei ein Terrorist, $P(T) = 0,00001$. Wie groß ist die Wahrscheinlichkeit, dass eine Festnahme tatsächlich einen Terroristen erbringt?

$$\begin{aligned} P(T|F) &= \frac{P(T) \cdot P(F|T)}{P(T) \cdot P(F|T) + P(\bar{T}) \cdot P(F|\bar{T})} = \frac{0,00001 \cdot 0,98}{0,00001 \cdot 0,98 + 0,99999 \cdot 0,001} \\ &= \frac{98}{10098} = 0,0097 < 0,01 \end{aligned}$$

Trotz der Zuverlässigkeit der Kontrollen erfolgen somit über 99% aller Festnahmen zu Unrecht.

- (5) **Herkunft von Ausschussware I:** Zwei Maschinen A und B produzieren Teekessel desselben Typs; A liefert 60%, B 40% der Produktion. Der Anteil einwandfreier (E) Teekessel betrage für A 99%, für B 95%. Aus der Gesamtproduktion wird anhand einer Zufallszahl Z der Z -te Teekessel ausgewählt. Wie groß ist die Wahrscheinlichkeit, dass dieser Teekessel A bzw. B entstammt, wenn er (a) einwandfrei oder (b) defekt ist?
Gegeben sind somit $P(A) = 0,6$ und $P(B) = 0,4$ sowie $P(E|A) = 0,99$ und $P(E|B) = 0,95$ [und damit $P(\bar{E}|A) = 1 - 0,99 = 0,01$ und $P(\bar{E}|B) = 1 - 0,95 = 0,05$]. Gesucht werden (a) $P(A|E)$ und $P(B|E)$ sowie (b) $P(A|\bar{E})$ und $P(B|\bar{E})$.

$$\begin{aligned} \text{(a)} \quad P(A|E) &= \frac{P(A) \cdot P(E|A)}{P(A) \cdot P(E|A) + P(B) \cdot P(E|B)} = \frac{0,6 \cdot 0,99}{0,6 \cdot 0,99 + 0,4 \cdot 0,95} = 0,610 \\ P(B|E) &= 1 - P(A|E) = 1 - 0,610 = 0,390 \end{aligned}$$

$$\begin{aligned} \text{(b)} \quad P(A|\bar{E}) &= \frac{P(A) \cdot P(\bar{E}|A)}{P(A) \cdot P(\bar{E}|A) + P(B) \cdot P(\bar{E}|B)} = \frac{0,6 \cdot 0,01}{0,6 \cdot 0,01 + 0,4 \cdot 0,05} = 0,231 \\ P(B|\bar{E}) &= 1 - P(A|\bar{E}) = 1 - 0,231 = 0,769. \end{aligned}$$

- (6) **Herkunft von Ausschussware II:** Zwei Maschinen einer Firma seien zu 10% und 90% an der Gesamtproduktion eines bestimmten Gegenstandes beteiligt. Angenommen, die Wahrscheinlichkeit, dass die erste Maschine (M_1) Ausschuss produziert, sei 0,01 und die Wahrscheinlichkeit, dass die zweite Maschine (M_2) Ausschuss liefert, sei 0,05. Wie groß ist die Wahrscheinlichkeit, dass ein zufällig der Tagesproduktion entnommener Gegenstand von M_1 stammt, vorausgesetzt, dass es sich um ein Ausschussprodukt handelt?
Es sei E = das Ereignis, dass ein Gegenstand Ausschussware ist, A_1 das Ereignis, dass er von M_1 hergestellt worden ist, und A_2 , dass er von M_2 stammt, d. h. $P(M_1 | \text{Ausschuss}) = P(A_1|E)$:

$$P(A_1|E) = \frac{P(A_1) \cdot (P(E|A_1))}{P(A_1) \cdot P(E|A_1) + P(A_2) \cdot P(E|A_2)} = \frac{0,10 \cdot 0,01}{0,10 \cdot 0,01 + 0,90 \cdot 0,05} = \frac{1}{46} \simeq 0,022.$$

- (7) **Aus welcher Urne entnommen?** Angenommen, es liegen zwei Urnen vor. Die Wahrscheinlichkeit, Urne I zu wählen, betrage 1/10; für Urne II beträgt sie dann 9/10. Nehmen wir weiter an, die Urnen enthalten schwarze und weiße Kugeln: Urne I enthalte zu 70% schwarze Kugeln, Urne II zu 40%. Wie gross ist die Wahrscheinlichkeit, dass eine mit verbundenen Augen entnommene schwarze Kugel der Urne I entstammt?

E = das Ereignis, dass die Kugel schwarz ist, A_1 = das Ereignis, dass sie aus Urne 1 entnommen ist, und A_2 , dass sie aus Urne II stammt.

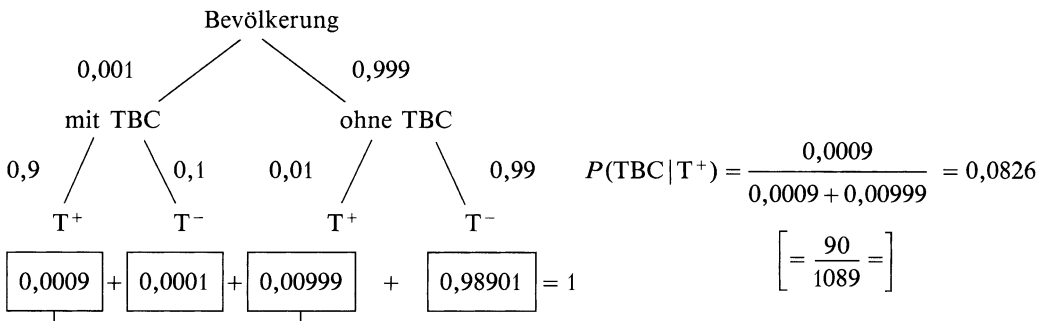
$$P(\text{aus Urne I}|\text{schwarz}) = \frac{0,10 \cdot 0,70}{0,10 \cdot 0,70 + 0,90 \cdot 0,40} = 0,163$$

Das heißt, nach vielen Versuchen wird man in 16,3% aller Fälle, in denen man eine schwarze Kugel zieht, mit Recht auf die Herkunft aus Urne I schließen.

- (8) **Durchleuchtung der Brust mit Röntgenstrahlen:** Nehmen wir an, die Verlässlichkeit einer Durchleuchtung der Brust mit Röntgenstrahlen zur Entdeckung einer Tbc betrage für Tbc-Träger 90%, d. h. 10% der Tbc-Träger bleiben bei der Untersuchung unerkannt; für Tbc-freie Personen betrage sie 99%, d. h. 1 % der Tbc-freien Personen werden fälschlich als Tbc-Träger diagnostiziert.

Aus einer großen Bevölkerung mit 0,1 % Tbc-Fällen sei eine Person durchleuchtet und als Tbc-Träger eingestuft worden. Wie groß ist die Wahrscheinlichkeit, dass diese Person eine Tbc hat?

Gesucht wird somit $P(\text{Tbc-Träger bei positivem Röntgentestbefund}) = P(\text{TBC}|\text{T}^+)$.



d. h. wir finden, dass von den röntgenologisch als Tbc-Träger eingestuft nur gut 8% wirklich eine Tbc aufweisen. Mit $P = 0,0826$ oder 0,083 (a posteriori) ist man jetzt aufgrund der Erfahrung gegenüber $P = 0,001$ (a priori) 83mal sicherer als vorher.

4.5 Der diagnostische Test

- ROC - Analyse
- Der Likelihoodquotient

Aufgrund eines diagnostischen Tests, der positiv (T^+) oder negativ (T^-) ausfallen kann, wird auf die Existenz einer bestimmten Krankheit (K) geschlossen bzw. angenommen, die Krankheit K liege nicht vor ($\bar{K}$).

Wichtige bedingte Wahrscheinlichkeiten, die die Brauchbarkeit eines diagnostischen Tests zur Erkennung einer Krankheit, eines vorliegenden speziellen Befundes, charakterisieren, sind:

- (1) **Spezifität:** $P(\text{T}^-|\bar{K})$, der Anteil negativer Testergebnisse unter den Gesunden
- (2) **Sensitivität:** $P(\text{T}^+|K)$, der Anteil positiver Testergebnisse unter den Kranken
- (3) **Negativer Voraussagewert:** $P(\bar{K}|\text{T}^-)$, der Anteil der Gesunden unter den Personen mit negativem Testergebnis.

- (4) **Positiver Voraussagewert:** $P(K|T^+)$, der Anteil der Kranken unter den Personen mit positivem Testergebnis.

Diese Wahrscheinlichkeiten – sie sollten bei 1 liegen – lassen sich aus umfangreichen Stichproben schätzen. Hierzu benutzt man das Vierfelderschema in Tabelle 4.4.

Tabelle 4.4. Zwei Vierfeldertabellen zum diagnostischen Test

	T^+	T^-	Σ	Befund	Test		Summe
					positiv	negativ	
K	a	b	$a+b$	vor-	a	$b = FN$	$a+b$
$\bar{K}$	c	d	$c+d$	handen		[falsch neg. Test]	[Kranke]
				fehlend	$c = FP$	d	$c+d$
					[falsch pos. Test]		[Gesunde]
Σ	$a+c$	$b+d$	n	Summe	$a+c$	$b+d$	$a+b+c+d$
					[Testpositive]	[Testnegative]	$= n$

Ist der diagnostische Test ideal, so gilt $a + d = n$ [„Konkordanzrate“ $(a + d)/n \leq 1$]. FN (oder b) seien die aufgrund des Tests falsch negativ eingestufteten Personen, FP (oder c) die falsch positiv eingestufteten. Anhand einer Zufallsstichprobe des Umfangs n lässt sich zunächst der Anteil an K erkrankter in der Grundgesamtheit durch die **Prävalenz** $(a + b)/n$ schätzen. Für großes n erhält man die entsprechende Wahrscheinlichkeit $P(K)$ [vgl. Punkt (1) in Übersicht 8]. Entsprechend erhält man auch die bedingten Wahrscheinlichkeiten (2) und (3), die Wahrscheinlichkeit, dass ein Kranker ein positives Testergebnis aufweist, wird **Sensitivität** des Tests genannt, die Wahrscheinlichkeit, dass ein Nicht-Kranker ein negatives Testergebnis aufweist, wird **Spezifität** genannt; beide Wahrscheinlichkeiten sollten möglichst groß sein, jedenfalls deutlich größer als 0,7.

Aus den Wahrscheinlichkeiten (1) bis (3) der Übersicht 8 lassen sich anhand des Satzes nach Bayes so genannte Voraussagewerte (4) und (5) berechnen: den prädiktiven Wert eines:

- positiven Testresultates, die Wahrscheinlichkeit $P(K|T^+)$,
- negativen Testresultates, die Wahrscheinlichkeit $P(\bar{K}|T^-)$.

Übersicht 8. Prävalenz, Sensitivität und Spezifität

	Begriff	relative Häufigkeit	für großes $\frac{\quad}{n}$	Wahrscheinlichkeit
(1)	Prävalenz	$\frac{a+b}{n}$		$P(K)$
(2)	Sensitivität	$\frac{a}{a+b}$		$P(T^+ K)$
(3)	Spezifität	$\frac{d}{c+d}$		$P(T^- \bar{K})$

Voraussagewert eines positiven Testergebnisses:

$$\begin{aligned}
 P(K|T^+) &= \frac{(\text{Prävalenz}) \cdot (\text{Sensitivität})}{(\text{Präv.}) \cdot (\text{Sens.}) + (1 - \text{Präv.}) \cdot (1 - \text{Spez.})} \\
 &= \frac{P(K) \cdot P(T^+|K)}{P(K) \cdot P(T^+|K) + ([1 - P(K)] \cdot [1 - P(T^-|\bar{K})])} \\
 &= \frac{P(K) \cdot P(T^+|K)}{P(K) \cdot P(T^+|K) + P(\bar{K}) \cdot P(T^+|\bar{K})}
 \end{aligned} \tag{4.41}$$

Voraussagewert eines negativen Testergebnisses:

$$\begin{aligned}
 P(\bar{K}|\bar{T}) &= \frac{(1 - \text{Prävalenz}) \cdot (\text{Spezifität})}{(1 - \text{Präv.}) \cdot (\text{Spez.}) + (\text{Präv.}) \cdot (1 - \text{Sens.})} \\
 &= \frac{P(\bar{K}) \cdot P(T^-|\bar{K})}{P(\bar{K}) \cdot P(T^-|\bar{K}) + P(K) \cdot [1 - P(T^+|K)]} \\
 &= \frac{P(\bar{K}) \cdot P(T^-|\bar{K})}{P(\bar{K}) \cdot P(T^-|\bar{K}) + P(K) \cdot P(T^-|K)}
 \end{aligned} \tag{4.42}$$

Der Anteil richtiger Testresultate ist dann die **Resultatvalidität**:

$$P(K|T^+) + P(\bar{K}|\bar{T}) \leq 2 \tag{4.43}$$

Beachtet sei der Unterschied zwischen der aus Sensitivität und Spezifität gebildeten „Testvalidität“ $P(T^+|K) + P(T^-|\bar{K})$ und der **entscheidenden „Resultatvalidität“**: $P(K|T^+) + P(\bar{K}|\bar{T})$.

Übersicht 9. Voraussagewerte und Resultatvalidität

Begriff	relative Häufigkeit	für großes n	Wahrscheinlichkeit
(1) Voraussagewert eines positiven Tests	$\frac{a}{a+c}$		$P(K T^+)$
(2) Voraussagewert eines negativen Tests	$\frac{d}{b+d}$		$P(\bar{K} \bar{T})$
(3) Anteil richtiger Resultate	$\frac{a}{a+c} + \frac{d}{b+d}$		$P(K T^+) + P(\bar{K} \bar{T})$

Hinweise:

- (1) Die Prävalenz $P(K)$ wird anhand einer Zufallsstichprobe aus der Bevölkerung geschätzt. Diese Wahrscheinlichkeit $P(K)$ wird als a-priori-Wahrscheinlichkeit für K bezeichnet. $P(K|X)$ als die a-posteriori-Wahrscheinlichkeit für K , z. B. $P(K|T^+)$.
- (2) $P(K|T^+)$, der positive prädiktive Wert eines diagnostischen Tests, der Voraussagewert eines positiven Tests, ist die Wahrscheinlichkeit, die Krankheit K aufzuweisen, wenn der Test positiv (T^+) ausfällt: $P(K|T^+) = P(K \cap T^+)/P(T^+)$. Dieser Wert sollte wie der Voraussagewert eines negativen Tests, der negative prädiktive Wert $P(\bar{K}|\bar{T}) = P(\bar{K} \cap \bar{T})/P(\bar{T})$ möglichst nahe bei Eins liegen.

- (3) Beachtet sei, daß $P(K|T^+)$ mit zunehmender Prävalenz ebenfalls zunimmt [vgl. (4.41)]; $P(\bar{K}|T^-)$ nimmt natürlich mit abnehmender Prävalenz zu [vgl. (4.42)]. Dieser Zusammenhang ist in Abbildung 4.5 für einen Test mit hoher Sensitivität und hoher Spezifität dargestellt (siehe auch Tabelle 4.5).

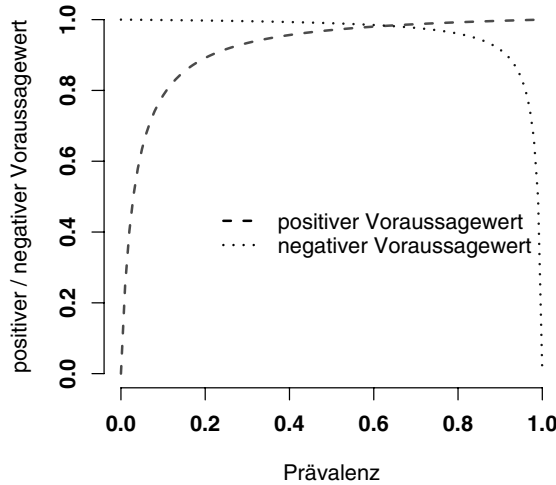


Abb. 4.5. Positiver und negativer Voraussagewert in Abhängigkeit von der Prävalenz (Sensitivität = 0.99, Spezifität = 0.97)

Sind beim Vorliegen einer ernsten Krankheit falsch positive Testergebnisse zu erwarten, die:

- (a) nicht zu sehr irritieren, dann sollte die Sensitivität des Tests, die ja der Power entspricht, größer sein als die Spezifität,
- (b) stark irritieren, dann sollte die Spezifität möglichst hoch sein,
- (c) ebenso stark irritieren wie eine unterlassene Behandlung aufgrund falsch negativer Testergebnisse, dann sollten beide Maße etwa gleichgroß sein (vgl. die folgenden Beispiele und Tabelle 4.5). Kann der Arzt die Krankheit nur lindern, so sollte die Spezifität hoch sein.

Beispiel: Von 10000 Personen weisen 150 die Krankheit K auf; bei 130 von den 150 Personen ist der Test positiv ausgefallen. Wie groß ist (a) die Sensitivität und (b) der positive Voraussagewert des Tests [gegeben sei seine Spezifität $P(T^-|\bar{K}) = 0,940$ und damit $P(T^+|\bar{K}) = 0,060$]? ✱

$$(a) P(T^+|K) = \frac{P(T^+ \cap K)}{P(K)} = \frac{130/10\,000}{150/10\,000} = 0,8667$$

In knapp 87% aller Fälle ist zu erwarten, daß der Test positiv ausfällt, vorausgesetzt, die Krankheit K liegt vor.

$$(b) P(K|T^+) = \frac{P(K) \cdot P(T^+|K)}{P(K) \cdot P(T^+|K) + P(\bar{K}) \cdot P(T^+|\bar{K})} = \frac{0,015 \cdot 0,8667}{0,015 \cdot 0,8667 + 0,985 \cdot 0,060} = 0,1803$$

In 18 % aller Fälle ist zu erwarten, dass die Krankheit K vorliegt, vorausgesetzt, der Test ist positiv ausgefallen.

Tabelle 4.5. Positiver $[P(K|T^+)]$ und negativer $[P(\bar{K}|T^-)]$ Voraussagewert eines Tests in Abhängigkeit von sechs Prävalenz-Stufen $[P(K)]$ bei Gleichheit von Sensitivität $[P(T^+|K) = Se]$ und Spezifität $[P(T^-|\bar{K}) = Sp]$ des diagnostischen Tests auf zwei hohen Niveaus ($P \approx 1$ bedeutet $P > 0,9999$)

$P(K)$	$Se = Sp = 0,95$		$Se = Sp = 0,99$	
	$P(K T^+)$	$P(\bar{K} T^-)$	$P(K T^+)$	$P(\bar{K} T^-)$
0,0001	0,0019	≈ 1	0,0098	≈ 1
0,001	0,019	≈ 1	0,090	≈ 1
0,01	0,161	0,9995	0,500	0,9999
0,10	0,679	0,994	0,917	0,999
0,50	0,950	0,950	0,990	0,990
0,90	0,994	0,679	0,999	0,917



Beispiel: Nach (4.41) und (4.42) ergeben sich die folgenden prädiktiven Werte oder Voraussagewerte für eine Sensitivität und Spezifität von jeweils 0,95 und einer Prävalenz von 0,005 bzw. 0,05 bzw. 0,5:

$$P(K) = \left\{ \begin{matrix} 0,005 \\ 0,05 \\ 0,5 \end{matrix} \right\} : P(K|T^+) = \left\{ \begin{matrix} 0,0872 \\ 0,5000 \\ 0,9500 \end{matrix} \right\} \text{ und } P(\bar{K}|T^-) = \left\{ \begin{matrix} 0,9997 \\ 0,9972 \\ 0,9500 \end{matrix} \right\}$$

Rechnet man somit in einer **Reihenuntersuchung** mit der Prävalenz einer Krankheit K von einem halben Prozent, dann beträgt der Aussagewert eines positiven Testergebnisses knapp 9%. Die Wahrscheinlichkeit dafür, dass bei einem Patienten trotz eines positiven Testresultates die Krankheit K nicht vorliegt, beträgt dann gut 91% (1-0,0872). Bei negativem Testresultat lässt sich dann K mit Sicherheit (99,97%) ausschließen. Weitere Beispiele sind in Tabelle 4.5 zusammengestellt. So ergibt sich nach (4.41) und aus Tabelle 4.5 für $P(K) = 0,001$ und bei Gleichheit von Sensitivität und Spezifität ($= 0,95$), dass bei einer so seltenen Krankheit nur knapp 2% aller Probanden mit positivem Testergebnis die Krankheit K wirklich aufweisen.

4.5.1 ROC - Analyse

Das Ergebnis eines (diagnostischen) Tests ist häufig eine kontinuierliche (quantitative) Messgröße (physikalische Messungen, Ergebnisse aus klinisch-chemischen Verfahren). Eine „natürliche“ Grenze zwischen „pathologisch“ und „normal“ gibt es in der Regel nicht. Sensitivität und Spezifität sind somit abhängig von der Festlegung eines geeigneten **Trennwertes** („cut off value“), der den Anteil falsch positiver und/oder falsch negativer Entscheidungen möglichst gering hält. Eine „optimale“ Trenngröße ist somit abhängig von den Risiken falscher Entscheidungen.

Eine häufig eingesetzte Methode zur Festlegung eines objektiven Trennwertes ist das **ROC-Verfahren** („receiver operating characteristic“). Dazu werden über den Definitionsbereich der Messgröße in diskreten Schritten möglichst viele (abhängig von der Anzahl der Messungen) Werte durchlaufen und die zugehörigen Sensitivitäten und Spezifitäten berechnet (Abbildung 4.6). Das Ergebnis wird in einem speziellen Diagramm, der sogenannten **ROC-Kurve**, aufgezeichnet.

Ein **optimaler Trennwert** ist insbesondere der Wert, für den Sensitivität und Spezifität möglichst hoch liegen (linke obere Ecke in Abbildung 4.7). Dazu kann eine Tangente an die ROC-Kurve parallel zu der Diagonalen (Winkelhalbierende) gelegt werden, die einen Wert mit maximalem Abstand von der Diagonalen festlegt (Berührungspunkt). In dem Beispiel für den Nüchternblutzucker gilt dieses für den Wert 100mg/dl , der zu einer Sensitivität von 85% und zu einer Spezifität von 81% führt. Ein Test trennt insgesamt um so besser, je größer die **Fläche unter der ROC-Kurve** ist. Diese ist maximal 1 bei einer 100%igen Sensitivität und Spezifität. Sie ist 0,5, wenn eine

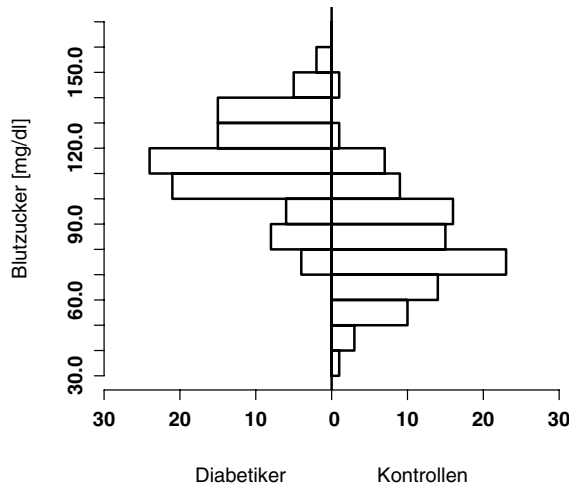


Abb. 4.6. Histogramm zum Nüchternblutzucker [mg/dl], gemessen bei jeweils n=100 Diabetikern und gesunden Kontrollpersonen

Trennung nicht möglich ist, d.h. wenn die ROC-Kurve entlang der Diagonalen verläuft. In dem Beispiel (Abbildung 4.7) ist die Fläche unter der Kurve (Area Under Curve, AUC) 0,885.

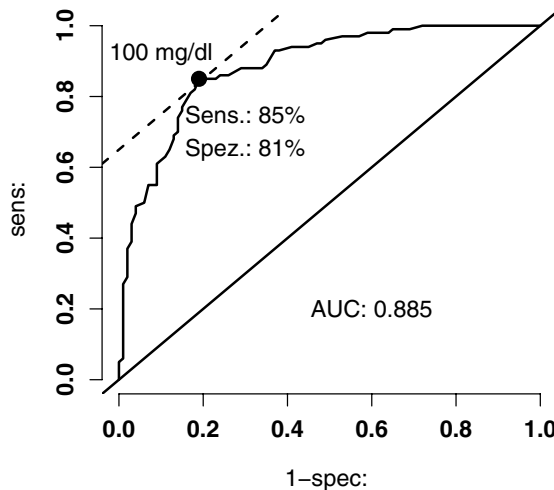


Abb. 4.7. ROC-Kurve zu den Blutzuckerwerten aus Abbildung 4.6

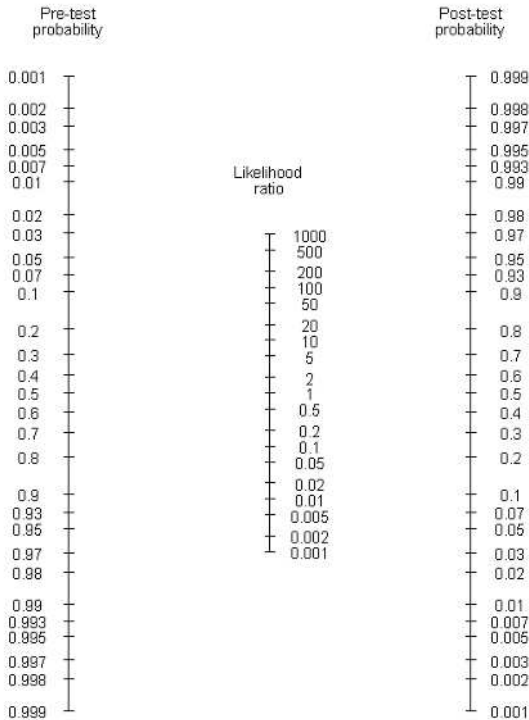
4.5.2 Der Likelihoodquotient

Der **Likelihood-Quotient** ist eine zusammenfassende, prävalenzunabhängige Maßzahl, die die Sensitivität und die Spezifität eines diagnostischen Tests berücksichtigt. Er beschreibt, wie sich das Resultat eines Tests auf die „Chancen“ (odds) auswirkt, dass tatsächlich die Erkrankung vorliegt (LR^+) oder nicht vorliegt (LR^-).

$$LR^+ = \frac{\text{Sensitivität}}{1 - \text{Spezifität}} \quad (4.44)$$

$$LR^- = \frac{1 - \text{Sensitivität}}{\text{Spezifität}} \quad (4.45)$$

In Verbindung mit einer **Prätest-Chance** für eine Krankheit, die aus der Prävalenz bestimmt werden kann, ermöglicht der Likelihood-Quotient die Berechnung der **Posttest-Chance**.



Ist beispielsweise die Prätest-Wahrscheinlichkeit (Prävalenz) 0,80, dann ist für eine beliebige Person aus dieser Population die Chance, die Krankheit zu haben, 80 zu 20 oder 4 zu 1 (Prätest-Chance). Der Likelihood-Quotient gibt an, wie sich die Chance für das Vorliegen der Krankheit durch das Testergebnis ändert. Mit einem Likelihood-Quotienten von z.B. $LR^+ = 2,5$ erhöht sich die Chance für das Vorliegen der Erkrankung um das 10fache, die Wahrscheinlichkeit beträgt dann 0,91 gegenüber der ursprünglich angenommenen Prävalenz von 0,8. Die Umrechnung von Wahrscheinlichkeiten in Chancen und umgekehrt nach 4.46 ist leicht verwirrend. Das **Fagan-Nomogramm** [Fag75] (vgl. Abbildung 4.8) ermöglicht auf einen Blick die Einschätzung der Wertigkeit eines Testverfahrens unter verschiedenen Annahmen zur Prävalenz oder auch den Vergleich verschiedener Testverfahren unter Annahme einer festen Prävalenz.

Abb. 4.8. Fagan-Nomogramm zur Bestimmung der Posttest-Wahrscheinlichkeit (Fagan 1975)

$\text{Prätest-Chance} = \frac{\text{Prätest-Wahrscheinlichkeit}}{1 - \text{Prätest-Wahrscheinlichkeit}}$ $\text{Posttest-Chance} = \text{Prätest-Chance} \cdot LR^+$ $\text{Posttest-Wahrscheinlichkeit} = \frac{\text{Posttest-Chance}}{\text{Posttest-Chance} + 1}$	(4.46)



Beispiel: Der „kontralaterale Loslassschmerz“ im Rahmen der Diagnostik einer akuten Appendizitis hat einen Likelihood-Quotienten von $LR^+ = 5,0$. Liegt ein Verdacht auf eine Appendizitis mit einer Prätest-Wahrscheinlichkeit von 0,6 vor (das entspricht einer Prätest-Chance von 1,5), dann ist die Posttest-Chance bei Vorliegen des Symptoms 7,5. Die Posttest-Wahrscheinlichkeit ist

dann 0,88. Weitere Symptome und Befunde, z.B. Fieber über 38° ($LR^{+} = 3$) oder Leukozytenzahl $> 15 \cdot 10^9/L$ ($LR^{+} = 7$), können schrittweise zu einer erhöhten Posttest-Wahrscheinlichkeit beitragen. Die einzelnen Schritte sind durch entsprechende Linien im Fagan-Nomogramm (Abbildung 4.8) nachzuvollziehen.

4.6 Maßzahlen in der Epidemiologie

- Prävalenz und Inzidenz
- Standardisierungen

Epidemiologie („die Lehre von dem, was über das Volk kommt“) ist die Wissenschaft vom Auftreten von Krankheiten und ihrer Determinanten in unterschiedlichen Bevölkerungsgruppen. Mit den Grundaufgaben:

- **Krankheitsursachen aufspüren und ihre relative Bedeutung herausarbeiten.**
- Bevölkerungsteile identifizieren, die besonderen **gesundheitlichen Risiken** unterliegen.
- Programme entwickeln und Prioritäten setzen, um (1) und (2) zu realisieren, um den **Gesundheitszustand der Bevölkerung zu verbessern**: Infektionsmöglichkeiten und präventive Maßnahmen zu erkunden und zu bewerten sowie Trends vorauszusagen.

Im 18. Jahrhundert wurde bekannt, dass katholische Nonnen einem höheren Brustkrebs-Risiko und dass Schornsteinfeger einem höheren Hodenkrebs-Risiko ausgesetzt sind. Zum anderen wird auch heute noch behauptet, der hohe Anteil Krebskranker unter den Rauchern habe nichts mit deren Rauchgewohnheiten zu tun – eher seien andere Faktoren, persönliche Merkmale, hierfür verantwortlich. Diese These wurde kurz nach dem 2. Weltkrieg auch von zwei führenden Statistikern und starken Rauchern vertreten, R. A. Fisher (1890–1962) und J. Neyman (1894–1981), die – ohne befreundet zu sein – unabhängig voneinander Front machten gegen Epidemiologen ihrer Zeit. R. A. Fisher publizierte sogar im Auftrag einer Tobacco Company. Heute wissen wir, dass Raucher, die das Rauchen aufgegeben haben, ein merklich geringeres Risiko haben, an Lungenkrebs zu erkranken.

Krankheiten sind abhängig (a) vom Grad der Exponiertheit der Person, (b) vom Grad der Anfälligkeit der Person und (c) von weiteren schädigenden Faktoren. Ziel der Epidemiologie ist es, jeder Krankheit mindestens einen sogenannten **Risikofaktor** zuzuordnen. Risikofaktoren sind Krankheitsursachen im weiteren oder engeren Sinne, identifiziert anhand von „Fall-Kontroll-Studien“ [BD80] und „Kohortenstudien“ [BD87] (vgl. Kapitel [7.6.3] auf Seite 487).

Ein Risikofaktor ist jedes Merkmal, das dem Krankheitsstatus zeitlich vorangeht und mit ihm nach Ansicht von Fachleuten kausal verknüpft ist; d. h. ein epidemiologisch belegter Gefährdungsfaktor heißt Risikofaktor. In epidemiologischen Studien nennt man bekannte Risikofaktoren für eine bestimmte Krankheit **Mitursachen** (confounding factors, vgl. z. B. Breslow und Day [BD80], S. 93–108). **Confounding** bezeichnet eine Verzerrung der Beziehung zwischen Exponiertheit und Krankheit durch Mitursachen. Mitursachen ermöglichen diagnostische Voraussagen: 12 Beispiele sowie Analogien und Unterschiede zwischen Fall-Kontroll-Studien und Kohorten-Studien geben Miettinen und Cook [MC81].

4.6.1 Prävalenz und Inzidenz

Unter Prävalenz (Präv) versteht man in der Epidemiologie den Anteil von Personen in einer bestimmten Bevölkerung (population), der eine bestimmte Krankheit aufweist, d. h. aus der Sicht dieses Kapitels auch die Wahrscheinlichkeit, dass eine Person, die zufällig aus der Bevölkerung

ausgewählt wurde, erkrankt ist. Die **Prävalenzrate** (Präv) (4.47) wird als ein **Bestandsmaß** somit durch die Zahl der Erkrankten im Verhältnis zur Gesamtbevölkerung zu einem bestimmten Zeitpunkt bestimmt.

$$\text{Präv} = \frac{\text{Zahl der Erkrankten}}{\text{Anzahl der Individuen in der Bevölkerung}} \quad (4.47)$$

Der Bezug auf die Zeit kann wie folgt präzisiert werden:

- Die **Punktprävalenz** wird durch einen genau bestimmten Zeitpunkt, z. B. „im Augenblick“ oder „zum gegebenen Stichtag“ definiert.
- Die **Periodenprävalenz** wird durch einen Zeitraum definiert, wie z. B. „im vergangenen Monat“, „im geschlechtsreifen Alter“ oder „im Senium“, im letzten Jahr (Jahresprävalenz) oder während des gesamten Lebens (Lebenszeitprävalenz)



Beispiel: Wenn in Amerika der Anteil der schwer Übergewichtigen in der erwachsenen Bevölkerung von 12,0% im Jahr 1991 auf 17,9% im Jahr 1999 stieg, kann die Dynamik des Geschehens mittels der Angabe zweier Punktprävalenzen weit besser vermittelt werden, als wenn der Anteil der Übergewichtigen nur durch eine Periodenprävalenz (zwischen 1991 und 1999) erfasst würde.

Die **Inzidenz** wird definiert als die Anzahl der **Neuerkrankungen** in einer Bevölkerung während einer bestimmten Zeit. Damit kennzeichnet die Inzidenz die **Veränderung des Krankenstandes** hinsichtlich einer bestimmten Krankheit in der Bevölkerung. Dabei unterscheidet man die folgenden Situationen:

- Die **kumulative Inzidenz** CI (4.48) gibt an, welcher Anteil einer gesunden Bevölkerung (population at risk) in einem bestimmten Zeitraum T neu erkrankt ist (häufig auf 100000 bezogen).

$$CI = \frac{\text{Anzahl der neu erkrankten Personen innerhalb eines Zeitraumes } T}{\text{Anzahl der gesunden Personen zu Beginn des Zeitraumes } T} \quad (4.48)$$

Die kumulative Inzidenz ist demnach eine Maßzahl für das Entstehen einer Krankheit. Sie erfordert die Beobachtung (follow up) einer Bevölkerung mindestens über den Referenzzeitraum (Longitudinalstudien, Kohortenstudien).

- Die Definition der **Inzidenzdichte** I (4.49) (hazard rate) erfolgt dagegen mit einem Bezug auf **Risikozeiten**. Unter Risikozeit versteht man die Zeit, in der ein einzelnes **nicht erkranktes** Individuum aus der Bevölkerung einer Exposition (Risiko) ausgesetzt ist. Diese Risikozeiten werden für alle Individuen zur **Personenzeit** unter Risiko addiert.

$$I = \frac{\text{Anzahl Erkrankungsfälle}}{\text{Personenzeit}} \quad (4.49)$$

Die Inzidenzdichte I ist eine **Beziehungszahl** und keine Verhältniszahl, kann also insbesondere auch nicht als eine Wahrscheinlichkeit interpretiert werden. I ist eine Maßzahl, die Neuerkrankungen pro Risikozeit angibt (Dimension: $1/\text{Zeitspanne}$) und kennzeichnet damit die Dynamik der Erkrankungen in der Bevölkerung.



Beispiel: In Stockholm ereigneten sich 1973 unter den 40-44 jährigen Männern (41532 Personenjahre) 29 Herzinfarkte.

Die Inzidenzdichte betrug $I = 29/41532 = 0,00071/\text{Jahr}$.

Im Falle einer konstanten Inzidenzdichte kann der Zusammenhang zwischen der kumulativen Inzidenz und der Inzidenzdichte durch $CI = 1 - \exp(-T \cdot I)$ beschrieben werden. Befindet sich

die Bevölkerung hinsichtlich der Zu- und Abgänge hinsichtlich einer Erkrankung im Gleichgewicht (steady state) und ist die Inzidenzdichte gering ($I \cdot T < 0,1$), dann folgt näherungsweise die wichtige Beziehung $CI \approx I \cdot T$.

Beispiel: Aus einer Inzidenzdichte von 0.005/J ergibt sich für die Erkrankungswahrscheinlichkeit innerhalb von 5 Jahren $CI_{5,J} = 1 - \exp((-0,005/j)5J) = 0,025$.



Der Zusammenhang zwischen der Prävalenz (Präv) und der Inzidenzdichte I kann in stabilen Bevölkerungen (Gleichgewicht) mit Hilfe der mittleren Krankheitsdauer D verdeutlicht werden.

$$\frac{\text{Präv}}{1 - \text{Präv}} = I \cdot D \quad (4.50)$$

Der „Anteil es Krankenstandes“ in einer Bevölkerung auf gesunde Personen bezogen (z. B. auf 100000) kann aus dem Produkt aus Inzidenzdichte und Krankheitsdauer ermittelt werden. Wird weiterhin angenommen, dass die Prävalenz sehr klein ist ($< 1\%$), dann ist $\text{Präv} \approx I \cdot D$.

Beispiel: Die jährliche Inzidenzrate für Typhus in der BRD beträgt 0.2 auf 100000 Personenjahre. Die durchschnittliche Erkrankungsdauer beträgt etwa ein Monat. Bei sehr seltenen Erkrankungen kann $(1 - \text{Präv}) = 1$ angenommen werden. Dann ist $\text{Präv} \approx I \cdot D = (0,2/100000)(1/12) = 0,00017\%$. Bei einer Bevölkerung von 81.5 Mio. sind somit zu einem festen Zeitpunkt ca. 14 Fälle von Typhus in der Bevölkerung zu erwarten.



Die Sterblichkeit oder **Mortalität** kann als ein Sonderfall für die Inzidenz betrachtet werden. Als Ereignisse werden dann nicht die Erkrankungen, sondern die Todesfälle an einer Erkrankung gezählt. Entsprechend gibt es auch eine kumulative Mortalität und eine Mortalität(-sdichte), hier als Sterblichkeit pro Zeitintervall (bezogen auf die Personenzeit der Population unter Risiko (hazard rate)).

Sonderfall: In der amtlichen Mortalitätsstatistik wird statt der Personenzeit nur die durchschnittliche Populationsgröße mit Bezug auf das Zeitintervall 1 Jahr berechnet.

4.6.2 Standardisierungen

Vorbemerkung: Eine Käsemasse mit $< 54\%$ Wasser heißt Hartkäse, etwa einer mit 40% Wasser, d. h. 60% Trockenmasse. Mit 50% Fett in der Trockenmasse liegt der absolute Fettgehalt dieses Käses bei $[0,50 \cdot 0,60 = 0,30]$ 30% ; 100 g dieses Käses enthalten 30 g Fett.

(1) Einfache Standardisierungen

Die standardisierten Anteile der Raucherinnen und Raucher in der Bevölkerung: Eine Bevölkerung Erwachsener bestehe zu einem bestimmten Zeitpunkt aus den Anteilen p_F und p_M . Die Anteile der Raucherinnen bzw. der Raucher betragen p_{RF} und p_{RM} ; dann ist der standardisierte Anteil der Raucherinnen in der Bevölkerung $p_{RFS} = p_F \cdot p_{RF}$ und der entsprechende standardisierte Anteil der Raucher $p_{RMS} = p_M \cdot p_{RM}$ und insgesamt der standardisierte Anteil beider in der Bevölkerung der Erwachsenen: $p_{RS} = p_{RFS} + p_{RMS}$.

Mit $\hat{p}_F = 0,53$ und $\hat{p}_M = 0,47$ sowie $\hat{p}_{RF} = 0,3$ und $\hat{p}_{RM} = 0,4$ erhält man $\hat{p}_{RFS} = 0,159$ und $\hat{p}_{RMS} = 0,188$ sowie $\hat{p}_{RS} = 0,347$.

Eine hervorragende Übersicht über Standardisierungen gibt Kalton [Kal68] (vgl. auch Fersch [Fer85], Fleiss [Fle81], Hochstädter [Hoc91] und Kitagawa [Kit64]); viele Lehrbücher der Epidemiologie bieten weitere Hinweise.

(2) Berechnung altersstandardisierter Sterbeziffern

Sind für ein bestimmtes Jahr die altersstandardisierten Sterbeziffern (für Inzidenzen gilt entsprechendes) von Gebieten (A, B, ...) mit unterschiedlicher Altersverteilung zu vergleichen, so bezieht man die Todesfälle dieser Gebiete auf eine Standardbevölkerung mit z. B. $10\,000$ Säuglingen

(< 1 J.,) ..., ... über 90-Jährigen (insgesamt 800 000 Personen). Sind in A von 182 Säuglingen 2 gestorben (vgl. Bundesrepublik 1990: 5 von 1000), so bedeutet dies für die säuglingsbezogene Sterbeziffer, jeweils auf 1000 Säuglinge bezogen ($x/1000 = 2/182$; $x = 11$), eine Sterbeziffer von 11 pro 1000. Entsprechende Werte berechnet man für sämtliche Altersgruppen und alle zu vergleichenden Gebiete.

Für jede Altersgruppe werden jetzt die entsprechenden erwarteten Todesfälle, bezogen auf die Standardbevölkerung, berechnet. Für Säuglinge (10 000 in der Standardbevölkerung) aus A ($x/10000 = 11/1000$; $x = 110$) erhält man 110 erwartete Todesfälle. Die Summe der erwarteten Todesfälle aller Altersgruppen (sagen wir 8000) wird jetzt durch die gesamte Standardbevölkerung (800 000) dividiert: $8000/800\,000 = 8/800 = x/1000$; 10 pro 1000; dies ist die altersstandardisierte Sterbeziffer für das Gebiet A, die mit denen anderer Gebiete direkt vergleichbar ist. Näheres ist z. B. Kahn und Sempas [KS89] (Chapter 5) zu entnehmen.

(3) Das Inzidenzdichte-Verhältnis

Das Inzidenzdichte-Verhältnis *IDR* (Incidence Density Ratio) ist das Verhältnis zweier Neuerkrankungsraten: „Fälle pro Personen-Jahre“ für eine Risikogruppe zu „Fälle pro Personen-Jahre“ für eine Gruppe, die diesem Risiko nicht ausgesetzt war. Man geht von einem Schadeffekt aus, sobald $IDR > 1$:

	Exponierte	Nicht Exponierte	Summe
Fälle	a	b	$m = a + b$
Länge*	L_1	L_0	$L = L_1 + L_0$

$$\widehat{IDR} = \frac{a/L_1}{b/L_0} \quad (A)$$

Schadeffekt,
sobald
 $IDR > 1$

* Beobachtungszeit für alle Personen der jeweiligen Studien-
gruppe

Beispiel:

$$\begin{array}{rcl} a = 10 & L_1 = & 2\,935 \\ b = 239 & L_0 = & 135\,130 \\ m = 249 & L = & 138\,065 \end{array}$$

Zu vergleichen sind 10 Fälle in 2935 Personen-Jahren (341 F pro 100 000 P-J) mit 239 Fällen in 135 130 Personen-Jahren (177 F pro 100 000 P-J). Daraus ergibt sich $IDR = 1,926$.

(4) Das standardisierte Mortalitätsverhältnis

Das standardisierte Mortalitätsverhältnis SMR (Standardized Mortality Ratio) ist eine geschlechts-spezifische Sterberate, die altersstandardisiert ist und indirekt berechnet wird:

$$SMR = 100 \frac{\text{Todesfälle in der untersuchten Bevölkerung}}{\left(\frac{\text{Todesfälle der Gesamtbev.}}{\text{pro 1000}} \right) \left(\frac{\text{Unters. Bev.}}{\text{in 1000}} \right)} \quad (4.51)$$

$$SMR = 100 \frac{\text{Beobachtete T.}}{\text{Erwartete T.}} = 100 \frac{B}{E} \quad (4.52)$$

Somit gilt für eine untersuchte Bevölkerung, dass sie rein formal (beschreibend) für eine

$$SMR \left\{ \begin{array}{l} > 100 \text{ einem höheren} \\ = 100 \text{ demselben} \\ < 100 \text{ einem niedrigeren} \end{array} \right\} \text{ Risiko}$$

ausgesetzt ist.

Beispiel: Ein stark vereinfachtes Beispiel, das nur drei Altersklassen (Männer) umfasst, enthält die folgende Tabelle.



Alters- gruppe in Jahren	Todesfälle der Gesamtbev. pro 1000	Untersuchte Bevölkerung in 1000	Erwartete Todesfälle (2)·(3)	Beobachtete Todesfälle [benöt. wird Σ]
(1)	(2)	(3)	(4)	(5)
21 – 35	0,042	25	1,05	3
36 – 50	0,507	40	20,28	38
51 – 65	2,072	10	20,72	12
Summe: $E = 42,05$				$B = 53$
$\widehat{SMR} = 100(53/42,05) = 126 > 100$				

Zufallsvariablen, Verteilungen

- Die Zufallsvariable
- Maßzahlen zur Kennzeichnung der Verteilung
- Diskrete Verteilungen
- Stetige Verteilungen
- Testverteilungen
- Verteilung zweidimensionaler Zufallsvariablen

5.1 Die Zufallsvariable

Eine Zufallsvariable ist eine Größe, die bei einem Zufallsexperiment auftritt, z. B. das Werfen einer „6“ mit einem Würfel oder die Länge der Brenndauer einer Glühbirne. Eine Zufallsvariable oder zufällige Variable ordnet jedem Ausgang des Experimentes eine Zahl zu (vgl. Abb. 5.1). Hat man ein Experiment gemacht, bei dem die Zufallsvariable X einen Wert x angenommen hat, so nennt man x eine **Realisierung von X** . Die **Grundgesamtheit** ist eine Menge aller möglichen Realisierungen einer Zufallsvariablen, die **Stichprobe** ist die n -fache Realisierung. Die Werte von x sind **reelle Zahlen**. Hierunter versteht man Zahlen, die sich durch Dezimalzahlen mit endlich $(2, -4)$ oder unendlich vielen Stellen [periodisch $(-7/3)$ oder nicht periodisch $(\sqrt{2}, \lg 3, \pi, e)$] darstellen lassen. Die Wahrscheinlichkeit des Ereignisses, dass X irgendeinen Wert in dem Intervall von a bis b annimmt, bezeichnen wir mit $P(a < X < b)$. Entsprechend ist $P(-\infty < X < +\infty)$ das sichere Ereignis, da X ja stets irgendeinen Wert auf der Zahlengeraden annehmen muss. Soll X irgendeinen Wert annehmen, der größer als c ist: $P(X > c)$, so gilt, da $P(X > c) + P(\leq c) = 1$, für beliebiges reelles c :

$$P(X > c) = 1 - P(X \leq c) \quad P(X \geq c) = 1 - P(X < c) \quad (5.1)$$

$$\begin{aligned} P(a < X < b) &= P(X < b) - P(X \leq a) & P(a < X \leq b) &= P(X \leq b) - P(X \leq a) \\ P(a \leq X < b) &= P(X < b) - P(X < a) & P(a \leq X \leq b) &= P(X \leq b) - P(X < a) \end{aligned} \quad (5.2)$$

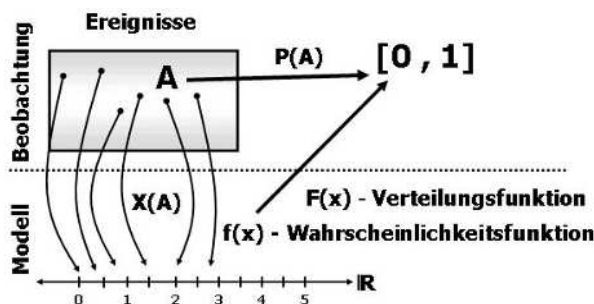


Abb. 5.1. Das Modell der diskreten Zufallsvariablen

Beispiel: Die beim Wurf eines regelmäßigen Würfels erzielte Augenzahl sei X , dann ist $P(X = 6)$ gleich $1/6$: $P(5 < X < 6) = 0$ $P(5 \leq X < 6) = 1/6$
 $P(1 \leq X \leq 6) = 1$ $P(5 < X \leq 6) = 1/6$
 $P(X > 1) = 1 - P(X \leq 1) = 1 - 1/6 = 5/6$.



Zusammenfassung:



- (1) Eine Abbildung oder Funktion, die den Elementen der Ergebnis- oder Ereignismenge eines Zufallsexperimentes reelle Zahlen zuordnet, heißt Zufallsvariable (vgl. Abbildung 5.1).
- (2) Zufallsvariable werden mit großen Buchstaben wie X , Y und Z bezeichnet; kleine Buchstaben wie x , y und z stehen für mögliche Werte: Realisierungen der betreffenden Zufallsvariablen.
- (3) Jede Zufallsvariable X ist eine Abbildung des Stichprobenraumes in die reellen Zahlen; treten hierbei nur ganze Zahlen auf, so heißt die Zufallsvariable **diskret**, andernfalls heißt sie **stetig**.
- (4) Eine Zufallsvariable X heißt diskret, wenn sie höchstens abzählbar viele Werte x_i mit den Einzelwahrscheinlichkeiten $P(X = x_i)$ annehmen kann, deren Summe 1 ist.
- (5) X wird als Zufallsvariable bezeichnet, wenn für jede reelle Zahl x die Wahrscheinlichkeit $P(X \leq x)$ existiert.

5.1.1 Wahrscheinlichkeitsfunktion, Wahrscheinlichkeitsdichte und Verteilungsfunktion

Die „Wahrscheinlichkeitsverteilung“ einer Zufallsvariablen gibt an, mit welcher Wahrscheinlichkeit die Werte der Zufallsvariablen angenommen werden. Die „Wahrscheinlichkeitsverteilung“ der Zufallsvariablen X wird durch die **Verteilungsfunktion**

$$F(x) = P(X \leq x) \quad (5.3)$$

eindeutig definiert. Sie gibt die Wahrscheinlichkeit an, dass die Zufallsvariable X einen Wert kleiner x oder gleich x annimmt.

Die Funktion F , die jeder reellen Zahl x die Wahrscheinlichkeit $P(X \leq x)$ des Ereignisses ($X \leq x$) zuordnet, nennt man **Verteilungsfunktion** der Zufallsvariablen X . $F(x)$ ist **monoton nicht fallend** (Abb. 5.4: links oben) mit den Grenzwerten $F(-\infty) = 0$ und $F(\infty) = 1$.

F ist damit für alle reellen Zahlen x definiert. $F(x)$ wird auch **Summenhäufigkeitsfunktion** oder **kumulierte Wahrscheinlichkeitsverteilung** genannt.

Die Funktion $\hat{F}_n(x) = [\text{Anzahl der } X_i \leq x]/n$ heißt **Stichproben-Verteilungsfunktion** oder **empirische Verteilungsfunktion**.

Beispiel (einfacher Würfelwurf): Die Zufallsvariable X ist die Zahl der geworfenen Augen. Die Wahrscheinlichkeiten für die zu werfenden Augen sind je $1/6$. Die Verteilungsfunktion $F(x)$ nimmt die folgenden Werte an:



x	$x < 1$	$1 \leq x < 2$	$2 \leq x < 3$	$3 \leq x < 4$	$4 \leq x < 5$	$5 \leq x < 6$	$x \geq 6$
$F(x)$	0	1/6	1/6+1/6=1/3	1/6+1/3=1/2	1/6+1/2=2/3	1/6+2/3=5/6	1/6+5/6=1

Man erhält eine sog. „**Treppenfunktion**“ (vgl. Abbildung 5.2). Sie springt genau an denjenigen Stellen x nach oben, an denen X einen Wert mit der Wahrscheinlichkeit $1/6$ annimmt. Zwischen

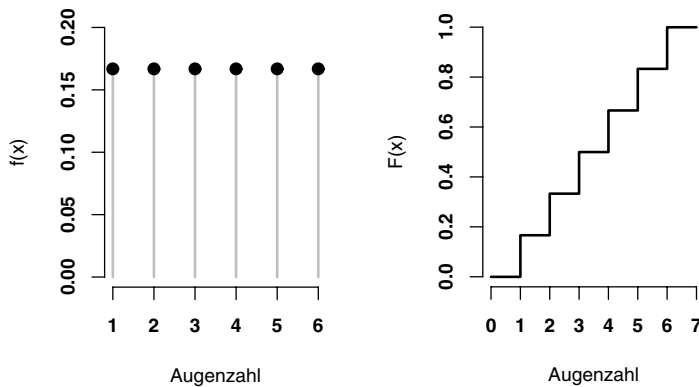


Abb. 5.2. Wahrscheinlichkeitsfunktion und Verteilungsfunktion für die Zufallsvariable: Augenzahl beim Werfen eines Würfels

zwei benachbarten Sprungstellen verläuft sie konstant. Man zeichne sich dies einmal auf (Abszisse: (x) die ganzen Zahlen von 0 bis 7; Ordinate: $[P(X \leq x)]$ in Sechstel geteilt von 0 bis 1).



Beispiel: Das Zufallsexperiment einmaliges Werfen von 2 Würfeln kann durch die Zufallsvariable X Summe der aufliegenden Augenzahlen beschrieben werden. Diese nimmt diskrete Werte $2, \dots, 12$ an, zu denen die Wahrscheinlichkeiten entsprechend abgeleitet werden können. Wahrscheinlichkeitsfunktion und Verteilungsfunktion sind in Abbildung 5.3 dargestellt.

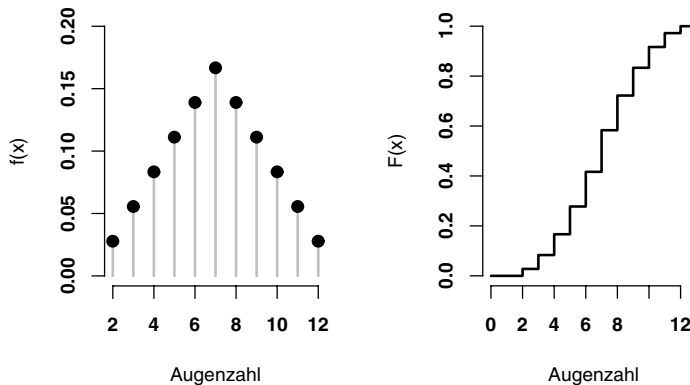


Abb. 5.3. Wahrscheinlichkeitsfunktion und Verteilungsfunktion für die Zufallsvariable: Augenzahl beim Werfen von zwei Würfeln

Es gibt einen weiteren Weg, die „Wahrscheinlichkeitsverteilung“ einer Zufallsvariablen zu beschreiben. Beispielsweise genügt es, im Würfelversuch die Wahrscheinlichkeiten anzugeben, mit der die betreffenden Augenzahlen geworfen werden $[P(X = x_i) = 1/6]$.

Allgemein bezeichnet man für diskrete Zufallsvariable die Zuordnung der Merkmale x_i zu den Wahrscheinlichkeiten $f(x_i)$ als **Wahrscheinlichkeitsfunktion** (probability function, frequency function). Für diskrete Zufallsvariable ermittelt man die Verteilungsfunktion durch einfaches Aufsummieren der Wahrscheinlichkeiten $f(x_i)$: $F(x) = \sum_i P(X = x_i)$ für $x_i \leq x$. Für stetige Zufallsvariable, also z. B. solche, deren Werte durch Längen-, Gewichts- oder Geschwindigkeits-

messungen zustande kommen, erhält man die Verteilungsfunktion durch Integration über die sog. **Wahrscheinlichkeitsdichte** (probability density function) oder Dichtefunktion. Sie legt die Verteilung ebenfalls eindeutig fest.

Verteilungsfunktionen geben Wahrscheinlichkeiten für Realisierungen einer Zufallsvariable im Bereich von $-\infty$ bis zu einer gewissen oberen Grenze x an: $F(x) = P(X \leq x)$. **Wahrscheinlichkeitsfunktionen** sind Zusammenstellungen der Wahrscheinlichkeiten für die *einzelnen* Realisierungen einer diskreten Zufallsvariable. Das Analogon dazu für stetige Zufallsvariable sind die **Wahrscheinlichkeitsdichten** oder **Dichtefunktionen**. Dichtefunktionen selbst sind keine Wahrscheinlichkeiten, dagegen sind Integrale einer Dichtefunktion Wahrscheinlichkeiten [vgl. (5.5), (5.8), Abbildung 5.4 und Übersicht 10].

Zwischen Wahrscheinlichkeitsfunktion bzw. Wahrscheinlichkeitsdichte und Verteilungsfunktion besteht der Zusammenhang [(5.4) und (5.5); für beide gilt: $F(-\infty) = 0$ und $F(\infty) = 1$; die Verwendung des Zeichens ∞ für unendlich geht zurück auf John Wallis, 1616–1703; zu (5.4): $F(x) = P(X \leq x) = \sum_{x_i \leq x} P_i$ mit $-\infty < x < \infty$]:

Diskrete Zufallsvariable: $F(x) = \sum_{x_i \leq x} f(x_i)$ $f(x_i)$ ist die Wahrscheinlichkeitsfunktion .	(5.4)
Stetige Zufallsvariable: $F(x) = \int_{-\infty}^x f(t) dt$ $f(t)$ ist die Wahrscheinlichkeitsdichte .	(5.5)

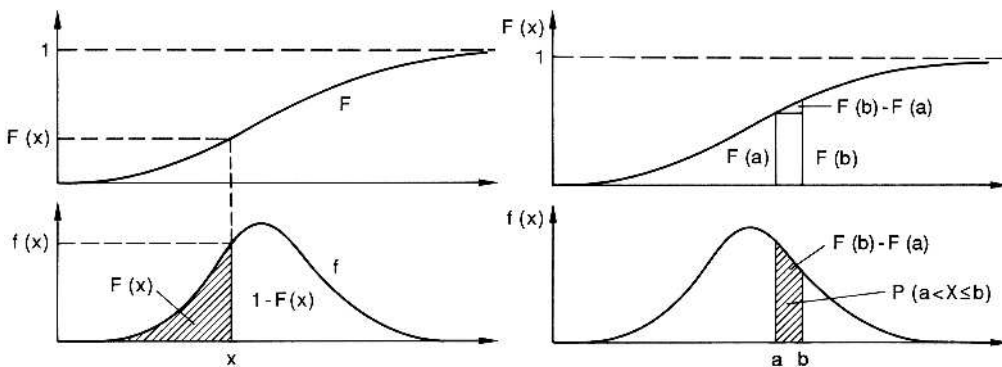


Abb. 5.4. Verteilungsfunktion F und Dichtefunktion f einer stetigen Zufallsvariablen X ; $f(x)$ ist der Wert der Dichtefunktion an der Stelle x , nach E. Walter [Walt88] modifiziert

Zur anschaulichen Bedeutung der Wahrscheinlichkeitsdichte ist zu sagen, dass für sehr kleine Intervalle dt die Wahrscheinlichkeit, dass X in das Intervall $(t, t + dt)$ fällt, näherungsweise durch das Differential $f(t)dt$ gegeben ist, das man auch als **Wahrscheinlichkeitselement** bezeichnet:

$$\boxed{f(t)dt \simeq P(t < X \leq t + dt)}$$
 (5.6)

Für jede Wahrscheinlichkeitsdichte gilt:

$$\int_{-\infty}^{+\infty} f(t)dt = 1 \quad (5.7)$$

und insbesondere gilt für die Wahrscheinlichkeit, dass die Zufallsvariable X mit der Verteilungsfunktion $F(x)$ einen Wert in dem Intervall größer als a und kleiner gleich b annimmt

$$P(a < X \leq b) = F(b) - F(a) = \int_a^b f(t)dt \quad (5.8)$$

Die Wahrscheinlichkeit des Ereignisses $a < X \leq b$ ist gleich der Fläche unter der Kurve der Wahrscheinlichkeitsdichte f zwischen $x > a$ und $x = b$.

Wichtige Eigenschaften und Zusammenhänge:

- (1) Die Dichtefunktion f beschreibt, wie die Gesamtwahrscheinlichkeit $P(-\infty \leq X \leq \infty) = 1$ über den Wertebereich der stetigen Zufallsvariablen X „verschmiert“ ist.
- (2) Für jede reelle Zahl x gilt: $f(x) \geq 0$ (f ist nicht negativ).
- (3) Die Wahrscheinlichkeit $P(a < x \leq b)$ entspricht der Fläche zwischen x -Achse und Wahrscheinlichkeitsdichte f oberhalb von a bis einschließlich b mit $b \geq a$:

$$P(a < X \leq b) = \int_a^b f(x)dx = F(b) - F(a)$$

$f(x)dx$ lässt sich als Wahrscheinlichkeit dafür auffassen, dass der Messwert x im Intervall dx auftritt [vgl. (5.6)].

- (4) Für die Gesamtfläche zwischen der Wahrscheinlichkeitsdichte f und der x -Achse muss wegen $F(\infty) = 1$

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

gelten (f ist normiert) [(5.7)].

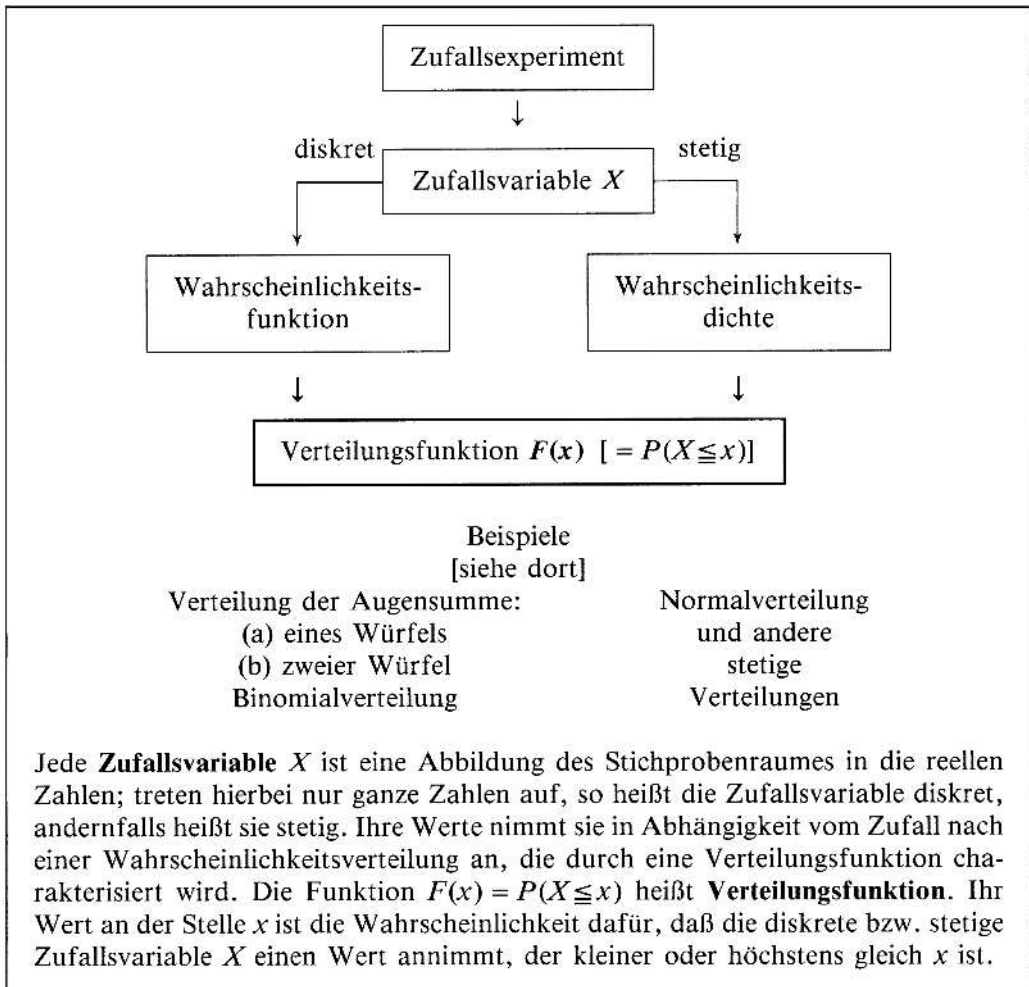
An jeder Stelle der Verteilung symbolisiert der Wert $F(x)$ den Inhalt des links von x liegenden Flächenstückes zwischen dem Graph der Dichtefunktion f und der x -Achse. Für $a = b$ wird $F(a) - F(b) = 0$, d. h. für alle x gilt die Punktwahrscheinlichkeit $P(X = x) = 0$. Dies gilt nur für stetige Zufallsvariablen (vgl. Abbildung 5.4).

Jetzt können wir auch diskrete und stetige Zufallsvariable (discrete random variables, continuous random variables) definieren:

1. Eine Zufallsvariable X , die nur endlich oder abzählbar viele Werte annehmen kann, nennen wir **diskret**. Diese Werte haben wir Sprungstellen genannt. Die zur Zufallsvariablen X gehörige Verteilungsfunktion weist abzählbar viele Sprungstellen auf (vgl. Abb. 5.2 und 5.3).
2. Eine Zufallsvariable X nennen wir **stetig**, wenn die zugehörige Verteilungsfunktion (5.3) in Integralform (5.5) dargestellt werden kann. Die Werte, die die stetige Variable X annehmen kann, bilden ein Kontinuum (vgl. Abb. 5.4).

Während die Wahrscheinlichkeit P eines bestimmten Ereignisses im Falle einer diskreten Verteilung meist bedeutungsvoll ist, kann dies im Falle einer stetigen Verteilung nicht behauptet werden (z. B. P , dass ein Ei 50,00123 g wiegt), daher interessieren hier Wahrscheinlichkeiten der Art, dass eine Variable X sagen wir $< a$ oder $\geq a$ ist [vgl. auch Formel (5.1), (5.2) und (5.3)].

Bei stetigen Zufallsvariablen mit der Prüfgröße X ist die Angabe eines Wertes $P(X \geq x)$ problemlos, nicht aber bei einer diskreten; man behilft sich dann durch Angabe eines **mittleren P -Wertes** (mid- P) $P_m = P(X > x) + 0,5(X = x)$.

Übersicht 10. Die Verteilungsfunktion**5.1.1.1 Rechenregeln zur Verteilungsfunktion für beliebige Zufallsvariablen**

Für die Zufallsvariable X mit der Verteilungsfunktion $F(x)$ an den Stellen $x = a$ und $x = b$ gelten die folgenden acht Beziehungen:

- (1) $P(X \leq a) = F(a)$
- (2) $P(X < a) = F(a) - P(X = a)$
- (3) $P(X > a) = 1 - F(a)$
- (4) $P(X \geq a) = 1 - F(a) + P(X = a)$
- (5) $P(a < X \leq b) = F(b) - F(a)$
- (6) $P(a \leq X \leq b) = F(b) - F(a) + P(X = a)$

$$(7) P(a < X < b) = F(b) - F(a) - P(X = b)$$

$$(8) P(a \leq X < b) = F(b) - F(a) + P(X = a) - P(X = b)$$

5.1.1.2 Empirische Verteilungsfunktion

$$\hat{F}_n(x) = \frac{1}{n} \cdot [\text{Anzahl der Stichprobenwerte } x_i \leq x] \quad (5.9)$$

Bezeichnet man die n ansteigend geordneten Werte mit $x_{(1)}, x_{(2)}, \dots, x_{(n)}$, dann lässt sich die empirische Verteilungsfunktion $\hat{F}_n(x)$ durch Auftragen von $i/(n+1)$ [oder $(i-0,5)/n$ oder $(i-3/8)/(n+1/4)$] gegen $x_{(i)}$ angeben. Graphisches Papier mit linearen Skalen: man erhält eine S-förmige Kurve [(5.9): Treppenfunktion mit Sprungstellen in $x_i, i = 1, \dots, n$], die empirische Verteilungsfunktion oder Stichproben-Verteilungsfunktion.

Satz von V.I. Glivenko und F.P. Cantelli: Die empirische Verteilungsfunktion $\hat{F}_n(x) = n_{\leq x}/n$, das ist der Anteil der Stichprobenwerte, die den Wert x nicht überschreiten, schätzt die Verteilungsfunktion $F(x)$ der Grundgesamtheit. Der Hauptsatz der Mathematischen oder Beurteilenden Statistik lautet:

$$\text{Die Differenz } |\hat{F}_n(x) - F(x)| \text{ strebt für großes } n \text{ mit der Wahrscheinlichkeit 1 gegen Null} \quad (5.10)$$

Empirische Verteilungen sind nach Glivenko-Cantelli für großen Stichprobenumfang praktisch gleich den theoretischen Verteilungen der Grundgesamtheiten.

Zur Anpassung einer Verteilungsfunktion an empirische Daten benötigt man einen mittleren Stichprobenumfang n ; ist n klein, so lassen sich unterschiedliche Verteilungsfunktionen gleichgut anpassen, ist n groß, so wird jede anhand eines Tests verworfen.

5.2 Maßzahlen zur Kennzeichnung der Verteilung

- Erwartungswert
- Varianz
- Momente: Schiefe und Exzess

Die Wahrscheinlichkeitsverteilung einer diskreten oder stetigen Zufallsvariablen ist eindeutig durch die Verteilungsfunktion $F(x)$ oder durch die Wahrscheinlichkeitsdichte (Wahrscheinlichkeitsfunktion) $f(x)$ beschrieben. Daneben gibt es spezielle Maßzahlen (Parameter), die das Zufallsexperiment qualitativ beschreiben. Man unterscheidet

- **Scharparameter**, Größen, die in der Formel der Wahrscheinlichkeits- oder der Verteilungsfunktion auftreten, und
- **Funktionalparameter**, Größen, die bestimmte Eigenschaften einer Wahrscheinlichkeitsverteilung kennzeichnen, wie z. B. Lage-, Streuungs- und Formparameter.

5.2.1 Erwartungswert

Beispiel: Ein spezieller Würfel, über dessen Qualität nichts bekannt ist, wird 8mal geworfen. Dabei werden die Augenzahlen 1, 2, 2, 3, 3, 3, 4, 6 beobachtet. Auf der Basis dieser Beobachtung kann eine zu erwartende Augenzahl empirisch durch den Mittelwert aus dem Zufallsexperiment berechnet werden:



$$\frac{1}{n} \sum_{i=1}^8 x_i = 24 \cdot \frac{1}{8} = 1 \cdot \frac{1}{8} + 2 \cdot \frac{2}{8} + 3 \cdot \frac{3}{8} + 4 \cdot \frac{1}{8} + 5 \cdot \frac{0}{8} + 6 \cdot \frac{1}{8} = \sum_{j=1}^6 x_j \frac{n_j}{n} = 3$$

Das bedeutet, dass eine gewichtete Summe der beobachteten Werte, der Realisierungen der Zufallsvariablen, gebildet wird, wobei die Gewichte entsprechend der relativen Häufigkeit gewählt werden. Diese Überlegung führt zur Definition des Erwartungswertes, wobei die Gewichtung mit den Wahrscheinlichkeiten für das Auftreten der entsprechenden Ereignisse, d.h. mit den Werten der Wahrscheinlichkeitsfunktion bzw. der Wahrscheinlichkeitsdichte, erfolgt.

1. Erwartungswert für diskrete Zufallsvariablen X :

$$E[X] = \mu = \sum_{i=1}^n x_i \cdot f(x_i) \quad (5.11)$$

2. Erwartungswert für stetige Zufallsvariablen X :

$$E[X] = \mu = \int_{-\infty}^{+\infty} x \cdot f(x) dx \quad (5.12)$$

Für den Erwartungswert wird allgemein der griechische Buchstabe μ (gelesen: mü) verwendet.

Beispiel: Die Augenzahl X eines regulären (guten) Würfels hat für das Zufallsexperiment eines einfachen Würfelwurfes den Erwartungswert



$$\mu = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3,5$$

Beispiel: Beim Wurf einer regulären (nicht gefälschten) Münze werde 1 Euro gesetzt. Erscheint das Ergebnis „Wappen“, dann erhält man den doppelten Einsatz zurück, anderenfalls ist der Einsatz verloren. Für die Zufallsvariable X = Gewinn, mit den Werten +1 Euro oder -1 Euro, ist der Erwartungswert somit



$$\mu = 1 \cdot \frac{1}{2} + (-1) \cdot \frac{1}{2} = 0$$

In diesem Fall spricht man auch von einem fairen Spiel.

Beispiel: Ein Zufallszahlengenerator erzeugt Zahlen zwischen 1 und 2 auf der Basis der folgenden Wahrscheinlichkeitsdichte.



$$f(x) = \begin{cases} 0 & \text{für } -\infty < x < 1 \\ 1 & \text{für } 1 \leq x \leq 2 \\ 0 & \text{für } 2 < x < +\infty \end{cases}$$

Da der Zufallszahlengenerator seine Werte mit gleicher Wahrscheinlichkeit im Bereich der reellen Zahlen zwischen 1 und 2 streuen soll, ist der Wert 1,5 zu erwarten:

$$\mu = \int_{-\infty}^1 x \cdot 0 dx + \int_1^2 x \cdot 1 dx + \int_2^{+\infty} x \cdot 0 dx = \frac{x^2}{2} \Big|_1^2 = 1,5$$

Hinweis: Hier wird zu der Funktion $x \cdot 1$ die Stammfunktion [vgl. (2.55) auf Seite 47] $x^2/2$ gebildet, mit deren Hilfe die Fläche unter der Kurve im Intervall $[1, 2]$ durch die Differenz der Werte der Stammfunktion bestimmt wird.



Beispiel: Die Lebensdauer eines speziellen Elektronikteils in einem Computer besitzt eine Verteilung der Form $F(x) = 1 - e^{-\frac{x}{10}}$. Die Wahrscheinlichkeitsdichte für eine entsprechende Zufallsvariable X - Lebensdauer in Jahren - ist dann definitionsgemäß $f(x) = F'(x) = \frac{1}{10}e^{-\frac{x}{10}}$. Diese hat den Erwartungswert:

$$\mu = \frac{1}{10} \int_0^{+\infty} x \cdot e^{-\frac{x}{10}} dx = \frac{1}{10} \cdot 100 = 10$$

Hinweis: Das Ergebnis folgt aus der Lösung des bestimmten Integrals $\int_0^{\infty} x \cdot e^{-ax} dx = 1/a^2$ mit $a = \frac{1}{10}$.

Rechenregeln für Erwartungswerte

$E[c]$	$= c$	c beliebige Konstante	
$E[c \cdot X]$	$= c \cdot E[X]$	X beliebige Zufallsvariable	
$E[a \cdot X + b]$	$= a \cdot E[X] + b$	a und b Konstanten	(5.13)
$E[X + Y]$	$= E[X] + E[Y]$	X und Y beliebige Zufallsvariablen	
$E[X \cdot Y]$	$= E[X] \cdot E[Y]$	X und Y sind stochastisch unabhängig!	

5.2.2 Varianz

Während der Erwartungswert ein Maß für die Erwartung an den Ausgang eines Zufallsexperimentes mit Bezug auf den Wert der Zufallsvariablen ist, soll mit der Varianz eine zu erwartende Streuung (Variation) in den Ausgängen bewertet werden. Dazu definiert man in Analogie zur empirischen Varianz aus der deskriptiven Statistik die Varianz der Zufallsvariablen wie folgt:

1. Varianz für diskrete Zufallsvariablen X :

$$\text{Var}[X] = \sigma^2 = \sum_{i=1}^n (x_i - \mu)^2 \cdot f(x_i) \quad (5.14)$$

2. Varianz für stetige Zufallsvariablen X :

$$\text{Var}[X] = \sigma^2 = \int_{-\infty}^{+\infty} (x - \mu)^2 \cdot f(x) dx \quad (5.15)$$



Damit drückt die Varianz den Erwartungswert der quadrierten Abweichungen vom Erwartungswert einer Zufallsvariablen aus. Die Wurzel aus der Varianz wird **Standardabweichung** (σ) genannt.

$$\begin{aligned}
 \sigma^2 &= E[(X - \mu)^2] \\
 &= E[(X^2 - 2X\mu + \mu^2)] \\
 &= E[X^2] - 2\mu \cdot E[X] + \mu^2 \\
 &= E[X^2] - \mu^2
 \end{aligned}
 \tag{5.16}$$

Für den praktischen Umgang mit und zur Berechnung von Varianzen zu speziellen Zufallsvariablen wird in der Regel (5.16) verwendet. Die Umformung basiert auf den Regeln zum Rechnen mit Erwartungswerten (5.13).

Beispiel: Bei einem Wurf mit zwei Würfeln ist die Varianz der Zufallsvariablen X „Summe der Augenzahlen“:



$$\begin{aligned}
 \sigma^2 &= \sum_{i=1}^n (x_i - \mu)^2 f(x_i) = \sum_{i=2}^{12} (x_i - 7)^2 f(x_i) \\
 &= 2(25 \frac{1}{36}) + 16 \frac{2}{36} + 9 \frac{3}{36} + 4 \frac{4}{36} + 1 \frac{5}{36} = \frac{210}{36} = 5,83
 \end{aligned}$$

Beispiel: Ein Zufallszahlengenerator, der reelle Zahlen zwischen 0 und 10 mit gleicher Wahrscheinlichkeit erzeugt ($f(x) = 1/10$ im Intervall $[0, 10]$), hat den Erwartungswert $\mu = 5$ und die Varianz:



$$\sigma^2 = E[X^2] - \mu^2 = \int_0^{10} x^2 \cdot \frac{1}{10} dx - 25 = \frac{1}{10} \left[\frac{x^3}{3} \right]_0^{10} - 25 = 33 \frac{1}{3} - 25 = 8 \frac{1}{3}$$

Hinweis: Unter Verwendung von (5.16) wird hier zu der Funktion x^2 die Stammfunktion [vgl. (2.55) auf Seite 47] $x^3/3$ gebildet, mit deren Hilfe die Fläche unter der Kurve im Intervall $[0, 10]$ durch die Differenz der Werte der Stammfunktion bestimmt wird.

Rechnen mit Varianzen

$Var[c]$	$= 0$	c beliebige Konstante
$Var[c \cdot X]$	$= c^2 \cdot Var[X]$	X beliebige Zufallsvariable
$Var[a \cdot X + b]$	$= a^2 \cdot Var[X]$	a und b Konstanten
$Var[X \pm Y]$	$= Var[X] + Var[Y]$	X und Y Zufallsvariablen, die stochastisch unabhängig sind

(5.17)

Anstelle eines Beweises von (5.17) sei darauf hingewiesen, dass

- die Streuung einer Konstanten, d.h. hier einer Zufallsvariablen, die nur einen Wert annehmen kann, gleich Null ist.
- die Streuung einer Zufallsvariablen, mit dem Faktor c multipliziert, in den quadrierten Abweichungen vom Erwartungswert mit dem Quadrat dieses Faktors berücksichtigt werden muss.
- die Streuung sich nicht ändert, wenn der Wertebereich der Zufallsvariablen um einen konstanten Wert verschoben wird.
- in die Summe oder Differenz von Streuungen zweier Zufallsvariablen eine so genannte Kovarianz eingeht, die gleich Null ist, wenn die Zufallsvariablen stochastisch unabhängig sind.

5.2.2.1 Ungleichung von Tschebyscheff

Eine Zufallsvariable X habe den Erwartungswert μ und die Varianz σ^2 . Die Wahrscheinlichkeit P , dass die Zufallsvariable X Werte x_i in einem symmetrisch zu μ liegenden Bereich annimmt,

der von $\mu - k\sigma$ bis $\mu + k\sigma$ reicht (einschließlich dieser Grenzen), wobei k eine positive reelle Zahl ($k > 0$) ist, wird durch die Ungleichung von Tschebyscheff bestimmt.

$$P(\mu - k\sigma \leq X \leq \mu + k\sigma) = P(|X - \mu| \leq k\sigma) \geq 1 - \frac{1}{k^2} \quad (5.18)$$

Für eine beliebige Verteilung einer Zufallsvariablen X , die den Erwartungswert μ und die Varianz σ hat, sind mindestens 75% der Ergebnisse in dem Bereich $\mu \pm 2\sigma$ zu erwarten.

$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) \geq 1 - \frac{1}{4} = 0,75$$

Diese Abschätzung wird genauer, wenn hinsichtlich der Verteilung zusätzliche Annahmen gemacht werden können, z.B. hinsichtlich einer symmetrisch-eingipfligen Verteilung oder einer Normalverteilung.

5.2.3 Momente: Schiefe und Exzess

Wichtige Funktionen einer Zufallsvariablen in der Statistik sind auch die Potenzen X^k . Der Erwartungswert dieser Potenz, $\mu'_k = E[X^k]$, heißt **k-tes Moment der Zufallsvariablen**.

$$\begin{aligned} \mu'_1 &= \mu \\ \mu'_2 &= \sigma^2 + \mu^2 \\ \mu'_3 &= 3\mu\sigma^2 + \mu^3 \\ \mu'_4 &= \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4 \end{aligned} \quad (5.19)$$

Das **erste Moment einer Zufallsvariablen entspricht somit dem Erwartungswert** ($\mu'_1 \equiv \mu$). Das zweite Moment einer Zufallsvariablen kann direkt aus der Definition für die Varianz abgeleitet werden. Aus $\sigma^2 = E[(X - \mu)^2] = E(X^2) - \mu^2$ folgt direkt $\mu'_2 = \sigma^2 + \mu^2$. Über die empirischen Momente können die Stichprobenvarianz und andere Verteilungscharakteristika isoliert beschrieben werden.

Werden Momente auf die Abweichungen vom Erwartungswert bezogen, dann spricht man von **zentralen Momenten** (zentrierte Momente).

$$\textbf{k-tes zentrales Moment: } \mu_k = E[(X - \mu)^k] \quad (5.20)$$

- Das 1. zentrale Moment hat den Wert 0, wie sich leicht aus den Regeln zum Rechnen mit Erwartungswerten schließen lässt.
- Das 2. zentrale Moment ist die Varianz (definitionsgemäß).
- Das 3. zentrale Moment charakterisiert die **Symmetrie / Asymmetrie** einer Verteilung und wird **Schiefe** (skewness) genannt.

Schiefe für diskrete / stetige Zufallsvariablen X :

$$\begin{aligned} \gamma_1 &= \frac{\sum_{i=1}^n (x_i - \mu)^3 \cdot f(x_i)}{\sigma^3} \\ \gamma_1 &= \frac{\int_{-\infty}^{+\infty} (x - \mu)^3 \cdot f(x) dx}{\sigma^3} \end{aligned} \quad (5.21)$$

Dabei erfolgt die Normierung mit Hilfe von σ^3 , um eine dimensionslose Kennzeichnung der Schiefe zu ermöglichen. Ist die Verteilung einer Zufallsvariablen symmetrisch zum Erwartungswert μ , dann hat die Schiefe den Wert Null. Liegt der Hauptanteil einer Verteilung auf

der linken Seite des Erwartungswertes konzentriert, dann spricht man ihr eine positive Schiefe zu ($\gamma_1 > 0$) und nennt sie **linksteil**. Anderenfalls, d.h. wenn gilt $\gamma_1 < 0$, bezeichnet man die Verteilung als **rechtssteil**. (Bemerkung: Linksteile Verteilungen kommen in der Biologie und Medizin wesentlich häufiger vor als rechtssteile Verteilungen.)

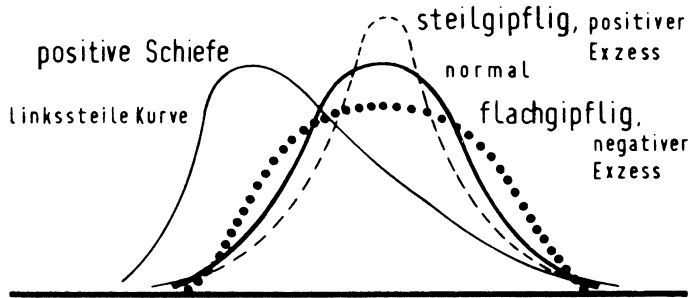


Abb. 5.5. Abweichungen von der symmetrischen Glockenkurve (Normalverteilung)

- Das 4. zentrale Moment charakterisiert die **Wölbung / Steilheit** einer Verteilung und wird **Exzess** (kurtosis) genannt.

Exzess für diskrete / stetige Zufallsvariable X :

$$\gamma_2 = \frac{\sum_{i=1}^n (x_i - \mu)^4 \cdot f(x_i)}{\sigma^4} - 3 \quad (5.22)$$

$$\gamma_2 = \frac{\int_{-\infty}^{+\infty} (x - \mu)^4 \cdot f(x) dx}{\sigma^4} - 3$$

Dabei erfolgt auch hier eine Normierung mit Hilfe von σ^4 , um eine dimensionslose Kennzeichnung der Steilheit einer Verteilung zu ermöglichen. Die zusätzliche **Zentrierung** (Subtraktion von 3) erfolgt, da die Form der Normalverteilung (s.d.) als Referenz herangezogen wird, die den Exzess 3 hat.

5.2.3.1 Berechnung der empirischen Momente

Die Berechnung des empirischen 1. und 2. zentralen Momentes, arithmetischer Mittelwert und empirische Varianz, ist ausführlich im Kapitel 3 (deskriptive Statistik) beschrieben. Für das **empirische 3. Moment** gilt entsprechend:

$$g_1 = \begin{cases} \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{s^3} \\ \frac{\frac{1}{n} \sum_{i=1}^k n_i (x_i - \bar{x})^3}{s^3} \end{cases} \quad (5.23)$$

Im unteren Ansatz wird auf Häufigkeiten n_i klassierter Beobachtungen Bezug genommen.

Die Größe der drei Maßzahlen zur zentralen Lage (Mittelwert ($\bar{x}$), Modalwert (D), Medianwert ($\tilde{x}$)) zeigt folgenden Zusammenhang mit der Schiefe einer Verteilung:

$$\begin{cases} g_1 > 0 & \text{linkssteile Verteilung:} & D < \bar{x}, \quad \tilde{x} < \bar{x}, \quad D < \tilde{x} \\ g_1 = 0 & \text{symmetrische Verteilung:} & D = \bar{x}, \quad \tilde{x} = \bar{x}, \quad D = \tilde{x} \\ g_1 < 0 & \text{rechtssteile Verteilung:} & D > \bar{x}, \quad \tilde{x} > \bar{x}, \quad D > \tilde{x} \end{cases}$$

Ein auf empirischen Momenten basierendes Exzessmaß ist das 4. Moment:

$$g_2 = \begin{cases} \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{s^4} - 3 \\ \frac{\frac{1}{n} \sum_{i=1}^k n_i (x_i - \bar{x})^4}{s^4} - 3 \end{cases} \quad (5.24)$$

Eine Verteilung heißt

- **platykurtisch** (flach gewölbt), wenn $g_2 < 0$.
- **mesokurtisch** (mittelmäßig gewölbt), wenn $g_2 = 0$.
- **leptokurtisch** (stark gewölbt), wenn $g_2 > 0$.

Für die Berechnung dieser (zentrierten!) Momente in **R** stehen in dem Paket **library(e1071)** (Dimitriadou [DHL⁺05]) die Funktionen **skewness()** und **kurtosis()** zur Verfügung:

```
> library(e1071)
> x <- c(2,3,4,4,4,5,5,5,5,6,8,10,20,40)
> skewness(x) # empirisches 3tes Moment (relativ)
[1] 2.198071
> kurtosis(x) # empirisches 4tes Moment (rel., zentriert)
[1] 3.89879
```

5.2.3.2 Potenzmomente

Für eine symmetrische Verteilung gilt $\gamma_1 = 0$, für die $N(0;1)$ -Verteilung $\gamma_2 = 0$. Ist γ_1 positiv, dann liegt eine linkssteile Verteilung vor, bei negativen Werten eine rechtssteile Verteilung. Eine Verteilung mit Hochgipfligkeit – steiler als die Normalverteilung – oder positivem Exzess weist einen positiven Wert γ_2 auf; eine Verteilung mit negativer Wölbung – flacher als die Normalverteilung – ist durch einen negativen Wert γ_2 charakterisiert, der, genau genommen, „peakedness combined with tailedness“ oder „lack of shoulders“ misst und daher bei einer bimodalen Kurve stark negativ ist (Finucan [Fin64], vgl. auch Chissom [Chi70] und Darlington [Dar70]). Die Rechteckverteilung mit ausgeprägter „Schulterpartie“ hat daher auch eine negative Wölbung ($\gamma_2 = -1,2$). Dies gilt sogar für jede Dreieckverteilung ($\gamma_2 = 0,6$), die gegenüber einer Normalverteilung mit gleicher Varianz eine stärker ausgebildete „Schulterpartie“ aufweist. Zunächst eine Bemerkung über **Potenzmomente**. Größen der Form

$$\frac{\sum f_i (x_i - \bar{x})^r}{n} = m_r \quad (5.25)$$

bezeichnet man als **Potenzmomente r-ten Grades** (m_r) der Stichprobe. Für $r = 2$ ergibt sich aus (5.25) angenähert die Stichprobenvarianz. Die beiden Momentenkoeffizienten lassen sich abgekürzt als

$$g_1 = m_3/s^3 \quad \text{und} \quad g_2 = m_4/s^4 - 3$$

schreiben. Ist die Klassenbreite nicht gleich eins ($b \neq 1$), so wird

$$m_r = \frac{\sum f_i \left(\frac{x_i - \bar{x}}{b} \right)^r}{n} \quad (5.26)$$

Zur Erleichterung der Rechnung war es üblich, die Potenzmomente nicht auf das arithmetische Mittel, sondern auf einen beliebigen Ursprung, sagen wir auf den Wert d , der am stärksten besetzten Klasse einer Häufigkeitsverteilung, zu beziehen. Die so erhaltenen Momente bezeichnen wir zur Unterscheidung von m_r mit m'_r . Nennen wir $(x - d)/b = z$, so erhalten wir für unsere eigentlich zu kleine Stichprobenverteilung aus Tabelle 5.1 ($n = 40$) die Potenzmomente ersten bis vierten Grades der Stichprobe nach

Moment 1. Grades

$$\begin{aligned} m'_1 &= \frac{\sum f_i \cdot z_i}{n} \\ &= \frac{18}{40} = 0,45 \end{aligned} \quad (5.27)$$

Moment 2. Grades

$$\begin{aligned} m'_2 &= \frac{\sum f_i \cdot z_i^2}{n} \\ &= \frac{110}{40} = 2,75 \end{aligned} \quad (5.28)$$

Moment 3. Grades

$$\begin{aligned} m'_3 &= \frac{\sum f_i \cdot z_i^3}{n} \\ &= \frac{216}{40} = 5,40 \end{aligned} \quad (5.29)$$

Moment 4. Grades

$$\begin{aligned} m'_4 &= \frac{\sum f_i \cdot z_i^4}{n} \\ &= \frac{914}{40} = 22,85 \end{aligned} \quad (5.30)$$

Zur Kontrolle der Rechnung enthält Tabelle 5.1 noch eine Spalte mit den Produkten $f_i \cdot (z_i + 1)^4$. Die Spaltensummen lassen sich dann nach

$$\sum f_i (z_i + 1)^4 = \sum f_i + 4 \sum f_i z_i + 6 \sum f_i z_i^2 + 4 \sum f_i z_i^3 + \sum f_i z_i^4 \quad (5.31)$$

$2550 = 40 + 72 + 660 + 864 + 914$ leicht überprüfen. Hieraus ergeben sich die Kennwerte der Verteilung:

Tabelle 5.1.

x_i	f_i	z_i	$f_i z_i$	$f_i z_i^2$	$f_i z_i^3$	$f_i z_i^4$	$f_i (z_i + 1)^4$
8,8	4	-2	-8	16	-32	64	4
9,3	8	-1	-8	8	-8	8	0
$d=9,8$	11	0	0	0		0	11
10,3	7	1	7	7	7	7	112
10,8	5	2	10	20	40	80	405
11,3	3	3	9	27	81	243	768
11,8	2	4	8	32	128	512	1250
Summe	40		18	110	216	914	2550

1. Mittelwert

$$\bar{x} = d + b m'_1 \quad (5.32)$$

$$\bar{x} = 9,8 + 0,5 \cdot 0,45 = 10,025$$

2. „Varianz“

$$s^2 = b^2 (m'_2 - m'^2_1) \quad (5.33)$$

$$s^2 = 0,5^2 (2,75 - 0,45^2) = 0,637$$

3. Schiefe

$$g_1 = \frac{b^3 (m'_3 - 3m'_1 m'_2 + 2m'^3_1)}{s^3} \quad (5.34)$$

$$g_1 = \frac{0,5^3 \cdot (5,40 - 3 \cdot 0,45 \cdot 2,75 + 2 \cdot 0,45^3)}{0,5082} = 0,460$$

4. Wölbung

$$g_2 = \frac{b^4 \cdot (m'_4 - 4 \cdot m'_1 m'_3 + 6 \cdot m'^2_1 m'_2 - 3 \cdot m'^4_1)}{s^4} - 3 \quad (5.35)$$

$$g_2 = \frac{0,5^4 \cdot (22,85 - 4 \cdot 0,45 \cdot 5,40 + 6 \cdot 0,45^2 \cdot 2,75 - 3 \cdot 0,45^4)}{0,4055} - 3 = -0,480$$

Für die Berechnung der Potenzmomente 2. und 4. Grades sollte man beim Vorliegen sehr großer Stichprobenumfänge und auch nur dann, wenn die Stichprobenverteilung keine Asymmetrie aufweist, eine nach Sheppard korrigierte Varianz wählen (b = Klassenbreite):

$$s_{\text{kor}}^2 = s^2 - b/12 \quad (5.36)$$

$$m'_{4,\text{kor}} = m'_4 - (1/2)m'_2 b^2 + (7/240)b^4 \quad (5.37)$$

Die Berechnungen können in **R** schrittweise elementar wie folgt nachvollzogen werden.

```
> x <- c(8.8, 9.3, 9.8, 10.3, 10.8, 11.3, 11.8)
> d <- 9.8; b <- 0.5
> f <- c(4, 8, 11, 7, 5, 3, 2)
> z <- (x - d) / b
> n <- sum(f)
>
```

```

> m1 <- sum(f*z) / n; m1
[1] 0.45
> m2 <- sum(f*z^2) / n; m2
[1] 2.75
> m3 <- sum(f*z^3) / n; m3
[1] 5.4
> m4 <- sum(f*z^4) / n; m4
[1] 22.85
>
> mittelwert <- d + b*m1; mittelwert
[1] 10.025
> varianz <- b^2*(m2 - m1^2); varianz
[1] 0.636875
> schiefe <- (b^3*(m3 - 3*m1*m2 + 2*m1^3))/varianz^(3/2); schiefe
[1] 0.4598458
> woelbung <- (b^4*(m4 - 4*m1*m3 + 6*m1^2*m2 - 3*m1^4))/varianz^2 - 3; woelbung
[1] -0.4809175

```

Ergänzend soll die Berechnung dieser Maßzahlen noch mit den in **R** verfügbaren Funktionen gezeigt werden, die jedoch insbesondere bei der Wölbung zu abweichenden Ergebnissen führen.

```

> library(e1071)
> x1 <- c(rep(8.8,4), rep(9.3,8), rep(9.8,11), rep(10.3,7), rep(10.8,5),
+       rep(11.3,3), rep(11.8, 2))
> mean(x1); var(x1); skewness(x1); kurtosis(x1)
[1] 10.025 [1] 0.6532051 [1] 0.4427098 [1] -0.6052972

```

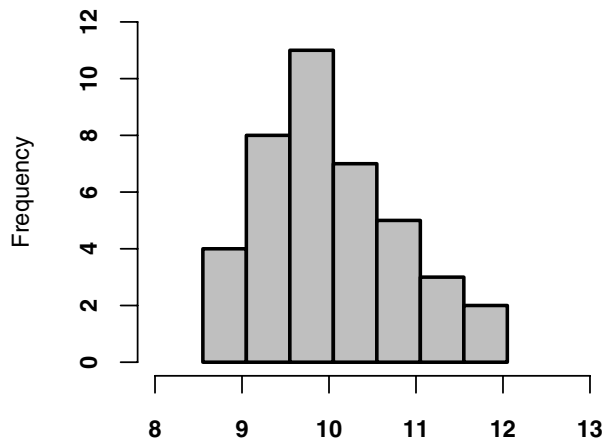


Abb. 5.6. Histogramm zur Verteilung der Daten aus Tabelle 5.1

Die über die Potenzmomente errechneten Maße für die Schiefe und den Exzess haben den Vorteil, dass die Standardfehler bekannt sind. Allerdings sind die Ausdrücke sehr unhandlich. Weitere Details geben Stuart und Ord ([SO94], Chapter 2: Lage- und Streuungsmaße, Chapter 3: Momentenmethode).

Tukeys Fünferregel besagt: man berechne das k -te Moment erst dann, wenn mindestens 5^k Beobachtungen vorliegen, d. h., der Mittelwert sollte auf $n \geq 5$ Beobachtungen basieren, die Varianz auf 25 Beobachtungen, usw.



Hinweis: Die Potenzmomenten-Methode hat Karl Pearson (1857–1936) eingeführt. Von ihm stammen auch die Begriffe Standardabweichung, Histogramm und Normalverteilung.

5.2.3.3 Quantilmaße zu Schiefe und Exzess

Schiefe und Exzess einer Verteilung können auch auf der Basis von Quantilen beurteilt werden.

$$\text{Skewness} = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1} \quad (5.38)$$

$$\text{Kurtosis} = \frac{(A_7 - A_5) + (A_3 - A_1)}{A_6 - A_2} \quad (5.39)$$

Formel (5.38) kennzeichnet den **Bowley-Koeffizient** (siehe Zar [Zar99]) für die Schiefe auf der Grundlage der **Quantile** Q_1 , Q_2 und Q_3 . Diese Maßzahl liegt zwischen -1 (extrem linkssteil), 0 (symmetrisch) und $+1$ (extrem rechtssteil).

Formel (5.39) kennzeichnet eine Maßzahl für den Exzess (die Wölbung) (Moors [Moo88]) auf der Grundlage der **Oktile** ("8" : $A_1, \dots, A_7$), d.h. einer Unterteilung in acht Stufen zu je 12,5%. Diese Maßzahl liegt zwischen 0 (extrem platykurtisch), $1,233$ (mesokurtisch) und $+\infty$ (extrem leptokurtisch).

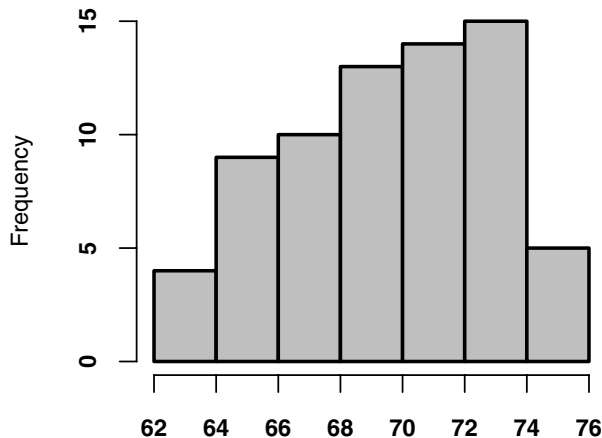


Abb. 5.7. Körpergröße [inch] von 70 Studenten (1inch=2,54cm); 7 Klassen für $n=70$ Beobachtungen sind optimal

Ein Beispiel in **R**, mit Körpergrößen von Studenten aus einem Statistikkurs (gemessen in inch, Abbildung 5.7), soll die Berechnung und Interpretation von Schiefe- und Wölbungsmaßzahlen verdeutlichen. Die Quantile, hier Quartile und Oktile geben einen guten Einblick in die Form einer Verteilung. Sie sind auch oft aussagekräftiger als Mittelwert und Standardabweichung, die durch Extremwerte stark beeinflusst werden. Bei mehrgipfligen Verteilungen gibt man zusätzlich auch die lokalen (sekundären) Dichtemittel und die Tiefpunkte an.

```
> y <- c(63, 63, 64, 64, rep(65,4), rep(66,5), rep(67,4), rep(68,6), rep(69,5),
+ rep(70,8), rep(71,7), rep(72,7), rep(73,10), rep(74,5), rep(75,3), rep(76,2))
> mean(y) # Mittelwert
[1] 70.04286
> var(y) # empirische Varianz
[1] 11.11408
> skewness(y) # empirisches 3tes Moment (relativ)
[1] -0.2843902
> kurtosis(y) # empirisches 4tes Moment (rel., zentriert)
```

```
[1] -0.8728042
> Q <- quantile(y, probs=seq(0, 1, 0.25), names=TRUE, type=7); Q # Quartile
 0% 25% 50% 75% 100%
 63 68 70 73 76
> Q1 <- as.numeric(Q[2]); Q2 <- as.numeric(Q[3]); Q3 <- as.numeric(Q[4])
> skew <- (Q3 + Q1 - 2*Q2)/(Q3-Q1); skew
[1] 0.2
> A <- quantile(y, probs=seq(0, 1, 0.125), names=TRUE, type=7); A # Oktile
 0% 12.5% 25% 37.5% 50% 62.5% 75% 87.5% 100%
 63 66 68 69 70 72 73 74 76
> A7 <- as.numeric(A[8]); A6 <- as.numeric(A[7]); A5 <- as.numeric(A[6])
> A3 <- as.numeric(A[4]); A2 <- as.numeric(A[3]); A1 <- as.numeric(A[2])
> kurt <- ((A7 - A5) + (A3 - A1))/(A6-A2); kurt
[1] 1
```

Damit sind wir nun in der Lage, eine eindimensionale Häufigkeitsverteilung ausführlich zu beschreiben. **Zur Übersicht ausreichend und für jeden Verteilungstyp geeignet sind:** $[x_{\min}]$, Q_1 , $\tilde{x}$, Q_3 , $[x_{\max}]$ und die aus ihnen gebildeten Maße (Tukey's five numbers). In R können diese einfach mit der Funktion `fivenum()` bestimmt werden.

```
> fivenum(y)
[1] 63 68 70 73 76
```

5.3 Diskrete Verteilungen

- Das Urnenmodell
- Gleichverteilung
- Binomialverteilung
- Poisson-Verteilung
- Negative Binomialverteilung
- Hypergeometrische Verteilung

5.3.1 Das Urnenmodell

Urnenmodelle garantieren die Realisierbarkeit einer reinen Zufallsstichprobe; außerdem ist dieses Experiment bei endlicher oder unendlicher Grundgesamtheit beliebig oft wiederholbar.

Münzen, Würfel und Karten sind die Elemente von Glücksspielen. Da sich jedes zufallsbeeinflusste Experiment oder jede zufallsartige Massenerscheinung näherungsweise durch ein Urnenmodell darstellen lässt, kann man, anstatt eine ideale Münze in die Luft zu werfen, auch Kugeln aus einer Urne ziehen, die genau zwei vollkommen gleiche Kugeln enthält, von denen die eine mit einem W und die andere mit einem Z (Wappen und Zahl) bezeichnet ist. Anstatt mit einem unverfälschten Würfel zu würfeln, können wir Kugeln aus einer Urne ziehen, die genau sechs mit 1, 2, 3, 4, 5 oder 6 Augen versehene Kugeln enthält. Anstatt eine Karte aus einem Kartenspiel zu ziehen, können wir Kugeln aus einer Urne ziehen, die genau 52 durchnummerierte Kugeln enthält. Diese Nummern sind ein Merkmal, eine definierte Eigenschaft einer statistischen Einheit, der Kugel. Die Kugeln, unsere Beobachtungs- oder Untersuchungseinheiten, die eine statistische Masse oder **Grundgesamtheit** bilden, heißen Merkmalsträger. Die verschiedenen Nummern der Kugeln, allgemein: der Kategorien, Abstufungen oder Werte eines Merkmals, heißen Merkmalsausprägungen; dokumentierte Merkmalsausprägungen sind **Daten**.

Aufgabe der Beurteilenden oder Schließenden Statistik ist es, aufgrund einer oder mehrerer Zufallsstichproben aus einer Urne Schlüsse zu ziehen hinsichtlich der Zusammensetzung des Inhaltes

(der Grundgesamtheit) dieser Urne. Diese Schlüsse sind Wahrscheinlichkeitssaussagen. Grundlage des statistischen Schlusses ist die **Wiederholbarkeit der Zufallsstichprobe** (Zufallsauswahl bedeutet streng genommen: „ein Ziehen mit Zurücklegen“). Die 52 Kugeln bilden die **Grundgesamtheit**. Wird der Urneninhalt gut durchgemischt („**randomisiert**“), dann erhält jedes Element der Grundgesamtheit, jede Kugel also, die gleiche Chance gezogen zu werden. Wir sprechen von dem Zufallscharakter der Stichprobe, von der zufälligen Stichprobe (random sample), kurz von der **Zufallsstichprobe**. Die Anzahl ausgewählter Elemente – 1 bis maximal 51 Kugeln – wird als **Stichprobenumfang** bezeichnet. Die Gesamtheit der möglichen Stichproben bildet den sog. **Stichprobenraum**. Die relative Häufigkeit der Spielkarten-Merkmale in der Grundgesamtheit ist die Wahrscheinlichkeit dieser Merkmale, gezogen zu werden: sie beträgt für die einer beliebigen Spielkarte entsprechenden Kugel $1/52$, für die den vier Königen entsprechenden Kugeln $4/52 = 1/13$ usw.

Demgegenüber ist die **relative Häufigkeit** der Merkmale in der Stichprobe eine **Schätzung der Wahrscheinlichkeit** dieser Merkmale. Die Schätzung ist um so genauer, je umfangreicher die Stichprobe ist. Vorausgesetzt werden **unabhängige Beobachtungen**. Bei endlichen Grundgesamtheiten ist die Unabhängigkeit dann gegeben, wenn nach jeder Einzelentnahme das entnommene Element wieder in die Grundgesamtheit zurückgelegt und neu gemischt wird: Urnenmodell der **Stichprobenentnahme mit Zurücklegen**. Die Zahl der Stichproben kann deshalb als unendlich groß angesehen werden, ein wichtiges Konzept der Beurteilenden Statistik.

Wird nach jeder Einzelentnahme aus einer endlichen Grundgesamtheit das entnommene Element nicht wieder zurückgelegt: **Urnenmodell ohne Zurücklegen**, so ändert sich laufend die Zusammensetzung der Restgesamtheit. Jede Beobachtung wird damit von der vorhergehenden abhängig. Wir sprechen von Wahrscheinlichkeitsansteckung oder von Wahrscheinlichkeitsverkettung. Modelle dieser Art werden durch so genannte **Markoffsche Ketten** (A. A. Markoff: 1856–1922) beschrieben: Jede Beobachtung ist nur von einer oder einer beschränkten Anzahl unmittelbar vorhergehender Beobachtungen abhängig. Diese und andere Klassen von Folgen nicht als unabhängig vorausgesetzter Zufallsvariablen in der Zeit bilden das mathematisch Interessierten vorbehaltene Gebiet der zufallsbedingten oder **stochastischen Prozesse**. Erwähnt seien Brownsche Molekularbewegung, Diffusionserscheinungen, Geburts-, Absterbe- und Einwanderungsprozesse; Theorie der Warteschlangen (Bedienungstheorie) und Servicesysteme.

Wenden wir uns wieder dem Urnenmodell der Stichprobenentnahme mit Zurücklegen zu. Die Verteilung der Wahrscheinlichkeiten auf die Merkmalsausprägungen bezeichnen wir als **Wahrscheinlichkeitsverteilung**, kurz als Verteilung. Charakteristische Größen von Verteilungen werden als Kenn- oder Maßzahlen bezeichnet. Maßzahlen wie relative Häufigkeit, Mittelwert oder Standardabweichung, die sich auf die Grundgesamtheit beziehen, bezeichnet man als **Parameter**. Die aus Zufallsstichproben errechneten Zahlenwerte heißen **Schätzwerte**. Parameter werden meistens mit griechischen Buchstaben bezeichnet (eine Tabelle mit dem griech. Alphabet befindet sich auf der Umschlaginnenseite), Schätzwerte durch lateinische Buchstaben.

So sind die Symbole für die relative Häufigkeit, Mittelwert und Standardabweichung, bezogen auf die Grundgesamtheit: π (pi), μ (mü), σ (sigma) – bezogen auf die Stichprobe: $\hat{p}$, $\bar{x}$ und s ; z. B. gilt nach Jakob Bernoulli (1713): $x/n = \hat{p} \xrightarrow[n \text{ groß}]{} \pi$.

Werden diese Werte aus Stichproben berechnet, die keine Zufallsstichproben sind, dann liegen keine Schätzwerte vor, sondern lediglich zur Beschreibung dienende Kenn- oder Maßzahlen. **Statistische Maßzahlen** („Statistiken“) zur zusammenfassenden Beschreibung von Daten sind allein eine Funktion der Daten; etwa der kleinste Wert oder die Summe der Daten. Statistische Maßzahlen sind gut, wenn man aufgrund der Originaldaten und aufgrund der Maßzahlen zu denselben Schlüssen gelangt.

5.3.2 Gleichverteilung

Wirft man einen Würfel, so kann die Anzahl der geworfenen Augen 1, 2, 3, 4, 5 oder 6 betragen. Dies gibt eine theoretische Verteilung, bei der die Werte 1 bis 6 die *gleiche* Wahrscheinlichkeit $1/6$ besitzen, d. h. $P(x) = 1/6$ für $x = 1, 2, \dots, 6$.

Die **diskrete Gleichverteilung** (engl. uniform distribution) ist definiert durch die folgende Wahrscheinlichkeits- und Verteilungsfunktion.

$$P(X = k) = 1/m \quad \text{für } k = 1, 2, \dots, m$$

$$F(x) = \begin{cases} 0 & \text{für } x < 1 \\ \frac{k}{m} & \text{für } k \leq x < k+1; \quad 1 \leq k < m \\ 1 & \text{für } x > m \end{cases} \quad (5.40)$$

Jeder Ausgang des Zufallsexperimentes tritt mit derselben Wahrscheinlichkeit auf. Am Beispiel des „**Urnenmodells**“ kann eine Urne (=Gefäß) betrachtet werden, in der sich Kugeln (m) gleicher Größe und Schwere befinden, die nur durch unterschiedliche Farben zu unterscheiden sind. Aus dieser Urne werden nun Kugeln „**unter Zurücklegen**“ gezogen. Somit hat jede Farbe bei jeder Ziehung die gleiche Wahrscheinlichkeit gezogen zu werden.

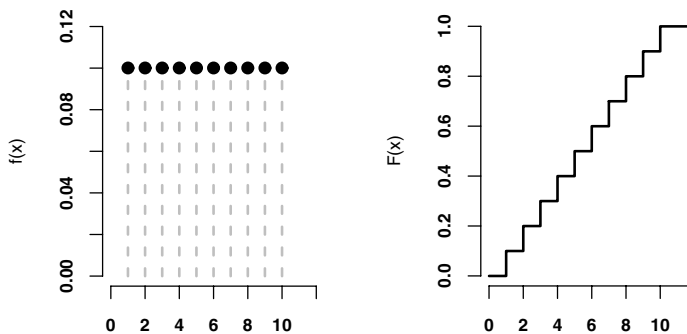


Abb. 5.8. Diskrete Gleichverteilung ($m=10$)

Für zahlreiche Verteilungsmodelle stehen im Statistikprogramm **R** spezielle Funktionen zur Verfügung, mit denen die wichtigsten Aspekte gezielt berechnet werden können. Dabei wird in der Benennung dieser Funktionen einheitlich über den **ersten Buchstaben** der jeweilige **Funktions-
typ** festgelegt:



- d** - **Wahrscheinlichkeitsdichte** (density function)
- p** - **Verteilungsfunktion** (engl. probability function)
- q** - **Quantilfunktion** (quantile function)
- r** - **Zufallszahlengenerator** (random number generator)

Darauf folgt eine kurze Kennzeichnung der Verteilung. Für diskrete Zufallsvariablen gibt es entsprechende Funktionen in dem Paket **library(e1071)** (Dimitriadou [DHL⁺05]) unter dem Namen **discrete()**. An dem Beispiel einer diskreten Gleichverteilung mit $m = 10$ (vgl. Abbildung 5.8) wird die Benutzung dieser Funktionen gezeigt.

```
> library(e1071)
> ddiscrete(1:10, rep(0.1, 10)) # Dichtefunktion
```

```
[1] 0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1 0.1
> pdiscrete(1:10, rep(0.1, 10)) # Verteilungsfunktion
[1] 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
> qdiscrete(c(0.25, 0.5, 0.75), rep(0.1, 10)) # Quantilfunktion (Quartile)
[1] 3 5 8
> rdiscrete(20, 1:10) # Zufallszahlen
[1] 9 5 6 4 9 6 4 10 7 8 6 3 9 10 5 1 6 10 9 4
```

Eine größere Bedeutung hat die diskreten Gleichverteilung für die **Erzeugung von Zufallszahlen**. Diese können als Realisierungen der 10 Ziffern 0 bis 9 aufgefasst werden, die alle mit der gleichen Wahrscheinlichkeit $1/10$ in einer „langen“ Sequenz von Ziffern vertreten sind.

Insbesondere der Erwartungswert und die Varianz einer diskreten, gleichverteilten Zufallsvariablen werden von zahlreichen statistischen Verfahren genutzt, die auf **Rangzahlen** (bzw. auf **Rangsummen**) aufbauen.

$$\begin{aligned}\mu &= \frac{m+1}{2} \\ \sigma^2 &= \frac{m^2-1}{12}\end{aligned}\tag{5.41}$$

Die Gleichverteilung tritt zum Beispiel auch bei **Abrundungsfehlern** auf. Hier ist jeweils

$$P(x) = 1/10 \quad \text{für } x = -0,4, -0,3, \dots, +0,5$$

Für den Erwartungswert und die Varianz gilt dann: $\mu = 0,05$ und $\sigma^2 = 0,287$.

5.3.3 Binomialverteilung

5.3.3.1 Bernoulli-Versuch

Urnenmodell: Einmaliges Ziehen ($n = 1$) aus einer Urne (mit Zurücklegen), in der N Kugeln zweierlei Typs (rot / schwarz) in einem bestimmten Verhältnis r/N ($N = r + s$) enthalten sind. Diese Form des Zufallsexperimentes wird auch Bernoulli-Versuch genannt. Die Wahl einer geeigneten Zufallsvariablen mit einem „dichotomen“ **Ausgang** wird in der Regel mit der Abbildung auf die **Werte 0 und 1** getroffen (Modell einer dichotomen Urne):

$$X = \begin{cases} 1 & \text{gezogene Kugel rot (=Erfolg)} \\ 0 & \text{gezogene Kugel schwarz (=Misserfolg)} \end{cases}$$

$$P(X = k) = \begin{cases} q := 1 - p & \text{für } k = 0 \\ p & \text{für } k = 1 \\ 0 & \text{sonst} \end{cases} \quad \text{mit } p = \frac{r}{r+s}$$

$$F(x) = \begin{cases} 0 & \text{für } x < 0 \\ 1 - p & \text{für } 0 \leq x < 1 \\ 1 & \text{für } x \geq 1 \end{cases}\tag{5.42}$$

Ein anderer Name für die Bernoulli-Verteilung ist auch **Null-Eins-Verteilung** (zero-one distribution). Die Bernoulli-Verteilung wird abgekürzt durch $X \sim Be(p)$ gekennzeichnet. Erwartungswert und Varianz einer $Be(p)$ verteilten Zufallsvariablen lassen sich einfach aus der Definition dieser Parameter ableiten:

$$\begin{aligned}\mu &= p \\ \sigma^2 &= p \cdot (1 - p) = p \cdot q\end{aligned}\quad (5.43)$$

Der Bernoulli-Verteilung kommt eine große Bedeutung in der Herleitung und dem Verständnis der Binomialverteilung zu.

5.3.3.2 Binomial-Verteilung

Urnenmodell: n -maliges Ziehen aus einer dichotomen Urne mit Zurücklegen, d.h. n -malige Wiederholung eines Bernoulli-Versuches. Es entsteht eine so genannte **Bernoulli-Kette vom Umfang n** . Als Ergebnis des Gesamtexperimentes wird die Anzahl der Versuche betrachtet, bei denen ein Erfolg (r) eintritt. Die Wahrscheinlichkeit, dass in einer beliebigen Folge **genau x -mal** Erfolg auftritt mit der Wahrscheinlichkeit p und genau $(n - x)$ -mal Misserfolg mit der Wahrscheinlichkeit $(1 - p)$, ist wegen der Unabhängigkeit der einzelnen Ereignisse gleich $p^x(1 - p)^{n-x}$.

Insgesamt gibt es $\binom{n}{x}$ gleichwahrscheinliche Anordnungen für genau x Erfolge in n Zufallsexperimenten. Somit gilt für Einzelwahrscheinlichkeiten der Binomialverteilung (Kurzschreibweise $B(n; p)$ -Verteilung): die Wahrscheinlichkeit, dass die $B(n, p)$ -verteilte Zufallsvariable X genau x Erfolge aufweist, ist:

$$P(X = k) = \begin{cases} \binom{n}{k} p^k (1 - p)^{(n-k)} & \text{für } k = 0, 1, 2, \dots, n \\ 0 & \text{sonst} \end{cases} \quad (5.44)$$

Der Ausdruck Binomialverteilung leitet sich von der **Binomialentwicklung** (5.45) her,

$$\left[(1 + x)^n = \underbrace{\binom{n}{0}}_1 + \binom{n}{1}x + \binom{n}{2}x^2 + \dots + \binom{n}{n}x^n = \sum_{i=0}^n \binom{n}{i} x^i \right]$$

hieraus folgt $(1 + x)^n \geq 1 + nx$ für $x \neq 0$, $x > -1$ und jede natürliche Zahl $n \geq 2$, die Ungleichung nach Isaac Barrow und Jakob Bernoulli.

$$([1 - p] + p)^n = \sum_{x=0}^n \binom{n}{x} p^x (1 - p)^{n-x} = 1 \quad (5.45)$$

die die **Summe der $n + 1$ Binomialwahrscheinlichkeiten** darstellt.

Beispiel: Wie groß ist die Wahrscheinlichkeit, mit einem intakten Würfel keine Dreifachsechs zu werfen?



$$P(\text{Dreifachsechs}) = \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{216}; \quad \text{d. h.}$$

$$P(\text{keine Dreifachsechs}) = 1 - \frac{1}{216} = \frac{215}{216}. \quad \text{Nach (5.45)}$$

$$\left(\left[1 - \frac{1}{6} \right] + \frac{1}{6} \right)^3 = \sum_{x=0}^3 \binom{3}{x} \left(\frac{1}{6} \right)^x \left(1 - \frac{1}{6} \right)^{3-x} = 1$$

lässt sich zeigen, dass die Wahrscheinlichkeit, mit drei Würfeln genau x Sechsen zu werfen, sich aus

Tabelle 5.2. Das Modell der Binomialverteilung für $n=3$ Wiederholungen

Binomialverteilung: $n = 3$, $p = P(A)$, d. h. $1 - p = P(\bar{A})$				
Mögliche Resultate	A tritt (jeweils) x -mal auf	$P(X = x)$ X nimmt den Wert A x -mal an	Zugehörige Wahrscheinlichkeit $[(1 - p) + p]^3$	Binomial-Wahrscheinlichkeit
(1)	(2)	(3)	(4)	(5)
$\bar{A}\bar{A}\bar{A}$	0	$P(X = 0)$	$(1 - p)^3$	$= \binom{3}{0} p^0 (1 - p)^{3-0}$
$A\bar{A}\bar{A} \quad \bar{A}A\bar{A} \quad \bar{A}\bar{A}A$	1	$P(X = 1)$	$3p(1 - p)^2$	$= \binom{3}{1} p^1 (1 - p)^{3-1}$
$AA\bar{A} \quad A\bar{A}A \quad \bar{A}AA$	2	$P(X = 2)$	$3p^2(1 - p)$	$= \binom{3}{2} p^2 (1 - p)^{3-2}$
AAA	3	$P(X = 3)$	p^3	$= \binom{3}{3} p^3 (1 - p)^{3-3}$
Summe der 4 Binomialwahrscheinlichkeiten: $1 = \sum_{x=0}^3 \binom{3}{x} p^x (1 - p)^{3-x}$				

[vgl. $(a + b)^3 = a^3 + 3a^2b + 3ab^2 + b^3$]

$$\left(\frac{5}{6} + \frac{1}{6}\right)^3 = \frac{125}{216} + \frac{75}{216} + \frac{15}{216} + \frac{1}{216} = 1$$

$x: \quad 0 \quad \quad 1 \quad \quad 2 \quad \quad 3$

ergibt und die gesuchte Wahrscheinlichkeit als Summe $(125 + 75 + 15)/216$ berechnet werden kann.

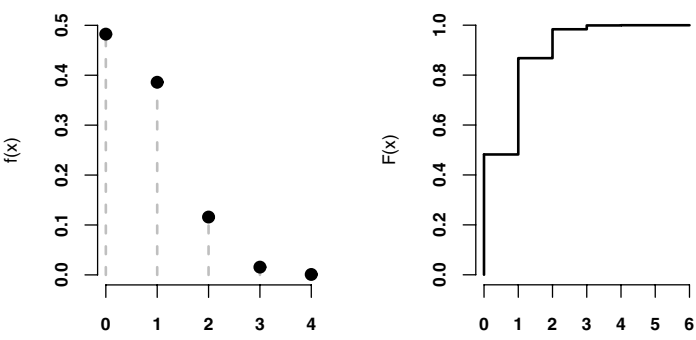


Abb. 5.9. Binomialverteilung für $n=4$ und $p=1/6$ (4maliger Wurf eines Würfels)

Abbildung 5.9 zeigt die Wahrscheinlichkeitsfunktion (links) und die Verteilungsfunktion (rechts) für die Zufallsvariable „Anzahl der Sechsen“ bei 4maligem Wurf eines Würfels.

Hat man einzelne Wahrscheinlichkeiten $P(x)$ zu berechnen (vgl. Beispiel zur Ausschussware), so kann man die sogenannten **Rekursionsformeln** (5.46) mit $q = 1 - p$ verwenden:

$$\begin{aligned} P(X = x + 1) &= \frac{n - x}{x + 1} \cdot \frac{p}{q} \cdot P(X = x) \\ \text{und } P(X = x - 1) &= \frac{x}{n - x + 1} \cdot \frac{p}{q} \cdot P(X = x) \end{aligned} \quad (5.46)$$

Mit $P(0) = q^n$ für gegebenes q und n (5.44), folgt nach der Rekursionsformel

$$P(1) = \frac{n}{1} \cdot \frac{p}{q} \cdot P(0), P(2) = \frac{n-1}{2} \cdot \frac{p}{q} \cdot P(1), \text{ usw.}$$

Summiert man die Binomialwahrscheinlichkeiten von $k = 0$ bis $k = x$, dann erhält man die Verteilungsfunktion $F(x)$ einer binomialverteilten Zufallsvariablen (vgl. auch Abschnitt [6.3.2.1]) nach:

$$\begin{aligned} F(x) &= \sum_{k=0}^{k=x} \binom{n}{k} p^k (1-p)^{n-k} \\ F(n) &= \sum_{k=0}^{k=n} \binom{n}{k} p^k (1-p)^{n-k} = 1 \end{aligned} \quad (5.47)$$

Die Berechnung der Binomialverteilung in **R** erfolgt durch die Funktionen



dbinom(x , size, prob) Wahrscheinlichkeitsfunktion

pbinom(q , size, prob) Verteilungsfunktion

qbinom(p , size, prob) Quantilfunktion

rbinom(n , size, prob) Zufallszahlen erzeugen

Die Parameter der Binomialverteilung werden durch die Anzahl der Wiederholungen (size) und durch die Erfolgswahrscheinlichkeit (prob) festgelegt. x und q sind Werte der Zufallsvariablen, zu denen die Wahrscheinlichkeitsfunktion ($P(X = x)$) bzw. Verteilungsfunktion ($P(X \leq q)$) berechnet werden soll. Für **p** sind Werte vorzugeben ($0 \leq p \leq 1$), zu denen die zugehörigen Quantile zu bestimmen sind. n bestimmt die Anzahl der Werte, die durch den Zufallszahlengenerator erzeugt werden sollen.

Beispiel (Münzwurf): Gefragt ist nach der Wahrscheinlichkeit, nach dreimaligem Wurf mit einer idealen Münze ($p = 1/2$) a) dreimal Zahl, b) zweimal Zahl [und einmal Wappen] zu erzielen.



$$\text{a) } P = \binom{3}{3} \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^0 = 1 \cdot \frac{1}{8} \cdot 1 = \frac{1}{8} = 0,125$$

$$\text{b) } P = \binom{3}{2} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^1 = 3 \cdot \frac{1}{4} \cdot \frac{1}{2} = \frac{3}{8} = 0,375.$$

```
> dbinom(3, 3, 1/2)
[1] 0.125
> dbinom(2, 3, 1/2)
[1] 0.375
```


Tabelle 5.3. Ausgewählte Binomialwahrscheinlichkeiten für $n = 2, \dots, 5$; zu beachten ist die Symmetrie in der letzten Spalte für $p = 0.50$ (erzeugt mit der Funktion **dbinom()** in R)

n	x	p=0.01	p=0.05	p=0.10	p=0.20	p=0.25	p=0.30	p=0.40	p=0.50
2	0	0.9801	0.9025	0.8100	0.6400	0.5625	0.4900	0.3600	0.2500
	1	0.0198	0.0950	0.1800	0.3200	0.3750	0.4200	0.4800	0.5000
	2	0.0001	0.0025	0.0100	0.0400	0.0625	0.0900	0.1600	0.2500
3	0	0.9703	0.8574	0.7290	0.5120	0.4219	0.3430	0.2160	0.1250
	1	0.0294	0.1354	0.2430	0.3840	0.4219	0.4410	0.4320	0.3750
	2	0.0003	0.0071	0.0270	0.0960	0.1406	0.1890	0.2880	0.3750
	3	0.0000	0.0001	0.0010	0.0080	0.0156	0.0270	0.0640	0.1250
4	0	0.9606	0.8145	0.6561	0.4096	0.3164	0.2401	0.1296	0.0625
	1	0.0388	0.1715	0.2916	0.4096	0.4219	0.4116	0.3456	0.2500
	2	0.0006	0.0135	0.0486	0.1536	0.2109	0.2646	0.3456	0.3750
	3	0.0000	0.0005	0.0036	0.0256	0.0469	0.0756	0.1536	0.2500
	4	0.0000	0.0000	0.0001	0.0016	0.0039	0.0081	0.0256	0.0625
5	0	0.9510	0.7738	0.5905	0.3277	0.2373	0.1681	0.0778	0.0312
	1	0.0480	0.2036	0.3280	0.4096	0.3955	0.3601	0.2592	0.1562
	2	0.0010	0.0214	0.0729	0.2048	0.2637	0.3087	0.3456	0.3125
	3	0.0000	0.0011	0.0081	0.0512	0.0879	0.1323	0.2304	0.3125
	4	0.0000	0.0000	0.0005	0.0064	0.0146	0.0284	0.0768	0.1562
	5	0.0000	0.0000	0.0000	0.0003	0.0010	0.0024	0.0102	0.0312

Beispiel (Ausschussware unter Bleistiften): Eine Maschine produziere 20% Ausschussbleistifte. Gefragt ist nach der Wahrscheinlichkeit, dass von 4 zufällig ausgewählten Bleistiften a) kein Bleistift, b) ein Bleistift, c) höchstens zwei Bleistifte Ausschussware sind. Die Wahrscheinlichkeit, Ausschussware zu produzieren, beträgt $p = 0,2$ – die Wahrscheinlichkeit, keine Ausschussware herzustellen, beträgt $q = 1 - p = 0,8$.

$$\begin{aligned} \text{a) } P(\text{nullmal Ausschuss}) &= \binom{4}{0} (0,2)^0 (0,8)^4 = 0,4096 \\ \text{b) } P(\text{einmal Ausschuss}) &= \binom{4}{1} (0,2)^1 (0,8)^3 = 0,4096 \\ \text{c) } P(\text{zweimal Ausschuss}) &= \binom{4}{2} (0,2)^2 (0,8)^2 = 0,1536 \end{aligned}$$

```
> dbinom(0, 4, 0.2)
[1] 0.4096
> dbinom(1, 4, 0.2)
[1] 0.4096
> dbinom(2, 4, 0.2)
[1] 0.1536
```

$P(\text{höchstens zweimal Ausschuss}) = P(\text{nullmal A.}) + P(\text{einmal A.}) + P(\text{zweimal A.}) = 0,4096 + 0,4096 + 0,1536 = 0,9728$. Hier ist $n = 4$, x durchläuft die Werte 0, 1, 2 für jeweils $p = 0,2$. Mit der Rekursionsformel (5.46) folgt:

$$p = 0,2 = \frac{1}{5} \quad \text{und} \quad n = 4; \quad \frac{p}{q} = \frac{1}{5} \bigg/ \frac{4}{5} = \frac{1}{4}; \quad P(x+1) = \frac{4-x}{x+1} \cdot \frac{1}{4} \cdot P_4(x)$$



$$\begin{aligned}
P(0) &= 0,8^4 = 0,4096 \\
P(1) &= \frac{4}{1} \cdot \frac{1}{4} \cdot 0,4096 = 0,4096 \\
P(2) &= \frac{3}{2} \cdot \frac{1}{4} \cdot 0,4096 = 0,1536 \\
P(3) &= \frac{2}{3} \cdot \frac{1}{4} \cdot 0,1536 = 0,0256 \\
P(4) &= \frac{1}{4} \cdot \frac{1}{4} \cdot 0,0256 = 0,0016
\end{aligned}
\left. \vphantom{\begin{aligned} P(0) \\ P(1) \\ P(2) \\ P(3) \\ P(4) \end{aligned}} \right\} 0,0272$$

Kontrolle: $\sum P = 1,0000$

$$(0,8 + 0,2)^4 = 0,4096 + 0,4096 + 0,1536 + 0,0256 + 0,0016 = 1$$

$$X = \quad 0 \quad \quad 1 \quad \quad 2 \quad \quad 3 \quad \quad 4$$

```
> dbinom(0:4, 4, 0.2)
[1] 0.4096 0.4096 0.1536 0.0256 0.0016
> pbinom(2, 4, 0.2)
[1] 0.9728
```

Beispiel (Chevalier de Méré): Was ist wahrscheinlicher: Beim Werfen a) mit 6 Würfeln wenigstens eine Sechs zu erzielen oder b) mit 12 Würfeln wenigstens zwei Sechsen zu erhalten? Ideale Würfel vorausgesetzt.



$$\begin{aligned}
\text{a) } P_{\text{Null Sechsen zu erzielen}} &= \binom{6}{0} \left(\frac{1}{6}\right)^0 \left(\frac{5}{6}\right)^6 \simeq 0,335 \\
P_{\text{Eine oder mehr Sechsen z. e.}} &= 1 - \binom{6}{0} \left(\frac{1}{6}\right)^0 \left(\frac{5}{6}\right)^6 \simeq 0,665 \\
\text{b) } P_{\text{zwei oder mehr Sechsen z. e.}} &= 1 - \left(\binom{12}{0} \left(\frac{1}{6}\right)^0 \left(\frac{5}{6}\right)^{12} + \binom{12}{1} \left(\frac{1}{6}\right)^1 \left(\frac{5}{6}\right)^{11} \right) \\
&\simeq 1 - (0,1122 + 0,2692) \simeq 0,619.
\end{aligned}$$

```
> 1 - pbinom(0, 6, 1/6, lower.tail=TRUE)
[1] 0.665102
> pbinom(1, 12, 1/6, lower.tail=FALSE)
[1] 0.6186674
```

Damit ist a) wahrscheinlicher als b). In der Lösung in **R** ist zu beachten, das mit der Funktion **pbinom()** über das zusätzliche Argument **lower.tail** sowohl die Verteilungsanteile links (**lower.tail=TRUE**, $\leq$) oder rechts (**lower.tail=FALSE**, $>$) berechnet werden können.

Beispiel: Ein idealer Würfel wird 120mal geworfen. Gefragt ist nach der Wahrscheinlichkeit, dass die Ziffer 4 achtzehnmal oder weniger häufig erscheint. Die manuelle Berechnung ist schon recht aufwendig. Die Lösung in **R** lautet:

```
> pbinom(18, 120, 1/6)
[1] 0.3657008
```





Beispiel (Mäusewürfe): Uns interessiert die Zahl der weiblichen Jungtiere in Würfen zu je 4 Mäusen (vgl. David, F. N.: A Statistical Primer, Ch. Griffin, London 1953, S. 187 ff.). Die Befunde von 200 Würfen dieser Art liegen vor:

Tabelle 5.4. Zahl weiblicher Mäuse in Würfen zu je 4 Mäusen

Zahl der weiblichen Mäuse/Wurf	0	1	2	3	4
Anzahl der Würfe (beobachtet insgesamt 200)	15	63	66	47	9
Anzahl der Würfe (erwartet mit $\hat{p}=0.465$)	16.38	56.96	74.26	43.04	9.36

Nehmen wir nun an, dass für den verwendeten Mäusestamm die Wahrscheinlichkeit, als Weibchen geboren zu werden, konstant, unabhängig von der Anzahl der bereits geborenen weiblichen Tiere ist und dass auch zweitens die Würfe unabhängig voneinander sind, also einem Zufallsprozess folgen, dann lässt sich der Prozentsatz weiblicher Tiere in der Grundgesamtheit aus der vorliegenden Stichprobe von 200 Würfen schätzen.



Hinweis: Die Schätzung für einen Parameter wird im folgenden stets durch ein „**Dach**“- oder „**Hut**“-Symbol über dem Parameter angezeigt.

Der Anteil weiblicher Jungtiere beträgt

$$\hat{p} = \frac{\text{Anzahl weiblicher Jungtiere}}{\text{Gesamtzahl der Jungtiere}}$$

$$\hat{p} = \frac{(0 \cdot 15 + 1 \cdot 63 + 2 \cdot 66 + 3 \cdot 47 + 4 \cdot 9)}{4 \cdot 200} = 0,465.$$

Wir wissen nun, das, wenn die Voraussetzungen der Binomialverteilung erfüllt sind, die Wahrscheinlichkeiten 0, 1, 2, 3, 4 weibliche Tiere in Würfen zu je 4 Tieren zu erhalten, mit Hilfe der binomischen Entwicklung $(0,535 + 0,465)^4$ ermittelt werden können. Die aufgrund dieser Entwicklung für 200 Vierlinge erwarteten Zahlen sind dann gegeben durch:

$$\begin{aligned} 200(0,535 + 0,465)^4 &= 200(0,0819 + 0,2848 + 0,3713 + 0,2152 + 0,0468) \\ &= 16,38 + 56,96 + 74,27 + 43,03 + 9,35. \end{aligned}$$

```
> round(200*dbinom(0:4, 4, 0.465), 2)
[1] 16.38 56.96 74.27 43.03 9.35
```

Die erwarteten Zahlen sind zum Vergleich mit den beobachteten Werten in Tabelle 5.4 eingetragen.

Erwartungswert und Varianz einer Binomialverteilung $B(n, p)$ lassen sich einfach aus den entsprechenden Parametern der Bernoulli-Verteilung (n -malige Wiederholung) ableiten.

Erwartungswert: $\mu = n \cdot p$

Varianz: $\sigma^2 = n \cdot p \cdot (1 - p) = n \cdot p \cdot q$

beachte $\mu > \sigma^2$

2. Moment: $n \cdot p \cdot (1 - p + n \cdot p)$

(5.48)

Schiefe und Exzess einer $B(n; p)$ -Verteilung sind:

$$\gamma_1 = \frac{(1 - p) - p}{\sigma} = \frac{1 - 2p}{\sqrt{np(1 - p)}} \quad \text{(Schiefe)}$$

$$\gamma_2 = \frac{1 - 6pq}{npq} \quad \text{(Exzess)}$$

(5.49)

Die Schiefe ist 0, wenn $p = 0,5$. Die Verteilung ist dann symmetrisch zum Erwartungswert. Die Schiefe wird auch sehr klein, wenn die Anzahl der Wiederholungen n sehr groß wird.

Beispiel (Behandlungserfolge): Die Wahrscheinlichkeit, dass eine bestimmte Behandlung erfolgreich ist, sei 0,8. Wie groß ist die Wahrscheinlichkeit, dass von



(a) 2 Behandlungen nur eine erfolgreich ist, (a) $P = \binom{2}{1} 0,8 \cdot 0,2 = 2 \cdot 0,8 \cdot 0,2 = 0,32$,

(b) 5 Behandlungen nur eine erfolgreich ist, (b) $P = \binom{5}{1} 0,8 \cdot 0,2^4 = 5 \cdot 0,8 \cdot 0,2^4 = 0,0064$,

(c) 5 Behandlungen alle erfolgreich sind? (c) $P = \binom{5}{5} 0,8^5 \cdot 0,2^0 = 0,8^5 = 0,32768$.

Für die $B(5; 0,8)$ -Verteilung ist nach (5.48) und (5.49) $\mu = 5 \cdot 0,8 = 4$; $\sigma^2 = 5 \cdot 0,8 \cdot 0,2 = 0,8$; Schiefe = $(0,2 - 0,8)/\sqrt{0,8} = -0,671$. Für die $B(500; 0,8)$ -Verteilung ist die Schiefe = $(0,2 - 0,8)/\sqrt{500 \cdot 0,8 \cdot 0,2} = -0,067$.

Beispiel (Fünfkinderfamilie): Wie groß ist die Wahrscheinlichkeit, dass in Familien mit 5 Kindern: (a) 2 Kinder Mädchen sind, (b) 5 Kinder Knaben sind, wenn die Wahrscheinlichkeit einer Mädchengeburt $p = 0,5$ gesetzt wird?



(a) $P(X=2|5; 0,5) = \binom{5}{2} 0,5^2 \cdot 0,5^{5-2} = 10 \cdot 0,5^5$ (b) $P(X=5|5; 0,5) = \binom{5}{5} 0,5^5 \cdot 0,5^{5-5} = 1 \cdot 0,5^5$
 $= 10/2^5 = 10/32$, $= 1/2^5 = 1/32$.

Hinweis: Es ist ein Trugschluss, unabhängige Ereignisse als abhängig aufzufassen: nach 5 Münzwürfen mit dem Ergebnis „Wappen“ gilt auch für den 6. Wurf nicht der „erwartete Ausgleich“ in der Art von „ $P(\text{Zahl}) > 0,5$ “ [bekannt als **Gambler's Fallacy**]; es bleibt bei $P(\text{Wappen}) = P(\text{Zahl}) = 0,5$. Dies entspricht einem „Ziehen“ mit Zurücklegen im Urnenmodell. Anders stellt sich die Situation, wenn über die Erfolgswahrscheinlichkeit keine gesicherte Annahme gemacht werden kann, etwa wenn die Münze gefälscht wurde, und dieses Phänomen im Laufe der Wiederholungen auffällt. In diesem Fall sollte man aus der Erfahrung lernen und mit einer modifizierten Erfolgswahrscheinlichkeit rechnen!



5.3.3.3 Approximation der Binomialverteilung durch die Standardnormalverteilung

Für $npq \geq 9$ kann

$$\hat{z} = (x - np) / \sqrt{npq} \quad (5.50)$$

als angenähert standardnormalverteilt (s.d.) gelten, mit

$$\begin{aligned} P(x_1 < X < x_2) &\approx P\left(\frac{x_1 + 0,5 - np}{\sqrt{np(1-p)}} < Z < \frac{x_2 - 0,5 - np}{\sqrt{np(1-p)}}\right) \\ P(x_1 \leq X \leq x_2) &\approx P\left(\frac{x_1 - 0,5 - np}{\sqrt{np(1-p)}} \leq Z \leq \frac{x_2 + 0,5 - np}{\sqrt{np(1-p)}}\right) \end{aligned} \quad (5.51)$$

Beispielsweise sei $P(16 < X \leq 26)$ für $n = 100$ und $p = 0,25$ zu bestimmen. Da $np(1-p) = 100 \cdot 0,25 \cdot 0,75 = 18,75 > 9$, berechnen wir $np = 25$ und $\sqrt{np(1-p)} = 4,330$ und

$$P\left(\frac{16 + 0,5 - 25}{4,330} < Z \leq \frac{26 + 0,5 - 25}{4,330}\right) = P(-1,963 < Z \leq 0,346) .$$

Hieraus erhält man für $P(16 < X \leq 26) = P(17 \leq X \leq 26)$ und Tabelle 5.9 die angenäherten Werte $(0,5 - 0,0248) + (0,5 - 0,3647) = 0,4752 + 0,1353 = 0,6105$ oder 0,61 (exakter Wert 0,62063).

$P(X \leq k|p; n) = \sum_{j=0}^k \binom{n}{j} p^j q^{n-j}$ lässt sich besser nach (5.52) approximieren [Mol70]:

$$\hat{z} - |\sqrt{q(4k+3,5)} - \sqrt{p(4n-4k-0,5)}| \quad (5.52)$$

In (5.52) ist (a) für $0,05 \leq P \leq 0,93$ 3,5 durch 3 und 0,5 durch 1 zu ersetzen; (b) für extremere P -Werte ersetze man 3,5 durch 4 und 0,5 durch 0.



Beispiel: $P(X \leq 13|0,6; 25) = 0,268$; $\hat{z} = |\sqrt{0,4(52+3,5)} - \sqrt{0,6(100-52-0,5)}| = 0,627$, d. h. $P = 0,265$; mit 3 und 1 ergibt sich über $\hat{z} = 0,620$ $P = 0,268$.



Beispiel(Würfelwurf): Ein idealer Würfel wird 120mal geworfen. Gefragt ist nach der Wahrscheinlichkeit, dass die Ziffer 4 achtzehnmal oder weniger häufig erscheint. Die Wahrscheinlichkeit dafür, dass die Vier null- bis achtzehnmal aufzeigt ($p = 1/6$; $q = 5/6$), ist genau gleich

$$\binom{120}{18} \left(\frac{1}{6}\right)^{18} \left(\frac{5}{6}\right)^{102} + \binom{120}{17} \left(\frac{1}{6}\right)^{17} \left(\frac{5}{6}\right)^{103} + \dots + \binom{120}{0} \left(\frac{1}{6}\right)^0 \left(\frac{5}{6}\right)^{120} .$$

Da der Rechenaufwand ziemlich groß ist, benutzen wir die Approximation über die Standardnormalverteilung (vgl. $npq = 120 \cdot 1/6 \cdot 5/6 = 16,667 > 9$). Betrachten wir die Zahlen als kontinuierlich, dann folgt, dass 0 bis 18 Vieren als $-0,5$ bis $18,5$ Vieren aufgefasst werden können, d. h.

$$\bar{x} = np = 120 \cdot \frac{1}{6} = 20 \quad \text{und} \quad s = \sqrt{npq} = \sqrt{16,667} = 4,08 .$$

$-0,5$ und $18,5$ werden dann in Standardeinheiten transformiert:

$$\begin{aligned} \hat{z} &= (x - \bar{x})/s, \text{ für } -0,5 \text{ erhält man } (-0,5 - 20)/4,09 = -5,01 \\ &\text{für } 18,5 \text{ erhält man } (18,5 - 20)/4,09 = -0,37 . \end{aligned}$$

Die gewünschte Wahrscheinlichkeit P ist dann durch die Fläche unter der Normalkurve zwischen $z = -5,01$ und $z = -0,37$ gegeben.

$$\begin{aligned} P &= (\text{Fläche zwischen } z = 0 \text{ und } z = -5,01) - \\ &\quad (\text{Fläche zwischen } z = 0 \text{ und } z = -0,37) \\ P &= 0,50000 - 0,1443 = 0,3557 . \end{aligned}$$

Hieraus folgt: Nehmen wir wiederholt Stichproben von 120 Würfeln, dann sollte die Vier in etwa 36% der Würfe 18mal oder seltener erscheinen.

5.3.3.4 Approximation der Binomialverteilung durch die Poisson-Verteilung

Liegt eine Binomialverteilung mit großem Stichprobenumfang n und kleiner Ereigniswahrscheinlichkeit p vor, so dass $q = 1 - p$ praktisch gleich 1 ist, sagen wir, wenn $p < 0,05$ und $n > 10$, dann kann die Poisson-Verteilung mit $\lambda = np$ als Approximation der Binomialverteilung dienen.

Beispiel (Häufigkeit von Bränden): In einem bestimmten Gebiet habe im Durchschnitt 1 von 2000 Häusern jährlich einen Brand. Wenn 4000 Häuser in diesem Gebiet sind, wie groß ist dann die Wahrscheinlichkeit, dass genau 5 Häuser im Verlauf des Jahres einen Brand haben?



$$\hat{\lambda} = np = 4000 \cdot \frac{1}{2000} = 2$$

$$P(X = 5, \hat{\lambda} = 2) = e^{-2} \cdot \frac{2^5}{5!} = 0,036$$

Die Wahrscheinlichkeit beträgt knapp 4%.

Beispiel (Gewinn-Chancen): Die Wahrscheinlichkeit, bei einem bestimmten Spiel zu gewinnen, betrage $p = 0,01$, es seien $n = 100$ Spiele durchgeführt worden. Wie gross ist die Wahrscheinlichkeit: (a) nullmal, (b) einmal und (c) mindestens zweimal zu gewinnen?



Mit $\lambda = np = 100 \cdot 0,01 = 1$ erhält man :

$$\text{a) } P(0|1) = e^{-1} = \frac{1}{e} \approx \frac{1}{2,72} = 0,368$$

$$\text{b) } P(1|1) = 1e^{-1} = 0,368$$

$$P(\geq 2|1) \approx 1 - 2 \cdot 0,368 = \underline{0,264}$$

$$\text{c) } P(\geq 0|1) = 1$$

Die exakten Binomialwahrscheinlichkeiten betragen 0,366; 0,370; 0,264.

5.3.4 Poisson-Verteilung

Setzen wir in (5.44) den ziemlich kleinen Wert $np = \lambda$ (gr. lambda) und lassen wir bei konstant gehaltenem $\lambda > 0$ die Zahl n beliebig wachsen ($n \rightarrow \infty$), so geht die Binomialverteilung mit dem Erwartungswert $np = \lambda$, in die so genannte Poisson-Verteilung mit dem Parameter λ über. Lambda ist der Erwartungswert dieser Verteilung; im allgemeinen ist $\lambda < 20$. Die Poisson-Verteilung ist durch den französischen Mathematiker S.D. Poisson (1781–1840) entwickelt worden. Sie war bereits A. de Moivre (1667–1754) bekannt; sie gilt, wenn die durchschnittliche Anzahl der Ereignisse das Ergebnis einer **sehr großen Zahl von Ereignismöglichkeiten** und einer **sehr kleinen Ereigniswahrscheinlichkeit** ist. Ein gutes Beispiel hierfür ist der radioaktive Zerfall: Von vielen Millionen Radiumatomen zerfällt in der Zeiteinheit nur ein sehr kleiner Prozentsatz. Wesentlich ist, dass der Zerfall ein Zufallsprozess ist und dass der Zerfall der einzelnen Atome unabhängig ist von der Zahl der schon zerfallenen Atome.

Die Poisson-Verteilung ist eine wichtige Verteilung. Sie wird – wie angedeutet – für die Lösungen der Probleme benutzt, die beim Zählen relativ seltener zufälliger und voneinander unabhängiger Ereignisse in der Zeit-, Längen-, Flächen- oder Raumeinheit auftreten. Man spricht auch von **isolierten Ereignissen in einem Kontinuum**. In brauchbarer bis guter Näherung folgen einer Poisson-Verteilung die Verteilung von (vor dem Backen ordentlich in Mehl geschwenkten und dadurch hoffentlich voneinander „unabhängigen“) Rosinen im Rosinenbrot, von Hefezellen in einer Suspension und von Erythrozyten in den einzelnen Feldern einer Zählkammer, die Anzahl der Druckfehler pro Seite, der Isolationsfehler an einer Verlängerungsschnur oder der Oberflächenfehler einer Tischplatte; die Anzahl der bei Rot vor einer Ampel wartenden Autos; die Ankunftsfolge von Flugzeugen auf dem Flughafen; die Häufigkeit von plötzlichen Unwettern (z. B. Überschwemmungen oder Erdbeben) in einem bestimmten Gebiet; die Verunreinigung von Samen durch Unkrautsamen oder Steine; die Anzahl der innerhalb einer bestimmten Zeitspanne eintreffenden Telefonanrufe, die Zahl der Elektronen, die von einer erhitzten Kathode in einer gegebenen Zeiteinheit emittiert werden; die Zahl der Pannen an den Fahrzeugen einer größeren

militärischen Einheit; die Zahl der Ausschusstücke innerhalb einer Produktion; die Zahl der Verkehrsmittel pro Weg- und Zeiteinheit; die Anzahl der Fehlerstellen in komplizierten Mechanismen - alles **pro Raum- oder Zeiteinheit**. Sowie die Wahrscheinlichkeit nicht konstant bleibt oder Ereignisse abhängig werden, resultieren Abweichungen von der Poisson-Verteilung. Werden diese Möglichkeiten ausgeschaltet, dann sind – das gilt für die gegebenen Beispiele – echte Poisson-Verteilungen zu erwarten. Selbstmordfälle oder Industrieunfälle pro Raum- und Zeiteinheit folgen nicht der Poisson-Verteilung, obwohl sie als seltene Ereignisse aufgefasst werden können. In beiden Fällen kann nicht von einer „gleichen Chance für jeden“ gesprochen werden, es gibt individuelle Unterschiede hinsichtlich der Unfallbereitschaft und Selbstmordanfälligkeit.

Denken wir uns ein **Rosinenbrot**, das in kleine gleich große Kost- oder Stichproben zerlegt wird. Infolge der jetzt von uns verlangten **zufälligen Verteilung** der Rosinen ist nicht zu erwarten, dass alle Stichproben genau die gleiche Anzahl von Rosinen enthalten. Wenn die mittlere Anzahl λ (lambda) der in diesen Stichproben enthaltenen Rosinen bekannt ist, so gibt die Poisson-Verteilung die Wahrscheinlichkeit $P(X = x)$ dafür an, dass eine beliebig herausgegriffene Stichprobe gerade x ($x = 0, 1, 2, 3, \dots$) Rosinen enthält. Anders ausgedrückt: Die Poisson-Verteilung gibt an, welcher prozentuale Anteil ($100 \cdot P(X = x)\%$) einer langen Serie nacheinander entnommener Stichproben mit (Rosinenbrot-Scheiben konstanter Breite) genau 0 bzw. 1 bzw. 2 usw. Rosinen besetzt ist.

Eine diskrete Zufallsvariable heißt **Poisson-verteilt**, wenn ihre Wahrscheinlichkeitsfunktion nach 5.53 beschrieben werden kann.

$$\boxed{P(X = x | \lambda) = P(x) = \frac{\lambda^x e^{-\lambda}}{x!}} \quad (5.53)$$

$$\lambda > 0, \quad x = 0, 1, 2, \dots$$

Dabei bedeutet:

$e = 2,718 \dots$ die Basis des natürlichen Logarithmus

λ = Erwartungswert (vgl. Übersicht 9)

$x = 0, 1, 2, 3 \dots$ die genaue Anzahl der Ereignisse (etwa der Rosinen) in einer einzelnen Stichprobe, der Wertebereich der Realisierungen

$x! = 1 \cdot 2 \cdot 3 \cdot \dots \cdot (x - 1)x$, (z. B. $4! = 1 \cdot 2 \cdot 3 \cdot 4 = 24$)

Die Zufallsvariable X bezeichnet die Anzahl der Stichprobenelemente n_R mit der Eigenschaft R unter den insgesamt n entnommenen Stichprobenelementen ($n \gtrapprox 30$; n_R ist relativ klein, etwa $n_R/n \lesssim 0,1$).

Durch die diskrete Wahrscheinlichkeitsfunktion (5.53) ist die Poisson-Verteilung definiert. Die Poisson-Verteilung wird durch den Parameter λ vollständig charakterisiert; er drückt die **Dichte von Zufallspunkten** innerhalb eines gegebenen Zeitintervalls, einer Längen-, einer Flächen- oder einer Raumeinheit aus. λ ist zugleich Erwartungswert und Varianz, d. h. $\mu = \lambda$, $\sigma^2 = \lambda$ [vgl. auch (5.48) mit $np = \lambda$ und $q = 1 - \frac{\lambda}{n}$; $\sigma^2 = \lambda \left(1 - \frac{\lambda}{n}\right)$, für großes n wird auch σ^2 gleich λ].

$$\boxed{\begin{array}{ll} \text{Erwartungswert: } \mu = \lambda \\ \text{Varianz: } \sigma^2 = \lambda \\ \text{2. Moment: } \lambda \cdot (\lambda + 1) \end{array}} \quad (5.54)$$

Der Parameter λ der Poisson-Verteilung wird für $q \simeq 1$ nach

$$\boxed{\hat{\lambda} = np} \quad (5.55)$$

geschätzt.



Beispiel (Radioaktivität): Ein radioaktives Präparat gebe durchschnittlich 10 Impulse pro Minute. Wie groß ist die Wahrscheinlichkeit, in einer Minute 5 Impulse zu erhalten?

$$P = \frac{\lambda^x \cdot e^{-\lambda}}{x!} = \frac{10^5 \cdot e^{-10}}{5!} = \frac{10^5 \cdot 4,54 \cdot 10^{-5}}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = \frac{4,54}{120} = 0,03783 \simeq 0,04$$

Man wird also in etwa 4% der Fälle mit 5 Impulsen pro Minute zu rechnen haben.

Übersicht 11. Binomialwahrscheinlichkeit und Poissonverteilung

Für sehr großes n , sehr kleines p und einen festen Wert $np = \lambda$ (d.h. $n \rightarrow \infty, p \rightarrow 0$ und $np \rightarrow \lambda > 0$) strebt die Binomialwahrscheinlichkeit $P(X = x|n, p)$ gegen die Wahrscheinlichkeit der Poisson-Verteilung mit dem Parameter λ [(5.53)]. Als Ereignisse kommen alle nichtnegativen ganzen Zahlen in Frage, so dass gilt:

$$\sum_{x=0}^{\infty} P(X = x|\lambda) = 1$$

Für $x = 0, 1, 2$ ergibt sich daraus speziell:

$$P(X = 0|\lambda) = e^{-\lambda}$$

$$P(X = 1|\lambda) = \lambda e^{-\lambda}$$

$$P(X = 2|\lambda) = \frac{\lambda^2}{2} e^{-\lambda}$$

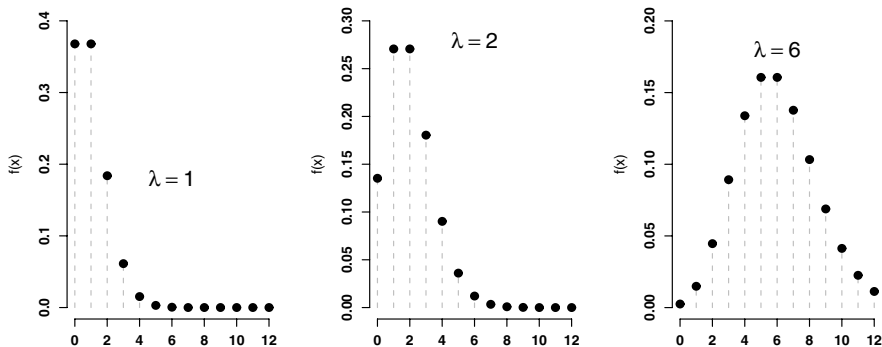


Abb. 5.10. Poisson-Verteilung mit verschiedenen Parametern $\lambda = 1, 2, 6$

Abbildung 5.10 und Tabelle 5.5 zeigen:

- Die Poisson-Verteilung ist eine diskrete linkssteile Verteilung. Ihre **Schiefte** ($1/\lambda$) strebt mit wachsendem λ gegen Null, d.h. die Verteilung wird dann nahezu symmetrisch. Die **Wölbung** der Poisson-Verteilung ($3 + 1/\lambda$) strebt mit wachsendem λ gegen den Wert 3.
- Die Einzelwahrscheinlichkeiten der Poissonverteilung nehmen für $\lambda < 1$ mit wachsendem X monoton ab (die Verteilung ist „J“-förmig); für $\lambda > 1$ zunächst zu und dann ab (die Verteilung ist eingipflig und schief).
- Das Maximum der Verteilung liegt bei der größten ganzen Zahl, die kleiner als λ ist. Bei positivem ganzzahligen λ treten zwei gleich große maximale Einzelwahrscheinlichkeiten auf.

Für den Fall, dass a) λ groß ist und b) $X = \lambda$, erhält man [vgl. die Stirlingsche Formel (die eigentlich von de Moivre stammt): $n! = n^n e^{-n} \sqrt{2\pi n}$ für $n \rightarrow \infty$] - approximativ:

$$P(\lambda) = \frac{e^{-\lambda} \cdot \lambda^\lambda}{\lambda!} \simeq \frac{e^{-\lambda} \cdot \lambda^\lambda}{\lambda^\lambda e^{-\lambda} \sqrt{2\pi\lambda}} = \frac{1}{\sqrt{2\pi\lambda}} \simeq \frac{0,4}{\sqrt{\lambda}}$$

$$\boxed{P(\lambda) \simeq \frac{0,4}{\sqrt{\lambda}}} \quad (5.56)$$

z. B. $P(X = \lambda = 8) \simeq 0,4/\sqrt{8} = 0,141$; der in Tabelle 5.5 notierte Wert lautet 0,1396.



Beispiel: Wenn die Anzahl der Druckfehler pro Zeitungsseite einer Poisson-Verteilung des Typs $\lambda = 0,2$ folgt, dann dürften von 100 Seiten etwa 82 Seiten keine, 16 einen und etwa 2 mehr als einen Druckfehler aufweisen. Tabelle 5.5 zeigt weiter, dass von 10000 Seiten etwa eine mit 4 Fehlern zu erwarten ist.

Tabelle 5.5. Wahrscheinlichkeiten der Poissonverteilung für ausgewählte Werte λ mit $x = 0, \dots, 29$; berechnet in **R** mit der Funktion **dpois()**

x	$\lambda = 0,2$	$\lambda = 0,5$	$\lambda = 0,8$	$\lambda = 1$	$\lambda = 3$	$\lambda = 5$	$\lambda = 8$	$\lambda = 12$	$\lambda = 20$
0	0.8187	0.6065	0.4493	0.3679	0.0498	0.0067	0.0003	0.0000	0.0000
1	0.1637	0.3033	0.3595	0.3679	0.1494	0.0337	0.0027	0.0001	0.0000
2	0.0164	0.0758	0.1438	0.1839	0.2240	0.0842	0.0107	0.0004	0.0000
3	0.0011	0.0126	0.0383	0.0613	0.2240	0.1404	0.0286	0.0018	0.0000
4	0.0001	0.0016	0.0077	0.0153	0.1680	0.1755	0.0573	0.0053	0.0000
5	0.0000	0.0002	0.0012	0.0031	0.1008	0.1755	0.0916	0.0127	0.0001
6		0.0000	0.0002	0.0005	0.0504	0.1462	0.1221	0.0255	0.0002
7			0.0000	0.0001	0.0216	0.1044	0.1396	0.0437	0.0005
8				0.0000	0.0081	0.0653	0.1396	0.0655	0.0013
9					0.0027	0.0363	0.1241	0.0874	0.0029
10					0.0008	0.0181	0.0993	0.1048	0.0058
11					0.0002	0.0082	0.0722	0.1144	0.0106
12					0.0001	0.0034	0.0481	0.1144	0.0176
13					0.0000	0.0013	0.0296	0.1056	0.0271
14						0.0005	0.0169	0.0905	0.0387
15						0.0002	0.0090	0.0724	0.0516
16						0.0000	0.0045	0.0543	0.0646
17							0.0021	0.0383	0.0760
18							0.0009	0.0255	0.0844
19							0.0004	0.0161	0.0888
20							0.0002	0.0097	0.0888
21							0.0001	0.0055	0.0846
22							0.0000	0.0030	0.0769
23								0.0016	0.0669
24								0.0008	0.0557
25								0.0004	0.0446
26								0.0002	0.0343
27								0.0001	0.0254
28								0.0000	0.0181
29									0.0125

Die Wahrscheinlichkeit, dass beim Vorliegen von $\lambda = 3$ keine, d. h. nur Null Fälle beobachtet werden, beträgt rund 0,05. Aufschlussreich sind die für steigende Werte λ drastisch abfallenden

Wahrscheinlichkeiten für Null-Ereignisse: von 82% ($\lambda = 0,2$) über 37% ($\lambda = 1$) und 5% ($\lambda = 3$) zu den nicht tabellierten Werten 1% ($\lambda = 4,6$) und 0,1% ($\lambda = 6,9$).

Mehrere aufeinanderfolgende Einzelwahrscheinlichkeiten können anhand der **Rekursionsformeln** (5.57) schnell berechnet werden:

$$P(X = x + 1) = \frac{\lambda}{x + 1} P(X = x) \quad \text{bzw.} \quad P(X = x - 1) = \frac{x}{\lambda} P(X = x) \quad (5.57)$$

Für die Berechnung ist außerdem wichtig die **Verteilungsfunktion** der Poisson-Verteilung in (5.58)

$$\begin{aligned} P(X \leq x) = F(x) &= e^{-\lambda} \sum_{k \leq x} \frac{\lambda^k}{k!} \quad \text{für } x \geq 0 \\ F(x) &= 0 \quad \text{für } x < 0 \end{aligned} \quad (5.58)$$

sowie:

$$P(X \geq x) = 1 - P(X \leq x - 1).$$

Beispielsweise für $\lambda = 1$:

$$\begin{aligned} P(X \geq 2|1) &= 1 - P(X \leq 2 - 1) = 1 - P(X \leq 1) = 1 - (0,368 + 0,368) \\ &= 1 - 0,736 = 0,264 = P(X > 1) \end{aligned}$$

In **R** werden Berechnungen zum Modell der Poisson-Verteilung durch die Funktionen **dpois()**, **ppois()**, **qpois()** und **rpois()** unterstützt. 

Beispiel (Geburtstagsproblem): Wie groß ist die Wahrscheinlichkeit, dass von 1000 Personen a) keiner, b) eine Person, c) zwei, d) drei Personen an einem bestimmten Tag Geburtstag haben? 

Da $q = \frac{364}{365} \simeq 1$, kann $\lambda = np = 1000 \cdot \frac{1}{365} = 2,7397$ geschätzt werden. Wir vereinfachen und setzen $\lambda = 2,74$.

$$P(X = 0) = \frac{\lambda^0 e^{-\lambda}}{0!} = e^{-\lambda} = e^{-2,74} = 0,06457 \simeq 0,065$$

$$P(X = 1) = \frac{\lambda^1 e^{-\lambda}}{1!} = \lambda e^{-\lambda} \simeq 2,74 \cdot 0,065 = 0,178$$

$$P(X = 2) = \frac{\lambda^2 e^{-\lambda}}{2!} = \frac{\lambda e^{-\lambda}}{2} \simeq \frac{2,74^2 \cdot 0,065}{2} = 0,244$$

$$P(X = 3) = \frac{\lambda^3 e^{-\lambda}}{3!} = \frac{\lambda^3 e^{-\lambda}}{6} = \frac{2,74^3 \cdot 0,065}{6} = 0,223$$

```
> dpois(0:3, 2.7397)
[1] 0.06458972 0.17695646 0.24240380 0.22137123
```

Liegt eine Stichprobe von 1000 Personen vor, so besteht eine Wahrscheinlichkeit von etwa 6%, dass keine Person an einem bestimmten Tag Geburtstag hat; die Wahrscheinlichkeit dafür, dass eine, zwei bzw. drei Personen an einem bestimmten Tage Geburtstag haben, sind rund 18%, 24% und 22%. Mit der Rekursionsformel (5.57) ergibt sich folgende Vereinfachung:

$P(0) = (\text{vgl. oben}) \simeq 0,065$	$P(1) \simeq \frac{2,74}{1} \cdot 0,065 = 0,178$	$P(2) \simeq \frac{2,74}{2} \cdot 0,178 = 0,244$	$P(3) \simeq \frac{2,74}{3} \cdot 0,244 = 0,223$

Multipliziert man diese Wahrscheinlichkeiten mit n , dann erhält man die durchschnittliche Anzahl der Personen, die in Stichproben von je 1000 Personen an einem bestimmten Tag Geburtstag haben.



Beispiel (Unverträglichkeit eines Serums): Die Wahrscheinlichkeit, dass ein Patient die Injektion eines gewissen Serums nicht verträgt, sei 0,001. Gefragt ist nach der Wahrscheinlichkeit, dass von 2000 Patienten a) genau drei, b) mehr als zwei Patienten die Injektion nicht vertragen. Da $q = 0,999 \simeq 1$, erhalten wir für $\hat{\lambda} = n \cdot p = 2000 \cdot 0,001 = 2$.

$$P(x \text{ Pat. vertragen die Inj. nicht}) = \frac{\lambda^x e^{-\lambda}}{x!} = \frac{2^x e^{-2}}{x!}$$

$$\text{a) } P(3 \text{ Pat. vertragen die Inj. nicht}) = \frac{2^3 e^{-2}}{3!} = \frac{4}{3e^2} = 0,180$$

$$\text{b) } P(0 \text{ Pat. vertragen die Inj. nicht}) = \frac{2^0 e^{-2}}{0!} = \frac{1}{e^2} = 0,135$$

$$P(1 \text{ Pat. verträgt die Inj. nicht}) = \frac{2^1 e^{-2}}{1!} = \frac{2}{e^2} = 0,271 = \frac{2}{e^2} = \frac{2^2 e^{-2}}{2!} = P(2 \text{ Pat. } \dots)$$

x	0	1	2	3	4	5	6	7	>7
$P(x)$	0,135	0,271	0,271	0,180	0,090	0,036	0,012	0,003	0,002

$$\begin{aligned} P(\text{mehr als 2 Pat. vertragen die Inj. nicht}) &= 1 - P(0 \text{ oder } 1 \text{ oder } 2 \text{ Pat. v. d. I. n}) \\ &= 1 - (1/e^2 + 2/e^2 + 2/e^2) = 1 - \frac{5}{e^2} = 0,323. \end{aligned}$$

```
> dpois(3, 2)
[1] 0.1804470
> 1-ppois(2, 2)
[1] 0.3233236
```

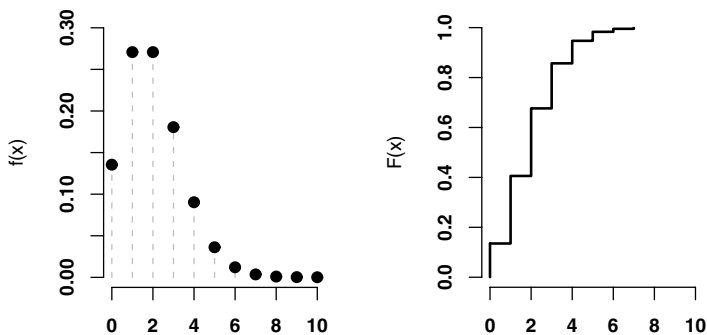


Abb. 5.11. Poisson-Verteilung zur Anzahl der Unverträglichkeiten mit $\lambda = 2$

Liegt eine größere Anzahl Stichproben zu je 2000 Patienten vor, dann dürften mit einer Wahrscheinlichkeit von etwa 18% drei Patienten und mit einer Wahrscheinlichkeit von etwa 32% mehr als zwei Patienten die Injektion nicht vertragen. Die Berechnung allein der Aufgabe a) mit Hilfe der Binomialverteilung wäre recht umständlich gewesen:

$$P(3 \text{ P. v. d. I. n.}) = \binom{2000}{3} \cdot 0,001^3 \cdot 0,999^{1997} = 0,1805; \text{ Aufgabe b) } P(> 2 \text{ Pat. } \dots) = 0,3233.$$

5.3.4.1 Der Dispersionsindex

Soll eine empirische Verteilung durch eine Poisson-Verteilung beschrieben werden, dann müssen die Daten die folgenden beiden Voraussetzungen erfüllen:

- (a) Es liegen unabhängige Ereignisse vor.
- (b) Die mittlere Zahl dieser Ereignisse pro Intervall (z. B. Zeit, Raum) ist der Länge des Intervalls proportional (und hängt nicht ab von der Lage des Intervalls).

Sind diese Bedingungen nicht oder nur teilweise erfüllt, dann ist die **Nullklasse** häufig stärker (schwächer) besetzt, als aufgrund der Poisson-Verteilung zu erwarten ist. Dann ist auch der Quotient (5.59) größer (kleiner) als 1.

$$\frac{\text{Stichprobenvarianz}}{\text{Theoretische Poisson-Varianz}} = \frac{\text{Stichprobenvarianz}}{\text{Theoretischer Poisson-Mittelwert}} = \frac{s^2}{\lambda} \quad (5.59)$$

Erwartungswert

Wie viel größer als 1 muss nun dieser Quotient sein, bevor wir entscheiden, dass die Verteilung „überdispers“ sei? Ist der Quotient $\approx 10/9$, dann kann angenommen werden, dass die vorliegende Verteilung durch eine Poisson-Verteilung approximiert werden kann. Ist er $\leq 9/10$, „underdispersed“ Verteilung, so liegt eher eine Binomialverteilung vor. Das nächste Beispiel wird uns Gelegenheit geben, diese Faustregel anzuwenden. Zur Prüfung, ob Daten (x_i) einer Poisson-Verteilung (mit dem Erwartungswert λ) entstammen dient der **Dispersionsindex** wobei

$$\hat{\chi}^2 = \sum_i \frac{(x_i - \bar{x})^2}{\bar{x}} = \frac{\sum_i x_i^2}{\bar{x}} - \sum_i x_i = \frac{n}{k} \left(\sum_{i=1}^n x_i^2 \right) - k \quad \begin{array}{l} \text{mit } k \text{ Ereignissen} \\ \text{in } n \text{ untersuchten} \\ \text{Intervallen:} \\ k/n = \bar{x} \end{array}$$

$$\hat{\chi}^2 = \frac{1}{\bar{x}} \sum_i f_i (x_i - \bar{x})^2 \quad \nu = n - 1 \quad \text{Dispersionsindex} \quad (5.60)$$

$n - 1$ Freiheitsgrade zur Verfügung stehen. Überschreitet der empirisch geschätzte $\hat{\chi}^2$ -Wert den tabellierten, ist also die Varianz wesentlich größer als der Mittelwert, dann liegt eine **zusammengesetzte Poisson-Verteilung** vor: Wenn überhaupt ein seltenes Ereignis eintritt, dann folgen häufig mehrere. Man spricht von positiver Wahrscheinlichkeitsansteckung. Tage mit Gewittern sind selten, treten aber gehäuft auf. Man erhält z. B. die so genannte **negative Binomialverteilung**, die im folgenden Abschnitt ausführlich behandelt wird.

Beispiel (Pferdehufschlagtote): Ein klassisches Beispiel für eine Poisson-Verteilung ist der Tod von Soldaten durch Pferdehufschlag in 10 preußischen Kavallerieregimentern während eines Zeitraumes von 20 Jahren (Preußisches Heer, 1875–1894) nach L. von Bortkiewicz [Bor98].

Tabelle 5.6. Tod durch Pferdehufschlag in 10 preußischen Kavallerieregimentern

Todesfälle	0	1	2	3	4	≥ 5	Σ
beobachtet	109	65	22	3	1	0	200
berechnet	108,7	66,3	20,2	4,1	0,6	0,1	200



$$\bar{x} = \frac{\sum x_i f_i}{n} = \frac{0 \cdot 109 + 1 \cdot 65 + 2 \cdot 22 + 3 \cdot 3 + 4 \cdot 1 + 5 \cdot 0}{200} = \frac{122}{200} = 0,61;$$

$$s^2 = \frac{\sum x_i^2 f_i - (\sum x_i f_i)^2 / n}{n - 1}$$

$$s^2 = \frac{0^2 \cdot 109 + 1^2 \cdot 65 + 2^2 \cdot 22 + 3^2 \cdot 3 + 4^2 \cdot 1 - 122^2 / 200}{200 - 1}$$

$$s^2 = \frac{196 - 74,42}{199} = \frac{121,58}{199} = 0,61; \text{ wir erhalten}$$

nach (5.59): $\frac{s^2}{\lambda} = \frac{0,61}{0,61} = 1 < \frac{10}{9}$ und

nach (5.60): $\hat{\chi}^2 = [109(0 - 0,61)^2 + 65(1 - 0,61)^2 + \dots + 0(5 - 0,61)^2] / 0,61$

$$\hat{\chi}^2 = 199,3 < 233 = \chi_{199;0,05}^2$$

Damit ist die Poisson-Verteilung ($\lambda = 0,61$) geeignet, die vorliegende Verteilung zu beschreiben. Im allgemeinen werden sich die Schätzungen von s^2 und λ unterscheiden.

Der Dispersionsindex liegt nahe bei 1, sobald die Ziffern 0, 1 und 2 häufig auftreten (insbesondere die Null und die Eins), selbst dann, wenn unterschiedliche Parameter λ_i vorliegen.

Für die vorliegende Poisson-Verteilung lassen sich nun aufgrund von $\lambda = 0,61$ Wahrscheinlichkeiten und erwartete Häufigkeiten berechnen.

$$P(0) = \frac{0,61^0 e^{-0,61}}{0!} = 0,5434; \quad 200 \cdot 0,5434 = 108,68 \text{ usw.}$$

```
> lambda <- 0.61
> n <- 200
> round(dpois(0:5, lambda) * n, 1)
[1] 108.7 66.3 20.2 4.1 0.6 0.1
```

Die relativen Häufigkeiten der Wahrscheinlichkeiten der Poisson-Verteilung sind durch die aufeinanderfolgenden Glieder der Beziehung

$$e^{-\lambda} \sum \frac{\lambda^x}{x!} = e^{-\lambda} \left(1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \dots + \frac{\lambda^x}{x!} \right) \quad (5.61)$$

gegeben. Die erwarteten Häufigkeiten erhält man als Produkt aus Einzelglied und gesamtem Stichprobenumfang. Man erhält also beispielsweise als Erwartungshäufigkeit für das dritte Glied

$$n \cdot e^{-\lambda} \cdot \frac{\lambda^2}{2!} = 200 \cdot 0,54335 \cdot \frac{0,3721}{2} = 20,2 \text{ usw.}$$

Liegen empirische Verteilungen vor, die Ähnlichkeit mit Poisson-Verteilungen aufweisen, dann kann λ , wenn die Nullklasse (Null Erfolge) die stärkste Besetzung aufweist, nach

$$-\ln \left(\frac{\text{Besetzung der Nullklasse}}{\text{Gesamtzahl aller Häufigkeiten}} \right) = \hat{\lambda} = -\ln \left(\frac{n_0}{n} \right) \quad (5.62)$$

geschätzt werden.

Auf das Pferdehufschlagbeispiel angewandt, erhalten wir anhand der Schnellschätzung

$$\hat{\lambda} = -\ln\left(\frac{109}{200}\right) = -\ln 0,545 = 0,60697 \quad \text{oder} \quad 0,61,$$

ein ausgezeichnetes Ergebnis!

5.3.4.2 Approximation der Poissonverteilung durch die Standardnormalverteilung

Die kumulierte Poisson-Wahrscheinlichkeit $P(X \leq k|\lambda) = \sum_{j=0}^k e^{-\lambda} \lambda^j / j!$ lässt sich nach (5.63) und wesentlich besser nach (5.64) [Mol70] approximieren.

Für $\lambda \geq 9$ gilt:

$$\hat{z} = |(k - \lambda) / \sqrt{\lambda}| \quad (5.63)$$

Beispiele 1. Für $P(X \leq 3|9)$ mit $\hat{z} = |(3-9)/\sqrt{9}| = 2,000$ erhält man $P = 0,0228$ (exakt: 0,021226).

2. Für $P(X \leq 4|10)$ mit $\hat{z} = |(4-10)/\sqrt{10}| = 1,897$ ergibt sich $P = 0,0289$ (exakt: 0,029253).

Für $\lambda \gtrapprox 5$ gilt:

$$\begin{aligned} \hat{z} &= |2\sqrt{k + (t+4)/9} - 2\sqrt{\lambda + (t-8)/36}| \\ \text{mit } t &= (k - \lambda + 1/6)^2 / \lambda \end{aligned} \quad (5.64)$$

Obiges 2. Beispiel: $t = (4-10+1/6)^2/10 = 3,403$

$$\hat{z} = |2\sqrt{4+7,403/9} - 2\sqrt{10-4,597/36}| = 1,892, \text{ d. h. } P = 0,0293.$$

5.3.5 Negative Binomial-Verteilung

Urnenmodell: Ziehen mit Zurücklegen aus einer dichotomen Urne solange, bis eine bestimmte Anzahl von Erfolgen (r) erstmals beobachtet wird. Die Anzahl der Wiederholungen ist somit nicht fest vorgegeben.

Genau $k+r$ Versuche sind notwendig, wenn beim $(k+r)$ -ten Versuch ein Erfolg eintritt und bei den vorangehenden $k+r-1$ Versuchen der Erfolg genau $(r-1)$ -mal auftrat. Die Wahrscheinlichkeit dafür, dass bei den ersten $k+r-1$ Versuchen der Erfolg genau $(r-1)$ -mal auftrat, kann mit der Binomialverteilung berechnet werden:

$$\binom{k+r-1}{r-1} p^{r-1} (1-p)^k = \binom{k+r-1}{k} p^{r-1} (1-p)^k$$

Da die Erfolgswahrscheinlichkeit auch bei dem $(k+r)$ -ten Versuch wieder p ist und die Versuche unabhängig voneinander sind, gilt die Wahrscheinlichkeitsfunktion:

$$P(X = k) = \begin{cases} \binom{k+r-1}{k} p^r (1-p)^k & \text{für } k = 0, 1, 2, \dots \\ 0 & \text{sonst} \end{cases} \quad (5.65)$$

Neben der „Erfolgswahrscheinlichkeit“ p wird diese Funktion zusätzlich auch durch die Zahl r der zu erzielenden Erfolge bestimmt. Der Name für diese Verteilung ist aus der Binomial-Entwicklung der folgenden speziellen Reihe abzuleiten:

$$p^{-r} = (1 - q)^{-r} = \sum \binom{r + k - 1}{k} q^k = \sum \binom{-r}{k} (-q)^k$$

Eine andere Form der Darstellung von (5.65) erinnert besonders an die Binomialverteilung.

$$P(X = k) = (-1)^k \binom{-r}{k} p^r (1 - p)^k$$

Hat man einzelne Wahrscheinlichkeiten $P(X = k)$ zu berechnen, so kann man auch hier eine **Rekursionsformel** (5.66) verwenden. Einfacher ist die Verwendung der entsprechenden Funktionen in **R**.

$$P(X = k + 1) = \frac{(r + k)(1 - p)}{k + 1} \cdot P(X = k) \quad (5.66)$$



Beispiel: Mit welcher Wahrscheinlichkeit müssen bei einer Lotterie bis zum 3. Gewinnlos eine bestimmte Anzahl 'Nieten' in Kauf genommen werden, wenn die Wahrscheinlichkeit für einen Gewinn 0,20 ist (jedes 5. Los gewinnt, vgl. auch Abbildung 5.12)?

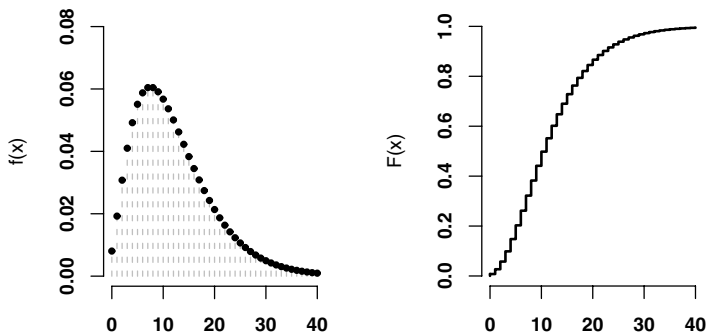


Abb. 5.12. Negative Binomialverteilung zur Anzahl der 'Nieten' bis zum 3. Gewinn ($p=0,20$)

Die Wahrscheinlichkeit dafür, spätestens beim 10ten Los den 3. Gewinn (7 Nieten) zu erzielen beträgt:

$$\sum_{i=0}^7 \binom{i+3-1}{i} 0.2^3 0.8^i = 0.3222$$

Ergänzend sollen einige Berechnungen in **R** angegeben werden. Für die negative Binomialverteilung stehen insbesondere die Funktionen **dnbinom()**, **pnbinom()**, **qnbinom()** und **rnbinom()** zur Verfügung.

```
> choose(7+3-1, 7) * 0.2^3 * 0.8^7
[1] 0.06039798
> dnbinom(7, 3, 0.2)
[1] 0.06039798
> p <- rep(NA, 8)
> for (i in 0:7) p[i+1] <- choose(i+3-1, i) * 0.2^3 * 0.8^i; sum(p)
[1] 0.3222005
> pnbinom(7, 3, 0.2)
[1] 0.3222005
```

Die Wahrscheinlichkeit dafür, den k ten Erfolg beim x ten Versuch zu erzielen, vorangegangen sind somit $k - 1$ Erfolge aus $x - 1$ Versuchen, kann auch nach (5.67) berechnet werden.

$$P(X = k) = \binom{x-1}{k-1} p^x (1-p)^{x-k} \quad \text{für } x = k, k+1, \dots \quad (5.67)$$

Beispiel: Ein Junge wirft Steine nach einem Ziel. Wie groß ist die Wahrscheinlichkeit dafür, dass sein 10. Wurf der 5. Treffer ist, wenn die Trefferwahrscheinlichkeit 0,4 beträgt?



$$P(X = k) = \binom{x-1}{k-1} p^x (1-p)^{x-k} = \binom{9}{4} 0,4^5 0,6^5 = 0,10$$

```
> p <- 0.4; x <- 10; k <- 5
> choose(x-1, k-1)*p^k*(1-p)^(x-k)
[1] 0.1003291
> # Zahl der Fehlversuche x-k und der Zahl der Erfolge k #
> dnbinom(x-k, k, 0.4)
[1] 0.1003291
```

Erwartungswert und Varianz einer negativen Binomialverteilung sind in (5.68) angegeben. Im Gegensatz zur Binomialverteilung ist die Varianz der negativen Binomialverteilung stets größer als ihr Erwartungswert.

Erwartungswert: $\mu = \frac{r(1-p)}{p}$

Varianz: $\sigma^2 = \frac{r(1-p)}{p^2}$

beachte $\mu < \sigma^2$

2. Moment: $\frac{r}{p} \cdot \left(\frac{r}{p} + \frac{1}{p} - 1 \right)$

(5.68)

Beispiel (Exposition und Krankheit): Eine klassische Sicht auf dieses spezielle Verteilungsmodell geht auf Greenwood und Yule [GY20] zurück.



Tabelle 5.7. Unfälle innerhalb 5 Wochen unter 647 Frauen, die mit hochexplosiven Sprengkapseln bei der Herstellung von Munition arbeiteten

Anzahl Unfälle	0	1	2	3	4	≥ 5	$\sum$
beobachtet	447	132	42	21	3	2	647
Poisson-Vert. (berechnet)	406	189	44	7	1	0	647
neg. Binomial-Vert. (berechnet)	443	139	44	14	5	2	647

Betrachtet wird das Auftreten einer Krankheit oder eines Ereignisses (z.B. Unfall) unter einer fortlaufenden Exposition. Kann die Erkrankung fatal (tödlich) enden, dann wird der Anteil der Personen, die unter der k ten Exposition versterben, aus den Patienten hervorgehen, die unter den vorangegangenen $(k - 1)$ Expositionen die Krankheit $(r - 1)$ mal überlebt haben und die die Krankheit unter der k ten Exposition das r te Mal erleiden.

Varianz und Mittelwert aus den beobachteten Daten sind nicht gleich! Das erklärt die recht starken Abweichungen zu den nach dem Modell der Poisson-Verteilung mit $\lambda = 0,47$ berechneten Häufigkeiten, insbesondere die stark besetzte „Nullklasse“ kann mit einer Poissonverteilung nicht erfasst werden.

```
> k <- c( 0, 1, 2, 3, 4, 5)
> obs <- c(447, 132, 42, 21, 3, 2); n <- sum(obs)
```



```
> m <- sum(obs*k)/n; round(m, 2)           # Mittelwert (Erwartungswert)
[1] 0.47
> round(dpois(k, m) * n, 0)                 # Poisson- Verteilung
[1] 406 189 44 7 1 0
> v <- sum((obs*(k - m)^2))/(n-1); v        # (emp.) Varianz
[1] 0.6919002
> p <- m / v; r <- m*p/(1-p)                # Modellparameter
> round(dnbinom(k, r, p) * n, 0)           # negative Binomialvert.
[1] 443 139 44 14 5 2
```

Wesentlich besser ist dagegen die Modellrechnung an Hand der negativen Binomialverteilung mit den Parametern $p = 0,67$ und $r = 0,95$, die sich aus den beobachteten Häufigkeiten schätzen lassen.

$$\begin{aligned}\hat{p} &= \frac{\bar{x}}{s^2} \\ \hat{r} &= \frac{\bar{x}\hat{p}}{1 - \hat{p}}\end{aligned}\tag{5.69}$$

Zahlreiche Anwendungen der negativen Binomialverteilung, besonders in der Ökonometrie, basieren auf einer **Parametrisierung der Wahrscheinlichkeitsfunktion** mit Hilfe des Erwartungswertes (Ehrenberg [Ehr86]).

Mit

$$p = \frac{r}{\mu + r}$$

wird aus (5.65):

$$P(X = k) = \binom{k+r-1}{k} \left(\frac{r}{\mu+r}\right)^r \left(\frac{\mu}{\mu+r}\right)^k\tag{5.70}$$

In dieser Darstellung ist zu erkennen, dass die Verteilung neben dem Erwartungswert μ durch einen weiteren Parameter r (häufig auch mit s (engl. size) bezeichnet) bestimmt wird. Die negative Binomialverteilung kann daher als eine **Verallgemeinerung der Poissonverteilung** betrachtet werden. Sie stellt dieselben Voraussetzungen an das Zufallsexperiment, erfordert aber nicht, dass die „durchschnittliche Rate“ konstant ist. Damit ergeben sich für die negative Binomialverteilung mehr Möglichkeiten der Anwendung. Die beiden Parameter der negativen Binomialverteilung können aus den Beobachtungen durch den arithmetischen Mittelwert ($\bar{x}$) und die empirische Varianz (s^2) wie folgt geschätzt werden:

$$\begin{aligned}\hat{\mu} &= \bar{x} \\ \hat{r} &= \frac{\bar{x}^2}{s^2 - \bar{x}}\end{aligned}\tag{5.71}$$



Beispiel: Die Anzahl der Käufe eines Markenartikels nach dem Prozentsatz der kaufenden Haushalte ist in Tabelle 5.8 angegeben (Ehrenberg [Ehr86]). Die beobachteten Anteile können mit dem Modell einer negativen Binomialverteilung (mit den Parametern $\mu = 3,4$ und $r = 0,5$) wesentlich besser erklärt werden als mit einer Poisson-Verteilung (mit dem Parameter $\lambda = 3,4$).

Tabelle 5.8. Anzahl der Käufe eines Markenartikels; Prozentsatz der kaufenden Haushalte

kaufende Haushalte (%)	Anzahl der Käufe in einem halben Jahr										
	0	1	2	3	4	5	6	7	8	9	10+
beobachtet	39	14	10	6	4	4	3	3	2	2	13
neg. binomial	36	16	10	7	6	4	4	3	2	2	2(+8)
Poisson	3	11	19	22	19	13	7	3	1	1	0

Die Käufe folgen einer Mischung von Poisson-Verteilungen, da verschiedene Verbraucher unterschiedlich oft kaufen. Insbesondere die extreme Schiefe wegen der stark besetzten „Null“-Klasse kann durch die Poisson-Verteilung nicht ausreichend erfasst werden.

Die Rechnungen sind mit Hilfe der entsprechenden Funktionen **dnbinom()** und **dpois()** in **R** einfach nach zu vollziehen.

```
> m <- 3.4; s <- 0.5; p <- s/(s+m)
>
> n=100; x <- 0:10
> round(dnbinom(x, s, p)*n, 0)
[1] 36 16 10 7 6 4 4 3 2 2 2
> round(dpois(x, m)*n, 0)
[1] 3 11 19 22 19 13 7 3 1 1 0
```

Beispiel: Ein frühes Anwendungsbeispiel der negativen Binomialverteilung geht auf R.A. Fisher zurück. Die Anzahl der beobachteten Zecken auf Schafen ist in der folgenden Tabelle zusammengestellt.



Anzahl der Zecken	0	1	2	3	4	5	6	7	8	9	10+	
Anzahl der Schafe (beobachtet)	7	9	8	13	8	5	4	3	0	1	2	60
Anzahl der Schafe (erwartet)	6	10	11	10	8	6	4	2	1	1	1	60

Fisher konnte zeigen, dass die Anzahl der Zecken X (Zufallsvariable), die man an einem Schaf findet, erstaunlich gut durch die Wahrscheinlichkeitsfunktion einer negativen Binomialverteilung mit den Parametern $p = 0,55$ und $r = 3,96$ beschrieben werden kann. Seine Beobachtung basierte auf einer Zahl von 60 Schafen, an denen ca. 200 Zecken gefunden wurden.

```
> beob <- c(rep(0,7), rep(1,9), rep(2,8), rep(3,13), rep(4,8), rep(5,5), rep(6,4),
+ rep(7,3), rep(8,0), rep(9,1), 10, 10)
> r.hat <- mean(beob)^2/(var(beob)-mean(beob)); r.hat
[1] 3.956746
> p.hat <- r.hat/(mean(beob)+r.hat); p.hat
[1] 0.5490336
> round(dnbinom(0:11, 3.96, 0.55)*60, 0)
[1] 6 10 11 10 8 6 4 2 1 1 1 0
```

5.3.5.1 Geometrische Verteilung

Die geometrische Verteilung ist ein Spezialfall der negativen Binomialverteilung. Hier wird eine Serie von Bernoulli-Versuchen so lange durchgeführt, bis das erste Mal ein Erfolg eintritt. Der Wertebereich der Zufallsvariablen, X -Anzahl der Versuche bis zum ersten Erfolg, ist abzählbar unendlich $W = 1, 2, 3, \dots$

$$P(X = k) = p \cdot (1 - p)^{k-1} \quad (5.72)$$

Die geometrische Verteilung ist eine der wenigen diskreten Verteilungsmodelle, in denen auch die Verteilungsfunktion explizit angegeben werden kann:

$$F(n) = P(X \leq n) = 1 - (1 - p)^n \quad \text{für } n = 1, 2, 3, \dots \quad (5.73)$$

Beispiel: Der erste Auftreten einer „Sechs“ im Spiel „Mensch ärgere Dich nicht“ kann mit dem Modell einer geometrischen Verteilung beschrieben werden. Wahrscheinlichkeitsfunktion und Verteilungsfunktion (nach (5.72) und (5.73)) sind in Abbildung 5.13 für die ersten 20 Versuche dargestellt.



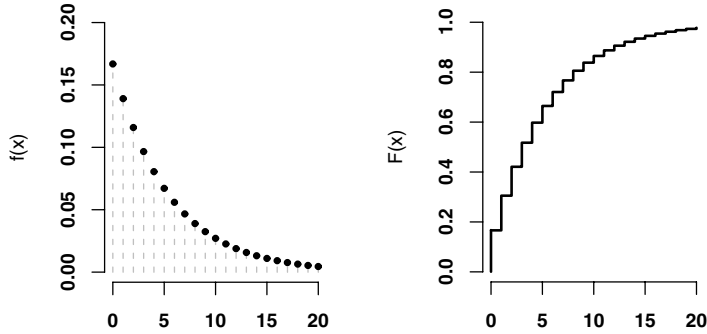


Abb. 5.13. Geometrische Verteilung zur Anzahl der Würfe bis zur 1. Sechs bis $n=20$

Erwartungswert und Varianz der geometrischen Verteilung sind in (5.74) angegeben (vgl. auch die negative Binomialverteilung in (5.68) mit $r = 1$):

Erwartungswert: $\mu = \frac{1-p}{p}$ Varianz: $\sigma^2 = \frac{(1-p)}{p^2}$ 2. Moment: $\frac{1}{p} \cdot \left(\frac{2}{p} - 1\right)$	(5.74)

5.3.6 Hypergeometrische Verteilung

Urnenmodell: Für das Modell Ziehen ohne Zurücklegen aus einer (dichotomen) Urne des Umfangs N mit W (weißen) und S (schwarzen) Kugeln von $n \geq 1$ Kugeln wird die hypergeometrische Verteilung anstelle der Binomialverteilung verwendet.

Hinweis: Werden solange Kugeln gezogen, bis eine bestimmte Anzahl weißer Kugeln (z.B. k) erreicht ist, dann führt das Modell auf die negative hypergeometrische Verteilung.

$$P(X = k) = \begin{cases} \frac{\binom{W}{k} \binom{S}{n-k}}{\binom{N}{n}} & \text{für } \begin{matrix} \max(0; W + n - N) \leq k \\ \text{und } k \leq \min(n; W) \end{matrix} \\ 0 & \text{sonst} \end{cases} \quad (5.75)$$

Intensiv angewendet wird die hypergeometrische Verteilung bei Problemen, die mit der Qualitätsüberwachung zusammenhängen. Betrachten wir eine Urne mit $W = 5$ weißen und $S = 10$ schwarzen Kugeln. Gefragt ist nach der Wahrscheinlichkeit, zufällig genau $w = 2$ weiße und $s = 3$ schwarze Kugeln zu ziehen.

Wir erhalten für $P(2 \text{ von } 5 \text{ weißen Kugeln und } 3 \text{ von } 10 \text{ schwarzen Kugeln}) =$

$$\frac{\binom{5}{2} \binom{10}{3}}{\binom{15}{5}} = \frac{(5!/3! \cdot 2!)(10!/7! \cdot 3!)}{15!/10! \cdot 5!} = \frac{5 \cdot 4 \cdot 10 \cdot 9 \cdot 8 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{2 \cdot 1 \cdot 3 \cdot 2 \cdot 1 \cdot 15 \cdot 14 \cdot 13 \cdot 12 \cdot 11} = 0,3996,$$

eine Wahrscheinlichkeit von rund 40%.

In **R** stehen für Rechnungen nach dem Modell der hypergeometrischen Verteilung die Funktionen **dhyper()**, **phyper()**, **qhyper()** und **rhyper()** zur Verfügung.



```
> # dhyper(k, W, S, n)
> dhyper(2, 5, 10, 5)
[1] 0.3996004
```

Beispiel (Urnenmodell): Gegeben sei eine Urne mit 4 blauen (B) und 3 roten (R) Kugeln, dann gelten für die entsprechenden Ereignisse die Wahrscheinlichkeiten:



- (1) eine blaue Kugel zu ziehen $P(B) = 4/7$
- (2) eine rote Kugel zu ziehen $P(R) = 3/7$
- (3) eine blaue Kugel sei bereits entnommen, anschließend ist eine rote zu ziehen $P(R|B) = 3/6 = 1/2$
- (4) eine rote Kugel sei bereits entnommen, anschließend ist eine blaue zu ziehen $P(B|R) = 4/6 = 2/3$
- (5) entweder zuerst eine blaue und dann eine rote zu ziehen oder zuerst eine rote und dann eine blaue zu ziehen (Multiplikationssatz)

$$P(B) \cdot P(R|B) = P(R) \cdot P(B|R) = P(B \cap R) \\ (4/7)(1/2) = (3/7)(2/3) = 2/7$$

- (6) beide zugleich zu ziehen
d.h. sowohl eine von den 4 blauen (1 von 4) und eine von den 3 roten (1 von 3)

$$P(1 \text{ von } 4, 1 \text{ von } 3) = \frac{\binom{4}{1} \binom{3}{1}}{\binom{7}{2}} = 4 \cdot 3 / [7 \cdot 6 / (1 \cdot 2)] = 4/7.$$

Erwartungswert und Varianz der hypergeometrischen Verteilung sind in (5.76) angegeben:

Erwartungswert: $\mu = n \frac{W}{N} = np$ Varianz: $\sigma^2 = np(1-p) \frac{N-n}{N-1}$ 2. Moment: $np \frac{(N-W) + n(W-1)}{N-1}$	(5.76)

Ist n/N klein, so wird diese Verteilung praktisch mit der Binomialverteilung identisch. Dementsprechend strebt auch die Varianz gegen die der Binomialverteilung (vgl. $\frac{N-n}{N-1} \simeq 1 - \frac{n}{N} \simeq 1$ für $N \gg n$).

Die **verallgemeinerte hypergeometrische Verteilung** (polyhypergeometrische Verteilung)

$$P(n_1, n_2, \dots, n_k | N_1, N_2, \dots, N_k) = \frac{\binom{N_1}{n_1} \binom{N_2}{n_2} \dots \binom{N_k}{n_k}}{\binom{N}{n}} \quad (5.77)$$

gibt die Wahrscheinlichkeit an, dass in einer Stichprobe vom Umfang n gerade $n_1, n_2, \dots, n_k$ Beobachtungen mit den Merkmalen $A_1, A_2, \dots, A_k$ auftreten, wenn in der Grundgesamtheit vom Umfang N die Häufigkeiten dieser Merkmalsausprägungen $N_1, N_2, \dots, N_k$ betragen und $\sum_{i=1}^k N_i =$

N und $\sum_{i=1}^k n_i = n$ gelten. Die Parameter (für die n_i) sind:

Erwartungswerte: $\mu_i = n \frac{N_i}{N}$ Varianzen: $\sigma_i^2 = np_i(1 - p_i) \frac{N - n}{N - 1}$	(5.78)

Die hypergeometrische Verteilung kann u. a. im Rahmen der Qualitätsüberwachung und für die **Abschätzung des unbekannten Umfangs N einer Population** (z. B. Wildbestände) verwendet werden: N_1 Individuen einfangen, markieren und wieder frei lassen, danach n Individuen einfangen und die Zahl der markierten (n_1) feststellen; dann ist $\hat{N} \approx nN_1/n_1$ („Wildlife Tracking“).

Beispiel (Studenten): Nehmen wir an, wir hätten 10 Studenten, von denen 6 Biochemie und 4 Statistik studieren. Eine Stichprobe von 5 Studenten sei ausgewählt. Wie groß ist die Wahrscheinlichkeit, dass unter den 5 Studenten 3 Biochemiker und 2 Statistiker sind?

$$\begin{aligned}
 P(3 \text{ von } 6 \text{ B., } 2 \text{ von } 4 \text{ S.}) &= \frac{\binom{6}{3} \binom{4}{2}}{\binom{6+4}{3+2}} = \frac{(6!/[3! \cdot 3!])(4!/[2! \cdot 2!])}{10!/[5! \cdot 5!]} \\
 &= \frac{6 \cdot 5 \cdot 4 \cdot 4 \cdot 3 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{3 \cdot 2 \cdot 1 \cdot 2 \cdot 1 \cdot 10 \cdot 9 \cdot 8 \cdot 7 \cdot 6} = \frac{20}{42} = 0,4762.
 \end{aligned}$$

Die Wahrscheinlichkeit beträgt damit erwartungsgemäß fast 50%.

Beispiel (Lotto): Gegeben seien die ganzen Zahlen von 1 bis 49. Hiervon sind 6 zu wählen. Wie groß ist die Wahrscheinlichkeit dafür, vier richtige Zahlen gewählt zu haben?

$$P(4 \text{ von } 6, 2 \text{ von } 43) = \frac{\binom{6}{4} \binom{43}{2}}{\binom{49}{6}} = \frac{15 \cdot 903}{13\,983\,816}$$

```
> dhyper(4, 6, 43, 6)
[1] 0.0009686197
```

$$P \simeq \frac{13,545 \cdot 10^3}{13,984 \cdot 10^6} \simeq 0,969 \cdot 10^{-3}, \text{ d.h. knapp } 0,001.$$

Die Wahrscheinlichkeit, mindestens 4 richtige Zahlen zu wählen, liegt ebenfalls noch unter 1 Promille. Die Wahrscheinlichkeit, 6 richtige Zahlen zu wählen, beträgt $1 / \binom{49}{6} = 1/13\,983\,816 \approx 7 \cdot 10^{-8}$. Ein vollständiges Modell ist in Abbildung 5.14 wiedergegeben.

Beispiel (Ausschussware): Eine Grundgesamtheit aus 100 Elementen enthalte 5% Ausschuss. Wie groß ist die Wahrscheinlichkeit, in einer 50 Elemente umfassenden Stichprobe (a) kein bzw. (b) ein Ausschussstück zu finden?

$$\begin{aligned}
 \text{Zu a: } P(50 \text{ von } 95, 0 \text{ von } 5) &= \frac{\binom{95}{50} \cdot \binom{5}{0}}{\binom{95+5}{50+0}} = \frac{95! \cdot 5! \cdot 50! \cdot 50!}{50! \cdot 45! \cdot 5! \cdot 0! \cdot 100!} \\
 &= \frac{95! \cdot 50!}{45! \cdot 100!} \\
 &= \frac{1,0330 \cdot 10^{148} \cdot 3,0414 \cdot 10^{64}}{1,1962 \cdot 10^{56} \cdot 9,3326 \cdot 10^{157}} = 0,02823
 \end{aligned}$$

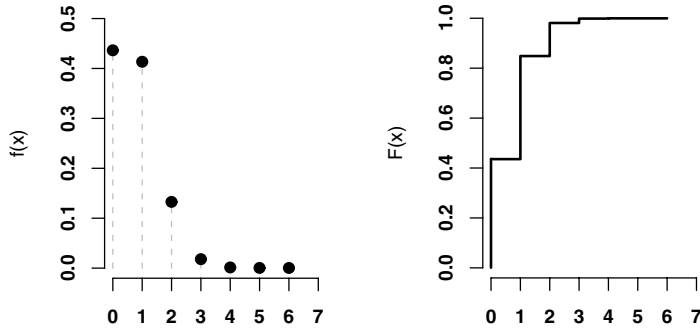


Abb. 5.14. Hypergeometrische Verteilung zur Anzahl der „Richtigen“ unter 6 aus 49 gezogenen Kugeln im Lotto

$$\begin{aligned} \text{Zu b: } P(49 \text{ von } 95, 1 \text{ von } 5) &= \frac{\binom{95}{49} \cdot \binom{5}{1}}{\binom{95+5}{49+1}} = \frac{95! \cdot 5! \cdot 50! \cdot 50!}{49! \cdot 46! \cdot 4! \cdot 1! \cdot 100!} \\ &= 5 \cdot \frac{95! \cdot 50! \cdot 50!}{49! \cdot 46! \cdot 100!} = 0,1529 \end{aligned}$$

```
> dhyper(50, 95, 5, 50)
[1] 0.02814225
> dhyper(49, 95, 5, 50)
[1] 0.152947
```

Beispiel (Annoncen in einer Zeitschrift): Werden im Laufe eines Jahres von $W = 52$ aufeinanderfolgenden Nummern einer Wochenzeitschrift $A = 10$ beliebige Ausgaben mit einer bestimmten Anzeige versehen, dann ist die Wahrscheinlichkeit, dass ein Leser von $w = 15$ beliebigen Nummern kein Heft mit einer Annonce erhält ($a = 0$)



$$P(a \text{ von } A, w \text{ von } W) = \frac{\binom{A}{a} \binom{W-A}{w-a}}{\binom{W}{w}}$$

Die „manuelle“ Rechnung über die entsprechenden Binomialkoeffizienten kann unter Verwendung der Logarithmen vereinfacht werden:

$$\begin{aligned} \text{oder } P(0 \text{ von } 10, 15 \text{ von } 52) &= \frac{\binom{10}{0} \binom{52-10}{15-0}}{\binom{52}{15}} & \lg 42! &= 51,14768 \\ & & \lg 15! &= 12,11650 \\ & & \lg 37! &= 43,13874 \\ & & \lg 52! &= 67,90665 \\ \text{d.h. vgl. } \binom{n}{0} &= 1, & & \\ P &= \frac{\binom{42}{15}}{\binom{52}{15}} = \frac{42! \cdot 15! \cdot 37!}{15! \cdot 27! \cdot 52!} & \lg 15! &= 12,11650 \\ & & \lg 27! &= 28,03698 \\ & & \lg 52! &= 67,90665 \\ & & \lg P &= 0,34279 - 2 \\ & & P &= 0,02202 \approx 2,2\% \end{aligned}$$

Damit beträgt die Wahrscheinlichkeit, mindestens eine Anzeige zu sehen, knapp 98%.

```
> dhyper(0, 10, 42, 15)
[1] 0.02201831
```

5.3.6.1 Approximationen der Hypergeometrischen Verteilung

1. Für großes N_1 und N_2 und im Vergleich hierzu kleines n ($n/N < 0,1$; $N \geq 60$) kann die hypergeometrische Verteilung durch die **Binomialverteilung** approximiert werden $p = N_1/(N_1 + N_2)$.

2. Für $np \geq 4$ kann

$$\hat{z} = (n_1 - np) / \sqrt{npq(N - n)/(N - 1)} \quad (5.79)$$

als angenähert **standardnormalverteilt** aufgefasst werden. Die kumulierte Wahrscheinlichkeit der hypergeometrischen Verteilung

$$P(X \leq k = n_1 | N; N_1; n) = \sum_{n_1=0}^{N_1} \binom{N_1}{n_1} \binom{N_2}{n_2} / \binom{N}{n}$$

lässt sich für $n \leq N_1 \leq N/2$ besser nach (5.80) approximieren ([Mol70]):

$$\hat{z} = \left| 2 \left[\sqrt{(k+0,9)(N-N_1-n+k+0,9)} - \sqrt{(n-k-0,1)(N_1-k-0,1)} \right] / \sqrt{N-0,5} \right| \quad (5.80)$$

In (5.80) ist für $0,05 \leq P \leq 0,93$ 0,9 durch 0,75; 0,1 durch 0,25 und 0,5 durch 0 zu ersetzen; für extremere P -Werte ersetze man 0,9 durch 1; 0,1 durch 0 und 0,5 durch 1. Beispiel: $P(X \leq 1 | 10; 5; 5) = 0,103$; $\hat{z}$ (nach 5.80) = 1,298, d. h. $P = 0,0971$; mit 0,75; 0,25 und 0 ergibt sich über $\hat{z} = 1,265$ $P = 0,103$.

3. Für kleines p , großes n und im Vergleich zu n sehr großes N ($n/N \leq 0,05$) lässt sich die hypergeometrische Verteilung durch die **Poisson-Verteilung** annähern ($\lambda = np$).

5.4 Stetige Verteilungen

- Gleichverteilung
- Normalverteilung
- Lognormalverteilung
- Exponentialverteilung
- Weibull-Verteilung

5.4.1 Gleichverteilung

Die konstante Wahrscheinlichkeitsdichte der stetigen Gleich- oder Rechteckverteilung (rectangular distribution) im Bereich von a bis b ist durch die Funktion

$$y = f(x) = \begin{cases} 1/(b-a) & \text{für } a < x < b \\ 0 & \text{für } x \leq a \text{ oder } x \geq b \end{cases} \quad (5.81)$$

gegeben; Erwartungswert und Varianz sind durch

Erwartungswert: $\mu = (a + b)/2$ Varianz: $\sigma^2 = (b - a)^2/12$ 2. Moment: $(a^2 + ab + b^2)/3$

(5.82)

definiert; außerdem gilt (neben $\tilde{\mu} = \mu$) für die Verteilungsfunktion:

$F(x) = \frac{x - a}{b - a}$

(5.83)

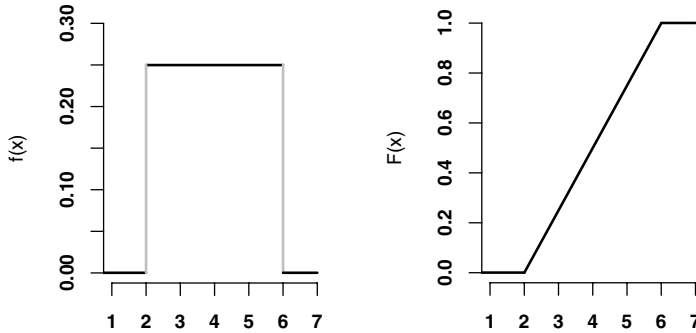


Abb. 5.15. Stetige Gleich- oder Rechteckverteilung im Intervall $[2, 6]$

Wichtig ist der Spezialfall (**Pseudozufallszahlen** im Intervall 0 bis 1): $0 < x < 1$; d. h. $f(x) = 1$ sowie

$$F(x) = \frac{x - 0}{1 - 0} = x; \quad \mu = (0 + 1)/2 = 1/2 = \tilde{\mu}; \quad \sigma^2 = (1 - 0)^2/12 = 1/12.$$

Die stetige Gleichverteilung hat in der angewandten Statistik eine gewisse Bedeutung: Einmal, wenn ein beliebiger Wert in einem Bereich von Werten gleichwahrscheinlich ist, zum anderen, für die Approximation relativ kleiner Spannweiten beliebiger kontinuierlicher Verteilungen. So ist z. B. die normalverteilte Variable X im Bereich

$\mu - \sigma/3 < X < \mu + \sigma/3$

(5.84)

angenähert gleichverteilt.

5.4.2 Normalverteilung

Stetige Zufallsvariable sind mitunter in ihrem mittleren Bereich angenähert **glockenförmig** verteilt. Eine typische Glockenkurve (vgl. Abb. 5.16) ist durch die Gleichung $y = e^{-x^2}$ (oder auch $y = \exp(-x^2)$) gegeben. Andere Glockenkurven werden durch

$y = a e^{-b x^2}$

(5.85)

(mit $a, b > 0$) dargestellt. In Abb. 5.16 sind die beiden Konstanten $a = b = 1$ bzw. $a = 5$ und $b = 1/3$: Eine Vergrößerung von a bewirkt eine Vergrößerung von y , die Kurve wird proportional vergrößert; eine Verkleinerung von b bewirkt ein „Flacherwerden“ der Glockenkurve.

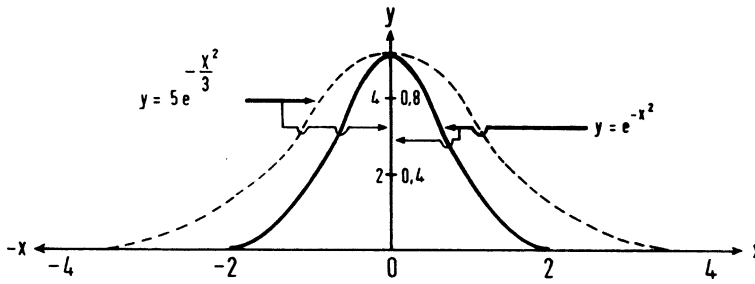


Abb. 5.16. Glockenkurven

Eine Familie von Verteilungen, die durch eine so genannte Glockenkurve mit dem Maximum an der Stelle μ und den Wendepunkten an den Stellen $\mu - \sigma$ und $\mu + \sigma$ charakterisiert ist (vgl. Abb. 5.17), hat als „**Normalverteilung**“ in der Statistik eine große Bedeutung: μ ist der Erwartungswert und Median der Verteilung, σ die Standardabweichung. Jede Normal- oder Gauß-Verteilung ist eingipflig, symmetrisch und nähert sich für sehr kleines und sehr großes x [asymptotisch] der x -Achse („Normal“ bedeutet nicht „häufig auftretend“).

Beispiele für angenähert normalverteilte Daten sind eigentlich selten; besonders selten sind sie in der Biologie. So soll das Gewicht in der 37.–43. Schwangerschaftswoche von nichtdiabetischen Müttern Geborener, die Körpergröße 18jähriger Männer und der Blutdruck gesunder 30–40jähriger Männer angenähert normalverteilt sein.

Die Normalverteilung gilt höchstens angenähert: auch nach dem Zentralen Grenzwertsatz kann man nur in gewissen Fällen eine ungefähre Normalverteilung erwarten und die Erfahrung mit großen Stichproben von Messdaten hoher Qualität zeigt häufig deutliche Abweichungen von der Normalverteilung in dem Sinne, dass beide Verteilungsenden stärker besetzt sind („heavy tailed distributions“).

Das **Sandexperiment**: Eine Normalverteilung lässt sich leicht experimentell annähern; trockenen Sand durch einen Trichter zwischen zwei parallele, senkrecht gestellte Glaswände einrinnen lassen.

Die zentrale Bedeutung der Normalverteilung besteht darin, dass eine **Summe von vielen unabhängigen, beliebig verteilten Zufallsvariablen gleicher Größenordnung angenähert normalverteilt ist**, und zwar um so besser angenähert, je größer ihre Anzahl ist (**Zentraler Grenzwertsatz**). Dieser Satz bildet die Grundlage dafür, dass Stichprobenverteilungen oberhalb eines bestimmten Stichprobenumfangs durch diese Verteilung approximiert werden können und dass für die entsprechenden Testverfahren die tabellierten Schranken der Standardnormalverteilung ausreichen.

Prinzipiell gesehen ist die Normalverteilung ein mathematisches Modell mit vielen günstigen mathematisch-statistischen Eigenschaften, das als ein Grundpfeiler der mathematischen Statistik angesehen werden kann. Seine grundlegende Bedeutung beruht darauf, dass sich viele zufällige Variable, die in der Natur beobachtet werden können, als Überlagerung vieler einzelner, weitgehend unabhängiger Einflüsse, von denen keiner dominieren darf, also als Summe vieler einzelner, voneinander unabhängiger zufälliger Variablen auffassen lassen (vgl. das Sandexperiment).

Die **Wahrscheinlichkeitsdichte der Normalverteilung** (vgl. Abb. 5.17) ist durch (5.86) gegeben.

$$y = f(x) = f(x|\mu, \sigma) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-1/2[(x-\mu)/\sigma]^2} \quad (-\infty < x < \infty, \quad -\infty < \mu < \infty, \quad \sigma > 0) \quad (5.86)$$

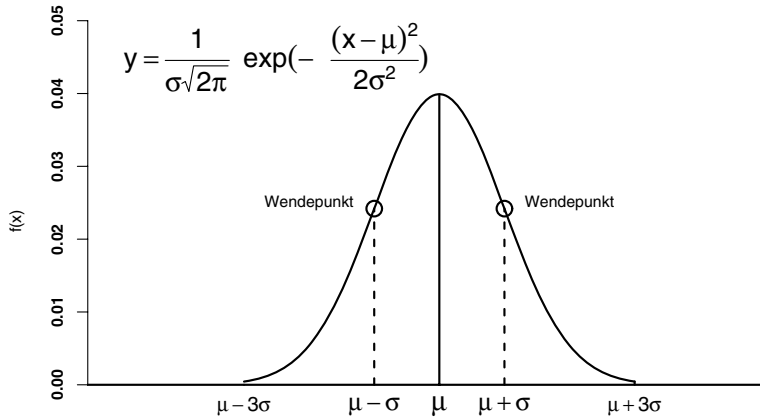


Abb. 5.17. Wahrscheinlichkeitsdichte y (Ordinate) für jeden Punkt x (Abszisse)

Hierin ist x eine beliebige Abszisse, y die zugehörige Ordinate (y ist eine Funktion von x : $y = f(x)$), σ die Standardabweichung der Verteilung, μ der Erwartungswert der Verteilung; π und e sind mathematische Konstanten mit den angenäherten Werten $\pi = 3,141593$ und $e = 2,718282$. Diese Formel enthält rechts die beiden Parameter μ und σ , die Variable x sowie die beiden Konstanten.

Wie (5.86) zeigt, ist die Normalverteilung durch die Parameter μ und σ vollständig charakterisiert. Der Erwartungswert μ bestimmt die Lage der Verteilung im Hinblick auf die x -Achse, die Standardabweichung σ die Form der Kurve (vgl. Abb. 5.17): Je größer σ ist, um so flacher ist der Kurvenverlauf, um so breiter ist die Kurve und um so niedriger liegt das Maximum. Dieses liegt an der Stelle des Erwartungswertes $E(X) = \mu$ [vgl. auch: $\text{Var}(X) = \sigma^2$; Schiefe $\gamma_1 = 0$ und Wölbung $\gamma_2 = 3$].

Normalverteilung: $N(\mu, \sigma)$	
$X = x$	$f(x \mu, \sigma)$
$-\infty$	0
$\mu - \sigma$	$\approx 0,6 \cdot y_{max}$
μ	$y_{max} = 1/[\sigma\sqrt{2\pi}]$
$\mu + \sigma$	$\approx 0,6 \cdot y_{max}$
$+\infty$	0
$f(\mu - x \mu, \sigma) = f(\mu + x \mu, \sigma)$ für jedes x ; Symmetrie	

Weitere Eigenschaften der Normalverteilung:

1. Die Kurve liegt symmetrisch zur Achse $x = \mu$, sie ist **symmetrisch** um μ . Die Werte $x' = \mu - a$ und $x'' = \mu + a$ haben die gleiche Dichte und damit denselben Wert y .
2. Das **Maximum** der Kurve beträgt $y_{max} = 1/(\sigma \cdot \sqrt{2\pi})$, für $\sigma = 1$ hat es den Wert $0,398942 \approx 0,4$. Für sehr großes x ($x \rightarrow \infty$) und sehr kleines x ($x \rightarrow -\infty$) strebt y gegen Null; die x -Achse stellt eine **Asymptote** dar. Sehr extreme Abweichungen vom Erwartungswert μ weisen eine so winzige Wahrscheinlichkeit auf, dass der Ausdruck „fast unmöglich“ gerechtfertigt erscheint.

3. Die Standardabweichung der Normalverteilung ist durch die **Abszisse der Wendepunkte** (Abb. 5.17) gegeben. Die Ordinate der Wendepunkte liegt bei etwa $0,6 \cdot y_{\max}$.

Rund **2/3** aller Beobachtungen liegen zwischen $\mu - \sigma$ und $\mu + \sigma$ bzw. im Bereich $\mu \pm \sigma$. Da μ und σ in der Formel für die Wahrscheinlichkeitsdichte der Normalverteilung beliebige Werte annehmen können, existieren **beliebig viele** unterschiedliche Normalverteilungen. Setzen wir in (5.86) folgende Transformation ein:

$$\boxed{\frac{X - \mu}{\sigma} = Z} \quad (5.87)$$

X ist in (5.87) dimensionsbehaftet - Z ist dimensionslos. Wir erhalten eine einzige, die **standardisierte Normalverteilung mit Erwartungswert Null und Standardabweichung Eins** [d. h. (5.86) geht wegen $f(x) dx = f(z) dz$ über in (5.88)]. Mit Abbildung 5.17 gilt für die Fläche unter der Kurve von $\mu - 3\sigma$ bis $\mu + 3\sigma$:

$$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = 0,9973 = P(-3 \leq Z \leq 3) = P(|Z| \leq 3).$$

Als Abkürzung für die Normalverteilung dient $N(\mu; \sigma)$ bzw. $N(\mu; \sigma^2)$, für die Standardnormalverteilung dementsprechend $N(0; 1)$. Für standardnormalverteilte Zufallsvariablen gilt somit: $\mu = 0$ und $\sigma = 1$. Angenähert **standardnormalverteilte Zufallsvariablen** erhält man, wenn von der Summe von 12 Zufallszahlen aus dem Bereich 0,0001 bis 0,9999 die Zahl 6 subtrahiert wird.

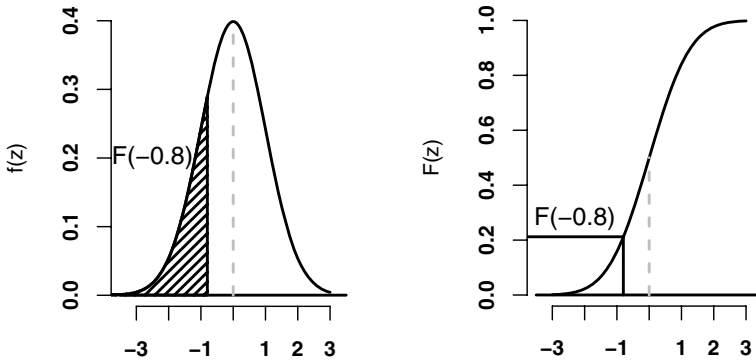


Abb. 5.18. Wahrscheinlichkeitsdichte und Verteilungsfunktion der Standardnormalverteilung $N(0, 1)$: $f(z)=0,2894$ und $F(z)=0,2119$

Die Wahrscheinlichkeitsdichte der Standardnormalverteilung ist durch (5.88) gegeben. Dabei wird diese spezielle Dichtefunktion häufig durch die Abkürzung $\varphi(z)$ [phi von z] besonders benannt.

$$\boxed{y = f(z) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{z^2}{2}} = \varphi(z)} \quad (5.88)$$

Die Werte der zugehörigen Verteilungsfunktion (5.89) [auch als $\Phi(z)$ (Phi von z) bezeichnet; vgl. (5.89): rechts] liegen tabelliert (Tabelle 5.9) vor oder können in dem Programm **R** direkt mit der Funktion **pnorm**(z, mean=0, sd=1) bestimmt werden.

$$\boxed{F(z) = P(Z \leq z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{\nu^2}{2}} d\nu = \Phi(z)} \quad (5.89)$$

Tabelle 5.9. Werte der Verteilungsfunktion $F(z)$ der Standardnormalverteilung im Bereich $[-2.99, 0]$; berechnet mit der Funktion **pnorm(z, mean=0, sd=1)** in R

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.49601	0.49202	0.48803	0.48405	0.48006	0.47608	0.47210	0.46812	0.46414
-0.1	0.46017	0.45620	0.45224	0.44828	0.44433	0.44038	0.43644	0.43251	0.42858	0.42465
-0.2	0.42074	0.41683	0.41294	0.40905	0.40517	0.40129	0.39743	0.39358	0.38974	0.38591
-0.3	0.38209	0.37828	0.37448	0.37070	0.36693	0.36317	0.35942	0.35569	0.35197	0.34827
-0.4	0.34458	0.34090	0.33724	0.33360	0.32997	0.32636	0.32276	0.31918	0.31561	0.31207
-0.5	0.30854	0.30503	0.30153	0.29806	0.29460	0.29116	0.28774	0.28434	0.28096	0.27760
-0.6	0.27425	0.27093	0.26763	0.26435	0.26109	0.25785	0.25463	0.25143	0.24825	0.24510
-0.7	0.24196	0.23885	0.23576	0.23270	0.22965	0.22663	0.22363	0.22065	0.21770	0.21476
-0.8	0.21186	0.20897	0.20611	0.20327	0.20045	0.19766	0.19489	0.19215	0.18943	0.18673
-0.9	0.18406	0.18141	0.17879	0.17619	0.17361	0.17106	0.16853	0.16602	0.16354	0.16109
-1.0	0.15866	0.15625	0.15386	0.15151	0.14917	0.14686	0.14457	0.14231	0.14007	0.13786
-1.1	0.13567	0.13350	0.13136	0.12924	0.12714	0.12507	0.12302	0.12100	0.11900	0.11702
-1.2	0.11507	0.11314	0.11123	0.10935	0.10749	0.10565	0.10383	0.10204	0.10027	0.09853
-1.3	0.09680	0.09510	0.09342	0.09176	0.09012	0.08851	0.08691	0.08534	0.08379	0.08226
-1.4	0.08076	0.07927	0.07780	0.07636	0.07493	0.07353	0.07215	0.07078	0.06944	0.06811
-1.5	0.06681	0.06552	0.06426	0.06301	0.06178	0.06057	0.05938	0.05821	0.05705	0.05592
-1.6	0.05480	0.05370	0.05262	0.05155	0.05050	0.04947	0.04846	0.04746	0.04648	0.04551
-1.7	0.04457	0.04363	0.04272	0.04182	0.04093	0.04006	0.03920	0.03836	0.03754	0.03673
-1.8	0.03593	0.03515	0.03438	0.03362	0.03288	0.03216	0.03144	0.03074	0.03005	0.02938
-1.9	0.02872	0.02807	0.02743	0.02680	0.02619	0.02559	0.02500	0.02442	0.02385	0.02330
-2.0	0.02275	0.02222	0.02169	0.02118	0.02068	0.02018	0.01970	0.01923	0.01876	0.01831
-2.1	0.01786	0.01743	0.01700	0.01659	0.01618	0.01578	0.01539	0.01500	0.01463	0.01426
-2.2	0.01390	0.01355	0.01321	0.01287	0.01255	0.01222	0.01191	0.01160	0.01130	0.01101
-2.3	0.01072	0.01044	0.01017	0.00990	0.00964	0.00939	0.00914	0.00889	0.00866	0.00842
-2.4	0.00820	0.00798	0.00776	0.00755	0.00734	0.00714	0.00695	0.00676	0.00657	0.00639
-2.5	0.00621	0.00604	0.00587	0.00570	0.00554	0.00539	0.00523	0.00508	0.00494	0.00480
-2.6	0.00466	0.00453	0.00440	0.00427	0.00415	0.00402	0.00391	0.00379	0.00368	0.00357
-2.7	0.00347	0.00336	0.00326	0.00317	0.00307	0.00298	0.00289	0.00280	0.00272	0.00264
-2.8	0.00256	0.00248	0.00240	0.00233	0.00226	0.00219	0.00212	0.00205	0.00199	0.00193
-2.9	0.00187	0.00181	0.00175	0.00169	0.00164	0.00159	0.00154	0.00149	0.00144	0.00139

Tabelle 5.9 zeigt die „linksseitigen“ Wahrscheinlichkeiten dafür, dass z nicht übertroffen wird. Beispielsweise entspricht einem Wert $z = 0,00$ eine Wahrscheinlichkeit von $P = 0,5$, d. h. unterhalb des Erwartungswertes (d. h. links von μ) liegt die halbe Fläche unter der Kurve; für $z = -1,53$ erhalten wir eine $P = 0,0630 = 6,3\%$, oder links von $z = -1,53$ liegt $6,3\%$ der Gesamtfläche: $P(Z \leq -1,53) = 0,0630$.

Für jeden Wert z kann man somit in Tabelle 5.9 die Wahrscheinlichkeit ablesen, die dem Ereignis zukommt, dass die zufällige Variable Z Werte kleiner oder gleich z mit $z \leq 0$ annimmt. Damit sind aber aus Gründen der Symmetrie (vgl. (5.90)) auch alle Werte für positive z -Werte ablesbar.

Einige **Beispiele** zu Tabelle 5.9:

1. $P(Z \geq 0,44) = 0,33$; d. h. $P(Z \leq 0,44) = 1 - 0,33 = 0,67 = F(0,44)$
2. Die Symmetrie der Standardnormalverteilung führt zu:

$$F(-z) = P(Z \leq -z) = P(Z \geq z) = 1 - P(Z \leq z) = 1 - F(z) \quad (5.90)$$



$$F(-1,0) = P(Z \geq 1,0) = 0,1587.$$

3. Berechnen Sie $P = P(-4,00 \leq Z \leq -0,04)$ auf 3 Stellen genau.

Wegen der Symmetrie gilt: $P = (0,04 \leq Z \leq 4,00) = P(Z \geq 0,04) - P(Z \geq 4,00) = 0,4840 - 0,0000 = 0,484$.

4. Ist $P(Z \geq 0,25) > 2P(-0,25 \leq Z \leq 0,25)$?

Ja, um $P = 0,0065$.



Beispiel: Der Nüchternblutzucker [mg/dl] sei eine normalverteilte Zufallsvariable mit dem Erwartungswert $\mu = 90$ und der Standardabweichung $\sigma = 10$. Wie groß ist die Wahrscheinlichkeit dafür, dass bei einer einer zufällig ausgewählten Person der gemessene Wert (a) unter 75 mg/dl, (b) über 100 mg/dl oder (c) zwischen 85 mg/dl und 105 mg/dl liegen wird? Die Antworten finden sich leicht anhand von Tabelle 5.9 (vgl. auch Abbildung 5.19).

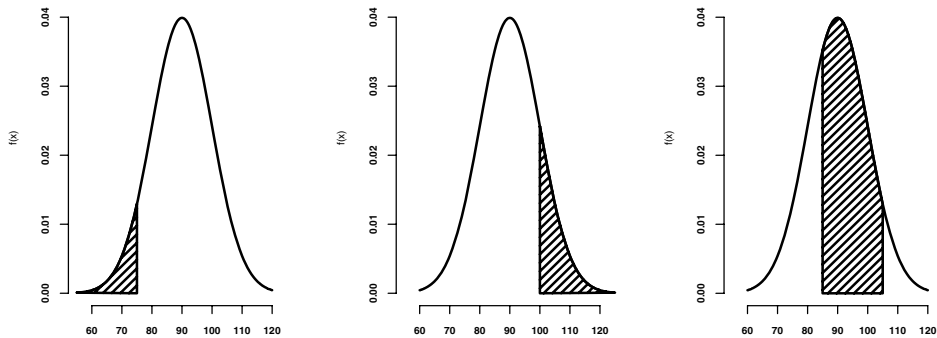


Abb. 5.19. Wahrscheinlichkeiten zum Beispiel Nüchternblutzucker

- (a) $P(X \leq 75[\text{mg/dl}]) = P(Z \leq -1,5) = 0,06681$ d.h. rund 0,07
 (b) $P(X > 100[\text{mg/dl}]) = P(Z > +1) = P(Z \leq -1) = 0,15866$
 (c) $P(85[\text{mg/dl}] \leq X \leq 105[\text{mg/dl}]) = P(Z \leq 1,5) - P(Z \leq -0,5)$
 $= 0,93319 - 0,30854 = 0,62465$

```
> pnorm(75, mean=90, sd=10)
[1] 0.0668072
> pnorm(100, mean=90, sd=10, lower.tail=FALSE)
[1] 0.1586553
> pnorm(105, mean=90, sd=10) - pnorm(85, mean=90, sd=10)
[1] 0.6246553
```

5.4.2.1 Zentrale Schwankungsintervalle

Aufschlussreich für die Beurteilung der Verteilung einer normalverteilten Zufallsvariablen sind die folgenden zentralen Schwankungsintervalle:

$\mu \pm 1,96\sigma$	oder	$z = \pm 1,96$	mit 95 % der Gesamtfläche	(5.91)
$\mu \pm 2,58\sigma$	oder	$z = \pm 2,58$	mit 99 % der Gesamtfläche	
$\mu \pm 3,29\sigma$	oder	$z = \pm 3,29$	mit 99,9 % der Gesamtfläche	

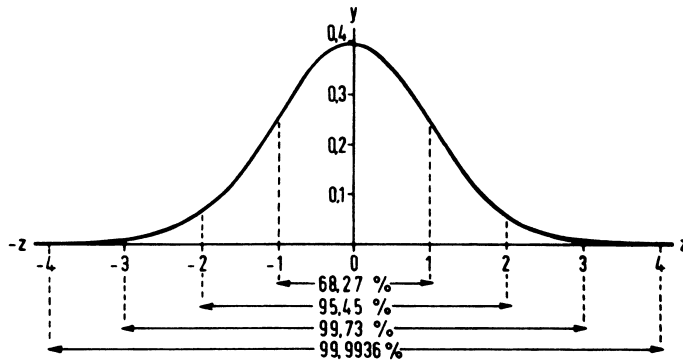


Abb. 5.20. Flächenanteile der Standardnormalverteilung - Schwankungsintervalle

Die zentralen Schwankungsintervalle der Standardnormalverteilung (Abb. 5.20) enthält Formeln (5.92). Darunterliegende (d. h. linksseitige) Flächenanteile sowie darüberliegende (d. h. rechtsseitige) Flächenanteile sind als Formeln in (5.95) angegeben. Weitere Wahrscheinlichkeiten für rechts und linksseitige Flächenanteile können natürlich auch mit Hilfe der Tabelle 5.9 bestimmt werden.

$\mu \pm 1\sigma$ oder $z = \pm 1$ mit 68,27 % der Gesamtfläche $\mu \pm 2\sigma$ oder $z = \pm 2$ mit 95,45 % der Gesamtfläche $\mu \pm 3\sigma$ oder $z = \pm 3$ mit 99,73 % der Gesamtfläche	(5.92)

Die sog. „**Drei-Sigma-Regel**“ besagt, dass beim Vorliegen einer Normalverteilung außerhalb der Grenzen $\mu \pm 3\sigma$ kaum noch gültige Beobachtungen vorliegen. Dies ist so nicht richtig; denn für kleine Stichproben ist diese „Regel“ zu konservativ und für große Stichproben ist sie zu liberal.

Eine Abweichung um mehr als σ vom Erwartungswert ist etwa einmal in je drei Versuchen zu erwarten, eine Abweichung um mehr als 2σ etwa nur einmal in je 22 Versuchen und eine Abweichung um mehr als 3σ etwa nur einmal in je 370 Versuchen, anders ausgedrückt: die Wahrscheinlichkeit, dass sich ein Wert X vom Erwartungswert absolut genommen um mehr als 3σ unterscheidet, ist mit knapp 0,3 % wesentlich kleiner als 0,01.

$P(X - \mu > 3\sigma) = 0,0027$	(5.93)

Lage der mittleren 95% der Zufallsvariablen X und Z

$P(\mu - 1,96\sigma \leq X \leq \mu + 1,96\sigma) = 0,95$ $P(-1,96 \leq Z \leq 1,96) = 0,95$	(5.94)

Die beiden folgenden Beziehungen sind gleichwertig:

1. Die Wahrscheinlichkeit, dass eine normalverteilte Zufallsvariable X mit den Parametern μ und σ zwischen $-1,96\sigma$ und $+1,96\sigma$ liegt, einschließlich dieser Grenzen, beträgt 0,95.
2. Die Wahrscheinlichkeit, dass eine standardnormalverteilte Zufallsvariable Z (mit den Parametern $\mu = 0$ und $\sigma = 1$) zwischen den z -Schranken $-1,96$ und $+1,96$ liegt, einschließlich dieser Grenzen, beträgt 0,95 .

Bevorzugt man als z -Schranken ganze Zahlen, so resultieren für die Wahrscheinlichkeiten der Standardnormalvariablen Z folgende zentrale Intervalle (5.95):

$P(-1 \leq Z \leq +1)$	=	0,68269
$P(-2 \leq Z \leq +2)$	=	0,95450
$P(-3 \leq Z \leq +3)$	=	0,99730
$P(-4 \leq Z \leq +4)$	=	0,99994

$P(Z \geq 1)$	=	0,15866	(5.95)
$P(Z \leq 1)$	=	0,84134	
$P(Z \geq 2)$	=	0,02275	
$P(Z \leq 2)$	=	0,97725	
$P(Z \geq 3)$	=	0,00135	
$P(Z \leq 3)$	=	0,99865	
$P(Z \geq 4)$	=	0,00003	
$P(Z \leq 4)$	=	0,99997	

Der Schrankenwert 1,96 bedeutet bei:

- (1) **zweiseitiger Fragestellung**, dass unterhalb von $-1,960$ und oberhalb von $1,960$ insgesamt („zweiseitige Schranken“) 5% aller Werte der Standardnormalvariablen liegen (Irrtumswahrscheinlichkeit $\alpha = 0,05$),
- (2) **einseitiger Fragestellung**, dass oberhalb von $1,960$ („einseitige obere Schranke“) genau 2,5 % aller Werte der Standardnormalvariablen liegen (Irrtumswahrscheinlichkeit $\alpha = 0,025$). Mitunter interessiert die einseitige Fragestellung in die andere Richtung: unterhalb von $-1,960$ („einseitige untere Schranke“) liegen dann ebenfalls 2,5 % aller Werte (Irrtumswahrscheinlichkeit $\alpha = 0,025$; $-1,960$ heißt 2,5. Perzentil, $1,960$ ist das 97,5. Perzentil und $1,645$ ist das 95. Perzentil).

5.4.2.2 Familie der Normalverteilungen

Ein Modell ist eine das Wesentliche eines Sachverhaltes erfassende formalisierte Darstellung. Ein statistisches Modell ist eine Beschreibung der **Wahrscheinlichkeitsverteilung** der Daten, die als beobachtete Zufallsvariablen (Zufallsveränderliche) aufgefasst werden. Meist ist man an den unbekannten Parametern dieser Wahrscheinlichkeitsverteilung und an Wahrscheinlichkeiten interessiert.

Der Name „**normal distribution**“ (Normalverteilung) wurde im Jahr 1893 durch Karl Pearson (1857–1936) eingeführt; entdeckt wurde sie 1721 durch Abraham De Moivre (1667–1754), wiederentdeckt und angewandt durch Laplace (1749–1827) und Gauß (1777–1855).

Das Modell der Normalverteilung $[N(\mu; \sigma)]$ ist:

- (a) ein idealisiertes Modell für empirische Häufigkeitsverteilungen,
- (b) bedeutungsvoll als theoretische Verteilung,
- (c) viele theoretische Verteilungen lassen sich durch eine $N(\mu; \sigma)$ gut annähern.

1. Stichprobenverteilungen in den exakten Naturwissenschaften sind bisweilen **angenähert normalverteilt** zumindest in ihrem mittleren Bereich.
2. Viele Stichprobenverteilungen lassen sich durch eine geeignete **Transformation** einer Normalverteilung annähern, beispielsweise indem man linkssteil verteilte Beobachtungswerte durch ihre Logarithmen ersetzt.
3. **Mittelwerte** und **Summen beliebiger Verteilungen** sind für hinreichend viele Beobachtungen angenähert normalverteilt.
4. Normalverteilung und Standardnormalverteilung sind mathematisch einfach zu handhaben. Tabellen der Standardnormalverteilung liegen vor.

5. Normalverteilung und Standardnormalverteilung sind Ausgangspunkt wichtiger abgeleiteter **Prüfverteilungen** (t , χ^2 , F).

6. Resultate, die für normalverteilte Zufallsvariable gelten, lassen sich in erster Annäherung auf nicht normalverteilte Zufallsvariable übertragen, vorausgesetzt der Stichprobenumfang ist hinreichend groß bzw. das Verfahren ist hinreichend robust gegenüber Abweichungen von der Normalverteilung.

5.4.2.3 Hinweise und Beispiele zur Normalverteilung

1. Mit Hilfe der Funktion **dnorm()** lässt sich eine *Normalkurve* in **R** leicht skizzieren.

```
> mue <- 80
> sig <- 8
> low <- mue - 3.5*sig; upp <- mue + 3.5*sig
> x <- seq(low, upp, by = 0.1)
> f <- dnorm(x, mean=mue, sd = sig)
> par(mfrow=c(1,1), lwd=2, font.axis=2, bty="n", ps=10)
> plot(x, f, type="l", xlim=c(low, upp), xlab=" ", ylab=" ")
```

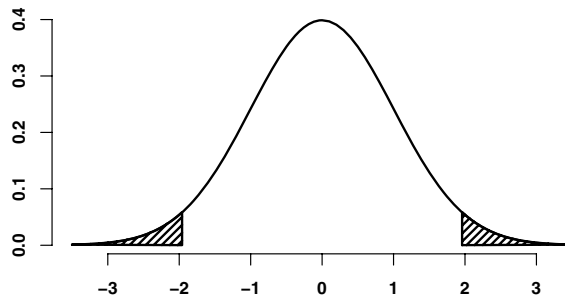


Abb. 5.21. Zeichnen der Normalkurve zur Länge eines Gegenstandes mit $\mu = 80$ cm und $\sigma = 8$ cm in **R**

2. Die Länge X eines Gegenstandes sei normalverteilt mit $\mu = 80$ cm und $\sigma = 8$ cm. Die entsprechende Normalverteilung lässt sich entsprechend Hinweis 1 leicht in **R** skizzieren (vgl. auch Hinweis 6). Zwischen welchen Längen liegen die mittleren 95% des Gegenstandes?

Formel (5.94) zeigt (mit $z = 1,96$): 95% der Gegenstände liegen im Bereich von 80 cm $\pm 1,96 \cdot 8$ cm, d. h. zwischen 64,3 cm und 95,7 cm bzw. $P(64,3 \text{ cm} < X < 95,7 \text{ cm}) \approx 0.95$. Der Wert 64,3 heißt 2,5tes Perzentil; denn unterhalb dieses Wertes liegen 2,5% aller Werte von X . Unterhalb von 95,7 liegen 97,5% aller Werte, 95,7 ist daher das 97,5te Perzentil dieser Normalverteilung. Die entsprechenden Bereiche sind in Abbildung 5.21 eingezeichnet.

```
> qnorm(0.025, mean=mue, sd=std)
[1] 64.32029
> qnorm(0.975, mean=mue, sd=std)
[1] 95.67971
```

3. Für die Normalverteilung $\mu = 100$ und $\sigma = 10$ interessiere

(a) $P(X > 115)$, (b) $P(90 < X < 115)$ und (c) $P(X < 90)$.

Für die Lösung mit Hilfe der Werte aus Tabelle 5.9 wird zunächst nach $z = (x - \mu)/\sigma$ transformiert.

Zu a) $x = 115$, $z = (115 - 100)/10 = 1,5$, d. h. $P(Z > 1,5) = P(X > 115) = 0,0668$ oder rund 7%.



Zu b) $x = 90$, $z = \frac{90-100}{10} = -1,0$; für $x = 115$ erhielten wir soeben $z = 1,5$. Gesucht ist der Anteil, besser der Flächenanteil unter der Normalkurve zwischen $z = -1,0$ und $z = 1,5$ (vgl. Abb. 5.22). Wir wissen, dass $P(Z \geq 0) = 0,5$ (vgl. Tab. 5.9), berechnen den Flächenanteil rechts von 0 als Differenz, „klappen“ den Flächenanteil links von 0 nach rechts (d. h. $z = 1$), berechnen es ebenfalls als Differenz und addieren beide Differenzen:

$$P = [0,5 - P(Z \geq 1,5)] + [0,5 - P(Z \geq 1,0)]$$

$$P = [0,5 - 0,06681] + [0,5 - 0,15866] = 0,77453.$$

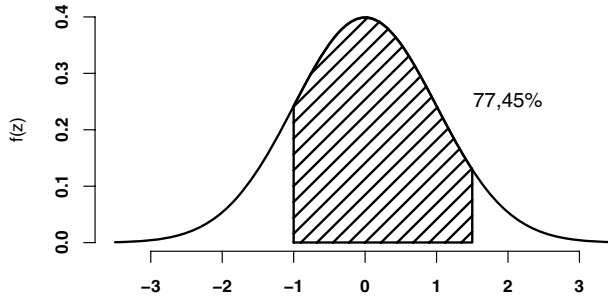


Abb. 5.22. Fläche unter der Dichte der Standardnormalverteilung im Intervall $[-1, 1.5]$

Zu c) Für $x = 90$ erhielten wir soeben $z = -1,0$. Oberhalb von $z = +1,0$ muss aus Symmetriegründen eine ebenso große Fläche liegen wie die gewünschte: 0,1587 oder 16%.

Eine Kontrolle dieser Rechnungen a, b, c ist gegeben: $0,0668 + 0,7745 + 0,1587 = 1$.

4. Für die Normalverteilung $\mu = 150$ und $\sigma = 10$ ist der Wert anzugeben, unterhalb dem 6% der Verteilung liegen; außerdem ist $P(130 < X < 160)$ zu bestimmen. Aus $(x - 150)/10 = -1,555$ folgt $x = 134,45$. Für $P(130 < X < 160)$ lässt sich schreiben $P([130 - 150]/10 < [X - 150]/10 < [160 - 150]/10) = P(-2 < Z < 1) = 1 - (0,0228 + 0,1587) = 0,8185$. Anhand von $\mu \mp 1,96\sigma = 150 \mp 1,96 \cdot 10$ erhält man die **Quantile** $x_{0,025} = 130,4$ und $x_{0,975} = 169,6$ und damit einen Bereich, in dem 95% aller Werte dieser Verteilung liegen.

```
> mue <- 150; sig <- 10
> qnorm(0.06, mean=mue, sd=sig)
[1] 134.4523
> pnorm(160, mean=mue, sd=sig) - pnorm(130, mean=mue, sd=sig)
[1] 0.8185946
> qnorm(0.025, mean=mue, sd=sig)
[1] 130.4004
> qnorm(0.975, mean=mue, sd=sig)
[1] 169.5996
```

5. Wahrscheinlichkeiten für normalverteilte Zufallsvariablen: Für eine normalverteilte Grundgesamtheit mit $\mu = 12,00$ und $\sigma = 2,00$ interessiere der Anteil der Werte oberhalb von $x = 15,11$. Mit $z = (15,11 - 12,00)/2,00 = 3,11/2,00 = 1,555$ sind $P = 0,06$ oder 6% aller Werte der Grundgesamtheit größer als 15,11.

Aus Symmetriegründen (Skizze!) sind dann auch:

- (1) 6% aller Werte kleiner als 8,89; denn
 $z = (8,89 - 12,00)/2,00 = -3,11/2,00 = -1,555$;
- (2) $100\% - 6\% = 94\%$ aller Werte nicht größer als 15,11;
- (3) $100\% - 2 \cdot 6\% = 88\%$ aller Werte im Bereich zwischen 8,89 und 15,11.

Für die normalverteilte Zufallsvariable X und für die entsprechende Standardnormalvariable Z lassen sich diese Resultate so formulieren:

- (1) $P(X > 15,11) = P(Z > 1,555) = 0,06$ (vgl. auch Tabelle 5.9);
 - (2) $P(X < 8,89) = P(Z < -1,555) = 0,06$;
 - (3) $P(X < 15,11) = 1 - 0,06 = 0,94$;
 - (4) $P(8,89 < X < 15,11) = 0,88$;
- da eine stetige Zufallsvariable vorliegt [$P(X = x) = 0$] gilt natürlich auch:
- (5) $P(8,89 \leq X \leq 15,11) = 0,88$.

Für nicht zu kleine Zufallsstichproben aus angenähert normalverteilten Grundgesamtheiten erhält man anhand von $\hat{z} = (x_f - \bar{x})/s$ eine **Schätzung für den Anteil P der Werte**, der größer (bzw. kleiner) ist als x_f (wobei x_f ein bestimmter fester Wert x ist). Für (1) würden wir dann z. B. $P(X > 15,11) \approx 0,06$ schreiben.

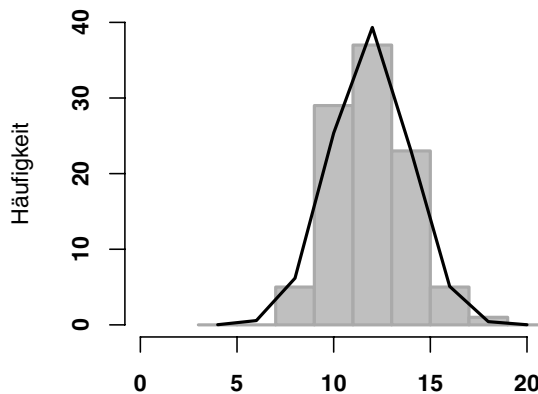


Abb. 5.23. Angepasste Normalverteilung auf der Grundlage von Histogrammdaten

6. Das Zeichnen einer angepassten Normalverteilung: Es liege eine Stichprobe des Umfangs n mit den Statistiken $\bar{x}$ und s vor. Über das zugehörige Histogramm mit der Klassenbreite b und der Fläche $b \cdot n$ lässt sich dann eine Normalkurve zeichnen, deren Ordinate $\hat{y} = (b n / s) f(z)$ beträgt; $f(z)$ kann in **R** mit der Funktion **dnorm()** (mit $\mu = 0$ und $\sigma = 1$) für $z = (x - \bar{x})/s$ (x sind jeweils die Klassenmitten) berechnet werden. Die Güte der Anpassung ist *abzuschätzen* (vgl. Abbildung 5.23).

```
> mue <- 12; sig <- 2; n <- 100;
> y.val <- rnorm(n, mean=mue, sd=sig)
> brk <- c(3, 5, 7, 9, 11, 13, 15, 17, 19, 21)
> hist(y.val, breaks = brk, ylim=c(0,40), xlim=c(0,20), main=" ",
+      border="darkgrey", xlab=" ", ylab="Häufigkeit", col="grey")
> mid <- c(4, 6, 8, 10, 12, 14, 16, 18, 20)
> z.val <- (mid - mean(y.val)) / sd(y.val)
> f.val <- dnorm(z.val, mean=0, sd=1)
> y.est <- (2 * n / sd(y.val)) * f.val
> lines(mid, y.est)
```

7. Typisches für eine Normalverteilung.

Typisch für eine Normalverteilung ist, dass

- (a) $\bar{X}$ und S^2 voneinander stochastisch unabhängig sind,
- (b) $\bar{X}$ exakt normalverteilt ist,
- (c) $\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} \right)^2$ nach χ^2_ν mit $\nu = n$ Freiheitsgraden verteilt ist.

Daneben gilt: (a) $\tilde{\mu} = \mu$ und (b) $\tilde{X}$ ist angenähert normalverteilt.

Sind X_1 und X_2 unabhängige normalverteilte Zufallsvariablen aus $N(\mu_1, \sigma_1)$ und $N(\mu_2, \sigma_2)$, dann ist:

$$\boxed{X = X_1 + X_2 \quad \text{normalverteilt mit} \quad \begin{array}{l} \mu = \mu_1 + \mu_2 \\ \sigma = \sqrt{\sigma_1^2 + \sigma_2^2} \end{array}} \quad (5.96)$$

Sind Z_1 und Z_2 unabhängige standardnormalverteilte Zufallsvariablen, dann ist

$$\boxed{Z = (Z_1 + Z_2)/\sqrt{2} \quad \text{standardnormalverteilt}} \quad (5.97)$$

Verallgemeinerung für wechselseitig unabhängige Standardnormalvariablen $Z_1, \dots, Z_n$:

$$\boxed{Z = (Z_1 + \dots + Z_n)/\sqrt{n} \quad \text{ist nach} \quad N(0, 1) \quad \text{verteilt}} \quad (5.98)$$

8. Verteilung einiger Schätzfunktionen aus normalverteilten Grundgesamtheiten.

Übersicht 12. Funktionen normalverteilter Zufallsvariablen X_i

Nr.	Schätzfunktion	Verteilung	Kommentar
(1)	X_i	$N(\mu; \sigma)$	Einzelwerte
(2)	$X_1 + X_2 + \dots + X_n$	$N(n\mu; \sqrt{n}\sigma)$	Summe
(3)	$(X_1 + X_2 + \dots + X_n)/n$	$N(\mu; \sigma/\sqrt{n})$	Mittelwert
(4)	$\frac{X_i - \mu}{\sigma}$	$N(0; 1)$	Transformation
(5)	$\frac{\bar{X}_i - \mu}{\sigma} \sqrt{n}$	$N(0; 1)$	Einstichproben-Gauß-Test (s.d.)

9. Für die Standardnormalvariable Z gilt für

(1) den Erwartungswert:

$$\boxed{E(Z) = E\left(\frac{X - \mu}{\sigma}\right) = \frac{1}{\sigma}E(X) - \frac{1}{\sigma}E(\mu) = \frac{\mu}{\sigma} - \frac{\mu}{\sigma} = 0} \quad (5.99)$$

(2) die **Varianz**:

$$\begin{aligned}
\text{Var}(Z) &= \text{Var}\left(\frac{X - \mu}{\sigma}\right) = \text{Var}\left(\frac{1}{\sigma}X - \frac{\mu}{\sigma}\right) = \frac{1}{\sigma^2} \text{Var}(X) = \frac{\sigma^2}{\sigma^2} = 1 \\
\text{Var}(Z) &= E[Z - E(Z)]^2 = E\left[\frac{X - \mu}{\sigma} - 0\right]^2 = \frac{1}{\sigma^2} E(X - \mu)^2 \\
&= \frac{1}{\sigma^2} \text{Var}(X) = \frac{\sigma^2}{\sigma^2} = 1 \\
\text{Var}(Z) &= E(Z^2) - \mu_z^2 = E(Z^2) - 0 = E\left[\left(\frac{X - \mu}{\sigma}\right)^2\right] \\
&= \frac{1}{\sigma^2} E[(X - \mu)^2] = \frac{\sigma^2}{\sigma^2} = 1
\end{aligned} \tag{5.100}$$

5.4.2.4 Ungleichungen von Bienaymé (1853) und Tschebyscheff (1874)

Die Wahrscheinlichkeit dafür, dass die absolute Differenz zwischen der Variablen und ihrem Erwartungswert größer oder gleich $k\sigma$ ist, ist kleiner oder gleich $1/k^2$:

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2} \quad \text{mit} \quad k > 0 \tag{5.101}$$

Für eine Standardnormalverteilung gilt:

$$P(|Z| \geq 2) = 1 - P(-2 < Z < 2) = 0,0455 \approx 1/20 \text{ und nach (5.101):}$$

$$P(|Z| \geq 2) \leq 1/4.$$

Bei kleinem σ sind größere Abweichungen von μ wenig wahrscheinlich.

Für (5.101) kann man auch das Komplement schreiben:

$$1 - P(|X - \mu| \geq k\sigma) \geq 1 - (1/k^2) \text{ bzw. } P(|X - \mu| < k\sigma) \geq 1 - (1/k^2).$$

oder:

$$\begin{aligned}
P(\mu - k\sigma < X < \mu + k\sigma) &\geq 1 - (1/k^2) \\
P(|X - \mu| < k\sigma) &\geq 1 - (1/k^2)
\end{aligned}$$

$$\text{z.B. } P(|X - \mu| < 2\sigma) \geq 1 - (1/4) = 3/4,$$

$$\text{und entsprechend } P(|X - \mu| < 3\sigma) \geq 8/9 = 0,8889.$$

Für symmetrische eingipflige Verteilungen gilt nach Gauß (1821) die schärfere Ungleichung:

$$P(|X - \mu| \geq k\sigma) \leq \frac{4}{9k^2} \quad k \gtrsim (2/\sqrt{3}) = 1,155 \tag{5.102}$$

oder:

$$P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - 4/(9k^2)$$

$$P(|X - \mu| < k\sigma) \geq 1 - 4/(9k^2)$$

$$\text{z.B. } P(|X - \mu| < 2\sigma) \geq 1 - (4/36) = 8/9$$

und entsprechend:

$$P(|X - \mu| < 3\sigma) \geq 0,9506$$

Für die Normalverteilung gilt:

$$P(|X - \mu| < 3\sigma) = 0,9973$$

Übersicht 13. Wahrscheinlichkeiten zentraler Anteile einer Verteilung: $k\sigma$ -Bereiche

Bereich	Verteilungstyp	Verteilungsanteil
$\mu \pm 1,96\sigma$	beliebig symmetrisch-eingipflig Normalverteilung	mindestens 74,0% mindestens 88,4% exakt 95,0%
$\mu \pm 3\sigma$	beliebig symmetrisch-eingipflig Normalverteilung	mindestens 88,9% mindestens 95,1% exakt 99,7%

5.4.2.5 Zentraler Grenzwertsatz

Der zentrale Grenzwertsatz (**central limit theorem**) enthält sowohl in theoretischer Hinsicht als auch im Hinblick auf praktische Anwendungen eine der wichtigsten Aussagen der Statistik. Er beschreibt eine sehr markante Eigenschaft der Normalverteilung, die sehr häufig die Rechtfertigung dafür liefert, Zufallserscheinungen, die sich aus dem **Zusammenwirken zahlreicher zufälliger Einzeleffekte** ergeben, eben durch das Modell der Normalverteilung darstellen und bewerten zu können. Der zentrale Grenzwertsatz unterstreicht damit die herausragende Bedeutung der Normalverteilung in der Wahrscheinlichkeitstheorie und der Statistik.

Von praktischem Interesse ist insbesondere nach J.W. Lindeberg und P. Levy (1922) der Spezialfall (5.103), in dem eine „große“ Zahl identisch verteilter Zufallsvariablen X_i betrachtet wird, die insbesondere alle denselben Erwartungswert $E[X_i] = \mu$ und dieselbe Varianz $Var[X_i] = \sigma^2$ haben.

$$P\left(\frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} < z\right) \rightarrow \Phi(z) \quad \text{für } n \rightarrow \infty \quad (5.103)$$

Hierbei ist $\Phi(z)$ die Verteilungsfunktion der $N(0; 1)$ -Verteilung. In kurzer Form beinhaltet (5.103) die Aussage, dass Summen (beliebiger) identisch verteilter Zufallsvariablen **angenähert als normalverteilt** betrachtet werden können, wenn ihre Anzahl n nur „genügend groß“ ist. Die Präzisierung, was hier als genügend groß angesehen werden kann, erfolgt durch entsprechende Konvergenzbetrachtungen von Fall zu Fall unterschiedlich.

Folgen die Zufallsvariablen X_i einer Bernoulli-Verteilung, d.h. $P(X_i = 1) = p$ und $P(X_i = 0) = q$ mit $p + q = 1$, dann beschreibt (5.103) den **klassischen zentralen Grenzwertsatz von de Moivre (1730) und Laplace (1812)**, der eine Rechtfertigung für zahlreiche Vereinfachungen bei der Behandlung binomialverteilter Zufallsvariablen liefert (5.104).

$$P\left(\frac{\sum_{i=1}^n X_i - np}{\sqrt{npq}} < z\right) \rightarrow \Phi(z) \quad \text{für } n \rightarrow \infty \quad (5.104)$$

5.4.3 Lognormalverteilung

Viele Verteilungen in der Natur laufen als positiv schiefe, **linkssteile Verteilungen rechts flach** aus (tailed to the right). Eine anschauliche Erklärung dafür, dass sich ein Merkmal nicht symmetrisch-normal verteilt, ist oft dadurch gegeben, dass das Merkmal einen bestimmten Schrankenwert nicht unter bzw. überschreiten kann und somit nach dieser Seite hin in seiner Variationsmöglichkeit gehemmt ist. Markantes Beispiel ist die Verteilung von Zeiten (untere Grenze: Null). Besonders dann, wenn die Verteilung links durch den Wert Null begrenzt ist, kommt man **durch Logarithmieren zu annähernd normalverteilten Werten**. Durch das Logarithmieren wird der Bereich zwischen 0 und 1 in den Bereich $-\infty$ bis 0 überführt, der linke Teil der

Verteilung stark gestreckt und der rechte stark gestaucht. Das gilt besonders dann, wenn die Standardabweichung groß ist im Vergleich zum Mittelwert, wenn der Variabilitätskoeffizient größer als 33% ist.

Die Entstehung einer logarithmischen Normalverteilung, kurz Lognormalverteilung genannt, kann darauf zurückgeführt werden, dass viele Zufallsgrößen **multiplikativ zusammenwirken**, die Wirkung einer Zufallsänderung also jeweils der zuvor bestehenden Größe proportional ist. Dagegen kommt die Normalverteilung durch additives Zusammenwirken vieler Zufallsgrößen zustande. Es ist somit verständlich, dass die Lognormalverteilung insbesondere bei Merkmalen aus Biologie und Wirtschaft vorherrscht. Beispielsweise die Empfindlichkeit von Tieren einer Art – Bakterien bis Großsäuger – gegenüber Pharmaka.

Merkmale beim Menschen: Körperlänge (Kinder), Herzgröße, Brustumfang, Pulsfrequenz, systolischer und diastolischer Blutdruck, Senkungsgeschwindigkeit der roten Blutkörperchen, prozentuale Anteile der einzelnen Arten weißer Blutkörperchen sowie der Gehalt vieler Serumbestandteile.

Wirtschaftsstatistische Merkmale: Bruttonomatsverdienst von Angestellten, Umsätze von Unternehmen, Anbauflächen verschiedener Fruchtarten in den Gemeinden.

Näherungsweise folgen der Lognormalverteilung oft auch solche Merkmale, die nur ganzzahlige Werte annehmen können, so z. B. die Zahl der Zuchtsauen auf den Zählflächen und die Zahl der Obstbäume in den Gemeinden.

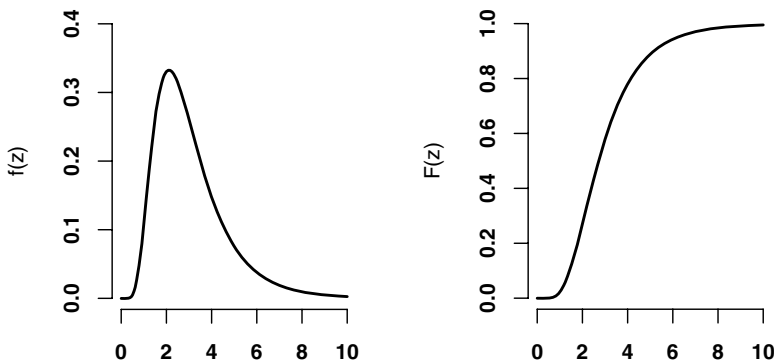


Abb. 5.24. Wahrscheinlichkeitsdichte und Verteilungsfunktion einer Lognormalverteilung mit $\mu = 1$, $\sigma = 0,5$ und $\tilde{\mu} = e^\mu \approx 2,72$

Williams [Wil40] untersuchte 600 Sätze aus G.B. Shaw's „An Intelligent Woman's Guide to Socialism“, jeweils die ersten 15 Sätze in den Abschnitten 1 bis 40, und erhielt

$$y = \frac{1}{0,29 \cdot \sqrt{2\pi}} e^{-\frac{(x-1,4)^2}{2 \cdot 0,29^2}}$$

(y = Häufigkeit und x = Logarithmus der Zahl der Wörter pro Satz) eine „lognormalverteilte“ Wahrscheinlichkeitsdichte. Überhaupt ist die Zahl der Buchstaben (und Phoneme) pro Wort der englischen Umgangssprache bemerkenswert gut lognormalverteilt. Lognormalverteilungen treten weiter, wie gesagt, bei **Zeitstudien und Lebensdaueranalysen** auf sowie in der **analytischen Chemie**: Bei Bestimmungen in einem sehr weiten Konzentrationsbereich (über mehrere Zehnerpotenzen), beim Arbeiten in der Nähe von null oder hundert Prozent (z. B. Reinheitsprüfungen) und wenn der Zufallsfehler eines Verfahrens mit den Messwerten selbst vergleichbar ist, z. B. bei der semiquantitativen Spektralanalyse.

Eine stetige Zufallsvariable $X (> 0)$ heißt **logarithmisch normalverteilt** (lognormal-verteilt), wenn $\ln x$ normalverteilt ist. Sie hat daher die Wahrscheinlichkeitsdichte:

$$y = f(x) = \begin{cases} \frac{1}{\sigma x \sqrt{2\pi}} \cdot e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} & \text{für } x > 0 \\ 0 & \text{für } x \leq 0 \end{cases} \quad (5.105)$$

Die Maßzahlen zur Beschreibung einer Lognormalverteilung sind in (5.106) zusammengefasst:

Erwartungswert:	$E[X] = e^{\mu + \frac{\sigma^2}{2}}$	(5.106)
Varianz:	$Var[X] = e^{2\mu + \sigma^2} (e^{\sigma^2} - 1)$	
2. Moment:	$e^{2(\mu + \sigma^2)}$	
Dichtemittel:	$e^{\mu - \sigma^2}$	
Median:	e^{μ}	

5.4.3.1 Berechnung der empirischen Maßzahlen zur Lognormalverteilung

Charakteristisch für eine Lognormalverteilung ist ihre Zentrale 68%-Masse, geschrieben

$$(\text{Median})(\text{Streufaktor})^{\pm 1} \quad (5.107)$$

den einen um die Extremwerte verminderten Bereich „noch typischer Werte“ enthält. Der Streufaktor wird in Formel (5.108) näher erläutert.

Für die rechnerische Ermittlung der Kennzahlen werden zu den in üblicher Weise mit konstanter Klassenbreite klassifizierten Daten die Logarithmen der Klassenmitten aufgesucht ($\lg x_j$), die Produkte $f_j \cdot \lg x_j$ und $f_j (\lg x_j)^2$ gebildet (f_j = Häufigkeiten pro Klasse), aufsummiert und in die folgenden Formeln eingesetzt.

Median_L = antilg $\bar{x}_{\lg x_i}$ = antilg $\left(\sum f_i \cdot \lg x_j / n \right)$	(5.108)
Streufaktor = antilg $\sqrt{s_{\lg x_j}^2}$ = antilg $\sqrt{\frac{\sum f_j (\lg x_j)^2 - (\sum f_j \lg x_j)^2 / n}{n - 1}}$	
Mittelwert_L = antilg $(\bar{x}_{\lg x_j} + 1,1513 s_{\lg x_j}^2)$	
Dichtemittel_L = antilg $(\bar{x}_{\lg x_j} - 2,3026 s_{\lg x_j}^2)$	

Bei kleinen Stichprobenumfängen werden statt der Logarithmen der Klassenmitten die Logarithmen der Einzelwerte verwendet; die Häufigkeit jeder Klasse (f_j) ist dann gleich Eins. Der Streufaktor ist eine Schätzung von antilg $s_{\lg x_j}$. Mit zunehmendem Streufaktor verschieben sich also das arithmetische Mittel vom Median nach rechts und das Dichtemittel um den doppelten Betrag nach links.

Beispiel: Die folgende Tabelle enthält 20 nach der Größe geordnete Messwerte x_i , die angenähert lognormalverteilt sind. Schätzen Sie die Kennwerte.



x_i	$\lg x_i$	$(\lg x_i)^2$
3	0.4771	0.2276
4	0.6021	0.3625
5	0.6990	0.4886
5	0.6990	0.4886
5	0.6990	0.4886
5	0.6990	0.4886
5	0.6990	0.4886
6	0.7782	0.6055
7	0.8451	0.7142
7	0.8451	0.7142
Fortsetzung rechts		

x_i	$\lg x_i$	$(\lg x_i)^2$
7	0.8451	0.7142
7	0.8451	0.7142
8	0.9031	0.8156
8	0.9031	0.8156
9	0.9542	0.9106
9	0.9542	0.9106
10	1.0000	1.0000
11	1.0414	1.0845
12	1.0792	1.1646
14	1.1461	1.3136
Σ	16.7141	14.5104

Der Variationskoeffizient der Originaldaten (x_i) liegt mit $V = \frac{2,83}{7,35} = 38,5\%$ deutlich oberhalb der 33%-Schranke. Die Kennwerte:

$$\text{Median}_L = \text{antilg} \left\{ \frac{16,7141}{20} \right\} = \text{antilg } 0,8357 = 6,850$$

$$\text{Streufaktor} = \text{antilg} \sqrt{\frac{14,5104 - 16,7141^2/20}{20 - 1}} = \text{antilg} \sqrt{0,02854}$$

$$\text{Streufaktor} = \text{antilg } 0,1690 = 1,476.$$

Die Zentrale 68%-Masse liegt zwischen $6,850/1,476 = 4,641$ und $6,850 \cdot 1,476 = 10,111$ (bzw. $6,850 \cdot 1,476^{\pm 1}$). Außerhalb dieses Bereiches liegen 5 Werte, zu erwarten wären $0,32 \cdot 20 = 6$ Werte.

$$\text{Mittelwert}_L = \text{antilg} (0,8357 + 1,1513 \cdot 0,02854) = \text{antilg } 0,8686 = 7,389$$

$$\text{Dichtemittel}_L = \text{antilg} (0,8357 - 2,3026 \cdot 0,02854)$$

$$\text{Dichtemittel}_L = \text{antilg } 0,7700 = 5,888.$$

```
> x <- c(3,4,5,5,5,5,5,6,7,7,7,7,8,8,9,9,10,11,12,14)
> lgx <- log10(x)
> lgx2 <- lgx^2
> median.L <- 10^mean(lgx); median.L
[1] 6.850103
> streufaktor <- 10^(sqrt(sd(lgx)^2)); streufaktor
[1] 1.475594
> mittelwert.L <- 10^(mean(lgx)+1.1513*sd(lgx)^2); mittelwert.L
[1] 7.388674
> dichtemittel.L <- 10^(mean(lgx)-2.3026*sd(lgx)^2); dichtemittel.L
[1] 5.88787
```

5.4.4 Exponentialverteilung

Eine stetige Zufallsvariable X heißt exponentialverteilt mit dem Parameter λ ($\lambda > 0$), wenn sie die Dichtefunktion in (5.109) besitzt.

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{für } x \geq 0 \\ 0 & \text{für } x < 0 \end{cases} \quad (5.109)$$

Typische Beispiele für exponentialverteilte Zufallsvariablen sind:

- Wartezeiten, Dauer von Telefongesprächen
- Zeitmessungen bei Zerfallsprozessen
- Arbeitszeit von Maschinen zwischen zwei Wartungen
- Lebensdauer/Funktionsdauer von Lebewesen/Bauteilen

Die Beispiele weisen darauf hin, dass die Zufallsvariable in der Regel durch Zeitmessungen (Zeitdauer) bestimmt sind (häufig wird t anstelle von x geschrieben).

Für die Verteilungsfunktion einer exponentialverteilten Zufallsvariablen folgt aus (5.109) entsprechend

$$F(x) = P(X \leq x) = \begin{cases} 1 - e^{-\lambda x} & \text{für } x \geq 0 \\ 0 & \text{für } x < 0 \end{cases} \quad (5.110)$$

In Abbildung 5.25 sind Wahrscheinlichkeitsdichte und Verteilungsfunktion der Exponentialverteilung für verschiedene Parameter ($\lambda = 1, 5, 10$) dargestellt. Der Parameter λ kennzeichnet eine (konstante) Ausfallrate (mittlere Zahl an Ausfällen pro Zeiteinheit). Die für die Verteilung bestimmende **charakteristische „Lebensdauer“** - die altersunabhängige Restlebensdauer - kann durch $T = 1/\lambda$ bestimmt werden (engl. **mean time between failures**). Bis zu dieser Zeit sind 63,2% der „Einheiten“ ausgefallen (d.h. $x_{0,632} = 1/\lambda$; $x_{0,5} = \ln 2/\lambda = 0,6931/\lambda$ - siehe auch exponentielles Wachstum auf Seite 77). Zu betonen ist, dass der Erwartungswert $\mu = 1/\lambda$ nicht die bereits vergangene Zeit berücksichtigt („**memoryless**“). Die zu erwartende Restlebensdauer eines Bauteils ist damit genauso groß wie bei seiner Inbetriebnahme, was selten der Fall sein wird!

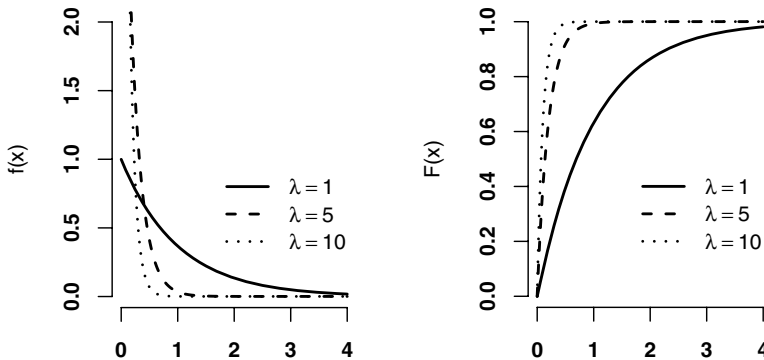


Abb. 5.25. Wahrscheinlichkeitsdichte und Verteilungsfunktion der Exponentialverteilung mit $\lambda = 1, 5, 10$

Erwartungswert und Varianz der Exponentialverteilung sind durch (5.111) gegeben (d.h. $\mu = \sigma$).

Erwartungswert: $\mu = \frac{1}{\lambda}$ Varianz: $\sigma^2 = \frac{1}{\lambda^2}$ 2. Moment: $\frac{2}{\lambda^2}$	(5.111)



Beispiel (Wartezeiten): An einer Kasse kommt durchschnittlich alle 2 Minuten ein Kunde an. Wie groß ist die Wahrscheinlichkeit dafür, dass der Abstand zwischen zwei Kunden größer als 4 Minuten ist, wenn der zeitliche Abstand zwischen der Ankunft zweier Kunden exponentialverteilt ist.

$$\mu = 2 = \frac{1}{\lambda} \rightarrow \lambda = 0,5$$

$$P(T > 4) = 1 - P(T \leq 4) = 1 - (1 - e^{-0.5 \cdot 4}) = 0,135$$

```
> 1- pexp(4, rate = 0.5)
[1] 0.1353353
```

Beispiel (Lebensdauer von Glühbirnen): Die mittlere Lebensdauer einer speziellen Sorte von Glühbirnen wird mit 100 Stunden angegeben. Wie groß ist die Wahrscheinlichkeit dafür, dass eine zufällig ausgewählte Glühbirne länger als 110 Stunden brennt.



$$P(T > 110) = 1 - P(T \leq 110) = 1 - (1 - e^{-110 \cdot 0.01}) = 0.333$$

5.4.5 Weibull-Verteilung

Eine stetige Zufallsvariable X heißt Weibull-verteilt mit den Parametern β (**Formparameter**, engl. shape) und α (**Skalenparameter**, engl. scale), wenn ihre Wahrscheinlichkeitsdichte $f(x)$ und entsprechend die Verteilungsfunktion $F(x)$ durch (5.112) gegeben sind.

$$\begin{aligned} f(x) &= \begin{cases} \left(\frac{\beta}{\alpha}\right) \left(\frac{x}{\alpha}\right)^{\beta-1} e^{-\left(\frac{x}{\alpha}\right)^\beta} & \text{für } x \geq 0 \\ 0 & \text{für } x < 0 \end{cases} \\ F(x) &= \begin{cases} 1 - e^{-\left(\frac{x}{\alpha}\right)^\beta} & \text{für } x \geq 0 \\ 0 & \text{für } x < 0 \end{cases} \end{aligned} \quad (5.112)$$

Während bei der Exponentialverteilung eine konstante Ausfallrate angenommen wird, kann bei der Weibull-Verteilung durch den zusätzlichen Formparameter β die Form der Wahrscheinlichkeitsdichte verändert werden, insbesondere können für $\beta < 1$ Frühausfälle und durch $\beta > 1$ sogenannte Verschleissausfälle besonders gewichtet werden. Für $\beta = 1$ erhält man als Spezialfall der Weibull-Verteilung die Exponentialverteilung mit dem Parameter $\lambda = 1/\alpha$. Verschiedene Wahrscheinlichkeitsdichten sind in Abbildung 5.26 dargestellt.

Erwartungswert und Varianz der Weibull-Verteilung sind durch (5.113) gegeben.

$$\begin{aligned} \text{Erwartungswert: } \mu &= \frac{\Gamma\left(\frac{1}{\beta} + 1\right)}{\frac{1}{\alpha}} \\ \text{Varianz: } \sigma^2 &= \frac{\Gamma\left(\frac{2}{\beta} + 1\right) - \left[\Gamma\left(\frac{1}{\beta} + 1\right)\right]^2}{\left(\frac{1}{\alpha}\right)^{\frac{2}{\beta}}} \\ \text{2. Moment: } &\alpha^2 \Gamma(2/\beta + 1) \end{aligned} \quad (5.113)$$

Darin bezeichnet Γ die Eulersche Gammafunktion, die in (5.118) und (5.119) näher beschrieben wird.

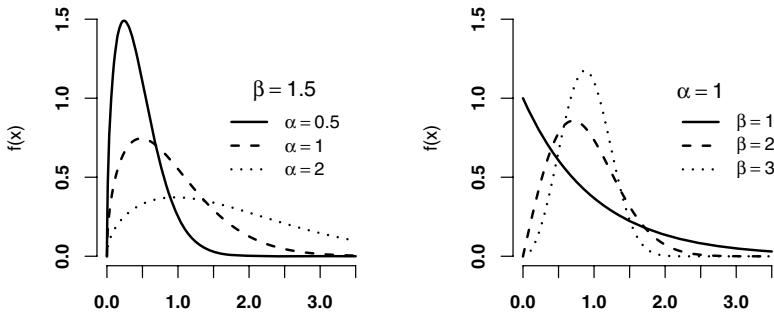


Abb. 5.26. Wahrscheinlichkeitsdichte der Weibull-Verteilung für unterschiedliche Skalen- und Formparameter



Beispiel: Die Bruchfestigkeit keramischer Werkstoffe, wie sie in der zahnärztlichen prothetischen Versorgung verwendet werden, kann mit dem Modell der Weibull-Verteilung beschrieben werden. Der Skalenparameter α kennzeichnet die charakteristische Festigkeit des Materials, β ist das so genannte Weibull-Modul. Wahrscheinlichkeitsdichte und Verteilungsfunktion für die Bruchlast von Zirkondioxidstiften mit einem Durchmesser von 1,4mm ($\alpha = 27$, $\beta = 7$) sind in Abbildung 5.27 dargestellt. Die Wahrscheinlichkeit für einen Bruch im Bereich von 30N (Newton) bis 35N beträgt $P=0.1215$ (schraffierte Fläche).

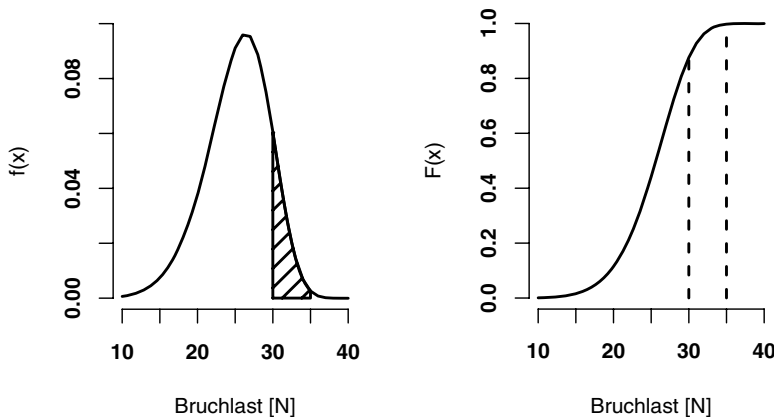


Abb. 5.27. Wahrscheinlichkeitsdichte und der Verteilungsfunktion zur Bruchlast[N] von Zirkondioxidstiften mit den Parametern $\alpha = 27$ und $\beta = 7$

5.5 Testverteilungen

- Student-Verteilung
- Chiquadrat-Verteilung
- Fisher-Verteilung
- Verteilung von Stichprobenfunktionen

Dieser Abschnitt behandelt die Verteilung von **Prüfgrößen**. Prüfgrößen sind Vorschriften, nach denen aus einer vorliegenden Stichprobe eine Zahl, der Wert der Prüfgröße für diese Stichprobe,

errechnet wird. So können der Stichprobenmittelwert, die Stichprobenvarianz oder das Verhältnis der Varianzen zweier Stichproben, alles dies sind Schätzwerte oder Funktionswerte von **Stichprobenfunktionen**, als Prüfgrößen aufgefasst werden. Die Prüfgröße ist eine zufällige Variable. Ihre Wahrscheinlichkeitsverteilungen bilden die Grundlage für die auf diesen Prüfgrößen basierenden Tests. **Prüfverteilungen sind Stichprobenfunktionen normalverteilter zufälliger Variablen.** Statt Prüfgröße sagt man häufig auch Teststatistik (test statistic).

5.5.1 Student-Verteilung (t)

W.S. Gosset (1876–1937) wies im Jahre 1908 unter dem Pseudonym „Student“ nach, dass die Verteilung des Quotienten aus der Abweichung eines Stichprobenmittelwertes vom Parameter der Grundgesamtheit und dem Standardfehler des Mittelwertes der Grundgesamtheit (5.114) nur dann der Standardnormalverteilung folgt, wenn die X_i normalverteilt sind und beide Parameter (μ , σ) bekannt sind. Die Maßzahl für die Abweichungen bei unbekannter Varianz (5.115) folgt dagegen der „Student“- t -Verteilung oder kurz t -Verteilung. Vorausgesetzt wird hierbei, dass die Einzelbeobachtungen X_i **unabhängig** und **normalverteilt** sind.

$$\frac{\text{Abweichung des Mittelwertes}}{\text{Standardfehler des Mittelwertes}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = Z \quad (5.114)$$

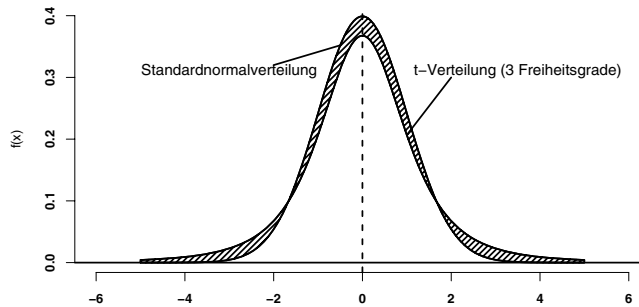


Abb. 5.28. Wahrscheinlichkeitsdichte der $N(0; 1)$ -Verteilung und der „Student“-Verteilung mit 3 Freiheitsgraden. Mit abnehmender Anzahl der Freiheitsgrade sinkt das Maximum der „Student“-Verteilung, die schraffierte Fläche nimmt zu. Im Gegensatz zur $N(0; 1)$ -Verteilung ist mehr Wahrscheinlichkeit in den Ausläufen und weniger im zentralen Teil konzentriert

(5.114) strebt mit zunehmendem n mehr oder weniger schnell gegen eine Normalverteilung, je nach dem Typ der Grundgesamtheit, aus der die Stichproben stammen; (5.115) ist dagegen (a) für kleines n und Grundgesamtheiten, die sich nicht stark von der Normalverteilung unterscheiden, approximativ wie t verteilt, (b) für großes n und fast alle Grundgesamtheiten angenähert standard-normalverteilt.

$$t = \frac{\bar{X} - \mu}{S/\sqrt{n}} \quad (5.115)$$

mit $S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}$

Die t -Verteilung (vgl. Abb. 5.28) ist der Standardnormalverteilung [$N(0; 1)$ -Verteilung] sehr ähnlich. Wie diese ist sie stetig, symmetrisch, glockenförmig, mit einem Variationsbereich von minus Unendlich bis plus Unendlich. Sie ist jedoch von μ und σ unabhängig.

Die **Form** der t -Verteilung wird nur von dem sogenannten **Freiheitsgrad** (FG) bestimmt, hier in (5.115) als $FG = n - 1 = \nu$. Der Parameter (gr. nü) charakterisiert somit die Familie der t -Verteilungen ($\nu = 1, 2, \dots$). Für $\nu \geq 2$ ist der Mittelwert der t -Verteilungen Null; für $\nu \geq 3$ ist ihre Varianz gleich $\nu/(\nu - 2)$, die für großes ν gleich Eins wird.

Freiheitsgrad: Die Anzahl der Freiheitsgrade FG oder ν (gr. nü) einer Zufallsgröße ist definiert durch die Zahl „frei“ verfügbarer Beobachtungen, dem Stichprobenumfang n minus der Anzahl a aus der Stichprobe geschätzter Parameter

$$FG = \nu = n - a \quad (5.116)$$

Anweisungen, wie der Freiheitsgrad für Spezialfälle dieser Zufallsgröße (und anderer Prüfgrößen) zu bestimmen ist, werden später von Fall zu Fall gegeben.

Je kleiner der Freiheitsgrad ist, um so stärker ist die Abweichung von der $N(0; 1)$ -Verteilung, um so flacher verlaufen die Kurven, d. h. im Gegensatz zur $N(0; 1)$ -Verteilung hat sie mehr Wahrscheinlichkeit in den Ausläufen und weniger im zentralen Teil konzentriert (vgl. Abb. 5.28). Bei großem Freiheitsgrad geht die t -Verteilung in die $N(0; 1)$ -Verteilung über.

Die **Student-Verteilung** hat im Verhältnis zur $N(0; 1)$ -Verteilung für kleine Freiheitsgrade bei geringer Höhe eine wesentlich **größere Ausbreitung**. Während bei der Normalkurve 5% und 1% der Gesamtfläche außerhalb der Grenzen $\pm 1,96$ und $\pm 2,58$ liegen, lauten die entsprechenden Werte für 5 Freiheitsgrade $\pm 2,57$ und $\pm 4,03$. Für 120 Freiheitsgrade werden mit $\pm 1,98$ und $\pm 2,62$ die Grenzen der $N(0; 1)$ -Verteilung fast erreicht.

Die Wahrscheinlichkeitsdichte der t -Verteilung ist durch (5.117) gegeben.

$$f(x) = \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{\pi n}} \left(1 - \frac{x^2}{n}\right)^{-\frac{n+1}{2}} \quad (5.117)$$

Dabei bezeichnet Γ die so genannte Eulersche **Gammafunktion**, die uns wiederholt begegnen wird. Die Gammafunktion ist durch (5.118) definiert.

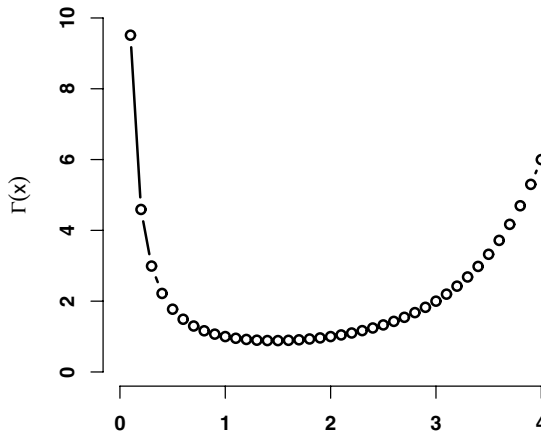


Abb. 5.29. Gamma-Funktion für den Bereich $0 < x \leq 4$

$$\Gamma(x) := \int_0^{\infty} t^{x-1} e^{-t} dt \quad \text{für } x > 0 \quad (5.118)$$

Wichtige Eigenschaften der Gammafunktion sind in 5.119 zusammengestellt.

$$\begin{array}{ll}
 1. & \Gamma(0) \text{ ist nicht definiert} \\
 2. & \Gamma(1) = \Gamma(2) = 1 \\
 3. & \Gamma(0,5) = \sqrt{\pi} \approx 1,77 \\
 4. & \Gamma(1,5) \approx 0,89 \\
 5. & \Gamma(3) = 2 \quad [\text{siehe auch 7.}] \\
 6. & \Gamma(\infty) = \infty \\
 7. & \Gamma(n+1) = n\Gamma(n) = n! \quad \text{für } n = 1, 2, 3, \dots \\
 & \quad \text{d.h. } \Gamma(n) = (n-1)! \\
 8. & \Gamma(x+1) = x\Gamma(x) \quad \text{für alle } x > 0 \\
 9. & \Gamma(x)\Gamma(x-1) = \frac{\pi}{\sin(\pi x)} \\
 10. & \Gamma(n+0,5) = \frac{(2n)!\sqrt{\pi}}{n!2^{2n}} \\
 11. & \binom{n+r-1}{n} = \frac{\Gamma(n+r)}{\Gamma(n+1)\Gamma(r)} \\
 12. & \frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)} = \int_0^1 u^{r-1}(1-u)^{s-1} du
 \end{array} \quad (5.119)$$

Für die Berechnung der Wahrscheinlichkeitsdichte und der Verteilungsfunktion einer t -verteilten Zufallsvariablen stehen in **R** die Funktionen **dt()** und **pt()** zur Verfügung.

Für 1, 3, und 8 Freiheitsgrade sind diese Funktionen in Abbildung 5.30 dargestellt.

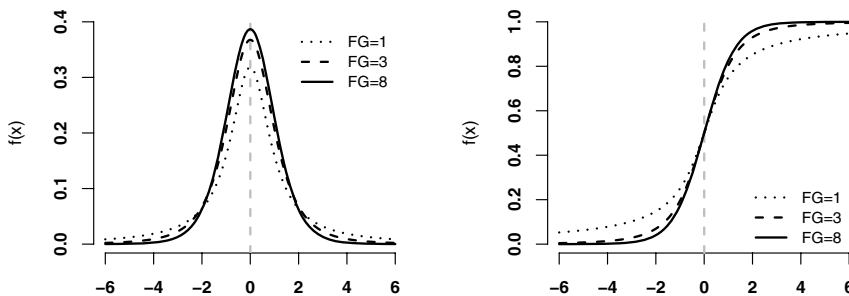


Abb. 5.30. Wahrscheinlichkeitsdichte und Verteilungsfunktion der t -Verteilung mit $FG = 1, 3, 8$

Die **Tabellierung von Werten der t -Verteilung** erfolgt nicht über die Verteilungsfunktion (wie bei der Standard-Normalverteilung), sondern über ausgewählte Quantile. Tabelle 5.10 enthält **Schranken der t -Verteilung**; über einen größeren Bereich von Freiheitsgraden sind die **Quantile zu ausgewählten Wahrscheinlichkeiten** (0,99, 0,975, 0,95 und 0,90) aufgelistet, d.h. **t -Werte**, die auf bestimmten Niveaus (Signifikanzniveaus) im Rahmen von statistischen Schätz- und Testverfahren zu überschreiten sind. Für das Ablesen geht man vom Freiheitsgrad aus; die Wahrscheinlichkeit, mit der die tabellierten t -Werte rein zufällig überschritten werden, sind im Kopf der Tabelle verzeichnet. So erhält man für 5 Freiheitsgrade ($FG = 5$ oder $\nu = 5$) die Überschreitungswahrscheinlichkeit P von $t = 2,571$ zu 0,975 oder 97,5%. Aus Symmetriegründen

sind diese Werte aus Tabelle 5.10 mit negativem Vorzeichen auch auf die Quantile zu 0,01, 0,025, 0,05 und 0,10 zu übertragen; hier müssen entsprechend die tabellierten (negativen) Werte „untersritten“ werden.

Tabelle 5.10. Ausgewählte Quantile (Signifikanzschranken) der t -Verteilung über einen größeren Bereich an Freiheitsgraden(berechnet mit der Funktion **qt()** in **R**); in Klammern gesetzt sind „linksseitige“ Verteilungsanteile bei negativem Vorzeichen

FG ν	Wahrscheinlichkeiten $P(t \leq t_{1-\alpha})$			
	0.99 (0.01)	0.975 (0.025)	0.95 (0.05)	0.90 (0.10)
1	31.8205	12.7062	6.3138	3.0777
2	6.9646	4.3027	2.9200	1.8856
3	4.5407	3.1824	2.3534	1.6377
4	3.7469	2.7764	2.1318	1.5332
5	3.3649	2.5706	2.0150	1.4759
6	3.1427	2.4469	1.9432	1.4398
7	2.9980	2.3646	1.8946	1.4149
8	2.8965	2.3060	1.8595	1.3968
9	2.8214	2.2622	1.8331	1.3830
10	2.7638	2.2281	1.8125	1.3722
11	2.7181	2.2010	1.7959	1.3634
12	2.6810	2.1788	1.7823	1.3562
13	2.6503	2.1604	1.7709	1.3502
14	2.6245	2.1448	1.7613	1.3450
15	2.6025	2.1314	1.7531	1.3406
16	2.5835	2.1199	1.7459	1.3368
17	2.5669	2.1098	1.7396	1.3334
18	2.5524	2.1009	1.7341	1.3304
19	2.5395	2.0930	1.7291	1.3277
20	2.5280	2.0860	1.7247	1.3253
22	2.5083	2.0739	1.7171	1.3212
24	2.4922	2.0639	1.7109	1.3178
26	2.4786	2.0555	1.7056	1.3150
28	2.4671	2.0484	1.7011	1.3125
	0.98	0.95	0.90	0.80
	„zweiseitige“ Wahrsch. $P(t \leq t_{1-\alpha/2})$			

FG ν	Wahrscheinlichkeiten $P(t \leq t_{1-\alpha})$			
	0.99 (0.01)	0.975 (0.025)	0.95 (0.05)	0.90 (0.10)
30	2.4573	2.0423	1.6973	1.3104
32	2.4487	2.0369	1.6939	1.3086
34	2.4411	2.0322	1.6909	1.3070
36	2.4345	2.0281	1.6883	1.3055
38	2.4286	2.0244	1.6860	1.3042
40	2.4233	2.0211	1.6839	1.3031
42	2.4185	2.0181	1.6820	1.3020
44	2.4141	2.0154	1.6802	1.3011
46	2.4102	2.0129	1.6787	1.3002
48	2.4066	2.0106	1.6772	1.2994
50	2.4033	2.0086	1.6759	1.2987
55	2.3961	2.0040	1.6730	1.2971
60	2.3901	2.0003	1.6706	1.2958
65	2.3851	1.9971	1.6686	1.2947
70	2.3808	1.9944	1.6669	1.2938
75	2.3771	1.9921	1.6654	1.2929
80	2.3739	1.9901	1.6641	1.2922
85	2.3710	1.9883	1.6630	1.2916
90	2.3685	1.9867	1.6620	1.2910
95	2.3662	1.9853	1.6611	1.2905
100	2.3642	1.9840	1.6602	1.2901
250	2.3414	1.9695	1.6510	1.2849
500	2.3338	1.9647	1.6479	1.2832
1000	2.3301	1.9623	1.6464	1.2824
	0.98	0.95	0.90	0.80
	„zweiseitige“ Wahrsch. $P(t \leq t_{1-\alpha/2})$			

Beispiele: siehe Text

Zum Ablesen von **ein- und zweiseitigen Schrankenwerten** (Sicherheitsgrenzen): Beispielsweise können wir für einen einseitigen Test die folgenden beiden t -Werte ablesen: $t_{30;0,05;\text{einseitig}} = t_{30;0,95} = 1,6973$ und $t_{100;0,01;\text{einseitig}} = t_{100;0,99} = 2,3642$. Bei zweiseitiger Sicht (linke und rechte Seite) auf die t -Verteilung gilt für die Flächenanteile: $P(|t| \geq t_{1-\alpha/2}) = P(|t| \geq 2,5706) = 0,05$ oder $t_{5;0,05;\text{zweiseitig}} = t_{5;0,975} = 2,5706$ (Tabelle 5.10 von unten gelesen).

Zweiseitige 5%-Schranken der t -Verteilung liegen oberhalb von

$$t_{\infty;0,05;\text{zweiseitig}} = t_{\infty;0,975} = z_{0,975} = 1,96$$

Für $29 \leq n \leq 70$ oder $28 \leq \nu \leq 69$ lassen sie sich durch den Wert 2,0 annähern; für 60 Freiheitsgrade ergibt sich $t_{60;0,05;\text{zweiseitig}} = t_{60;0,975} = 2,0003$ oder $P(t_{60} \leq -2,0003 \text{ sowie } t_{60} \geq 2,0003) = 0,05$ oder $P(|t_{60}| \geq 2,0003) = 0,05$.

Erwartungswert und Varianz der t -Verteilung sind durch (5.120) gegeben.

Erwartungswert: $\mu = E[t_n] = 0$ für $n \geq 2$	
Varianz: $\sigma^2 = Var[t_n] = \frac{n}{n-2}$ für $n \geq 3$	(5.120)
2. Moment: $\frac{n}{n-2}$ für $n \geq 3$	

5.5.2 Chi-Quadrat-Verteilung (χ^2)

Sind $Z_1, Z_2, \dots, Z_n$ unabhängige, standardnormalverteilte Zufallsvariablen ($N(0; 1)$), dann heißt die Zufallsvariable

$$\chi_n^2 = Z_1^2 + Z_2^2 + \dots + Z_n^2 = \sum_{i=1}^n Z_i^2 \quad (5.121)$$

Chi-Quadrat-verteilt mit n Freiheitsgraden (kurz χ_n^2 oder χ_ν^2). Die Wahrscheinlichkeitsdichte dieser Zufallsvariablen wird für $x \geq 0$ durch (5.122) gekennzeichnet.

$$f(x) = \frac{1}{2^{n/2} \Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \quad (5.122)$$

Dabei bezeichnet Γ die Gammafunktion, die schon bei der Dichte der t-Verteilung (5.118) erwähnt wurde.

Die χ^2 -Verteilung (vgl. auch Abb. 5.31) ist eine stetige unsymmetrische Verteilung. Ihr Variationsbereich erstreckt sich von Null bis Unendlich. Sie nähert sich mit wachsenden Freiheitsgraden („langsam“) einer Normalverteilung $N(\mu = \nu; \sigma^2 = 2\nu)$. Die Form der χ^2 -Verteilung hängt somit ebenfalls wie die der Student-Verteilung nur vom Freiheitsgrad ab. Nimmt dieser zu, so wird die schiefe, eingipflige Kurve flacher und symmetrischer. Eine wesentliche Eigenschaft der χ^2 -Verteilung ist ihre **Additivität**: Wenn zwei unabhängige Größen χ^2 -Verteilungen mit ν_1 und ν_2 Freiheitsgraden haben, so hat die Summe eine χ^2 -Verteilung mit $\nu_1 + \nu_2$ Freiheitsgraden. Entdeckt worden ist diese Verteilung von I.-J. Bienaymé (1858), E. Abbe (1863), F. R. Helmert (1876) und K. Pearson (1900).

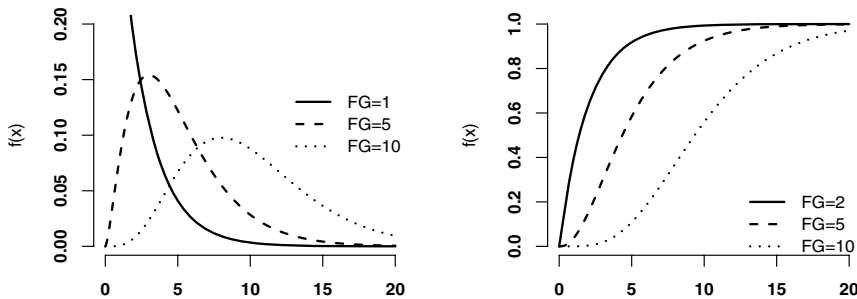


Abb. 5.31. Wahrscheinlichkeitsdichte und Verteilungsfunktion der χ^2 -Verteilung mit $FG = 1, 5, 10$

Die Verteilungsfunktion der χ^2 -Verteilung ist nicht elementar darstellbar. Sie lässt sich für $\nu > 30$ Freiheitsgrade mit der Standardnormalverteilung nach (5.123) recht gut approximieren (Wilson-Hilferty).

$$F(x|\nu) \approx \Phi\left(3 \cdot \sqrt{\frac{\nu}{2}} \left[\frac{2}{9\nu} + \left(\frac{x}{\nu}\right)^{1/3} - 1 \right]\right) \quad (5.123)$$

Für die Berechnung der Wahrscheinlichkeitsdichte und der Verteilungsfunktion einer χ^2 -verteilten Zufallsvariablen stehen in **R** die Funktionen **dchisq()** und **pchisq()** zur Verfügung. Für 2, 5, und 10 Freiheitsgrade sind diese Funktionen in Abbildung 5.31 dargestellt.

Die **Tabellierung** von Werten der χ^2 -Verteilung erfolgt wie bei der t -Verteilung über die **Quantile zu ausgewählten Wahrscheinlichkeiten** (0,01, 0,025, 0,05, 0,10, 0,90, 0,95, 0,975, 0,99). Da die χ^2 -Verteilung nicht symmetrisch ist, sind in Tabelle 5.11 „untere“ und „obere“ Quantile parallel aufgeführt.

Die **Berechnung „exakter“ Wahrscheinlichkeiten** für bestimmte Werte, d.h. $P(\chi^2 \leq x)$ bzw. $P(\chi^2 > x)$, erfolgen in **R** mit der Funktion **pchisq()**, wobei für „rechtsseitige“ Grenzen (Überschreitungswahrscheinlichkeiten) der optionale Parameter **„lower.tail = FALSE“** gesetzt werden muss.

```
> pchisq(2, 5, lower.tail = TRUE)
[1] 0.1508550
> pchisq(3.841458, 1, lower.tail=FALSE)
[1] 0.05000002
```

Für das Ablesen geht man vom Freiheitsgrad aus; die Wahrscheinlichkeit, mit der die tabellierten χ^2 -Werte rein zufällig überschritten werden, sind im Kopf der Tabelle verzeichnet. So erhält man für 10 Freiheitsgrade ($FG = 10$ oder $\nu = 10$) die Überschreitungswahrscheinlichkeit $P = 0,95$ für den Wert $\chi^2 = 18,307$, d.h. $\chi^2_{10;0,95} = 18,307$.



Hinweis: Ältere Auflagen der Angewandten Statistik verwenden hinsichtlich der Quantile zu den Testverteilungen häufig eine andere Symbolik [z. B. $\chi^2_{10;0,05} = 18,307$], die sich an den entsprechenden oberen Überschreitungswahrscheinlichkeiten orientiert.

Erwartungswert und Varianz der χ^2 -Verteilung sind durch (5.124) gegeben.

Erwartungswert: $\mu = E[\chi_n^2] = n$ Varianz: $\sigma^2 = Var[\chi_n^2] = 2n$ 2. Moment: $n(2 + n)$	(5.124)

Ist $(X_1, X_2, \dots, X_n)$ eine Stichprobe mit $n > 1$ unabhängigen Beobachtungen aus einer normalverteilten Grundgesamtheit $N(\mu, \sigma^2)$, und ist S^2 die Varianz dieser Stichprobe, dann folgt der Quotient

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2}$$

einer χ^2 -Verteilung mit $\nu = n-1$ Freiheitsgraden. Die χ^2 -Verteilung kann daher insbesondere zur Herleitung von **Vertrauensbereichen bei der Schätzung von Varianzen** herangezogen werden. Eine besondere Bedeutung kommt der χ^2 -Verteilung auch bei der statistischen **Bewertung von Häufigkeitsdaten** zu.

5.5.3 Fisher-Verteilung (F)

Sind χ_m^2 und χ_n^2 zwei unabhängige χ^2 -verteilte Zufallsvariablen mit m und n Freiheitsgraden, dann heißt die Zufallsvariable

Tabelle 5.11. Ausgewählte Quantile (Signifikanzschranken) der χ^2 -Verteilung über einen größeren Bereich an Freiheitsgraden(berechnet mit der Funktion `qchisq()` in **R**)

ν	0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99
1	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635
2	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345
4	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277
5	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086
6	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812
7	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475
8	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090
9	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666
10	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209
11	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725
12	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217
13	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688
14	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141
15	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578
16	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000
17	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409
18	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805
19	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191
20	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566
22	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289
24	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980
26	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642
28	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278
30	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892
32	16.362	18.291	20.072	22.271	42.585	46.194	49.480	53.486
34	17.789	19.806	21.664	23.952	44.903	48.602	51.966	56.061
36	19.233	21.336	23.269	25.643	47.212	50.998	54.437	58.619
38	20.691	22.878	24.884	27.343	49.513	53.384	56.896	61.162
40	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691
42	23.650	25.999	28.144	30.765	54.090	58.124	61.777	66.206
44	25.148	27.575	29.787	32.487	56.369	60.481	64.201	68.710
46	26.657	29.160	31.439	34.215	58.641	62.830	66.617	71.201
48	28.177	30.755	33.098	35.949	60.907	65.171	69.023	73.683
50	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154
55	33.570	36.398	38.958	42.060	68.796	73.311	77.380	82.292
60	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379
65	41.444	44.603	47.450	50.883	79.973	84.821	89.177	94.422
70	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425
75	49.475	52.942	56.054	59.795	91.061	96.217	100.839	106.393
80	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329
85	57.634	61.389	64.749	68.777	102.079	107.522	112.393	118.236
90	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116
95	65.898	69.925	73.520	77.818	113.038	118.752	123.858	129.973
100	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807
250	200.939	208.098	214.392	221.806	279.050	287.882	295.689	304.940
500	429.388	439.936	449.147	459.926	540.930	553.127	563.852	576.493
1000	898.912	914.257	927.594	943.133	1057.724	1074.679	1089.531	1106.969

Beispiel: $\chi^2_{16;0.99}=32.000$. Das ist die obere 1%-Signifikanzschranke für 16 Freiheitsgrade.

$$F_{m,n} = \frac{\chi_m^2/m}{\chi_n^2/n} \quad (5.125)$$

Fisher-verteilt (benannt nach R. A. Fisher) oder kurz **F-verteilt** mit (m, n) Freiheitsgraden. Die Wahrscheinlichkeitsdichte dieser Zufallsvariablen wird für $x \geq 0$ durch (5.126) gegeben.

$$f(x) = \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right) \Gamma\left(\frac{n}{2}\right)} \left(\frac{m}{n}\right)^{\frac{m}{2}} \cdot \frac{x^{\frac{m}{2}-1}}{\left(1 + \frac{m}{n}x\right)^{\frac{m+n}{2}}} \quad (5.126)$$

Wahrscheinlichkeitsdichte und Verteilungsfunktion zweier F -Verteilungen mit $(2, 5)$ bzw. $(10, 10)$ Freiheitsgraden sind in Abbildung 5.32 dargestellt. Die F -Verteilung ist eine stetige, **unsymmetrische** Verteilung, mit einem Variationsbereich von Null bis Unendlich. Sie ist für $m \leq 2$ L -förmig, für $m > 2$ eingipflig bis glockenförmig.

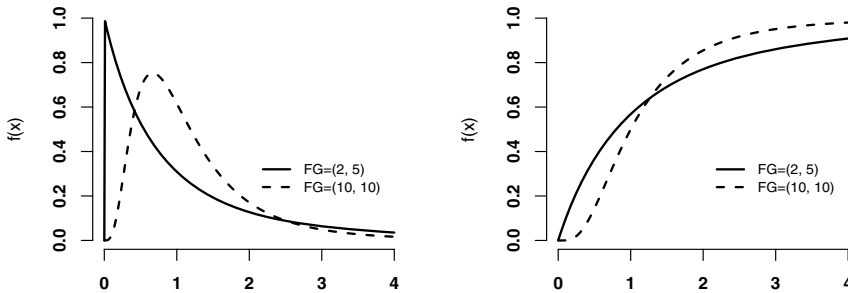


Abb. 5.32. Wahrscheinlichkeitsdichte und Verteilungsfunktion der F -Verteilung mit $FG = (2, 5)$ und $FG = (10, 10)$

Erwartungswert und Varianz der F -Verteilung sind durch (5.127) gegeben.

Erwartungswert: $\mu = E[F_{m,n}] = \frac{n}{n-2}$ für $n > 2$ Varianz: $\sigma^2 = Var[F_{m,n}] = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$ für $n > 4$ 2. Moment: $\frac{2n^2}{(n-2)^2} \left(\frac{m+n+2}{m(n-4)} + \frac{1}{2} \right)$ für $n > 4$	(5.127)

Wenn S_1^2 und S_2^2 Varianzen unabhängiger zufälliger Stichproben der Umfänge n_1 und n_2 aus zwei normalverteilten Grundgesamtheiten mit gleicher Varianz σ^2 sind, dann folgt die zufällige Variable

$$F = \frac{S_1^2}{S_2^2}$$

einer F -Verteilung mit den Parametern $m = n_1 - 1$ und $n = n_2 - 1$. Damit wird diese spezielle Testverteilung insbesondere bei der Überprüfung von Hypothesen bezüglich von Varianzen oder Varianzkomponenten einzusetzen sein.

Bei der Tabellierung der F -Verteilung sind zwei Freiheitsgrade zu berücksichtigen. Außerdem ist zu beachten, dass die Verteilung nicht symmetrisch ist. In Tabelle 5.12 werden nur die 0, 95-Quantile (einseitige obere 5%-Signifikanzschranken) der F -Verteilung über einen größeren Bereich von Freiheitsgraden angegeben. F als Verhältnis zweier Quadrate kann nur Werte zwischen

Tabelle 5.12. Ausgewählte 0.95-Quantile (obere 5%Schranken) der F -Verteilung über einen größeren Bereich an Freiheitsgraden(berechnet mit der Funktion **qf()** in **R**)

FG	n=1	2	3	4	5	6	7	8	9	10
m=1	161.45	18.51	10.13	7.71	6.61	5.99	5.59	5.32	5.12	4.96
2	199.50	19.00	9.55	6.94	5.79	5.14	4.74	4.46	4.26	4.10
3	215.71	19.16	9.28	6.59	5.41	4.76	4.35	4.07	3.86	3.71
4	224.58	19.25	9.12	6.39	5.19	4.53	4.12	3.84	3.63	3.48
5	230.16	19.30	9.01	6.26	5.05	4.39	3.97	3.69	3.48	3.33
6	233.99	19.33	8.94	6.16	4.95	4.28	3.87	3.58	3.37	3.22
7	236.77	19.35	8.89	6.09	4.88	4.21	3.79	3.50	3.29	3.14
8	238.88	19.37	8.85	6.04	4.82	4.15	3.73	3.44	3.23	3.07
9	240.54	19.38	8.81	6.00	4.77	4.10	3.68	3.39	3.18	3.02
10	241.88	19.40	8.79	5.96	4.74	4.06	3.64	3.35	3.14	2.98
12	243.91	19.41	8.74	5.91	4.68	4.00	3.57	3.28	3.07	2.91
14	245.36	19.42	8.71	5.87	4.64	3.96	3.53	3.24	3.03	2.86
16	246.46	19.43	8.69	5.84	4.60	3.92	3.49	3.20	2.99	2.83
18	247.32	19.44	8.67	5.82	4.58	3.90	3.47	3.17	2.96	2.80
20	248.01	19.45	8.66	5.80	4.56	3.87	3.44	3.15	2.94	2.77
25	249.26	19.46	8.63	5.77	4.52	3.83	3.40	3.11	2.89	2.73
30	250.10	19.46	8.62	5.75	4.50	3.81	3.38	3.08	2.86	2.70
40	251.14	19.47	8.59	5.72	4.46	3.77	3.34	3.04	2.83	2.66
50	251.77	19.48	8.58	5.70	4.44	3.75	3.32	3.02	2.80	2.64
100	253.04	19.49	8.55	5.66	4.41	3.71	3.27	2.97	2.76	2.59

FG	n=12	14	16	18	20	25	30	40	50	100
m=1	4.75	4.60	4.49	4.41	4.35	4.24	4.17	4.08	4.03	3.94
2	3.89	3.74	3.63	3.55	3.49	3.39	3.32	3.23	3.18	3.09
3	3.49	3.34	3.24	3.16	3.10	2.99	2.92	2.84	2.79	2.70
4	3.26	3.11	3.01	2.93	2.87	2.76	2.69	2.61	2.56	2.46
5	3.11	2.96	2.85	2.77	2.71	2.60	2.53	2.45	2.40	2.31
6	3.00	2.85	2.74	2.66	2.60	2.49	2.42	2.34	2.29	2.19
7	2.91	2.76	2.66	2.58	2.51	2.40	2.33	2.25	2.20	2.10
8	2.85	2.70	2.59	2.51	2.45	2.34	2.27	2.18	2.13	2.03
9	2.80	2.65	2.54	2.46	2.39	2.28	2.21	2.12	2.07	1.97
10	2.75	2.60	2.49	2.41	2.35	2.24	2.16	2.08	2.03	1.93
12	2.69	2.53	2.42	2.34	2.28	2.16	2.09	2.00	1.95	1.85
14	2.64	2.48	2.37	2.29	2.22	2.11	2.04	1.95	1.89	1.79
16	2.60	2.44	2.33	2.25	2.18	2.07	1.99	1.90	1.85	1.75
18	2.57	2.41	2.30	2.22	2.15	2.04	1.96	1.87	1.81	1.71
20	2.54	2.39	2.28	2.19	2.12	2.01	1.93	1.84	1.78	1.68
25	2.50	2.34	2.23	2.14	2.07	1.96	1.88	1.78	1.73	1.62
30	2.47	2.31	2.19	2.11	2.04	1.92	1.84	1.74	1.69	1.57
40	2.43	2.27	2.15	2.06	1.99	1.87	1.79	1.69	1.63	1.52
50	2.40	2.24	2.12	2.04	1.97	1.84	1.76	1.66	1.60	1.48
100	2.35	2.19	2.07	1.98	1.91	1.78	1.70	1.59	1.52	1.39

Beispiel: $F_{m=12;n=6;0,95} = 4.00$

Tabelle 5.13. Ausgewählte 0.975-Quantile (obere 2,5%Schranken) der F -Verteilung über einen größeren Bereich an Freiheitsgraden(berechnet mit der Funktion **qf()** in **R**)

FG	n=1	2	3	4	5	6	7	8	9	10
m=1	647.79	38.51	17.44	12.22	10.01	8.81	8.07	7.57	7.21	6.94
2	799.50	39.00	16.04	10.65	8.43	7.26	6.54	6.06	5.71	5.46
3	864.16	39.17	15.44	9.98	7.76	6.60	5.89	5.42	5.08	4.83
4	899.58	39.25	15.10	9.60	7.39	6.23	5.52	5.05	4.72	4.47
5	921.85	39.30	14.88	9.36	7.15	5.99	5.29	4.82	4.48	4.24
6	937.11	39.33	14.73	9.20	6.98	5.82	5.12	4.65	4.32	4.07
7	948.22	39.36	14.62	9.07	6.85	5.70	4.99	4.53	4.20	3.95
8	956.66	39.37	14.54	8.98	6.76	5.60	4.90	4.43	4.10	3.85
9	963.28	39.39	14.47	8.90	6.68	5.52	4.82	4.36	4.03	3.78
10	968.63	39.40	14.42	8.84	6.62	5.46	4.76	4.30	3.96	3.72
12	976.71	39.41	14.34	8.75	6.52	5.37	4.67	4.20	3.87	3.62
14	982.53	39.43	14.28	8.68	6.46	5.30	4.60	4.13	3.80	3.55
16	986.92	39.44	14.23	8.63	6.40	5.24	4.54	4.08	3.74	3.50
18	990.35	39.44	14.20	8.59	6.36	5.20	4.50	4.03	3.70	3.45
20	993.10	39.45	14.17	8.56	6.33	5.17	4.47	4.00	3.67	3.42
25	998.08	39.46	14.12	8.50	6.27	5.11	4.40	3.94	3.60	3.35
30	1001.41	39.46	14.08	8.46	6.23	5.07	4.36	3.89	3.56	3.31
40	1005.60	39.47	14.04	8.41	6.18	5.01	4.31	3.84	3.51	3.26
50	1008.12	39.48	14.01	8.38	6.14	4.98	4.28	3.81	3.47	3.22
100	1013.17	39.49	13.96	8.32	6.08	4.92	4.21	3.74	3.40	3.15

FG	n=12	14	16	18	20	25	30	40	50	100
m=1	6.55	6.30	6.12	5.98	5.87	5.69	5.57	5.42	5.34	5.18
2	5.10	4.86	4.69	4.56	4.46	4.29	4.18	4.05	3.97	3.83
3	4.47	4.24	4.08	3.95	3.86	3.69	3.59	3.46	3.39	3.25
4	4.12	3.89	3.73	3.61	3.51	3.35	3.25	3.13	3.05	2.92
5	3.89	3.66	3.50	3.38	3.29	3.13	3.03	2.90	2.83	2.70
6	3.73	3.50	3.34	3.22	3.13	2.97	2.87	2.74	2.67	2.54
7	3.61	3.38	3.22	3.10	3.01	2.85	2.75	2.62	2.55	2.42
8	3.51	3.29	3.12	3.01	2.91	2.75	2.65	2.53	2.46	2.32
9	3.44	3.21	3.05	2.93	2.84	2.68	2.57	2.45	2.38	2.24
10	3.37	3.15	2.99	2.87	2.77	2.61	2.51	2.39	2.32	2.18
12	3.28	3.05	2.89	2.77	2.68	2.51	2.41	2.29	2.22	2.08
14	3.21	2.98	2.82	2.70	2.60	2.44	2.34	2.21	2.14	2.00
16	3.15	2.92	2.76	2.64	2.55	2.38	2.28	2.15	2.08	1.94
18	3.11	2.88	2.72	2.60	2.50	2.34	2.23	2.11	2.03	1.89
20	3.07	2.84	2.68	2.56	2.46	2.30	2.20	2.07	1.99	1.85
25	3.01	2.78	2.61	2.49	2.40	2.23	2.12	1.99	1.92	1.77
30	2.96	2.73	2.57	2.44	2.35	2.18	2.07	1.94	1.87	1.71
40	2.91	2.67	2.51	2.38	2.29	2.12	2.01	1.88	1.80	1.64
50	2.87	2.64	2.47	2.35	2.25	2.08	1.97	1.83	1.75	1.59
100	2.80	2.56	2.40	2.27	2.17	2.00	1.88	1.74	1.66	1.48

Beispiel: $F_{m=25;n=4;0.975} = 8,50$

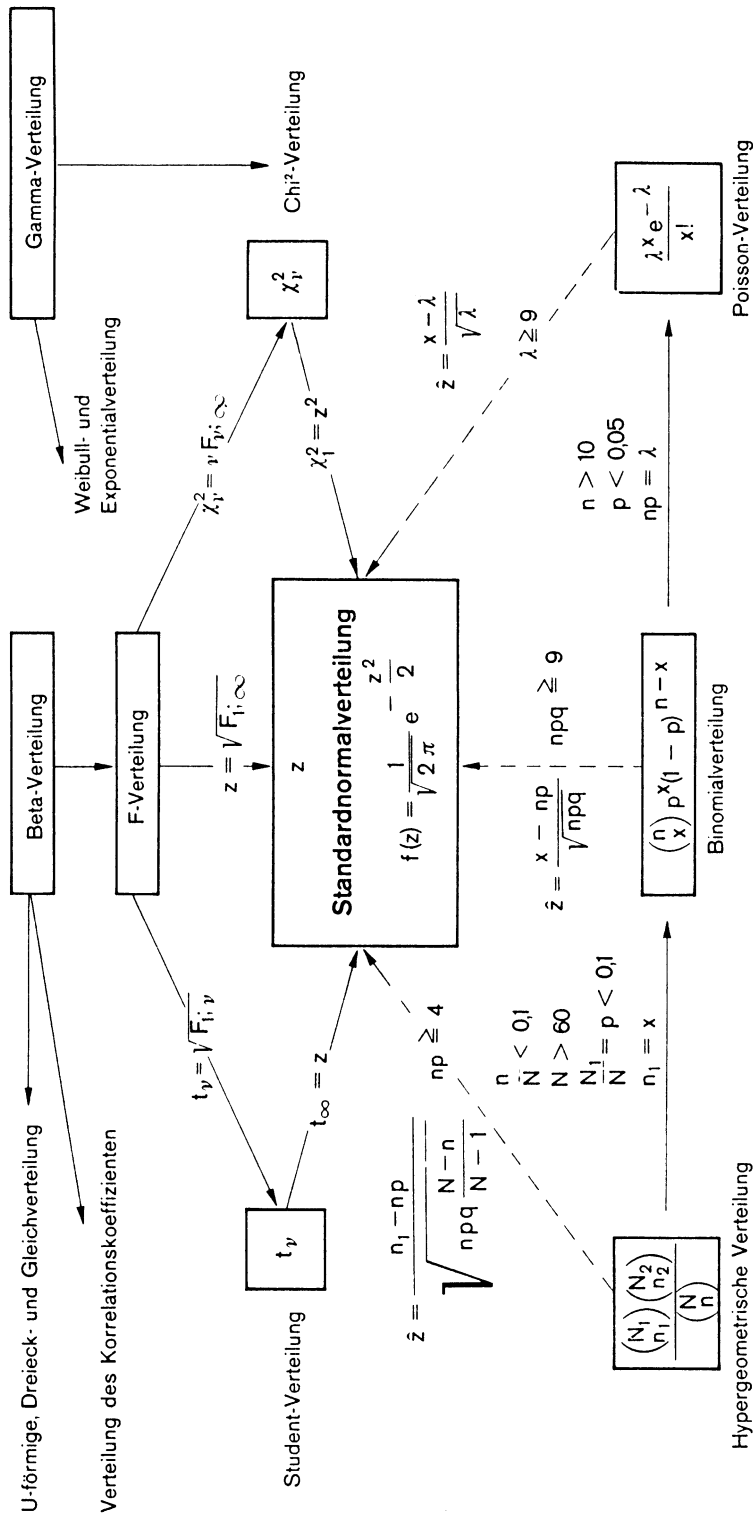


Abb. 5.33. Zusammenhänge zwischen einigen Verteilungen

Null und plus Unendlich annehmen, sich also wie die χ^2 -Verteilung nur rechts des Koordinatenanfangspunktes erstrecken. An die Stelle einer spiegelbildlich symmetrischen Verteilungskurve, wie sie bei der t -Verteilung vorliegt, tritt hier gewissermaßen eine „reziproke Symmetrie“. Wie $+t$ mit $-t$, so kann hier F mit $1/F$ und zugleich m mit n vertauscht werden. Es gilt

$$F(m, n; 1 - \alpha) = 1/F(n, m; \alpha) \quad (5.128)$$

Nach dieser Beziehung lässt sich beispielsweise aus $F_{0,95}$ leicht $F_{0,05}$ ermitteln.

5.5.4 Verteilungen wichtiger Stichprobenfunktionen aus normalverteilten Grundgesamtheiten

In den folgenden Hinweisen A bis D werden exakte Verteilungen einiger wichtiger Stichprobenfunktionen aus normalverteilten Grundgesamtheiten gegeben.

A. Verteilung des Stichprobenmittelwertes

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i; \quad E(\bar{X}) = \mu; \quad \text{Var}(\bar{X}) = \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} :$$

$$\begin{aligned} \bar{X} &\text{ ist } N\left(\mu; \frac{\sigma}{\sqrt{n}}\right)\text{-verteilt} \\ Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} &\text{ ist } N(0; 1)\text{-verteilt} \end{aligned} \quad (5.129)$$

Beispiel: 95%-Konfidenzintervall für μ :

$$\begin{aligned} P\left(-1,96 \leq \frac{\mu - \bar{X}}{\sigma/\sqrt{n}} \leq 1,96\right) &= 0,95 \\ &= P(-1,96\sigma/\sqrt{n} \leq \mu - \bar{X} \leq 1,96\sigma/\sqrt{n}) \\ &= P(\bar{X} - 1,96\sigma/\sqrt{n} \leq \mu \leq \bar{X} + 1,96\sigma/\sqrt{n}) \end{aligned}$$

Bei unbekanntem σ und bei Verwendung der Standardabweichung der Stichprobe

$$S = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}} \text{ gilt: } t = \frac{\bar{X} - \mu}{S/\sqrt{n}} \text{ ist } t\text{-verteilt mit } \nu = n-1 \text{ FG} \quad (5.130)$$

B. Verteilung der Stichprobenvarianz

(1) Erwartungswert der Grundgesamtheit **unbekannt**:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2; \quad E(S^2) = \sigma^2; \quad \text{Var}(S^2) = \frac{2\sigma^4}{n-1} :$$

$$V = \sum_{i=1}^n \left(\frac{X_i - \bar{X}}{\sigma} \right)^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2} \text{ ist } \chi^2\text{-verteilt mit } \nu = n-1 \text{ FG} \quad (5.131)$$

(2) Erwartungswert μ der Grundgesamtheit **bekannt**:

$$S'^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2; \quad E(S'^2) = \sigma^2; \quad \text{Var}(S'^2) = \frac{2\sigma^4}{n} :$$

$$V' = \sum_{i=1}^n \left(\frac{(X_i - \mu)}{\sigma} \right)^2 = \frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2} = \frac{nS'^2}{\sigma^2} \quad \text{ist } \chi^2\text{-verteilt} \\ \text{mit } \nu = n \text{ FG} \quad (5.132)$$

C. Verteilung der Differenz von Stichproben-Mittelwerten

Gegeben seien zwei voneinander unabhängige nach $N(\mu_1; \sigma_1)$ bzw. $N(\mu_2; \sigma_2)$ verteilte Zufallsstichproben mit $\bar{X}_1$ und $\bar{X}_2$:

$$D = \bar{X}_1 - \bar{X}_2, \delta = \mu_1 - \mu_2; \quad E(D) = \delta; \quad \text{Var}(D) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

(1) D ist verteilt wie $N\left(\delta; \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$

$$Z = \frac{D - \delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \quad \text{ist } N(0; 1)\text{-verteilt} \quad (5.133)$$

(2) $\sigma_1 = \sigma_2 = \sigma$ und σ **bekannt**

$$Z = \frac{D - \delta}{\sigma \cdot \sqrt{\frac{n_1 + n_2}{n_1 n_2}}} \quad \text{ist } N(0; 1)\text{-verteilt} \quad (5.134)$$

(3) $\sigma_1 = \sigma_2 = \sigma$ und σ **unbekannt**

$$t = \frac{D - \delta}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \left[\frac{n_1 + n_2}{n_1 n_2} \right]}} \quad \text{ist } t\text{-verteilt mit } \nu = n_1 + n_2 - 2 \text{ FG} \quad (5.135)$$

(4) $\sigma_1 \neq \sigma_2$ und beide **unbekannt**

$$t = \frac{D - \delta}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \quad \text{ist angenähert } t\text{-verteilt mit} \\ \nu = \frac{\left[\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} \right]^2}{\frac{S_1^4}{n_1^2(n_1 - 1)} + \frac{S_2^4}{n_2^2(n_2 - 1)}} \text{ FG} \quad (5.136)$$

D. Verteilung des Quotienten von Stichproben-Varianzen S_1^2 und S_2^2 gegeben:(1) $\sigma_1 = \sigma_2$

$$F = \frac{S_1^2}{S_2^2} \text{ ist } F\text{-verteilt mit } \begin{matrix} m = n_1 - 1 \\ n = n_2 - 1 \end{matrix} \text{ FG} \quad (5.137)$$

(2) $\sigma_1 \neq \sigma_2$

$$F = \frac{S_1^2}{S_2^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} \text{ ist } F\text{-verteilt mit } \begin{matrix} m = n_1 - 1 \\ n = n_2 - 1 \end{matrix} \text{ FG} \quad (5.138)$$

5.6 Verteilung zweidimensionaler Zufallsvariablen

- Modellbildung
- Randverteilungen und Unabhängigkeit
- Korrelationskoeffizient
- Zweidimensionale Normalverteilung
- Multinomialverteilung (Polynomialverteilung)

5.6.1 Modellbildung**5.6.1.1 Einführendes Beispiel**

Die **Ergebnismenge** zum Zufallsexperiment „Werfen von zwei Würfeln“, z.B. ein roter und ein blauer Würfel, wird durch eine Menge von Wertepaaren

$$\Omega = \{(i, j) | 1 \leq i \leq 6, 1 \leq j \leq 6, i \in N, j \in N\}$$

beschrieben, wobei i und j für die jeweilige Augenzahl stehen. Es lassen sich zwei Zufallsvariablen definieren, X - **Augenzahlsumme** und Y - **absolute Differenz in der Augenzahl**, deren Wahrscheinlichkeitsfunktion (eindimensional) elementar aus dem Zufallsexperiment ableitbar ist. Zum Beispiel gilt:

$$P(X = 10) = \frac{3}{36}$$

$$P(Y = 0) = \frac{6}{36}$$

Die Verknüpfung der beiden Zufallsvariablen in einer **gemeinsamen Wahrscheinlichkeit** führt dann auf die **zweidimensionale Wahrscheinlichkeitsfunktion** $P(X = x, Y = y)$:

$$P(X = 10, Y = 2) = \frac{2}{36}$$

$$P(X = 10, Y = 1) = 0$$

$$P(X \geq 10, Y \geq 1) = \frac{4}{36}$$

5.6.1.2 Verteilungsfunktion

Die zweidimensionale Verteilung der Zufallsvariablen $(X, Y) : \Omega \rightarrow \mathbb{R}^2$ wird auf der Grundlage der **Wahrscheinlichkeiten** (diskret, X und Y nehmen höchstens abzählbar viele Werte an) bzw. der **Dichte** (stetig, $f : \mathbb{R}^2 \rightarrow \mathbb{R}_+$) in (5.139) beschrieben.

Diskrete Zufallsvariable X, Y :

$$\begin{aligned} P(X = x, Y = y) &= P(x, y) \\ &= P(\{\omega \in \Omega \mid X(\omega) = x, Y(\omega) = y\}) \end{aligned}$$

Stetige Zufallsvariable X, Y :

$$\begin{aligned} P(X \in [a, b], Y \in [c, d]) &= P([a, b] \times [c, d]) \\ &= \int_a^b \int_c^d f(w, r) dr dw \\ \text{mit } a \leq b \quad \text{und} \quad c \leq d \end{aligned} \tag{5.139}$$

Sei (X, Y) eine zweidimensionale Zufallsvariable, dann kann die **Verteilungsfunktion** auf der Grundlage von (5.139) wie folgt definiert werden:

Diskrete Zufallsvariable X, Y :

$$F(x, y) = \sum_{x_i \leq x} \sum_{y_j \leq y} P(x_i, y_j) \tag{5.140}$$

Stetige Zufallsvariable X, Y :

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(w, r) dr dw$$

Die Verteilungsfunktion F einer zweidimensionalen Zufallsvariablen (X, Y) hat die folgenden **Eigenschaften**:

- (a) $F(x, y) \rightarrow 0$ für $x \rightarrow -\infty$ **oder** $y \rightarrow -\infty$.
- (b) $F(x, y) \rightarrow 1$ für $x \rightarrow +\infty$ **und** $y \rightarrow +\infty$
- (c) F ist in jeder Koordinate monoton wachsend.
- (d) Insbesondere gilt:

$$P(a < X \leq b, c < Y \leq d) = F(b, d) - F(a, d) - F(b, c) + F(a, c)$$

mit $a, b, c, d \in \mathbb{R}$ und $a < b$ und $c < d$.

Beispiel: Die Zahl der Kunden in der Schlange vor zwei Kassen ($\mathbf{X}, \mathbf{Y}$) in einem Supermarkt zu einem bestimmten Zeitpunkt, z.B. eine Stunde vor Geschäftsschluss, ist durch die gemeinsamen Wahrscheinlichkeiten und die Verteilungsfunktion in Tabelle 5.14 gegeben.

Tabelle 5.14. Wahrscheinlichkeiten und Verteilungsfunktion zweier diskreter Zufallsvariablen X, Y am Beispiel der Zahl der Kunden in der Schlange vor zwei Kassen

gemeinsame Wahrscheinlichkeiten				
X / Y	0	1	2	3
0	0,175	0,025	0,000	0,000
1	0,050	0,250	0,050	0,000
2	0,000	0,025	0,200	0,025
3	0,000	0,000	0,050	0,150

Verteilungsfunktion				
X / Y	0	1	2	3
0	0,175	0,200	0,200	0,200
1	0,225	0,500	0,550	0,550
2	0,225	0,525	0,775	0,800
3	0,225	0,525	0,825	1,000



Die Wahrscheinlichkeit dafür, dass zu einem bestimmten Zeitpunkt an beiden Kassen die gleiche Anzahl Kunden ansteht ist gleich (vgl. die Diagonalsumme in Tabelle 5.14):

$$P(X - Y = 0) = \sum_{x=y} \sum_{x=y} f_{x,y} = 0,175 + 0,250 + 0,200 + 0,150 = 0,775$$



Beispiel: Die Zeit, die ein Teenager täglich Sendungen im Fernsehen verfolgt (X) und die Zeit, die für die Bearbeitung von Hausaufgaben aufgewendet wird (Y), X und Y sind stetige Zufallsvariablen, wird durch die gemeinsame Dichtefunktion $f_{x,y} = xy e^{-(x+y)}$ beschrieben (vgl. Abbildung 5.34).

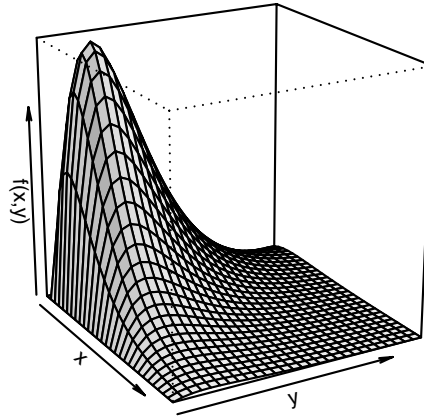


Abb. 5.34. Zweidimensionale Wahrscheinlichkeitsdichte zum Beispiel Fernsehen und Hausaufgaben (fauler Schlingel)

Die Berechnung der gemeinsamen Wahrscheinlichkeiten von zwei stetigen Zufallsvariablen über die entsprechenden Integrale ist aufwendig und soll nur beispielhaft gezeigt werden. Die Wahrscheinlichkeit dafür, dass ein Teenager höchstens eine Stunde am Fernseher verbringt und höchstens eine Stunde für die Hausaufgaben aufwendet ist:

$$\begin{aligned} P(X \leq 1, Y \leq 1) &= \int_0^1 \int_0^1 xy e^{-(x+y)} dx dy \\ &= \int_0^1 y e^{-y} \left(\int_0^1 x e^{-x} dx \right) dy \\ &= \int_0^1 y e^{-y} \left(\left[e^{-x} (-x - 1) \right]_0^1 \right) dy \\ &= 0,264 \int_0^1 y e^{-y} dy = 0,264^2 \approx 0,07 \end{aligned}$$

Hinweis: In der Rechnung wird das bestimmte Integral $\int x e^{ax} dx = a^{-2} e^{ax} (ax - 1)$ verwendet.

5.6.2 Randverteilungen und Unabhängigkeit

Sei (X, Y) eine zweidimensionale Zufallsvariable, dann können aus der gemeinsamen Verteilung $P(x, y)$ die **Randverteilungen** P_1 (für X) und P_2 (für Y) bestimmt werden.

Diskrete Zufallsvariable X, Y :

$$\begin{aligned} P_1(x) &= P(X = x, Y \in R) \\ &= \sum_{y_i} P(x, y_i) \\ P_2(y) &= P(X \in R, Y = y) \\ &= \sum_{x_i} P(x_i, y) \end{aligned} \quad (5.141)$$

Stetige Zufallsvariable X, Y :

$$\begin{aligned} P_1([a, b]) &= P(X \in [a, b], Y \in R) \\ &= \int_a^b \int_{-\infty}^{+\infty} f(w, r) dr dw = \int_a^b f_1(w) dw \\ P_2([c, d]) &= P(X \in R, Y \in [c, d]) \\ &= \int_c^d \int_{-\infty}^{+\infty} f(w, r) dw dr = \int_c^d f_2(r) dr \end{aligned} \quad (5.142)$$

Hinweis: Aus der gemeinsamen Wahrscheinlichkeit $P(X, Y)$ lassen die Randverteilungen bzw. Randdichten ableiten. Die Umkehrung ist jedoch nicht möglich!



Beispiel: Zur Anzahl der Kunden in der Schlange vor zwei Kassen (vgl. Tabelle 5.14) kann die Randverteilung durch die Zeilensummen (X) und die Spaltensummen (Y) bestimmt werden.



Tabelle 5.15. Randverteilungen zweier diskreter Zufallsvariablen X, Y am Beispiel der Zahl der Kunden in der Schlange vor zwei Kassen

$P_1(X = x_i, Y \in R)$				
X	0	1	2	3
P	0,200	0,350	0,250	0,200

$P_2(X \in R, Y = y_j)$				
Y	0	1	2	3
P	0,225	0,300	0,300	0,175

Sei (X, Y) eine zweidimensionale Zufallsvariable, dann heißen die Zufallsvariablen X **und** Y **unabhängig**, wenn im diskreten Fall gilt:

$$P(x, y) = (P(X = x, Y = y) = P_1(x) \cdot P_2(y) \quad \text{für alle } x, y) \quad (5.143)$$

Die Übertragung auf die Randdichten stetiger Zufallsvariablen X und Y führt auf

$$f(w, r) = f_1(w) \cdot f_2(r) \quad w, r \in R \quad (5.144)$$

Der direkte Zusammenhang mit der Definition der Unabhängigkeit von Ereignissen wird deutlich, wenn man formal zwei Ereignisse $A = \{\omega | X(\omega) = x\}$ und $B = \{\omega | Y(\omega) = y\}$ betrachtet, für die dann gilt:

$$P(A \cap B) = P(X = x, Y = y)$$

$$P(A) = P_1(x) \quad \text{und} \quad P(B) = P_2(y)$$

$$X \text{ und } Y \text{ sind unabhängig} \Leftrightarrow P(A \cap B) = P(A) \cdot P(B) \text{ für alle } x, y$$

X und Y sind unabhängig genau dann, wenn $F(x, y) = F_1(x) \cdot F_2(y)$ für alle x, y mit $F_1(x) = F(x, +\infty)$ der Verteilungsfunktion zur Randverteilung von X $F_2(y) = F(+\infty, y)$ der Verteilungsfunktion zur Randverteilung von Y	(5.145)

5.6.2.1 Bedingte Verteilung und Unabhängigkeit

Seien X und Y diskrete Zufallsvariablen, dann sind die **bedingten Verteilungen** definiert durch:

X gegeben y_j $P(X = x_i Y = y_j) = \frac{P(x_i, y_j)}{P_2(y_j)}, \quad i = 1, 2, \dots$ Y gegeben x_i $P(Y = y_j X = x_i) = \frac{P(x_i, y_j)}{P_1(x_i)}, \quad j = 1, 2, \dots$	(5.146)

Seien X und Y stetige Zufallsvariablen, dann sind die **bedingten Dichten** definiert durch:

X gegeben y $f(w y) = \frac{f(w, y)}{f_2(y)}, \quad \text{für } f_2(y) > 0$ Y gegeben x $f(r x) = \frac{f(x, r)}{f_1(x)}, \quad \text{für } f_1(x) > 0$	(5.147)

Der Begriff **Unabhängigkeit zweier Ereignisse** wurde im Kapitel Wahrscheinlichkeiten ausführlich dargestellt. Insbesondere heißen zwei Ereignisse A und B unabhängig, wenn gilt:

$$P(A \cap B) = P(A) \cdot P(B)$$

Daraus folgt für die bedingten Wahrscheinlichkeiten

$$P(A|B) = P(A) \quad \text{für } P(B) > 0 \quad \text{und} \quad P(B|A) = P(B) \quad \text{für } P(A) > 0$$

Zwei **Zufallsvariablen X und Y sind unabhängig**, wenn die folgenden Bedingungen erfüllt sind (die Begründung folgt direkt aus (5.146) bzw. (5.147)):

$$\begin{aligned}
 P(X = x_i | Y = y_j) &= P(X = x_i) = P_1(x_i) \\
 P(Y = y_j | X = x_i) &= P(Y = y_j) = P_2(y_j) \\
 &\text{für alle } x_i, y_j \text{ mit } P(X = x_i) > 0, P(Y = y_j) > 0
 \end{aligned}$$

und

$$\begin{aligned}
 f(w|y) &= f_1(w) \\
 f(r|x) &= f_2(r) \\
 &\text{für alle } x, y \text{ mit } f_2(y) > 0, f_1(x) > 0.
 \end{aligned}$$

Tabelle 5.16. Bedingte Wahrscheinlichkeiten zweier diskreter Zufallsvariablen X, Y am Beispiel der Zahl der Kunden in der Schlange vor zwei Kassen

$P(X = x_i Y = y_j)$					
	0	1	2	3	
0	0,875	0,125	0,000	0,000	1
1	0,143	0,714	0,143	0,000	1
2	0,000	0,100	0,800	0,100	1
3	0,000	0,000	0,250	0,750	1

$P(Y = y_j X = x_i)$				
	0	1	2	3
0	0,778	0,083	0,000	0,000
1	0,222	0,833	0,167	0,000
2	0,000	0,083	0,667	0,143
3	0,000	0,000	0,167	0,857
	1	1	1	1

Beispiel: Die bedingten Verteilungen zum Beispiel der Anzahl der Kunden in der Schlange vor den Kassen (vgl. Tabelle 5.14) werden auf die Zeilen bzw. Spalten bezogen bestimmt. Tabelle 5.16 gibt die entsprechenden bedingten Wahrscheinlichkeiten wieder. Es ist leicht zu erkennen, dass die beiden Zufallsvariablen nicht unabhängig sind, da die bedingten Verteilungen nicht übereinstimmen.

Beispiel: Die bedingten Dichtefunktionen zum Beispiel der Fernsehzeiten X und der Zeit für Hausaufgaben Y lassen sich aus der gemeinsamen Verteilung durch feste Werte für X bzw. Y ableiten (vgl. Abbildung 5.34). Anschaulich entspricht dieses den vertikalen Schnittkurven durch die gemeinsame Dichtefunktion, die durch den Inhalt der Schnittfläche zu normieren sind. Zwei Beispiele für $f(y|X = 2)$ und $f(x|Y = 4)$ sind in Abbildung 5.35 dargestellt. Die Unabhängigkeit der beiden Zufallsvariablen X und Y lässt sich leicht direkt aus der gemeinsamen Dichtefunktion ableiten.

$$f(x, y) = xye^{-(x+y)} = xe^{-x} \cdot ye^{-y} = f_1(x) \cdot f_2(y)$$

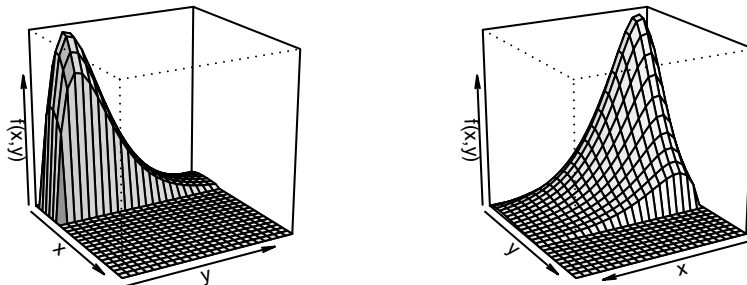


Abb. 5.35. Bedingte Dichtefunktionen $f(y|X = 2)$ (links) und $f(x|Y = 4)$ (rechts) für das Beispiel Fernsehzeiten und Hausaufgaben

5.6.2.2 Satz von Bayes für Zufallsvariablen

Der Satz von Bayes ermöglicht die Bestimmung von „a posteriori“ Wahrscheinlichkeiten auf der Grundlage von „a priori“ (bekannten oder unter bestimmten Annahmen festgelegten) Wahrscheinlichkeiten.

Diskrete Zufallsvariablen X und Y :

$$P(X = x_i | Y = y_j) = \frac{P(Y = y_j | X = x_i)P(X = x_i)}{\sum_k P(Y = y_j | X = x_k)P(X = x_k)}, \quad i = 1, 2, \dots$$

$$P(Y = y_j | X = x_i) = \frac{P(X = x_i | Y = y_j)P(Y = y_j)}{\sum_k P(X = x_i | Y = y_k)P(Y = y_k)}, \quad j = 1, 2, \dots$$

(5.148)

Stetige Zufallsvariablen X und Y :

$$f(w|y) = \frac{f(y|w)f_1(w)}{\int f(y|w)f_1(w)dw} \quad \text{Dichte für } X \text{ gegeben } Y = y.$$

$$f(r|x) = \frac{f(x|r)f_2(r)}{\int f(x|r)f_2(r)dr} \quad \text{Dichte für } Y \text{ gegeben } X = x.$$

5.6.3 Korrelationskoeffizient

Das gemeinsame Moment zweier Zufallsvariablen (um den Erwartungswert) heißt **Kovarianz** $Cov(X, Y)$ (oder auch σ_{XY}) und wird über die Erwartungswerte wie folgt definiert.

$$Cov(X, Y) = E((X - E(X))(Y - E(Y))) = E(X \cdot Y) - E(X) \cdot E(Y)$$

$$= \begin{cases} \sum_i \sum_j (x_i - E(X))(y_j - E(Y))P_{XY}(x_i, y_j) \\ \text{falls } (X, Y) \text{ diskret, und} \\ \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - E(X))(y - E(Y))f(x, y)dxdy \\ \text{falls } (X, Y) \text{ stetig} \end{cases} \quad (5.149)$$

Die Kovarianz ist ein Maß für die **gemeinsame Variation** zweier Zufallsvariablen. Sie ist positiv, wenn die Zufallsvariablen X und Y gemeinsam (gleichzeitig) überwiegend größere Werte als der jeweilige Erwartungswert oder überwiegend kleinere Werte als der jeweilige Erwartungswert annehmen. Treten hinsichtlich der Abweichung vom Erwartungswert überwiegend positive Differenzen in der einen Zufallsvariablen und überwiegend negative Differenzen in der anderen Zufallsvariablen auf, dann ist die Kovarianz negativ.

Die Größe der Kovarianz sagt nichts über die Stärke eines Zusammenhangs (Abhängigkeit oder Unabhängigkeit) zwischen zwei Zufallsvariablen aus. Sie ist abhängig von der Größe (den Wertebereichen) der Zufallsvariablen und ist zudem mit deren Dimensionen behaftet. Die **Normierung der Kovarianz** durch das Produkt der Varianzen von X und Y führt auf die Definition des **Korrelationskoeffizienten** ρ_{XY} in (5.150).

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}} \quad (5.150)$$

Der Korrelationskoeffizient ρ_{XY} ist somit ein **dimensionsloses Maß für die Stärke des Zusammenhangs** zweier Zufallsvariablen X und Y . Insbesondere gilt

$$-1 \leq \rho_{XY} \leq +1.$$

Ein Zusammenhang ist an den **Höhenlinien** (das sind Linien, auf denen die gemeinsame Dichtefunktion $f(x, y)$ den gleichen Wert hat, also horizontale Schnittlinien parallel zur X, Y -Ebene) zu erkennen. Je größer $|\rho_{XY}|$ ist, desto mehr zeigt sich in diesen Linien eine Abhängigkeit zwischen den Zufallsvariablen. Abbildung 5.36 zeigt für das Beispiel **Fernsehzeiten und Hausaufgaben**, dass die Zufallsvariablen unabhängig sind, dass also $\rho_{XY} = 0$ gilt (vgl. auch Abbildung 5.37 und den Text darunter).

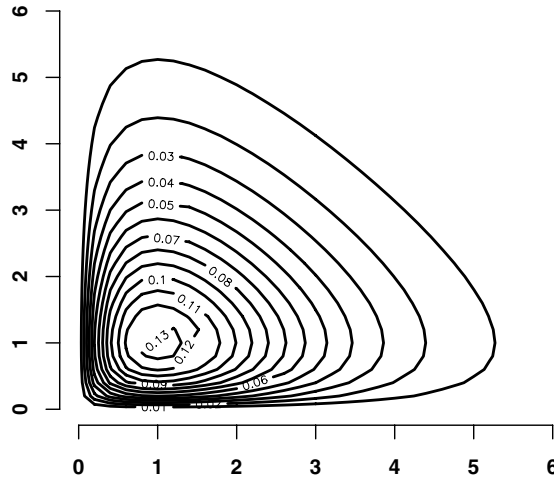


Abb. 5.36. Linien gleicher Wahrscheinlichkeit (Höhenlinien) für die gemeinsame Dichte zweier Zufallsvariablen am Beispiel Fernsehen und Hausaufgaben

Zwei Zufallsvariablen heißen **unkorreliert**, wenn $\rho_{XY} = 0$.

$$\begin{aligned} \rho_{XY} = 0 &\Leftrightarrow \text{Cov}(X, Y) = 0 \\ &\Leftrightarrow E(X \cdot Y) - E(X) \cdot E(Y) = 0 \\ &\Leftrightarrow E(X \cdot Y) = E(X) \cdot E(Y) \end{aligned} \quad (5.151)$$

5.6.4 Zweidimensionale Normalverteilung

Die Dichte einer zweidimensionalen Normalverteilung ist durch

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\frac{x-\mu_x}{\sigma_x}\frac{y-\mu_y}{\sigma_y} + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]\right)$$

gegeben. Die zweidimensionale Normalverteilung $N(\mu_x, \mu_y, \sigma_x, \sigma_y, \rho)$ wird somit durch fünf Parameter bestimmt, deren Bedeutung in der folgenden Zusammenstellung erklärt ist:

Parameter	Bedeutung
μ_x	Erwartungswert der Zufallsvariablen X
μ_y	Erwartungswert der Zufallsvariablen Y
σ_x^2	Varianz der Zufallsvariablen X
σ_y^2	Varianz der Zufallsvariablen Y
ρ	Korrelationskoeffizient von X und Y

Die Standardisierung mit

$$z_1 = \frac{x - \mu_x}{\sigma_x} \text{ und } z_2 = \frac{y - \mu_y}{\sigma_y}$$

führt auf die Standardform einer zweidimensionalen Normalverteilung $N(0, 0, 1, 1, \rho)$

$$f(z_1, z_2) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left(z_1^2 - 2\rho z_1 z_2 + z_2^2\right)\right) \quad (5.152)$$

Beispiele für Dichtefunktionen der standardisierten zweidimensionalen Normalverteilung für $\rho = 0$, $\rho = 0,5$ und $\rho = 0,9$ sind in Abbildung 5.37 dargestellt

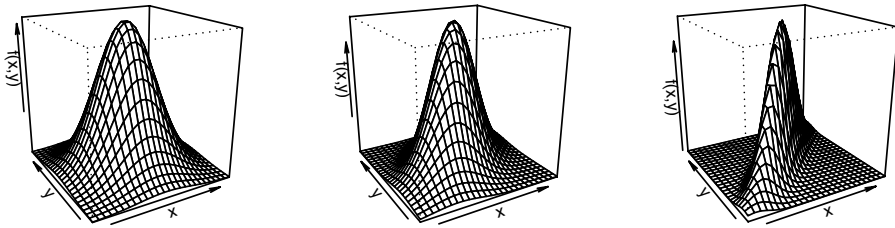


Abb. 5.37. Zweidimensionale standardisierte Normalverteilung für $\rho = 0$, $\rho = 0,5$ und $\rho = 0,9$

Die Bedeutung des Korrelationskoeffizienten ρ hinsichtlich der Form dieser Verteilung wird durch die Höhenlinien (Linien gleicher Wahrscheinlichkeit) deutlich (vgl. Abbildung 5.38). **Je größer $|\rho|$, desto mehr nähern sich die Höhenlinien einer Geraden.** Das Vorzeichen des Korrelationskoeffizienten bestimmt die Orientierung dieser Geraden - positive oder negative Steigung. Der Korrelationskoeffizient kann daher als ein **Maß für die Stärke eines linearen Zusammenhangs** angesehen werden.

Sind X und Y gemeinsam normalverteilt, dann gilt für die **Randverteilungen** von X und Y

$$X \sim N(\mu_x, \sigma_x) \text{ und } Y \sim N(\mu_y, \sigma_y)$$

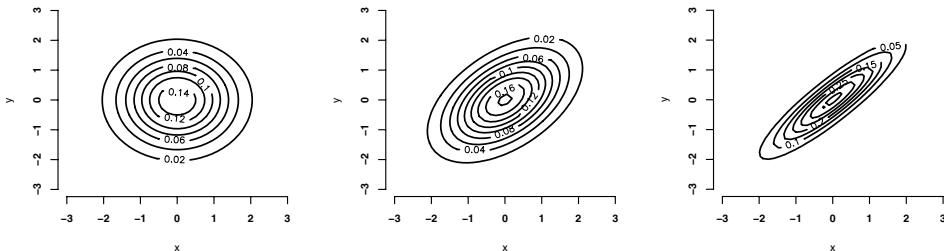


Abb. 5.38. Linien gleicher Wahrscheinlichkeit (Höhenlinien) zur zweidimensionalen standardisierten Normalverteilung für $\rho = 0$, $\rho = 0,5$ und $\rho = 0,9$

Für die **bedingten Verteilungen** von X gegeben $Y = y$, bzw. Y gegeben $X = x$ gilt:

$$N\left(\mu_x + \rho \sigma_x (y - \mu_y) / \sigma_y, \sigma_x \sqrt{1 - \rho^2}\right)$$

$$N\left(\mu_y + \rho \sigma_y (x - \mu_x) / \sigma_x, \sigma_y \sqrt{1 - \rho^2}\right)$$

5.6.5 Multinomialverteilung (Polynomialverteilung)

Wir wissen, wenn die Wahrscheinlichkeit, einen Raucher auszuwählen, p beträgt und die Wahrscheinlichkeit, einen Nichtraucher auszuwählen, $1 - p$ beträgt, dann ist die Wahrscheinlichkeit, genau x Raucher in n Interviews zu erhalten, gegeben durch

$$P(X = x|n, p) = \binom{n}{x} p^x (1 - p)^{n-x} \quad (5.153)$$

Sind statt 2 Ereignisse deren mehrere – sagen wir $A_1, A_2, \dots, A_k$ – möglich mit den entsprechenden Wahrscheinlichkeiten $p_1, p_2, \dots, p_k$, dann sind in n Versuchen mit $n_1, n_2, \dots, n_k$ Realisierungen von $A_1, A_2, \dots, A_k$ die Wahrscheinlichkeiten, genau $x_1, x_2, \dots, x_k$ Ereignisse zu erzielen, gegeben durch (5.154), d. h. sind mehr als zwei Merkmalsausprägungen möglich, besteht also die Grundgesamtheit aus den Merkmalsausprägungen $A_1, A_2, \dots, A_k$ mit den Wahrscheinlichkeiten

$p_1, p_2, \dots, p_k$, wobei $\sum_{i=1}^k p_i = 1$, so ergibt sich für die Wahrscheinlichkeit, dass in einer Stichprobe von n unabhängigen Beobachtungen gerade n_1 -mal die Ausprägung A_1 , n_2 -mal die Ausprägung A_2 usw. auftritt, die so genannte **Multinomialverteilung**

$$P(n_1, n_2, \dots, n_k | p_1, p_2, \dots, p_k | n) = \frac{n!}{n_1! \cdot n_2! \cdot \dots \cdot n_k!} \cdot p_1^{n_1} p_2^{n_2} \cdot \dots \cdot p_k^{n_k} \quad (5.154)$$

deren k Zufallsvariablen n_i der Bedingung $\sum_{i=1}^k n_i = n$ genügen. Die Funktional-Parameter sind für die n_i :

Erwartungswerte: $\mu_i = np_i$ Varianzen: $\sigma_i^2 = np_i(1 - p_i) = np_i q_i$	(5.155)

Ereignis*	$A_1 \cdots A_k$	Summe
Wahrscheinlichkeit	$p_1 \cdots p_k$	1
Häufigkeit**	$N_1 \cdots N_k$	n

* $S = A_1 \cup \dots \cup A_k$ mit $i = 1, 2, \dots, k$

** Zufallsvariablen: N_i ; Realisierungen: n_i ; wegen der Bedingung $\sum N_i = \sum n_i = n$ sind sie nicht stochastisch unabhängig

Für $k = 2$ erhält man als Spezialfall wieder die Binomialverteilung. (5.154) lässt sich auch aus der verallgemeinerten hypergeometrischen Verteilung (5.77) bei festem n und wachsendem N gewinnen.

Multinomialverteilung

- (1) Insgesamt werden n voneinander stochastisch unabhängige Versuche unternommen.
- (2) Bei jedem Versuch resultiert eines von k verschiedenen Ereignissen A_i ($i = 1, \dots, k$).
- (3) Die Wahrscheinlichkeit eines bestimmten Ereignisses A_i ist p_i mit $p_i > 0$; es gilt: $p_1 + p_2 + \dots + p_k = 1$.
- (4) Die entsprechenden Realisierungen der interessierenden diskreten Zufallsvariablen N_i sind $n_1, n_2, \dots, n_k$ mit $n_i = 0, 1, \dots, n$.
- (5) Formel (5.154) [die linke Seite lässt sich schreiben
 $P(N_1 = n_1, N_2 = n_2, \dots, N_k = n_k | p_1, p_2, \dots, p_k | n) =$
gibt die Wahrscheinlichkeit dafür an, dass genau n_i -mal das Ergebnis A_i eintritt.
- (6) Die expliziten Parameter sind n und p_i , Die Erwartungswerte der N_i sind die np_i .



Beispiel (Perlen): Eine Schachtel enthalte 100 Perlen, von denen 50 rot, 30 grün und 20 schwarz gefärbt seien. Wie groß ist die Wahrscheinlichkeit, zufällig 6 Perlen, und zwar 3 rote, 2 grüne und 1 schwarze, auszuwählen?

Da die Auswahl jeweils mit Zurücklegen erfolgt, ist die Wahrscheinlichkeit 1 rote, 1 grüne und 1 schwarze Perle auszuwählen $p_1 = 0,5$, $p_2 = 0,3$ und $p_3 = 0,2$. Die Wahrscheinlichkeit, 6 Perlen der gegebenen Zusammensetzung zu ziehen, ist gegeben durch

$$P = [6!/(3!2!1!)](0,5)^3(0,3)^2(0,2)^1 = 0,135.$$



Beispiel (Würfelspiel): Ein regelmäßiger Würfel wird zwölfmal geworfen. Die Wahrscheinlichkeit, die 1, die 2 und die 3 je einmal und die 4, die 5 und die 6 je dreimal zu werfen (beachte: $1 + 1 + 1 + 3 + 3 + 3 = 12$), ist

$$P = \frac{12!}{1! \cdot 1! \cdot 1! \cdot 3! \cdot 3! \cdot 3!} \left(\frac{1}{6}\right)^1 \left(\frac{1}{6}\right)^1 \left(\frac{1}{6}\right)^1 \left(\frac{1}{6}\right)^3 \left(\frac{1}{6}\right)^3 \left(\frac{1}{6}\right)^3 = 0,001.$$



Beispiel (Wahl eines Kandidaten): Zehn Personen sollen sich für einen von drei Kandidaten (A , B , C) entscheiden. Wie groß ist die Wahrscheinlichkeit für die Wahl: $8A$, $1B$ und $1C$?

$$P = \frac{10!}{8! \cdot 1! \cdot 1!} \left(\frac{1}{3}\right)^8 \left(\frac{1}{3}\right)^1 \left(\frac{1}{3}\right)^1 = 90 \cdot \frac{1}{6561} \cdot \frac{1}{3} \cdot \frac{1}{3} = 0,00152$$

Wahrscheinlichstes Ergebnis wäre: $3A$, $3B$, $4C$ (bzw. $3A$, $4B$, $3C$ bzw. $4A$, $3B$, $3C$) mit

$$P = \frac{10!}{3! \cdot 3! \cdot 4!} \left(\frac{1}{3}\right)^3 \left(\frac{1}{3}\right)^3 \left(\frac{1}{3}\right)^4 = \frac{3\,628\,800}{6 \cdot 6 \cdot 24} \cdot \frac{1}{27} \cdot \frac{1}{27} \cdot \frac{1}{81} = \frac{4200}{59\,049}$$

$$P = 0,07113 \text{ d. h. knapp 47mal häufiger als } P_{8A,1B,1C}.$$

Schätzen

- Zufallsstichproben und Zufallszahlen
- Das Schätzen von Parametern
- Schätzverfahren für Maßzahlen einer Verteilung
- Konfidenzintervalle
- Toleranzgrenzen
- Übereinstimmung von Messwerten nach Bland-Altman

6.1 Zufallsstichproben und Zufallszahlen

Eine Stichprobe ist repräsentativ für die Grundgesamtheit, wenn sie den Schluss auf bestimmte Merkmale ermöglicht: d. h. „... ist repräsentativ für ... hinsichtlich der Merkmale ...“.

Die Beurteilende Statistik setzt stets Zufallsstichproben voraus. Diese meinen wir auch, wenn wir in den folgenden Kapiteln von „Stichproben“, „Daten“, „Beobachtungen“, „Messreihen“ und „Messwerten“ sprechen. Daher noch einmal:

Zufallsstichproben sind Teile einer Grundgesamtheit, die durch einen Auswahlprozess mit Zufallsprinzip aus dieser entnommen und stellvertretend, repräsentativ für die Grundgesamtheit sind. Ein Teil einer Grundgesamtheit kann auch dann als repräsentative Stichprobe angesehen werden, wenn das den Teil bestimmende Teilungs- oder Auswahlprinzip zwar nicht zufällig, aber von den auszuwertenden Merkmalen **stochastisch unabhängig** ist.

Übersicht 14. Datenbeschreibung und Verallgemeinerung

Aktion	Voraussetzung	Ziel	Tätigkeit
(1) Beschreiben		Zusammenfassung	einen Datenkörper knapp charakterisieren
(2) Schätzen	Zufallsstichprobe aus einer definierten Grundgesamtheit	Konfidenzintervall	einen Parameter mit vorgegebener Ungenauig- keit schätzen
(3) Entscheiden		Statistischer Test	eine Nullhypothese mit vor- gegebener Unsicherheit ablehnen

Verallgemeinerungen aufgrund von „**Stichproben, die gerade zur Hand sind**“ und die nicht als Zufallsstichproben angesehen werden können, sind nicht möglich. Ergebnisse aus statistischen Analysen auf Grund derartiger Stichproben müssen sehr vorsichtig / umsichtig interpretiert werden! Mitunter ist wenigstens eine Verallgemeinerung auf eine durch beliebige Vermehrung der vorliegenden Stichprobeneinheiten angenommene **gedachte Grundgesamtheit** möglich, die sich mehr oder weniger von der uns aufgrund der Fragestellung interessierenden Grundgesamtheit unterscheiden wird.



Eine Methode, echte **Zufallsstichproben** zu erzeugen, bietet das Lotterieverfahren. Beispielsweise sollen von 652 Personen einer Grundgesamtheit zwei Stichproben (I und II) zu je 16 Elementen ausgewählt werden. Man nimmt 652 Zettel, beschreibt je 16 mit einer I, je 16 mit einer II; die restlichen 620 Zettel bleiben leer. Lässt man jetzt 652 Personen Lose ziehen, dann erhält man die geforderten Stichproben.

Übersicht 15. Zufallszahlen und Zufälligkeit

Zufallszahlen sind stochastisch unabhängig und gleichverteilt: Jede Ziffer 0,1, ..., 9 ist von ihren Vorgängern stochastisch unabhängig und jede tritt mit der gleichen Wahrscheinlichkeit $P = 0,1$ auf [ihr Erwartungswert ist 4,5, ihre Varianz 8,25]. Durch Ablesen von z. B. 3 Ziffern zugleich erhält man gleichverteilte Zufallszahlen von 000 bis 999.

Anhand der folgenden Tabelle lassen sich Pseudozufallsziffern auf Zufälligkeit prüfen:

Ziffern	Beispiel	Wahrscheinlichkeit
ungleich	7329	$(10 \cdot 9 \cdot 8 \cdot 7)/10^4 = 0,504$
1 Paar	1281	$(6 \cdot 10 \cdot 9 \cdot 8)/10^4 = 0,432$
3 gleiche	5855	$(4 \cdot 10 \cdot 9)/10^4 = 0,036$
2 Paare	2442	$(3 \cdot 10 \cdot 9)/10^4 = 0,027$
4 gleiche	6666	$10/10^4 = 0,001$

Die entsprechenden relativen Häufigkeiten sollten in der Nähe dieser Wahrscheinlichkeiten liegen.

Einfacher löst man Aufgaben dieser Art mit Hilfe einer **Zufallszahlen-Tabelle** (Tabelle 6.1); notiert sind jeweils fünfstellige Zifferngruppen. Angenommen, 16 Zufallszahlen kleiner als 653 werden benötigt. Man liest die Zahlen von links nach rechts, jeweils als Dreizifferngruppe und notiert sich nur diejenigen dreistelligen Zahlen, die kleiner sind als 653. Die sechzehn Zahlen lauten, wenn wir beispielsweise rein zufällig mit der Bleistiftspitze in der 6. Zeile von oben die erste Ziffer der 3. Spalte treffen und mit ihr beginnen : 202, [unberücksichtigt bleibt 881 > 653], 244, 187, 052, 512, 355, 631, 211, 542 usw.

Wenn aus einer Grundgesamtheit von N Elementen eine Stichprobe von n Elementen ausgewählt werden soll, kann allgemein folgende Vorschrift befolgt werden:

1. Ordne den N Elementen der Grundgesamtheit Zahlen von 1 bis N zu. Wenn $N = 600$, dann wären die Einzelelemente von 001 bis 600 zu nummerieren, wobei jedes Element durch eine dreistellige Zahl bezeichnet ist.
2. Wähle eine beliebige Ziffer der Tafel zum Ausgangspunkt und lies die folgenden Ziffern, jeweils als Dreiergruppe, wenn die Grundgesamtheit eine dreistellige Zahl ist. Ist die Grundgesamtheit eine z -stellige Zahl, dann sind Gruppen aus je z Ziffern zusammenzufassen.
3. Wenn die in der Tabelle abgelesene Zahl kleiner oder gleich N ist, wird das so bezeichnete Element der Grundgesamtheit in die Zufallsstichprobe von n Elementen übernommen. Ist die abgelesene Zahl größer als N oder ist das Element schon in die Stichprobe aufgenommen, dann wird diese Zahl nicht berücksichtigt; man wiederhole den Prozess, bis die n Elemente der Zufallsstichprobe ausgewählt sind.

Zufallsstichproben aus von 1 bis N durchnummerieren Merkmalsträgern einer Grundgesamtheit lassen sich anhand von Zufallszahlen gewinnen. Soll z. B. aus einer Gruppe von $N = 800$ Personen eine 15%ige Zufallsstichprobe (Auswahlsatz: $n/N = 0,15$) gewonnen werden, d. h. $n = 0,15 \cdot 800 = 120$, dann entnimmt man einer Tabelle 3-stellige Zufallszahlen, die die Nummern der auszuwählenden Personen bezeichnen.

Hinweis: Eine der ältesten Methoden zur Erzeugung von Zufallszahlen, man spricht besser von Pseudozufallsziffern, ist die auf von Neumann zurückgehende „**Middle-Square**“-Methode: eine s -zifferige Zahl (s gerade) wird quadriert, ausgewählt werden die mittleren s Ziffern des $2s$ -

Tabelle 6.1. Tabelle mit Zufallszahlen in Blöcken zu je 5 Ziffern

Zeile Nr.	Spalte Nr.										Zeile Nr.
	1-5	6-10	11-15	16-20	21-25	26-30	31-35	36-40	41-45	46-50	
1	26759	86367	75096	13609	16110	73533	42564	67362	43218	50076	1
2	79924	25651	73254	96067	50717	13878	03216	09060	64297	51674	2
3	02510	26113	29832	96118	75792	25326	22940	53548	13564	59089	3
4	32989	74014	67160	39408	97056	43517	84426	25860	86355	33941	4
5	53412	09013	81060	19422	65596	59787	47939	16275	07100	92063	5
6	51202	88124	41870	52689	51275	83556	31211	54288	39296	37318	6
7	26123	05155	59194	52799	28225	85762	33216	19358	02591	54263	7
8	85205	41001	12535	12133	14645	23541	12489	51924	86871	92446	8
9	71899	15475	95434	98227	21824	19585	77940	39298	97838	95145	9
10	47348	20203	18534	03862	78095	50136	89295	59062	39404	13198	10
11	82486	84846	99254	20673	37800	63835	71051	84724	52492	22342	11
12	21885	32906	92431	72438	01174	42159	11392	20724	54322	36923	12
13	60336	98782	07408	18148	81386	80431	90628	52506	02016	85151	13
14	43937	46891	24010	99805	10419	76939	25993	03544	21560	83471	14
15	97656	63175	89303	55835	38835	59399	13790	35112	01324	39520	15
16	74976	14631	35908	28221	39470	91548	12854	30166	09073	75857	16
17	35893	71628	70189	26436	63407	91178	90348	55359	80392	41012	17
18	35676	12797	51434	82976	42010	26344	92920	92155	95407	54644	18
19	74815	67523	72985	23183	02446	63503	92924	20633	58842	85961	19
20	45246	95048	65173	50989	91060	89894	36063	32819	68559	99221	20

stelligen Quadrats. Diese Zahl wird quadriert usw.; die s -zifferigen Zahlen stellen dann Folgen von **Pseudozufallsziffern** dar. Aber: mit 3792 beginnend erhält man mit $3792^2 = 14379264$ die Zufallsziffern, 37923792 . . . Entsprechendes passiert, wenn man von 495475 und 971582 ausgeht. Gute Zufallszahlen sind die unperiodischen Folgen der Dezimalentwicklungen gewisser Irrationalzahlen, wie etwa $\sqrt{2}$, $\sqrt{3}$, $\pi = 3,141592653589793238462643$ und die meisten Logarithmen.

Für die Gewinnung von Zufallszahlen kann in **R** die Funktion **sample()** verwendet werden. Mit dem zusätzlichen Argument „replace=FALSE“ wird verhindert, dass einzelne Zahlen mehrfach auftreten. Zum Beispiel können 20 Zahlen aus dem Bereich der Zahlen zwischen 1 und 80 wie folgt bestimmt werden:

```
> sample(1:80, 20, replace = FALSE)
[1] 21 45 13 18 27 39 25 79 61 69 57 34 1 17 60 74 8 53 38 29
```

Im allgemeinen ist die Grundgesamtheit sehr umfangreich und nicht durchnummeriert. Lassen sich die Merkmalsträger der Grundgesamtheit vor der Ziehung in eine Zufallsanordnung bringen, dann ist die **systematische Auswahl mit Zufallsstart** angemessen: sie besteht darin, z. B. bei einem Auswahlatz von 1%, jeden 100. Merkmalsträger auszuwählen, wobei man etwa mit dem 7. (Zufallsstart) beginnt. Kann keine Zufallsanordnung der Grundgesamtheit vorausgesetzt werden, so ist bei systematischen Stichproben mit Verzerrungen zu rechnen, man spricht von **selektierten Stichproben**, d. h. gewisse Teilmengen der Grundgesamtheit sind stärker als angemessen in der Stichprobe vertreten, diese ist dann nicht mehr repräsentativ für jene. Daher sind Auswahlverfahren z. B. nach dem Anfangsbuchstaben des Familiennamens (vgl. Häufigkeiten nord-, süddeutscher, ausländischer Namen) oder nach dem Geburtstag (mögliche Bevorzugung „runder“ Daten) nicht unbedingt repräsentativ. Ist die Grundgesamtheit hinsichtlich der interessierenden Merkmale sehr unterschiedlich, sehr heterogen, gibt es zusätzliche Schwierigkeiten, die man mit Hilfe der sog. **geschichteten Auswahl** überwindet.



6.1.1 Spezielle Stichprobenverfahren

Wissen wir einiges über die zu erwartende Heterogenität innerhalb der Grundgesamtheit, die wir untersuchen wollen, dann gibt es wirksamere Verfahren als die Auswahl zufälliger Stichproben. Wichtig ist die Verwendung **geschichteter** oder **stratifizierter** Stichproben; hier wird die Grundgesamtheit in relativ homogene Teilgrundgesamtheiten, **Schichten oder Strata** unterteilt, und zwar jeweils nach den Gesichtspunkten, die für das Studium der zu untersuchenden Variablen von Bedeutung sind. Geht es um die Voraussage von Wahlergebnissen, dann wird man die Stichprobe so wählen, dass sie ein verkleinertes Modell der Gesamtbevölkerung darstellt. Dabei werden in erster Linie Altersschichtung, das Verhältnis zwischen Männern und Frauen und die Einkommensgliederung berücksichtigt. So gliedern sich die Erwerbstätigen in der BRD im April 1990 (Statistisches Jahrbuch 1992, S. 114) nach der Stellung im Beruf etwa in 37% Arbeiter, 43% Angestellte, 9% Selbständige und 9% Beamte sowie 2% mithelfende Familienangehörige¹. Stratifizierung verteuert meist die Stichprobenerhebung, ist jedoch ein wichtiges Hilfsmittel. Der Stichprobenumfang pro Schicht ist um so kleiner, je kleiner die Schicht, je kleiner die Varianz und je teurer die Erhebung in der betreffenden Schicht ist.

Demgegenüber wird in der **systematischen Stichprobe** so vorgegangen, dass jedes q -te Individuum der Grundgesamtheit nach einer Liste ausgewählt wird. Hierbei ist q der auf eine ganze Zahl aufgerundete Quotient, den man bei der Division der Gesamtbevölkerung durch den Stichprobenumfang erhält. Bei der Auswahl einer systematischen Stichprobe kann man Volkszählungen, Wahllisten sowie Karteien der Einwohnermeldeämter oder der Gesundheitsbehörden verwenden. Vorausgesetzt wird allerdings, dass die zugrundeliegende Liste frei von periodischen Schwankungen ist. Eine *einwandfreie Zufallsauswahl* ist allerdings nur dann möglich, wenn die Einheiten – etwa Karteikarten – durch *Mischen* in eine *Zufallsanordnung* gebracht werden und dann *systematisch jede q -te Karte* gezogen wird. Die Verwendung einer systematischen Stichprobe hat den Vorteil, dass es oft leichter ist, jedes q -te Individuum herauszugreifen als rein zufällig auszuwählen. Außerdem bringt die Methode in bestimmten Fällen eine **indirekte Stratifikation** mit sich, beispielsweise wenn die Ausgangsliste nach Wohnorten, Berufen oder Einkommensgruppen geordnet wird.

Besonders bei geographischen Problemstellungen verwendet man die Stichprobe mit geschlossenen Erfassungsgruppen, das **Klumpen-Verfahren**. Die Grundgesamtheit wird hier in kleine relativ homogene Gruppen oder Klumpen unterteilt, die man mit wirtschaftlichem Vorteil gemeinsam untersuchen kann. Untersucht wird dann eine zufällige Stichprobe der Klumpen (Familien, Schulklassen, Häuser, Dörfer, Straßenblöcke, Stadtteile). Mehrstufige Zufallsauswahlen sind hier gut möglich (z. B. Dörfer und hieraus wieder zufällig Häuser). Erhebungsgrundlagen für Klumpen (Gemeinden, Betriebe, Kliniken, Haushalte) liegen meist vor. Klumpen sind auch stabiler in der Zeit als die jeweiligen Untersuchungseinheiten (Haushalte, Beschäftigte, Patienten [bzw. Sektionsfälle], Personen).

Dass es nicht einfach ist, **auswahlbedingte (selektionsbedingte) Trugschlüsse** zu vermeiden, zeigt folgendes Beispiel: Angenommen, zwischen zwei Krankheiten bestehe stochastische Unabhängigkeit und die Wahrscheinlichkeiten, in die Klinik aufgenommen bzw. seziert zu werden, seien für beide Krankheiten unterschiedlich. Hierdurch werden einzelne Gruppen unterschiedlich selektiert, wodurch künstlich Abhängigkeiten geschaffen werden. Diese **Selektionskorrelation**, die für die Bevölkerung wie gesagt nicht zutrifft, ist von J. Berkson als Trugschluss erkannt worden. Sie wird **Berkson's Fallacy** genannt. Sie basiert auf der Nichtberücksichtigung unterschiedlicher Zutritts- und Austrittswahrscheinlichkeiten.

¹ D, April 2001, Statistisches Jahrbuch 2002, S. 102; [wie oben]: 33%; 50%; 10%; 6%; 1%.

6.1.1.1 Schlussziffern Auswahl

Bei der Schlussziffern Auswahl gelangen z. B. für einen Auswahlatz von 20% (2%, 0,2%) die Nummern mit den aus einer Zufallszahlentabelle entnommenen Schlussziffern 6 und 9 (11 und 53; 008 und 729) in die Stichprobe (Übersicht 14).

Das Schlussziffernverfahren setzt eine von 1 bis N durchnummerierte Grundgesamtheit voraus. Bei der Auswahl nach Schlussziffern sind diese nach einem Zufallsverfahren festzulegen. Dann erfolgt, dem Auswahlatz und der zugehörigen Schlussziffernkomination entsprechend, die Auswahl aller nummerierten Einheiten, deren Schlussziffern zutreffen.

Übersicht 16. Schlussziffern Auswahl

Gewählter Auswahlatz n/N	Geeignete Schlussziffernkonbination	Die Stichprobe besteht z.B. aus Elementen mit der/den Schlussziffern
0,20	zwei beliebige Ziffern	6 und 9
0,10	eine beliebige Ziffer	7
0,05	fünf zweistellige Zahlen	02, 13, 48, 77, 90
0,03	drei zweistellige Zahlen	05 und 11 und 89
0,01	eine zweistellige Zahl	68
0,002	zwei dreistellige Zahlen	273 und 634
0,001	eine dreistellige Zahl	341

Bei der Auswahl anhand von zufällig festgelegten 3-, 2- oder 1stelligen Schlussziffern (keine echte Zufallsauswahl!) kann somit ein geplanter Auswahlatz genau eingehalten werden. Ebenso wie Geburtstags- und Buchstabenverfahren (Personenauswahl nach dem Geburtsdatum bzw. dem Namensanfang) handelt es sich auch beim Schlussziffernverfahren um eine „Klumpenstichprobe“. Der Vorteil des Schlussziffernverfahrens gegenüber den anderen beiden besteht darin, dass ein vorgegebener Auswahlatz genau eingehalten werden kann, insbesondere an Personengruppen. Studien dieser Art dienen dazu, Ursachen auf Wirkungen zurückzuführen, ohne dass eine randomisierte Zuordnung von Behandlungen zu den Personen möglich ist.

6.1.1.2 Geburtstagsverfahren

Bei diesem Auswahlverfahren werden alle Personen in die Stichprobe einbezogen, die an bestimmten Tagen im Jahr geboren sind. Werden z. B. alle am 11. eines Monats Geborenen ausgewählt, so erhält man eine Stichprobe mit einem Auswahlatz von etwa $12 : 365 = 0,033$, d. h. rund 3%. Das Verfahren kann nur dann benutzt werden, wenn geeignete Auswahlgrundlagen (z. B. Liste, Kartei) für den zu erfassenden Personenkreis vorliegen.

6.2 Das Schätzen von Parametern

- Wünschenswerte Eigenschaften von Schätzfunktionen
- Gesetz der großen Zahlen
- Der mittlere quadratische Fehler

Die Vorschrift, unbekannte Parameter mit Hilfe von Zufallsstichproben angenähert zu bestimmen, genauer, die Vorschrift zur Berechnung eines Schätzwertes aus den Zufallsvariablen X heißt **Schätzfunktion**. Diese Schätzfunktion (estimator), z. B. $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ zur Schätzung des Erwartungswertes $E(X) = \mu$, ist als Funktion von Zufallsvariablen selbst eine Zufallsvariable, hat also eine eigene Verteilung. Für die Verteilung von $\bar{X}$ in unserem Beispiel gilt: (1) der Erwartungswert von $\bar{X}$ ist gleich μ , (2) die Varianz von $\bar{X}$ ist gleich σ^2/n , sie strebt mit zunehmendem n gegen Null.

Von einer brauchbaren Schätzfunktion erwartet man, dass die Schätzung mit zunehmendem n immer besser wird (Gesetz der großen Zahlen), d. h., dass der **Schätzwert** (estimate) mit zunehmender Sicherheit den unbekannten Parameter beliebig genau trifft. Die Punktschätzung ist somit eine Funktion der Realisierungen x_i der Zufallsvariablen X_i , die anhand einer konkreten Zufallsstichprobe einen bestimmten Schätzwert $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ für den Parameter μ liefert.

6.2.1 Vorbemerkungen

6.2.1.1 Der Schluss von der Stichprobe auf die Gesamtheit

In der Beurteilenden oder Schließenden Statistik werden die untersuchten Merkmale, über die man Näheres erfahren möchte, als Zufallsgrößen aufgefasst: X_i ist dann die Zufallsvariable, die angibt, welchen Wert

- (1) X beim i -ten Zufallsexperiment annehmen wird,
- (2) von X das i -te Zufallsstichproben-Element haben wird.

In beiden Fällen ist x_i der beobachtete Wert von X , d. h. die Realisierung von X_i , wobei $i = 1, 2, \dots, n$.

Man nennt $X_1, X_2, \dots, X_n$ **Zufallsvariablen**,
 $x_1, x_2, \dots, x_n$ **Stichprobenwerte**;
 n heißt **Stichprobenumfang**.

Anhand von n Stichprobenwerten bemüht man sich um Aufschluss über die Verteilung von X in der unbekannten experimentell erzeugten (1) bzw. bereits vorliegenden (2) Grundgesamtheit und ihre Parameter, wobei ein vereinfachtes Modell der Wirklichkeit entsteht. Schätzfunktionen, Konfidenzintervalle und statistische Tests sind hier entscheidende Hilfen.

Zur Gewinnung von Informationen über unbekannte Parameter eines den Daten zugrundegelegten Wahrscheinlichkeitsmodells, etwa das einer stetigen Verteilung vom Typ $\dots$, dienen

- (1) **Schätzfunktionen**: sie schätzen anhand von **Punktschätzungen** Parameter aus einer gegebenen Zufallsstichprobe.
- (2) **Konfidenzintervalle**: sie gestatten anhand von **Schätzfunktionen** eine Charakterisierung des betreffenden Parameters durch einen Bereich.
- (3) **Tests**: sie gestatten anhand von Schätzfunktionen den Vergleich von mindestens zwei Parametern.

6.2.1.2 Punktschätzung und Intervallschätzung

Schätzen heißt das Festlegen von Werten für unbekannte Parameter der zugrundegelegten Verteilung mittels eines statistischen Experiments bzw. einer Stichprobe. Man unterscheidet die Punktschätzung eines Parameters von der entsprechenden Intervallschätzung.

Ein **Schätzwert** ist das Ergebnis einer Punktschätzung für einen unbekannten Parameter. Der Schätzwert ist die Realisierung der **Schätzfunktion** in einer konkreten Stichprobe. Unbekannt bleibt, um welchen Betrag der erhaltene Schätzwert von dem betreffenden Parameter abweicht. Daher ergänzt man in vielen Fällen eine Punktschätzung durch eine Intervallschätzung.

Bei der **Intervallschätzung** konstruiert man Intervalle, die den unbekannten Parameter im Mittel mit vorgegebener Vertrauenswahrscheinlichkeit enthalten werden. Diese Intervalle heißen **Konfidenzintervalle** oder **Vertrauensbereiche**. Sie bestehen aus all denjenigen Schätzwerten, die noch

mit dem Wert des Parameters verträglich sind. Die Vertrauensbereichsgrenzen sind Realisierungen von Zufallsvariablen; sie werden aus der Stichprobe berechnet und bilden somit ein Zufallsintervall, das im konkreten Einzelfall als realisierter Vertrauensbereich den Parameter enthält oder auch nicht. Als allgemeine Vorschrift enthält z. B. der 95%-Vertrauensbereich (95%-VB) bzw. das 95%-Konfidenzintervall (95%-KI), dann, wenn alle Voraussetzungen erfüllt sind, bei häufiger Anwendung dieser Vorschrift in rund 19 von 20 Fällen den unbekannten Parameter. Mit wachsendem Stichprobenumfang n lässt sich ein Konfidenzintervall beliebig klein machen. Man geht davon aus, dass die Grundgesamtheit (Umfang N) sehr groß ist. Trifft dies nicht zu, dann erhält man konservative Konfidenzintervalle, d.h. sie sind länger als notwendig. Gilt $n \leq 0,1N$, so sind die Resultate noch angenähert korrekt. Auch wenn die Voraussetzungen erfüllt sind, so gibt das KI lediglich eine untere Grenze der Unsicherheit an.

6.2.1.3 Schätzwert und Schätzfunktion

Die Folge von Beobachtungen $x_1, x_2, \dots, x_n$ einer Zufallsstichprobe seien Realisierungen von n unabhängigen Zufallsvariablen $X_1, X_2, \dots, X_n$, die alle dieselbe Verteilung besitzen; n heißt Stichprobenumfang. Der Erwartungswert [„Mittelwert“] dieser Zufallsvariablen sei $E(X) = \mu$. Meist ist dieser Parameter unbekannt. Geschätzt wird er anhand des **Schätzwertes**

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (6.1)$$

aus einer Zufallsstichprobe. Wird aus einer Zufallsstichprobe ein einziger Schätzwert für den unbekannten Parameter berechnet, so spricht man von einer **Punktschätzung**. Die Vorschrift, unbekannte Parameter mit Hilfe von Zufallsstichproben angenähert zu bestimmen, genauer die Vorschrift zur Berechnung eines Schätzwertes aus den Zufallsvariablen X , heißt **Schätzfunktion**. Diese **Schätzfunktion**

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad (6.2)$$

ist als Funktion von Zufallsvariablen selbst eine Zufallsvariable, hat also eine **Verteilung**. Von einer brauchbaren Schätzfunktion erwartet man, dass die Schätzung mit zunehmendem n immer besser wird (siehe auch Gesetz der großen Zahlen) d. h. dass der Schätzwert (estimate) mit zunehmender Sicherheit den unbekannten Parameter beliebig genau trifft. Die Punktschätzung ist somit eine Funktion der Realisierungen x_i der Zufallsvariablen X_i , die einen bestimmten Schätzwert $\bar{x}$ für den Parameter μ liefert.

Allgemein: Der **Schätzwert** t für den Parameter ϑ (gr. Theta) ist eine spezielle Realisierung der mit T_n oder nur T bezeichneten Schätzfunktion. Für Schätzfunktion (estimator) sagt man auch kurz **Schätzer**.



6.2.2 Wünschenswerte Eigenschaften von Schätzfunktionen

Eine Schätzfunktion T_n oder T sollte:

- (1) einen möglichst geringen systematischen Fehler aufweisen, d.h. **erwartungstreu** sein:

$$E(T) = \vartheta \quad (6.3)$$

- (2) zusätzlich zu Punkt eins auch schon für kleine Stichprobenumfänge eine möglichst geringe Varianz aufweisen, d. h. **effizient** sein:

$$\text{Var}(T) = \text{minimal} \quad (6.4)$$

- (3) einen möglichst kleinen zufälligen Fehler für große Stichprobenumfänge aufweisen, d. h. **konsistent** sein:

$$\begin{aligned} &\text{für großes } n : T_n \approx \vartheta \\ &T_n \text{ strebt stochastisch gegen } \vartheta \end{aligned} \quad (6.5)$$

- (4) alle Informationen in der Stichprobe über den Parameter nutzen, d. h. **suffizient** sein und **robust** sein gegenüber **Abweichungen vom angenommenen Modell**.

Hinweis: Beispielsweise ist der Median $\tilde{X}$ unempfindlich gegenüber der Modellvoraussetzung „Normalverteilung“, nicht aber das arithmetische Mittel $\bar{X}$.

- (5) normalverteilt (zumindest asymptotisch) sein, d.h.

$$\frac{T - E(T)}{\sqrt{\text{Var}(T)}} \approx N(0, 1) \quad (6.6)$$

Diese einzelnen Eigenschaften treten häufig kombiniert auf. So ist z.B. ein

- **gleichmäßig bester unverzerrter Schätzer** erwartungstreu und am effizientesten (uniformly best unbiased estimator).
- **best asymptotisch normaler Schätzer** asymptotisch normalverteilt und besitzt die kleinste Varianz (best asymptotic normal estimator).



Der **Stichprobenmittelwert** $\bar{X}_n$ ist erwartungstreu, effizient und, wenn σ^2 endlich ist, auch konsistent für den Erwartungswert μ der Grundgesamtheit, da die **Varianz des Mittelwertes**

$$\boxed{\text{Var}(\bar{X}) = \sigma_X^2 = \frac{\sigma^2}{n}} \quad (6.7)$$

mit wachsendem n gegen Null strebt. Außerdem ist $\bar{X}_n$ auch suffizient.

6.2.2.1 S^2 ist eine unverzerrte Schätzfunktion, nicht aber S

Um zu zeigen, dass der Erwartungswert von S^2 gleich σ^2 ist, benötigen wir folgende Zerlegung der Summe der Abweichungsquadrate:

Mit $X_i - \bar{X} = (X_i - \mu) + (\mu - \bar{X})$
wird $(X_i - \bar{X})^2 = (X_i - \mu)^2 + (\mu - \bar{X})^2 + 2(X_i - \mu)(\mu - \bar{X})$;
entsprechend:

$$\begin{aligned} \sum (X_i - \bar{X})^2 &= \sum (X_i - \mu)^2 + n(\mu - \bar{X})^2 + 2(\mu - \bar{X}) \sum (X_i - \mu) \\ [\text{vgl. } \sum (X_i - \mu) &= \sum X_i - n\mu = n\bar{X} - n\mu] \\ \sum (X_i - \bar{X})^2 &= \sum (X_i - \mu)^2 + n(\bar{X} - \mu)^2 - 2n(\bar{X} - \mu)^2 \\ &= \sum (X_i - \mu)^2 - n(\bar{X} - \mu)^2 \quad [\text{Verschiebungssatz von Steiner (1)}] \end{aligned}$$

dann gilt: $E(S^2) = E\left[\frac{1}{n-1} \sum (X_i - \bar{X})^2\right]$; die X_i sind stochastisch unabhängig und identisch verteilt;

$$\begin{aligned}
 &= \frac{1}{n-1} E\left[\sum (X_i - \bar{X})^2\right] \\
 &= \frac{1}{n-1} E\left[\sum (X_i - \mu)^2 - n(\bar{X} - \mu)^2\right] \\
 &= \frac{1}{n-1} \left[\sum E(X_i - \mu)^2 - nE(\bar{X} - \mu)^2\right]
 \end{aligned}
 \quad \longrightarrow \quad
 \begin{aligned}
 &= \frac{1}{n-1} \left[n\sigma^2 - n\frac{\sigma^2}{n}\right] \\
 &= \frac{1}{n-1} [(n-1)\sigma^2] \\
 &= \sigma^2.
 \end{aligned}$$

Dagegen ist S keine unverzerzte Schätzfunktion für σ :

$$\begin{aligned}
 \text{Mit } \text{Var}(X) &= E((X - \mu)^2) = E(X^2 - 2\mu X + \mu^2) \\
 &= E(X^2) - 2\mu E(X) + \mu^2 = E(X^2) - 2\mu^2 + \mu^2 \\
 &= E(X^2) - \mu^2 \quad [\text{Verschiebungssatz von Steiner (2)}]
 \end{aligned}$$

$$\text{gilt: } \text{Var}(S) = E(S^2) - [E(S)]^2 > 0 \quad \text{d. h.} \quad [E(S)]^2 < E(S^2) = \sigma^2 \quad E(S) < \sigma.$$

6.2.3 Gesetz der großen Zahlen

Ein Ereignis E habe die Wahrscheinlichkeit π . Bei n unabhängigen Zufallsexperimenten sei die relative Häufigkeit des Auftretens von E gleich $\hat{p}_n$. Für jede beliebig kleine fest vorgegebene positive Zahl ε (gr. epsilon) gilt

$$P(|\hat{p}_n - \pi| < \varepsilon) \rightarrow 1 \quad \text{für } n \rightarrow \infty \quad (6.8)$$

d. h. die Wahrscheinlichkeit, dass $\hat{p}_n$ weniger als ε vom theoretischen Wert π abweicht, strebt (konvergiert) mit wachsendem n gegen Eins. Man kann also mit einer beliebig nahe an 1 gelegenen Wahrscheinlichkeit erwarten, dass bei einer hinreichend großen Zahl n von Versuchen sich die relative Häufigkeit $\hat{p}_n$ beliebig wenig von ihrem Parameter π unterscheidet. Eine Folge dieses sogenannten **schwachen Gesetzes der großen Zahlen**, ist z. B. auch die **stochastische Konvergenz** (Konsistenz) des Stichprobenmittelwertes gegen den Erwartungswert der Grundgesamtheit: $\bar{X}_n$ konvergiert stochastisch gegen μ .

Liegen n unabhängige Zufallsvariablen mit derselben Verteilungsfunktion und endlichem Erwartungswert μ vor, dann strebt das arithmetische Mittel $\bar{X}_n$ mit wachsendem n gegen μ ; und zwar **fast sicher**, d. h. mit **Wahrscheinlichkeit eins**.

Dies ist das **starke Gesetz der großen Zahlen** von Cantelli und von Kolmogoroff.

Auf den Gesetzen der großen Zahlen (qualitativen Konvergenzaussagen) basieren

- 1) die Möglichkeit, Parameter aufgrund von Stichproben beliebig genau zu schätzen, und
- 2) die Monte-Carlo-Technik, auch Monte-Carlo-Simulation oder Simulation genannt.

6.2.4 Der mittlere quadratische Fehler

Der mittlere quadratische Fehler MSE (**mean squared error**) oder auch die mittlere quadratische Abweichung (mean squared deviation) ist ein Maß für die Güte der Schätzfunktion T , mit der der unbekannte Parameter ϑ geschätzt wird:

$$\text{MSE}(\vartheta; T) = E[(T - \vartheta)^2] \quad (6.9)$$

Für Erwartungswert und Varianz einer Normalverteilung gelten (6.10) und (6.11):

$$\text{MSE}(\mu; \bar{X}) = E[(\bar{X} - \mu)^2] = \text{Var}(\bar{X}) = \frac{\sigma^2}{n} \quad (6.10)$$

Dieser MSE ist abhängig von σ^2 und unabhängig von μ .

$$\text{MSE}(\sigma^2; S^2) = \text{Var}(S^2) = \frac{2}{n-1} \sigma^4 \quad (6.11)$$

Formel (6.9) lässt sich mit Hilfe von

$$\begin{aligned} E[(T - \vartheta)^2] &= E[(T - E(T) + E(T) - \vartheta)^2] \\ &\quad \downarrow \\ &= E[(T - E(T))^2] + 2 \underbrace{[E(T) - E(T)][E(T) - \vartheta]}_0 + [E(T) - \vartheta]^2 \\ &= \text{Var}(T) + [E(T) - \vartheta]^2 = \text{Var}(T) + [\text{Bias}(T)]^2 \end{aligned}$$

auch schreiben:

$$\text{MSE}(\vartheta; T) = \text{Var}(T) + [E(T) - \vartheta]^2 \quad (6.12)$$

Der Ausdruck $[E(T) - \vartheta]$ wird als **Bias** oder Verzerrung bezeichnet; d.h.

$$\text{MSE} = \text{Varianz der Schätzfunktion} + \text{Quadrat der Verzerrung}$$



Der mittlere quadratische Fehler setzt sich somit additiv aus zwei Komponenten zusammen. Entfällt die Verzerrung (Bias), dann ist die Schätzfunktion erwartungstreu (unbiased); der MSE einer erwartungstreuen Schätzfunktion ist gleich ihrer Varianz.

Für die Beurteilung der **Güte einer Schätzung** wird als Vergleichskriterium der mittlere quadratische Fehler herangezogen. Eine Schätzfunktion T_1 heißt danach effizienter als eine Schätzfunktion T_2 , wenn gilt:

$$\text{MSE}(\vartheta; T_1) \leq \text{MSE}(\vartheta; T_2)$$

Aus dieser Definition ist jedoch im Allgemeinen keine klare Präferenz zwischen verschiedene Schätzfunktionen ableitbar. Allerdings kann in der Klasse aller erwartungstreuen Schätzfunktionen die Schätzfunktion mit der kleinsten Varianz als effizienteste Schätzfunktion ausgezeichnet werden, da hier eine Verzerrung (Bias) entfällt. Idealerweise sucht man daher nach einer erwartungstreuen Schätzfunktion mit gleichmäßig minimaler Varianz (uniformly minimum variance unbiased estimator). Ein Fehler einer Schätzung (Schätzfehler) wird nach diesen Überlegungen durch die Standardabweichung der Schätzfunktion auf der Grundlage der Stichprobenwerte quantifiziert.

Für die Schätzung der Parameter aus den Stichprobenwerten ist eine umfangreiche Schätzmethodik entwickelt worden. Von besonderer Wichtigkeit ist die **Maximum-Likelihood-Methode** (R.A. Fisher): Sie ist die universellste Methode zur optimalen Schätzung unbekannter Parameter. Sie ist allerdings nur anwendbar, wenn der Typ der Verteilungsfunktion der Variablen bekannt ist; dann bestimmt sie diejenigen Werte als Schätzwerte für die unbekannten Parameter, **die dem erhaltenen Stichprobenresultat die größte Wahrscheinlichkeit des Auftretens verleihen**; d.h. als Schätzwerte werden die Werte mit maximaler Likelihood-Funktion für die Parameter ermittelt, vorausgesetzt die Parameter existieren. Diese Methode zur Konstruktion von Punktschätzungen für Parameter steht in engem Zusammenhang mit der außerordentlich wichtigen **Methode der kleinsten Quadrate**.

6.3 Schätzverfahren für Maßzahlen einer Verteilung

- Momentenmethode
- Schätzung nach der größten Erwartung (MLE)
- Kleinster Fehler (OLS)

6.3.1 Momentenmethode

Bei dem Schätzverfahren nach der Momentenmethode (Karl Pearson 1857-1936) werden die Momente der Grundgesamtheit gleich den Stichprobenmomenten gesetzt. Dabei ist das **k-te Moment** einer Zufallsvariablen X definiert durch:

$$E[X^k] = \begin{cases} \sum_x x^k f(x) & \text{wenn } X \text{ diskret} \\ \int_{-\infty}^{+\infty} x^k f(x) dx & \text{wenn } X \text{ stetig} \end{cases} \quad (6.13)$$

Sei nun $X_1, \dots, X_n$ eine Folge identisch verteilter Zufallsvariablen, dann ist der **Momentenschätzer** (MOM) (method of moments) für das k-te Moment $E[X^k]$ gegeben durch:

$$\boxed{\frac{\sum_{i=1}^n X_i^k}{n}} \quad (6.14)$$

Einige **MOM-Schätzer**, die sich direkt aus der Definition ergeben sind:

(1) $\mu = E[X_i]$ wird durch

$$\bar{X} = \sum_{i=1}^n X_i / n \quad \text{geschätzt.}$$

(2) $E[X_i^2]$ wird durch

$$\sum_{i=1}^n X_i^2 / n \quad \text{geschätzt.}$$

(3) $\sigma^2 = \text{Var}[X_i] = E[X_i^2] - E[X_i]^2$ wird durch

$$\frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 = \frac{\sum_{i=1}^n X_i^2 - n\bar{X}^2}{n} = \frac{n-1}{n} S^2 \quad \text{geschätzt.}$$

(4) Für eine Folge von identisch Poisson-verteilten Zufallsvariablen $X_1, \dots, X_n$ gilt $\mu = E[X_i] = \lambda$; der Parameter der Poisson-Verteilung λ wird entsprechend durch $\bar{X}$ geschätzt.

Hinweis: Da auch $\sigma^2 = \lambda$ gilt, kann auch $\frac{n-1}{n} S^2$ als eine MOM-Schätzung betrachtet werden.

Ein wesentlicher Vorteil von MOM-Schätzern (sofern verfügbar) ist deren einfache Berechnung aus den Stichprobenmomenten. Hinsichtlich der wünschenswerten Eigenschaften sind MOM-Schätzer

- fast stets asymptotisch normalverteilt
- zumindest asymptotisch erwartungstreu.
- immer konsistent.
- oft nicht effizient.
- oft nicht suffizient.

Bessere Eigenschaften als die Momentenschätzer haben in der Regel die nach der Maximum-Likelihood-Methode abgeleiteten Schätzfunktionen.



6.3.2 Schätzung nach der größten Erwartung (MLE)

Das Schätzverfahren nach der größten Erwartung, in der Regel wird die englische Bezeichnung **Maximum Likelihood Estimation (MLE)** bevorzugt, basiert auf der Likelihood-Funktion der Beobachtungsdaten. Darin ist die Likelihood (auch Mutmaßlichkeit) von Beobachtungen durch die Wahrscheinlichkeit bestimmt, die eine konkreten Stichprobe unter einem bestimmten Wahrscheinlichkeitsmodell hat. In der Likelihood-Funktion (6.15) treten die Parameter des Modells als Unbekannte auf und müssen geeignet bestimmt werden. Dazu wird das Maximum der Funktion mit den Stichprobenwerten bestimmt. Umfassende und ausführliche Darstellungen dieses Konzepts sind in Y. Pawitan [Paw01] und T.A. Severini [Sev00] nachzulesen.

Für die zufällige Stichprobe $(X_1, \dots, X_n)$ aus einer diskreten Verteilung, die durch den Parameter ϑ charakterisiert ist, wird die Likelihood-Funktion für eine Realisierung der Zufallsvariablen X_i durch x_i wie folgt definiert:

$$L = L(\vartheta) = \prod_{i=1}^n P(X_i = x_i | \vartheta) \quad (6.15)$$



Beispiel (Münzwurf): Eine Münze wird zehnmal ($n = 10$) geworfen, dabei wird 9mal das Ergebnis „Kopf“ (K) beobachtet. Wie kann aus dieser Beobachtung heraus die „Erfolgswahrscheinlichkeit“ p für K geschätzt werden? Mit dem Modell der Binomialverteilung kann die Likelihood-Funktion wie folgt gebildet werden:

$$\begin{aligned} L(p) &= \binom{10}{9} p^9 (1-p)^1 \\ \ln(L) &= \ln \binom{10}{9} + 9 \ln(p) + 1 \ln(1-p) \\ \frac{\partial \ln(L)}{\partial p} &= \frac{9}{p} - \frac{1}{1-p} = 0 \\ \hat{p} &= \frac{9}{10} = 0,9 \end{aligned}$$

Durch die Transformation mit dem natürlichen Logarithmus wird die Bestimmung der partiellen Ableitungen wesentlich erleichtert. Die Ableitung nach dem unbekannten Parameter wird gleich 0 gesetzt und die Auflösung dieser Gleichung führt auf den ML-Schätzer von p . Da die zweite Ableitung an der Stelle $9/10$ negativ ist, ist diese Schätzung tatsächlich ein Maximum für die Likelihood-Funktion.

Die Wahrscheinlichkeitsfunktion für das Zufallsexperiment 10maliges Werfen einer „regulären“ Münze ($p = 0,5$) und die Likelihood-Funktion für das beobachtete Ergebnis (9mal Kopf) sind in Abbildung 6.1 gegenübergestellt.

Für den Fall einer stetigen Zufallsvariablen kann in (6.15) die Dichtefunktion der Verteilung eingesetzt werden. Die Lösung erfolgt dann analog zum diskreten Fall.

Vorteile:



- (1) Maximum Likelihood Schätzungen können nach einem einheitlichen Muster für eine Vielzahl von Schätzproblemen in der Statistik eingesetzt werden.
- (2) Maximum Likelihood Schätzungen zeigen in der Regel die wünschenswerten mathematischen Eigenschaften. Sie sind
 - stets konsistent,
 - zumindest asymptotisch erwartungstreu,

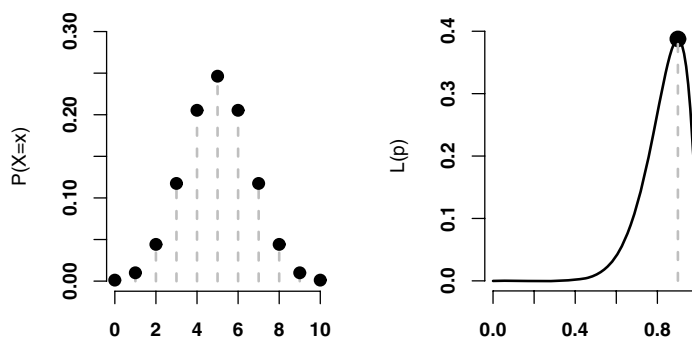


Abb. 6.1. Wahrscheinlichkeitsfunktion für das 10malige Werfen einer regulären Münze und Likelihood-Funktion für das beobachtete Ergebnis von 9mal Kopf

- zumindest asymptotisch effizient,
- suffizient,
- bester asymptotisch normaler Schätzer.

Hinweis: Die letztgenannte Eigenschaft ermöglicht es, neben einer Punktschätzung auch die Verteilung der Schätzfunktion (zumindest asymptotisch) anzugeben. Damit können auch Intervalle berechnet werden, in denen der geschätzte Parameter mit einer vorgegebenen Wahrscheinlichkeit liegt (Konfidenzintervall).

- (3) Maximum Likelihood Schätzungen können nach einem einheitlichen Muster für eine Vielzahl von Schätzproblemen in der Statistik eingesetzt werden.

Nachteile:

- (1) Die Likelihood-Funktionen müssen für eine bestimmte Verteilung und ein spezielles Schätzproblem formuliert werden. Die mathematischen Verfahren sind häufig nicht trivial, insbesondere wenn auch Konfidenzintervalle zu bestimmen sind.
- (2) Die numerischen Verfahren zur Lösung sind in der Regel nicht trivial. Oft ist die Likelihood-Funktion eine hochgradig nicht-lineare Funktion der zu schätzenden Parameter, so dass eine analytische Lösung nicht existiert und numerische Maximierungsverfahren in geeigneten Computerprogrammen eingesetzt werden müssen. Einige einfachere Lösungsansätze werden in den folgenden Abschnitten dargestellt.
- (3) Maximum Likelihood Schätzungen können insbesondere bei kleinen Stichprobenumfängen **erheblich verzerrt** sein. Ihre Optimalität gilt nicht für kleine Stichproben. So erhält man für den Parameter σ^2 einer Normalverteilung mit $S^2 = \frac{1}{n} \sum (X_i - \bar{X})^2$ eine Schätzfunktion, die wegen des Faktors $\frac{1}{n}$ nur asymptotisch erwartungstreu ist.

6.3.2.1 ML-Schätzer zur Binomialverteilung

Sei k die Anzahl von Erfolgen aus einer n -maligen Wiederholung eines Bernoulli-Versuches (K ist eine binomial verteilte Zufallsvariable), dann kann die Wahrscheinlichkeit π für den Erfolg mit dem MLE-Verfahren wie folgt geschätzt werden:

$$\begin{aligned}
 L(\pi|k, n) &= \binom{n}{k} \pi^k (1 - \pi)^{n-k} \\
 \ln(L) &= \ln \binom{n}{k} + k \ln(\pi) + (n - k) \ln(1 - \pi) \\
 \frac{\partial \ln L}{\partial \pi} &= \frac{k}{\pi} - \frac{n - k}{1 - \pi} \\
 \Rightarrow \hat{\pi} &= \frac{k}{n}
 \end{aligned} \tag{6.16}$$

Für die Maximum-Likelihood-Schätzung in **R** steht die Funktion **mle()** zur Verfügung. Auf der Grundlage einer vorgegebenen Likelihood-Funktion (das Argument in der Funktion **mle()** muss durch $-\log L$ spezifiziert werden) werden die Lösungen für die unbekannten Parameter numerisch bestimmt. Das Prinzip soll am Beispiel einer Binomialverteilung gezeigt werden, auch wenn hierfür eine analytische Lösung leicht abzuleiten ist (vgl. (6.16)).

```

> library(mle)
> x <- 16 # Beobachtung: 16mal die Sechs
> size <- 24 # Anzahl der Wuerfe (24)
> # Likelihoodfunktion - initial geschaetzt p=1/6 (regulaerer Wurfel)
> logL <- function(p=0.167) -sum(dbinom(x, size, p, log = TRUE))
> mle(logL)

Call: mle(minuslogl = logL)

Coefficients:
      p
0.6666661

```

6.3.2.2 ML-Schätzer zur negativen Binomialverteilung

Die Schätzung der Parameter einer negativen Binomialverteilung aus den Beobachtungen einer Zufallsstichprobe kann direkt nach der Momentenmethode erfolgen. Ist die Verteilung durch

$$P(n) = \binom{n+k-1}{k-1} p^k (1-p)^n$$

gegeben, dann gilt für die beiden ersten Momente

$$m_1 = \frac{kq}{p} \quad (,,\text{Mittelwert}'' \text{ und}$$

$$m_2 = \frac{kq}{p^2} \quad (,,\text{Varianz}'').$$

Aus diesen kann eine MOM-Schätzung für die beiden Parameter direkt abgeleitet werden.

$$\hat{p} = \frac{\bar{x}}{s^2} \quad \text{und} \quad \hat{k} = \frac{\bar{x}\hat{p}}{1 - \hat{p}} = \frac{\bar{x}^2}{s^2 - \bar{x}} \tag{6.17}$$

Für eine Schätzung nach dem MLE-Verfahren wird für die Stichprobe die folgende Notation verwendet:

$$\bar{x} = \frac{1}{N} \sum_{n=0}^z n \cdot r_n \quad \text{mit} \quad N = \sum_{n=0}^z r_n$$

Dabei bedeutet z die höchste für ein n beobachtete Anzahl. Die Likelihoodfunktion ist dann durch (6.18) gegeben.

$$L = \prod_{n=0}^z [P(n)]^{r_n} \quad (6.18)$$

Die partielle Ableitung von (6.18) nach dem Parameter p führt auf eine analytische Lösung, die mit der MOM-Schätzung übereinstimmt:

$$\begin{aligned} \frac{\partial L}{\partial p} &= \frac{N \cdot k}{p} - \frac{\sum n \cdot r_n}{(1-p)} \\ \hat{\mu} &= \frac{k \cdot q}{p} \end{aligned} \quad (6.19)$$

Der Mittelwert der Stichprobe $\bar{x}$ ist somit eine MLE-Schätzung für den Erwartungswert der Verteilung. Die partielle Ableitung nach dem Parameter k führt dagegen nicht eindeutig auf eine analytische Lösung. Praktisch wird eine numerische Lösung (iterativ) mit den Stichprobenwerten gesucht, wobei als Ausgangsnäherung der Wert aus der Momentenschätzung zu empfehlen ist. Alternativ kann das k auch aus dem Anteil der „0-Werte“ direkt geschätzt werden.

$$\begin{aligned} P(0) &= p^k \\ \hat{P}(0) &= \frac{r_0}{N} \\ \hat{k} &= \frac{\ln\left(\frac{r_0}{N}\right)}{\ln(\hat{p})} \end{aligned} \quad (6.20)$$

Beispiel: Die Zahl von kariös/gefüllten Zahnflächen (d_3f -Flächen) je Kind in einer Stichprobe von 467 Kindern ist in der folgenden Tabelle wiedergegeben (nach Städtler, P., Oralprophylaxe, peter.stadtler@kfunigraz.ac.at).



d_3f -Flächen	0	1	2	3	4	5	6	7	8	9	10	11	12	13
Kinder	221	32	42	27	27	13	11	9	8	14	6	5	4	7

d_3f -Flächen	14	15	16	17	18	19	20	21	22	23	24	25	>25
Kinder	6	4	4	1	1	3	3	3	3	-	1	1	11

Die Schätzung der Parameter p und k einer negativen Binomialverteilung auf der Grundlage der Beispieldaten kann im Programm **R** wie folgt durchgeführt werden:

```
> d3f <- 0:47
> n <- c(221, 32, 42, 27, 27, 13, 11, 9, 8, 14, 6, 5, 4, 7, 6, 4, 4, 1, 1,
+       3, 3, 3, 3, 0, 1, 1, 0, 1, 1, 0, 0, 1, 1, 0, 1, 1, 1, 2, 1, 0, 0,
+       0, 0, 0, 0, 0, 0, 1)
> N <- sum(n) # Momentenschätzung
> m <- sum(n*d3f)/N; m # Mittelwert
[1] 3.989293
> v <- (sum(n*(d3f^2)) - (sum(n*d3f))^2/N)/(N-1); v # Varianz
[1] 48.82607
> prob <- m/v; prob # p geschätzt
[1] 0.08170417
> size <- m^2/(v-m); size # k geschätzt
[1] 0.3549422
> library(mle)
> #### Likelihood-Funktion - initiale Werte aus der Momentenschätzung #####
> logL <- function(k=0.3, p=0.08) -sum(dnbinom(n, k, p, log=TRUE))
> summary(mle(logL))
```

Maximum likelihood estimation

```
Call: mle(minuslogl = logL)
```

```
Coefficients:
```

```
      Estimate Std. Error
k  0.2948594  0.061332931
p  0.0294216  0.009714858
```

```
-2 log L: 276.2133
```

In dem Beispiel mit **R** wird aus den Daten zunächst eine analytische Lösung $\hat{p}_0 = 0,082$ und $\hat{k}_0 = 0,355$ berechnet (Momentenschätzung). Diese Ergebnisse werden dann als initiale Werte für die numerische Lösung durch die **mle()** Funktion verwendet und führen zum Ergebnis $\hat{p} = 0,029$ und $\hat{k} = 0,295$.

6.3.2.3 ML-Schätzer zur Poisson-Verteilung

Sei $X_1, \dots, X_n$ eine Zufallsstichprobe aus einer Poisson-verteilten Grundgesamtheit (unabhängig und identisch verteilt mit dem Erwartungswert λ) mit den Realisierungen $x_i, i = 1, \dots, n$, dann können Erwartungswert und Varianz nach (6.21) geschätzt werden.

$$\begin{aligned}
 L(\lambda|x_i) &= e^{-n\lambda} \frac{\lambda^{\sum_{i=1}^n x_i}}{x_1! \cdots x_n!} \quad \text{für alle } \lambda > 0 \\
 \ln(L) &= -n\lambda + \left(\sum_{i=1}^n x_i \right) \ln \lambda - \ln(x_1! \cdots x_n!) \\
 \frac{\partial \ln L}{\partial \lambda} &= -n + \left(\sum_{i=1}^n x_i \right) \frac{1}{\lambda} = 0 \\
 \Rightarrow \hat{\lambda} &= \frac{\sum_{i=1}^n x_i}{n} = \bar{x}
 \end{aligned} \tag{6.21}$$

6.3.2.4 ML-Schätzer zur Normalverteilung

Sei $X_1, \dots, X_n$ eine Zufallsstichprobe aus einer normalverteilten Grundgesamtheit (unabhängig und identisch verteilt nach $N(\mu, \sigma)$ mit den Realisierungen $x_i, i = 1, \dots, n$), dann können Erwartungswert und Varianz nach (6.22) geschätzt werden.

$$\begin{aligned}
 L(\mu, \sigma) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right) \\
 \ln L &= -\frac{n}{2} \ln(2\pi) - \frac{N}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \\
 \frac{\partial \ln L}{\partial \mu} &= \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) = 0 \\
 \frac{\partial \ln L}{\partial \sigma} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 = 0 \\
 \Rightarrow \hat{\mu} &= \frac{1}{n} \sum_{i=1}^n x_i = \bar{x} \\
 \Rightarrow \hat{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2
 \end{aligned} \tag{6.22}$$

Die Maximum-Likelihood-Schätzung für die Varianz ist nur asymptotisch erwartungstreu. Eine erwartungstreue Schätzung erhält man durch die Multiplikation mit $\frac{n}{n-1}$. An einem Beispiel mit **R** soll die analytische Lösung mit der numerischen Lösung mittels der Funktion **mle()** verglichen werden.

```
> library(mle) # Realisierungen der Zufallsvariablen
> x <- c(23, 25, 30, 18, 17, 24, 23, 20, 19)
> #### Likelihood-Funktion – initiale Werte Erwartungswert=20 und Varianz=16
> logL <- function(m=20, s=4) -sum(dnorm(x, mean=m, sd=s, log=TRUE))
> mle(logL)

Call: mle(minuslogl = logL)

Coefficients:
      m      s
22.111221  3.842649
> mean(x); sd(x) # analytische Loesung
[1] 22.11111 [1] 4.075673
```

Die numerische Lösung durch die Funktion **mle()** in **R** ist $\hat{\mu} = 22,11$ und $\hat{\sigma} = 3,84$. Während die entsprechende analytische Lösung für μ nahezu identisch ist, zeigt sich bei σ ein deutlicher Unterschied (der MLE-Schätzer für σ ist nicht erwartungstreu).

6.3.3 Kleinster Fehler (OLS)

Die Schätzung nach der Methode der kleinsten **Fehlerquadrate** (OLS) (**Ordinary Least Squares**) geht auf Laplace und Gauß zurück. In ihr wird die Stichprobe als Summe einer Funktion des Parameters (oder der Parameter) $f(\vartheta)$ plus eines Fehlers (Rauschen) angesehen. Der Parameter wird aus der Stichprobe so geschätzt, dass dieser Fehler minimiert wird.

$$S(\vartheta) = \sum_{i=1}^n [x_i - f(\vartheta)]^2 \quad (6.23)$$

Dazu wird die Summe der Abstandsquadrate (6.23) gebildet und anschließend minimiert, indem die Ableitung nach dem Parameter gleich Null gesetzt wird, bzw. die partiellen Ableitungen nach den Parametern gleich Null gesetzt werden. Das Prinzip wird in (6.24) an einem einfachen Beispiel der Schätzung des Erwartungswertes deutlich.

$$\begin{aligned} S(\mu) &= \sum_{i=1}^n (x_i - \mu)^2 \\ &= \sum_{i=1}^n (x_i^2 - 2\mu x_i + \mu^2) \\ &= \sum_{i=1}^n x_i^2 - 2\mu \sum_{i=1}^n x_i + n\mu^2 \\ \frac{\partial S}{\partial \mu} &= 2 \sum_{i=1}^n x_i + 2n\mu = 0 \\ \Rightarrow \hat{\mu} &= \frac{\sum_{i=1}^n x_i}{n} = \bar{x} \end{aligned} \quad (6.24)$$

Dieses Verfahren findet besonders bei der Schätzung von Parametern in linearen und nichtlinearen Modellen Anwendung. Sollen zum Beispiel die beiden Parameter einer linearen Regression an Hand einer Stichprobe geschätzt werden, dann lautet der Ansatz nach der OLS-Methode:

$$\begin{aligned}
S(\alpha, \beta) &= \sum_{i=1}^n [y_i - (\alpha + \beta x_i)]^2 \\
\frac{\partial S}{\partial \alpha} &= -2 \sum_{i=1}^n (y_i - \alpha - \beta x_i) = 0 \\
\frac{\partial S}{\partial \beta} &= -2 \sum_{i=1}^n (y_i - \alpha - \beta x_i) x_i = 0 \\
\Rightarrow \hat{\beta} &= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{s_{xy}}{(s_x)^2} \\
\Rightarrow \hat{\alpha} &= \bar{y} - \hat{\beta} \bar{x}
\end{aligned} \tag{6.25}$$

Die Lösung der entstehenden Gleichungssysteme ist in der Regel nicht so einfach wie in (6.25) analytisch möglich. Allerdings stehen in den meisten Statistikprogrammen Funktionen zu numerischen Bearbeitung zur Verfügung. In **R** werden zwei Beispiele für eine einfache lineare Funktion $y_1 = f(x_1)$ und für eine exponentielle Funktion $y_2 = f(x_2)$ mit den Funktionen **lm()** (linear model) und **nls()** (nonlinear functions) gezeigt (Abbildung 6.2).

```

> x1 <- seq(0, 10, by=0.5)
> n1 <- length(x1)
> e1 <- rnorm(n1, mean=0, sd=3) # zufaellige Abweichungen (Rauschen)
> y1 <- 20 - 5*x1 + e1          # Parameter a=20 und b=5
> lm(y1 ~ x1)                  # lineares Modell

Call:
lm(formula = y1 ~ x1)

Coefficients:
(Intercept)          x1
      21.695         -5.261
#####
> x2 <- seq(0, 10, by=0.2)
> n2 <- length(x2)
> e2 <- rnorm(n2, mean=0, sd=0.5) # zufaellige Abweichungen (Rauschen)
> y2 <- 5/exp(0.5*x2) + e2        # Parameter p1=5 und p2=0.5
> nls(y2 ~ p1/exp(p2*x2), start=list(p1=1, p2=1)) # nichtlineares Modell
Nonlinear regression model
  model: y2 ~ p1/exp(p2 * x2)
 data:  parent.frame()
      p1      p2
5.4888316 0.6507575
residual sum-of-squares: 9.877245

```

6.4 Konfidenzintervalle

Der Begriff **Konfidenzintervall** oder **Vertrauensbereich** ist von J. Neyman und E.S. Pearson (vgl. Neyman [Ney50]) eingeführt worden. Man versteht darunter ein aus Stichprobenwerten berechnetes [d. h. in Lage und Breite zufälliges] Intervall, das den wahren aber unbekannten Parameter mit einer vorgegebenen Wahrscheinlichkeit, der Vertrauenswahrscheinlichkeit, überdeckt. Als Vertrauenswahrscheinlichkeit wird meist 95% gewählt; diese Wahrscheinlichkeit besagt, dass **bei häufiger berechtigter Anwendung** dieses Verfahrens die berechneten Konfidenzintervalle in etwa 95% der Fälle den Parameter überdecken und ihn in nur 5% der Fälle nicht erfassen. Für den konkreten Einzelfall gilt mit $P = 1$: „der Parameter wird überdeckt oder nicht“. Dies ist natürlich für den Praktiker wenig tröstlich. Häufig ist das Konfidenzintervall auch breiter als erwartet bzw. erfordert mehr Beobachtungen als verfügbar sind.

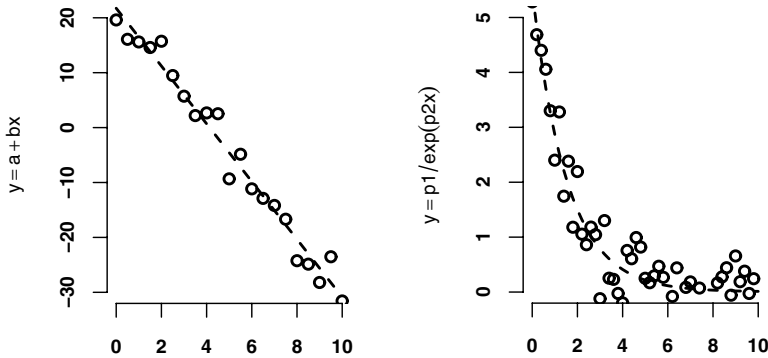


Abb. 6.2. OLS-Schätzer zur linearen Regression (links) und zur nichtlinearen Regression (rechts) mit den Funktionen `lm()` und `nls()` in R: $\hat{y} = -5,26 + 21,7x$ und $\hat{y} = 5,49 / \exp(0,65x)$

Präzisierung: Die Schätzung des Fehlers einer konkreten Schätzung $\hat{\vartheta}$ (anhand einer Stichprobe) für den wahren Parameter ϑ einer Verteilung soll dahingehend betrachtet werden, eine **Intervallschätzung** (Bereichsschätzung) für diesen Parameter zu konstruieren. Grundlage dafür kann die Standardabweichung des verwendeten Schätzers $\sigma_{\hat{\vartheta}}$ sein (vgl. auch mittlerer quadratischer Fehler), der die Unsicherheit der Schätzung wiedergibt. Es ist zu erwarten, dass der wahre Wert ϑ zwischen den beiden Werten

$$\hat{\vartheta} \pm k \cdot \sigma_{\hat{\vartheta}}$$

liegt, sofern der Wert für das $k > 0$ nur hinreichend groß gewählt wird. Dabei können grundsätzlich auch unsymmetrische Intervalle zur Eingrenzung von Interesse sein;

$$[\hat{\vartheta} - k_1 \cdot \sigma_{\hat{\vartheta}}, \quad \hat{\vartheta} + k_2 \cdot \sigma_{\hat{\vartheta}}]$$

mit „geeigneten“ Zahlen $k_1 > 0$ und $k_2 > 0$.

Allgemein versteht man unter einer Intervallschätzung für einen Parameter ϑ ein Intervall $I = [\hat{\vartheta}_u, \hat{\vartheta}_o]$, dessen Grenzen aus den Beobachtungen einer Zufallsstichprobe $x_1, x_2, \dots, x_n$ mit einer geeigneten Methode so zu berechnen sind, dass der wahre Wert mit großer Wahrscheinlichkeit in diesem Intervall liegt, z.B. mit einer Wahrscheinlichkeit von 95%. Dabei sollte das Intervall nicht zu groß sein, da es sonst wenig über den unbekannten wahren Parameter aussagt. Die Suche nach einem geeigneten Intervall verfolgt somit gegenläufige Ziele: Hohe Anforderung hinsichtlich der „**Überdeckungswahrscheinlichkeit**“ erfordert breitere Intervalle, schmalere Intervalle führen zwangsläufig zu einer kleineren Überdeckungswahrscheinlichkeit.

Eine Intervallschätzung $I = [\hat{\vartheta}_u, \hat{\vartheta}_o]$ heißt **(1 - α)-Konfidenzintervall** oder auch **(1 - α)-Vertrauensbereich**, wenn die zugehörige Überdeckungswahrscheinlichkeit mindestens $1 - \alpha$ beträgt:

$$P(\hat{\vartheta}_u \leq \vartheta \leq \hat{\vartheta}_o) \geq 1 - \alpha$$

Beispiel: Eine Schätzfunktion T für den unbekannten Parameter ϑ sei **normalverteilt**, dann bedeutet jedes **95 %-Konfidenzintervall** (95%-KI):



$$\begin{aligned} P(-1,96 \leq \frac{T - \vartheta}{\sigma_T} \leq 1,96) &= 0,95 \\ P(T - 1,96 \cdot \sigma_T \leq \vartheta \leq T + 1,96 \cdot \sigma_T) &= 0,95 = P(a \leq \vartheta \leq b) \end{aligned} \quad (6.26)$$

Zwischen den Intervallgrenzen a und b ist in 95% aller Intervalle der Parameter ϑ enthalten. Für einen 90%-KI ist der Faktor 1,96 durch 1,645 zu ersetzen, für einen 99%-KI durch 2,576 (vgl. die entsprechenden Quantile der Standardnormalverteilung).



Das Konfidenzintervall wird mit wachsendem Stichprobenumfang n enger. Bei gleichem n erhält man durch Verringerung der Vertrauenswahrscheinlichkeit (d.h. Vergrößerung von α) ebenfalls engere Bereiche, also **schärfere**, aber **weniger sichere Aussagen** über den unbekannten Parameter ϑ . Übliche Vertrauensbereiche sind: **90%-KI** (mit $\alpha = 0,10$), **95%-KI** (mit $\alpha = 0,05$) und **99%-KI** (mit $\alpha = 0,01$). Der Vorteil der Bereichsschätzung besteht in der Quantifizierung ihrer Unschärfe.

6.5 Konfidenzintervall für einen Anteilswert aus einer dichotomen Grundgesamtheit (π)

- Approximation durch die Normalverteilung
- Sonderfälle mit $\hat{p} = 0$ bzw. $\hat{p} = 1$
- Schnellschätzung nach Clopper und Pearson
- Angenähertes 95%-Konfidenzintervall für $\pi_1 - \pi_2$
- Mindestumfang einer Stichprobe bei ausgezählten Werten

Bedeutet x die Anzahl der Treffer unter n Beobachtungen der Stichprobe, dann ist $\hat{p} = x/n$ die relative Häufigkeit. Die prozentuale Häufigkeit der Treffer in der Stichprobe ist

$$\hat{p} = \frac{x}{n} \cdot 100 \quad \text{mit} \quad n \geq 100$$

Für $n < 80$ gibt man „ x von n “ oder x/n an, jedoch keine Prozentzahlen. (6.27)

und für $80 \lesssim n < 150$ ohne Kommastelle zu schreiben [für $80 \leq n < 100$ schreibt man, falls aus Vergleichsgründen erwünscht, z. B. $29/80 = 0,3625$ als „(36%)“], erst ab etwa $n = 2000$ mit zwei Stellen nach dem Komma. Beispiel: $\hat{p} = 33/189 = 0,17460$ wird als Prozentzahl 17,5% angegeben.

Beachtet sei der Unterschied zwischen „Prozente“ und „Prozentpunkte“, der Differenz zweier Prozente: nimmt z. B. etwas von 70% auf 77% zu, so wächst es um 10% bzw. um sieben Prozentpunkte. Steigt dagegen etwas von 143% auf 144% an, d.h. akkurat um einen Prozentpunkt, so erhöht es sich nur um $1/143$, d.h. $100/143\% = 0,7\%$. Nur für die Zunahme von 100% auf 101% erhöht sich etwas um genau $1/100$, d.h. um $100/100\% = 1\%$. Je nachdem, welche Tendenz dem Leser suggeriert werden soll, wird mit Prozentpunkten oder mit Erhöhung um ... % „argumentiert“.

$\hat{p} = x/n$ ist eine **erwartungstreue, konsistente Schätzung für π** ; beachte $x/n \rightarrow \pi$ sobald $n \rightarrow \infty$ (Gesetz der großen Zahlen). Exakte zweiseitige Grenzen, untere und obere Vertrauensgrenzen ($\pi_u; \pi_o$), für das Konfidenzintervall (KI) des Parameters π lassen sich nach (6.28) berechnen. Dabei werden die $\alpha/2$ -Quantile der Fisher-Verteilung (F) mit der entsprechenden Anzahl an Freiheitsgraden verwendet.

$$\begin{aligned} \pi_o &= \frac{(x+1)F}{n-x+(x+1)F} \quad \text{mit} \quad F_{\{FG_1=2(x+1), FG_2=2(n-x), \alpha/2\}} \\ \pi_u &= \frac{x}{x+(n-x+1)F} \quad \text{mit} \quad F_{\{FG_1=2(n-x+1), FG_2=2x, \alpha/2\}} \end{aligned} \quad (6.28)$$

Ausgewählte 95%-Konfidenzintervalle sind in Tabelle 6.2 zusammengestellt. Für $x/n > 0,5$ lese man das 95%-KI für $(1-x/n)$ ab und subtrahiere beide Grenzen von 100; z. B. $\hat{p} = x/n = 20/25$,

für $(1 - 20/25) = (5/25)$ liest man 6,83 bis 40,70 ab und erhält das 95%-KI: $100 - 40,70 = 59,30$ bis $93,17 = 100 - 6,83$, d. h. $0,593 \leq \pi \leq 0,932$.

Tabelle 6.2. Ausgewählte 95%-Konfidenzintervalle für π : n = Stichprobenumfang, x = Trefferzahl; z. B. $\hat{p} = x/n = 10/300$ oder 3,3%, 95%-VB: $1,60\% \leq \pi \leq 6,07\%$ oder besser $0,016 \leq \pi \leq 0,061$

x	n : 25	50	75	100	200	300	400	500	1000	x
0	0,00–13,72	0,00– 7,11	0,00– 4,80	0,00– 3,62	0,00– 1,83	0,00– 1,22	0,00– 0,92	0,00– 0,74	0,00– 0,37	0
1	0,10–20,35	0,05–10,65	0,03– 7,21	0,03– 5,45	0,01– 2,75	0,01– 1,84	0,01– 1,38	0,01– 1,11	0,00– 0,56	1
2	0,98–26,03	0,49–13,71	0,32– 9,30	0,24– 7,04	0,12– 3,57	0,08– 2,39	0,06– 1,79	0,05– 1,44	0,02– 0,72	2
3	2,55–31,22	1,25–16,55	0,83–11,25	0,62– 8,52	0,31– 4,32	0,21– 2,89	0,16– 2,18	0,12– 1,74	0,06– 0,87	3
4	4,54–36,08	2,22–19,23	1,47–13,10	1,10– 9,93	0,55– 5,04	0,36– 3,38	0,27– 2,54	0,22– 2,04	0,11– 1,02	4
5	6,83–40,70	3,33–21,81	2,20–14,88	1,64–11,28	0,80– 5,78	0,53– 3,88	0,40– 2,92	0,32– 2,34	0,16– 1,17	5
6	9,36–45,13	4,53–24,31	2,99–16,60	2,23–12,60	1,09– 6,46	0,73– 4,33	0,54– 3,26	0,43– 2,61	0,22– 1,31	6
7	12,07–49,39	5,82–26,74	3,84–18,29	2,86–13,89	1,40– 7,12	0,93– 4,77	0,70– 3,59	0,56– 2,88	0,28– 1,44	7
8	14,95–53,50	7,17–29,11	4,72–19,94	3,52–15,16	1,73– 7,76	1,15– 5,21	0,86– 3,92	0,69– 3,14	0,34– 1,58	8
9	17,97–57,48	8,58–31,44	5,64–21,56	4,20–16,40	2,07– 8,40	1,37– 5,64	1,03– 4,25	0,82– 3,40	0,41– 1,71	9
10	21,13–61,33	10,03–33,72	6,58–23,16	4,90–17,62	2,41– 9,03	1,60– 6,07	1,20– 4,57	0,96– 3,66	0,48– 1,84	10
11	24,40–65,07	11,53–35,96	7,56–24,73	5,62–18,83	2,77– 9,66	1,84– 6,49	1,37– 4,88	1,10– 3,92	0,55– 1,97	11
12	27,80–68,69	13,06–38,17	8,55–26,28	6,36–20,02	3,13–10,28	2,08– 6,90	1,55– 5,20	1,24– 4,17	0,62– 2,09	12
13	31,31–72,20	14,63–40,34	9,57–27,81	7,11–21,20	3,50–10,89	2,32– 7,32	1,74– 5,51	1,39– 4,42	0,69– 2,22	13
14	34,93–75,60	16,23–42,49	10,60–29,33	7,87–22,37	3,88–11,49	2,57– 7,73	1,92– 5,82	1,54– 4,67	0,77– 2,34	14
15	38,67–78,87	17,86–44,61	11,65–30,83	8,65–23,53	4,26–12,09	2,82– 8,13	2,11– 6,12	1,69– 4,91	0,84– 2,47	15
16	42,52–82,03	19,52–46,70	12,71–32,32	9,43–24,68	4,64–12,69	3,08– 8,53	2,30– 6,43	1,84– 5,16	0,92– 2,59	16
17	46,50–85,05	21,21–48,77	13,79–33,79	10,23–25,82	5,03–13,29	3,33– 8,94	2,49– 6,73	1,99– 5,40	0,99– 2,71	17
18	50,61–87,93	22,92–50,81	14,89–35,25	11,03–26,95	5,42–13,88	3,59– 9,33	2,69– 7,03	2,14– 5,64	1,07– 2,84	18
19	54,87–90,64	24,65–52,83	15,99–36,70	11,84–28,07	5,82–14,46	3,85– 9,73	2,88– 7,33	2,30– 5,88	1,15– 2,96	19
20	59,30–93,17	26,41–54,82	17,11–38,14	12,67–29,18	6,22–15,04	4,12–10,12	3,08– 7,63	2,46– 6,12	1,22– 3,08	20
21	63,92–95,46	28,19–56,79	18,24–39,56	13,49–30,29	6,62–15,62	4,38–10,52	3,28– 7,93	2,62– 6,36	1,30– 3,20	21
22	68,78–97,45	29,99–58,75	19,38–40,98	14,33–31,39	7,03–16,20	4,65–10,91	3,48– 8,22	2,78– 6,60	1,38– 3,32	22
23	73,97–99,02	31,81–60,68	20,53–42,38	15,17–32,49	7,44–16,78	4,92–11,30	3,68– 8,51	2,94– 6,83	1,46– 3,44	23
24	79,65–99,90	33,66–62,58	21,69–43,78	16,02–33,57	7,85–17,35	5,19–11,68	3,88– 8,81	3,10– 7,07	1,54– 3,55	24
25	86,28–100	35,53–64,47	22,86–45,17	16,88–34,66	8,26–17,92	5,47–12,07	4,08– 9,10	3,26– 7,30	1,62– 3,67	25

Beispiel: Berechne das 95%-KI für π aufgrund von $\hat{p} = x/n = 7/20 = 0,35$ (F -Werte als 0.025-Quantile der Fisher-Verteilung aus der Funktion **qf()** in **R** oder aus einer entsprechenden Tabelle).

F -Werte: $2(7+1) = 16$; $2(20-7) = 26$; $F_{16;26;0,025} = 2,36$
 $2(20-7+1) = 28$; $2 \cdot 7 = 14$; $F_{28;14;0,025} = 2,75$

KI-Grenzen: $\pi_u = \frac{7}{7+(20-7+1)2,75} = 0,154$ und $\pi_o = \frac{(7+1)2,36}{20-7+(7+1)2,36} = 0,592$
 95%-KI: $0,154 \leq \pi \leq 0,592$, d. h. $0,15 \leq \pi \leq 0,60$.

Anhand von $F_{16;26;0,05} = 2,05$ und $F_{28;14;0,05} = 2,32$ (vgl. interpolierte Werte aus Tab. 5.12) lässt sich auch das entsprechende 90%-KI: $0,177 \leq \pi \leq 0,558$ als $0,18 \leq \pi \leq 0,56$ angeben.

```
> x <- 7; n <- 20; p <- x / n
> Fu <- qf(0.975, 2*(n-x+1), 2*x); Fu
[1] 2.748669
> Fo <- qf(0.975, 2*(x+1), 2*(n-x)); Fo
[1] 2.359684
> piu <- x/(x+(n-x+1)*Fu); piu
[1] 0.1539092
> pio <- (x+1)*Fo/(n-x+(x+1)*Fo); pio
[1] 0.5921885
```

Hinweise:

- (1) Vorausgesetzt wird, dass $\hat{p} = x/n$ anhand einer **Zufallsstichprobe** geschätzt worden ist.
- (2) Nur für $\hat{p} = 0,5$ erhält man **symmetrisch** liegende Vertrauensgrenzen (vgl. obiges Beispiel: $0,592 - 0,350 = 0,242 > 0,196 = 0,350 - 0,154$).



6.5.1 Approximation durch die Normalverteilung

Eine gute Approximation für $0,3 \leq \pi \leq 0,7$ mit $n \geq 10$ und $0,05 \leq \pi \leq 0,95$ mit $n \geq 60$ ist, als 95%-Konfidenzintervall geschrieben, (6.29) [mit dem 0,95-Quantil der Standardnormalverteilung $z_{0,95} = 1,96$; $1,95 = (1,96^2 + 2)/3$ sowie $0,18 = (7 - 1,96^2)/18$] (Molenaar [Mol70]).

$$\begin{aligned}\pi_o &= [x+1,95+1,96\sqrt{(x+1-0,18)(n-x-0,18)/(n+11 \cdot 0,18-4)}]/(n+2 \cdot 1,95-1) \\ \pi_u &= [x-1+1,95-1,96\sqrt{(x-0,18)(n+1-x-0,18)/(n+11 \cdot 0,18-4)}]/(n+2 \cdot 1,95-1)\end{aligned}\quad (6.29)$$



Beispiel: 95%-KI für π aufgrund von $\hat{p} = x/n = 7/20$.

$$\begin{aligned}\pi_o &= [7+1,95+1,96\sqrt{(7+1-0,18)(20-7-0,18)/(20+11 \cdot 0,18-4)}]/(20+2 \cdot 1,95-1) \\ \pi_u &= [7-1+1,95-1,96\sqrt{(7-0,18)(20+1-7-0,18)/(20+11 \cdot 0,18-4)}]/ \\ &\quad (20+2 \cdot 1,95-1)\end{aligned}$$

95%-KI: $0,151 \lesssim \pi \lesssim 0,593$, d.h. $0,15 \lesssim \pi \lesssim 0,59$.

Das 90%-KI nach (6.29) mit „1,96“ durch 1,645 ersetzt (entsprechend auch „1,95“ durch 1,57 und „0,18“ durch 0,24), erhält man für unser Beispiel ($\hat{p} = x/n = 7/20$) als 90%-KI: $0,176 \lesssim \pi \lesssim 0,555$ bzw. 90%-KI: $0,18 \lesssim \pi \lesssim 0,56$.

Für nicht zu kleine Stichprobenumfänge n und nicht zu extreme relative Häufigkeiten $\hat{p}$; d. h. für $n\hat{p} > 5$ und $n(1 - \hat{p}) > 5$ kann zur groben Übersicht (6.30) benutzt werden [Stichprobenziehen mit Zurücklegen, unendliche Grundgesamtheit, vgl. auch (6.31)].

$$\begin{aligned}\pi_o &\approx \left(\hat{p} + \frac{1}{2n}\right) + z \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \\ \pi_u &\approx \left(\hat{p} - \frac{1}{2n}\right) - z \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\end{aligned}\quad (6.30)$$

Diese Approximation dient zur Groborientierung; sind die Bedingungen von Tab. 6.3 erfüllt, dann ist sie zwar schlechter als (6.29), aber noch brauchbar; deutlich besser ist

$$\frac{1}{n+z^2} \left(x + \frac{z^2}{2} - z \cdot \sqrt{x(1-\hat{p}) + \frac{z^2}{4}} \right) \lesssim \pi \lesssim \frac{1}{n+z^2} \left(x + \frac{z^2}{2} + z \cdot \sqrt{x(1-\hat{p}) + \frac{z^2}{4}} \right) \quad (6.31)$$

Mit $z = 1,96$ für das 95%-KI bzw. $z = 1,645$ für das 90%-KI.



Beispiel: 95%-KI für π aufgrund von $\hat{p} = x/n = 70/200 = 0,35$.

Mit $z = 1,96$ erhält man das angenäherte 95%-KI für π über

$$\begin{aligned}\frac{1}{200 + 1,96^2} \left(70 + \frac{1,96^2}{2} - 1,96 \sqrt{70(1 - 0,35) + \frac{1,96^2}{4}} \right) &= 0,287 \\ \frac{1}{200 + 1,96^2} \left(70 + \frac{1,96^2}{2} + 1,96 \sqrt{70(1 - 0,35) + \frac{1,96^2}{4}} \right) &= 0,418\end{aligned}$$

als 95%-KI: $0,287 \lesssim \pi \lesssim 0,418$, nach „aussen“ gerundet $0,28 \lesssim \pi \lesssim 0,42$; die exakten Grenzen sind 0,2844 und 0,4206.

In **R** erfolgt die Berechnung von Konfidenzintervallen in zahlreichen Funktionen, insbesondere für Parameter im Rahmen der statistischen Modellbildung und ergänzend in den statistischen Testverfahren. In **R** wird mit der Funktion **binom.test()** das Konfidenzintervall nach Clopper und Pearson [CP34] bestimmt, welches das vorgegebene Konfidenzniveau in jedem Fall einhält, allgemein jedoch nicht das kürzeste Konfidenzintervall darstellt.

```
> binom.test(70, 200, p=0.4, conf.level = 0.95)

Exact binomial test

data: 70 and 200
number of successes = 70, number of trials = 200, p-value = 0.1701
....
95 percent confidence interval: 0.2840707 0.4204506
```

Hinweise:

1. Die Größe $1/2n$ in (6.30) wird als **Kontinuitätskorrektur** bezeichnet. Sie weitet das Konfidenzintervall! Die Ausgangswerte sind Häufigkeiten, also diskrete Variable; für das Konfidenzintervall benutzen wir die Standardnormalvariable, eine stetige Zufallsvariable. Der Fehler, den wir durch den Übergang von der diskreten auf die Normalverteilung machen, wird durch die Kontinuitätskorrektur verringert.

2. Für **endliche Grundgesamtheiten** des Umfangs N kann man zur Groborientierung (6.32) benutzen; $\sqrt{(N-n)/(N-1)}$ ist eine Größe, die für $N \rightarrow \infty$ den Wert 1 erreicht (vgl. $\sqrt{1-n/N}/(1-1/N) \rightarrow \sqrt{1} = 1$) und dann vernachlässigt werden darf (vgl. z. B. (6.30)). Das gilt auch für den Fall, wenn N gegenüber n ausreichend groß ist, d. h. wenn z. B. n kleiner als 5% von N ist. Die Approximation (6.32) darf nur angewandt werden, wenn die in Tabelle (6.3) angegebenen Voraussetzungen (Cochran [Coc63]) erfüllt sind.

Für endliche Grundgesamtheit, d. h. **Stichprobenziehen ohne Zurücklegen**, gilt:

$$\begin{aligned}\pi_o &\approx \left(\hat{p} + \frac{1}{2n}\right) + z \cdot \sqrt{\left\{\frac{\hat{p}(1-\hat{p})}{n}\right\} \left\{\frac{N-n}{N-1}\right\}} \\ \pi_u &\approx \left(\hat{p} - \frac{1}{2n}\right) - z \cdot \sqrt{\left\{\frac{\hat{p}(1-\hat{p})}{n}\right\} \left\{\frac{N-n}{N-1}\right\}}\end{aligned}\quad (6.32)$$

Tabelle 6.3. (Aus W.G. Cochran, Sampling Techniques, 2nd edition, J. Wiley, New York, 1963, p. 57, table 3.3)

Für $\hat{p}$ gleich	und sowohl $n\hat{p}$ als auch $n(1-\hat{p})$ mindestens gleich	bei n gleich oder größer als
0,5	15	30
0,4 oder 0,6	20	50
0,3 oder 0,7	24	80
0,2 oder 0,8	40	200
0,1 oder 0,9	60	600
0,05 oder 0,95	70	1400
darf (6.32) angewendet werden		

3. Wird eine relative Häufigkeit auf rund 70% geschätzt, d. h. man ist fast sicher, dass sie zwischen 60% und 80% liegt, so heißt dies, dass das 95%-KI für π , d. h. ($z = 1,96$ oder 2) mit

$$\begin{aligned}\pm 2 \cdot \sqrt{0,3 \cdot 0,7/n} &= \pm 0,1 \\ 2 \cdot \sqrt{0,21/n} &= 0,1 | \cdot 10 \quad (\text{d.h. mit 10 multipliziert}) \\ 20 \cdot \sqrt{0,21/n} &= 1 \\ 400 \cdot 0,21 &= 84 = n,\end{aligned}$$

auf, sagen wir 85, Beobachtungen basieren dürfte mit $\hat{p} = 60/85 = 0,7$. Nach den Ciba-Geigy Tabellen [CG80] lautet das entsprechende 95%-KI: $0,60 \leq \pi \leq 0,80$.

Wenn in einer Stichprobe des Umfangs n aus einer Grundgesamtheit des Umfangs N genau Null Stichprobenelemente mit einer bestimmten Eigenschaft gefunden werden und wir mit einer Vertrauenswahrscheinlichkeit von sagen wir 95% auf die Grundgesamtheit schließen wollen, dann sollte hierfür der Auswahlssatz n/N mindestens 0,95 betragen (Wright [Wri90]).

6.5.2 Sonderfälle mit $\hat{p} = 0$ bzw. $\hat{p} = 1$

Die einseitige **obere Vertrauensgrenze** (VG) für $\hat{p} = 0$ (Nullergebnis, vgl. Tab. 6.4) erhält man nach

$$\pi_o = \frac{F}{n + F} \quad \text{mit} \quad F_{(FG1=2; FG2=2n)} \quad (6.33)$$



Beispiel: Berechne die einseitige obere 95%-Vertrauensgrenze π_o aufgrund von $\hat{p} = 0$ für $n = 60$. Mit $F_{2;120;0,95} = 3,07$ erhält man

$$95\text{-VG: } \pi_o = \frac{3,07}{60 + 3,07} = 0,0487 \quad [\text{d. h. } \pi \leq 0,049]$$

Die einseitige **untere Vertrauensgrenze** für $\hat{p} = 1$ (Vollergebnis, vgl. Tab. 6.4) ist durch (6.34) gegeben.

$$\pi_u = \frac{n}{n + F} \quad \text{mit} \quad F_{FG1=2; FG2=2n)} \quad (6.34)$$



Beispiel: Berechne die einseitige untere 99%-Vertrauensgrenze π_u aufgrund von $\hat{p} = 1$ für $n = 60$. Mit $F_{2;120;0,01} = 4,79$ erhält man

$$99\text{-VG: } \pi_u = \frac{60}{60 + 4,79} = 0,9261 \quad [\text{d. h. } \pi \geq 0,93]$$

Für die **einseitigen 95%-Vertrauensgrenzen** mit $n > 50$ und

$$\begin{aligned} \hat{p} = 0 & \quad \text{gilt näherungsweise} \quad \pi_o \approx \frac{3}{n} \\ \hat{p} = 1 & \quad \text{gilt näherungsweise} \quad \pi_u \approx 1 - \frac{3}{n} \end{aligned} \quad (6.35)$$

$$\hat{p} = 0, n = 100; 95\text{-VG: } \pi_o \approx 3/100 = 0,03$$

$$\hat{p} = 1, n = 100; 95\text{-VG: } \pi_u \approx 1 - (3/100) = 0,97$$

Zum Vergleich: $F_{2;200;0,05} = 3,04$ und damit nach (6.33, 6.34)

$$\hat{p} = 0; 95\text{-VG: } \pi_o = 3,04/(100 + 3,04) = 0,0295 = 0,03$$

$$\hat{p} = 1; 95\text{-VG: } \pi_u = 100/(100 + 3,04) = 0,9705 = 0,97.$$



Beispiel: Wenn bei 100 Patienten, die mit einem bestimmten Medikament behandelt worden sind, keine unerwünschten Nebenwirkungen auftreten, dann ist mit höchstens 3% unerwünschten Nebenwirkungen zu rechnen ($\alpha = 0,05$) (vgl. auch Tab. 6.4).

Wenn bei n Patienten **keine Nebenwirkungen** beobachtet werden, dann ist es:

- (1) unwahrscheinlich ($P \approx 0,05$), dass Nebenwirkungen bei mehr als $(3/n)$ 100% auftreten;
- (2) treten doch Nebenwirkungen auf, so ist deren Wahrscheinlichkeit fast sicher ($P \approx 0,99$) nicht größer als $(5/n)$ 100% .

Tabelle 6.4. Einseitige untere und obere 90%-, 95%- und 99%-Vertrauensgrenzen ($\alpha = 0,10$; $\alpha = 0,05$; $\alpha = 0,01$) in % für ausgewählte Stichprobenumfänge n und Nullergebnis bzw. Vollergebnis

α	n	10	30	50	80	100	150	200	300	500	1000
10%	π_0	21	7,4	4,5	2,8	2,3	1,5	1,1	0,77	0,46	0,23
	π_u	79	92,6	95,5	97,2	97,7	98,5	98,9	99,23	99,54	99,77
5%	π_0	26	9,5	5,8	3,7	3,0	2,0	1,5	0,99	0,60	0,30
	π_u	74	90,5	94,2	96,3	97,0	98,0	98,5	99,01	99,40	99,70
1%	π_0	37	14	8,8	5,6	4,5	3,0	2,3	1,5	0,92	0,46
	π_u	63	86	91,2	94,4	95,5	97,0	97,7	98,5	99,08	99,54

Für $n = 100$ Fälle ohne Nebenwirkungen ist somit mit höchstens 3% Nebenwirkungen in der Grundgesamtheit aller Fälle zu rechnen ($P \approx 0,95$). Will man eine schärfere Aussage ($P \approx 0,99$), so steigt der Anteil von 3% auf maximal 5%. Um den Anteil von Nebenwirkungen in der Grundgesamtheit aller Fälle fast sicher ($P \approx 0,99$) auf weniger als 0,5% anzusetzen, müssten demnach 1000 Fälle ohne Nebenwirkungen vorliegen.

Beispiel für $\alpha = 5\%$ und $n = 100$: sind an 100 Objekten einer Zufallsstichprobe keine Fehler (Nullergebnis) festgestellt worden, so ist für den betrachteten Zeitraum in der Grundgesamtheit aller Objekte dieses Types mit höchstens 3% Fehlern (obere 95%-Vertrauensgrenze) zu rechnen, d. h. mindestens 97% aller Objekte (untere 95%-Vertrauensgrenze für das Vollergebnis) sind als fehlerfrei aufzufassen. Dies setzt voraus, dass das Null-Fehler-Resultat exakt bestimmbar ist, d. h. eine Sensitivität und Spezifität von 100% aufweist. Sinken diese Werte, so ist die Drei durch größere Werte zu ersetzen, wobei der Effekt der Sensitivität deutlich größer ist als der der Spezifität: einige gerundete Werte nach Reibnegger und Mitarbeitern ([RHW89]). Entsprechende Ausweitungen gelten dann auch für die anderen Vertrauensgrenzen der Tabelle 6.4, deren Zahlen im gestrichelten Rechteck ($n < 80$) als relative Häufigkeiten zu interpretieren sind: z.B. ist für $n = 30$ und $\hat{p} = 0$ die 95%-VG für π_0 gleich 0.095.

Sensitivität (%)	Spezifität (%)			
	100	90	80	70
100				
90	3		4	
80			5	6
70			6	8

Zur klinischen Prüfung von Arzneimitteln auf Nebenwirkungen

Für die präklinische Prüfung (Tierversuch) gilt, dass Wahrscheinlichkeitsaussagen vom Tier auf den Menschen nicht möglich sind. Unerwünschte Nebenwirkungen gehören zum Wirkungsspektrum einer Substanz. Ihre Unerwünschtheit ist ein subjektiver Maßstab. Der Verdacht, dass eine Substanz beim Menschen schädliche Nebenwirkungen verursacht, lässt sich ohne kontrollierten Versuch mit Zufallszuteilung weder bestätigen noch widerlegen; die Unschädlichkeit lässt sich nicht „beweisen“. Die Problematik der Entscheidung von zufälligen Zusammenhängen, Assoziationen durch Drittvariable und möglichen Kausalzusammenhängen spielt eine wichtige Rolle. Allen Aussagen haftet eine erhebliche Unsicherheit an, die nur aufgrund von Plausibilitätsüberlegungen eingengt werden kann.

6.5.3 Schnellschätzung der Vertrauensgrenzen anhand einer beobachteten relativen Häufigkeit nach Clopper und Pearson

Eine schnelle Methode, aus dem Anteil oder dem Prozentsatz in der Stichprobe auf den Parameter in der Grundgesamtheit zu schließen (indirekter Schluss), bietet Abbildung 6.3 von Clopper und Pearson [CP34]. Diese Zeichnung gibt die Vertrauensgrenzen für π anhand von $\hat{p} = x/n$ mit einer

Vertrauenswahrscheinlichkeit von 95%, d. h. das 95%- KI für π . Die Zahlen an den Kurven bezeichnen den Stichprobenumfang. Die Vertrauensgrenzen werden mit zunehmendem Stichprobenumfang n enger und symmetrischer, da die Binomialverteilung in eine Normalverteilung übergeht, für $\hat{p} = 0,5$ ist der Vertrauensbereich auch bei kleinen n -Werten symmetrisch. Aus der Abbildung lässt sich auch das zur Erreichung einer bestimmten Genauigkeit notwendige n abschätzen. Für den praktischen Gebrauch bevorzuge man die exakten Werte aus Tabelle 6.2 oder man berechne die Grenzen mit **R**.

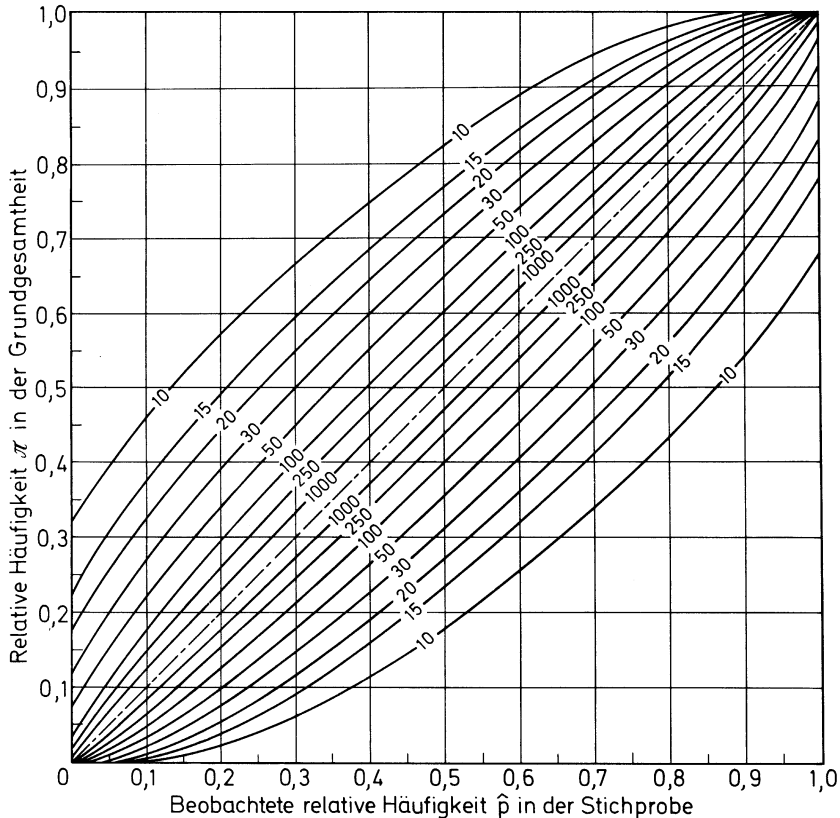


Abb. 6.3. 95%-Kle für π anhand relativer Häufigkeiten $\hat{p}$. Die Zahlen an den Kurven bezeichnen den Stichprobenumfang n (aus C.J. Clopper and E.S. Pearson: The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* **26** (1934) 404–413, p. 410)



Beispiel 1: In einer Stichprobe von $n = 10$ Werten sei das Ereignis x 7mal beobachtet worden, d. h. $\hat{p} = \frac{x}{n} = \frac{7}{10} = 0,7$. Abbildung 6.3: Die Schnittpunkte der Vertikalen über 0,7 mit der unteren und der oberen Kurve $n = 10$ bestimmen dann die Grenzen des 95%-Konfidenzintervalls für den Parameter π der Grundgesamtheit. Erst ein außerhalb dieses Intervalls $0,34 \lesssim \pi \lesssim 0,93$ auftretendes $\hat{p}$ würde (mit einer Vertrauenswahrscheinlichkeit von 95%) auf eine Stichprobe hinweisen, die einer anderen Grundgesamtheit entstammt (direkter Schluss von dem Parameter der Grundgesamtheit auf den Variationsbereich der Stichproben-Kennzahl).



Beispiel 2: Ein in der Nähe von 40% liegender Prozentsatz soll so geschätzt werden, dass der resultierende 95%-KI einen 20%-Bereich bildet. Nach Abbildung 6.3 ist diese Bedingung bei etwa $n \approx 100$ erfüllt.

6.5.4 Angenähertes 95%-Konfidenzintervall für $\pi_1 - \pi_2$ (n_1 und n_2 groß)

Für $n_1 \gtrapprox 50$ und $n_2 \gtrapprox 50$ mit $\hat{p}_1 > \hat{p}_2$ lässt sich mit $\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}$ ein angenähertes 95%-KI für $\pi_1 - \pi_2$ angeben:

$$\boxed{\hat{p}_1 - \hat{p}_2 - 1,96\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}} \lesssim \pi_1 - \pi_2 \lesssim \hat{p}_1 - \hat{p}_2 + 1,96\sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}} \quad (6.36)$$

(6.36) lässt sich verbessern, indem für $\hat{p}_1 - \hat{p}_2 < 0$ zu dieser Differenz der Wert $0,5[(1/n_1) + (1/n_2)]$ addiert (bzw. für $\dots > 0, \dots$ subtrahiert) wird.

Vorausgesetzt werden (wie üblich) zwei unabhängige Zufallsstichproben. Für das entsprechende 90%-KI wird 1,96 (wie üblich) durch 1,645 ersetzt.

Beispiel: $\hat{p}_1 = 140/200 = 0,7$ und $\hat{p}_2 = 150/250 = 0,6$; $\hat{p}_1 - \hat{p}_2 = 0,1$;

$$\sqrt{\frac{0,7 \cdot 0,3}{200} + \frac{0,6 \cdot 0,4}{250}} = 0,0448 ; \quad 1,96 \cdot 0,0448 = 0,0878$$

$$\begin{aligned} 0,5[(1/200) + (1/250)] &= 0,0045 & 0,0955 \pm 0,0878 \\ 0,7 - 0,6 - 0,0045 &= 0,0955 ; & 95\text{-KI: } 0,008 \lesssim \pi_1 - \pi_2 \lesssim 0,183 . \end{aligned}$$

Da die Null ausgeschlossen ist, besteht zwischen den Wahrscheinlichkeiten π_1 und π_2 auf dem 5%-Niveau ein statistisch gerade noch erfassbarer signifikanter Unterschied.

6.5.4.1 95%-Konfidenzintervalle für die prozentuale Zunahme eines kleinen Anteils

Zufällige oder nicht zufällige Zunahme?

Liegen zwei Stichproben der Umfänge n_1 und n_2 (jeweils > 120) mit den kleinen prozentualen Anteilen $\hat{p}_1$ und $\hat{p}_2$ vor, $\hat{p}_1 = 100(k_1/n_1)\%$, $\hat{p}_2 = 100(k_2/n_2)\%$ mit $\hat{p}_2 > \hat{p}_1$, $\hat{p}_1 < (1 - \hat{p}_1)$ und $\hat{p}_2 < (1 - \hat{p}_2)$, dann lassen sich eine prozentuale Zunahme $\hat{h}$

$$\boxed{\hat{h} = 100(\hat{p}_2 - \hat{p}_1)/\hat{p}_1 [\%]}$$

und nach Bross [Bro54] der Vertrauensbereich (VB) für h angeben.

Hierzu bilden wir die relative Häufigkeit $k_1/(k_1 + k_2)$ und den zugehörigen 95%-KI mit der unteren Grenze G_1 und der oberen Grenze G_2 . Dann ist der 95%-KI für h :

$$\boxed{\begin{aligned} 100 \frac{n_1 - (n_1 + n_2)G_2}{n_2 G_2} &\leq h \leq 100 \frac{n_1 - (n_1 + n_2)G_1}{n_2 G_1} \\ \text{bzw. für } n_1 &= n_2 \\ 100 \frac{1 - 2G_2}{G_2} &\leq h \leq 100 \frac{1 - 2G_1}{G_1} \end{aligned}} \quad (6.37)$$

Umschließt dieses 95%-KI den Wert „0%“, dann kann eine auf dem 5%-Niveau statistisch signifikante prozentuale Zunahme ausgeschlossen werden.

Beispiel:

$$\begin{aligned} \hat{p}_1 &= 100 \frac{14}{200} = 7\% & \hat{h} &= 100(14 - 7)/7 = 100\% \\ \hat{p}_2 &= 100 \frac{28}{200} = 14\% \end{aligned}$$

$$k_1/(k_1 + k_2) = 14/(14 + 28) = 14/42 (= 0,333)$$

Der entsprechende 95%-KI reicht von $G_1 = 0,196$ bis $G_2 = 0,496$; der 95%-KI für h ist dann:

$$100 \frac{1 - 2 \cdot 0,496}{0,496} = 1,61\% \leq h \leq 310,2\% = 100 \frac{1 - 2 \cdot 0,196}{0,196}$$

d. h. $0,02 \leq h/100 \leq 3,10$
und damit oberhalb von Null.

Somit liegt eine auf dem 5%-Niveau statistisch gerade noch erfassbare signifikante prozentuale Zunahme vor.

6.5.5 Schätzung des Mindestumfangs einer Stichprobe bei ausgezählten Werten

Aus der Formel für die Vertrauensgrenzen einer prozentualen Häufigkeit

$\hat{p} \pm z \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ folgt, wenn $z \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ der Abweichung a gleichgesetzt wird

$$\sqrt{n} = \frac{z \sqrt{\hat{p}(1-\hat{p})}}{a} \text{ und } n = \frac{z^2 \hat{p}(1-\hat{p})}{a^2}$$

Setzt man $z = 2$ (d. h. $\alpha = 0,0456$ oder $1 - \alpha \approx 0,9544$ oder $> 0,95$), dann wird der geschätzte Mindestumfang

$$\hat{n} = \frac{4 \cdot \hat{p}(1-\hat{p})}{a^2} \quad (6.38)$$

Da n sein Maximum erreicht, wenn $\hat{p}(1-\hat{p})$ am größten ist – dies ist für $\hat{p} = 50\%$ der Fall –, wird, wenn wir $\hat{p} = 50\%$ setzen, der Stichprobenumfang größer, als im allgemeinen notwendig ist und

$$\hat{n} = \frac{4 \cdot 0,5^2}{a^2} \quad \boxed{\hat{n} = \frac{1}{a^2}} \quad (6.39)$$

Hätten wir das Konfidenzintervall der prozentualen Häufigkeit mit der vereinfachten Formel für die Endlichkeitskorrektur

$$\sqrt{\frac{N-n}{n}} \text{ anstatt } \sqrt{\frac{N-n}{N-1}} \text{ geschrieben, } \hat{p} \pm z \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \sqrt{\frac{N-n}{N}}$$

dann erhielten wir für den geschätzten Mindestumfang

$$\hat{n} = \frac{N}{1 + a^2 N} \quad (6.40)$$



Beispiel 1: Angenommen, wir interessieren uns für den Prozentsatz von Familien eines genau lokalisierten ländlichen Gebietes, die ein bestimmtes Fernsehprogramm sehen. Es wohnen dort etwa 1000 Familien. Alle Familien zu befragen erscheint zu umständlich. Die Untersucher entschließen sich, eine Stichprobe zu ziehen und mit einer Abweichung a von $\pm 10\%$ und einer Vertrauenswahrscheinlichkeit von 95% zu schätzen. Wie groß muss die Stichprobe sein? Nach (6.40) erhalten wir

$$\hat{n} = \frac{1000}{1 + 0,10^2 \cdot 1000} \simeq 91$$

Damit sind nur 91 Familien zu befragen. Man erhält eine Schätzung von π mit einem Fehler von $a = 0,10$ und einer Vertrauenswahrscheinlichkeit von 95%. Nach (6.39) hätten wir ganz grob $n = \frac{1}{0,10^2} = \frac{1}{0,01} = 100$ erhalten. Erwarteten wir, dass $\pi = 0,30$ beträgt, dann ist unser geschätzter

Stichprobenumfang natürlich zu hoch, wir benötigen dann nur etwa $n' = 4n \cdot \pi(1 - \pi) = 4 \cdot 91 \cdot 0,3 \cdot 0,7 = 76$ Einzelwerte.

$$\hat{n}' = 4n\hat{p}(1 - \hat{p}) \quad (6.41)$$

Für $\hat{n} > 0,5N$ wird (6.38) ersetzt durch (6.42)

$$\hat{n}_{\text{kor.}} = \frac{N(a^2/4) + Np - Np^2}{N(a^2/4) + p - p^2} \quad (6.42)$$

d. h.
$$\hat{n}_{\text{kor.}} = \frac{1000(0,10^2/4) + 1000 \cdot 0,30 - 1000 \cdot 0,30^2}{1000(0,10^2/4) + 0,30 - 0,30^2} \approx 74$$

Bei Bedarf ist in beiden Formeln die 4 durch den entsprechenden Wert z^2 zu ersetzen: 2,6896 ($\alpha = 0,10$), 3,8416 ($\alpha = 0,05$) und 6,6564 ($\alpha = 0,01$).

Beispiel 2: Es wird nach dem Prozentsatz von Familien gefragt, die in einer kleinen Stadt von 3000 Einwohnern ein bestimmtes Fernsehprogramm gesehen haben. Gefordert wird eine Vertrauenswahrscheinlichkeit von 95% mit einer Abweichung von $\pm 3\%$.



$$\hat{n} = \frac{N}{1 + a^2 N} = \frac{3000}{1 + 0,0009 \cdot 3000} \simeq 811$$

Nach Entnahme einer Zufallsstichprobe von 811 Familien ergibt sich, daß 243 Familien dem Fernsehprogramm gefolgt waren, d. h. $\hat{p} = \frac{243}{811} \simeq 0,30$. Damit erhalten wir das 95%-Konfidenzintervall zu

$$\begin{aligned} 0,30 - 0,03 &\leq \pi \leq 0,30 + 0,03 \\ 0,27 &\leq \pi \leq 0,33. \end{aligned}$$

6.6 Konfidenzintervalle für μ bei Normalverteilung

- Vertrauenswahrscheinlichkeit und Irrtumswahrscheinlichkeit
- Konfidenzintervall für den Erwartungswert μ
- Konfidenzintervall für die Differenz $\mu_1 - \mu_2$
- Konfidenzintervall für den Erwartungswert aus Paardifferenzen
- Konfidenzintervall für das Verhältnis μ_1/μ_2
- Mindestzahl von Beobachtungen

6.6.1 Vertrauenswahrscheinlichkeit und Irrtumswahrscheinlichkeit

Bei verschiedenen Stichproben werden im allgemeinen die aus den Stichproben ermittelten Schätzwerte variieren. Daher ist der aus einer Stichprobe ermittelte Mittelwert $\bar{x}$ nur eine Schätzung des Erwartungswertes μ der Grundgesamtheit, der die Stichprobe entstammt. Zu diesem Schätzwert lässt sich nun ein Intervall angeben, das sich über die nächst kleineren und -größeren Werte erstreckt und das vermutlich auch den Parameter der Grundgesamtheit enthält. Dieses Intervall um den Schätzwert, das den Parameter mit einer vorgegebenen Wahrscheinlichkeit mit einschließen soll, heißt **Konfidenzintervall** (confidence interval). Tabelle 6.5 gibt einen Überblick über Konfidenzintervalle für den Erwartungswert μ der Grundgesamtheit:

$$\bar{X} \pm z \frac{\sigma}{\sqrt{n}} \text{ bzw. } P \left(\bar{X} - z \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z \frac{\sigma}{\sqrt{n}} \right) = P = 1 - \alpha \quad (6.43)$$

Der Wert z entspricht dem $(1 - \alpha/2)$ -Quantil der Standardnormalverteilung und ist in **R** mit der Funktion **qnorm()** zu berechnen oder der Tabelle 5.9 zu entnehmen. Sigma (σ) ist die bekannte (oder aus einer umfangreichen Stichprobe geschätzte) Standardabweichung; $\sigma/\sqrt{n}$ ist der Standardfehler des Mittelwertes.

Man irrt sich also beim sogenannten Konfidenzschluss mit der Wahrscheinlichkeit α , d. h. spricht man n -mal die Behauptung aus, der unbekannte Parameter liege im Konfidenzintervall, so hat man im Mittel αn Fehlschlüsse zu erwarten.

Tabelle 6.5. Konfidenzintervalle für den Erwartungswert μ - sichere Aussagen (α klein, P groß) sind unscharf (breites Konfidenzintervall)

Konfidenzintervall für den Erwartungswert μ einer normalverteilten Grundgesamtheit (σ bekannt)	Vertrauenswahrscheinlichkeit $P = 1 - \alpha$	Irrtumswahrscheinlichkeit α
$\bar{X} \pm z(\sigma/\sqrt{n})$	vgl. (6.43)	
$\bar{X} \pm 1,645 \frac{\sigma}{\sqrt{n}}$	90% = 0,90	10% = 0,10
$\bar{X} \pm 1,960 \frac{\sigma}{\sqrt{n}}$	95% = 0,95	5% = 0,05
$\bar{X} \pm 2,576 \frac{\sigma}{\sqrt{n}}$	99% = 0,99	1% = 0,01

Sehen wir uns Tabelle 6.5 näher an, so erkennen wir, dass P (oder α , beide ergänzen sich zu 100% oder zum Wert 1) die Unsicherheit der statistischen Aussage bestimmt. Je größer P ist, um so größer wird bei gegebener Standardabweichung und bei gegebenem Stichprobenumfang das Konfidenzintervall sein. Daraus folgt: Es besteht ein Gegensatz zwischen der Schärfe einer Aussage und der Sicherheit, die dieser Aussage zukommt: sichere Aussagen (d. h. mit hoher P) sind unscharf (d. h. weite Bereiche); scharfe Aussagen sind unsicher. Übliche Irrtumswahrscheinlichkeiten (Signifikanzniveaus) sind $\alpha = 0,05$, $\alpha = 0,01$ und $\alpha = 0,001$, je nachdem, wie schwer wiegend die Entscheidung ist, die man aufgrund der Stichprobe fällen will.

Der Schluss vom Parameter auf die Schätzfunktion. Die Parameter einer $N(\mu; \sigma)$ -Verteilung seien bekannt. Gefragt sei nach dem zentralen Bereich (um μ), in dem mit vorgegebener Vertrauenswahrscheinlichkeit $P = 1 - \alpha$ die Werte $\bar{X}_i$ liegen werden.

Ein Bereich, der mit vorgegebener Wahrscheinlichkeit einen bestimmten Anteil einer Grundgesamtheit (hier: der Verteilung der $\bar{X}_i$) enthält, heißt **Toleranzintervall**. Die Toleranzgrenzen für $\bar{X}$ sind durch

$$\mu \pm z \frac{\sigma}{\sqrt{n}} \quad \text{bzw.} \quad P \left(\mu - z \frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq \mu + z \frac{\sigma}{\sqrt{n}} \right) = P = 1 - \alpha \quad (6.44)$$

gegeben. Vertauscht man in Tabelle 6.5 die Symbole μ und $\bar{X}$, so ist sie auch für diesen Zusammenhang gültig. Mit der Vertrauenswahrscheinlichkeit $P = 1 - \alpha$ wird ein beliebiger Stichprobenmittelwert $\bar{X}$ vom Toleranzintervall überdeckt, d. h. in $(P \cdot 100)\%$ aller Fälle ist $\bar{X}$ innerhalb der angegebenen Toleranzgrenzen zu erwarten. Fällt der Stichprobenmittelwert $\bar{X}$ in das Toleranzintervall, so wird man die Abweichung vom Erwartungswert μ der Grundgesamtheit als **zufällig** betrachten, während man sie andernfalls als auf dem **100 α %-Niveau statistisch gesichert** ansieht

und dann schließt, die vorliegende Stichprobe entstamme mit der Vertrauenswahrscheinlichkeit $P = 1 - \alpha$ einer anderen als der betrachteten Grundgesamtheit.

6.6.2 Konfidenzintervall für den Erwartungswert μ

Gegeben sei eine Zufallsstichprobe $x_1, x_2, \dots, x_n$ aus einer normalverteilten Grundgesamtheit. Der Erwartungswert der Grundgesamtheit sei unbekannt und soll aus der Stichprobe geschätzt werden. Der Mittelwert $\bar{x}$ und die Standardabweichung s_n aus der Stichprobe sind daher in den folgenden Ausführungen als Zufallsvariablen $\bar{X}_n$ und S_n aufzufassen. Wir suchen zwei aus der Stichprobe zu errechnende Werte x_{links} und x_{rechts} , die mit einer bestimmten, nicht zu kleinen Wahrscheinlichkeit den unbekannten Parameter μ zwischen sich einschließen: $x_{\text{links}} \leq \mu \leq x_{\text{rechts}}$. Diese Grenzen nennt man **Vertrauensgrenzen** (confidence limits), sie bestimmen den sogenannten **Vertrauens-**, oder **Konfidenzbereich** (confidence interval), den wir für den Fall, dass σ bekannt ist, in (6.43) und Tabelle 6.5 kennengelernt haben. Mit der Vertrauenswahrscheinlichkeit $(1 - \alpha)$ liegt dann, wenn σ durch s geschätzt werden muss, der gesuchte Parameter μ zwischen den Vertrauensgrenzen (6.45)

$$\bar{x} \pm t_{n-1; 1-\alpha/2} \cdot \frac{s}{\sqrt{n}} \quad (6.45)$$

mit $t_{n-1; 1-\alpha/2}$ (Quantil der Student-Verteilung, das mit der Funktion **qt()** in **R** berechnet oder der Tabelle 5.10 entnommen werden kann), d. h. in durchschnittlich $100 \cdot (1 - \alpha)\%$ aller Stichproben werden diese Grenzen den wahren Wert des Parameters einschließen:

$$P(\bar{x} - t_{n-1; 1-\alpha/2} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{x} + t_{n-1; 1-\alpha/2} \frac{s}{\sqrt{n}}) = 1 - \alpha \quad (6.46)$$

d.h. in höchstens $100 \cdot \alpha\%$ aller Stichproben wird der unbekannte Parameter nicht überdeckt. In durchschnittlich $100\alpha\%$ aller Stichproben werden diese Grenzen den Parameter nicht erfassen, und zwar wird er in durchschnittlich $100 \cdot \alpha/2\%$ aller Stichproben unterhalb und in durchschnittlich $100 \cdot \alpha/2\%$ aller Stichproben oberhalb des Vertrauensbereiches liegen. Es sei daran erinnert, dass für das vorliegende zweiseitige Konfidenzintervall $\alpha/2 + (1 - \alpha) + \alpha/2 = 1$ gilt. **Einseitige Konfidenzintervalle** (z. B. obere Vertrauensgrenze $\mu_{\text{ob.}} = \bar{x} + t_{n-1; 1-\alpha} s / \sqrt{n}$):

untere VG	obere VG
$P(\bar{x} - t \frac{s}{\sqrt{n}} \leq \mu) = (1 - \alpha) \quad \text{bzw.} \quad P(\mu \leq \bar{x} + t \frac{s}{\sqrt{n}}) = (1 - \alpha)$	(6.47)

mit $t_{n-1; 1-\alpha}$ schließen in durchschnittlich $100 \cdot \alpha\%$ aller Stichproben den Parameter nicht ein, überdecken ihn dagegen in durchschnittlich $100 \cdot (1 - \alpha)\%$ aller Fälle.

In **R** kann das Konfidenzintervall für den Erwartungswert direkt nach (6.47) berechnet oder mit der Funktion **t.test()** (hier für den Fall einer Stichprobe) bestimmt werden.

```
> x <- c(95, 84, 105, 96, 86, 86, 95, 94, 75, 93)
> n <- length(x)
> m <- mean(x); m
[1] 90.9
> s <- sd(x); s
[1] 8.305955
> m - qt(0.975, n-1)*s/sqrt(n)      # untere Vertrauensgrenze
[1] 84.95828
> m + qt(0.975, n-1)*s/sqrt(n)      # obere Vertrauensgrenze
[1] 96.84172
> t.test(x, mu = 90, conf.level = 0.95)
      One Sample t-test
data:  x ... interval:  84.95828 96.84172 ...
```

Die Schätzung für den Erwartungswert ist demnach $\hat{\mu} = 90,9$; der wahre Wert für μ wird dagegen mit den Vertrauensgrenzen von 84,96 und 96,84 (95%) überdeckt.

Ist σ bekannt oder wird s aus großem n berechnet, dann wird (6.45) durch (6.48) ersetzt (z = Standardnormalvariable):

$$\boxed{\bar{x} \pm z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}} \quad \text{Stichprobenziehen mit Zurücklegen} \quad (6.48)$$

mit $z_{1-\alpha/2} = 1,96$ ($\alpha = 0,05$), $z_{1-\alpha/2} = 2,58$ ($\alpha = 0,01$) und $z_{1-\alpha/2} = 3,29$ ($\alpha = 0,001$).

Vorausgesetzt wird auch hier, dass die Stichprobe entweder einer unendlich großen Grundgesamtheit entstammt, oder einer endlichen Grundgesamtheit entnommen ist und ihr anschließend wieder zugeteilt wird. Entstammt die Stichprobe einer **endlichen Grundgesamtheit** des Umfangs N und wird sie nach Entnahme und Auswertung nicht wieder zu ihrer Grundgesamtheit zurückgelegt, so gelten die Vertrauensgrenzen

$$\boxed{\bar{x} \pm z \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}} \quad \text{Stichprobenziehen ohne Zurücklegen} \quad (6.49)$$

Der Quotient $\frac{N-n}{N-1}$ heißt „Korrekturfaktor für endliche Grundgesamtheiten“, kurz

„**Endlichkeitskorrektur**“. Es gilt:
(vgl. 6.32)

$$\boxed{\frac{N-n}{N-1} \approx \frac{N-n}{N} = 1 - \frac{n}{N}} \quad (6.50)$$

Man bezeichnet $\frac{n}{N}$ als **Auswahlsatz** (s. auch Seite 236). Je kleiner er ist, um so eher wird man auf den Korrekturfaktor verzichten, d. h. für hinreichend große Grundgesamtheiten wird bei gegebenem Stichprobenumfang die Unterscheidung in Stichproben-Entnahme mit bzw. ohne Zurücklegen überflüssig. Für $n/N < 0,05$ ist die Endlichkeitskorrektur im allgemeinen zu vernachlässigen.

Der Quotient $\frac{\sigma}{\sqrt{n}} = \sigma_{\bar{X}}$ kennzeichnet den Standardfehler des Mittelwertes ($\bar{X}$). Die entsprechende Varianz, **Varianz des Mittelwertes $\bar{X}$**

$$\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n} \left(\frac{N-n}{N-1} \right) \approx \frac{\sigma^2}{n} \left(1 - \frac{n}{N} \right) \quad (6.51)$$

wird gleich Null, sobald $n = N$ ist. Das Konfidenzintervall (KI) für μ kann daher (6.48) bzw. (6.45) **in kurzer Form** auch wie folgt geschrieben werden:

$$\boxed{\bar{x} \pm z \sigma_{\bar{X}}} \quad \text{bzw.} \quad \boxed{\bar{x} \pm t s_{\bar{x}}}$$



Beispiel: Gegeben sei die Stichprobe $n = 200$, $\bar{x} = 320$, $s = 20$ aus einer Normalverteilung. Bestimme das 95%-Konfidenzintervall (95%-KI) für den Erwartungswert μ .

$$\frac{t_{199;0,975}=1,972}{t \cdot s_{\bar{x}}=1,972 \cdot 1,414=2,77} \quad s_{\bar{x}}=\frac{s}{\sqrt{n}}=\frac{20}{\sqrt{200}}=1,414 \quad \frac{z=1,96}{z \cdot s_{\bar{x}}=1,96 \cdot 1,414=2,77}$$

$$317 \leq \mu \leq 323$$

Das seltener gebrauchte prozentuale Konfidenzintervall errechnet man bei Bedarf nach

$$\frac{t}{\bar{x}} \cdot s_{\bar{x}} = \frac{1,972}{320} \cdot 1,414 = 0,0087 = 0,9\% \text{ bzw. } \frac{z}{\bar{x}} \cdot s_{\bar{x}} = \frac{1,96}{320} \cdot 1,414 = 0,0087 = 0,9\%.$$

KI für μ :	Weitere Beispiele				Differenz
$\bar{x} \pm ts/\sqrt{n}$	80%-KI:	$10 \pm 0,59$	oder	$9,41 \leq \mu \leq 10,59$	1,18
für:	90%-KI:	$10 \pm 0,77$	oder	$9,23 \leq \mu \leq 10,77$	1,54
$n = 20$	95%-KI:	$10 \pm 0,94$	oder	$9,06 \leq \mu \leq 10,94$	1,88
$\bar{x} = 10$	99%-KI:	$10 \pm 1,28$	oder	$8,72 \leq \mu \leq 11,28$	2,56
$s = 2$	99,9%-KI:	$10 \pm 1,74$	oder	$8,26 \leq \mu \leq 11,74$	3,48



Das 95%-KI für μ gibt man an als „95%-KI: $\bar{x} \pm ts_{\bar{x}}$ “ [mit $t = t_{n-1;0,975}$] bzw. besser als „95%-KI: $a \leq \mu \leq b$ “; z. B. (95%-KI: 320 ± 3), 95%-KI: $317 \leq \mu \leq 323$.

Hinweis (Rückschluss und direkter Schluss):

Schließen wir nach (6.52) von den Werten der Stichprobe auf den Erwartungswert der Grundgesamtheit



$$\bar{X} - t \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t \frac{S}{\sqrt{n}} \quad (6.52)$$

so liegt ein **Rückschluss** oder, da die Stichprobe die Grundgesamtheit „repräsentiert“, ein **Repräsentationsschluss** vor. Umgekehrt ist der Schluss von den Parametern der Grundgesamtheit auf den Mittelwert $\bar{X}$ (Schätzfunktion) einer Stichprobe

$$\mu - z \frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq \mu + z \frac{\sigma}{\sqrt{n}} \quad (6.53)$$

ein **direkter Schluss** oder, da die Grundgesamtheit die Stichprobe mit „einschließt“, ein **Inklusionsschluss**. Schließt man von den Werten einer Stichprobe auf die einer anderen Stichprobe derselben Grundgesamtheit, dann liegt ein sogenannter **Transponierungsschluss** vor.

6.6.3 Konfidenzintervall für die Differenz $\mu_1 - \mu_2$

Das Konfidenzintervall für die Differenz zweier Erwartungswerte unabhängiger Stichproben aus normalverteilten Grundgesamtheiten mit gleicher Varianz ($\sigma_1^2 = \sigma_2^2$) ist durch (6.54) gegeben:

$$(\bar{x}_1 - \bar{x}_2) - t_{\nu;1-\alpha/2} A \leq \mu_1 - \mu_2 \leq (\bar{x}_1 - \bar{x}_2) + t_{\nu;1-\alpha/2} A$$

mit $\nu = n_1 + n_2 - 2$ und

$$A = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \left[\frac{1}{n_1} + \frac{1}{n_2} \right]}$$

Für $n_1 = n_2 = n$ gilt: $\nu = 2n - 2$

und $A = \sqrt{\frac{s_1^2 + s_2^2}{n}}$

(6.54)

Ein Unterschied zwischen μ_1 , und μ_2 ist auf dem verwendeten Niveau statistisch signifikant, sobald das Konfidenzintervall den Wert $\mu_1 - \mu_2 = 0$ nicht einschließt. Statistische Testverfahren und Konfidenzintervalle führen beide zu Entscheidungen. Das Konfidenzintervall bietet darüber hinaus noch zusätzliche Informationen über den oder die Parameter!

(1) Schließt ein Konfidenzintervall die **Null** aus, so sind Vorzeichen, Lage und Breite des Konfidenzintervalles aufschlussreich.

- (2) Konfidenzintervalle charakterisieren den Parameter, sind **einfacher zu verstehen als Tests und diesen praktisch gleichwertig**: eine H_0 bezüglich eines Parameters, die auf dem 5%-Niveau abgelehnt wird, wird auch abgelehnt, wenn der entsprechende 95%-KI den Nullwert-Parameter ausschließt.
- (3) Im **Ein-Parameter-Fall** ist das Konfidenzintervall dem Test fast stets überlegen. Liegen zwei oder mehr Parameter vor, dann ist der P -Wert meist einfacher zu berechnen und wird dann routinemäßig angegeben.
- (4) Sind **mehrere Konfidenzintervalle** zu vergleichen, so bevorzugt man 99%-Konfidenzintervalle.



Beispiel: Für den Vergleich zweier Stichproben mit

$$\begin{aligned} n_1 &= 30 \quad \bar{x}_1 = 42,76 \quad s_1^2 = 33,44 \\ n_2 &= 30 \quad \bar{x}_2 = 40,21 \quad s_2^2 = 22,55 \end{aligned}$$

folgt mit $t_{58;0,975} = 2,002$ und $A = 1,366$

$$\begin{aligned} 2,55 - 2,002 \cdot 1,366 &\leq \mu_1 - \mu_2 \leq 2,55 + 2,002 \cdot 1,366 \\ -0,19 &\leq \mu_1 - \mu_2 \leq 5,29 \end{aligned}$$

Da das 95%-KI die Null mit einschließt, kann die Nullhypothese, hier die Übereinstimmung der Erwartungswerte, nicht abgelehnt werden.

Sind die Varianzen unterschiedlich ($\sigma_1^2 \neq \sigma_2^2$) und sind $n_1 \geq 6$ und $n_2 \geq 6$, dann kann ein **Konfidenzintervall für die Differenz der Erwartungswerte nach Hsu** wie folgt bestimmt werden:

$$\begin{aligned} (\bar{x}_1 - \bar{x}_2) - t_{\nu;1-\alpha/2}B &\leq \mu_1 - \mu_2 \leq (\bar{x}_1 - \bar{x}_2) + t_{\nu;1-\alpha/2}B \\ \text{mit } \nu &= \min(n_1, n_2) - 1 \quad \text{und} \\ B &= \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} \\ \text{Für } n_1 = n_2 = n \text{ gilt: } B &= \sqrt{\frac{s_1^2 + s_2^2}{n}} \end{aligned} \tag{6.55}$$

Die Abschätzung der Freiheitsgrade nach Welch, wie in früheren Auflagen verwendet, gewährleistet nicht immer die Einhaltung des α -Niveaus.



Beispiel: Für den Vergleich zweier Stichproben mit

$$\begin{aligned} n_1 &= 66 \quad \bar{x}_1 = 183 \quad s_1^2 = 5 \\ n_2 &= 40 \quad \bar{x}_2 = 175 \quad s_2^2 = 7 \end{aligned}$$

folgt mit $\nu = 39$, $t_{39;0,975} = 2,023$ und $B = 1,2664$

$$\begin{aligned} 8 - 2,023 \cdot 1,2664 &\leq \mu_1 - \mu_2 \leq 8 + 2,023 \cdot 1,2664 \\ 5,44 &\leq \mu_1 - \mu_2 \leq 10,56 \end{aligned}$$

Die Nullhypothese, hier Gleichheit der Erwartungswerte, kann abgelehnt werden, da der 95%-KI für die Differenz $\mu_1 - \mu_2$ die Null deutlich ausschließt.

Weitere Verfahren zur Bestimmung von Konfidenzintervallen für die Differenz aus Erwartungswerten im Rahmen von multiplen paarweisen Vergleichen werden bei den Mehrstichprobentests angegeben.

6.6.4 Das Konfidenzintervall für den Erwartungswert μ_d der Paardifferenzen

Das Konfidenzintervall für den Erwartungswert μ_d der Paardifferenzen ist durch

$$\boxed{\bar{d} \pm (t_{n-1; 1-\alpha/2}) s_{\bar{d}}} \quad (6.56)$$

gegeben, mit $\bar{d} = \frac{\sum d}{n}$ (Mittelwert der Paardifferenzen) und

$$s_{\bar{d}} = \frac{s_d}{\sqrt{n}} = \sqrt{\frac{\sum d_i^2 - (\sum d_i)^2/n}{n(n-1)}} \quad (\text{Standardfehler der mittleren Differenz}) \text{ sowie dem ent-}$$

sprechenden Quantil der t-Verteilung $t_{n-1; 1-\alpha/2}$.

Beispiel: Die folgende Tabelle enthält Messwerte (x_i, y_i) für ein Material, das nach zwei Verfahren behandelt wurde. Ein Behandlungseffekt kann durch die Angabe des 95%-Konfidenzintervalles für den Erwartungswert der Paardifferenzen bewertet werden.



Nr.	x_i	y_i	$d_i = x_i - y_i$	d_i^2
1	4,0	3,0	1,0	1,00
2	3,5	3,0	0,5	0,25
3	4,1	3,8	0,3	0,09
4	5,5	2,1	3,4	11,56
5	4,6	4,9	-0,3	0,09
6	6,0	5,3	0,7	0,49
7	5,1	3,1	2,0	4,00
8	4,3	2,7	1,6	2,56
$n = 8$			$\sum d_i = 9,2$	$\sum d_i^2 = 20,4$

Man erhält man das 95%Konfidenzintervall: $1,15 \pm 2,365 \cdot 0,411$ bzw. $1,15 \pm 0,97$, d. h. 95%-KI: $0,18 \leq \mu_d \leq 2,12$, d.h. μ_d ist deutlich von Null verschieden ($P < 0,05$).

```
> x <- c(4.0, 3.5, 4.1, 5.5, 4.6, 6.0, 5.1, 4.3)
> y <- c(3.0, 3.0, 3.8, 2.1, 4.9, 5.3, 3.1, 2.7)
> d <- x - y; d
[1] 1.0 0.5 0.3 3.4 -0.3 0.7 2.0 1.6
> t.test(x, y, mu=0, paired=TRUE, con.level = 0.95)
```

Paired t-test

```
data: x and y,
...
95 percent confidence interval:
 0.1781177 2.1218823
...
```

Einseitige Vertrauensgrenzen lassen sich auch angeben. Als obere 95%-Vertrauensgrenze erhielt man nach (6.56) mit $t_{7;0,95} = 1,895$ über $1,15 + 1,895 \cdot 0,411 = 1,15 + 0,78$ den Wert 1,93, d. h. 95%-VG_{oben}: $\mu_d = 1,93$ bzw. $\mu_d \leq 1,93$.

6.6.5 Konfidenzintervall für das Verhältnis μ_1/μ_2

Für $\lambda = \mu_1/\mu_2$, mit $\mu_2 \neq 0$, lässt sich nach Chakravarti [Cha71] ein $(1 - \alpha)$ -Konfidenzintervall angeben:

$$\begin{aligned}
 &P(\lambda_- \leq \lambda \leq \lambda_+) \geq 1 - \alpha \quad \text{mit} \\
 &\lambda_{\pm} = (b \pm \sqrt{b^2 - ac})/a \quad \text{und} \\
 &a = \bar{x}_2^2 - \frac{t_{n_2-1, 1-\alpha/2}^2 s_2^2}{n_2}, \quad b = \bar{x}_1 \bar{x}_2, \quad c = \bar{x}_1^2 - \frac{t_{n_1-1, 1-\alpha/2}^2 s_1^2}{n_1}
 \end{aligned} \tag{6.57}$$

Zu beachten ist, dass das Konfidenzintervall mindestens gleich und nicht genau gleich $1 - \alpha$ ist.



Beispiel: Für den Vergleich zweier Gruppen folgt

Untersuchungsgruppe: $n_1 = 42$, $\bar{x}_1 = 11,33$, $s_1 = 7,59$

Kontrollgruppe: $n_2 = 32$, $\bar{x}_2 = 15,25$, $s_2 = 8,55$

$t_{41;0,975} = 2,02$, $t_{31;0,975} = 2,04$

$$\hat{\lambda} = \frac{\bar{x}_1}{\bar{x}_2} = \frac{11,33}{15,25} = 0,743$$

$$a = 15,25^2 - \frac{2,04^2 \cdot 8,55^2}{32} = 223,06$$

$$b = 11,33 \cdot 15,25 = 172,78$$

$$c = 11,33^2 - \frac{2,02^2 \cdot 7,59^2}{42} = 122,77$$

$$\lambda_{\pm} = (172,78 \pm \sqrt{29852,928 - 223,06 \cdot 122,77})/223,06$$

$$\lambda_+ = 0,9973, \lambda_- = 0,5519$$

$$95\text{-KI: } 0,552 \leq \lambda \leq 0,997$$

6.6.5.1 Die Schätzung von Verhältniszahlen

Eine Verhältniszahl ist ein Quotient zweier Kennziffern, der Auskunft gibt über Beziehungen zwischen zwei Erscheinungen. Etwa die mittlere Zahl der Einwohner pro Arzt, die mittlere Höhe der Ausgaben für Nahrungsmittel pro Haushaltseinkommen oder die mittlere Wertänderung von Einfamilienhäusern nach zwei Jahren, jeweils bezogen auf ein bestimmtes Gebiet.

Für die Formeln werden die folgenden Bezeichnungen verwendet:

$$\text{Grundgesamtheit: } v = \frac{\sum_{i=1}^N Y_i}{\sum_{i=1}^N X_i} = \frac{\mu_y}{\mu_x}$$

$$\text{Zufallsstichprobe: } \hat{v} = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i} = \frac{\bar{y}}{\bar{x}}$$

Für nicht zu kleine Stichprobenumfänge ($n > 20$) und für den Fall, dass für beide Variationskoeffizienten gilt: $s_x/\bar{x} < 0,1$ und $s_y/\bar{y} < 0,1$, lässt sich ein angenähertes **95%-Konfidenzintervall** für das Verhältnis v in der Grundgesamtheit angeben:

$$\hat{v} \pm 1,96 \sqrt{\left[\frac{N-n}{Nn} \right] \left[\frac{1}{\mu_x^2} \right] [\hat{v}^2 s_x^2 + s_y^2 - 2\hat{v} r s_x s_y]} \tag{6.58}$$

Ist μ_x unbekannt, so ersetze man μ_x durch $\bar{x}$. Hierbei sind s_x^2 und s_y^2 die beiden Stichprobenvarianzen, s_x und s_y die entsprechenden Standardabweichungen und r ist der Stichprobenkorrelationskoeffizient. Sollte der r zugrundeliegende Parameter ρ bekannt sein, so wird r durch ρ ersetzt. Bei unendlich großer Grundgesamtheit ist $[(N-n)/(Nn)]$ durch $[1/n]$ zu ersetzen. Liegen μ_x und r nicht vor, so ersetze man (6.58) durch (6.59) oder (6.60).

$$\hat{v} \pm 1,96 \sqrt{\frac{N-n}{nN} \left[\frac{1}{\bar{x}^2} \right] \left[\left(\hat{v}^2 \sum_{i=1}^n x_i^2 + \sum_{i=1}^n y_i - 2\hat{v} \sum_{i=1}^n x_i y_i \right) / (n-1) \right]} \quad (6.59)$$

$$\hat{v} \pm 1,96 \sqrt{\frac{N-n}{Nn} \left[\frac{1}{\bar{x}^2} \sum_{i=1}^n \frac{(y_i - x_i \bar{y}/\bar{x})^2}{n-1} \right]} \quad (6.60)$$

Beispiele sind bei Bedarf z.B. Mendenhall und Mitarbeitern [MOS71] zu entnehmen.

6.6.6 Mindestzahl von Beobachtungen zur Schätzung eines Mittelwertes

Formel (6.61) gibt mit vorgegebener Genauigkeit (d) und vorgegebener statistischer Sicherheit minimale Stichprobenumfänge an (auf der Normalverteilung basierende Näherungen!) zur Schätzung des Mittelwertes ($n_{\bar{x}}$) [mit $d = \bar{x} - \mu$]:

$$n_{\bar{x}} > \left(\frac{z_{\alpha}}{d} \right)^2 \cdot \sigma^2 \quad (6.61)$$

$z_{1-\alpha/2}$ ist das entsprechende Quantil der Standardnormalverteilung für die gewünschte Vertrauenswahrscheinlichkeit $1 - \alpha$ (d. h. die Irrtumswahrscheinlichkeit α). Für die Beispiele benutzen wir $z_{0,975} = 1,96$ für $\alpha = 0,05$ (95%KI) und $z_{0,995} = 2,58$ für $\alpha = 0,01$ (99%KI).

(6.61) **unterschätzt** den benötigten Stichprobenumfang. Muss σ^2 erst noch anhand einer Zufallsstichprobe des Umfangs m geschätzt werden, dann sollte $m \gtrapprox 60$ gelten; für $m \leq 60$ geben Shiffler und Adams [SA87] Korrekturfaktoren c , mit denen $n_{\bar{x}}$, berechnet aus (6.61) mit s_m^2 anstatt von σ^2 , zu multiplizieren ist, um $n_{\bar{x},\text{korrt.}}$ zu erhalten. Einige Werte c mit zugehörigen in Klammern gesetzten m -Werten sind: 1,011 (60); 1,017 (40); 1,036 (20); 1,049 (15); 1,064 (12); 1,071 (10); ...; 1,443 (3).

Beispiel: Zur Schätzung eines **Mittelwertes** bei bekannter Varianz $\sigma^2 = 3$ mit einer Irrtumswahrscheinlichkeit $\alpha = 0,01$ und mit einer Genauigkeit von $d = 0,5$ benötigt man mehr als $n_{\bar{x}} = (2,58/0,5)^2 \cdot 3 = 80$ Beobachtungen; d. h. mit etwa 90 Beobachtungen erhält man den 99%-KI für μ



$(\bar{x} - 0,5 \leq \mu \leq \bar{x} + 0,5$ bzw. $\mu = \bar{x} \pm 0,5$) mit der Länge $2d$.

Zu $n_{\bar{x}}$, jetzt kurz n genannt: Ist n größer als 10% der Grundgesamtheit N ($n > 0,1 \cdot N$), so benötigt man nicht n , sondern nur $n' = n/[1 + (n/N)]$ Beobachtungen: Für $N = 750$ benötigt man somit nicht 90, sondern $90/[1 + (90/750)] = 81$ Beobachtungen.

6.7 Konfidenzintervall für die mittlere absolute Abweichung

Seien X_i mit $i = 1, \dots, n$ stetige, unabhängige und identisch verteilte Zufallsvariablen, die den Erwartungswert $E[X_i] = \mu$ und die Varianz $\text{var}[X_i] = \sigma^2$ haben, dann wird die mittlere absolute Abweichung vom Median $\tilde{\mu}$ (MD) durch (6.62) definiert.

$$MD = \frac{\sum_{i=1}^n |X_i - \tilde{\mu}|}{n} \quad (6.62)$$

Eine konsistente Schätzung für den Erwartungswert $E[MD] = \tau$ der mittleren absoluten Abweichung anhand einer Stichprobe ist nach (6.63) möglich.

$$\hat{\tau} = \frac{\sum_{i=1}^n |x_i - \tilde{x}|}{n} \quad (6.63)$$

Darin bezeichnet $\tilde{x}$ den Medianwert aus der Stichprobe. Der Schätzer ist allerdings verzerrt und es nicht möglich einen einfachen, allgemein gültigen Ansatz für eine unverzerrte Schätzung anzugeben. Empirische Untersuchungen lassen vermuten, dass eine Korrektur mit dem Faktor $c = n/(n-1)$ (ähnlich der Varianzschätzung) zur einer geringen Verzerrung führt.

$$\hat{\tau}' = \hat{\tau} \cdot c = \frac{\sum_{i=1}^n |x_i - \tilde{x}|}{n-1} \quad (6.64)$$

Die Verteilung von $\hat{\tau}$ zeigt eine positive Schiefe, die durch eine Logarithmustransformation ausgeglichen werden kann. Somit wird die Schätzung für die Varianz von $\hat{\tau}$ nach Bonett und Seier [BS03] durch (6.65) berechnet.

$$\begin{aligned} \widehat{Var}[\log(\hat{\tau})] &= (\hat{\delta}^2 + \hat{\gamma} - 1)/n \\ \text{mit } \hat{\delta} &= (\bar{x} - \tilde{x})/\hat{\tau} \\ \text{und } \hat{\gamma} &= s^2/\hat{\tau}^2 \end{aligned} \quad (6.65)$$

In (6.65) bezeichnet $\bar{x}$ den arithmetischen Mittelwert und s^2 die empirische Varianz aus den Werten der Stichprobe.

Das $(1-\alpha)100\%$ -Konfidenzintervall für τ kann anhand von (6.65) mit dem entsprechenden Quantil aus der Standardnormalverteilung $z_{1-\alpha/2}$ nach (6.66) bestimmt werden.

$$\exp \left[\log(\hat{\tau}') \pm z_{1-\alpha/2} \sqrt{\widehat{Var}[\log(\hat{\tau})]} \right] \quad (6.66)$$

Für den Fall von zwei Stichproben kann ein $(1-\alpha)100\%$ -Konfidenzintervall für das Verhältnis τ_1/τ_2 entsprechend nach (6.67) hergeleitet werden.

$$\exp \left[\log(\hat{\tau}'_1/\hat{\tau}'_2) \pm z_{1-\alpha/2} \sqrt{\widehat{Var}[\log(\hat{\tau}_1)] + \widehat{Var}[\log(\hat{\tau}_2)]} \right] \quad (6.67)$$

Einseitige Konfidenzintervalle erhält man, wenn in (6.66) bzw. in (6.67) $z_{1-\alpha/2}$ durch $z_{1-\alpha}$ ersetzt wird. Die Berechnung soll in **R** in einem einfachen Beispiel für das 95%-Konfidenzintervall in einzelnen Schritten verdeutlicht werden.

```
> x <- c(10, 15, 20, 16, 13, 12, 15, 21, 11, 24, 17, 14, 12, 10, 30)
> n <- length(x)
> medi <- median(x)
> c <- n / (n-1)
> tau.h <- sum(abs(x-medi))/n; tau.h*c
[1] 4.357143
> d <- (mean(x) - medi)/tau.h; g <- var(x) / tau.h^2
> varln.tau <- (d^2 + g - 1)/n
> upper <- exp(log(tau.h*c) + qnorm(0.975)*sqrt(varln.tau)); upper
[1] 7.203192
> lower <- exp(log(tau.h*c) - qnorm(0.975)*sqrt(varln.tau)); lower
[1] 2.635595
```

Die Schätzung für die mittlere absolute Abweichung ist $\hat{\tau}' = 4,36$. Das 95%-Konfidenzintervall hat die Grenzen [2,64 bis 7,20].

Tabelle 6.6. Schranken für den Vorzeichentest (aus B.L. Van der Waerden: Mathematische Statistik, Springer, Berlin 1957, S. 345, Tafel 9)

Zweiseitig	5%	2%	1%	Zweiseitig	5%	2%	1%
n = 5	0 5	0 5	0 5	n = 53	19 34	18 35	17 36
6	1 5	0 6	0 6	54	20 34	19 35	18 36
7	1 6	1 6	0 7	55	20 35	19 36	18 37
8	1 7	1 7	1 7	56	21 35	19 37	18 38
9	2 7	1 8	1 8	57	21 36	20 37	19 38
10	2 8	1 9	1 9	58	22 36	20 38	19 39
11	2 9	2 9	1 10	59	22 37	21 38	20 39
12	3 9	2 10	2 10	60	22 38	21 39	20 40
13	3 10	2 11	2 11	61	23 38	21 40	21 40
14	3 11	3 11	2 12	62	23 39	22 40	21 41
15	4 11	3 12	3 12	63	24 39	22 41	21 42
16	4 12	3 13	3 13	64	24 40	23 41	22 42
17	5 12	4 13	3 14	65	25 40	23 42	22 43
18	5 13	4 14	4 14	66	25 41	24 42	23 43
19	5 14	5 14	4 15	67	26 41	24 43	23 44
20	6 14	5 15	4 16	68	26 42	24 44	23 45
21	6 15	5 16	5 16	69	26 43	25 44	24 45
22	6 16	6 16	5 17	70	27 43	25 45	24 46
23	7 16	6 17	5 18	71	27 44	26 45	25 46
24	7 17	6 18	6 18	72	28 44	26 46	25 47
25	8 17	7 18	6 19	73	28 45	27 46	26 47
26	8 18	7 19	7 19	74	29 45	27 47	26 48
27	8 19	8 19	7 20	75	29 46	27 48	26 49
28	9 19	8 20	7 21	76	29 47	28 48	27 49
29	9 20	8 21	8 21	77	30 47	28 49	27 50
30	10 20	9 21	8 22	78	30 48	29 49	28 50
31	10 21	9 22	8 23	79	31 48	29 50	28 51
32	10 22	9 23	9 23	80	31 49	30 50	29 51
33	11 22	10 23	9 24	81	32 49	30 51	29 52
34	11 23	10 24	10 24	82	32 50	31 51	29 53
35	12 23	11 24	10 25	83	33 50	31 52	30 53
36	12 24	11 25	10 26	84	33 51	31 53	30 54
37	13 24	11 26	11 26	85	33 52	32 53	31 54
38	13 25	12 26	11 27	86	34 52	32 54	31 55
39	13 26	12 27	12 27	87	34 53	33 54	32 55
40	14 26	13 27	12 28	88	35 53	33 55	32 56
41	14 27	13 28	12 29	89	35 54	34 55	32 57
42	15 27	14 28	13 29	90	36 54	34 56	33 57
43	15 28	14 29	13 30	91	36 55	34 57	33 58
44	16 28	14 30	14 30	92	37 55	35 57	34 58
45	16 29	15 30	14 31	93	37 56	35 58	34 59
46	16 30	15 31	14 32	94	38 56	36 58	35 59
47	17 30	16 31	15 32	95	38 57	36 59	35 60
48	17 31	16 32	15 33	96	38 58	37 59	35 61
49	18 31	16 33	16 33	97	39 58	37 60	36 61
50	18 32	17 33	16 34	98	39 59	38 60	36 62
51	19 32	17 34	16 35	99	40 59	38 61	37 62
52	19 33	18 34	17 35	100	40 60	38 62	37 63
Einseitig	2,5%	1%	0,5%	Einseitig	2,5%	1%	0,5%

6.8 Konfidenzintervall für den Median

Dieses Konfidenzintervall ist besonders wichtig, da es bei nicht normalverteilten Grundgesamtheiten (6.45) und (6.46) ersetzt!

Sei W eine binomialverteilte Zufallsvariable mit einer Erfolgswahrscheinlichkeit $p = 0.5$ und bei n Versuchen, dann bezeichnet $P(h \leq W \leq n - h)$ für jedes h zwischen 0 und $[n/2]$ die Wahrscheinlichkeit dafür, dass die Zahl der Erfolge W zwischen h und $n - h$ (einschließlich) liegt.

Dieser Zusammenhang wird auf die Ordnungszahlen (Ränge) hinsichtlich der Lage des Medians übertragen.

Bezeichnet man die der Größe nach **aufsteigend geordneten** n Beobachtungen einer Stichprobe mit

$$x_{(1)}, x_{(2)}, x_{(3)}, \dots, x_{(n)} \quad ,$$

dann ist ein Konfidenzintervall für den Median allgemein durch

$$\boxed{x_{(h)} \leq \tilde{\mu} \leq x_{(n-h+1)}} \quad (6.68)$$

gegeben. Da die Binomialverteilung diskret ist, ist es nicht möglich ein h so zu bestimmen, dass die Überdeckungswahrscheinlichkeit exakt durch $1 - \alpha$ eingehalten wird. So kann für $n=10$ exakt nur ein 89,1%- oder ein 97,8%- Konfidenzintervall berechnet werden, aber nicht ein 95%-Konfidenzintervall, wie es oftmals gewünscht wird. Abschätzungen für die Grenzen von Konfidenzintervallen erfolgen dann durch lineare Interpolation oder bei ausreichenden Stichprobenumfängen unter Ausnutzung des zentralen Grenzwertsatzes über Quantile der Standardnormalverteilung. Die Werte von h für ausgewählte Konfidenzintervalle (90%-KI, der 95%-KI und 99%-KI für $\tilde{\mu}$) können der Tabelle 6.6 entnommen werden.



Beispiel: Den 95%-KI für $\tilde{\mu}$ erhält man für $n \leq 100$ anhand von Tabelle 6.6 nach:

$$LS \leq \tilde{\mu} \leq 1 + RS ;$$

$$\text{z. B. } n = 60, 95\text{-KI: } (22. \text{ Wert}) \leq \tilde{\mu} \leq (39. \text{ Wert})$$

Für () setze man dann die entsprechenden geordneten Messwerte.

Für $n \geq 30$ und die Vertrauenswahrscheinlichkeiten 90%, 95%, 99% kann h nach (6.69)

$$\boxed{h = \frac{n - 1 - z\sqrt{n}}{2} \quad \begin{array}{l} \text{auf die ganze} \\ \text{Zahl aufgerundet} \end{array}} \quad (6.69)$$

approximiert werden (mit $z = 1,64; 1,96; 2,58$). So liegt für $n = 300$ das 95%-Konfidenzintervall zwischen dem 133. und dem 168. Wert der aufsteigend geordneten Stichprobe ($h = \lceil 300 - 1 - 1,96\sqrt{300} \rceil / 2 \approx 133$, $n - h + 1 = 300 - 133 + 1 = 168$), z. B. 95%-KI: $[x_{(133)} =] 21,3 \leq \tilde{\mu} \leq 95,4 [= x_{(168)}]$.



Hinweis: Ein konservatives 95%-Konfidenzintervall (d. h. $P > 0,95$) für $\tilde{\mu}$ lässt sich anhand der **Faustregel** $\tilde{x} \pm 1,58(Q_3 - Q_1)/\sqrt{n}$ angeben. Für $x_i: 1, 2, 3, \dots, 100$, d. h. $n = 100$ erhält man $50,5 \pm 1,58(75,75 - 25,25)/\sqrt{100}$ somit $42 \lesssim \tilde{\mu} \lesssim 59$.

6.8.1 Angenäherte verteilungsunabhängige Konfidenzintervalle für beliebige Quantile

Gilt $F(x) = p$, so heißt x das **Quantil** von p , symbolisiert durch x_p . Quantile sind somit die **Umkehrfunktion** von $F(x)$. In einer stetigen Grundgesamtheit liegen $100p\%$ der Grundgesamtheit unter dem Quantil und $100(1 - p)\%$ über dem Quantil. Besonders wichtige Quantile der Grundgesamtheit wie die Quartile oder die Dezile der Grundgesamtheit schreiben wir mit dem griechischen Buchstaben ξ (Xi), z. B. Median $\tilde{\mu} = \xi_{0,5}$, 3. Quartil $= \xi_{0,75}$ und 1. Dezil $= \xi_{0,10}$.

Angenäherte 95%-Konfidenzintervalle für nicht zu extreme ξ_p -Quantile: $0,1 \leq p \leq 0,9$ ($n \gtrsim 100$):

$$\left. \begin{matrix} a \\ b \end{matrix} \right\} = \left\{ \begin{matrix} - \\ + \end{matrix} np \cdot 1,96 \cdot \sqrt{np(1-p)} \begin{matrix} \text{abgerundet} \\ \text{aufgerundet} \end{matrix} \right\} \begin{matrix} \text{zur nächsten} \\ \text{ganzen Zahl} \end{matrix} \quad (6.70)$$

Für das 90%-Konfidenzintervall wird 1,96 durch 1,645 ersetzt.

Soll ein Konfidenzintervall für ein extrem liegendes ξ_p -Quantil bestimmt werden, so werden natürlich deutlich mehr als 100 Beobachtungen benötigt; dann ist (6.70) anwendbar.

Beispiel: Das erste Dezil $\xi_{0,10}$ für $n = 100$:

$$100 \cdot 0,1 \mp 1,96 \cdot \sqrt{100 \cdot 0,1(1 - 0,1)}$$

$$10 \mp 5,9; \quad \text{d. h. } a = 4 \text{ und } b = 16$$

$$95\text{-KI: } (4. \text{ Wert}) \lesssim \xi_{0,10} \lesssim (16. \text{ Wert}).$$



In **R** erfolgt die Bestimmung von Konfidenzintervallen für den Median unter anderem in der Funktion **wilcox.test()** auf der Grundlage eines Algorithmus von Bauer [Bau72]. Für den Fall von Bindungen, d.h. gleicher Werte in der Rangfolge, wird eine angenäherte (asymptotisch) Abschätzung durchgeführt. Eine Alternative ist auch die Funktion **wilcox.exact()** aus dem Zusatzpaket **library(exactRankTests)** (Hothorn [HH05]).

```
> x <- c(95, 84, 105, 96, 86, 86, 95, 94, 75, 93)
> wilcox.test(x, mu = 0, conf.int = TRUE, conf.level = 0.95)
Wilcoxon signed rank test with continuity correction
...
95 percent confidence interval: 85.0 95.5
...
Warning messages: Cannot compute exact confidence interval with ties in:
wilcox.test.default(x, mu = 0, conf.
...
> library(exactRankTests)
> wilcox.exact(x, mu = 0, conf.int = TRUE, conf.level = 0.95)
Exact Wilcoxon signed rank test
...
95 percent confidence interval: 84.5 96.0
...
```

6.9 Konfidenzintervalle nach dem Bootstrap-Verfahren

Die Schätzung von Parametern einer Verteilung kann nach Efron [ET93] durch ein spezielles **Resampling-Verfahren** erfolgen. Die wesentlichen Vorteile dieses Ansatzes betreffen die folgenden Punkte:

- **Verständnis:** Das Vorgehen fördert ein „intuitives“ Verständnis theoretischer Konzepte durch die konkrete Analogie (Variation in Stichproben).
- **Voraussetzungen:** Das Verfahren erfordert nicht, dass ein spezielles Verteilungsmodell (z.B. die Normalverteilung) für die Zufallsvariablen vorliegt. Die Stichproben müssen auch nicht so groß sein, dass Folgerungen aus dem zentralen Grenzwertsatz Anwendung finden können: $n > 20$ sollte allerdings schon eingehalten sein.
- **Verallgemeinerung:** Das Verfahren ist für eine Vielzahl von (auch komplexen) Statistiken / Schätzfunktionen einsetzbar, ohne dass neue Verteilungsmodelle und deren Eigenschaften herzuleiten und zu überprüfen sind.

Gegeben sei eine Stichprobe $X_1, \dots, X_n$ vom Umfang n , für die eine Realisierung $x_1, \dots, x_n$ vorliege. Ist diese Stichprobe repräsentativ für die zugrundeliegende Wahrscheinlichkeitsverteilung, dann können Aussagen über spezielle Schätzfunktionen (Parameterschätzungen) $\hat{\vartheta}_n = (x_1, \dots, x_n)$ anhand der Werte $x_1, \dots, x_n$ durch das folgende Vorgehen begründet werden. Speziell für die Bestimmung von Vertrauensbereichen ist die **Bootstrap-Perzentilmethode** eine intuitiv verständliche und leicht nachvollziehbare Prozedur:

- (1) **Resampling:** Erzeugen neuer **Stichproben durch Ziehen mit Zurücklegen** aus den vorliegenden Beobachtungen, jeweils mit dem gleichen Stichprobenumfang n :

$$x_1^{*k}, \dots, x_n^{*k} \quad \text{für} \quad k = 1, \dots, K \quad (6.71)$$

Dabei sollte in jedem Fall $K \geq 500$ (keinesfalls weniger) gewählt werden.

- (2) **Schätzung:** Berechnen der Statistik oder des Parameters (z.B. Mittelwert oder Medianwert) auf der Grundlage der unter 1. gewonnenen Stichproben führt auf wiederholte **Bootstrap-Schätzungen**

$$\hat{\vartheta}_n^{*k} = (x_1^{*k}, \dots, x_n^{*k}) \quad (6.72)$$

- (3) **Verteilung:** Die $\hat{\vartheta}_n^{*k}$ werden als **Bootstrap-Stichprobe** für die Beurteilung der Verteilung des zu schätzenden Parameters herangezogen. Insbesondere können ausgewählte Quantile dieser Verteilung (z.B. 0,05-0,95 für das 90%-KI und 0,025-0,975 für das 95%-KI) bestimmt und im Sinne der entsprechenden Konfidenzintervalle interpretiert werden.

Aus diesem Ansatz ergeben sich zahlreiche Möglichkeiten der Anwendung im Rahmen von Parameterschätzungen und Hypothesentests (M.R. Chernick [Che99]).

Beispiel: In einer Stichprobe (hier Ausgangs- oder Original-Stichprobe) werden $n = 8$ Werte beobachtet: 68, 69, 69, 70, 71, 72, 72, 74 (z.B. Körpergrößen in inch gemessen). An diesen Beispieldaten soll das Prinzip verdeutlicht werden: die statistische Validität des Bootstrap-Ansatzes ist allerdings nur für Stichproben mit $n > 20$ gegeben. Aus diesen Daten werden 5 Bootstrap-Stichproben erzeugt (Ziehen mit Zurücklegen).

```
> x <- c(68, 69, 69, 70, 71, 72, 72, 74); mean(x)
[1] 70.625
> b1 <- sample(x, 8, replace = TRUE); b1; mean(b1)
[1] 71 70 72 71 69 71 68 68 [1] 70
> b2 <- sample(x, 8, replace = TRUE); b2; mean(b2)
[1] 72 69 72 69 68 74 72 68 [1] 70.5
> b3 <- sample(x, 8, replace = TRUE); b3; mean(b3)
[1] 71 72 72 74 69 69 74 70 [1] 71.375
> b4 <- sample(x, 8, replace = TRUE); b4; mean(b4)
[1] 68 68 71 69 72 72 71 70 [1] 70.125
> b5 <- sample(x, 8, replace = TRUE); b5; mean(b5)
[1] 70 70 74 71 69 74 71 69 [1] 71
> sd(c(mean(b1), mean(b2), mean(b3), mean(b4), mean(b5)))
[1] 0.5822907
```

Der Mittelwert aus der Originalstichprobe ist 70,625 und der Standardfehler für den Mittelwert ist 0,706. Die Mittelwerte der Bootstrap-Stichproben sind 70,70,5,71,375,70,125 und 71. Die Standardabweichung dieser 5 Werte beträgt 0,582 und kann als eine Schätzung für den Standardfehler betrachtet werden.

Die „übliche“ Bestimmung eines 95%-Konfidenzintervalls mit Hilfe des entsprechenden Quantils der t -Verteilung ($t_{7,0,975} = 2,365$) führt zu den Grenzen (68,96 bis 72,29). Wie zu erwarten, liegen alle Mittelwerte aus den 5 Bootstrap-Stichproben in diesem Intervall. Andererseits erwarten wir, dass unter 100 Bootstrap-Stichproben etwa 5 Mittelwerte resultieren, die kleiner als 68,96 oder größer als 72,29 sind. Diese Vertrauensgrenzen (Grundlage t -Verteilung) sind nur solange verlässlich, wie die Annahme einer Normalverteilung zumindest angenähert berechtigt ist.

Werden nun 1000 Bootstrap-Stichproben erzeugt, dann kann auf ein 95%-Konfidenzintervall aus der Verteilung (Quantile) der berechneten Mittelwerte geschlossen werden, ohne dass die Annahme hinsichtlich der Normalverteilung erforderlich wäre. Mit einigen einfachen Anweisungen kann diese Idee in **R** umgesetzt werden und führt mit den Beispieldaten zu dem 95%-Konfidenzintervall $68,2 \leq \mu \leq 73,65$.

```
> b <- rep(NA, 1000)
> for (i in 1:1000) b[i] <- mean(sample(x, 8, replace=TRUE))
> quantile(x, probs = c(0.025, 0.975))
 2.5% 97.5%
68.175 73.650
```

Das Bootstrap-Verfahren setzt somit den Einsatz eines geeigneten Computerprogrammes voraus. In **R** steht mit der Funktion **bootstrap()** (aus dem gleichnamigen Zusatzpaket von R. Tibshirani [TF05]) ein sehr flexibles Werkzeug zur Verfügung, mit dem die aufgeführten Schritte 1-3 für „beliebige“ Schätzfunktionen durchgeführt werden können. Ein Beispiel zur Bestimmung des **95%-Konfidenzintervalles für den Median** aus einer Stichprobe mit 15 Werten ist:

```
> library(bootstrap)
> x <- c(10, 10, 11, 12, 12, 13, 14, 15, 15, 16, 17, 20, 21, 24, 30)
> n <- length(x)
> boot <- bootstrap(x, 500, median) # Median aus 500 Stichproben
> quantile(boot$varthetastar, probs=c(.025, .975)) # Quantile der Verteilung
 2.5% 97.5%
 12 17
```

Das Ergebnis für das 95%-Konfidenzintervall aus den entsprechen Quantilen ist somit $12 \leq \tilde{\mu} \leq 17$.

Ein **Bootstrap-Standardfehler** der Schätzung kann aus der Standardabweichung der Bootstrap-Stichprobe berechnet werden:

$$SE_{boot, \hat{\vartheta}} = \sqrt{\frac{1}{K-1} \sum_{k=1}^K \left(\hat{\vartheta}_n^{*k} - \frac{1}{K} \sum_{k=1}^K \hat{\vartheta}_n^{*k} \right)^2} \quad (6.73)$$

Analog zum Ansatz der Bestimmung von Konfidenzintervallen bei Vorliegen einer Normalverteilung (z.B. für den Erwartungswert mit $\bar{x} \pm t_{n-1, 1-\alpha/2} \frac{s}{\sqrt{n}}$) können **Konfidenzintervalle** nach der **Bootstrap t-Methode** auch wie folgt definiert werden:

$$\hat{\vartheta} \pm t^* \cdot SE_{boot, \hat{\vartheta}} \quad (6.74)$$

Für t^* kann das entsprechende Quantil der t-Verteilung (vgl. Tabelle 5.10) mit $n-1$ Freiheitsgraden verwendet werden oder es wird eine t^* -Verteilung aus Bootstrap-Stichproben erzeugt ($t^{*k} = (\hat{\vartheta}_n^{*k} - \hat{\vartheta}) / SD_{boot, \hat{\vartheta}}$), deren Quantile dann in (6.74) eingesetzt werden. Für das Resampling dieser „empirischen“ t-Verteilung sollten mehr als 1000 ($K > 1000$) Stichproben gezogen werden.

Für diesen Ansatz steht im Programm **R** eine spezielle Funktion **boott()** in der **library(bootstrap)** zur Verfügung. In dem Beispiel wird wiederum der 95%-Konfidenzintervall für den Median aus einer Stichprobe mit 15 Werten bestimmt. Der Rechenaufwand ist erheblich höher gegenüber der Perzentilmethode, da hier 50·1000 Bootstrap-Stichproben erhoben und berechnet werden müssen.

```
> library(bootstrap)
> x <- c(10, 15, 20, 16, 13, 12, 15, 21, 11, 24, 17, 14, 12, 10, 30)
> boott(x, median, nbootsd=50, nboott=1000, perc=c(0.025, 0.975))
$confpoints
 0.025 0.975
[1,] 11.96863 20.71497
```

6.10 Konfidenzintervall für σ^2 bzw. σ

Konfidenzintervalle und Tests, die σ , σ^2 und σ_1^2/σ_2^2 betreffen, sind gegenüber Abweichungen von der Normalverteilung empfindlicher als Verfahren, die zweiseitige Konfidenzintervalle und Tests für μ und $\mu_1 - \mu_2$ (t -Verteilung) betreffen.

Das **Konfidenzintervall** für σ^2 lässt sich anhand der χ^2 -Verteilung nach

$$\frac{s^2(n-1)}{\chi_{n-1;\alpha/2}^2} \leq \sigma^2 \leq \frac{s^2(n-1)}{\chi_{n-1;1-\alpha/2}^2} \quad (6.75)$$

schätzen.

Beispiel: Wir erhalten für $n = 51$ und $s^2 = 2$ das 95%-Konfidenzintervall ($\alpha = 0,05$), d. h.

$$\chi_{50;0,025}^2 = 71,42 \text{ und } \chi_{50;0,975}^2 = 32,36:$$

$$\frac{2 \cdot 50}{71,42} \leq \sigma^2 \leq \frac{2 \cdot 50}{32,36}$$

$$1,40 \leq \sigma^2 \leq 3,09.$$

Den **Schätzwert** für σ^2 erhält man nach

$$\hat{\sigma}^2 = \frac{s^2(n-1)}{\chi_{n-1;0,5}^2} \quad \text{z. B. } \hat{\sigma}^2 = \frac{2 \cdot 50}{49,335} \simeq 2,03. \quad (6.76)$$

Mitunter ist der Bereich für die Standardabweichung erwünscht, d.h. für das Beispiel: $\sqrt{1,40} < \sigma < \sqrt{3,09}$; $1,18 < \sigma < 1,76$. Da die χ^2 -Verteilung unsymmetrisch ist, liegt der geschätzte Parameter (σ) nicht in der Mitte des Konfidenzintervalles. Man rechnet dann anhand von (6.77).

Liegt eine **Normalverteilung** vor, so berechnet man das Konfidenzintervall für σ aus einer Zufallsstichprobe des Umfangs n mit der Stichproben-Standardabweichung s nach:

$$\frac{s}{\sqrt{F_{n-1;\infty;\alpha/2}}} \leq \sigma \leq s \cdot \sqrt{F_{\infty;n-1;\alpha/2}} \quad (6.77)$$

Mit den oberen Schranken der F -Verteilung (vgl. Tabelle 5.12 und 5.13) benötigen wir z. B. für das 99%-Konfidenzintervall die oberen $(1 - 0,99)/2 = 0,005$ -Schranken, etwa für $n = 10$ mit $\bar{x} = 1,506$ und $s = 0,0942$:

$$F_{9;\infty;0,005} = 2,62 \quad \text{sowie} \quad F_{\infty;9;0,005} = 5,19$$

$$\text{d. h. } \frac{0,0942}{\sqrt{2,62}} \leq \sigma \leq 0,0942 \sqrt{5,19}$$

$$99\text{-Konfidenzintervall: } 0,058 \leq \sigma \leq 0,215.$$

Zum Vergleich sei das entsprechende deutlich engere 95%-Konfidenzintervall berechnet:

$$F_{9;\infty;0,025} = 2,11; F_{\infty;9;0,025} = 3,33$$

$$\frac{0,0942}{\sqrt{2,11}} \leq \sigma \leq 0,0942 \sqrt{3,33}$$

$$95\text{-KI: } 0,065 \leq \sigma \leq 0,172.$$

Im allgemeinen wird man das 95%-KI angeben. Dies entspricht einem Test auf dem 5%-Niveau bei zweiseitiger Fragestellung. Ist das 95%-KI, es wird gern nach außen gerundet, d. h. hier $0,06 \leq \sigma \leq 0,18$, zu weit, so muss der Stichprobenumfang n erhöht werden.



Hinweis: Für $n > 150$ gelten auch die auf der Normalverteilung basierenden Approximationen zum 95%-KI für σ^2 bzw. σ :

$$\begin{aligned} \frac{s^2}{1 + 1,96\sqrt{\frac{2}{n-1}}} &\lesssim \sigma^2 \lesssim \frac{s^2}{1 - 1,96\sqrt{\frac{2}{n-1}}} \\ \frac{s}{1 + \frac{1,96}{\sqrt{2(n-1)}}} &\lesssim \sigma \lesssim \frac{s}{1 - \frac{1,96}{\sqrt{2(n-1)}}} \end{aligned} \quad (6.78)$$

Für das 90%-Konfidenzintervall wird in (6.78) 1,96 durch 1,64 ersetzt, für das 99%-KI durch 2,58.

6.10.1 Konfidenzintervall für den Variationskoeffizienten γ

Die Vertrauensgrenzen des Variationskoeffizienten können nach Johnson und Welch [JW40] bestimmt werden. Für $n \gtrsim 25$ und $V < 0,4$ genügt die Approximation (6.79):

$$\frac{V}{1 + z \cdot \sqrt{\frac{1 + 2V^2}{2(n-1)}}} \lesssim \gamma \lesssim \frac{V}{1 - z \cdot \sqrt{\frac{1 + 2V^2}{2(n-1)}}} \quad (6.79)$$

90%-KI: $z = 1,64$; 95%-KI: $z = 1,96$; 99%-KI: $z = 2,58$.

Für die häufig interessierende (einseitige) obere Vertrauensgrenze (VG_0) (6.79 rechts) γ_0 benötigt man 90%- VG_0 : $z = 1,28$; 95%- VG_0 : $z = 1,64$; 99%- VG_0 : $z = 2,33$.

Beispiel: Berechne das 90%-KI für γ anhand von $n = 25$ und $V = 0,30$.

$$1,64\sqrt{(1 + 2 \cdot 0,3^2)/[2(25 - 1)]} = 0,257$$

$$0,3/1,257 = 0,239 \quad 0,3/0,743 = 0,404; \quad 90\text{-KI: } 0,24 \lesssim \gamma \lesssim 0,40$$

0,40 ist zugleich die angenäherte obere 95%-VG, d. h. 95%- VG_0 : $\gamma_0 \approx 0,40$; der Variationskoeffizient γ liegt mit einer Vertrauenswahrscheinlichkeit von 95% unter 0,40.

6.10.2 Konfidenzintervall für den Quotienten zweier Varianzen σ_1^2/σ_2^2

Das 95%-Konfidenzintervall für das Verhältnis zweier Varianzen σ_1^2/σ_2^2 anhand zweier auf den Freiheitsgraden $\nu_1 = n_1 - 1$ und $\nu_2 = n_2 - 1$ basierender Stichprobenvarianzen s_1^2 und s_2^2 angenähert normalverteilter Grundgesamtheiten erhält man mit Hilfe der oberen 2,5%-Quantile der F-Verteilung (vgl. Tabelle 5.13) nach

$$\frac{s_1^2}{s_2^2} \cdot \frac{1}{F_{\nu_1, \nu_2}} \leq \sigma_1^2/\sigma_2^2 \leq F_{\nu_2, \nu_1} \cdot \frac{s_1^2}{s_2^2} \quad (6.80)$$

Für den 90%-KI verwende man entsprechend die oberen 5%-Quantile der F-Verteilung.

6.10.3 Mindestzahl von Beobachtungen zur Schätzung einer Standardabweichung

Formel (6.81) gibt mit vorgegebener Genauigkeit (d) und vorgegebener statistischer Sicherheit **minimale Stichprobenumfänge** an (auf der Normalverteilung basierende Näherungen; ergänzt 6.61) zur Schätzung der Standardabweichung (n_s):

$$n_s \approx 1 + 0,5 \left(\frac{z_\alpha}{d} \right)^2 \quad \text{mit } d = \frac{(s - \sigma)}{\sigma} \quad (6.81)$$

$z_{1-\alpha/2}$ ist das entsprechende Quantil der Standardnormalverteilung für die gewünschte Vertrauenswahrscheinlichkeit $1 - \alpha$ (d. h. die Irrtumswahrscheinlichkeit α). Für die Beispiele benutzen wir $z_{0,975} = 1,96$ für $\alpha = 0,05$ und $z_{0,995} = 2,58$ für $\alpha = 0,01$.



Beispiel: Zur Schätzung einer Standardabweichung mit einer Vertrauenswahrscheinlichkeit von 95% ($\alpha = 0,05$) und einer Genauigkeit von $d = 0,2$ benötigt man etwa $n_s \approx 1 + 0,5(1,96/0,2)^2 = 49$ Beobachtungen. Für $\alpha = 0,05$ und $d = 0,14$ benötigt man etwa $n_s \approx 1 + 0,5(1,96/0,14)^2 = 99$ Beobachtungen. Tabelle 6.7 liefert $n_s = 100$. Exakter als (6.81) ist (z. B. für dieses Beispiel): $n_s \approx 1 + 0,5[1,960/(\ln 1,14)]^2 = 113$. Dieser Ansatz ist brauchbar, sobald $n_s > 15$ resultiert. Häufig begnügt man sich mit $\alpha = 0,10$ (d.h. $P = 0,90$ und $z_{0,95} = 1,645$).

Tabelle 6.7. Die halbe Länge des Konfidenzintervalles für den relativen Fehler der Standardabweichung $[(s - \sigma)/\sigma]$ einer normalverteilten Grundgesamtheit für ausgewählte Irrtumswahrscheinlichkeiten α und Stichprobenumfänge n_s . Vergleiche das zweite Beispiel zu Formel (6.81). (Aus Thompson, W. A., Jr. and J. Endriss: The required Sample size when estimating variances. The American Statistician **15** (June 1961) 22–23, p. 22, Table I)

n_s	α			
	0,01	0,05	0,10	0,20
4	0,96	0,75	0,64	0,50
6	0,77	0,60	0,50	0,40
8	0,66	0,51	0,43	0,34
10	0,59	0,45	0,38	0,30
12	0,54	0,41	0,35	0,27
15	0,48	0,37	0,31	0,24
20	0,41	0,32	0,27	0,21
25	0,37	0,28	0,24	0,18
30	0,34	0,26	0,22	0,17
100	0,18	0,14	0,12	0,09
1000	0,06	0,04	0,04	0,03

6.11 Konfidenzintervall für den Erwartungswert λ einer Poisson-Verteilung

Es gibt zwei Arten von Konfidenzintervallen (Kle) für λ ; Man verwende entweder (1) oder (2), nicht aber beide gemeinsam:

- (1) **Nicht-zentrale (kürzeste) Kle** nach Crow und Gardner [CG59], die für $x \leq 300$ als 95%-KI bzw. als 99%-KI der Tabelle 6.8 entnommen werden können.



Beispiel: In der **Beobachtungseinheit** von 8 Stunden seien 26 Ereignisse registriert worden. Die 95%-Grenzen ($x = 26$) für (a) die Beobachtungseinheit sind $16,77 \simeq 17$ und $37,67 \simeq 38$ Ereignisse und für (b) eine Stunde sind $16,77/8 \simeq 2$ und $37,67/8 \simeq 5$ Ereignisse.

- (2) **Zentrale Kle**, die sich nach (6.82) berechnen und nach (6.83) approximieren lassen, hier als 90%-Kle geschrieben; die entsprechenden 95%- und 99%-Kle lassen sich anhand der Schranken der Tabellen 50 und 24 oder 29 bilden, etwa der 95%-KI für λ anhand von $x = 10$: $\chi_{20;0,975}^2 = 9,59$ und $\chi_{22;0,025}^2 = 36,78$, d. h. 95%-KI: $4,80 \leq \lambda \leq 18,39$.

$$90\%-KI: \frac{1}{2}\chi_{0,95;2x}^2 \leq \lambda \leq \frac{1}{2}\chi_{0,05;2(x+1)}^2 \quad (6.82)$$

$$90\%-KI: \left(\frac{1,645}{2} - \sqrt{x} \right)^2 \lesssim \lambda \lesssim \left(\frac{1,645}{2} + \sqrt{x+1} \right)^2 \quad (6.83)$$

Rechts in (6.82) und (6.83) stehen zugleich die (einseitigen) oberen 95%-Vertrauensgrenzen: So ist z. B. für $x = 50$ nach (6.82) $2(50+1) = 102$, $\chi_{0,05;102}^2 = 126,57$ d. h. $\lambda \leq 63,3$ und nach (6.83) $(1,645/2 + \sqrt{50+1})^2 = 63,4$, d. h. $\lambda \lesssim 63,4$. Entsprechend erhält man auch z. B. die oberen 90%-Vertrauensgrenzen (6.82): mit $\chi_{0,10}^2$ anstatt $\chi_{0,05}^2$; (6.83): mit 1,282 anstatt 1,645 (entsprechende Quantile der Standardnormalverteilung).

Tabelle 6.8. Konfidenzintervalle für den Erwartungswert einer Poisson-Verteilung (auszugsweise entnommen aus E.L. Crow und R.S. Gardner: Confidence intervals for the expectation of a Poisson variable, Biometrika **46** (1959) 441–453). Diese Tabelle gestattet nicht die Angabe einseitiger Vertrauensgrenzen.

x	95	99	x	95	99	x	95	99						
0	0	3,285	0	4,771	100	80,25	120,36	76,61	127,31	200	172,38 ⁵	227,73	164,31	238,01
1	0,051	5,323	0,010	6,914	101	81,61	121,06	76,61	128,70	201	173,79	228,99	165,33	239,46
2	0,355	6,686	0,149	8,727	102	83,14	122,37	77,15	130,27 ⁵	202	175,48 ⁵	230,28	166,71	241,32
3	0,818	8,102	0,436	10,473	103	84,57	123,77	78,71	131,50	203	176,23	231,65	168,29	241,32
4	1,366	9,598	0,823	12,347	104	84,57	125,46	80,06	131,82	204	176,23	233,19	169,49	242,01
5	1,970	11,177	1,279	13,793	105	84,67	126,26	80,06	133,21	205	176,23	234,53	169,49	243,31 ⁵
6	2,613	12,817	1,785	15,277	106	86,01	126,48	80,65	134,79	206	177,48	234,53	169,64	244,69
7	3,285	13,765	2,330	16,801	107	87,48	127,78	82,21	135,99	207	178,77	235,14 ⁵	170,98	246,24
8	3,285	14,921	2,906	18,362	108	89,23	129,14	83,56	136,30	208	180,14	236,39	172,41	247,54 ⁵
9	4,460	16,768	3,507	19,462	109	89,23	130,68	83,56	137,68	209	181,67	237,67	174,36	247,54 ⁵
10	5,323	17,633	4,130	20,676	110	89,23	132,03	84,12	139,24	210	183,05	239,00	174,36	248,62
11	5,323	19,050	4,771	22,042	111	90,37	132,03	85,65	140,54	211	183,05	240,45	174,36	249,94
12	6,686	20,335	4,771	23,765	112	91,78	133,14 ⁵	87,12	140,76	212	183,05	242,27	175,25	251,35
13	6,686	21,364	5,829	24,925	113	93,48	134,48	87,12	142,12	213	183,86	242,27	176,61	253,14
14	8,102	22,945	6,668	25,992	114	94,23	135,92	87,55	143,64	214	185,13	242,53	178,11	253,65
15	8,102	23,762	6,914	27,718	115	94,23	137,79	89,05	145,13	215	186,46	243,76	179,67	253,92
16	9,598	25,400	7,756	28,852	116	94,70 ⁵	137,79	90,72	145,19	216	187,89	245,02	179,67	255,20
17	9,598	26,306	8,727	29,900	117	96,06	138,49	90,72	146,54	217	189,83	246,32 ⁵	179,67	256,54
18	11,177	27,735	8,727	31,839	118	97,54 ⁵	139,79	90,96	148,01	218	189,83	247,70	180,84	258,00
19	11,177	28,966	10,009	32,547	119	99,17	141,16	92,42	149,76	219	189,83	249,28	182,22	259,78
20	12,817	30,017	10,473	34,183	120	99,17	142,70	94,34 ⁵	149,76	220	190,21	250,43	183,81	259,78
21	12,817	31,675	11,242	35,204	121	99,17	144,01	94,34 ⁵	150,93	221	191,46	250,43	184,97 ⁵	260,47
22	13,765	32,277	12,347	36,544	122	100,32	144,01	94,35	152,35 ⁵	222	192,76	251,11	184,97 ⁵	261,77
23	14,921	34,048	12,347	37,819	123	101,71	145,08	95,76	154,18	223	194,11 ⁵	252,35	185,08	263,12 ⁵
24	14,921	34,665	13,793	38,939	124	103,31 ⁵	146,39	97,42	154,60	224	195,63	253,63	186,40	264,63
25	16,768	36,030	13,793	40,373	125	104,40	147,80	98,36	155,31	225	197,09	254,95	187,81	266,15
26	16,77	37,67	15,28	41,39	126	104,40	149,53	98,36	156,69	226	197,09	256,37	189,50	266,15
27	17,63	38,16 ⁵	15,28	42,85	127	104,58	150,19	99,09	158,25	227	197,09	258,34	190,28	267,01
28	19,05	39,76	16,80	43,91	128	105,90 ⁵	150,36	100,61	159,53	228	197,78	258,34	190,28	268,31
29	19,05	40,94	16,80	45,26	129	107,32	151,63	102,16 ⁵	159,67	229	199,04	258,45	190,61 ⁵	269,68
30	20,33 ⁵	41,75	18,36	46,50	130	109,11	152,96	102,16 ⁵	161,01	230	200,35	259,67	191,94	271,22
31	21,36	43,45	18,36	47,62	131	109,61	154,39	102,42	162,46	231	201,73	260,92	193,36	272,56
32	21,36	44,26	19,46	49,13	132	109,61	156,32	103,84	164,31	232	203,35 ⁵	262,20	195,19	272,56
33	22,94 ⁵	45,28	20,28 ⁵	49,96	133	110,11	156,32	105,66	164,31	233	204,36	263,54	195,59	273,53
34	23,76	47,02 ⁵	20,68	51,78	134	111,44	156,87	106,12	165,33	234	204,36	265,00	195,59	274,83
35	23,76	47,69	22,04	52,28	135	112,87	158,15	106,12	166,71	235	204,36	266,71	196,13	276,20 ⁵
36	25,40	48,74	22,04	54,03	136	114,84	159,48	107,10	168,29	236	205,31 ⁵	266,71	197,46	277,77
37	26,31	50,42	23,76 ⁵	54,74	137	114,84	160,92 ⁵	108,61 ⁵	169,49	237	206,58	266,97	198,88	279,01 ⁵
38	26,31	51,29	23,76 ⁵	56,14	138	114,84	162,79	110,16	169,64	238	207,90	268,19	200,84	279,01 ⁵
39	27,73 ⁵	52,15	24,92 ⁵	57,61 ⁵	139	115,60 ⁵	162,79	110,16	170,98	239	209,30	269,44	200,94	280,02
40	28,97	53,72	25,83	58,35	140	116,93	163,35	110,37	172,41	240	211,03	270,73	200,94	281,32
41	28,97	54,99	25,99	60,39	141	118,35	164,63	111,78	174,36	241	211,69	272,08	201,62	282,70
42	30,02	55,51	27,72	60,59	142	120,36	165,96	113,45	174,36	242	211,69	273,57	202,94	284,25
43	31,67 ⁵	56,99	27,72	62,13	143	120,36	167,39	114,33	175,25	243	211,69	275,15	204,36	285,53
44	31,67 ⁵	58,72	28,85	63,63 ⁵	144	120,36	169,33	114,33	176,61	244	212,82	275,15	206,19	285,53
45	32,28	58,84	29,90	64,26	145	121,06	169,33	114,99	178,11	245	214,09	275,46	206,60	286,50
46	34,05	60,24	29,90	65,96	146	122,37	169,80	116,44	179,67	246	215,40	276,69	206,60	287,79
47	34,66 ⁵	61,90	31,84	66,81 ⁵	147	123,77	171,07	118,33	179,67	247	216,81	277,94	207,08	289,16
48	34,66 ⁵	62,81	31,84	67,92	148	125,46	172,38 ⁵	118,33	180,84	248	218,56	279,22	208,40	290,68
49	36,03	63,49	32,55	69,83	149	126,26	173,79	118,33	182,22	249	219,16	280,57	209,81	292,10
50	37,67	64,95	38,18	70,05	150	126,26	175,48 ⁵	119,59	183,81	250	219,16	282,05	211,50	292,10
51	37,67	66,76	34,18	71,56	151	126,48	176,23	121,09	184,97 ⁵	251	219,16	283,67	212,29	292,95
52	38,16 ⁵	66,76	35,20	73,20	152	127,78	176,23	122,69	185,08	252	220,29	283,67	212,29	294,24
53	39,76	68,10	36,54	73,62	153	129,14	177,48	122,69	186,40	253	221,56	283,93	212,53	295,59
54	40,94	69,62	36,54	75,16	154	130,68	178,77	122,78	187,81	254	222,86 ⁵	285,15	213,84	297,07
55	40,94	71,09	37,82	76,61	155	132,03	180,14	124,16	189,50	255	224,26	286,40	215,22	298,71
56	41,75	71,28	38,94	77,15	156	132,03	181,67	125,70	190,28	256	225,90 ⁵	287,68	216,80	298,71
57	43,45	72,66	38,94	78,71	157	132,03	183,05	127,07	190,61 ⁵	257	226,81	289,01	217,98	299,39
58	44,26	74,22	40,37	80,06	158	133,14 ⁵	183,05	127,07	191,94	258	226,81	290,46	217,98	300,67
59	44,26	75,49	41,39	80,65	159	134,48	183,86	127,31	193,36	259	226,81	292,26	217,98	302,00
60	45,28	75,78 ⁵	41,39	82,21	160	135,92	185,13	128,70	195,19	260	227,73	292,26	219,25	303,43
61	47,02 ⁵	77,16	42,85	83,56	161	137,79	186,46	130,27 ⁵	195,59	261	228,99	293,37	220,61	305,35
62	47,69	78,73	43,91	84,12	162	137,79	187,89	131,50	196,13	262	230,28	293,59	222,10 ⁵	305,35
63	47,69	79,98	43,91	85,65	163	137,79	189,83	131,50	197,46	263	231,65	294,82 ⁵	223,67 ⁵	305,81
64	48,74	80,25	45,26	87,12	164	138,49	189,83	131,82	198,88	264	233,19	296,09	223,67 ⁵	307,07
65	50,42	81,61	46,50	87,55	165	139,79	190,21	133,21	200,84	265	234,53	297,41	223,67 ⁵	308,38
66	51,29	83,14	46,50	89,05	166	141,16	191,46	134,79	200,94	266	234,53	298,81	224,65	309,77 ⁵
67	51,29	84,57	47,62	90,72	167	142,70	192,76	135,99	201,62	267	234,53	300,56	225,98	311,41
68	52,15	84,67	49,13	90,96	168	144,01	194,11 ⁵	135,99	202,94	268	235,14 ⁵	301,16	227,41	312,38

Tabelle 6.8. (Fortsetzung)

x	95	99	x	95	99	x	95	99
69	53,72	86,01	49,13	92,42	169	144,01	195,63	136,30
70	54,99	87,48	49,96	94,34 ⁵	170	144,01	197,09	137,68
71	54,99	89,23	51,78	94,35	171	145,08	197,09	139,24
72	55,51	89,23	51,78	95,76	172	146,39	197,78	140,54
73	56,99	90,37	52,28	97,42	173	147,80	199,04	140,54
74	58,72	91,78	54,03	98,36	174	149,53	200,35	140,76
75	58,72	93,48	54,74	99,09	175	150,19	201,73	142,12
76	58,84	94,23	54,74	100,61	176	150,19	203,35 ⁵	143,64
77	60,24	94,70 ⁵	56,14	102,16 ⁵	177	150,36	204,36	145,13
78	61,90	96,06	57,61 ⁵	102,42	178	151,63	204,36	145,13
79	62,81	97,54 ⁵	57,61 ⁵	103,84	179	152,96	205,31 ⁵	145,19
80	61,81	99,17	58,35	105,66	180	154,39	206,58	146,54
81	63,49	99,17	60,39	106,12	181	156,32	207,90	148,01
82	64,95	100,32	60,39	107,10	182	156,32	209,30	149,76
83	66,76	101,71	60,59	108,61 ⁵	183	156,32	211,03	149,76
84	66,76	103,31 ⁵	62,13	110,16	184	156,87	211,69	149,76
85	66,76	104,40	63,63 ⁵	110,37	185	158,15	211,69	150,93
86	68,10	104,58	63,63 ⁵	111,78	186	159,48	212,82	152,35 ⁵
87	69,62	105,90 ⁵	64,26	113,45	187	160,92 ⁵	214,09	154,18
88	71,09	107,32	65,96	114,33	188	162,79	215,40	154,60
89	71,09	109,11	66,81 ⁵	114,99	189	162,79	216,81	154,60
90	71,28	109,61	66,81 ⁵	116,44	190	162,79	218,56	155,31
91	72,66	110,11	67,92	118,33	191	163,35	219,16	156,69
92	74,22	111,44	69,83	118,33	192	164,63	219,16	158,25
93	75,49	112,87	69,83	119,59	193	165,96	220,29	159,53
94	75,49	114,84	70,05	121,09	194	167,39	221,56	159,53
95	75,78 ⁵	114,84	71,56	122,69	195	169,33	222,86 ⁵	159,67
96	77,16	115,60 ⁵	73,20	122,78	196	169,33	224,26	161,01
97	78,73	116,93	73,20	124,16	197	169,33	225,90 ⁵	162,46
98	79,98	118,35	73,62	125,70	198	169,80	226,81	164,31
99	79,98	120,36	75,16	127,07	199	171,07	226,81	164,31
100	80,25	120,36	76,61	127,31	200	172,38 ⁵	227,73	164,31



Beispiel: Für ein bestimmtes Gebiet seien in einem Jahrhundert vier Sturmfluten beobachtet worden. Angenommen, die Zahl der Sturmfluten in verschiedenen Jahrhunderten folge einer Poisson-Verteilung, dann kann damit gerechnet werden, dass nur in einem von 20 Jahrhunderten ($P = 0,95$; Tabelle 6.8) die Zahl der Sturmfluten außerhalb der Grenzen $1,366 \approx 1$ und $9,598 \approx 10$ liegen wird; d. h. 95%-KI: $1 \lesssim \lambda \lesssim 10$.



Beispiel: Eine Telefonzentrale erhalte während einer Minute 23 Anrufe. Gewünscht sind die 95%-Vertrauensgrenzen für die erwarteten Anrufe in 1 Min. bzw. in 1 Stunde. Nehmen wir an, dass die Zahl der Anrufe im betrachteten Zeitraum relativ konstant ist und (da die Anlage sagen wir 1000 Anrufe/min vermitteln kann) einer Poisson-Verteilung folgt, dann sind die 95%-Vertrauensgrenzen für 1 Minute (nach Tabelle 6.8) $14,921 \approx 15$ und $34,048 \approx 34$. In einer Stunde ist mit $60 \cdot 14,921 \approx 895$ bis $60 \cdot 34,048 \approx 2043$ Anrufen zu rechnen ($P = 0,95$); d. h. 95%-KI: $15 \lesssim \lambda_{1 \text{ min}} \lesssim 34$ bzw. $895 \lesssim \lambda_{1 \text{ h}} \lesssim 2043$.



Hinweis: Tabelle 6.8 dient auch zur Prüfung der Nullhypothese: $\lambda = \lambda_x$ (λ ist vorgegeben; x ist die beobachtete Erfolgszahl, λ_x ist der zugehörige Parameter). Wenn das KI für λ_x den Parameter λ nicht überdeckt, wird die Nullhypothese: $\lambda = \lambda_x$ zugunsten der Alternativhypothese $\lambda \neq \lambda_x$ verworfen.

Obere Vertrauensgrenzen für Lambda für die Vertrauenswahrscheinlichkeiten $P = 0,75$ bis $0,999$ und für $x = 0$ bis 50 Ereignisse [d. h. z. B. $P(X > 0 | \lambda = 3) = 0,95$] enthält Tabelle 6.9.



Beispiel: $P(X > 31 | \lambda = 44,00) = 0,975$ sowie $P(X > 7 | \lambda = 16,00) = 0,990$.

Tabelle 6.9. Obere Vertrauensgrenzen für Lambda für die Vertrauenswahrscheinlichkeiten $P = 0,75$ bis $0,999$ und für $x \leq 50$

x	$P:$	0,750	0,900	0,950	0,975	0,990	0,995	0,999
0		1,39	2,30	3,00	3,69	4,61	5,30	6,91
1		2,69	3,89	4,74	5,57	6,64	7,43	9,23
2		3,92	5,32	6,30	7,22	8,41	9,27	11,23
3		5,11	6,68	7,75	8,77	10,05	10,98	13,06
4		6,27	7,99	9,15	10,24	11,60	12,59	14,79
5		7,42	9,27	10,51	11,67	13,11	14,15	16,45
6		8,56	10,53	11,84	13,06	14,57	15,66	18,06
7		9,68	11,77	13,15	14,42	16,00	17,13	19,63
8		10,80	12,99	14,43	15,76	17,40	18,58	21,16
9		11,91	14,21	15,71	17,08	18,78	20,00	22,66
10		13,02	15,41	16,96	18,39	20,14	21,40	24,13
11		14,12	16,60	18,21	19,68	21,49	22,78	25,59
12		15,22	17,78	19,44	20,96	22,82	24,14	27,03
13		16,31	18,96	20,67	22,23	24,14	25,50	28,45
14		17,40	20,13	21,89	23,49	25,45	26,84	29,85
15		18,49	21,29	23,10	24,74	26,74	28,16	31,24
16		19,57	20,45	24,30	25,98	28,03	29,48	32,62
17		20,65	23,61	25,50	27,22	29,31	30,79	33,99
18		21,73	24,76	26,69	28,45	30,58	32,09	35,35
19		22,81	25,90	27,88	29,67	31,85	33,38	36,70
20		23,88	27,05	29,06	30,89	33,10	34,67	38,04
21		24,96	28,18	30,24	32,10	34,35	35,95	39,37
22		26,03	29,32	31,41	33,31	35,60	37,22	40,70
23		27,10	30,45	32,59	34,51	36,84	38,48	42,02
24		28,17	31,58	33,75	35,71	38,08	39,75	43,33
25		29,23	32,71	34,92	36,90	39,31	41,00	44,64
26		30,30	33,84	36,08	38,10	40,53	42,25	45,94
27		31,36	34,96	37,23	39,28	41,76	43,50	47,23
28		32,43	36,08	38,39	40,47	42,98	44,74	48,52
29		33,49	37,20	39,54	41,65	44,19	45,98	49,80
30		34,55	38,32	40,69	42,83	45,40	47,21	51,08
31		35,61	39,43	41,84	44,00	46,61	48,44	52,36
32		36,67	40,54	42,98	45,17	47,81	49,67	53,63
33		37,73	41,65	44,13	46,34	49,01	50,89	54,90
34		38,79	42,76	45,27	47,51	50,21	52,11	56,16
35		39,85	43,87	46,40	48,68	51,41	53,32	57,42
36		40,90	44,98	47,54	49,84	52,60	54,54	58,67
37		41,96	46,08	48,68	51,00	53,79	55,75	59,93
38		43,01	47,19	49,81	52,16	54,98	56,96	61,17
39		44,07	48,29	50,94	53,31	56,16	58,16	62,42
40		45,12	49,39	52,07	54,47	57,35	59,36	63,66
41		46,17	50,49	53,20	55,62	58,53	60,56	64,90
42		47,22	51,59	54,32	56,77	59,71	61,76	66,14
43		48,27	52,69	55,45	57,92	60,88	62,96	67,37
44		49,33	53,78	56,57	59,07	62,06	64,15	68,60
45		50,38	54,88	57,69	60,21	63,23	65,34	69,83
46		51,42	55,97	58,82	61,36	64,40	66,53	71,06
47		52,47	57,07	59,94	62,50	65,57	67,72	72,28
48		53,52	58,16	61,05	63,64	66,74	68,90	73,51
49		54,57	59,25	62,17	64,78	67,90	70,08	74,72
50		55,62	60,34	63,29	65,92	69,07	71,27	75,94

Beispiele siehe Seite 282

6.12 Weibull-Verteilung

6.12.1 Bestimmung der Parameter

Schätzungen für die Weibull-Parameter α (scale) und β (shape) ergeben sich durch die Bestimmung der Ausgleichsgeraden in einem linearisierten **Weibull-Diagramm**. Die Koordinaten der Punkte lassen sich durch Umstellen der 2-parametrischen Weibull-Funktion bestimmen. Wird die Verteilung in der Form

$$F(t) = 1 - e^{-\left(\frac{t}{\alpha}\right)^\beta} \quad (6.84)$$

geschrieben, dann lautet die **linearisierende Transformation**:

$$\begin{aligned} x &= \log(t) \\ y &= \log\left(\log\left(\frac{1}{1 - F(t)}\right)\right) \end{aligned} \quad (6.85)$$

Die **empirische Verteilung** kann durch die Näherungsformeln (6.86) bestimmt werden. Dazu werden die Ränge zu den aufsteigend sortierten Stichprobenwerten $t_{(1)} \leq t_{(2)} \leq \dots \leq t_{(n)}$ verwendet:

$$\begin{aligned} F(t_{(i)}) &= \frac{i - 0,3}{n + 0,4} \quad \text{für } n < 50 \\ F(t_{(i)}) &= \frac{i}{n + 1} \quad \text{für } n \geq 50 \end{aligned} \quad (6.86)$$

Die **Verteilungsfunktion** kann dann im linearisierten Maßstab durch (6.87) charakterisiert werden.

$$y = -\beta \cdot \log(\alpha) + \beta \cdot x \quad (6.87)$$

Für den Fall, dass die Beobachtungen im Rahmen einer Zufallsstichprobe tatsächlich aus einer Weibull-Verteilung stammen, müssen die Punkte der zugehörigen Wertepaare im Koordinatensystem **auf einer Linie** liegen. Die Schätzung $\hat{a}$ für den **Achsenabschnitt** und $\hat{b}$ für die **Steigung** im linearen Modell $y = a + bx$ (kleinste Abweichungsquadrate) führt somit direkt auf die Schätzung der Parameter der Weibull-Verteilung mit:

$$\begin{aligned} y &= a + bx \Rightarrow \text{Schätzung von } \hat{a} \text{ und } \hat{b} \\ \hat{\beta} &= \hat{b} \\ \hat{\alpha} &= \exp\left(-\frac{\hat{a}}{\hat{b}}\right) \end{aligned} \quad (6.88)$$



Beispiel: Zu der Prüfung eines Garns auf „Scheuertüchtigkeit“ wurde die Anzahl der Scheuerzyklen bis zum Bruch beobachtet (Graf, Henning und Wilrich [GHW74]).

550	760	830	890	1100	1150	1200	1350	1400	1600
1700	1750	1800	1850	1850	2200	2400	2850	3200	

Die Analyse der Daten nach (6.88) in **R** zeigt folgendes Ergebnis:

```

> garn <- c(550, 760, 830, 890, 1100, 1150, 1200, 1350, 1400, 1600,
+          1700, 1750, 1800, 1850, 1850, 2200, 2400, 2850, 3200)
> garn <- sort(garn); n <- length(garn)
> F <- (rank(garn) - 0.3) / (n+0.4) # empirische Verteilungsfunktion
> x <- log(garn) # Transformation
> y <- log(log(1/(1-F)))
> z <- lm(y ~ x); z # lineare Regression
Call: lm(formula = y ~ x) Coefficients: (Intercept) x
-18.813 2.509
> coef(z)[2] # shape
x
2.508568
> exp(-(coef(z)[1]/coef(z)[2])) # scale
(Intercept)
1807.446

```

Die Werte der Stichprobe sind im **Weibull-Diagramm** in Abbildung 6.4 dargestellt. Die Berechnung mit dem Programm **R** führt zu den Schätzungen $\hat{\alpha} = 1807,45$ und $\hat{\beta} = 2,51$. Eine Maximum-Likelihood-Schätzung der beiden Parameter mit der Funktion **mle()** in **R** führt zu vergleichbaren Ergebnissen:

```

> library(mle)
> ll <- function(shape=1.5, scale=2000)
+   -sum(dweibull(garn, shape, scale, log = TRUE))
> mle(ll)
Call: mle(minuslogl = ll)
Coefficients:
  shape      scale
2.549477 1893.728286

```

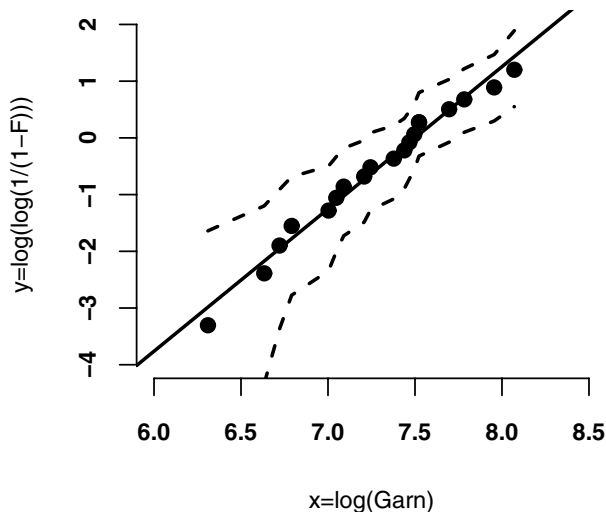


Abb. 6.4. Weibull-Diagramm zur Schauerfestigkeit eines Garns; Weibull-Gerade zu den geschätzten Parametern und 95%-Konfidenzintervall

6.12.2 Das Konfidenzintervall für die Weibull-Gerade

Die Weibull-Analyse erfolgt in der Regel auf der Basis einer Stichprobe. Die Gerade im Weibull-Diagramm (vgl. Abbildung 6.4) ist somit abhängig von einer speziellen Stichprobe, insbesondere von der Anzahl und von der Streuung der Werte (Versuche). Die Punkte „streuen“ um die

Weibull-Gerade. Durch das Konfidenzintervall soll eine Abschätzung für die Gerade über den Bereich der Grundgesamtheit gemacht werden. Da im Rahmen der Schätzung beide Parameter der Weibull-Verteilung einem Schätzfehler unterliegen, verlaufen die Vertrauensgrenzen nicht parallel zur Weibull-Geraden, sondern sie laufen im unteren und oberen Bereich mehr oder weniger auseinander. Eine Berechnung der Vertrauensgrenzen kann mit (6.89) erfolgen. Dabei werden die Rangzahlen i zu den Werten der Stichprobe und die entsprechenden Quantile der Fisher-Verteilung eingesetzt. Das Ergebnis für den 95%-KI, nach Transformation (6.85), ist durch die gestrichelten Linien in Abbildung 6.4 dargestellt.

$$\begin{aligned} V_{i,unten} &= \frac{1}{\frac{n-i+1}{i} F_{2(n-i+1), 2i, \alpha/2} + 1} \\ V_{i,oben} &= 1 - \frac{1}{1 + \frac{i}{n-i+1} F_{2i, 2(n-i+1), \alpha/2}} \end{aligned} \quad (6.89)$$

6.13 Konfidenzintervalle für die Parameter einer linearen Regression

- Die Schätzung einiger Standardabweichungen
- Konfidenzintervalle für den Regressionskoeffizienten, für den Achsenabschnitt und für die Restvarianz
- Konfidenzintervalle und Prädiktionsintervalle für die Regressionsgerade
- Inverse Prädiktion aus einer linearen Regression
- Das Konfidenzintervall für den Korrelationskoeffizienten ρ

6.13.1 Die Schätzung einiger Standardabweichungen

Die Standardabweichungen s_x und s_y werden über die Summen der Abweichungsquadrate der Variablen x und y ermittelt:

$$\begin{aligned} Q_x &= \sum (x - \bar{x})^2 = \sum x^2 - (\sum x)^2/n \text{ und } Q_y = \sum (y - \bar{y})^2 = \sum y^2 - (\sum y)^2/n \\ s_x &= \sqrt{\frac{Q_x}{n-1}} \quad s_y = \sqrt{\frac{Q_y}{n-1}} \end{aligned}$$

Jede Beobachtung einer bivariaten oder zweidimensionalen Häufigkeitsverteilung besteht aus einem Paar von Beobachtungswerten (x, y) . Das Produkt der beiden Abweichungen vom jeweiligen Mittelwert ist daher ein geeignetes Maß für den Grad des „Miteinandervariierens“ der Beobachtungen:

$$Q_{xy} = \sum (x - \bar{x})(y - \bar{y})$$

Das „mittlere“ Abweichungsprodukt s_{xy} ist die **empirische Kovarianz**:

$$\frac{\sum (x - \bar{x})(y - \bar{y})}{n-1} = \frac{Q_{xy}}{n-1} = s_{xy} \quad (6.90)$$

Diese schätzt die **Kovarianz** σ_{xy} . Die Berechnung der Summe der Abweichungsprodukte, kurz Q_{xy} genannt, wird durch folgende Identitäten erleichtert:

$$Q_{xy} = \sum xy - \bar{x} \sum y$$

$$Q_{xy} = \sum xy - \bar{y} \sum x$$

$$Q_{xy} = \sum xy - \frac{\sum x \sum y}{n} \quad (6.91)$$



(6.91) ist rechentechnisch am günstigsten. Über Q_{xy} erhält man die Schätzung für den Korrelationskoeffizienten $\hat{\rho} = r$ sowie für die beiden Regressionskoeffizienten $\hat{\beta}_{yx} = b_{yx}$ und $\hat{\beta}_{xy} = b_{xy}$ nach

$$r = \frac{Q_{xy}}{\sqrt{Q_x \cdot Q_y}} = \frac{s_{xy}}{s_x \cdot s_y} = \frac{1}{n-1} \sum \left(\frac{x - \bar{x}}{s_x} \right) \left(\frac{y - \bar{y}}{s_y} \right) \quad (6.92)$$

Dem Summenzeichen in (6.92) rechts folgt dann, wenn X und Y normalverteilt sind, das Produkt zweier Standardnormalvariablen (**normierte Produktsumme von Standardnormalvariablen**). Die Regressionskoeffizienten erhält man nach:

$$b_{yx} = \frac{Q_{xy}}{Q_x} = \frac{s_{xy}}{s_x^2} = r \frac{s_y}{s_x} \quad (6.93)$$

$$b_{xy} = \frac{Q_{xy}}{Q_y} = \frac{s_{xy}}{s_y^2} = r \frac{s_x}{s_y} \quad (6.94)$$

Die Standardabweichung für Y unter der Bedingung, dass X bestimmte Werte annimmt, ist

$$s_{y \cdot x} = \sqrt{\frac{\sum (y - a_{yx} - b_{yx}x)^2}{n-2}} = \sqrt{\frac{\sum (y - \hat{y})^2}{n-2}} \quad (6.95)$$

Hinweis:

$$\left[\begin{array}{l} \text{MSE als unverzerrter Schätzer} \\ \text{der Restvarianz} \end{array} \quad \text{MSE} = \sum_{i=1}^n \frac{(Y_i - \hat{Y})^2}{n-2} \right]$$

Das Symbol $s_{y \cdot x}$, die **Standardabweichung der $\hat{y}$ -Werte** für ein gegebenes x , wird gelesen „*sy* Punkt x “. Der Zähler unter der Wurzel stellt die Summe der Quadrate der Abweichungen der beobachteten y -Werte von den entsprechenden Werten auf der Regressionsgeraden dar. Diese Summe wird durch $n-2$ und nicht durch $n-1$ dividiert, da wir aus den Daten *zwei* Kennwerte a_{yx} und b_{yx} geschätzt haben. Der Wert $s_{y \cdot x}$ könnte erhalten werden, indem man für jeden Wert x anhand der Regressionsgeraden den zugehörigen $\hat{y}$ -Wert ermittelt, die Quadrate der einzelnen Differenzen $(y - \hat{y})^2$ summiert und durch den um zwei verminderten Stichprobenumfang teilt. Die Wurzel aus der Restvarianz wäre dann $s_{y \cdot x}$. Schneller und exakter erhält man diese Standardabweichung nach

$$s_{y \cdot x} = \sqrt{\frac{Q_y - (Q_{xy})^2 / Q_x}{n-2}} \quad (6.96)$$

Da $s_{y.x}$ ein Maß für die Fehler ist, die man bei der Schätzung oder Voraussage von Y aus vorgegebenen Werten X macht, wird diese Standardabweichung auch als **Standardschätzfehler** oder als **Standardfehler der Voraussage** bezeichnet.

Ein **Variationskoeffizient für die Regression** ist $V_R = s_{y.x}/\bar{y}$. Mit Hilfe von $s_{y.x}$ lässt sich nach Dixon und Massey [DM83] ein Toleranzintervall approximieren.

Bezeichnen wir nun die Standardabweichung des Achsenabschnitts a (auf der Ordinatenachse) mit s_a und die Standardabweichung des Regressionskoeffizienten $b_{yx} = b$ mit s_b , dann ist ihr Quotient durch (6.100) gegeben; die exakt indizierten Standardabweichungen erhält man aus (6.97) bis (6.99):

$$s_{a_{yx}} = s_{y.x} \cdot \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{Q_x}} \quad (6.97)$$

$$s_{b_{yx}} = \frac{s_{y.x}}{\sqrt{Q_x}} = \sqrt{s_{y.x}^2 / Q_x} \quad (6.98)$$

Für die Standardabweichung des Regressionskoeffizienten, für $s_{b_{yx}}$ gilt, dass sie bei konstanter Restvarianz um so kleiner wird, je größer der Definitionsbereich $x_{\max} - x_{\min}$ ist, je größer Q_x ist, weil dann r meist groß und die Steigung der Regressionsgeraden besser geschätzt werden kann.

$$s_{a_{yx}} = s_{b_{yx}} \cdot \sqrt{\frac{\sum x^2}{n}} \quad (6.99)$$

Damit ist eine Kontrolle für s_a und s_b möglich;

$$\frac{s_a}{s_b} = \sqrt{\frac{\sum x^2}{n}} \quad (6.100)$$

Zwischen der **empirischen Restvarianz** $s_{y.x}^2$ und s_y^2 besteht eine interessante Beziehung: (6.101) rechts.

$$s_{y.x}^2 = (s_y^2 - b_{yx}^2 s_x^2) \frac{n-1}{n-2} = s_y^2 (1 - r^2) \frac{n-1}{n-2} \quad (6.101)$$

Beachten Sie: $s_{y.x}^2 = s_y^2 - (s_{xy}^2 / s_x^2)$, hier ist (...) durch den **Regressionseffekt von Y auf X** erklärt.

Für große Stichprobenumfänge gilt:

$$s_{y.x} \approx s_y \sqrt{1 - r^2} \quad (6.102)$$

$$s_{x.y} \approx s_x \sqrt{1 - r^2} \quad (6.103)$$

$$\begin{array}{ccc} s_{y.x} \approx s_y & & s_{y.x} \approx 0 \\ \text{Für } r = 0 \text{ wird} & \text{und für } r \approx 1 \text{ wird} & \\ s_{x.y} \approx s_x & & s_{x.y} \approx 0 \end{array} \quad (6.104)$$

Nach der Quadrierung von (6.102) und (6.103), der Division durch s_y^2 bzw. s_x^2 , der Subtraktion von 1 ergibt sich die Varianz-Interpretation des Bestimmtheitsmaßes $\hat{B} = r^2$:

$$r^2 \approx 1 - \frac{s_{y,x}^2}{s_y^2} \approx 1 - \frac{s_{x,y}^2}{s_x^2} \quad (6.105)$$

Kontrollen

Zur Kontrolle der Rechnungen bediene man sich der folgenden Beziehungen:

$$\sum (x + y)^2 = \sum x^2 + \sum y^2 + 2 \sum xy \quad (6.106)$$

$$\sum (x + y)^2 - \frac{1}{n} [\sum (x + y)]^2 = Q_x + Q_y + 2Q_{xy} \quad (6.107)$$

$$s_{y,x}^2 = \frac{\sum (y - \hat{y})^2}{n - 2} \quad (6.108)$$

Beispiel: Die Berechnung der verschiedenen Standardabweichungen und Möglichkeiten zur Rechenkontrolle (bei manuellen Rechnungen) sollen an einem Zahlenbeispiel mit $n = 7$ Wertepaaren gezeigt werden (Tabelle 6.10).



Tabelle 6.10. Zahlenbeispiel zur Berechnung von Standardabweichungen bei Schätzungen im Rahmen der Regressions- und Korrelationsrechnung

	x	y	x^2	y^2	xy	$x + y$	$(x + y)^2$	$\hat{y}$	$y - \hat{y}$	$(y - \hat{y})^2$
1	13	12	169	144	156	25	625	13,267	-1,267	1,6053
2	17	17	289	289	289	34	1156	14,971	2,029	4,1168
3	10	11	100	121	110	21	441	11,989	-0,989	0,9781
4	17	13	289	169	221	30	900	14,971	-1,971	3,8848
5	20	16	400	256	320	36	1296	16,249	-0,249	0,0620
6	11	14	121	196	154	25	625	12,415	1,585	2,5122
7	15	15	225	225	225	30	900	14,119	0,881	0,7762
Σ	103	98	1593	1400	1475	201	5943	-	0	13,9354

Für die Summen gilt:

$$\begin{aligned} \sum x &= 103, \sum y = 98 \\ \sum x^2 &= 1593, \sum y^2 = 1400 \\ \sum xy &= 1475 \end{aligned}$$

Zunächst berechnen wir

$$\begin{aligned} Q_x &= 1593 - (103)^2/7 = 77,429 \\ Q_y &= 1400 - (98)^2/7 = 28 \\ Q_{xy} &= 1475 - 103 \cdot 98/7 = 33, \end{aligned}$$

und hieraus bei Bedarf den Korrelationskoeffizienten nach (6.92)

$$r = \frac{Q_{xy}}{\sqrt{Q_x Q_y}} = \frac{33}{\sqrt{77,429 \cdot 28}} = 0,709$$

Aus Q_x und Q_y erhält man schnell die entsprechenden Standardabweichungen

$$s_x = \sqrt{\frac{77,429}{6}} = 3,592$$

$$s_y = \sqrt{\frac{28}{6}} = 2,160 ;$$

dann ermitteln wir die Standardabweichung der y -Werte für ein gegebenes x (6.96)

$$s_{y.x} = \sqrt{\frac{28 - 33^2/77,429}{5}} = 1,670$$

und hiermit die Standardabweichung des Achsenabschnitts $s_{a_{yx}}$ und die Standardabweichung des Regressionskoeffizienten $s_{b_{yx}}$:

$$s_{a_{yx}} = 1,670 \cdot \sqrt{\frac{1}{7} + \frac{14,714^2}{77,429}} = 2,862$$

$$s_{b_{yx}} = \frac{1,670}{\sqrt{77,429}} = 0,190$$

Kontrolle:
$$\frac{s_{a_{yx}}}{s_{b_{yx}}} = \frac{2,862}{0,190} \simeq 15 \simeq \sqrt{\frac{1593}{7}} = \sqrt{\frac{\sum x^2}{n}}$$

Wir kontrollieren die Resultate des Beispiels in Tabelle 6.10 und ermitteln $\sum(x+y)$ und $\sum(x+y)^2$. Bekannt sind $\sum x^2 = 1593$, $\sum y^2 = 1400$ und $\sum xy = 1475$. Haben wir richtig gerechnet, dann muss nach der ersten Kontrollgleichung (6.106) $5943 = 1593 + 1400 + 2 \cdot 1475 = 5943$ sein.

Nun zur Kontrolle der Abweichungsquadratsummen $Q_x = 77,429$, $Q_y = 28$, $Q_{xy} = 33$ nach der zweiten Kontrollgleichung (6.107) $5943 - (1/7)201^2 = 171,429 = 77,429 + 28 + 2 \cdot 33$.

Für die letzte Kontrolle benötigen wir die aufgrund der Regressionsgeraden $\hat{y} = 7,729 + 0,426x$ für die 7 gegebenen x -Werte erhaltenen Schätzwerte $\hat{y}$. Für $s_{y.x}$ hatten wir 1,67 erhalten, in die dritte Kontrollgleichung (6.108) eingesetzt:

$$1,67^2 = 2,79 = \frac{13,9354}{5}$$

Analog sollen die Berechnungen in den einzelnen Schritten mit **R** gezeigt werden. Die Wahl der Variablennamen entspricht den in Formeln gewählten Kürzeln.

```
> n <- 7
> x <- c(13, 17, 10, 17, 20, 11, 15); sum(x); sum(x^2)
[1] 103 [1] 1593
> y <- c(12, 17, 11, 13, 16, 14, 15); sum(y); sum(y^2)
[1] 98 [1] 1400
> xy <- x * y; sum(xy)
[1] 1475
> Qx <- sum(x^2) - sum(x)^2/n; Qx
[1] 77.42857
> Qy <- sum(y^2) - sum(y)^2/n; Qy
[1] 28
> Qxy <- sum(xy) - sum(x)*sum(y)/n; Qxy
[1] 33
> #####
> r <- Qxy / sqrt(Qx*Qy); r
[1] 0.7087357
> #####
> sx <- sqrt(Qx/(n-1)); sx
```

```
[1] 3.59232
> sy <- sqrt(Qy / (n - 1));          sy
[1] 2.160247
> sy.x <- sqrt((Qy - Qxy^2 / Qx) / (n - 2)); sy.x
[1] 1.669456
> #####
> byx <- Qxy / Qx;                  byx
[1] 0.4261993
> sbyx <- sy.x / sqrt(Qx);           sbyx
[1] 0.1897250
> #####
> ayx <- mean(y) - byx * mean(x);    ayx
[1] 7.728782
> sayx <- sy.x * sqrt(1 / n + mean(x)^2 / Qx); sayx
[1] 2.862090
```

Die vorangehenden Erklärungen und Berechnen machen das Prinzip deutlich, nach dem Schätzfehler (Residuen) und Standardfehler der Regressionskoeffizienten bestimmt werden. Praktisch wird man in **R** die Analyse mit der Funktion **lm()** durchführen. Die Ergebnisse können dann mit der generischen Funktion **summary()** ausgegeben werden:

```
> summary(lm(y ~ x))

Call: lm(formula = y ~ x)

Residuals:
    1      2      3      4      5      6      7
-1.2694  2.0258 -0.9908 -1.9742 -0.2528  1.5830  0.8782

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   7.7288     2.8621   2.700  0.0428 *
x              0.4262     0.1897   2.246  0.0746 .
...

```

Auf der Grundlage dieser Maßzahlen können dann Konfidenzintervalle zur Regression bestimmt werden.

6.13.2 Konfidenzintervalle für den Regressionskoeffizienten, für den Achsenabschnitt und für die Restvarianz

Die Konfidenzintervalle für den Regressionskoeffizienten und für den Achsenabschnitt sind durch (6.109) gegeben. Darin bezeichnet t das entsprechende Quantil der t -Verteilung mit $FG = n - 2$ Freiheitsgraden.

$$b_{yx} \pm t \cdot s_{b_{yx}} \quad \text{und} \quad a_{yx} \pm t \cdot s_{a_{yx}} \quad (6.109)$$

Beispiel: (für 95%-Konfidenzintervalle):

Gegeben: $b_{yx} = 0,426$; $s_{b_{yx}} = 0,190$; $n = 80$; d. h. $t_{78;0,05} = 1,99$ $1,99 \cdot 0,19 = 0,378$

$$b_{yx} \pm t s_{b_{yx}} = 0,426 \pm 0,378 \quad 95\text{-KI: } 0,048 \leq \beta_{yx} \leq 0,804$$

Gegeben: $a_{yx} = 7,729$; $s_{a_{yx}} = 2,862$; $n = 80$; d. h. $t_{78;0,05} = 1,99$ $1,99 \cdot 2,862 = 5,695$

$$a_{yx} \pm t s_{a_{yx}} = 7,729 \pm 5,695 \quad 95\text{-KI: } 2,034 \leq \alpha_{yx} \leq 13,424$$

**Das Konfidenzintervall
für die Restvarianz $\sigma_{y \cdot x}^2$**
erhält man nach:

$$\frac{s_{y \cdot x}^2 (n - 2)}{\chi_{(n-2; \alpha/2)}^2} \leq \sigma_{y \cdot x}^2 \leq \frac{s_{y \cdot x}^2 (n - 2)}{\chi_{(n-2; 1-\alpha/2)}^2} \quad (6.110)$$



**Beispiel:**

Gegeben: $s_{y \cdot x} = 0,138$; $n = 80$; $P = 95\%$ (d. h. $\alpha = 5\% = 0,05$; $\alpha/2 = 0,025$;

$$1 - 0,025 = 0,975) \quad \chi^2_{78;0,025} = 104,31 \quad \chi^2_{78;0,975} = 55,47$$

Das 95%-Konfidenzintervall lautet damit: $\frac{0,138 \cdot 78}{104,31} \leq \sigma^2_{y \cdot x} \leq \frac{0,138 \cdot 78}{55,47}$

$$95\text{-KI: } 0,103 \leq \sigma_{y \cdot x} \leq 0,194.$$

Gibt man im konkreten Fall jeweils die drei 95%-Kle (6.109 und 6.110) an, so gilt nach Bonferroni für die Wahrscheinlichkeit, dass diese drei die entsprechenden Parameter erfassen oder überdecken $P \geq 1 - (0,05 + 0,05 + 0,05) = 0,85$ (simultane Wahrscheinlichkeit: bei Unabhängigkeit $P_{\text{sim}} = 0,95^3 = 0,8573$). Für α und β allein lässt sich mühelos anhand der beiden 95%-Konfidenzintervalle ($0,95^2 = 0,9025$) ein **simultanes 90%-Konfidenzintervall** angeben. Sonst bevorzugt man bei t -Tests generell die Maximum Modulus t Prozedur, d. h. zweiseitige Schranken der **Studentisierten Maximum Modulus-Verteilung** für Rho gleich Null (Hahn [HH71], Tabelle 6.11). Erhält man z. B. für n Beobachtungspaare (x_i, y_i) die vier Schätzwerte in (6.109), dann ist der Wert $t_{n-2;0,05;\text{zweiseitig}}$ durch den Wert $|M|_{10;2;0,05}$ zu ersetzen, um Konfidenzintervalle zu erhalten, die mit einer Wahrscheinlichkeit von mindestens $[1 - 0,05 =] 0,95$ beide Parameter enthalten.

6.13.3 Konfidenzintervalle und Prädiktionsintervalle für die Regressionsgerade

Jede gegebene Regressionsgerade erfährt durch Veränderung von $\bar{y}$ eine Parallelverschiebung nach oben oder unten. Verändert man den Regressionskoeffizienten b , so beginnt die Gerade um ihren Mittelpunkt $(\bar{x}, \bar{y})$ zu rotieren (vgl. Abb. 6.5).

Wir benötigen zunächst zwei Standardabweichungen:

1. Die Standardabweichung für einen geschätzten Mittelwert $\hat{y}$ an der Stelle x

$$s_{\hat{y}} = s_{y \cdot x} \cdot \sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{Q_x}} \quad (6.111)$$

2. Die Standardabweichung für einen vorausgesagten Einzelwert $\hat{y}$ an der Stelle x

$$s_{\hat{y}} = s_{y \cdot x} \cdot \sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{Q_x}} \quad (6.112)$$

Folgende Konfidenzintervalle (KI) gelten für:

1. die **gesamte Regressionsgerade** (simultanes Konfidenzintervall):

$$\hat{y} \pm \sqrt{2F_{(2,n-2)} s_{\hat{y}}} \quad (6.113)$$

2. den **Erwartungswert von Y [$E(Y)$] an der Stelle $X = x$:**

$$\hat{y} \pm t_{(n-2)} s_{\hat{y}} \quad (6.114)$$

Tabelle 6.11. Zweiseitige Schranken der Studentisierten Maximum Modulus Verteilung $t_{\nu;k;\alpha}^{\rho=0} = |M|_{\nu;k;\alpha}$ (SMM-Verteilung) mit dem Parameter k und den Freiheitsgraden ν für den Korrelationskoeffizienten $\rho = 0$ und die Signifikanzstufen $\alpha = 0,05$ und $\alpha = 0,01$ (aus Hahn, G. J. and Hendrickson, R.W. (1971): A table of percentage points of the distribution of the largest absolute value of k Student t variates and its applications. Biometrika 58, 323-332, Table 1, p. 325; mit freundlicher Erlaubnis)

	$k=1$	2	3	4	5	6	8	10	12	15	20
$\alpha=0,05$											
$\nu=3$	3,183	3,960	4,430	4,764	5,023	5,233	5,562	5,812	6,015	6,259	6,567
4	2,777	3,382	3,745	4,003	4,203	4,366	4,621	4,817	4,975	5,166	5,409
5	2,571	3,091	3,399	3,619	3,789	3,928	4,145	4,312	4,447	4,611	4,819
6	2,447	2,916	3,193	3,389	3,541	3,664	3,858	4,008	4,129	4,275	4,462
7	2,365	2,800	3,056	3,236	3,376	3,489	3,668	3,805	3,916	4,051	4,223
8	2,306	2,718	2,958	3,128	3,258	3,365	3,532	3,660	3,764	3,891	4,052
9	2,262	2,657	2,885	3,046	3,171	3,272	3,430	3,552	3,651	3,770	3,923
10	2,228	2,609	2,829	2,984	3,103	3,199	3,351	3,468	3,562	3,677	3,823
11	2,201	2,571	2,784	2,933	3,048	3,142	3,288	3,400	3,491	3,602	3,743
12	2,179	2,540	2,747	2,892	3,004	3,095	3,236	3,345	3,433	3,541	3,677
15	2,132	2,474	2,669	2,805	2,910	2,994	3,126	3,227	3,309	3,409	3,536
20	2,086	2,411	2,594	2,722	2,819	2,898	3,020	3,114	3,190	3,282	3,399
25	2,060	2,374	2,551	2,673	2,766	2,842	2,959	3,048	3,121	3,208	3,320
30	2,042	2,350	2,522	2,641	2,732	2,805	2,918	3,005	3,075	3,160	3,267
40	2,021	2,321	2,488	2,603	2,690	2,760	2,869	2,952	3,019	3,100	3,203
60	2,000	2,292	2,454	2,564	2,649	2,716	2,821	2,900	2,964	3,041	3,139
$\alpha=0,01$											
3	5,841	7,127	7,914	8,479	8,919	9,277	9,838	10,269	10,616	11,034	11,559
4	4,604	5,462	5,985	6,362	6,656	6,897	7,274	7,565	7,801	8,087	8,451
5	4,032	4,700	5,106	5,398	5,625	5,812	6,106	6,333	6,519	6,744	7,050
6	3,707	4,271	4,611	4,855	5,046	5,202	5,449	5,640	5,796	5,985	6,250
7	3,500	3,998	4,296	4,510	4,677	4,814	5,031	5,198	5,335	5,502	5,716
8	3,355	3,809	4,080	4,273	4,424	4,547	4,742	4,894	5,017	5,168	5,361
9	3,250	3,672	3,922	4,100	4,239	4,353	4,532	4,672	4,785	4,924	5,103
10	3,169	3,567	3,801	3,969	4,098	4,205	4,373	4,503	4,609	4,739	4,905
11	3,106	3,485	3,707	3,865	3,988	4,087	4,247	4,370	4,470	4,593	4,750
12	3,055	3,418	3,631	3,782	3,899	3,995	4,146	4,263	4,359	4,475	4,625
15	2,947	3,279	3,472	3,608	3,714	3,800	3,935	4,040	4,125	4,229	4,363
20	2,845	3,149	3,323	3,446	3,541	3,617	3,738	3,831	3,907	3,999	4,117
25	2,788	3,075	3,239	3,354	3,442	3,514	3,626	3,713	3,783	3,869	3,978
30	2,750	3,027	3,185	3,295	3,379	3,448	3,555	3,637	3,704	3,785	3,889
40	2,705	2,969	3,119	3,223	3,303	3,367	3,468	3,545	3,607	3,683	3,780
60	2,660	2,913	3,055	3,154	3,229	3,290	3,384	3,456	3,515	3,586	3,676

3. Voraussagebereich (prediction interval) für eine zukünftige Beobachtung Y an der Stelle $X = x$:

$$\hat{y} \pm t_{(n-2)} s_{\hat{y}} \quad (6.115)$$

Diese Bereiche gelten nur für den Messbereich. Sie werden in Abhängigkeit von x durch Hyperbeläste begrenzt. Das Konfidenzintervall (6.113) ist von den drei Bereichen das weiteste, (6.114) ist das engste; für $n \rightarrow \infty$ schrumpfen (6.113) und (6.114) gegen Null, (6.115) schrumpft gegen einen Streifen der Breite $z \cdot \sigma_{y \cdot x}$.



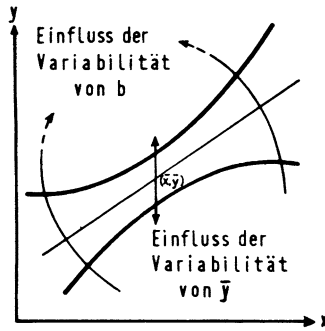


Abb. 6.5. Konfidenzintervall für die lineare Regression



Beispiel: Wir nehmen wieder das einfache Modellbeispiel (Tabelle 6.10), wählen vier x -Werte aus, zu denen die entsprechenden Punkte des simultanen Vertrauensbandes ermittelt werden sollen (95%-KI: d. h. $F_{(2;5;0,025)} = 8,43$). Die x -Werte sollten innerhalb des gemessenen Bereiches liegen, sie mögen gleiche Abstände voneinander aufweisen. In Tabelle 6.12 bilden diese vier x -Werte Spalte 1, ihre Abweichungen vom Mittelwert ($\bar{x} = 14,714$) sind in der folgenden Spalte notiert. Spalte 3 enthält die aufgrund der Regressionsgeraden $\hat{y} = 7,729 + 0,426x$ für die ausgewählten x -Werte geschätzten $\hat{y}$ -Werte. Die Abweichungen der x -Werte von ihrem Mittelwert werden quadriert, durch $Q_x = 77,429$ dividiert und um $(1/n) = (1/7)$ vermehrt. Die Quadratwurzel aus diesem Zwischenergebnis liefert, mit $\sqrt{2F} \cdot s_{y \cdot x} = \sqrt{2 \cdot 8,43} \cdot 1,67 = 6,857$ multipliziert, die entsprechenden B_x Werte (vgl. $\hat{y} \pm B_x$ mit $B_x = \sqrt{2F_{(2;n-2)} s_{\hat{y}}}$). Verbindet man die erhaltenen Punkte des Vertrauensbereiches ($y \pm B_x$) (Tabelle 6.12) durch einen die oberen Punkte und einen die unteren Punkte erfassenden Kurvenzug, dann erhält man ein **simultanes 95%-Vertrauensband** für die gesamte Regressionsgerade. Werden mehr Punkte benötigt, dann sollte beachtet werden, dass aus Symmetriegründen die vier B_x -Werte praktisch acht B_x -Werte darstellen, es sind dann jeweils nur noch die vier restlichen $\hat{y}$ -Werte zu ermitteln. Beispielsweise hat B_x denselben Wert bei $x = 14$, d. h. $(\bar{x} - 0,714)$ und bei $x = 15,428$, d. h. $(\bar{x} + 0,714)$.

Tabelle 6.12. Fortsetzung zum Zahlenbeispiel in Tabelle 6.10

x	$x - \bar{x}$	$\hat{y}$	$\sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{Q_x}}$	B_x	$y - B_x$	$y + B_x$
12	-2,714	12,84	0,488	3,35	9,49	16,19
14	-0,714	13,69	0,387	2,65	11,04	16,34
16	1,286	14,54	0,405	2,78	11,76	17,32
18	3,286	15,40	0,531	3,64	11,76	19,07

Häufiger finden allerdings die beiden anderen Konfidenzintervalle Anwendung, die mit dem Quantil der t-Verteilung, in unserem Beispiel $t_{5;0,975} = 2,57$, bestimmt werden. Für den Punkt $x = 16$ sollen die Vertrauensgrenzen ermittelt werden, wobei wir zunächst $B_{x=16}$ nach (6.114) und anschließend $B'_{x=16}$ nach (6.115) berechnen wollen:

$$B_{x=\text{konst.}} = t_{s_{y \cdot x}} \cdot \sqrt{\frac{1}{n} + \frac{(x - \bar{x})^2}{Q_x}}, \text{ d.h. z.B. } B_{16} = 2,57 \cdot 1,67 \cdot \sqrt{\frac{1}{7} + \frac{(16 - 14,714)^2}{77,429}} = 1,74$$

Das 95%-Konfidenzintervall für eine Schätzung des Mittelwertes von y an der Stelle $x = 16$ ist dann durch das Intervall $14,54 \pm 1,74$ gegeben. Die Grenzwerte des Bereiches sind 12,80 und 16,28. In **R** kann die Bestimmung des Konfidenzintervalles durch die Funktion **predict()** erfolgen; die Schätzung ist mit *est*, die untere und obere Vertrauensgrenze mit *lwr* und *upr* bezeichnet.


```
> new <- data.frame(x = c(12,14,16,18))
> predict(lm(y~x), new, int="c", level = 0.95)
      fit      lwr      upr
1 12.84317 10.74953 14.93681
2 13.69557 12.03656 15.35458
3 14.54797 12.80896 16.28698
4 15.40037 13.12028 17.68046
```

$$B'_{x=\text{konst.}} = t_{s_{y \cdot x}} \cdot \sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{Q_x}}, \text{ z. B.}$$

$$B'_{16} = 2,57 \cdot 1,67 \cdot \sqrt{1 + \frac{1}{7} + \frac{(16 - 14,714)^2}{77,429}} = 4,63$$

Das 95%-Konfidenzintervall für eine Schätzung des Wertes y (Prädiktionsintervall) an der Stelle $x = 16$ ist durch das Intervall $14,54 \pm 4,63$ gegeben. Die Grenzwerte dieses Bereiches sind 9,91 und 19,17. Dieses Intervall ist als Bereich für Einzelwerte wesentlich größer als der oben berechnete Mittelwert-Bereich. Auch dieser Bereich kann in **R** mit der Funktion **predict()** bestimmt werden, wenn der Funktionsparameter **int="p"** gesetzt wird.

```
> predict(lm(y~x), new, int="p", level = 0.95)
      fit      lwr      upr
1 12.84317  8.068231 17.61812
2 13.69557  9.094586 18.29656
3 14.54797  9.917538 19.17840
4 15.40037 10.540783 20.25996
```

Beispiel (Flügelweite): Zusammenfassend soll an einem Beispiel die einfache lineare Regression mit dem Programm **R** gezeigt werden. Bei 13 Sperlingen unterschiedlichen Alters (Tage) wurden die Flügelweiten (in cm) gemessen. Die Ergebnisse sind in Tabelle 6.13 zusammengefasst.

Tabelle 6.13. Flügelweite und Alter von 13 Sperlingen

Alter (Tage)	3	4	5	6	8	9	10	11	12	14	15	16	17
Flügel (cm)	1,4	1,5	2,2	2,4	3,1	3,2	3,2	3,9	4,1	4,7	4,5	5,2	5,0

Die Darstellung der Werte in einer Punktwolke (vgl. Abbildung 6.6) weist deutlich auf eine lineare Abhängigkeit hin, die durch die Funktion **lm()** (fitting linear models) näher analysiert wird. Der Achsenabschnitt ist $a = 0,7134$, der Regressionskoeffizient (Steigung) beträgt $b = 0,27$. Zeichnet man die zugehörige Regressionsgerade in die Punktwolke ein, so ist wird die gute Annäherung von Beobachtung und Modell deutlich. Die Standardfehler für die Schätzung von Achsenabschnitt und Regressionskoeffizient (im Beispiel $s_{a_{yx}} = 0,14790$ bzw. $s_{b_{yx}} = 0,01349$ können durch die Funktion **summary()** angezeigt werden.

```
> Alter <- c( 3, 4, 5, 6, 8, 9, 10, 11, 12, 14, 15, 16, 17) # Tage
> Fluegel <- c(1.4,1.5,2.2,2.4,3.1,3.2,3.2,3.9,4.1,4.7,4.5,5.2,5.0) # cm
> ##### Zeichnen der Punkte
> plot(Alter, Fluegel, xlim=c(0, 20), ylim=c(0,6), pch=16, cex=1.4,
+      xlab="Alter in Tagen", ylab="Fluegelspannweite in cm")
> linreg <- lm(Fluegel ~ Alter)
> ##### lineares Regressionsmodell
> a <- linreg$coef[1]; a # Achsenabschnitt
(Intercept)
0.7130945
> b <- linreg$coef[2]; b # Regressionskoeffizient (Steigung)
Alter
0.270229
> ##### Regressionsgerade
```



```
> Flueg.est <- a + b*Alter
> lines(Alter, Flueg.est, lty=1, cex=1.2, col="red")
> ##### Parameterschaetzung
> summary(linreg)
```

Call:

```
lm(formula = Fluegel ~ Alter)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.30699	-0.21538	0.06553	0.16324	0.22507

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.71309	0.14790	4.821	0.000535 ***
Alter	0.27023	0.01349	20.027	5.27e-10 ***

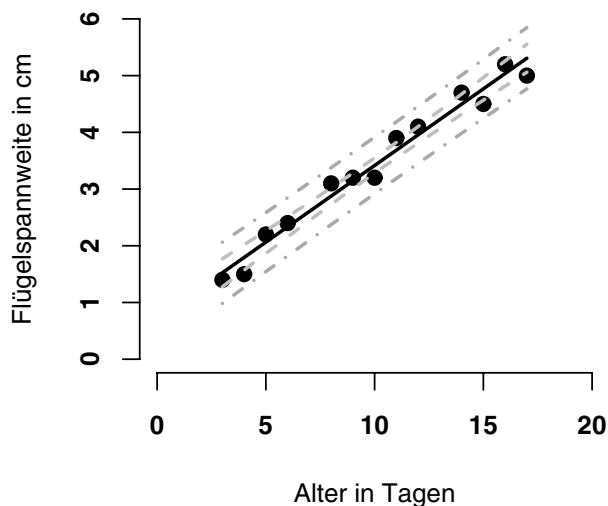


Abb. 6.6. Lineare Regression zu Flügelspannweite und Alter von Sperlingen; Regressionsgerade (Linie), 95%-Konfidenzintervall (Strich) und 95%-Prädiktionsintervall (Punkt-Strich)

Die Schätzung der Flügelweite aus dem Alter nach dem linearen Modell kann durch die Funktion `predict()` angezeigt werden. Insbesondere werden Konfidenzintervalle bzw. Prädiktionsintervalle berechnet und können ebenfalls in Abbildung 6.6 eingezeichnet werden.

```
> ##### Konfidenz- und Praediktions-Intervall #####
> new <- data.frame(Alter = seq(3, 17, by=1))
> conf <- predict(lm(Fluegel ~ Alter), new, int="c", se.fit=TRUE, level= 0.95)
> lines(new$Alter, conf$fit[,2], lty=2, cex=1.1, col="blue")
> lines(new$Alter, conf$fit[,3], lty=2, cex=1.1, col="blue")
> pred <- predict(lm(Fluegel ~ Alter), new, int="p", se.fit=TRUE, level= 0.95)
> lines(new$Alter, pred$fit[,2], lty=4, cex=1.1, col="darkgreen")
> lines(new$Alter, pred$fit[,3], lty=4, cex=1.1, col="darkgreen")
```

6.13.4 Inverse Prädiktion aus einer linearen Regression

Unter bestimmten Voraussetzungen kann es wünschenswert und sinnvoll sein, einen Wert der unabhängigen Variablen (x_i) zu bestimmen, der nach einem fest vorgegebenen Wert der abhängigen Variablen (y_i) unter dem gewählten (hier linearen) Modell zu erwarten ist (**inverse Prädiktion**). Für das Beispiel in Tabelle 6.13 bedeutet dies, dass auf der Basis einer gemessenen Flügelweite

auf das Alter des Vogels geschlossen werden soll. Wie alt ist ein Vogel, dessen Flügelweite z.B. 5,2cm beträgt? Eine einfache algebraische Umformung der linearen Regressionsgleichung führt auf

$$\hat{x}_i = \frac{y_i - a_{yx}}{b_{yx}} \quad (6.116)$$

Die Übertragung dieser Überlegung auf die Bestimmung eines geeigneten Konfidenzintervalles ist allerdings nicht so einfach, da dieses nicht symmetrisch verläuft, wie in Abbildung 6.6 zu erkennen ist. Das $1 - \alpha$ -Konfidenzintervall für ein x_i , zu einem gegebenen y_i kann nach 6.117 berechnet werden.

$$\begin{aligned} & \bar{x} + \frac{b_{yx}(y_i - \bar{y})^2}{K} \pm \frac{t}{K} s_{y \cdot x} \sqrt{\frac{(y_i - \bar{y})^2}{\sum x^2} + K \left(1 + \frac{1}{n}\right)} \\ & \text{mit } K = b_{yx}^2 - t_{n-2, 1-\alpha/2}^2 s_{b_{yx}}^2 \end{aligned} \quad (6.117)$$

6.13.5 Das Konfidenzintervall für den Korrelationskoeffizienten ρ

Das 95%-Konfidenzintervall für ρ entnimmt man Abbildung 6.7 (aus David [Dav38]) an der über r errichteten Senkrechten zwischen den beiden mit dem betreffenden n versehenen Kurven. Nur dann, wenn der Vertrauensbereich den Wert $\rho = 0$ nicht einschließt, kann von einer echten Korrelation ($\rho \neq 0$) gesprochen werden.

Beispiel 1: Ein extremes Beispiel mit $r = 0,5$ und $n = 3$ mag dies illustrieren.

Wir gehen mit $r = +0,5$ (Abszisse: Mitte der rechten Hälfte) in das Nomogramm ein und lesen über $r = 0,5$ die Höhen der beiden Kurven $n = 3$ auf der Ordinate ab: $\rho_1 \simeq -0,91$ und $\rho_2 \simeq +0,98$. Das Konfidenzintervall ist riesig (95%-KI: $-0,91 \lesssim \rho \lesssim +0,98$) und lässt praktisch keine Aussage zu.



Beispiel 2: Für $r = 0,68$ und $n = 50$ (vgl. Abb. 6.7) erhalten wir den 95%-KI: $0,50 \lesssim \rho \lesssim 0,80$ und damit die Bestätigung einer echten formalen Korrelation ($P = 0,05$).



Vertrauensgrenzen für großes n gehen von folgender Überlegung aus. Wenn sich der Korrelationskoeffizient statistisch signifikant von Null unterscheidet, weicht seine Verteilung um so stärker von der zweidimensionalen Normalverteilung ab, je kleiner die Anzahl der Beobachtungspaare n und je größer sein Absolutwert ist. Durch die **z-Transformation** nach R. A. Fisher (6.118) wird die Verteilung des Korrelationskoeffizienten approximativ normalisiert. Diese Approximation ist umso besser, je kleiner $|\rho|$ ist und je größer n ist (vgl. auch Abschnitt [7.7.2]).

$$\dot{z} = 0,5 \ln[(1+r)/(1-r)] = 1,1513 \lg[(1+r)/(1-r)] \text{ mit } s_{\dot{z}} = 1/\sqrt{n-3} \quad (6.118)$$

Die Umrechnung (Rückrechnung) eines transformierten $\dot{z}$ -Wertes in den entsprechenden Wert für r erfolgt nach (6.119).

$$r = (e^{2\dot{z}} - 1)/(e^{2\dot{z}} + 1) = (10^{\dot{z}/1,1513} - 1)/(10^{\dot{z}/1,1513} + 1) \quad (6.119)$$

Wir haben damit das Intervall für den Korrelationskoeffizienten $-1 < r < +1$ zu $-\infty < \dot{z} < +\infty$ geweitet. Dieses Transformations- $\dot{z}$ (r ist der Tangens hyperbolicus von $\dot{z}$, $r = \tanh \dot{z}$ und $\dot{z} = \tanh^{-1} r$), $\dot{z}$ heisst **Korrelationsziffer**, darf nicht mit der Standardnormalvariablen z verwechselt werden. Man benutze diese Transformation nur für $n > 10$. Für $n < 50$ empfiehlt Hotelling [Hot53] $\dot{z}$ durch $\dot{z}_H$ und $s_{\dot{z}}$ durch $s_{\dot{z}_H}$ zu ersetzen:



Abb. 6.7. Vertrauensgrenzen des Korrelationskoeffizienten: 95%-Konfidenzintervall für ϱ : Die Zahlen an den Kurven bezeichnen den Stichprobenumfang (aus F. N. David: Tables of the Ordinates and Probability Integral of the Distribution of the Correlation Coefficient in Small Samples, The Biometrika Office, London 1938)

$$\dot{z}_H = \dot{z} - (3\dot{z} + r)/4n; \quad s\dot{z}_H = 1/\sqrt{n-1} \quad (6.120)$$

Der Quotient

$$\hat{z} = \frac{\dot{z}}{s_{\dot{z}}} = \dot{z}\sqrt{n-3}$$

kann demnach als asymptotisch standardnormal-verteilt angenommen und das 95%-Konfidenzintervall für ϱ durch (6.121) angegeben werden.

$$\boxed{\dot{z} \pm 1,960s_{\dot{z}}} \quad (6.121)$$



Beispiel 3: Für einen Korrelationskoeffizienten von $r = 0,687$ aus $n = 50$ Beobachtungspaaren soll das 95%-Konfidenzintervall berechnet werden. Nach (6.118) folgt $\hat{z} = 0,842$ und weiter $\hat{z} = \hat{z}\sqrt{n-3} = 0,842\sqrt{47} = 5,772$. Das 95%-Konfidenzintervall erhält man dann über

$$s_z = \frac{1}{\sqrt{n-3}} = \frac{1}{\sqrt{50-3}} = 0,146$$

$$\text{und } \hat{z} \pm 1,96 \cdot 0,146 = \hat{z} \pm 0,286$$

$$0,556 \leq \hat{z} \leq 1,128$$

$$\text{zu 95\%-KI: } 0,505 \leq \varrho \leq 0,810.$$

```
> n <- 50
> r <- 0.687
>
> zp <- 0.5*log((1+r)/(1-r)); zp
[1] 0.842252
> szp <- 1/sqrt(n-3)
> lwr.z <- zp - qnorm(0.975)*szp; upr.z <- zp + qnorm(0.975)*szp
> lwr.z; upr.z
[1] 0.5563618      [1] 1.128142
>
> lwr.r <- (exp(2*lwr.z)-1)/(exp(2*lwr.z)+1)
> upr.r <- (exp(2*upr.z)-1)/(exp(2*upr.z)+1)
> lwr.r; upr.r
[1] 0.5052731      [1] 0.8103824
```

6.14 Toleranzgrenzen

Vertrauensgrenzen betreffen einen Parameter. **Grenzen für einen Anteil der Grundgesamtheit** werden als Toleranzgrenzen bezeichnet; sie grenzen statistische Anteilsbereiche ab und schaffen damit Toleranzbereiche; man spricht auch von Toleranzintervallen.

Toleranzgrenzen geben an, innerhalb welcher Grenzen ein bestimmter Anteil der Grundgesamtheit mit vorgegebener Wahrscheinlichkeit $P = (1 - \alpha)$ erwartet werden kann. Für eine **normalverteilte Grundgesamtheit** sind diese Grenzen von der Form $\bar{x} \pm ks$, wobei k eine geeignete Konstante ist. Beispielsweise entnehmen wir zur Ermittlung eines Toleranzbereiches – in dem in durchschnittlich 95% aller Fälle ($P = 0,95$; $\alpha = 0,05$) wenigstens der Anteil $\gamma = 0,90$ der Grundgesamtheit liegt – der Tabelle 6.14 für einen Stichprobenumfang von $n = 50$ den Faktor $k = 2,00$. Der gewünschte Toleranzbereich erstreckt sich damit von $\bar{x} - 2,00s$ bis $\bar{x} + 2,00s$. Hierbei ist s die aus den 50 Stichprobenelementen geschätzte Standardabweichung und $\bar{x}$ der zugehörige Mittelwert.

Zweiseitige Toleranzgrenzen für Normalverteilungen sind robust (Canavos [CK84]) für $P \leq 0,9$ [vgl. z. B. (6.122)], nicht zu stark ausgeprägter Schiefe und nicht extrem stark besetzten Verteilungsenden.

$$\text{Anteil} = \int_{\bar{X}-kS}^{\bar{X}+kS} f(t)dt = 0,9 \quad \text{oder} \quad P(\bar{X} - kS \leq X \leq \bar{X} + kS) = 0,9 \quad (6.122)$$

Übrigens sind einseitige Konfidenzintervalle für Quantile identisch mit einseitigen Toleranzgrenzen (vgl. z. B. Conover [Con99], S. 153).

Odeh [Ode78] gibt eine Tabelle der Toleranzfaktoren k (zweiseitig) für Zufallsstichproben des Umfangs n aus einer **normalverteilten Grundgesamtheit**. Der Toleranzbereich

$$\bar{X} - kS \leq X_i \leq \bar{X} + kS \quad (6.123)$$

mit dem Stichprobenmittelwert $\bar{X}$ und der Stichprobenstandardabweichung S enthält mit der Vertrauenswahrscheinlichkeit γ den Anteil P einer normalverteilten Grundgesamtheit [P und γ jeweils in 7 Stufen für $n = 2(1)100$].

Tabelle 6.14. Toleranzfaktoren für die Normalverteilung. Faktoren k für den zweiseitigen Toleranzbereich normalverteilter Grundgesamtheiten: Mit der Vertrauenswahrscheinlichkeit P liegen wenigstens γ Prozent der Elemente der Grundgesamtheit innerhalb des Toleranzbereiches $\bar{x} \pm ks$; hierbei sind $\bar{x}$ und s aus einer Stichprobe vom Umfang n berechnet. Ausgewählte, gerundete Werte (aus A. H. Bowker: Tolerance Factors for Normal Distributions, p. 102, in (Statistical Research Group, Columbia University), Techniques of Statistical Analysis (edited by Churchill Eisenhart, Millard W. Hastay, and W. Allen Wallis) New York and London 1947, McGrawHill Book Company Inc.) (Copyright vom 1. März 1966)

γ n	$P = 0,95$				$P = 0,99$			
	0,90	0,95	0,99	0,999	0,90	0,95	0,99	0,999
3	8,38	9,92	12,86	16,21	18,93	22,40	29,06	36,62
6	3,71	4,41	5,78	7,34	5,34	6,35	8,30	10,55
12	2,66	3,16	4,15	5,29	3,25	3,87	5,08	6,48
24	2,23	2,65	3,48	4,45	2,52	3,00	3,95	5,04
30	2,14	2,55	3,35	4,28	2,39	2,84	3,73	4,77
50	2,00	2,38	3,13	3,99	2,16	2,58	3,39	4,32
100	1,87	2,23	2,93	3,75	1,98	2,36	3,10	3,95
300	1,77	2,11	2,77	3,54	1,82	2,17	2,85	3,64
500	1,74	2,07	2,72	3,48	1,78	2,12	2,78	3,56
1000	1,71	2,04	2,68	3,42	1,74	2,07	2,72	3,47
∞	1,65	1,96	2,58	3,29	1,65	1,96	2,58	3,29

Erinnert sei: $P(\mu - 2,5\sigma \leq X_i \leq \mu + 2,5\sigma) = 1 - 2 \cdot 0,00621 = 0,98758$ oder fast 99%. Liegt nun anstatt der Grundgesamtheit eine Stichprobe des Umfangs n vor und postuliert man für (6.123), dass k den Wert 2,5 annehmen sollte, dann ergeben sich für die folgenden vier γ -Stufen und jeweils zwei P -Stufen die in Tabelle 6.15 genannten Stichprobenumfänge.

Tabelle 6.15. Benötigter Stichprobenumfang nach Odeh (1978) für vorgegebene Werte P , γ und $k = 2,5$. Demnach umfasst das konkrete Intervall $\bar{x} - 2,5s$ bis $\bar{x} + 2,5s$ z.B. den Anteil ($P =$) 95% einer normalverteilten Grundgesamtheit mit einer Vertrauenswahrscheinlichkeit (von $\gamma =$) 90%, vorausgesetzt, eine Zufallsstichprobe des Umfangs $n = 24$ mit den aus ihr berechneten Werten $\bar{x}$ und s liegt vor

Vertrauenswahrsch.	0,75		0,90		0,95		0,99	
Anteil der GG	0,90	0,95	0,90	0,95	0,90	0,95	0,90	0,95
Stichprobenumfang	6	11	11	24	15	35	26	62

Sobald der Stichprobenumfang n genügend groß ist, gilt näherungsweise $\bar{x} \pm z \cdot s$. Strenggenommen gilt dieser Ausdruck nur für $n \rightarrow \infty$. Für unbekannte Verteilungen ist die Ermittlung des Wertes k irrelevant. Hier geht man so vor, dass man denjenigen minimalen Stichprobenumfang angibt, bei dem mit einer Vertrauenswahrscheinlichkeit P angenommen werden darf, daß der Anteil γ der Grundgesamtheit zwischen dem kleinsten und dem größten Wert der Stichprobe liegen wird.

Toleranzintervalle enthalten einen genau spezifizierten Anteil β einer Grundgesamtheit. Man unterscheidet insbesondere auch beim Vorliegen einer Normalverteilung Toleranzintervalle, die **angenähert** 100 β % der Verteilung enthalten (β -expectation tolerance intervals, siehe Odeh u. Mitarb. [OCO89]), und solche, die **mindestens** 100 β % der Verteilung enthalten (β -content tolerance intervals, siehe Odeh u. Mitarb. [OCO87]).

Bei geringen Abweichungen von der Normalverteilung sind verteilungsunabhängige Toleranzgrenzen zu bevorzugen.

6.14.1 Verteilungsunabhängige Toleranzgrenzen

Wünschen wir mit einer Vertrauenswahrscheinlichkeit $P = 1 - \alpha$, dass der Anteil γ der Elemente einer beliebigen Grundgesamtheit zwischen dem größten und dem kleinsten Stichprobenwert liegt, so lässt sich der benötigte Stichprobenumfang n leicht abschätzen:

Aus den Tabellen 6.16 und 6.17 folgt, dass man mit rund 30 Beobachtungen einer Zufallsstichprobe und einer Wahrscheinlichkeit von 95% etwa 85% der Werte **einer beliebigen Grundgesamtheit** zwischen den Extremwerten bzw. oberhalb des kleinsten Extremwertes bzw. unterhalb des größten Extremwertes zu erwarten hat.

Tabelle 6.16. Stichprobenumfänge n für **zweiseitige** verteilungsunabhängige Toleranzgrenzen: zwischen den Extremwerten einer Zufallsstichprobe liegt mit der Wahrscheinlichkeit P mindestens der Anteil γ der Werte einer beliebigen Grundgesamtheit

γ	P=0,95	P=0,90	P=0,70	P=0,50
0,99	473	388	244	168
0,95	93	77	49	34
0,90	46	38	24	17
0,85	30	25	16	11

Tabelle 6.16 enthält Stichprobenumfänge n für zweiseitige nichtparametrische Toleranzgrenzen, die der **Gleichung von Wilks** ([Wil41] und [Wil42]) $n\gamma^{n-1} - (n-1)\gamma^n = 1 - P = \alpha$ genügen. Im Mittel liegt mit der Vertrauenswahrscheinlichkeit P mindestens der Anteil γ einer beliebigen Grundgesamtheit zwischen dem größten und dem kleinsten Wert einer der Grundgesamtheit entstammenden Zufallsstichprobe. Das heißt, in etwa $P100\%$ der Fälle, in denen einer beliebigen Grundgesamtheit Stichproben des Umfangs n entnommen werden, schließen die **Extremwerte der Stichprobe** mindestens $\gamma \cdot 100\%$ der Werte der Grundgesamtheit in sich ein.

Tabelle 6.17. Stichprobenumfänge für **einseitige** verteilungsunabhängige Toleranzgrenzen: unterhalb des größten Stichprobenwertes bzw. oberhalb des kleinsten Stichprobenwertes einer Zufallsstichprobe liegt mit der Wahrscheinlichkeit P mindestens der Anteil γ der Werte einer beliebigen Grundgesamtheit

γ	P=0,95	P=0,90	P=0,70	P=0,50
0,99	459	299	120	69
0,95	90	59	24	14
0,90	44	29	12	7
0,85	29	19	8	5

Ordnet man also die Werte einer Stichprobe der Größe nach, dann liegen mit einer durchschnittlichen Vertrauenswahrscheinlichkeit $P = 1 - \alpha$ innerhalb des durch den kleinsten und den größten Wert gegebenen Intervalls mindestens $\gamma \cdot 100\%$ der Elemente der Grundgesamtheit. Tabelle 6.19 gibt Werte von γ für verschiedene Irrtumswahrscheinlichkeiten α und Stichprobenumfänge n .

Die Tabellen 6.16 und 6.17 setzen eine sehr große Grundgesamtheit voraus; ist dies nicht der Fall, dann benötigt man kleinere Stichprobenumfänge wie Tabelle 6.18 [FC91] zeigt: z. B. $n = 87$ für $\gamma = 0,95$ [jetzt β genannt], $P = 0,95$ und $N = 1000$ anstatt der 93 Beobachtungen (für $N \rightarrow \infty$).

Beispiel (zu den Tabellen 6.16 und 6.19: Für $P = 0,95$ und $\gamma = 0,85$ ergibt sich ein Stichprobenumfang von $n = 30$, d. h. eine zufällige Stichprobe des Umfangs $n = 30$ enthält in durchschnittlich 95% aller Fälle mindestens 85% der Grundgesamtheit. Zwischen dem kleinsten und dem größten



Tabelle 6.18. Kleinsten Stichprobenumfang n für ein zweiseitiges β -content-Toleranzintervall $[x_{(1)}, x_{(n)}]$ für eine endliche Grundgesamtheit des Umfangs N . Einige Werte aus Fountain und Chou (1991)

N	90%-TI		95%-TI	
	$\beta = 0,90$	$\beta = 0,95$	$\beta = 0,90$	$\beta = 0,95$
100	28	45	33	51
200	32	57	39	67
300	34	62	41	74
500	35	67	43	81
700	36	70	44	84
1 000	37	72	44	87
3 000	37	75	45	91
5 000	38	76	46	92
10 000	38	76	46	93
∞	38	77	46	93

Wert einer Zufallsstichprobe des Umfangs $n = 30$ aus jeder beliebigen Grundgesamtheit liegen in durchschnittlich 95% aller Fälle mindestens 85% der Werte der betreffenden Grundgesamtheit. Legt man beide Prozentsätze auf 90% (95%) fest, so benötigt man eine Zufallsstichprobe des Umfangs $n = 38$ (93).

6.15 Übereinstimmung von Messwerten nach Bland-Altman

Die Darstellung und Bewertung der Übereinstimmung (agreement) von Messwerten (z. B. nach zwei verschiedenen Methoden oder durch zwei Untersucher bestimmt) kann sehr übersichtlich nach dem Verfahren von Bland und Altman [BA86] erfolgen.

Der Korrelationskoeffizient ist ein Maß für die Stärke eines linearen Zusammenhangs und orientiert sich dabei an der allgemeinen Geradengleichung ($y = a + bx$). Als ein Maß für die Übereinstimmung ist er somit nicht geeignet, da diese sich an der Winkelhalbierenden ($y = x$) orientieren muss. Ein hoher Korrelationskoeffizient weist somit nicht notwendig auf eine gute Übereinstimmung hin.

Bland und Altman schlagen daher eine Analyse der Abweichungen (Differenzen) zwischen den Messwertreihen in Abhängigkeit von der Höhe der Messungen (Mittelwert) vor. Sie definieren mit den Grenzen der Übereinstimmung (6.124) (limits of agreement) einen Bereich in Anlehnung an den Normbereich (Toleranzintervall).

$$L_{1/2} = \bar{d} \pm 2 \cdot s_d \quad (6.124)$$

In (6.124) bezeichnet $\bar{d}$ den Mittelwert und s_d die Standardabweichung der Differenzen. Die statistische Bewertung der so berechneten (geschätzten) Grenzwerte kann durch die Bestimmung entsprechender Konfidenzintervalle anhand der Quantile der t -Verteilung erfolgen.

$$L_{1/2} \pm t_{n-1, 1-\alpha/2} \frac{s_d}{\sqrt{n}} \quad (6.125)$$

Die Anwendung dieser Methode soll in **R** an einigen, mit der Funktion `rnorm()` künstlich erzeugten, Daten schrittweise gezeigt werden.

Tabelle 6.19. Verteilungsunabhängige Toleranzgrenzen (auszugsweise aus Wetzels, W.: Elementare Statistische Tabellen, Kiel 1965; Berlin, De Gruyter 1966, S. 31)

α	n	0,200	0,150	0,100	0,090	0,080	0,070	0,060	0,050	0,040	0,030	0,020	0,010	0,005	0,001
3	3	0,2871	0,2444	0,1958	0,1850	0,1737	0,1617	0,1490	0,1354	0,1204	0,1036	0,0840	0,0589	0,0414	0,0184
4	4	0,4175	0,3735	0,3205	0,3082	0,2950	0,2809	0,2656	0,2486	0,2294	0,2071	0,1794	0,1409	0,1109	0,0640
5	5	0,5098	0,4679	0,4161	0,4038	0,3906	0,3762	0,3603	0,3426	0,3222	0,2979	0,2671	0,2221	0,1851	0,1220
6	6	0,5776	0,5387	0,4897	0,4779	0,4651	0,4512	0,4357	0,4182	0,3979	0,3734	0,3417	0,2943	0,2540	0,1814
7	7	0,6291	0,5933	0,5474	0,5363	0,5242	0,5109	0,4961	0,4793	0,4596	0,4357	0,4044	0,3566	0,3151	0,2375
8	8	0,6696	0,6365	0,5938	0,5833	0,5719	0,5594	0,5453	0,5293	0,5105	0,4875	0,4570	0,4101	0,3685	0,2887
9	9	0,7022	0,6715	0,6316	0,6218	0,6111	0,5993	0,5861	0,5709	0,5530	0,5309	0,5017	0,4560	0,4150	0,3349
10	10	0,7290	0,7004	0,6632	0,6540	0,6439	0,6328	0,6202	0,6058	0,5888	0,5678	0,5398	0,4956	0,4557	0,3763
11	11	0,7514	0,7247	0,6898	0,6811	0,6716	0,6611	0,6493	0,6356	0,6195	0,5995	0,5727	0,5302	0,4914	0,4134
12	12	0,7704	0,7454	0,7125	0,7043	0,6954	0,6855	0,6742	0,6613	0,6460	0,6269	0,6013	0,5605	0,5230	0,4466
13	13	0,7867	0,7632	0,7322	0,7245	0,7160	0,7066	0,6959	0,6837	0,6691	0,6509	0,6264	0,5872	0,5510	0,4766
14	14	0,8008	0,7787	0,7493	0,7420	0,7340	0,7250	0,7149	0,7033	0,6894	0,6720	0,6485	0,6109	0,5760	0,5037
15	15	0,8132	0,7923	0,7644	0,7575	0,7499	0,7414	0,7317	0,7206	0,7073	0,6907	0,6683	0,6321	0,5984	0,5282
16	16	0,8242	0,8043	0,7778	0,7712	0,7639	0,7558	0,7467	0,7360	0,7234	0,7075	0,6859	0,6512	0,6186	0,5505
17	17	0,8339	0,8150	0,7898	0,7835	0,7765	0,7688	0,7600	0,7499	0,7377	0,7225	0,7018	0,6684	0,6370	0,5708
18	18	0,8426	0,8246	0,8005	0,7945	0,7879	0,7805	0,7721	0,7623	0,7507	0,7361	0,7162	0,6840	0,6537	0,5895
19	19	0,8505	0,8332	0,8102	0,8045	0,7981	0,7910	0,7830	0,7736	0,7624	0,7484	0,7293	0,6982	0,6689	0,6066
20	20	0,8576	0,8411	0,8190	0,8135	0,8074	0,8006	0,7929	0,7839	0,7731	0,7596	0,7412	0,7112	0,6829	0,6224
21	21	0,8640	0,8482	0,8271	0,8218	0,8159	0,8093	0,8019	0,7933	0,7829	0,7699	0,7521	0,7232	0,6957	0,6370
22	22	0,8699	0,8547	0,8344	0,8293	0,8237	0,8174	0,8102	0,8019	0,7919	0,7793	0,7622	0,7342	0,7076	0,6506
23	23	0,8753	0,8607	0,8412	0,8362	0,8308	0,8247	0,8178	0,8098	0,8002	0,7880	0,7715	0,7443	0,7186	0,6631
24	24	0,8803	0,8663	0,8474	0,8426	0,8374	0,8315	0,8249	0,8171	0,8078	0,7961	0,7800	0,7538	0,7287	0,6748
25	25	0,8849	0,8713	0,8531	0,8485	0,8435	0,8378	0,8314	0,8239	0,8149	0,8035	0,7880	0,7625	0,7382	0,6858
26	26	0,8892	0,8761	0,8585	0,8540	0,8491	0,8437	0,8374	0,8302	0,8215	0,8105	0,7954	0,7707	0,7471	0,6960
27	27	0,8931	0,8805	0,8634	0,8591	0,8544	0,8491	0,8431	0,8360	0,8276	0,8169	0,8023	0,7783	0,7554	0,7056
28	28	0,8968	0,8845	0,8681	0,8639	0,8593	0,8542	0,8483	0,8415	0,8333	0,8230	0,8088	0,7854	0,7631	0,7146
29	29	0,9002	0,8884	0,8724	0,8683	0,8639	0,8589	0,8532	0,8466	0,8387	0,8286	0,8148	0,7921	0,7704	0,7231
30	30	0,9035	0,8919	0,8764	0,8725	0,8682	0,8633	0,8578	0,8514	0,8437	0,8339	0,8205	0,7984	0,7772	0,7311
31	31	0,9065	0,8953	0,8802	0,8764	0,8722	0,8675	0,8622	0,8559	0,8484	0,8389	0,8258	0,8043	0,7837	0,7387
32	32	0,9093	0,8984	0,8838	0,8801	0,8760	0,8714	0,8662	0,8602	0,8528	0,8436	0,8309	0,8099	0,7898	0,7458
33	33	0,9120	0,9014	0,8872	0,8836	0,8796	0,8751	0,8701	0,8641	0,8570	0,8480	0,8356	0,8152	0,7956	0,7526
34	34	0,9145	0,9042	0,8903	0,8868	0,8830	0,8786	0,8737	0,8679	0,8610	0,8522	0,8401	0,8202	0,8010	0,7590
35	35	0,9169	0,9069	0,8934	0,8899	0,8862	0,8819	0,8771	0,8715	0,8647	0,8562	0,8444	0,8249	0,8062	0,7651
36	36	0,9191	0,9094	0,8962	0,8929	0,8892	0,8851	0,8804	0,8749	0,8683	0,8599	0,8484	0,8290	0,8111	0,7709
37	37	0,9212	0,9117	0,8989	0,8956	0,8921	0,8880	0,8834	0,8781	0,8716	0,8635	0,8522	0,8337	0,8158	0,7764
38	38	0,9232	0,9140	0,9015	0,8983	0,8948	0,8909	0,8864	0,8811	0,8748	0,8669	0,8559	0,8377	0,8202	0,7817
39	39	0,9252	0,9161	0,9039	0,9008	0,8974	0,8935	0,8892	0,8840	0,8779	0,8701	0,8594	0,8416	0,8244	0,7867
40	40	0,9270	0,9182	0,9062	0,9032	0,8998	0,8961	0,8918	0,8868	0,8808	0,8732	0,8627	0,8453	0,8285	0,7915
41	41	0,9287	0,9201	0,9084	0,9055	0,9022	0,8985	0,8943	0,8894	0,8836	0,8761	0,8658	0,8488	0,8323	0,7961
42	42	0,9304	0,9219	0,9105	0,9076	0,9044	0,9008	0,8967	0,8920	0,8862	0,8789	0,8688	0,8521	0,8360	0,8005
43	43	0,9320	0,9237	0,9125	0,9097	0,9066	0,9031	0,8990	0,8944	0,8887	0,8816	0,8717	0,8554	0,8396	0,8047
44	44	0,9335	0,9254	0,9145	0,9117	0,9086	0,9052	0,9012	0,8967	0,8911	0,8841	0,8745	0,8584	0,8430	0,8087
45	45	0,9349	0,9270	0,9163	0,9136	0,9106	0,9072	0,9034	0,8989	0,8934	0,8866	0,8771	0,8614	0,8462	0,8126
46	46	0,9363	0,9286	0,9181	0,9154	0,9124	0,9091	0,9054	0,9010	0,8957	0,8889	0,8796	0,8642	0,8493	0,8163
47	47	0,9376	0,9300	0,9197	0,9171	0,9142	0,9110	0,9073	0,9030	0,8978	0,8912	0,8821	0,8669	0,8523	0,8199
48	48	0,9389	0,9315	0,9214	0,9188	0,9160	0,9128	0,9092	0,9049	0,8998	0,8934	0,8844	0,8695	0,8552	0,8233
49	49	0,9401	0,9328	0,9229	0,9204	0,9176	0,9145	0,9110	0,9068	0,9018	0,8954	0,8866	0,8721	0,8579	0,8266
50	50	0,9413	0,9341	0,9244	0,9220	0,9192	0,9162	0,9127	0,9086	0,9037	0,8974	0,8888	0,8745	0,8606	0,8298
60	60	0,9509	0,9449	0,9367	0,9346	0,9323	0,9298	0,9268	0,9234	0,9192	0,9139	0,9066	0,8944	0,8826	0,8562
70	70	0,9578	0,9526	0,9456	0,9438	0,9418	0,9396	0,9370	0,9340	0,9304	0,9258	0,9195	0,9089	0,8986	0,8756
80	80	0,9630	0,9585	0,9522	0,9507	0,9489	0,9470	0,9447	0,9421	0,9389	0,9348	0,9292	0,9199	0,9108	0,8903
90	90	0,9671	0,9630	0,9575	0,9561	0,9545	0,9527	0,9507	0,9484	0,9455	0,9419	0,9369	0,9285	0,9203	0,9020
100	100	0,9704	0,9667	0,9617	0,9604	0,9590	0,9574	0,9556	0,9534	0,9509	0,9476	0,9431	0,9355	0,9280	0,9114
200	200	0,9851	0,9832	0,9807	0,9800	0,9793	0,9785	0,9776	0,9765	0,9752	0,9735	0,9712	0,9673	0,9634	0,9548
300	300	0,9901	0,9888	0,9871	0,9867	0,9862	0,9856	0,9850	0,9843	0,9834	0,9823	0,9807	0,9781	0,9755	0,9696
400	400	0,9925	0,9916	0,9903	0,9900	0,9896	0,9892	0,9887	0,9882	0,9875	0,9867	0,9855	0,9835	0,9816	0,9772
500	500	0,9940	0,9933	0,9922	0,9920	0,9917	0,9914	0,9910	0,9905	0,9900	0,9893	0,9884	0,9868	0,9852	0,9817
600	600	0,9950	0,9944	0,9935	0,9933	0,9931	0,9928	0,9926	0,9921	0,9917	0,9911	0,9903	0,9890	0,9877	0,9847
700	700	0,9957	0,9952	0,9945	0,9943	0,9941	0,9938	0,9936	0,9932	0,9929	0,9924	0,9917	0,9906	0,9894	0,9869
800	800	0,9963	0,9958	0,9951	0,9950	0,9948	0,9946	0,9944	0,9941	0,9937	0,9933	0,9927	0,9917	0,9907	0,9885
900	900	0,9967	0,9963	0,9957	0,9955	0,9954	0,9952	0,9950	0,9947	0,9944	0,9941	0,9935	0,9926	0,9918	0,9898
1000	1000	0,9970	0,9966	0,9961	0,9960	0,9958	0,9957	0,9955	0,9953	0,9950	0,9947	0,9942	0,9934	0,9926	0,9908
1500	1500	0,9980	0,9978	0,9974	0,9973	0,9972	0,9971	0,9970	0,9968	0,9967	0,9964	0,9961	0,9956	0,9951	0,9939

```

> x1 <- rnorm(20, mean=10, sd=5); x2 <- 0.95*x1 + rnorm(20, mean=0, sd=2)
> cor(x1, x2)
[1] 0.9290261
> diff <- x1 - x2; mittel <- (x1 + x2)/2
> mdiff <- mean(diff); mdiff
[1] 0.725
> sdiff <- sd(diff); sdiff
[1] 1.980397
> uplim <- mdiff + 2*sdiff; uplim
[1] 4.685795
> lowlim <- mdiff - 2*sdiff; lowlim

```

```

[1] -3.235795
>
> n      <- length(diff)
> tval   <- qt(0.025, n-1, lower.tail=F)
>
> upp95u <- upplim + tval * sqrt(sdiff^2/n); upp95u
[1] 5.612649
> upp95l <- upplim - tval * sqrt(sdiff^2/n); upp95l
[1] 3.75894
>
> low95u <- lowlim + tval * sqrt(sdiff^2/n); low95u
[1] -2.30894
> low95l <- lowlim - tval * sqrt(sdiff^2/n); low95l
[1] -4.162649

```

Die mittlere Differenz aus den beiden Messwertreihen ist $\bar{d} = 0,73$, mit einer Standardabweichung von $s_d = 1,98$. Der Korrelationskoeffizient beträgt $r = 0,93$ und zeigt somit, dass die beiden Messwertreihen zusammenhängen. Die Grenzwerte für die Übereinstimmung mit Bezug auf die mittlere Differenz nach (6.124) sind $-3,24$ und $4,69$. Die Grenzen unter Beachtung der 95%-Konfidenzintervalle dieser Grenzwerte (6.125) sind dann durch $-4,16$ und $5,61$ gegeben. In Abbildung 6.8 sind die Messwerte und die daraus abgeleiteten Grenzwerte für die Übereinstimmung dargestellt. Neben der Höhe der Abweichungen zeigen sich systematische Unterschiede (Fehler) zwischen den Messungen einerseits durch eine Verschiebung der mittleren Differenz (> 0 oder < 0) und andererseits durch eine Abhängigkeit der Differenz von der Größe der Messungen (z. B. größere Messwerte führen auch zu größeren Fehlern).

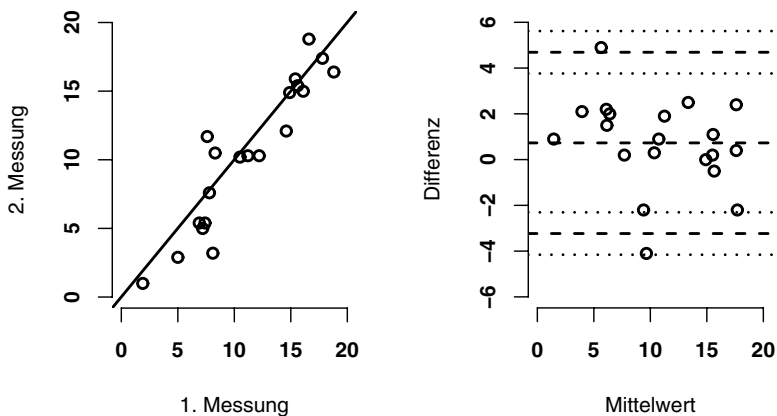


Abb. 6.8. Analyse der Übereinstimmung von Messungen nach Bland-Altman. Punktwolke mit den Messwerten links; Differenz in Abhängigkeit von der Größe der Messungen und Grenzwerte für die Übereinstimmung rechts

Hypothesentest

- Der statistische Test
- Tests der Verteilung (goodness of fit)
- Einstichprobenverfahren
- Zweistichprobenverfahren
- Mehrstichprobenverfahren, varianzanalytische Methoden
- Die Analyse von Häufigkeiten
- Die Bewertung von Zusammenhängen

7.1 Der statistische Test

Folgende nette Geschichte stammt von R.A. Fisher [Fis60]. auf einer Gesellschaft behauptet eine Dame X : Setze man ihr eine Tasse Tee vor, der etwas Milch beigegeben wurde, so könne sie im allgemeinen einwandfrei schmecken, ob zuerst Tee oder ob zuerst Milch eingegossen worden sei. Wie prüft man diese Behauptung? Sicher nicht so: Zwei äußerlich völlig gleichartige Tassen vorsetzen, wobei in die erste zuerst Milch und dann Tee (Reihenfolge MT) und in die zweite zuerst Tee und dann Milch (TM) eingegossen wurde. Würde man jetzt die Dame wählen lassen, so hätte sie offenbar eine Chance von 50% die richtige Antwort zu geben, auch wenn ihre Behauptung falsch ist.

Besser ist folgendes Vorgehen: Acht äußerlich gleiche Tassen nehmen, vier davon in der Reihenfolge MT , die vier anderen in der Reihenfolge TM füllen. Die Tassen zufällig über den Tisch verteilen; dann die Dame herbeirufen und ihr mitteilen, dass von den Tassen je vier vom Typ TM bzw. MT sind, ihre Aufgabe sei, die vier TM -Tassen herauszufinden. Jetzt ist die Wahrscheinlichkeit, ohne eine besondere Begabung die richtige Auswahl zu treffen, sehr gering geworden. Aus 8 Tassen kann man nämlich auf $\frac{8 \cdot 7 \cdot 6 \cdot 5}{4 \cdot 3 \cdot 2} = 70$ Arten 4 auswählen; nur eine dieser 70 Kombinationen ist die richtige.

Die Wahrscheinlichkeit, ohne besondere Begabung, also zufällig, die richtige Auswahl zu treffen, ist daher mit $1/70 = 0,0143$ oder etwa 1,4% sehr gering. Wählt die Dame nun wirklich die 4 richtigen Tassen, so werden wir die **Nullhypothese** – Lady X hat diese „Sonderbegabung“ nicht – fallen lassen und ihr diese besondere Fähigkeit zuerkennen. Dabei nehmen wir eine **Irrtumswahrscheinlichkeit** von 1,4% in Kauf. Natürlich können wir diese Irrtumswahrscheinlichkeit dadurch noch weiter verringern, dass wir die Anzahl der Tassen erhöhen (z. B. auf 12, je zur Hälfte nach TM bzw. nach MT gefüllt, Irrtumswahrscheinlichkeit $\alpha = 0,001$). **Charakteristisch ist für unser Vorgehen:** Wir stellen zunächst die Nullhypothese auf und verwerfen sie genau dann, wenn sich ein Ergebnis einstellt, das bei Gültigkeit der Nullhypothese unwahrscheinlich ist. Stellen wir eine Hypothese auf, die wir mit statistischen Methoden prüfen wollen, so interessiert uns, ob eine vorliegende Stichprobe die Hypothese stützt oder nicht. Im Teetassen-Beispiel würden wir die Nullhypothese verwerfen, wenn die Dame die 4 richtigen Tassen wählt. In jedem anderen Fall behalten wir die Nullhypothese bei. Wir müssen also bei jeder möglichen Stichprobe eine Entscheidung treffen. Im Beispiel wäre auch die Entscheidung vertretbar, die Nullhypothese zu verwerfen, wenn die Dame mindestens 3 richtige Tassen wählt.

Um der Schwierigkeit zu entgehen, sich in jedem konkreten Fall die Entscheidung vorher überlegen zu müssen, sucht man nach Verfahren, die eine solche Entscheidung **stets** herbeiführen. Ein solches Verfahren, das für jede Stichprobe die Entscheidung, ob das Stichprobenergebnis die Hypothese stützt oder nicht, herbeiführt, heißt **statistischer Test**. Die Standardtests in der Statistik sind dadurch ausgezeichnet, dass sie in gewisser Weise optimal sind. Viele Tests setzen voraus, dass die Beobachtungen **unabhängig** sind, wie es in sogenannten Zufallsstichproben der Fall ist. Die meisten statistischen Tests werden mit Hilfe einer **Prüfgröße** (oder Teststatistik) durchgeführt. Eine solche Prüfgröße ist eine Vorschrift, nach der aus einer gegebenen Stichprobe eine Zahl errechnet wird. Der Test besteht nun darin, dass je nach dem Wert der Prüfgröße entschieden wird.

- Entscheidungsprinzipien
- Statistische Hypothesen und Testentscheidungen
- Statistischer Test - Schritt für Schritt
- Powerfunktion und Operationscharakteristik
- Die Formulierung von Hypothesen
- Der P-Wert nach R.A. Fisher
- Äquivalenztests
- Verteilungsunabhängige Verfahren

7.1.1 Entscheidungsprinzipien

Viele unserer Entscheidungen werden gemäß der sogenannten **Minimax-Philosophie** von Abraham Wald (1902–1950) gefällt. Nach dem Minimax-Prinzip wird diejenige Entscheidung bevorzugt, die den maximalen Verlust, der im ungünstigsten Falle zu erwarten ist, zu einem Minimum macht. Der größtmögliche Verlust gibt den Ausschlag. Das ist optimal bei größtmöglicher Risikoscheu; dies führt in vielen Fällen zu einer kaum tragbaren Außerachtlassung großer Chancen. Nur ein chronischer Pessimist wird stets so handeln. Andererseits minimalisiert dieses Prinzip die Chancen eines katastrophalen Verlustes.

Ein „**Minimaxer**“ ist also jemand, der sich so entscheidet, dass er sich möglichst gut (maximal) gegen die denkbar schlimmste Situation (Minimum) verteidigt. Nach dem Minimax-Kriterium wird es jeder Richter vermeiden, unschuldige Personen ins Gefängnis zu schicken. Freisprüche von nicht vollständig überführten Kriminellen sind die Kosten dieses Verfahrens. Ohne „Minimaxer“ gäbe es keine Versicherungen: Nehmen wir an, eine Werkstatt im Werte von DM 100 000 sei zu einer Prämie von DM 5000 gegen Feuer versichert. Die Wahrscheinlichkeit für ein die Werkstatt zerstörendes Feuer betrage 1%. Soll der Verlust möglichst gering sein, dann ist zu bedenken, dass durch den Abschluss der Versicherung ein sicherer Verlust von DM 5000 eintritt, während man – ohne Versicherung – mit einem erwarteten Verlust in Höhe von einem Prozent, das sind nur DM 1000, zu rechnen hat. Der wirkliche Verlust beträgt jedoch Null oder DM 100 000. Daher bevorzugt man vernünftigerweise den sicheren Verlust von DM 5000.

Ist nicht nur ein Objekt zu versichern, sondern handelt es sich um viele – sagen wir 80 Schiffe einer großen Reederei – dann kann es zweckmäßig sein, nur einzelne Schiffe versichern zu lassen oder auch überhaupt keine Versicherung abzuschließen. Schuldenfreie Objekte brauchen nicht versichert zu werden. Der Staat versichert nichts.

Der Vollblutoptimist – in unserer Ausdrucksweise ein „**Maximaxer**“ – wählt die Entscheidung, die unter den günstigsten Umständen (Maximum) die besten Resultate liefert (Maximum) und verzichtet auf den Abschluss einer Versicherung, da ein Werkstattbrand „unwahrscheinlich“ ist. Das Maximax-Kriterium verspricht dann Erfolg, wenn bei relativ kleinen Verlusten große Gewinne möglich sind. Der „Maximaxer“ spielt im Toto und Lotto, da der fast sichere unbedeutende Verlust durch den höchst unwahrscheinlichen großen Gewinn mehr als wettgemacht wird. Dieses Entscheidungsprinzip – bei dem der größtmögliche Gewinn den Ausschlag gibt – geht auf

Bayes (1702–1761) und Laplace (1749–1827) zurück. Bei der **Entscheidungstheorie** geht es um die Analyse von Entscheidungssituationen mit dem Ziel, zu erkunden, wie Entscheidungen gefällt werden, sowie **Handlungsempfehlungen zu geben**.

Entscheidungen und Schlussfolgerungen: Über Entscheidungen gelangt die Wissenschaft zu Schlussfolgerungen. **Entscheidungen** haben den Charakter des „wir entscheiden jetzt als ob“. Mit den Einschränkungen „handeln als ob“ und „jetzt“ tun wir in der besonderen gegenwärtig vorliegenden Situation „unser Bestes“, ohne hiermit zugleich ein Urteil über die „Wahrheit“ im Sinne des $6 > 4$ abzulegen.

Demgegenüber werden **Schlussfolgerungen** – die Maximen der Wissenschaft – unter sorgfältiger Beachtung des aus spezifischen Beobachtungen und Experimenten gewonnenen Beweismaterials gezogen. Nur der „Wahrheitsgehalt“ entscheidet. Fehlt ausreichendes Beweismaterial, so werden Schlussfolgerungen zurückgestellt. Eine Schlussfolgerung ist eine Feststellung, die als anwendbar auf Bedingungen des Experiments oder einer Beobachtung akzeptiert werden kann, solange nicht ungewöhnlich starkes Beweismaterial ihr widerspricht. Diese Definition stellt drei entscheidende Punkte heraus: Sie betont „Annahme“ im eigentlichen Sinne des Wortes, spricht von „ungewöhnlich starkem Beweismaterial“ und enthält die Möglichkeit späterer Ablehnung (vgl. Tukey [Tuk60]). Übrigens, Wahrheit ist eher konstruiert als entdeckt.

7.1.2 Statistische Hypothesen und Testentscheidungen

Eine statistische Hypothese ist eine Behauptung über Eigenschaften einer oder mehrerer Zufallsvariablen, z.B. über deren Parameter (Parameterhypothesen) oder über deren Verteilung (Verteilungshypothesen). Derartige Hypothesen sind in der Regel nur indirekt prüfbar. Beispiele machen sie zwar empirisch sicherer, ohne sie jedoch beweisen zu können. Zur Widerlegung genügt dagegen oft schon ein Gegenbeispiel. Da eine **Arbeitshypothese** (H_A) nie direkt bestätigt werden kann, stellt man eine **Gegenhypothese** (Nicht- H_A oder H_0) auf und versucht, diese zu widerlegen. Hierdurch lässt sich die Arbeitshypothese indirekt bestätigen.

Stellen wir, etwa für den Vergleich zweier Stichprobenmittelwerte bezüglich ihrer Parameter, als Verneinung der eigentlichen Arbeitshypothese H_A (Ungleichheit beider Parameter μ_1 und μ_2) eine Nullhypothese H_0 auf (μ_1 und μ_2 sind gleich: $\mu_1 = \mu_2$ oder $\mu_1 - \mu_2 = 0$), die wir mit statistischen Methoden prüfen wollen, so interessiert uns, ob eine vorliegende Stichprobe die Nullhypothese stützt oder nicht. Ein Verfahren, das für jede Stichprobe diese Entscheidung herbeiführt, heißt **statistischer Test**. Viele statistische Tests setzen voraus, dass die Beobachtungen unabhängig sind, wie es in Zufallsstichproben der Fall ist. Die meisten statistischen Tests werden mit Hilfe einer **Teststatistik** durchgeführt. Das ist eine Vorschrift (Formel), nach der aus einer bzw. häufig aus zwei gegebenen Stichproben eine Zahl (**Prüfgröße**) errechnet wird. Der Test besteht nun darin, dass je nach dem Wert der Prüfgröße **für oder gegen die Nullhypothese** entschieden wird.

Zu der berechneten Prüfgröße kann dann ein **P-Wert** bestimmt werden. Er gibt an, mit welcher Wahrscheinlichkeit eine ebenso große und größere Prüfgröße, wie die berechnete, erwartet werden kann, wenn die Nullhypothese richtig ist, d.h. in unserem Beispiel zwischen den beiden Mittelwerten μ_1 und μ_2 kein Unterschied vorliegt. Ist diese Wahrscheinlichkeit kleiner als üblicherweise 5% (oder 1%), so schließt man: Anhand der vorliegenden Stichproben wird die Nullhypothese auf dem **5%-Signifikanzniveau**, kurz 5%-Niveau (1%-Niveau) abgelehnt. Hierbei wird in 100 Fällen ohne Mittelwertunterschied ($\mu_1 = \mu_2$) im Durchschnitt 5mal (1mal) irrtümlich das Urteil „echter Unterschied“ gefällt. In diesem Fall hat man eine richtige Nullhypothese mit einer **Irrtumswahrscheinlichkeit** α (alpha) in Höhe von 5% (1%) abgelehnt und damit einen **Fehler 1. Art** begangen. Urteilt man im Sinne der Nullhypothese „es besteht kein Mittelwertunterschied ($\mu_1 \neq \mu_2$)“, wenn doch einer vorliegt, so begeht man einen **Fehler 2. Art** mit der Wahrscheinlichkeit β (beta). Die Wahrscheinlichkeit, eine richtig spezifizierte Arbeits- oder Alternativhypothese als solche zu erkennen $P(H_A|H_A)$, wird als **Power** oder Teststärke bezeichnet; diese Wahrscheinlichkeit ist

gleich $1 - \beta$. Für α wählt man im allgemeinen die Werte $\alpha = 0,05$ (5%) oder $\alpha = 0,01$ (1%), seltener auch $\alpha = 0,001$ (0,1%). Je kleiner α ist, desto seltener wird zwar H_0 fälschlicherweise abgelehnt (Fehler 1. Art), aber um so häufiger wird H_0 fälschlicherweise beibehalten (Fehler 2. Art).

Wird die Nullhypothese durch das Stichprobenergebnis (durch den Wert der Prüfgröße) nicht widerlegt, so wird man sich - aus Mangel an Beweisen, nicht etwa wegen erwiesener Richtigkeit - für ein „vorläufiges“ Beibehalten der Nullhypothese entscheiden müssen: Die Beobachtungen sind mit der Nullhypothese vereinbar. Grundsätzlich sind jedoch (z.B. auf Grund extrem ungünstiger Stichprobenergebnisse) zwei Fehlentscheidungen möglich:

- (1) Die **unberechtigte Ablehnung** der Nullhypothese: Fehler 1. Art
- (2) Das **unberechtigte Beibehalten** der Nullhypothese: Fehler 2. Art

Wird also z.B. bei einem Vergleich festgestellt, dass ein neues Medikament besser ist, obwohl in Wirklichkeit dem alten gleichwertig, so liegt ein Fehler 1. Art vor; stellt sich durch den Vergleich heraus, dass beide Medikamente gleichwertig sind, obwohl tatsächlich das neue besser ist, so wird ein Fehler 2. Art begangen.

Tabelle 7.1. Möglichkeiten zur Fehlentscheidung im statistischen Test

Entscheidung des Tests	Wirklichkeit	
	H_0 wahr	H_0 falsch
H_0 abgelehnt (H_A angenommen)	Fehler 1. Art α -Fehler	richtige Entscheidung mit der Power ^a $1 - \beta$
H_0 beibehalten (H_A abgelehnt)	richtige Entscheidung ($1 - \alpha$)	Fehler 2. Art β -Fehler

^a Die Power (Teststärke, Trennschärfe) eines Tests ist die Wahrscheinlichkeit, die der H_0 gegenübergestellte richtige Alternativhypothese H_A , die meist auch die Arbeitshypothese ist, als solche zu erkennen.

Fehler 1. und 2. Art völlig vermeiden, hieße: restlose Ausschaltung von Zufallswirkungen. Das würde den logisch zwingenden Schluss vom Teil auf das Ganze bedeuten, was in vielen Fällen grundsätzlich unmöglich ist. Je nachdem, welche Fehlentscheidung folgenschwer ist, wird man in einem konkreten Fall α und β nach Möglichkeit so festlegen, dass die kritische Wahrscheinlichkeit $\leq 0,01$ und die andere $\leq 0,10$ ist. So wird z.B. bei einer Herstellung eines Impferserums äußerste Konstanz des Serums gefordert. Nicht einwandfreie Chargen müssen rechtzeitig erkannt und eliminiert werden. Das unberechtigte Beibehalten der Nullhypothese H_0 (Serum in Ordnung) bedeutet einen gefährlichen Herstellungsfehler. Man wird also β möglichst klein wählen, während das Verwerfen guter Chargen zwar Unkosten mit sich bringt, im übrigen aber keine ernststen Folgen hat. Ist man gezwungen, α und β auf $\leq 0,01$ festzulegen, und unterscheiden sich beide Hypothesen H_0 und H_A nur wenig voneinander, so lässt sich eine Entscheidung nur mit umfangreichen Stichproben erzwingen. Andererseits lassen sich mit genügend kleinen Stichproben (und großen Varianzen) fast alle H_0 „bestätigen“; und das, obwohl man, insbesondere wenn H_0 vom Typ „kein Unterschied“ ist, mitunter schon vor der Datengewinnung weiß, dass H_0 falsch ist. Nicht nur in diesen Fällen sollten die Vertrauensbereiche für die Differenzen der Parameter angegeben werden. Betont sei auch, dass **formale statistische Signifikanz** (nur diese kann ein statistischer Test nachweisen!) und **wirkliche (praktische) Bedeutung** - Relevanz - nicht miteinander verwechselt werden dürfen.



Beispiel (Prüfung eines Arzneimittels auf Wirksamkeit und auf Unbedenklichkeit): Bei der „Wirksamkeitsprüfung“ von Arzneimitteln besteht H_0 darin, dass das Mittel keine Wirkung hat. Bei der „Unbedenklichkeitsprüfung“ besteht H_0 darin, dass das Mittel keine Nebenwirkung hat. Vergleichen wir jeweils die Fehler 1. und 2. Art. Bei der Untersuchung von Arzneimitteln folgt die Prüfung auf Wirksamkeit der Prüfung auf Unbedenklichkeit, wobei unterschiedliche Akzente gesetzt werden können. Bei der Prüfung auf

(1) **Wirksamkeit** lautet H_0 : Das Mittel ist ohne Wirkung, d. h. mit **kleinem** α werden nur deutlich wirksame Mittel zugelassen; da dann β nicht klein und die Power klein wird, ist ein Abwürgen möglicher zukunftssträchtiger Mittel nicht auszuschließen.

(2) **Unbedenklichkeit** lautet H_0 : Das Mittel ist ohne Nebenwirkung, d. h. mit **nicht kleinem** α werden alle Nebenwirkungen ernst genommen, β wird klein und die Power groß.

Übersicht 17. Die Nullhypothese, der α -Fehler, der β -Fehler und die Power eines Hypothesentests

Beim Hypothesentest gibt es zwei Fehler, die **Nullhypothese**

- fälschlich abzulehnen, α -Fehler genannt;
- fälschlich beizubehalten, β -Fehler genannt.

Der α -Fehler ist ein „falscher Alarm“.

Der β -Fehler ist ein „versäumter Alarm“.

Die Nullhypothese, kurz „ H_0 “:

- abzulehnen, ist eine „starke“ Aussage, besonders, wenn sie „deutlich falsch“ ist;
- beizubehalten, ist eine „schwache“ Aussage.

Ist H_0 richtig, so wird es kaum gelingen, H_0 abzulehnen; **ist jedoch H_0 deutlich falsch**, so wird ein Test eine gute Chance haben, eine korrekte Entscheidung herbeizuführen, d. h. H_0 mit der Wahrscheinlichkeit $1 - \beta$, der sogenannten „**Power**“ (Teststärke oder Trennschärfe) zugunsten der richtigen Alternativhypothese abzulehnen, d. h. diese als solche zu erkennen. Aus der Power, die als Funktion von α , Stichprobenumfang n und Effektgröße betrachtet werden kann, schließt man auf den erforderlichen Stichprobenumfang, d. h. bei großem Effekt wird man mit $\alpha = 0,05$ und kleinem n schon eine gute Power erzielen.

Da α größer als Null sein muss, für $\alpha = 0$ würde man die Nullhypothese immer beibehalten, besteht stets ein α -Fehler. Bei vorgegebenem Stichprobenumfang n und α wird β um so größer, je kleiner wir α vorgeben. Nur wenn n unbeschränkt wachsen darf, können α und β beliebig klein gewählt werden, d. h. bei sehr kleinem α und β kann man die Entscheidung nur mit sehr großen Stichprobenumfängen erzwingen! Bei kleinen Stichprobenumfängen und kleinem α ist die Möglichkeit, tatsächlich vorhandene Unterschiede nachzuweisen, gering: das Ergebnis, es liege kein statistisch signifikanter Unterschied vor, muss dann mit Vorsicht beurteilt werden. Aus der Nichtablehnung einer Nullhypothese lässt sich nichts über deren Gültigkeit erschließen, solange β unbekannt ist!



Wenn wir in diesem Buch den Begriff „signifikant“ verwenden, so stets nur im Sinne von „statistisch signifikant“ (auf dem verwendeten Niveau).

Da „kaum falsche“ Nullhypothesen (H_0) anhand eines statistischen Tests nicht zu erkennen sind, nehmen wir an, H_0 sei entweder „richtig“ oder „deutlich falsch“. Die Entscheidung, H_0 nicht abzulehnen (sondern sie „beizubehalten“) bedeutet nicht, dass H_0 wahrscheinlich richtig ist, sondern, dass H_0 richtig sein könnte (vgl. Übersicht 17). Es ist somit eine „**schwache**“ Aussage. Demgegenüber kann bei der Ablehnung von H_0 – einer „**starken**“ Aussage – davon ausgegangen werden, dass H_0 wahrscheinlich falsch ist (und H_A wahrscheinlich richtig ist). Ist dies nicht der Fall, so kommentiert man z. B. „... gelang es nicht, H_0 auf dem 5%-Niveau abzulehnen.“ Nebenbei bemerkt, den Wissenschaftlern ist im allgemeinen die relative Stärke der Ablehnung von H_0 willkommen, insbesondere dann, wenn H_0 „deutlich“ falsch ist.

Wählt man $\alpha = \beta$, so sind die Wahrscheinlichkeiten für Fehlentscheidungen erster und zweiter Art gleich. Nicht selten wählt man lediglich ein festes α und billigt der Nullhypothese eine Sonderstellung zu, da die Alternativhypothese im allgemeinen nicht genau festliegt. So entscheiden einige Standardverfahren der Statistik mit fest vorgegebenem α und unbestimmtem β zugunsten der Nullhypothese: man bezeichnet sie daher als **konservative Tests**.



Beispiel (Tausend Münzwürfe zur Prüfung der Nullhypothese $\pi = 0,5$): Angenommen, wir kennen aufgrund sehr vieler Versuche mit einer bestimmten Münze deren Wahrscheinlichkeit π für das Ereignis „Wappen“ – sagen einem Freunde aber lediglich, dass π entweder gleich 0,4 oder gleich 0,5 ist. Unser Freund entschließt sich zur Prüfung der Nullhypothese $\pi = 0,5$ für folgenden Versuchsplan. Die Münze wird $n = 1000$ mal geworfen. Ist $\pi = 0,5$, so würden sich vermutlich etwa 500 „Wappen“ einstellen. Unter der Alternativhypothese $\pi = 0,4$ wären etwa 400 „Wappen“ zu erwarten. Der Freund wählt daher folgendes Entscheidungsverfahren: Tritt das Ereignis „Wappen“ weniger als 450mal auf, so lehnt er die Nullhypothese $\pi = 0,5$ ab und akzeptiert die Alternativhypothese $\pi = 0,4$. Tritt es dagegen 450mal oder häufiger auf, so behält er die Nullhypothese bei.

Ein Fehler 1. Art – Ablehnung einer richtigen Nullhypothese – liegt dann vor, wenn π tatsächlich gleich 0,5 ist und trotzdem bei einem speziellen Versuch weniger als 450 „Wappen“ ermittelt werden. Ein Fehler 2. Art wird dann begangen, wenn tatsächlich $\pi = 0,4$ ist und sich bei der Prüfung 450 oder mehr „Wappen“ ergeben. In diesem Beispiel haben wir α und β etwa gleich groß gewählt (vgl. npq ist einmal gleich 250 und zum anderen gleich 240). Man kann aber auch bei vorgegebenem Stichprobenumfang n durch Vergrößerung des Annahmebereiches für die Nullhypothese den Fehler 1. Art verkleinern. Beispielsweise lässt sich verabreden, dass die Nullhypothese $\pi = 0,5$ nur dann abgelehnt wird, wenn sich weniger als 430 „Wappen“ einstellen. Damit wird aber bei konstantem Stichprobenumfang n der Fehler, 2. Art – das Beibehalten der falschen Nullhypothese – um so größer.

7.1.3 Statistischer Test - Schritt für Schritt

Die einzelnen Arbeitsschritte bei einem statistischen Test sollen am Beispiel eines **Parametertests** zusammenfassend dargestellt und in einem Beispiel erläutert werden. Dabei geht man von einer Verteilungsfunktion $F_X(x|\vartheta)$ aus, die von einem **unbekannten Parameter** $\vartheta \in \Omega$ (ϑ griech. theta, Ω griech. Omega) abhängt. Ω bezeichnet dabei den sogenannten **Parameterraum**, d.h. eine Menge von Werten, die der Parameter annehmen kann.

(1) Hypothesenbildung:

Für den unbekannten Parameter werden zwei Hypothesen in alternativer Form formuliert

$$H_0 : \vartheta \in \Omega_0 \quad \text{und} \quad H_A : \vartheta \in \Omega_A = \overline{\Omega_0} \quad (7.1)$$

Die Hypothese H_0 oder H_A heißt

- **einfach** (simple hypothesis), wenn die entsprechende Menge Ω_0 oder Ω_A nur ein Element enthält.
- **zusammengesetzt** (composite hypothesis), wenn die entsprechenden Parametermengen mehr als nur ein Element enthalten; in der Regel bezeichnen Ω_0 oder Ω_A in diesem Fall Intervalle.

Ein Test zu diesen Hypothesen heißt

- **einseitig** (engl. one-sided test), wenn Ω_A einseitig beschränkt ist.
- **zweiseitig** (engl. two-sided test), wenn Ω_A nach oben und nach unten unbeschränkt ist.

Beispiel:

$$H_0 : \pi = 0,5 \quad \text{und} \quad H_A : \pi \neq 0,5$$

bezeichnet bei einer zweiseitigen Fragestellung eine einfache Hypothese gegen eine zusammengesetzte Alternative.

$$H_0 : \pi \leq 0,5 \quad \text{und} \quad H_A : \pi > 0,5$$

bezeichnet bei einer einseitigen Fragestellung eine zusammengesetzte Hypothese gegen eine zusammengesetzte Alternative.



(2) **Fallzahl:**

Die Festlegung der Fallzahl (sample size) n beeinflusst die Varianz der **Prüfgröße** T und damit auch die **Güte des Tests** (power). Je größer n gewählt wird, desto kleiner wird die Varianz der Prüfgröße ausfallen und desto „trennschärfer“ wird ein Test sein (s.u.)

(3) **Signifikanzniveau:**

Das Signifikanzniveau (Fehler 1. Art, α) begrenzt den Fehler, H_0 zu Unrecht zu verwerfen („sup“; lat. supremum, obere Grenze).

$$\alpha = \sup_{\vartheta \in \Omega_0} P(T \in K_\alpha | \vartheta) \quad (7.2)$$

Die Ablehnung der Nullhypothese wird durch das Ereignis $T \in K_\alpha$ beschrieben, wobei K_α einen durch α festgelegten kritischen Bereich (**Ablehnungsbereich**) bezeichnet. Je kleiner α gewählt wird, desto unwahrscheinlicher wird eine Fehlentscheidung. Allerdings steigt dann der Fehler 2. Art (β) die Hypothese H_0 fälschlicherweise beizubehalten. Die Tabelle 7.1 kann danach formal auch wie folgt formuliert werden:

Tabelle 7.2. Wahrscheinlichkeiten^a zu falschen oder richtigen Entscheidungen im statistischen Test

Entscheidung des Tests	Wirklichkeit	
	H_0 wahr	H_0 falsch
H_0 abgelehnt (H_A angenommen)	$P(T \in K_\alpha H_0) \leq \alpha$ Fehler 1. Art	$P(T \in K_\alpha H_A)$ richtige Entscheidung
H_0 beibehalten (H_A abgelehnt)	$P(T \notin K_\alpha H_0) \geq 1 - \alpha$ richtige Entscheidung	$P(T \notin K_\alpha H_A)$ Fehler 2. Art

^a Die Wahrscheinlichkeiten in der Tabelle sind bedingte Wahrscheinlichkeiten, die sich nur in den Spalten zu Eins ergänzen.

(4) **Prüfgröße:**

Bei Wahl einer geeigneten Prüfgröße T (Stichprobenfunktion) ist zu beachten, dass diese von den zu prüfenden Hypothesen abhängt, insbesondere aber, dass die Verteilung von T unter der Nullhypothese H_0 bekannt sein muss. Nur so ist gewährleistet, dass das Signifikanzniveau bei der Testentscheidung eingehalten werden kann.

(5) **Ablehnungsbereich:**

Ein **Ablehnungsbereich** K_α (auch kritischer Bereich, engl. critical region) bzw. sein Komplement der **Annahmebereich** $\bar{K}_\alpha$ (engl. acceptance region) wird so festgelegt, dass das Signifikanzniveau (zumindest) eingehalten wird, d.h. dass $\sup_{\vartheta \in \Omega_A} P(T \in K_\alpha | \vartheta) \leq \alpha$ gilt, und dass andererseits $\sup_{\vartheta \in \Omega_A} P(T \in \bar{K}_\alpha | \vartheta)$ möglichst klein ausfällt. Die Wahrscheinlichkeit dafür, dass ein beobachteter Wert $\hat{t}$ der Prüfgröße T unter der Nullhypothese H_0 (d. h. bei Gültigkeit der H_0) in den Ablehnungsbereich fällt, soll möglichst klein sein und damit eine Entscheidung gegen H_0 begründen.

(6) **Testentscheidung:**

Die Nullhypothese wird abgelehnt, wenn der beobachtete Wert $\hat{t}$ der Prüfgröße T in den kritischen Bereich K_α fällt. Die Beobachtung steht damit in einem signifikanten Widerspruch zu der unter H_0 gemachten Annahme und berechtigt zu deren Ablehnung. In diesem Fall wird die Alternativhypothese H_A angenommen bzw. bestätigt. Die Entscheidung darüber, ob H_A nun richtig oder falsch ist, wird unter der Kontrolle des Fehlers 1. Art getroffen. Wird H_0 nicht abgelehnt, dann kann daraus nicht geschlossen werden, dass H_0 richtig ist. Vielmehr reicht in diesem Fall die Beobachtung nicht aus, um zu einer Ablehnung zu gelangen. Da eine Entscheidung getroffen werden muss, behält man auch weiterhin H_0 bei (unter weiter bestehenden Zweifeln; in der Rechtsprechung entscheidet man sich mangels eines Beweises

für den Angeklagten). Der Fehler, der möglicherweise in dieser Entscheidung liegt, ist jedoch nicht unter statistischer Kontrolle.



Beispiel (Einstichproben-Gauß-Test): Es liegt eine normalverteilte Grundgesamtheit vor, deren Standardabweichung $\sigma = \sigma_0 = 10$ bekannt ist. In einem Parametertest sollen 3 Hypothesen (H_0) bezüglich des unbekannten Erwartungswertes μ überprüft werden.

- (a) $H_A : \mu < \mu_0 = 20$ und $H_0 : \mu \geq \mu_0$
 (b) $H_A : \mu > \mu_0 = 20$ und $H_0 : \mu \leq \mu_0$
 (c) $H_A : \mu \neq \mu_0 = 20$ und $H_0 : \mu = \mu_0$

(a) und (b) bezeichnen einseitige Hypothesen, (c) eine zweiseitige Hypothese. Der Stichprobenumfang wird mit $n = 25$ Beobachtungen festgelegt. Für den Fehler 1. Art wird $\alpha = 0,05$ gewählt. Als Prüfgröße kann der Stichprobenmittelwert $\bar{X}_n$ verwendet werden. Er ist eine Schätzfunktion für μ und ist unter der Annahme der Nullhypothese normalverteilt mit dem Erwartungswert μ_0 und der Varianz σ^2/n . Somit können entsprechend der Hypothesenstellung die folgenden Ablehnungsbereiche K_α konstruiert werden.

- (a) $K_\alpha = (-\infty; \mu_0 - z_{1-\alpha} \frac{\sigma}{\sqrt{n}})$ bzw. $\bar{X}_n < \mu_0 - z_{1-\alpha} \frac{\sigma}{\sqrt{n}}$
 (b) $K_\alpha = (\mu_0 + z_{1-\alpha} \frac{\sigma}{\sqrt{n}}; +\infty)$ bzw. $\bar{X}_n > \mu_0 + z_{1-\alpha} \frac{\sigma}{\sqrt{n}}$
 (c) $K_\alpha = (-\infty; \mu_0 - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}) \cup (\mu_0 + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}; +\infty)$
 bzw. $\bar{X}_n < \mu_0 - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$ oder $\bar{X}_n > \mu_0 + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$

Für einen beobachteten Mittelwert $\bar{x}_n$ aus einer Stichprobe mit $n = 25$ Beobachtungen können dann entsprechend die folgenden Entscheidungen getroffen werden (für die Berechnung beachte $z_{1-\alpha} \cdot \sigma/\sqrt{n} = 1,645 \cdot 10/\sqrt{25} = 3,29$ bzw. $z_{1-\alpha/2} \cdot \sigma/\sqrt{n} = 1,960 \cdot 10/\sqrt{25} = 3,92$):

- (a) Ablehnung von H_0 wenn: $\bar{x}_n \in (-\infty; 16,71)$
 (b) Ablehnung von H_0 wenn: $\bar{x}_n \in (23,29; +\infty)$
 (c) Ablehnung von H_0 wenn: $\bar{x}_n \in (-\infty; 16,08) \cup (23,92; +\infty)$

In äquivalenter Weise können auch die entsprechenden Wahrscheinlichkeiten (P-Werte) betrachtet werden. Für einen beobachteten Stichprobenmittelwert $\bar{x}_{25} = 16$ folgt zum Beispiel:

$$(a) \quad P(\bar{x}_{25} < 16 | \mu = 20; \frac{\sigma}{\sqrt{n}} = \frac{10}{5} = 2) \\ = P(Z < \frac{16 - 20}{2}) = P(Z < -2) = 0,0228$$

$$(b) \quad P(\bar{x}_{25} > 16 | \mu = 20; \frac{\sigma}{\sqrt{n}} = \frac{10}{5} = 2) \\ = P(Z < \frac{16 - 20}{2}) = P(Z > -2) = 0,9332$$

- (c) Sei $D = \bar{x}_{25} - \mu_0$. Unter der Nullhypothese ist D normalverteilt mit dem Erwartungswert 0 und der Varianz $\sigma^2/n = 4$:

$$P(|D| > |\bar{x}_{25} - \mu_0|) = P(|D| > 4) \\ = P(D < -4 \cup D > +4) \\ = P(Z < -2 \cup Z > +2) = 2P(Z < -2) = 0,0455$$

7.1.3.1 Der Likelihood-Quotient; das Neyman-Pearson Lemma

Ein **Niveau- α -Test** ($\alpha \in [0, 1]$) für die Nullhypothese H_0 ($\vartheta \in \Omega_0$) gegen die Alternative H_A ($\vartheta \in \Omega_A$) ist ein **Entscheidungsverfahren** der Form:

- Entscheidung für H_A wenn $T \in K_\alpha$, z.B. $T \geq c_\alpha$
- Entscheidung für H_0 wenn $T \notin K_\alpha$, z.B. $T < c_\alpha$

Dabei ist $T : \Omega \rightarrow \mathbb{R}$ eine Zufallsgröße, die aus den Beobachtungen zu berechnen ist und $c_\alpha \in \mathbb{R}$ muss so gewählt werden, dass der Fehler 1. Art entsprechend (7.2) zumindest eingehalten wird. Die Herleitung und die Begründung von Prüfgrößen mit optimalen Eigenschaften kann auf der Grundlage des **Neyman-Pearson-Lemmas** erfolgen. Danach lässt sich für einfache Hypothesen

$$H_0 : \vartheta = \vartheta_0 \quad \text{und} \quad H_A : \vartheta = \vartheta_A \quad (7.3)$$

ein **gleichmäßig bester Test** (s. auch im nächsten Abschnitt), d.h. ein **optimaler Ablehnungsbereich** K_α aus dem Quotienten der Likelihoodfunktionen ableiten.

$$LQ = \frac{L(\vartheta_A|x)}{L(\vartheta_0|x)} > c_\alpha \quad (7.4)$$

Bei der Entscheidung für die Alternative können der Fehler 1. Art und die Power auch durch die Likelihoodfunktionen für den unbekannten Parameter unter der Annahme / Voraussetzung vorliegender Beobachtungen x beschrieben werden.

$$\alpha = \int_{K_\alpha} L(\vartheta_0|x) dx \quad \text{und} \quad (1 - \beta) = \int_{K_\alpha} L(\vartheta_A|x) dx \quad (7.5)$$

Es ist einsehbar, dass das Verhältnis $(1 - \beta)/\alpha$ ein sinnvolles Kriterium für die Kennzeichnung eines optimalen Ablehnungsbereichs liefert. Die Herleitung einer geeigneten Prüfgröße kann nun so erfolgen, dass für ein festes α eine konstante Größe c_α so festzulegen ist, dass die Bedingung 7.4 für alle Werte $x \in K_\alpha$ eingehalten wird.

Beispiel (Einstichproben-Gauß-Test): Es liegen Beobachtungen x_j ($j = 1, \dots, n$) aus einer normalverteilten Grundgesamtheit vor. Es wird angenommen, dass die Varianz σ^2 bekannt ist. Die einfachen Hypothesen zum unbekannten Erwartungswert sind:

$$H_0 : \mu = \mu_0 \quad \text{und} \quad H_A : \mu = \mu_1 \neq \mu_0$$

Die Likelihood-Quotienten-Bedingung führt (in kurzer Form dargestellt) auf :

$$L(\mu_i|x) = \left(\frac{1}{\sigma\sqrt{2\pi}} \right)^n \cdot \exp \left[-\frac{1}{2\sigma^2} \sum_j (x_j - \mu_i)^2 \right] \quad \text{für } i = 0, 1$$

$$LQ = \frac{L(\mu_1|x)}{L(\mu_0|x)} = \exp \left[\frac{n}{\sigma^2} \left(\bar{x}(\mu_1 - \mu_0) + \frac{1}{2}(\mu_1^2 - \mu_0^2) \right) \right] > c_\alpha$$

$$\bar{x}(\mu_1 - \mu_0) > \frac{1}{2}(\mu_1^2 - \mu_0^2) + \frac{\sigma^2}{n} \cdot \log c_\alpha$$

$$\bar{x} > \frac{1}{2}(\mu_0 + \mu_1) + \frac{\sigma^2}{n} \cdot \frac{\log c_\alpha}{(\mu_1 - \mu_0)} =: A$$

Soll der Test auf dem α -Niveau erfolgen, dann lassen sich aus den folgenden Beziehungen

$$P(\bar{x}_n > A | \mu_0) = \alpha = 1 - \Phi \left(\frac{A - \mu_0}{\sigma/\sqrt{n}} \right)$$



$$P(\bar{x}_n \leq A | \mu_1) = \beta = \Phi\left(\frac{A - \mu_1}{\sigma/\sqrt{n}}\right)$$

die Werte für A (Grenzwert für den Ablehnungsbereich) und n (Anzahl der Beobachtungen) direkt ableiten

$$\begin{aligned} A &= \mu_0 + z_{1-\alpha} \frac{\sigma}{\sqrt{n}} \\ n &= \frac{(z_\beta - z_{1-\alpha})^2}{(\mu_0 - \mu_1)^2} \cdot \sigma^2 \end{aligned} \quad (7.6)$$

7.1.4 Powerfunktion und Operationscharakteristik

Für die Überprüfung einer Nullhypothese bieten sich häufig verschiedene Testverfahren an. Die zusammenfassende Beurteilung der Güte (Qualität) eines Tests erfolgt durch die **Gütefunktion** (power function) und durch die **Operationscharakteristik** (operation characteristic). Anhand dieser Kriterien ist die Auswahl einer speziellen Prüfgröße bzw. Teststatistik objektiv zu rechtfertigen.

Die Gütefunktion beschreibt die Ablehnungswahrscheinlichkeit in Abhängigkeit von der dem zu schätzenden Parameter ϑ .

$$G(\vartheta) = P(T \in K_\alpha | \vartheta) \quad (7.7)$$



Danach kennzeichnet die Gütefunktion die Wahrscheinlichkeit für eine Fehlentscheidung (Fehler 1. Art, α), wenn $\vartheta \in \Omega_0$ und für eine richtige Entscheidung (Power, $1 - \beta$) wenn $\vartheta \in \Omega_A$.

$$\sup_{\vartheta \in \Omega_0} G(\vartheta) = \alpha \quad (7.8)$$

Ein Test mit dem Signifikanzniveau α heißt **konservativ** (conservative test), wenn er die vorgegebene Irrtumswahrscheinlichkeit nicht voll ausschöpft (Zurückgehen auf die sichere Seite), d.h. für die Powerfunktion gilt:

$$\sup_{\vartheta \in \Omega_0} G(\vartheta) < \alpha \quad (7.9)$$

Ein Test mit dem Signifikanzniveau α heißt **unverfälscht** (unbiased test), wenn die Nullhypothese, sofern sie nicht zutrifft, mit mindestens so hoher Wahrscheinlichkeit verworfen wird, wie im Falle ihres Zutreffens, d.h. für die Gütefunktion gilt:

$$G(\vartheta) \geq \alpha \quad \text{für alle } \vartheta \in \Omega_A \quad (7.10)$$

Ein Test heißt **konsistent** (consistent test), wenn mit wachsendem Stichprobenumfang die Ablehnung der Nullhypothese immer wahrscheinlicher wird.

$$\lim_{n \rightarrow \infty} P(T_n \in K_\alpha | \vartheta \in \Omega_A) = 1 \quad (7.11)$$

Ein Test heißt **gleichmäßig bester Test** (uniformly most powerful), auch trennscharfer Test, wenn die Werte der Gütefunktion für Parameter $\vartheta \in \Omega_A$ mindestens so groß sind, wie für einen beliebigen anderen Test auf dem gleichen Niveau α .

Zwei Stichprobenverteilungen einer gegebenen Prüfgröße oder Teststatistik seien durch die beiden Glockenkurven dargestellt (Abbildung 7.1), die linke repräsentiere die Nullhypothese (H_0), die rechte eine spezifizierte (einseitige) Alternativhypothese (H_A). Erhalten wir nun aufgrund eines bestimmten Entscheidungsverfahrens einen kritischen Wert für die Teststatistik, dann sind – je nach Lage der aus einer Stichprobe empirisch ermittelten Teststatistik – zwei Entscheidungen möglich. Erreicht oder überschreitet dieser Wert der Teststatistik den kritischen Wert, dann wird

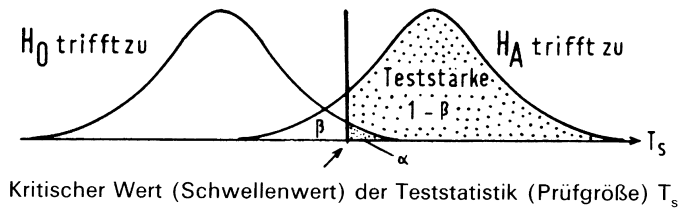


Abb. 7.1. Die Teststärke, Trennschärfe oder Power

die Nullhypothese abgelehnt, d.h. die Alternativhypothese akzeptiert. Wird der kritische Wert durch die Teststatistik nicht erreicht, dann besteht keine Veranlassung, die Nullhypothese abzulehnen, d.h. sie wird beibehalten. Abb. 7.2 zeigt, dass je nach Lage des kritischen Wertes der Teststatistik – bei konstantem Abstand zwischen den mittleren Teststatistiken für $H_0(T_{s1})$ und $H_A(T_{s2})$ – mit kleiner werdender Irrtumswahrscheinlichkeit α der β -Fehler zunimmt und die Power $1 - \beta$ abnimmt.

Der möglichst kleine β -Fehler, eine falsche Nullhypothese beizubehalten, hängt ab:

- (1) Vom **Umfang der Stichprobe** n : Je größer die Stichprobe ist, um so eher wird bei gegebener Irrtumswahrscheinlichkeit α ein Unterschied zwischen zwei Grundgesamtheiten entdeckt werden.
- (2) Vom **Grad des Unterschieds** δ (griech. delta) zwischen dem hypothetischen und dem wahren Zustand des zu erfassenden Effektes, das ist der Betrag δ , um den die Nullhypothese falsch ist.
- (3) Von der Eigenart des Tests, die man als **Power** bezeichnet. Die Teststärke, **Trennschärfe** oder **Power** ist um so größer:
 - a) Je höher der vom Test verwendete Informationsgehalt der Ausgangsdaten ist – nimmt also in der Reihe: Häufigkeiten, Rangplätze und Messwerte zu.
 - b) Und je mehr Voraussetzungen über die Verteilung der Werte gemacht werden: Ein Test, der Normalverteilung und Varianzhomogenität erfordert, ist im allgemeinen wesentlich stärker als einer, der keinerlei Voraussetzungen macht.

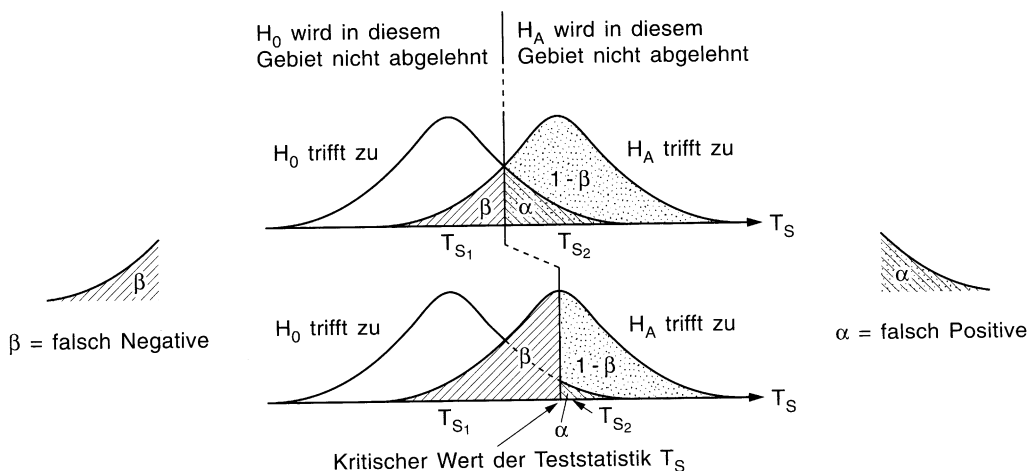


Abb. 7.2. Kritischer Wert der Teststatistik (Prüfgröße) in Abhängigkeit von α (und β)

Die **Power eines Tests** ist die Wahrscheinlichkeit H_0 abzulehnen, wenn die spezielle einfache H_A richtig ist. Sie hängt damit zumindest ab von δ , α , n und von der Gerichtetheit oder Seitigkeit des Tests (zwei- oder einseitiger Test).

$$\text{Power} = P(\text{Entscheidung } H_0 \text{ abzulehnen} \mid H_A \text{ trifft zu}) = 1 - \beta \quad (7.12)$$

Je kleiner bei vorgegebenem α die Wahrscheinlichkeit β ist, desto schärfer trennt der Test H_0 und H_A . Ein Test heißt **trennscharf** (powerful), wenn er im Vergleich zu anderen möglichen Tests bei vorgegebenem α eine relativ hohe Trennschärfe aufweist. Wenn H_0 wahr ist, ist die Maximalpower eines Tests gleich α .

Wählen wir $\alpha = 0,05$ und eine Power von 0,8, d. h. $\beta/\alpha = 0,20/0,05 = 4$, so bedeutet dies, dass der α -Fehler viermal so wichtig ist wie der β -Fehler (vgl. dagegen 0,1/0,1).

Wie viele Beobachtungen sind erforderlich?

Zu kleine Stichprobenumfänge sind nicht einmal in der Lage, **große Unterschiede** zwischen zwei Parametern zu erfassen; zu große Stichprobenumfänge entdecken **winzige Unterschiede**, die praktisch bedeutungslos sind. Daher muss man sich zunächst überlegen, welcher Unterschied (oder Effekt), falls vorhanden, unbedingt gefunden werden soll. Danach ist festzulegen, mit welcher Wahrscheinlichkeit oder Power zumindest dieser Unterschied/Effekt gefunden werden soll: im allgemeinen wird man einen Test nur durchführen, wenn die Power deutlich größer ist als 0,5. Für die Irrtumswahrscheinlichkeit wird bei zwei- bzw. einseitiger Fragestellung in vielen Fällen das 5%-Niveau ($\alpha = 0,05$) bevorzugt. Will man hiervon abweichen, so sollten alle Konsequenzen bedacht und die Abweichung kommentiert werden.

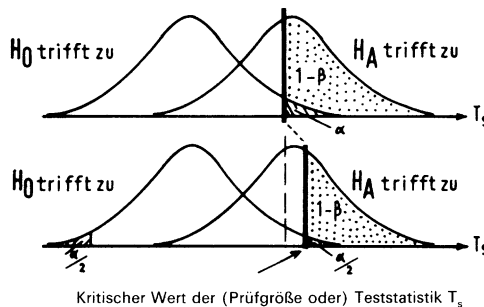


Abb. 7.3. Abhängigkeit der Power von der ein oder zweiseitigen Fragestellung

Nur bei großem n oder bei großem Unterschied δ wird sich dann, wenn ein sehr kleines α vorgegeben wird, statistische Signifikanz ergeben. Daher begnügt man sich häufig mit dem 5%-Niveau und einer Power von mindestens 70%, besser von etwa 80%. Beliebig lässt sich die Trennschärfe nur durch wachsenden Stichprobenumfang erhöhen. Es sei daran erinnert, dass Zufallsstichproben mit unabhängigen Beobachtungen vorausgesetzt werden.

Beim Übergang von der einseitigen auf die zweiseitige Fragestellung vermindert sich die Power. Für Abb. 7.3 würde das bedeuten: Das „Dreieck“ α wird halbiert, der kritische T_S -Wert wandert nach rechts, erhöht sich, β wird größer und die Power $1 - \beta$ kleiner. Bei gleichem Stichprobenumfang ist ein **einseitiger Test stets trennschärfer** als der zweiseitige.

Die in Abb. 7.4 stark schematisiert gezeichneten Teststärkekurven zeigen die **Power als Funktion des Unterschieds zwischen zwei Erwartungswerten**. Ein Test ist bei gegebener Parameter-Differenz um so stärker, je größer n und α werden. Für α ist der uns zur Verfügung stehende Variationsbereich natürlich nur klein, da wir das Risiko, eine wahre Nullhypothese abzulehnen, im Normalfall nur ungenügend über 5% anwachsen lassen werden:

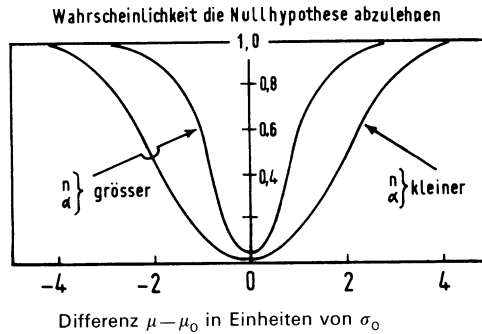


Abb. 7.4. Teststärkekurven (Gütefunktionen) für unterschiedliche Bedingungen bei zweiseitiger Fragestellung, die mittlere Ordinate gibt für beide Kurven die Irrtumswahrscheinlichkeiten ($\alpha \approx 0,01$ bzw. $\alpha \approx 0,03$), mit zunehmendem α und n nähern sich die napfförmigen Kurven ihrer Symmetrieachse, der Ordinate; alles schematisiert

- (1) Besteht zwischen den Erwartungswerten der Grundgesamtheiten kein Unterschied, so werden wir, wenn wir mit der Irrtumswahrscheinlichkeit (dem Signifikanzniveau) α arbeiten, in $\alpha\%$ der Fälle die Nullhypothese zu Unrecht aufgeben.
- (2) Besteht zwischen den Erwartungswerten ein Unterschied von **1,5 Einheiten** von σ_0 , so wird der stärkere Test, die **engere** umgekehrte Glockenkurve der Abb. 7.4, bei 100 Stichproben **80mal** den bestehenden Unterschied nachweisen (Power = 0,80). Dagegen wird der schwächere Test – die weite umgekehrte Glockenkurve – ziemlich versagen; er wird nur in 30% der Fälle den Unterschied aufdecken (Power = 0,30).
- (3) Besteht zwischen den Erwartungswerten ein sehr großer Unterschied, dann haben beide Kurven die Power 1.

Wir haben somit gesehen, dass beim zweiseitigen Test mit zunehmendem Abstand $\mu - \mu_0$ die Wahrscheinlichkeit, die Nullhypothese abzulehnen, zunimmt und dass es mit kleiner werdendem Signifikanzniveau und mit kleiner werdendem Stichprobenumfang schwieriger wird, eine wahre Alternativhypothese zu akzeptieren. Auch hieraus ersehen wir, dass zur Erzielung einer guten Teststärke möglichst große Stichprobenumfänge verwendet werden sollten. Ist der Stichprobenumfang klein, dann sollte das Signifikanzniveau nicht zu klein sein, da sowohl die kleine Stichprobe als auch ein kleines Signifikanzniveau sich durch unerwünschte Senkung der Power bemerkbar machen.

Der einseitige Test ist, wie wir gesehen haben, durch eine größere Power ausgezeichnet als der zweiseitige. Da der einseitige Test damit bestehende Unterschiede eher aufdeckt als der zweiseitige, wird die einseitige Fragestellung bevorzugt, wenn die zweiseitige Fragestellung offensichtlich sinnwidrig ist. Wird beispielsweise eine neue Therapie mit einer allgemein praktizierten verglichen, Vorprüfungen haben bereits stattgefunden, dann ist nur die Frage interessant, ob die neue Therapie besser ist. Ist die neue Methode weniger wirksam oder genau so wirksam, dann besteht keine Veranlassung, von der alten Methode abzugehen. Stehen aber zwei neue Methoden im Vergleich, dann ist die zweiseitige Fragestellung die einzig brauchbare. Nicht zuletzt deshalb, weil der einseitige Test gegenüber der anderen oder sagen wir „falschen“ Alternativhypothese nahezu unempfindlich ist.

Verteilungsfreie Tests, besser **verteilungsunabhängige Tests** genannt, besonders Schnelltests, sind gegenüber den parametrischen Tests durch eine geringere Power charakterisiert. Hat man wirklich einmal normalverteilte oder homogen variante Messwerte zu analysieren, so nimmt man bei Anwendung verteilungsfreier Tests einen höheren Fehler 2. Art in Kauf. Die statistische Entscheidung ist dann **konservativ**, d. h. man hält „länger“ als geboten an der Nullhypothese fest und



kommt etwas seltener zu statistisch signifikanten Befunden, oder anders ausgedrückt: zur Verwerfung der Nullhypothese sind größere Stichproben nötig. Liegen kleine Stichproben vor ($n \lesssim 20$), dann sind verteilungsunabhängige Tests nicht selten wirksamer als die sonst optimalen parametrischen Tests, die für umfangreiche Stichproben meist wirksamer sind. Kommen für eine Analyse mehrere Tests in Frage, so ist im allgemeinen derjenige Test zu bevorzugen, der den Informationsgehalt der Ausgangsdaten am vollständigsten ausschöpft. Verlangt wird natürlich, dass die Grundvoraussetzungen des dem Test zugrundeliegenden statistischen Modells (Zufallsstichproben bzw. randomisierte Beobachtungen) seitens der Ausgangsdaten erfüllt sind, eine Verallgemeinerung berechtigt ist und der Test genau der Fragestellung entspricht. Bei jedem Test ist es sinnvoll (wenn möglich), neben einem festen α **nach Neyman und Pearson** auch den **P-Wert nach R.A. Fisher** anzugeben.

Stets ist zu beachten: Aussagen in der Statistik sind höchstens so sicher wie die Voraussetzungen dieser Aussagen.

Je mehr Voraussetzungen ein Test hat, um so höher ist im allgemeinen seine Power. Prüft man einen Teil oder sämtliche Voraussetzungen anhand mehrerer **Vortests** auf einem bestimmten Signifikanzniveau α bzw. auf unterschiedlichen Niveaus, so ist die Irrtumswahrscheinlichkeit des eigentlichen Tests nicht mehr korrekt, da sie durch die Vortests in undurchsichtiger Weise modifiziert worden ist.

Sind die **Voraussetzungen eines Testverfahrens nicht oder nur teilweise erfüllt**, so muss dieses in der entsprechend vorsichtigen Interpretation des Resultates berücksichtigt werden. Es ist zu empfehlen, die Voraussetzungen, deren Erfüllung unsicher ist, namentlich zu nennen, etwa:

- „Unter der Voraussetzung, dass die beiden Stichproben normalverteilten Grundgesamtheiten entstammen, besteht ...“
- „Wegen des beeinträchtigten Zufallsstichproben-Charakters der vorliegenden Stichproben lassen sich die Resultate lediglich auf eine hypothetische Grundgesamtheit verallgemeinern, die sich durch Vergrößerung der Fallzahlen vorstellen lässt.“
- „Dieses ist als formalisierte Datenbeschreibung aufzufassen. Sie gibt lediglich eine grobe Übersicht über Die angeführten Wahrscheinlichkeits- und Signifikanzaussagen treffen im strengen Sinn nicht zu.“
- „Geht man davon aus, dass keine echten Zufallsstichproben vorliegen und dass die hier vorliegenden Kollektive für den betrachteten Zusammenhang nicht untypisch sind, dann lassen sich die in der Arbeit angegebenen Signifikanzaussagen als formalisierte Datenbeschreibung auffassen: streng genommen treffen diese Aussagen nicht zu, trotzdem möchte man ungern auf sie verzichten.“



Im Zweifelsfall ist es besser auf statistische Tests zu verzichten und sich mit einer Beschreibung der Daten anhand tabellarischer und graphischer Darstellungen zu begnügen. Folgende Warnung hilft mit, Irrtümer und Nachlässigkeiten zu vermeiden:

Ein „Durchprobieren“ der Tests ist nicht zulässig. Durch fast ausschließliche Verwendung von einseitigen Tests und die Auswahl eines Tests aufgrund der Resultate wird praktisch erreicht, dass die effektive Irrtumswahrscheinlichkeit mitunter mehr als doppelt so gross ist wie die vorgegebene Irrtumswahrscheinlichkeit [Wal64].

7.1.5 Die Operationscharakteristik

Abbildung 7.4 gibt die Teststärke an in Abhängigkeit von der Differenz der Erwartungswerte in Einheiten der Standardabweichung $[(\mu - \mu_0)/\sigma_0]$, die Teststärkefunktion (power function) oder wie man auch sagt, die Trennschärfe- oder Gütefunktion. Ihr Komplement, die Wahrscheinlichkeit,

eine falsche Nullhypothese beizubehalten, d. h. einen Fehler 2. Art zu begehen, wird **Operationscharakteristik** OC, OC-Kurve (operating characteristic curve) oder **Annahmekennlinie** genannt.

$$OC(\vartheta) = P(T \in K_\alpha | \vartheta) = 1 - G(\vartheta) \quad (7.13)$$

Sie beschreibt damit die Wahrscheinlichkeit einer Fehlentscheidung (Fehler 2. Art, β) wenn $\vartheta \in \Omega_A$ und für eine richtige Entscheidung ($1 - \alpha$) wenn $\vartheta \in \Omega_0$.

$$\sup_{\vartheta \in \Omega_A} OC(\vartheta) = \beta \quad (7.14)$$

OC-Kurven sind bei zweiseitiger Fragestellung „glockenförmige Komplemente“ der napfförmigen Gütefunktionen. Wir können nun zur Kennzeichnung eines Tests eine dieser beiden Funktionen heranziehen und z. B. anhand der OC für gegebenes α und n das zur Unterscheidung zwischen Null- und Alternativhypothese, zur Entdeckung des Unterschieds δ (griech. delta) unvermeidbare β ablesen. Wenn für gegebenes α bei kleinem β der benötigte Stichprobenumfang zur Aufdeckung von δ zu groß wird, muss das α vergrößert werden. Mitunter kann man allerdings auch einen trennschärferen oder mächtigeren Test verwenden. Die OC würde bei gleichem Stichprobenumfang **steiler** verlaufen und damit einen Unterschied besser erkennen lassen. Ist ein Versuch abgeschlossen, dann zeigt die OC, welche Chance man hat, um δ zu entdecken. War bei kleinem Stichprobenumfang ebenfalls ein kleines α vorgegeben worden, dann ist ein großes β zu erwarten und ein Beibehalten der Nullhypothese nur mit Vorsicht zu akzeptieren, da unter diesen Bedingungen auch ein deutlicher Unterschied kaum hätte entdeckt werden können. Große Bedeutung hat die OC für die Festlegung von **Stichprobenplänen** im Rahmen der Qualitätsüberwachung, insbesondere der Abnahmeprüfung. 

7.1.5.1 Die OC-Kurve in der Qualitätskontrolle

Die Prüfung der Qualität einer Charge (Produktionseinheit, lot) ist nicht erforderlich, wenn von einer bekannten Güte ausgegangen werden kann oder wenn die Qualität keine Rolle spielt. Die gelegentlich Überprüfung dieser Voraussetzung bzw. Annahme ist zu empfehlen. Eine 100%-Prüfung ist dazu im Gegenteil notwendig, wenn von der Qualität das Leben oder die Gesundheit von Menschen abhängt. Zwischen diesen Extremen liegt die **regelmäßige Annahme- oder Abnahmeprüfung** nach einem festen Stichprobenplan (acceptance sampling). Mit diesem ist eine **Entscheidungsregel** verbunden, nach der eine Charge angenommen oder zurückgewiesen werden kann. Besondere Gründe für eine Abnahmeprüfung sind

- zerstörende Prüfungen.
- hohe Kosten (für eine 100%-Püfung).
- hoher Zeitaufwand.

Ein einfacher **Prüfplan** (lot by lot) wird durch das Tripel

$$(N, n, c) \quad (7.15)$$

vollständig festgelegt. Dabei bezeichnet N den Umfang einer Charge, n die Größe der Stichprobe, die einer Charge entnommen wird und c die **Annahmehzahl** (kritischer Wert). Die Charge wird zurückgewiesen, wenn die Zahl d der defekten Elemente in der Stichprobe größer ist als die Annahmehzahl c ($d > c$), anderenfalls wird die Charge angenommen. Die diesem Vorgehen zugrunde liegende Entscheidungsregel kann formal auch als Hypothesentest aufgefasst werden.

$$\begin{array}{ll} H_0 : \pi \leq p_0 & \text{Annahme} \\ H_A : \pi > p_0 & \text{Ablehnung} \end{array} \quad (7.16)$$

Auf der Grundlage des Verteilungsmodells einer Binomialverteilung (exakter wäre hier das Modell der hypergeometrischen Verteilung, insbesondere wenn N klein ist), bezeichnet π in (7.16) den

unbekannten Anteil fehlerhafte Elemente in der Charge und p_0 ist ein Anteil für den maximal zulässigen (vertretbaren) Ausschuss (**kritische Qualitätslage**).

Die Risiken, die Nullhypothese H_0 fälschlicherweise abzulehnen (α -Fehler) oder fälschlicherweise beizubehalten (β -Fehler) erhalten in diesem Zusammenhang eine besondere Bedeutung:

- α begrenzt das **Produzentenrisiko**, fälschlicherweise eine Charge zurückzuweisen.
- β begrenzt das **Konsumentenrisiko**, fälschlicherweise eine Charge zu akzeptieren.

Diesen Risiken können, zum Beispiel dem Verteilungsmodell einer Binomialverteilung (7.17) entsprechend, Kennzahlen zur Bewertung der Qualitätslage gegenübergestellt werden.

- AQL legt die Grenze für eine **tolerable bzw. akzeptable Ausschussquote** (acceptance quality level) fest. Damit wird diejenige (gute) Qualitätslage in einer Charge bezeichnet, bei der die Wahrscheinlichkeit für eine Annahme zum „Schutz des Produzenten“ einen fest vorgegebenen Wert, oft 95% oder 99%, nicht unterschreitet. Je größer der Wert für AQL ist, desto geringer ist auch die Wahrscheinlichkeit, Chargen anzunehmen.
- RQL kennzeichnet die Grenze für eine **nicht mehr zu akzeptierende Ausschussquote** (rejectable quality level). Damit wird die (schlechte) Qualitätslage in einer Charge bezeichnet, bei der die Wahrscheinlichkeit für eine Annahme zum „Schutz des Konsumenten“ einen fest vorgegebenen Wert, oft 5% oder 1%, nicht überschreitet.

$$\begin{aligned} (1 - \alpha) &= \sum_{d=0}^c \frac{n!}{d!(n-d)!} AQL^d (1 - AQL)^{n-d} = P_{AQL} \\ (\beta) &= \sum_{d=0}^c \frac{n!}{d!(n-d)!} RQL^d (1 - RQL)^{n-d} = P_{RQL} \end{aligned} \quad (7.17)$$

Für festes α und β (bzw. AQL und RQL) können die Gleichungen (7.17) nach den für den Stichprobenplan erforderlichen Werten n und c gelöst werden. Anschaulich kann der Zusammenhang in einer OC-Kurve (Abbildung 7.5) dargestellt werden.

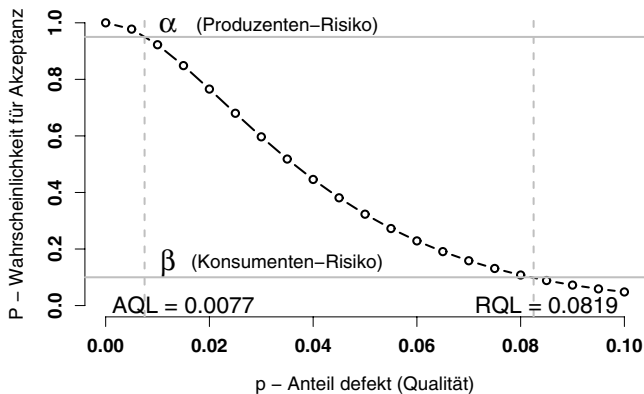


Abb. 7.5. OC-Kurve für einen Stichprobenplan ($N = 1000$, $n = 46$, $c = 1$), mit $\alpha = 0,05$ ($AQL = 0,0077$) und $\beta = 0,10$ ($RQL = 0,0810$)



Beispiel: In einer Stichprobe von 46 Elementen (aus einer Charge mit 1000 Elementen; das Modell einer Binomialverteilung ist näherungsweise zulässig) wird ein fehlerhaftes Element akzeptiert. Damit kann das Produzentenrisiko auf $\alpha=5\%$ (akzeptable Qualitätslage $AQL=0,77\%$) und das Konsumentenrisiko auf $\beta=10\%$ (nicht akzeptable Qualitätslage $RQL=8,1\%$) begrenzt werden (vgl. Abbildung 7.5).

Die Qualität einer Produktion oder eines Prozesses wird durch den Stichprobenplan (N, n, c) verbessert, da ein bestimmter Anteil an Chargen zurückgewiesen und genauer untersucht wird. Die Kurve zum „**mittleren Durchschlupf**“ (average outgoing quality, AOQ) (7.18) beschreibt die Verbesserung der Qualität aufgrund des Stichprobenplans.

$$AOQ = p \cdot P_A \cdot \frac{N - n}{N} \quad (7.18)$$

Dabei bezeichnet p die Wahrscheinlichkeit defekter Elemente in der Produktion (incoming quality) und P_A die Wahrscheinlichkeit, eine Charge nach dem vorliegenden Stichprobenplan zu akzeptieren (vgl. Abbildung 7.5). Für große Chargen kann $(N - n)/N \sim 1$ angenommen werden. Abbildung 7.6 zeigt die Kurve für den mittleren Durchschlupf für die Zahlen aus dem Beispiel. Das Maximum dieser Kurve AOQL (average outgoing quality limit) kann als ein Maß für die Verbesserung der Qualität herangezogen werden.

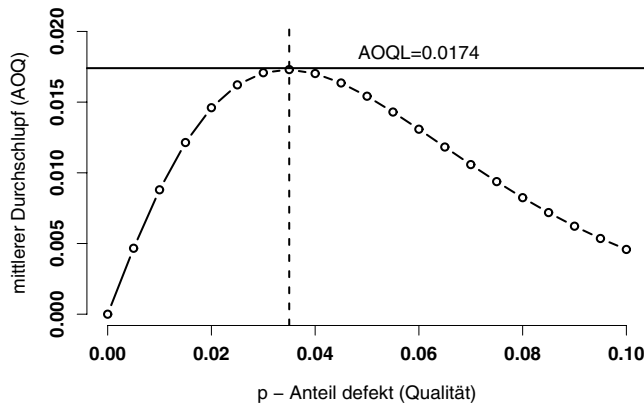


Abb. 7.6. AOQ-Kurve für den mittleren Durchschlupf zum Stichprobenplan ($N = 1000$, $n = 46$, $c = 1$)

7.1.6 Die Formulierung von Hypothesen

7.1.6.1 Alternativhypothese als Gegensatz zur Nullhypothese

Beim statistischen Test werden Alternativ- und Nullhypothese ungleich behandelt. Mit der Festlegung des Fehlers 1. Art zielt man darauf ab, dass die Nullhypothese möglichst selten und nur mit einer kontrollierten Fehlerwahrscheinlichkeit irrtümlich abgelehnt wird. **Hinter der Nullhypothese steht somit häufig ein anerkannter Sachverhalt, den man nicht leichtfertig in Frage stellt oder verwerfen möchte.** Die Ablehnung oder die Beibehaltung der Nullhypothese stellen sehr unterschiedliche Qualitäten (Risiken) dar, die besonders bei der Wahl von α zu berücksichtigen sind. Als Faustregel gilt, dass man diejenige Hypothese unter der Alternative H_A formuliert, die bestätigt bzw. geprüft werden soll, mit der man also zu **neuer Erkenntnis** gelangen will. Somit wird die Alternativhypothese in der Regel durch diejenigen vertreten (festgelegt), der ein Interesse an der Untersuchung dieser Fragestellung hat (z.B. der Geldgeber). Bei Ablehnung von H_0 gilt dann H_A als statistisch bestätigt! Insbesondere kann die Beibehaltung von H_0 nicht als statistische Bestätigung dieser selbst formulierten Annahme aufgefasst werden!



7.1.6.2 Durch Daten angeregte Hypothesen

Ein noch zu häufig begangener Fehler ist die Festlegung der „eigentlichen Fragestellung“ nach der Beobachtung, im krassesten Fall die „Statistische Bestätigung“ eines auffälligen Ergebnisses an denselben Beobachtungen.

Statistische Tests setzen voraus, dass sie nicht erst aufgrund sorgfältiger Betrachtung der Daten ausgewählt werden, sondern bereits vor der Datengewinnung in allen Einzelheiten festliegen. Denn jeder Datenkörper wird auch bei echten, identisch verteilten Zufallsvariablen Anomalien irgendwelcher Art aufweisen, die auf dem üblichen Niveau statistisch signifikant sind, obwohl diese Anomalien in der Grundgesamtheit nicht auftreten. Prüft man auf dem 5%-Niveau, so wird man bei Gültigkeit der Nullhypothese in 5 von 100 Fällen statistisch signifikante Befunde finden. Da viele Abweichungsmuster möglich sind und selten echte Zufallsstichproben vorliegen, wird wenigstens eine Anomalie viel häufiger auftreten. Benutzt man also dieselben Daten zur Auswahl und zugleich zur Prüfung von Hypothesen, so wird eine verlässliche statistische Aussage unmöglich. Im allgemeinen wird man Voruntersuchungen (Pilotstudien, explorative Studien) durchführen und im Anschluss hieran die statistischen Hypothesen formulieren und an neuen Beobachtungen überprüfen. Gegen die Möglichkeit, dass unbekannte Faktoren die Untersuchung stören oder das Resultat verfälschen, sichert man sich z.B. durch Randomisierung. Während der Auswertung auftretende Fragen sind erst aufgrund weiterer neuer Untersuchungen zu prüfen. Häufig ist es möglich, einen Teil der Daten zur Gewinnung und den Hauptteil zur Prüfung der Hypothesen zu verwenden.

Zufallsstichproben gestatten es, zuvor aufgestellte Hypothesen auf ihre mutmaßliche, im besten Falle wahrscheinliche, Richtigkeit hin zu prüfen. Sind die Voraussetzungen eines Hypothesentests weitgehend erfüllt, so gilt ein Ergebnis als „wahrscheinlich“; sind sie nur teilweise erfüllt, was die Regel sein dürfte, so gilt es, sich mit einer höchstens „mutmaßlichen“ Richtigkeit zu begnügen, insbesondere auch dann, wenn man vor Datengewinnung keine Überlegungen bezüglich des Fehlers 1. Art (α), der Power ($1 - \beta$) und der notwendigen Stichprobenumfänge anstellt.

7.1.7 Der P-Wert nach R.A. Fisher

P-Wert: Signifikanztest und Hypothesentest

Der *P*-Wert nach R.A. Fisher [Fis73] ist unter der Annahme einer Nullhypothese H_0 die Wahrscheinlichkeit (Signifikanzwahrscheinlichkeit, genauer: Likelihood der H_0 bei gegebenen Daten) für eine gegebene (beobachtete) experimentelle Situation und **extremere** Situationen. Ist diese Wahrscheinlichkeit, als **Überschreitungswahrscheinlichkeit** bezeichnet,

die, entsprechend dem Vorwissen und der Problemlage als Evidenzmaß für die Glaubwürdigkeit der H_0 der gerade hier vorliegenden und extremeren Daten aufzufassen ist,

und nicht als beobachtete Irrtumswahrscheinlichkeit eines Hypothesentests nach Neyman und Pearson [NP33],

klein, so lässt sich H_0 ablehnen; diese Prozedur wird Signifikanztest genannt.

Dagegen ist der genannte Hypothesentest eine **Entscheidungshilfe**, die mit fest vorgegebener Irrtumswahrscheinlichkeit bei wiederholter Anwendung eine falsche H_0 hinreichend häufig zugunsten der Alternativhypothese ablehnen wird.

Es ist üblich, bei der Anwendung statistischer Testverfahren nicht nur die Entscheidung hinsichtlich der Hypothesenstellung zu einem festen Signifikanzniveau anzugeben, sondern auch die *P*-Werte einzeln aufzuführen. Ein Grund liegt sicherlich in der Verfügbarkeit statistischer Programm-

systeme, in denen zu der ermittelten Teststatistik die Werte der entsprechenden Verteilungsfunktion direkt berechnet werden können. Derartige **P-Werte** haben ihren Ursprung schon in den Arbeiten von Pearson, der zum χ^2 -Anpassungstest P-Werte berechnete.

Übersicht 18. *P*-Wert und Sternsymbolik; historisch - aber auch heute noch gebräuchlich

Ist diese Forderung der Vorgabe einer festen Irrtumswahrscheinlichkeit ausnahmsweise nicht zu erfüllen, dann kann man (1) aber auch den aufgrund der Daten zur Ablehnung der Nullhypothese erreichten **P-Wert**, das nominelle Signifikanzniveau angeben. Es hat den Vorteil, einen vollständigen Lagebericht zu geben. Außerdem gestattet es dem Leser, sein (eigenes) problemgerechtes Signifikanzniveau festzulegen und beide zu vergleichen. Besser ist folgendes Vorgehen (2): Im allgemeinen bezeichnet man ein $P > 0,05$ als statistisch nicht signifikant (ns). Für $P \leq 0,05$ gibt man anhand der kritischen 5%-, 1%- und 0,1%-Schranken an, zwischen welchen Grenzen P liegt und kennzeichnet statistisch signifikante Befunde durch die **dreistufige Sternsymbolik**:

$$[*]0,05 \geq P > 0,01 \quad [**]0,01 \geq P > 0,001 \quad [***]P \leq 0,001 .$$

Der P -Wert ist die Wahrscheinlichkeit, eine mindestens so große Prüfgröße wie die aus den Daten berechnete zu erhalten, wenn in Wirklichkeit die H_0 gilt; d. h. der P -Wert ist die **Überschreitungswahrscheinlichkeit**, mit der man sich irrt, wenn man die Nullhypothese ablehnt. Der P -Wert ist somit die wahre aus den Daten gewonnene (empirische) Irrtumswahrscheinlichkeit, die nicht verwechselt werden darf mit der vor Testbeginn festgelegten Irrtumswahrscheinlichkeit α (auch α -Fehler genannt). Gilt H_0 , dann ist der Wert P gleichverteilt, andernfalls liegt er nahe bei 0. Beachtet werden muss, dass die Wahrscheinlichkeit, die Nullhypothese beizubehalten, obwohl die genau beschriebene Alternativhypothese gilt (der β -Fehler), hier beim P -Wert unberücksichtigt bleibt.

Bei der Beurteilung von P -Werten ist zu unterscheiden, ob die Hypothesen vor der Datengewinnung formuliert wurden oder erst danach; in diesem Fall ist ihre Aussage sehr begrenzt.

Der **P-Wert** ist ein nützliches und informatives Maß für die Evidenz einer Hypothese, insbesondere dann, wenn bei einseitiger Hypothesenstellung der Ablehnungsbereich an nur einer Seite der Verteilung der Teststatistik liegt. Bei rechtsseitigem Ablehnungsbereich gilt für den P -Wert der Prüfgröße (X):

$$P(X) = 1 - F(X|H_0)$$

Dabei ist $F(X|H_0)$ die Verteilungsfunktion von X bei Gültigkeit der Nullhypothese. **Insbesondere ist der P-Wert danach eine Zufallsvariable, deren (asymptotische) Verteilung unter der Nullhypothese durch eine Rechteckverteilung im Intervall $[0, 1]$ beschrieben werden kann.** Welcher Wert als ausreichend angesehen wird, um „berechtigte Zweifel“ an der Nullhypothese (also Evidenz gegen die Annahme), zu hegen, ist aus der Situation / Fragestellung sehr unterschiedlich zu begründen. Fisher hat als Orientierungshilfe die Werte 0,05 und 0,01 empfohlen, aber auch größere oder kleinere Werte sind unter Umständen gerechtfertigt.

Bei zweiseitigem Ablehnungsbereich ist die Angabe von P -Werten nicht frei von Fehlinterpretationen und Absurditäten. Gegen die Verwendung des P -Wertes spricht insbesondere, dass er nichts über die Güte (Power) einer Entscheidung (Test) aussagt. Mit der Beibehaltung der Nullhypothese ist ebenso ein Risiko für eine Fehlentscheidung verbunden (vgl. Fehler 2. Art) wie für deren Ablehnung. Diese Sicht auf das Testproblem ist durch Neyman und Pearson formalisiert worden.

Bedenkenswerte **Alternativen** zum 5%-Signifikanzniveau und **Argumentationen** zum P -Wert enthalten die folgenden Tabellen:

Nr.	n	H_0	α
1	klein	[nicht plausibel	0,10]
2		plausibel	0,05]
3	mittel	nicht plausibel	0,05
4		plausibel	0,01
5	groß	nicht plausibel	0,01
6		[plausibel	0,001]

P	gegen H_0 spricht
$> 15\%$	nichts
10% bis $< 15\%$	kaum etwas
5% bis $< 10\%$	einiges, manches
1% bis $< 5\%$	vieles
$< 1\%$	sehr vieles

Die **Wiederholung** und Veröffentlichung eines Befundes bringt wenig Ehre ist aber wichtig zur Bestätigung und verringert dadurch die Chancen für einen α -Fehler; außerdem gewährt jede Wiederholung Kenntnisse über den zunehmenden Bereich unterschiedlicher Realisierungsbedingungen und damit auch erste **Verallgemeinerungen**. Zusätzlich sollten: der Zufallsfehler sinken und mögliche Verzerrungen erkannt und beseitigt werden.

7.1.8 Äquivalenztests

Das Äquivalenzproblem unterscheidet sich von dem klassischen Testproblem aus den vorangehenden Abschnitten durch eine spezielle Form der Formulierung der Alternativhypothese. Dabei wird unter Äquivalenz die **Gleichheit eines Parameters ϑ mit einem geeigneten Sollwert ϑ_0 „bis auf praktisch irrelevante Abweichungen“** verstanden. Hinsichtlich der Äquivalenz (Übereinstimmung) zweier Parameter, z.B. der Erwartungswerte μ_1 und μ_2 aus zwei Grundgesamtheiten, kann ϑ als Differenz dieser Parameter aufgefasst ($\vartheta = \mu_1 - \mu_2$) und $\vartheta_0 = 0$ gesetzt werden.

Die Präzisierung dieses Äquivalenzbegriffes führt zur **Festlegung eines Äquivalenzbereiches** $[\vartheta_0 - \varepsilon_1, \vartheta_0 + \varepsilon_2]$, der häufig auch symmetrisch um den Sollwert angenommen wird ($\varepsilon_1 = \varepsilon_2 = \varepsilon$). Die zu prüfende Alternativhypothese (7.19) behauptet somit, dass der unbekannte Parameter im Äquivalenzbereich liegt (die H_0 verneint dies).

$$\begin{aligned} H_A : \vartheta_0 - \varepsilon_1 < \vartheta < \vartheta_0 + \varepsilon_2 \\ H_0 : \vartheta \leq \vartheta_0 - \varepsilon_1 \quad \text{oder} \quad \vartheta \geq \vartheta_0 + \varepsilon_2 \end{aligned} \quad (7.19)$$

Der Annahmebereich zu H_0 in (7.19) besteht aus zwei Teilbereichen, die bei der Definition geeigneter Teststatistiken zu berücksichtigen sind. Eine Konstruktion von Teststatistiken nach dem Prinzip der **Intervallinklusion** liegt nahe und führt unter anderem zu einem besseren Verständnis von Äquivalenztests. Ausgehend von zwei $(1 - 2\alpha)$ -Konfidenzintervallen kann ein Äquivalenztest als **Kombination aus zwei einseitigen Tests** angesehen werden, ohne dabei hinsichtlich der Äquivalenzaussage das Signifikanzniveau α zu überschreiten.

Für Äquivalenztests gibt es drei Hauptanwendungsgebiete, Beispiele werden in den folgenden Abschnitten zu konkreten Fragestellungen angegeben.

- (1) Tests zur Überprüfung von **Modellvoraussetzungen** statistischer Verfahren, wie z.B. die Annahme der Varianzhomogenität in der Varianzanalyse (praktisch werden heute immer noch die klassischen Testverfahren auf Varianzunterschiede als Vortests gerechnet!).
- (2) Nachweis von **Bioäquivalenz** in Studien zur **Bioverfügbarkeit**.
- (3) Nachweis von **therapeutischer Äquivalenz** (nicht Unterlegenheit) in klinischen Studien.

7.1.9 Verteilungsunabhängige Verfahren

Die klassischen statistischen Verfahren setzen allgemein Normalverteilung voraus, die streng genommen jedoch nie vorliegt, so daß jede Anwendung ein mehr oder weniger unbefriedigendes Gefühl hinterlässt. Aus diesem Grund wurde die Entwicklung verteilungsfreier oder verteilungsunabhängiger Methoden, die die **Normalverteilung nicht voraussetzen**, mit großem Interesse verfolgt. Bei diesen Tests wird über die Form der Verteilung keinerlei Voraussetzung gemacht, wenn nur gewährleistet oder zumindest plausibel erscheint, dass die zu vergleichenden Zufallsstichproben derselben Grundgesamtheit angehören. Man bezeichnet verteilungsunabhängige Methoden, da Parameter kaum eine Rolle spielen (nichtparametrische Hypothesen!), auch als **parameterfreie oder nichtparametrische Methoden**. Sie sind meistens numerisch einfach zu handhaben. Ihr Vorteil besteht darin, dass man praktisch überhaupt **keine Kenntnisse über die den Daten zugrundeliegende Verteilungsfunktion** zu haben braucht. Darüber hinaus können diese meist leichter verständlichen Verfahren auch auf **Rangdaten** und qualitative Informationen angewendet werden. Unter den folgenden Voraussetzungen ist beispielsweise der klassische Mittelwertvergleich nach „Student“ anwendbar:

- (1) Unabhängigkeit der Beobachtungsdaten (Zufallsstichproben!).
- (2) Das Merkmal muss in Einheiten einer metrischen Skala messbar sein.
- (3) Die Grundgesamtheiten müssen (zumindest angenähert) normalverteilt sein.
- (4) Die Varianzen müssen gleich sein ($\sigma_1^2 = \sigma_2^2$).

Die dem „Student“-Test entsprechenden verteilungsunabhängigen Verfahren fordern lediglich unabhängige Daten. Ob die Beobachtungsdaten voneinander **unabhängig** sind, muss aus der Art ihrer Gewinnung geschlossen werden. So ist die praktisch einzige Voraussetzung lediglich, dass alle Daten oder Datenpaare zufallsmäßig und unabhängig voneinander aus ein und derselben Grundgesamtheit von Daten entnommen worden sind, was durch den Aufbau und die Durchführung der Untersuchung gewährleistet sein muss.

Da ein verteilungsunabhängiger Test, wenn man ihn auf normalverteilte Messwerte anwendet, stets schwächer ist als der entsprechende parametrische Test, wird nach Pitman [Pit49] der Index E_n

$$E_n = \frac{n \text{ für den parametrischen Test}}{n \text{ für den nichtparametrischen Test}} \quad (7.20)$$

als „**Effizienz**“ (Wirksamkeit) des nichtparametrischen Tests bezeichnet. Hierbei bezeichnet n den jeweils erforderlichen Stichprobenumfang zur Erzielung einer gegebenen Teststärke. Der Begriff „**asymptotische Effizienz**“ meint die Wirksamkeit des Tests im Grenzfall einer unendlich großen Stichprobe normalverteilter Messwerte. In diesem Index kommt zum Ausdruck, wie wirksam oder wie leistungsfähig ein verteilungsunabhängiger Test ist, wenn er anstelle eines klassischen Tests auf normalverteilte Daten angewendet wird. Eine asymptotische Effizienz von $E = 0,95$ – wie ihn beispielsweise der U -Test aufweist – bedeutet: Wenn man bei Anwendung des nichtparametrischen Tests im Durchschnitt eine Stichprobe von $n = 100$ Messwerten für eine bestimmte Signifikanzstufe benötigt, so käme man bei Anwendung des entsprechenden parametrischen Tests mit $n = 95$ Messwerten aus. Die so genannten **Rangsummentests** setzen stetige Verteilungen voraus; mehrfach auftretende gleiche Messwerte erschüttern weniger diese Kontinuitätsannahme, sie betonen eher die Ungenauigkeit der Messmethode. Da die Wahrscheinlichkeit für den Fehler 2. Art von der speziellen Verteilung unter der Alternativhypothese abhängt, lässt sie sich nicht allgemein angeben. Für die Prüfung eines Unterschieds wird diese Wahrscheinlichkeit um so kleiner sein, je größer der tatsächliche Unterschied ist. Rangsummentests haben beim Vorliegen einer Normalverteilung zwar einen etwas größeren Fehler 2. Art; ihr Fehler 1. Art, die Irrtumswahrscheinlichkeit α , gilt jedoch für jede Verteilung.



Übersicht 19. Nichtparametrische Tests und verteilungsunabhängige Verfahren. Übrigens ermöglichen auch verteilungsunabhängige Verfahren, die, wie der U -Test zumindest angenähert formgleiche Grundgesamtheiten voraussetzen, die Prüfung der Gleichheit zweier Parameter, der Mediane $\tilde{\mu}_1$ und $\tilde{\mu}_2$ ($H_0 : \tilde{\mu}_1 = \tilde{\mu}_2$)

Parameter sind Größen zur Charakterisierung bestimmter Zufallsvariablen; so sind μ , und σ die Parameter einer Normalverteilung $[N(\mu; \sigma)]$.

Die Verteilung einer Zufallsvariablen X wird durch deren Wahrscheinlichkeitsdichte (X_{stetig}) bzw. durch deren Wahrscheinlichkeitsfunktion (X_{diskret}) charakterisiert. Jede Zufallsvariable X wird insbesondere durch ihre Verteilungsfunktion $F(x)$ charakterisiert.

HYPOTHESEN

Parametrische Hypothesen

z. B. $H_0 : \mu_1 = \mu_2$

TESTS

verteilungsgebundenen Verfahren
wenn ..., dann t -Test

werden
geprüft
anhand
von

Nichtparametrische Hypothesen

z. B. $H_0 : F_1(x) = F_2(x)$

verteilungsunabhängigen Verfahren
wenn ..., dann U -Test

So ist z. B. auch die Angabe eines 95%-Vertrauensbereiches für den Median einer Grundgesamtheit unbekannter Verteilung ein verteilungsunabhängiges Verfahren.

Verteilungsunabhängige Verfahren sind dann angezeigt, wenn (a) das parametrische Verfahren wenig robust gegenüber gewissen Abweichungen von den Voraussetzungen ist, oder wenn (b) die Erzwingung dieser Voraussetzungen durch eine geeignete Transformation (b_1) bzw. durch Beseitigung von Ausreißern (b_2) Schwierigkeiten bereitet; allgemein: bei Nicht-Normalität (1), bei Daten, die einer Rangskala oder einer Nominalskala (vgl. weiter unten) entstammen (2) sowie zur Kontrolle eines parametrischen Tests (3) und als Schnelltest (4). Verteilungsfreie Tests, die sich durch die Kürze des Rechengangs auszeichnen, werden als **Schnelltests** bezeichnet. Die Eigenart dieser Tests ist neben ihrer Rechenökonomie ihre weitgehend voraussetzungsfreie Anwendbarkeit. Ihr Nachteil ist geringe Power, denn nur ein Teil der im Zahlenmaterial enthaltenen Informationen wird zur statistischen Entscheidung herangezogen! Verglichen mit dem einschlägigen optimalen parametrischen oder nichtparametrischen Test ist die statistische Entscheidung eines **Schnelltests konservativ**; d. h. er hält länger als geboten an der Nullhypothese fest oder anders formuliert: es sind größere Stichproben von Messwerten, Rang- oder Alternativdaten erforderlich, um die Nullhypothese zu verwerfen.

7.2 Tests der Verteilung (goodness of fit)

- Der Quotient R/s
- Überprüfung des 3. und 4. Momentes
- Das Wahrscheinlichkeitsnetz, QQ-Plot
- Der Chiquadrat-Anpassungstest
- Kolmogoroff-Smirnoff-Anpassungstest
- Shapiro-Wilk Test
- Anderson-Darling Test
- Ausreißerproblem

Zahlreiche statistische Testverfahren setzen das Modell einer normalverteilten Zufallsvariablen in der Grundgesamtheit voraus. Diese Annahme ist zu begründen bzw. anhand der vorliegenden Beobachtungen (Messwerte) zu überprüfen, bevor weitere Hypothesen aufgestellt und bearbeitet werden können.

7.2.1 Der Quotient R/s

Im Bereich $\mu \pm 3\sigma$ ist ein Anteil von 99,73% einer Normalverteilung enthalten. Dieser Bereich kann erweitert werden auf $\pm 4\sigma$ (99,9937%) und noch extremer auf $\pm 5\sigma$ (99,999942%). Ein sehr einfaches Kriterium („quick-and-dirty“) für einen Test auf Normalverteilung ergibt sich somit, wenn man die Spannweite einer Stichprobe durch 6 dividiert und das Ergebnis mit der empirischen Standardabweichung vergleicht. Allerdings wird die Spannweite gerade bei kleinen Stichproben recht groß ausfallen und sehr stark variieren. Nur bei sehr großen Stichproben wird sich die Spannweite (bei Vorliegen einer Normalverteilung) dem Wert 6σ annähern.

David und Mitarbeiter [DHP54] haben für einen Test auf Normalverteilung diese Überlegung präzisiert. Ist das Verhältnis von Spannweite zur Standardabweichung zu klein oder zu groß, dann zeigt dieses eine Abweichung von der Normalverteilung an.

$$\frac{\text{Spannweite}}{\text{Standardabweichung}} = \frac{R}{s} \quad (7.21)$$

Tabelle 7.3 (Pearson und Stephens [PS64]) enthält ausgewählte Quantile zu dem Verhältnis der Spannweite zur Standardabweichung beim Vorliegen einer Normalverteilung. Liegt das Verhältnis bei einer Stichprobe nicht innerhalb der kritischen Grenzen in dieser Tabelle, dann kann die Hypothese (Nullhypothese) hinsichtlich des Vorliegens einer Normalverteilung auf den entsprechenden Signifikanzniveau abgelehnt werden.

Dieser Test besitzt eine sehr gute Power für den Test auf Normalverteilung gegen zahlreiche alternative, speziell symmetrische Verteilungsformen, wie z.B. gegen eine Gleichverteilung mit ähnlichen Skalen- und Lageparametern. Die Power ist allerdings schlecht bei Vorliegen einer Verteilung, die steiler („spitzer“) als eine Normalverteilung ist, insbesondere wenn gegen eine „log-normale“ Verteilung getestet werden soll.

Beispiel: Aus einer Stichprobe mit $n = 40$ Beobachtungen ergibt sich für $R = 5$ und $s = 1,27$. Das Verhältnis ist $R/s = 5/1,27 = 4,44$. Für $n = 40$ lassen sich aus Tabelle 7.3 folgende Bereiche ablesen:

α	Bereich
0%	1,98 - 8,83
1%	3,47 - 5,56
5%	3,67 - 5,16
10%	3,79 - 4,96



Tabelle 7.3. Kritische Grenzen des Quotienten R/s (aus E.S. Pearson and M.A. Stephens: The ratio of range to standard deviation in the normal sample. Biometrika 51 (1964) 484-487, p. 486, table 3)

n	untere Quantile						obere Quantile					
	Signifikanz-Niveau											
	0,000	0,005	0,01	0,025	0,05	0,10	0,10	0,05	0,025	0,01	0,005	0,000
3	1,732	1,735	1,737	1,745	1,758	1,782	1,997	1,999	2,000	2,000	2,000	2,000
4	1,732	1,83	1,87	1,93	1,98	2,04	2,409	2,429	2,439	2,445	2,447	2,449
5	1,826	1,98	2,02	2,09	2,15	2,22	2,712	2,753	2,782	2,803	2,813	2,828
6	1,826	2,11	2,15	2,22	2,28	2,37	2,949	3,012	3,056	3,095	3,115	3,162
7	1,871	2,22	2,26	2,33	2,40	2,49	3,143	3,222	3,282	3,338	3,369	3,464
8	1,871	2,31	2,35	2,43	2,50	2,59	3,308	3,399	3,471	3,543	3,585	3,742
9	1,897	2,39	2,44	2,51	2,59	2,68	3,449	3,552	3,634	3,720	3,772	4,000
10	1,897	2,46	2,51	2,59	2,67	2,76	3,57	3,685	3,777	3,875	3,935	4,234
11	1,915	2,53	2,58	2,66	2,74	2,84	3,68	3,80	3,903	4,012	4,079	4,472
12	1,915	2,59	2,64	2,72	2,80	2,90	3,78	3,91	4,02	4,134	4,208	4,690
13	1,927	2,64	2,70	2,78	2,86	2,96	3,87	4,00	4,12	4,244	4,325	4,899
14	1,927	2,70	2,75	2,83	2,92	3,02	3,95	4,09	4,21	4,34	4,431	5,099
15	1,936	2,74	2,80	2,88	2,97	3,07	4,02	4,17	4,29	4,44	4,53	5,292
16	1,936	2,79	2,84	2,93	3,01	3,12	4,09	4,24	4,37	4,52	4,62	5,477
17	1,944	2,83	2,88	2,97	3,06	3,17	4,15	4,31	4,44	4,60	4,70	5,657
18	1,944	2,87	2,92	3,01	3,10	3,21	4,21	4,37	4,51	4,67	4,78	5,831
19	1,949	2,90	2,96	3,05	3,14	3,25	4,27	4,43	4,57	4,74	4,85	6,000
20	1,949	2,94	2,99	3,09	3,18	3,29	4,32	4,49	4,63	4,80	4,91	6,164
25	1,961	3,09	3,15	3,24	3,34	3,45	4,53	4,71	4,87	5,06	5,19	6,93
30	1,966	3,21	3,27	3,37	3,47	3,59	4,70	4,89	5,06	5,26	5,40	7,62
35	1,972	3,32	3,38	3,48	3,58	3,70	4,84	5,04	5,21	5,42	5,57	8,25
40	1,975	3,41	3,47	3,57	3,67	3,79	4,96	5,16	5,34	5,56	5,71	8,83
45	1,978	3,49	3,55	3,66	3,75	3,88	5,06	5,26	5,45	5,67	5,83	9,38
50	1,980	3,56	3,62	3,73	3,83	3,95	5,14	5,35	5,54	5,77	5,93	9,90
55	1,982	3,62	3,69	3,80	3,90	4,02	5,22	5,43	5,63	5,86	6,02	10,39
60	1,983	3,68	3,75	3,86	3,96	4,08	5,29	5,51	5,70	5,94	6,10	10,86
65	1,985	3,74	3,80	3,91	4,01	4,14	5,35	5,57	5,77	6,01	6,17	11,31
70	1,986	3,79	3,85	3,96	4,06	4,19	5,41	5,63	5,83	6,07	6,24	11,75
75	1,987	3,83	3,90	4,01	4,11	4,24	5,46	5,68	5,88	6,13	6,30	12,17
80	1,987	3,88	3,94	4,05	4,16	4,28	5,51	5,73	5,93	6,18	6,35	12,57
85	1,988	3,92	3,99	4,09	4,20	4,33	5,56	5,78	5,98	6,23	6,40	12,96
90	1,989	3,96	4,02	4,13	4,24	4,36	5,60	5,82	6,03	6,27	6,45	13,34
95	1,990	3,99	4,06	4,17	4,27	4,40	5,64	5,86	6,07	6,32	6,49	13,71
100	1,990	4,03	4,10	4,21	4,31	4,44	5,68	5,90	6,11	6,36	6,53	14,07
150	1,993	4,32	4,38	4,48	4,59	4,72	5,96	6,18	6,39	6,64	6,82	17,26
200	1,995	4,53	4,59	4,68	4,78	4,90	6,15	6,39	6,60	6,84	7,01	19,95
500	1,998	5,06	5,13	5,25	5,37	5,49	6,72	6,94	7,15	7,42	7,60	31,59
1000	1,999	5,50	5,57	5,68	5,79	5,92	7,11	7,33	7,54	7,80	7,99	44,70

Das Verhältnis liegt noch im kleinsten dieser Bereiche und spricht damit nicht gegen die Annahme einer Normalverteilung.

7.2.2 Überprüfung des 3. und 4. Momentes

Die Abweichung einer empirischen Verteilung von dem Modell einer Normalverteilung (Nullhypothese) kann durch die **Schiefte** (skewness) oder durch die **Wölbung** (kurtosis) beurteilt wer-

den. Die entsprechenden empirischen Maßzahlen sind in (7.22) angegeben (vgl. auch Abschnitt [5.2.3]). Die Schreibweisen $\sqrt{b_1}$ und b_2 stellen den Bezug zu Tabelle 7.4 her.

$$\begin{aligned}
 \text{Schiefe: } g_1 &= \frac{\sqrt{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\sqrt{\left[\sum_{i=1}^n (x_i - \bar{x})^2 \right]^3}} = \sqrt{b_1} \\
 \text{Wölbung: } g_2 &= \frac{n \sum_{i=1}^n (x_i - \bar{x})^4}{\left[\sum_{i=1}^n (x_i - \bar{x})^2 \right]^2} - 3 = b_2
 \end{aligned} \tag{7.22}$$

Die Bezeichnungen für die Parameter im Modell einer Normalverteilung sind:

$$\gamma_1 = 0 \quad \text{und} \quad \gamma_2 = 3 \tag{7.23}$$

Tabelle 7.4 (aus Sachs [Sac84]) enthält Quantile für die Verteilung der 3. und 4. Momente einer Normalverteilung. Da die Verteilung von $\sqrt{b_1}$ symmetrisch zum Wert Null ist, können die tabellierten Werte mit negativem Vorzeichen auch für die unteren Quantile verwendet werden.

Für eine symmetrische Verteilung gilt $g_1 \simeq 0$, speziell für die $N(0;1)$ -Verteilung **$g_2 \simeq 3$.** Häufig wird der Wert für das 4. Moment durch $g'_2 = g_2 - 3$ transformiert, um die Interpretation zu erleichtern. Ist g_1 positiv, dann liegt eine linkssteile Verteilung vor, bei negativen Werten eine rechtssteile Verteilung. Eine Verteilung mit **Hochgipfligkeit** – steiler als die Normalverteilung – oder positivem Exzess weist einen positiven Wert g'_2 auf; eine Verteilung mit negativer Wölbung – flacher als die Normalverteilung – ist durch einen negativen Wert g'_2 charakterisiert, der, genau genommen, „peakedness combined with tailedness“ oder „lack of shoulders“ misst und daher **bei einer bimodalen Verteilung stark negativ** ist. Die Rechteckverteilung mit ausgeprägter „Schulterpartie“ hat daher auch eine negative Wölbung ($g'_2 = -1,2$). Dies gilt sogar für jede Dreieckverteilung ($g'_2 = 0,6$), die gegenüber einer Normalverteilung mit gleicher Varianz eine stärker ausgebildete „Schulterpartie“ aufweist.

Die Berechnung von Schiefe und Steilheit sollte in **R** direkt erfolgen. Bei der Verwendung von vorgegebenen Funktionen, z.B. die Funktionen **skewness()** und **kurtosis()** in **library(e1071)** (Dimitriadou [DHL⁺05]), muss auf deren spezielle Definition geachtet werden. Insbesondere die Verwendung der empirischen Standardabweichung **sd()** kann, wie das folgende Beispiel zeigt, hier zu unterschiedlichen Ergebnissen führen!

```

> x <- c(rep(30, 16), 50, 70, 90, 110)           # Daten
> n <- length(x); m <- mean(x)
> sqrt(n)*sum((x-m)^3) / sqrt(sum((x-m)^2)^3)     # skewness
[1] 2.146625
> n * sum((x-m)^4) / (sum((x-m)^2))^2             # kurtosis
[1] 6.248
> library(e1071)
> skewness(x)                                     # Definition in e1071
[1] 1.987658
> kurtosis(x)+3                                   # Definition in e1071
[1] 5.63882

```

Für die Überprüfung einseitige oder zweiseitiger Hypothesen zu den 3. und 4. Momenten ($\sqrt{b_1}$ und b_2) einer Normalverteilung (H_0) geben d'Agostino und Mitarbeiter [dBd90] ein Verfahren zur Berechnung spezieller Teststatistiken an. Aus diesen resultiert insbesondere ein gemeinsamer

Tabelle 7.4. Ausgewählte untere und obere Quantile der standardisierten 3. und 4. Momente ($\sqrt{b_1}$ und b_2) für den Test einer Abweichung vom Modell der Normalverteilung (aus Pearson, E.S. and H.O. Hartley (Eds.): *Biometrika Tables for Statisticians*. Vol I 3rd ed., Cambridge Univ. Press 1970, pp. 207-208, Table 34 B and C; und aus d'Agostino, R.B. and G.L. Tietjen (a): *Approaches to the null distribution of $\sqrt{b_1}$* . *Biometrika* **60** (1973), 169-173, Table 2. (b) *Simulation probability points of b_2 for small samples*. *Biometrika* **58** (1971), 669-672, p. 670, Table 1; und aus F. Gebhardt: *Verteilung und Signifikanzschranken des 3. und 4. Stichprobenmomentes bei normalverteilten Variablen*. *Biom. Z.* **8** (1966), 219-241, S. 235, Tabelle 4, S. 238-39, Tabelle 6)

n	Schiefe [$\sqrt{b_1}$]			Steilheit [b_2]					
	obere Quantile			untere Quantile			obere Quantile		
	10%	5%	1%	1%	5%	10%	10%	5%	1%
7	0,787	1,008	1,432	1,25	1,41	1,53	3,20	3,55	4,23
10	0,722	0,950	1,397	1,39	1,56	1,68	3,53	3,95	5,00
15	0,648	0,862	1,275	1,55	1,72	1,84	3,62	4,13	5,30
20	0,593	0,777	1,152	1,65	1,82	1,95	3,68	4,17	5,36
25	0,543	0,714	1,073	1,72	1,91	2,03	3,68	4,16	5,30
30	0,510	0,664	0,985	1,79	1,98	2,10	3,68	4,11	5,21
35	0,474	0,624	0,932	1,84	2,03	2,14	3,68	4,10	5,13
40	0,45	0,587	0,870	1,89	2,07	2,19	3,67	4,06	5,04
45	0,43	0,558	0,825	1,93	2,11	2,22	3,65	4,00	4,94
50	0,41	0,534	0,787	1,95	2,15	2,25	3,62	3,99	4,88
70	0,35	0,459	0,673	2,08	2,25	2,35	3,58	3,88	4,61
75	0,34			2,08	2,27			3,87	4,59
100	0,30	0,389	0,567	2,18	2,35	2,44	3,52	3,77	4,39
125		0,350	0,508	2,24	2,40	2,50	3,48	3,71	4,24
150	0,249	0,321	0,464	2,29	2,45	2,54	3,45	3,65	4,13
175		0,298	0,430	2,33	2,48	2,57	3,42	3,61	4,05
200	0,217	0,280	0,403	2,37	2,51	2,59	3,40	3,57	3,98
250		0,251	0,360	2,42	2,55	2,63	3,36	3,52	3,87
300	0,178	0,230	0,329	2,46	2,59	2,66	3,34	3,47	3,79
400		0,200	0,285	2,52	2,64	2,70	3,30	3,41	3,67
500	0,139	0,179	0,255	2,57	2,67	2,73	3,27	3,37	3,60
700		0,151	0,215	2,62	2,72	2,77	3,23	3,31	3,50
1000	0,099	0,127	0,180	2,68	2,76	2,81	3,19	3,26	3,41
2000	0,070	0,090	0,127	2,77	2,83	2,86	3,14	3,18	3,28

Test (Omnibus-Test; lat. omnibus = für alle), der beide Aspekte gemeinsam, d.h. Abweichungen sowohl hinsichtlich der Schiefe als auch der Steilheit erfasst.

7.2.3 Das Wahrscheinlichkeitsnetz, QQ-Plot

Mit Hilfe des Wahrscheinlichkeitsnetzes kann man sich einen ersten Überblick verschaffen, ob eine Stichprobenverteilung angenähert normalverteilt ist. Außerdem erhält man Mittelwert und Standardabweichung der Verteilung. Das Wahrscheinlichkeitsnetz, eine besondere Art von Zeichenpapier, ist so eingerichtet, dass sich beim Einzeichnen der in Prozent ausgedrückten, jeweils fortlaufend addierten Häufigkeiten einer Normalverteilung eine *Gerade* ergibt. Die Ordinate des Netzes ist nach der Verteilungsfunktion der $N(0; 1)$ -Verteilung geteilt, sie enthält die **Summenhäufigkeitsprozente**. Die Abszisse kann linear (in Millimetern) oder logarithmisch eingeteilt sein (vgl. Abb. 7.7). Die Ordinatenwerte 0% und 100% sind im Wahrscheinlichkeitsnetz nicht enthalten. Prozentuale Häufigkeiten mit diesen Werten bleiben daher bei der graphischen Darstellung unberücksichtigt.

Aus $z = (x - \mu)/\sigma$ folgt die Gerade $z = -(\mu/\sigma) + (1/\sigma)x$ mit den Punkten $F(\mu) = 0,5$ und $F(\mu + \sigma) = 0,8413$. Die Ordinatenskala enthält die entsprechenden Prozentwerte (50%; 84%) der Verteilungsfunktion der Standardnormalverteilung (z. B. liegen 15,87% der Verteilung unterhalb von $z = -1$):

y	0%	10%	15,87%	...	50%	...	84,13%	90%	100%
z	$-\infty$	-1,28	-1	...	0	...	+1	+1,28	$+\infty$

$$N(\mu; \sigma)$$

$$F(\mu + \sigma) = P(X \leq \mu + \sigma) \approx 0,84$$

$$F(\mu - \sigma) = P(X \leq \mu - \sigma) \approx 0,16$$

$$F(\mu + \sigma) - F(\mu - \sigma) = P(\mu - \sigma \leq X \leq \mu + \sigma) \approx 0,68$$

Abb. 7.7. Wahrscheinlichkeitsnetz

Man berechnet zu der empirischen Häufigkeitsverteilung die Summenverteilung in Prozent und zeichnet diese Werte in das Netz ein. Hierbei ist zu beachten, dass auf der Abszisse Klassengrenzen abzutragen sind. Beim Vorliegen einer Normalverteilung liegen die Punkte, von Zufallsschwankungen abgesehen, insbesondere im mittleren Bereich zwischen 10% und 90% auf einer Geraden, die nach Augenmaß eingezeichnet wird.

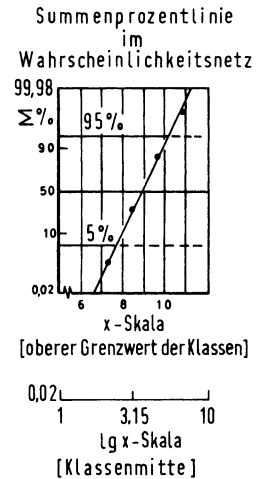
Die Schnittpunkte der Geraden mit den Ordinaten 50% und 84 (den 50%- und 84%-Linien) haben als Abszisse Näherungswerte für $\bar{x}$ und $\bar{x} + s$.

Eine Normalverteilung mit kleiner Standardabweichung hat eine steile Gerade; eine Normalverteilung mit gleicher Standardabweichung aber anderem Erwartungswert erscheint im Wahrscheinlichkeitsnetz als parallel verschobene Gerade. Die **Summenlinie der Normalverteilung**, auch **Hazensche Gerade** genannt, erhält man auf umgekehrte Weise durch die folgenden charakteristischen Werte: Für $x = \mu$ gilt $y = 50\%$; für $x = \mu + \sigma$ gilt $y \approx 84\%$; für $x = \mu - \sigma$ gilt $y \approx 16\%$. Der untere Teil von Abb. 7.7 (rechts) soll darauf hinweisen, dass auf diese Weise prinzipiell auch die Überprüfung hinsichtlich des Modells einer Lognormalverteilung möglich.

Das Verfahren des Wahrscheinlichkeitsnetzes ist hier eher aus historischer Sicht dargestellt. Ein moderner (methodisch vergleichbarer) Ansatz wird in Statistikprogrammsystemen in Form der sogenannten **QQ-Plots (Quantile-Quantile Plots)** umgesetzt. Allgemein bieten QQ-Plots eine graphische Möglichkeit zu entscheiden, ob zwei Messwertreihen aus Grundgesamtheiten mit der gleichen Verteilung stammen. Dazu werden einfach die Quantile der ersten Messwertreihe gegen die Quantile der zweiten Reihe im Koordinatensystem aufgezeichnet. Zusätzlich wird dann die Winkelhalbierende (45° -Line) eingezeichnet. Für den Fall, dass beide Messwertreihen aus Grundgesamtheiten mit gleicher Verteilung stammen, sollten die Punkte angenähert entlang dieser Referenzlinie liegen. Je ausgeprägter die Abweichung von der Referenzlinie ist, desto stärker ist auch die Evidenz für den Schluss, dass die beiden Messwertreihen nicht die gleiche Verteilung aufweisen.

Ein wesentlicher Vorteil dieser Technik liegt darin, dass (a) die Anzahl der Beobachtungen in den zu vergleichenden Reihen nicht gleich sein muss und dass (b) verschiedene Aspekte der Verteilung, insbesondere Verschiebungen in der Lage und der Steilheit, Unterschiede in der Symmetrie und das Vorliegen von Ausreißern oder Extremwerten, in einem Bild beurteilt werden können.

Wird für die Überprüfung eines speziellen „Verteilungsmodells“ eine der Messwertreihen durch die Quantile dieser theoretischen Verteilung ersetzt, dann erhält man ein **Wahrscheinlichkeits-Plot** (probability plot), welches der Idee des Wahrscheinlichkeitsnetzes entspricht. Für die Anpas-



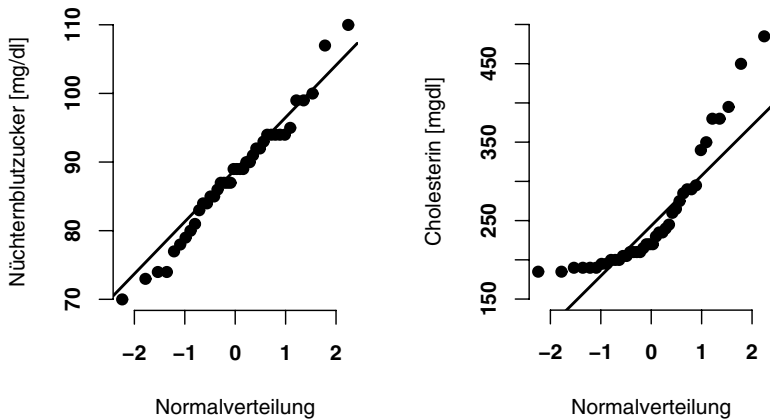


Abb. 7.8. QQ-Plot für je $n = 40$ Werte zum Blutzucker und zum Cholesterin

sung an das Modell einer Normalverteilung kann (7.24) als Approximation für die Wahrscheinlichkeitswerte p_i zu den nach der Größe geordneten Beobachtungen $x_{(i)}$ verwendet werden (Blom [Blo58]).

$$(x_{(i)}, p_i) \quad \text{mit} \quad p_i = \frac{i - 0,375}{n + 0,25} \quad (7.24)$$

In **R** werden diese Darstellungen durch die Funktionen **qqplot()**, **qqnorm()** und **qqline()** erzeugt. Abbildung 7.8 zeigt ein Beispiel mit je $n = 40$ Blutzuckerwerten, die als normalverteilt angenommen werden können sowie Cholesterinwerten, die sicher nicht normalverteilt sind (Tabelle 7.5).

Tabelle 7.5. Beispieldaten zum Nüchternblutzucker und zum Cholesterin ($n = 40$)

Blutzucker [mg/dl]	90	74	94	79	100	87	87	84	78	94
	73	99	85	83	70	84	91	99	85	89
	80	89	81	95	89	94	77	87	89	86
	94	110	92	92	93	94	87	90	107	74
Cholesterin [mg/dl]	195	205	245	190	260	190	340	195	285	380
	220	240	235	215	190	275	205	290	200	210
	220	265	235	200	350	220	450	230	185	295
	380	200	485	210	185	210	395	290	190	210

Viele empirische Verteilungen sind **inhomogene Mischverteilungen**. Unsymmetrische Mischverteilungen entstehen durch Überlagerung mehrerer Verteilungen mit gegenüber der Normalverteilung besonders **stark besetzten Verteilungsenden**. Aus der Tatsache, dass eine Stichprobenverteilung einen homogenen Eindruck macht und z.B. angenähert normalverteilt ist, darf nicht auf das Vorliegen einer einheitlichen Verteilung der Merkmale geschlossen werden. Nicht selten erweist sich eine gefundene Normalverteilung als zusammengesetzt. Grundsätzlich lässt sich die Homogenität eines Untersuchungsmaterials nicht beweisen! **Nur Inhomogenitäten lassen sich feststellen!** Inhomogenität bedeutet nicht Unbrauchbarkeit des Materials, sondern erfordert Berücksichtigung der Inhomogenität in der Auswertung, meist durch **Untergruppenbildung** (Stratifizierung)!

Zu bedenken ist, dass z. B. bei mehreren Untergruppen, die eine einheitliche Struktur aufweisen (H_0), bei 10 Prüfungen auf dem 5%-Niveau mit einer Wahrscheinlichkeit von 0,4 mindestens ein statistisch signifikanter Effekt fälschlich zu erwarten ist (s. multiples Testen). Wenn eine Gruppe von n Individuen oder Objekten rein zufällig in k (mit $k \leq 15$) gleichgroße Untergruppen zerlegt wird, weist die Differenz zwischen dem größten und dem kleinsten Untergruppenmittel einen Erwartungswert auf, der etwa k mal so groß ist wie der Standardfehler des Mittelwertes der n Beobachtungen. Daher ist bei „statistisch signifikanten“ Untergruppeneffekten Vorsicht angebracht, um nicht dem Risiko falsch positiver Resultate zu erliegen.

7.2.4 Der Chiquadrat-Anpassungstest

Gegeben sei eine Stichprobe aus einer Grundgesamtheit mit unbekannter Verteilungsfunktion $F(x)$ und eine ganz bestimmte theoretische Verteilungsfunktion $F_0(x)$. Ein Anpassungstest prüft die Nullhypothese (H_0): $F(x) = F_0(x)$ gegen die Alternativhypothese: $F(x) \neq F_0(x)$. Wird H_0 nicht abgelehnt, so sind – allein aufgrund des Tests – Folgerungen derart: beim Zustandekommen der empirischen Verteilung sind die gleichen Ursachen wirksam, die der empirischen Verteilung zugrunde liegen, nur unter Vorbehalt zu ziehen.

Die Prüfgröße (7.25), knapp als $\hat{\chi}^2$ bezeichnet,

$$\sum_{i=1}^k \frac{(B_i - E_i)^2}{E_i} \quad \text{bzw.} \quad \sum_{i=1}^k \frac{(n_i - np_i)^2}{np_i} = \frac{1}{n} \sum_{i=1}^k \frac{n_i^2}{p_i} - n \quad (7.25)$$

ist unter H_0 asymptotisch (für $n \rightarrow \infty$) χ^2 -verteilt mit ν Freiheitsgraden; H_0 wird abgelehnt, sobald für nicht zu kleines n (vgl. weiter unten) die Prüfgröße (7.25) d. h. $\hat{\chi}^2 > \chi_{\nu; 1-\alpha}^2$ mit $\nu = k - 1$ (Tab. 5.11).

Nun zur Erläuterung von (7.25):

- k = Klassenzahl der Stichprobe des Umfangs n ;
- $B_i = n_i$ = Beobachtete Häufigkeit (Besetzungszahl) der Klasse i , d. h. $n = \sum_{i=1}^k n_i$;
- $E_i = np_i$ = (unter H_0) Erwartete (angepasste) Häufigkeit;

Für eine diskrete Verteilung und für jedes i liege unter H_0 eine bestimmte gegebene oder hypothetische Wahrscheinlichkeit p_i vor $\left(\sum_{i=1}^k p_i = 1 \right)$, dann lassen sich die n_i mit den erwarteten np_i

vergleichen. Werden anhand der Zufallsstichprobe (die p_i als $\hat{p}_i$ bzw.) insgesamt a unbekannte Parameter geschätzt, dann verringert sich ν auf $\nu = k - 1 - a$; bei der Anpassung an eine Binomialverteilung oder an eine Poisson-Verteilung ist $a = 1$, bei der Anpassung an eine Normalverteilung ist $a = 3, 2$ oder 1 [vgl. weiter unten].

Bei **Anpassungstests** dieser Art sollten die Stichproben als Ganzes nicht zu klein und die der Nullhypothese entsprechenden erwarteten Häufigkeiten E nicht unter 1 liegen ($E > 1$). Sind sie kleiner, so werden sie durch Zusammenlegen von 2, 3, ... benachbarten Klassen auf das geforderte Niveau erhöht. Dies ist aber nur dann nötig, wenn die Anzahl der Klassen klein ist. Für den Fall $\nu \gtrsim 8$ und einem nicht zu kleinen Stichprobenumfang $n \gtrsim 40$ dürfen die Erwartungshäufigkeiten in vereinzelter Klassen bis unter 1 absinken. Bei großem n und $\alpha = 0,05$ wähle man 16 Klassen.

Bei der Berechnung von $\hat{\chi}^2$ sind die *Vorzeichen der Differenzen* $B - E$ zu beachten: + und – sollten miteinander abwechseln und keine systematischen Zyklen zeigen. Wir werden hierauf im Beispiel zur Poisson-Verteilung noch einmal zurückkommen.

Erwartungshäufigkeiten gleich Eins:

Wenn N unabhängige Objekte auf N Felder verteilt werden – n_i ($i = 1, 2, \dots, N$) Objekte auf Feld i [$0 \leq n_i \leq N$; $\sum n_i = N$] – jedes Feld weist die Erwartungshäufigkeit 1 auf, gilt für $N > 25$ die nach χ^2_{N-1} verteilte Approximation zur Prüfung der Gleichverteilung (H_0) gegenüber deutlichen Abweichungen (H_A):

$$\begin{aligned}\hat{\chi}^2 &= \sum_{\text{Felder}} (B_i - E_i)^2 / E_i \\ &= \sum_{i=1}^N (n_i - 1)^2 / 1 = \sum_{i=1}^N n_i^2 - N = \sum_{i=1}^N n_i(n_i - 1)\end{aligned}\quad (7.26)$$



Beispiel: Angenommen 30 Schüler werden nach dem beliebtesten gefragt, jeder darf nur 1 Person nennen, wobei Selbstnennung erlaubt ist. Wie übereinstimmend urteilt die Klasse? Genannt werden: eine Person 5mal, eine Person 4mal, zwei Personen jeweils 3mal, 4 Personen jeweils 2mal und 7 Personen jeweils 1mal:

$$\hat{\chi}^2 = \sum_{i=1}^N n_i^2 - N = (1 \cdot 5^2 + 1 \cdot 4^2 + 2 \cdot 3^2 + 4 \cdot 2^2 + 7 \cdot 1^2) - 30 = 82 - 30 = 52$$

Mit dem Quantil $\chi^2_{29;0,99} = 49,59$, d.h. $P(\chi^2_{29} \geq 52) < 0,01$ folgt, dass auf dem 1%-Niveau keine Gleichverteilung auftritt sondern deutliche Bevorzugungen (Klumpungen) [vgl. auch: $\hat{\chi}^2 = (30 \cdot 1^2) - 30 = 0$ sowie $\hat{\chi}^2 = (1 \cdot 30^2) - 30 = 870$]. Näheres sowie kritische Schranken sind Zahn und Roberts [ZR71] zu entnehmen.

**Vergleich von beobachteten Häufigkeiten mit Spaltungsziffern:**

Bei einem als Vorversuch geplanten Spaltungsversuch werden 3 Phänotypen im Verhältnis 1 : 2 : 1 erwartet; gefunden werden die Häufigkeiten 14 : 50 : 16. Entspricht das gefundene Verhältnis der 1 : 2 : 1-Spaltung? Die Rechnung (mit $FG = k - 1 = 3 - 1 = 2$) zeigt ($\hat{\chi}^2 = 5,10 < 5,99 = \chi^2_{2;0,95}$ [Tab. 5.11]), dass die H_0 : Verträglichkeit der beobachteten mit den theoretischen Häufigkeiten, auf dem 5%-Niveau nicht abzulehnen ist.

B	E	$B - E$	$(B - E)^2$	$\frac{(B - E)^2}{E}$
14	20	-6	36	1,80
50	40	10	100	2,50
16	20	-4	16	0,80
80	80	$\hat{\chi}^2 = \sum \frac{(B - E)^2}{E} = 5,10$		

Vergleich einer empirischen Verteilung mit einer Gleichverteilung:

Zur Prüfung eines Würfels werden 60 Würfe durchgeführt. Die beobachteten Häufigkeiten (B) für die 6 Augenzahlen sind:

Augenzahl	1	2	3	4	5	6
Häufigkeit	7	16	8	17	3	9

Die Nullhypothese – es liegt ein „guter“ Würfel vor – sagt für jede Augenzahl eine theoretische oder [unter H_0] erwartete Häufigkeit (E) von 10 voraus, eine sogenannte Gleichverteilung. Wir testen auf dem 5%-Niveau und erhalten nach (7.25), vorausgesetzt $E \geq 0,3$:

$$\hat{\chi}^2 = \sum \frac{(B-E)^2}{E} = \frac{(7-10)^2}{10} + \frac{(16-10)^2}{10} + \dots + \frac{(9-10)^2}{10}$$



$\hat{\chi}^2 = 14,8$, ein Wert, der größer ist als der für $k-1 = 6-1 = 5$ Freiheitsgrade auf dem 5%-Niveau tabellierte χ^2 -Wert (11,07): H_0 wird abgelehnt.

Oder: mit den auf k Klassen verteilten Besetzungszahlen n_i und $n = \sum_{i=1}^k n_i$ nach

$$\hat{\chi}^2 = \left[(k/n) \sum_{i=1}^k n_i^2 \right] - n = [(6/60)(7^2 + 16^2 + \dots + 9^2)] - 60 = 14,8.$$

```
> obs <- c(7, 16, 8, 17, 3, 9); summe <- sum(obs)
> exp <- rep(summe/6, 6)
>
> stat <- sum((obs-exp)^2/exp); stat; qchisq(0.95, 5)
[1] 14.8 [1] 11.07050
```

Vergleich einer empirischen Verteilung mit einer Normalverteilung:

Erfahrungsgemäß sind Stichprobenverteilungen höchstens im mittleren Bereich angenähert normalverteilt. Das folgende, einfache und ziemlich grobe, Verfahren hat daher für den Praktiker einige Bedeutung, wenn man vom Wahrscheinlichkeitsnetz bzw. QQ-Plot absieht. Ausgehend von der empirischen Verteilung, die graphisch durch das Histogramm gekennzeichnet wird, kann unter der Annahme einer Normalverteilung eine „erwartete Häufigkeitsverteilung“ bestimmt werden.

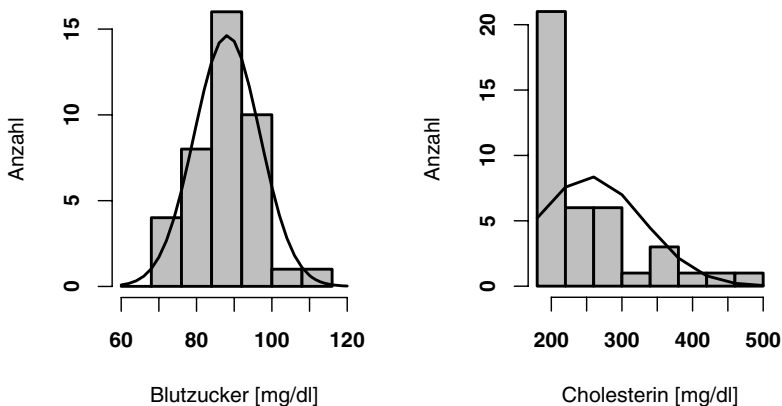


Abb. 7.9. Histogramm (Normalverteilung) für je $n = 40$ Werte zum Blutzucker und zum Cholesterin

Abbildung 7.9 zeigt die Übereinstimmung bzw. die Abweichung zwischen Beobachtung (empirischer Verteilung) und Verteilungsmodell am Beispiel von je $n = 40$ Messungen zum Nüchternblutzucker und zum Cholesterin (Daten in Tabelle 7.5, vgl. auch QQ-Plot in Abbildung 7.8). Zur Erklärung geben wir ein sehr einfaches Zahlenbeispiel. Spalte 1 der Tabelle 7.6 enthält die Klassenmitten x , die Klassenbreite b beträgt $b = 1$. Die beobachteten Häufigkeiten sind in Spalte 2 notiert. Die 3., 4. und 5. Spalte dienen zur Berechnung von $\bar{x}$ und s . In den Spalten 6, 7 und 8 wird der Weg über die Standardnormalvariable z zur Ordinate von z gezeigt. Die Multiplikation mit der Konstanten K in Spalte 9 dient zur Anpassung der Gesamtzahl der Erwartungshäufigkeiten. Klassen mit $E < 1$ sind mit den Nachbarklassen zusammenzufassen. Dann liegen insgesamt k Klassen vor (vgl. Tab. 7.6: $k = 5$). Exakte Ansätze sind Greenwood und Nikulin [GN96] zu entnehmen.

Für die Überprüfung der Normalitätshypothese kann hier der χ^2 -Anpassungstest verwendet werden. Geschätzt werden $\bar{x}$ und s aus den klassierten Daten [hierfür werden 3 FG benötigt (würden $\bar{x}$ und s direkt aus den Originaldaten berechnet, so benötigte man 2 FG, ist μ oder σ bekannt



Tabelle 7.6. Empirische Verteilung (vgl. die Besetzungszahlen in Spalte (2)) und Normalverteilung

x	B	x^2	Bx	Bx^2	$x - \bar{x}$	$\left \frac{x - \bar{x}}{s} \right = \hat{z}$	Ordinate von $\hat{z} = f(\hat{z})$	$f(\hat{z}) \cdot K$	E	$B - E$	$(B - E)^2$	$(B - E)^2 / E$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
1	1	1	1	1	-2,6	2,31	0,0277	0,983	} 6,15	-1,15	1,322	0,215
2	4	4	8	16	-1,6	1,42	0,1456	5,168		3,70	13,690	1,113
3	16	9	48	144	-0,6	0,53	0,3467	12,305		13,32	11,022	0,827
4	10	16	40	160	0,4	0,35	0,3752	13,317	6,56	0,44	0,194	0,030
5	7	25	35	175	1,4	1,24	0,1849	6,562	1,47	0,53	0,281	0,191
6	2	36	12	72	2,4	2,13	0,0413	1,466				
Σ	40		144	568					39,80	+0,2	$\hat{\chi}^2 = 2,376$	
									= 40	= 0	$\nu = 5 - 4 = 1$	

$$\bar{x} = \frac{\sum Bx}{n} = \frac{144}{40} = 3,60 \quad K = \frac{nb}{s} = \frac{40 \cdot 1}{1,127} = 35,492$$

$$s = \sqrt{\frac{\sum Bx^2 - (\sum Bx)^2/n}{n-1}} = \sqrt{\frac{568 - 144^2/40}{39}} = 1,127$$

Dieser Ansatz ist eine grobe Approximation. Exakte Methoden behandeln Greenwood und Nikulin (1996).

$\hat{\chi}^2 = 2,376 < 2,706 = \chi^2_{0,10}$

und wird der unbekannte Parameter aus den Originaldaten geschätzt, so benötigte man nur noch 1 FG)], so dass insgesamt $\nu = k - 1 - a$, hier $k - 1 - 3 = 5 - 4 = 1$ Freiheitsgrad zur Verfügung steht. Mit $2,376 < 2,706 = \chi^2_{1;0,10}$ ist gegen die Normalitätshypothese nichts einzuwenden.

Im praktischen Fall einer **groben Prüfung** auf Abweichungen vom Typ der Normalverteilung anhand des χ^2 -Anpassungstests sollte gelten:

$$1) n \geq 60 \quad 2) k \geq 7 \quad 3) \alpha = 0,10 \quad \text{bzw. } 0,05 \quad [\text{oder } 0,01]$$

Faustregel: Für $0,9 < (\tilde{x}/\bar{x}) < 1,1$ und $3s < \bar{x}$ wird eine Stichprobenverteilung als angenähert normalverteilt aufgefasst. Mit den Daten der Tabelle 7.6 ergibt sich: $\tilde{x} = 2,5 + 1\{([40/2] - 5)/16\} = 3,4375$ oder $3,44$ und $\tilde{x}/\bar{x} = 3,44/3,60 = 0,956$ oder $0,96$; $0,9 < 0,96 < 1,1$ und $3s = 3 \cdot 1,127 = 3,381 < 3,60 = \bar{x}$.



Der χ^2 -Anpassungstest ist für die Überprüfung der Normalitätshypothese allgemein nicht zu empfehlen, da er hinsichtlich der Power anderen Testverfahren (vgl. in den folgenden Abschnitten) unterlegen ist. In **R** steht im Paket **nortest** (Gross [Gro05]) eine spezielle Funktion **pearson.test()** zur Verfügung, die wir am Beispiel der Blutzucker- und Cholesterinwerte (Tabelle 7.5, Abbildung 7.8 und 7.9) vorstellen.

```
> library(nortest)
> pearson.test(nblz, n.classes=8, adjust=TRUE)
Pearson chi-square normality test
data:  nblz      P = 7.6,    p-value = 0.1797
> pearson.test(chol, n.classes=8, adjust=TRUE)
Pearson chi-square normality test
data:  chol      P = 21.6,   p-value = 0.0006237
```

Die Angabe einer Anzahl von Klassen (k) in der Funktion **pearson.test()** beeinflusst die Zahl der Freiheitsgrade ($k - 1$) der χ^2 -Verteilung und damit den berechneten P-Wert. Zusätzlich kann

durch den Wert „adjust=TRUE“ die Zahl der Freiheitsgrade korrigiert werden ($k - 3$), da Erwartungswert und Varianz aus den Daten abgeleitet werden müssen. Die (Null-)Hypothese einer den Beobachtungen zugrunde liegenden Normalverteilung kann für die Blutzuckerwerte auf dem 5%-Signifikanzniveau nicht abgelehnt werden. Dagegen kann angenommen werden, dass die Cholesterinwerte nicht normalverteilt sind ($P < 0,05$).

Vergleich einer empirischen Verteilung mit der Poisson-Verteilung:

Wir nehmen das Pferdehufschlagbeispiel (vgl Poisson-Verteilung, Tabelle 5.6), fassen die schwach besetzten drei Endklassen zusammen und erhalten die folgende Tabelle:



B	E	$B - E$	$(B - E)^2$	$(B - E)^2/E$
109	108,7	0,3	0,09	0,001
65	66,3	-1,3	1,69	0,025
22	20,2	1,8	3,24	0,160
4	4,8	-0,8	0,64	0,133
200	200,0	$\hat{\chi}^2 = 0,319$		

Es liegen $k = 4$ Klassen vor, geschätzt wurde $a = 1$ Parameter (λ aus $\hat{\lambda} = \bar{x}$). Damit stehen insgesamt $\nu = k - 1 - a = 4 - 1 - 1 = 2$ zur Verfügung. Der ermittelte $\hat{\chi}^2$ -Wert ist so niedrig ($\chi^2_{2;0,95} = 5,991$), dass die Übereinstimmung als gut zu bezeichnen ist.

7.2.5 Kolmogoroff-Smirnoff-Anpassungstest

Der Kolmogoroff-Smirnoff-Test wird verwendet, um zu überprüfen, ob die beobachteten Daten einer speziellen Verteilung entsprechen. Dazu bezieht sich der Test auf die **empirische Verteilungsfunktion**. Diese ist für die Beobachtungen $x_1, x_2, \dots, x_n$ gegeben durch:

$$E_n = n(i)/n \quad (7.27)$$

Dabei ist $n(i)$ die Anzahl der Beobachtungen, die kleiner als x_i sind. Die empirische Verteilungsfunktion ist daher eine Stufenfunktion, die mit einer Stufenhöhe von $1/n$ an den beobachteten Werten ansteigt.

Der K-S-Test basiert nun auf dem maximalen Abstand zwischen der empirischen Verteilungsfunktion und einer ausgewählten speziellen Verteilungsfunktion, z.B. auch einer Normalverteilung. Ein wesentlicher Vorteil der K-S-Teststatistik ist, dass sie exakt bestimmt werden kann und dass sie nicht von der beobachteten Verteilung abhängt. Sie unterliegt jedoch den folgenden Bedingungen:

- Der K-S-Test kann nur für kontinuierliche (stetige) Verteilungen eingesetzt werden.
- Der K-S-Test entscheidet sensibler in der Mitte der Verteilung als an deren Rändern.
- Die Modellverteilung muss vollständig spezifiziert sein, d.h. für den Fall, dass Lage (Erwartungswert) und Streuung (Varianz) aus den Beobachtungen einer Stichprobe geschätzt werden müssen, sind die kritischen Werte des K-S-Tests nicht exakt.

Insbesondere die letzten beiden Bedingungen führen dazu, dass der Anderson-Darling-Test häufig für die Prüfung einer Verteilung bevorzugt wird. Allerdings ist dieser nur für einige spezielle Verteilungsmodelle verfügbar. Der K-S-Test prüft die folgende Hypothesenstellung:

$$\begin{aligned} H_0 &: \text{Die Daten folgen einer speziellen Verteilung} \\ H_A &: \text{Die Daten folgen nicht einer speziellen Verteilung} \end{aligned}$$

Die Teststatistik lautet:

$$\hat{D} = \max_{1 \leq i \leq N} \left| F(y_i) - \frac{i}{N} \right| \quad (7.28)$$

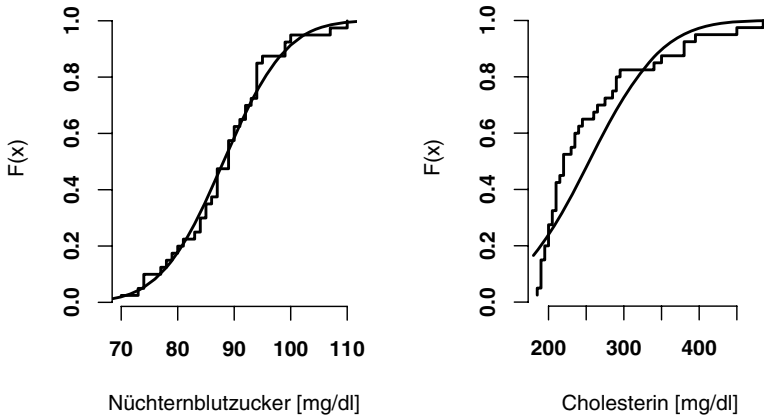


Abb. 7.10. Empirische Verteilung und Normalverteilung am Beispiel von Blutzucker und Cholesterin

Dabei ist F eine theoretische Verteilungsfunktion (Modell), die mit der empirischen Verteilung verglichen werden soll. F ist stetig und muss hinsichtlich der Verteilungsparameter (Lage, Schiefe, Steilheit, Varianz) vollständig definiert sein. Für den Fall, dass die Parameter aus der Stichprobe geschätzt werden, ist der K-S-Test sehr konservativ. Besser ist dann die nach Lilliefors benannte Modifikation, die weiter unten dargestellt wird.

Abbildung 7.10 zeigt die im K-S-Test verwendeten Differenzen zwischen empirischer und der unter dem Modell der Normalverteilung angenommenen Verteilung am Beispiel der in den vorangehenden Abschnitten bereits dargestellten Werte zum Nüchternblutzucker und zum Cholesterin (Tabelle 7.5, Abbildung 7.8 und 7.9).

Tabelle 7.7. Kritische Werte für den Kolmogoroff-Smirnoff-Test ($n > 35$)

Schranken für D	Signifikanzniveau α
$1,037/\sqrt{n}$	0,20
$1,138/\sqrt{n}$	0,15
$1,224/\sqrt{n}$	0,10
$1,358/\sqrt{n}$	0,05
$1,517/\sqrt{n}$	0,02
$1,628/\sqrt{n}$	0,01
$1,731/\sqrt{n}$	0,005
$1,949/\sqrt{n}$	0,001

Die Nullhypothese wird verworfen, wenn der Wert der Teststatistik größer ist als der kritische Wert, der entsprechenden Tabellen entnommen werden kann. Für Stichprobenumfänge $n > 35$ kann der Wert der Teststatistik anhand der kritischen Werte aus Tabelle 7.7 beurteilt werden.

Miller [Mil56] gibt für $n = 1$ bis 100 und $\alpha = 0,20, 0,10, 0,05, 0,02$ und $0,01$ exakte kritische Werte: Die besonders wichtigen 10%- und 5%-Grenzen für kleine und mittlere Stichprobenumfänge haben wir gerundet notiert (Tabelle 7.8). Ein beobachteter $\hat{D}$ -Wert, der den Tabellenwert erreicht oder überschreitet, ist auf dem entsprechenden Niveau statistisch signifikant. Für andere Werte α erhält man den Zähler der Schranke als $\sqrt{-0,5 \cdot \ln(\alpha/2)}$ (z. B. $\alpha = 0,10$; $\ln(0,10/2) = \ln 0,05 = -2,996$, d. h. $\sqrt{(-0,5)(-2,996)} = 1,224$).

In den Statistikprogrammen werden die kritischen Werte (p-Werte) teilweise direkt exakt berechnet oder sie werden durch Quantile der Standardnormalverteilung angenähert ersetzt. In **R** kann die Funktion **ks.test()** verwendet werden. Diese ist standardmäßig für den Vergleich zweier Verteilungen (vgl. auch Zweistichproben-test) vorgesehen, kann aber mit entsprechenden Spezifikationen auch zur Überprüfung eines speziellen Verteilungsmodells verwendet werden. Die Anwendung für die Beispieldaten zum Nüchternblutzucker und zum Cholesterin zeigt das folgende Beispiel.

```
> ks.test(nblz, "pnorm", mean(nblz), sd(nblz))
One-sample Kolmogorov-Smirnov test
data:  nblz    D = 0.1006, p-value = 0.8127    ...
...
> ks.test(chol, "pnorm", mean(chol), sd(chol))
One-sample Kolmogorov-Smirnov test
data:  chol    D = 0.1997, p-value = 0.08232    ...
```

Tabelle 7.8. Kritische Werte D für den Kolmogoroff-Smirnoff-Anpassungstest (aus Miller, L. H.: Table of percentage points of Kolmogorov statistics. J. Amer. Statist. Assoc. **51** (1956) 111–121, p. 113–115, part of table 1)

n	$D_{0,10}$	$D_{0,05}$	n	$D_{0,10}$	$D_{0,05}$	n	$D_{0,10}$	$D_{0,05}$	n	$D_{0,10}$	$D_{0,05}$
3	0,636	0,708	23	0,247	0,275	13	0,325	0,361	33	0,208	0,231
4	0,565	0,624	24	0,242	0,269	14	0,314	0,349	34	0,205	0,227
5	0,509	0,563	25	0,238	0,264	15	0,304	0,338	35	0,202	0,224
6	0,468	0,519	26	0,233	0,259	16	0,295	0,327	36	0,199	0,221
7	0,436	0,483	27	0,229	0,254	17	0,286	0,318	37	0,196	0,218
8	0,410	0,454	28	0,225	0,250	18	0,278	0,309	38	0,194	0,215
9	0,387	0,430	29	0,221	0,246	19	0,271	0,301	39	0,191	0,213
10	0,369	0,409	30	0,218	0,242	20	0,265	0,294	40	0,189	0,210
11	0,352	0,391	31	0,214	0,238	21	0,259	0,287	50	0,170	0,188
12	0,338	0,375	32	0,211	0,234	22	0,253	0,281	100	0,121	0,134

Lilliefors-Modifikation des Kolmogoroff-Smirnoff-Tests

Müssen für die Anpassung an eine Normalverteilung Mittelwert und Varianz aus den Stichprobenwerten geschätzt werden, dann sind die auf Tabelle 7.7 basierenden Resultate sehr konservativ; exakte Schranken (vgl. Lilliefors [Lil67]) geben Dallal und Wilkinson ([DW86]) an; einige Werte sind in Tabelle 7.9 zusammengefasst.

Tabelle 7.9. Kritische Werte D für den nach Lilliefors modifizierten Kolmogoroff-Smirnoff-Test

n	5	8	10	12	15	17	20	25	30	40
10%	0,319	0,265	0,241	0,222	0,201	0,190	0,176	0,159	0,146	0,128
5%	0,343	0,288	0,262	0,242	0,219	0,207	0,192	0,173	0,159	0,139
1%	0,397	0,333	0,304	0,281	0,254	0,240	0,223	0,201	0,185	0,162

Für $n > 30$ gelten nach Mason und Bell [MB86] mit $d_n = \sqrt{n} - 0,01 + 0,83/\sqrt{n}$ die approximierten Schranken $0,741/d_n$ ($\alpha = 0,20$), $0,775/d_n$ ($\alpha = 0,15$), $0,819/d_n$ ($\alpha = 0,10$), $0,895/d_n$ ($\alpha = 0,05$) und $1,035/d_n$ ($\alpha = 0,01$). Beispiel: $n = 40$, $\alpha = 0,05$, d. h. $d_n = (\sqrt{40} - 0,01 + 0,83/\sqrt{40}) = 6,446$ und damit $0,895/6,446 = 0,139$.

Das Paket **nortest** (Gross [Gro05]) stellt in **R** eine entsprechende Funktion **lillie.test()** zur Verfügung, in der der P-Wert nach der Formel von Dallal-Wilkinson berechnet wird.

```
> library(nortest)
> lillie.test(nblz)
Lilliefors (Kolmogorov-Smirnov) normality test
data:  nblz    D = 0.1006, p-value = 0.3897
> lillie.test(chol)
```

Tabelle 7.10. Kritische Schranken D für den Kolmogoroff-Smirnoff-Test: Anpassung an eine Poisson-Verteilung mit aus der Zufallsstichprobe geschätztem Mittelwert $\bar{x}$ ($\bar{x} \leq 10$) für 5 Mittelwertsbereiche, 6 Stichprobenumfänge ($n \geq 6$) und drei Signifikanzniveaus: ist $\bar{x}$ keine ganze Zahl, so benutze den nächst größeren vorgegebenen Wert und die entsprechende Spalte

						α : 0,10 0,05 0,01
n	$\bar{x} \leq 1$	$1 < \bar{x} \leq 2$	$2 < \bar{x} \leq 3$	$3 < \bar{x} \leq 5$	$5 < \bar{x} \leq 10$	
6	0,202	0,214	0,226	0,237	0,254	
	0,234	0,242	0,254	0,265	0,281	
	0,290	0,300	0,310	0,324	0,334	
12	0,152	0,166	0,172	0,179	0,185	
	0,180	0,188	0,194	0,199	0,206	
	0,223	0,234	0,236	0,243	0,249	
20	0,120	0,132	0,140	0,144	0,149	
	0,141	0,151	0,156	0,160	0,165	
	0,176	0,185	0,188	0,195	0,197	
30	0,100	0,112	0,116	0,120	0,124	
	0,116	0,125	0,129	0,134	0,140	
	0,149	0,154	0,158	0,160	0,168	
40	0,087	0,097	0,102	0,106	0,110	
	0,101	0,108	0,113	0,118	0,122	
	0,130	0,135	0,137	0,143	0,146	
> 40	$0,55/\sqrt{n}$	$0,61/\sqrt{n}$	$0,65/\sqrt{n}$	$0,67/\sqrt{n}$	$0,70/\sqrt{n}$	
	$0,64/\sqrt{n}$	$0,69/\sqrt{n}$	$0,72/\sqrt{n}$	$0,75/\sqrt{n}$	$0,77/\sqrt{n}$	
	$0,82/\sqrt{n}$	$0,86/\sqrt{n}$	$0,87/\sqrt{n}$	$0,90/\sqrt{n}$	$0,93/\sqrt{n}$	

Lilliefors (Kolmogorov–Smirnov) normality test
data: chol D = 0.1997, p-value = 0.0003435



Beispiel 1 (Vergleich der Spalten (2) und (9) der Tabelle 7.6): Wir verwenden das Beispiel in Tabelle 7.6 und erhalten über $2,55/40 = 0,063 < 0,127 = 0,819/6,446$ [mit $d_{40} = (\sqrt{40} - 0,01 + 0,83/\sqrt{40}) = 6,446$] ebenfalls das Resultat: Die Nullhypothese lässt sich auf dem 10%-Niveau nicht ablehnen.

B	1	4	16	10	7	2
E	0,98	5,17	12,30	13,32	6,56	1,47
F_B	1	5	21	31	38	40
F_E	0,98	6,15	18,45	31,77	38,33	39,80
$ F_B - F_E $	0,02	1,15	2,55	0,77	0,33	0,20



Beispiel 2 (Idealer Würfel): Ein Würfel wird zur Kontrolle 120mal geworfen. Die Häufigkeiten für die 6 Augen sind: 18, 23, 15, 21, 25, 18. Entspricht das gefundene Verhältnis der Nullhypothese, nach der ein idealer Würfel vorliegt? Wir prüfen mit $\alpha = 0,01$ die aufsteigend geordneten Häufigkeiten: 15, 18, 18, 21, 23, 25 [d.h. $15 + 18 = 33$; $33 + 18 = 51$; ...].

F_E	20	40	60	80	100	120
F_B	15	33	51	72	95	120
$ F_B - F_E $	5	7	9	8	5	0

Da $9/120 = 0,075 < 0,1486 = 1,628/\sqrt{120} = D_{120;0,01}$ (Tabelle 7.7) ist, wird die Nullhypothese auf dem 1%-Niveau nicht abgelehnt.

7.2.5.1 Anpassung an eine Poisson-Verteilung

Die Nullhypothese, nach der die Daten einer Poisson-Verteilung entstammen, wird auf dem $100\alpha\%$ -Niveau abgelehnt, sobald das empirische $\hat{D}$ größer ist als ein kritischer Wert D . In Tabelle 7.10 sind einige Werte für D aufgeführt [CO79].

Beispiel (Verteilung von Hefezellen): Wir prüfen die Stichprobenverteilung der Tabelle 7.11 [die Beobachtungen B], ob sie einer Poisson-Verteilung mit $\bar{x} = (0 \cdot 75 + 1 \cdot 103 + 2 \cdot 121 + \dots + 9 \cdot 1)/400 = 720/400 = 1,8$ entstammen könnte ($\alpha = 0,01$). Wir berechnen die empirische Verteilungsfunktion $F_B(x)$, die angepasste ($\lambda = 1,8$) Verteilungsfunktion $F_A(x)$ und die maximale Differenz $\hat{D}$.

Tabelle 7.11. Verteilung von Hefezellen auf 400 Quadrate eines Hämazytometers mit $\bar{x} = 1,8$ Zellen pro Quadrat

Zellzahl pro Quadrat x	B	Empir. Vert.- Funktion $F_B(x)$	Angepaßte Vert.-F. $F_A(x)$	$ F_B(x) - F_A(x) $
(1)	(2)	(3)	(4)	(5)
0	75	0,1875	0,165	0,0225 = $\hat{D}$
1	103	0,4450	0,463	0,0180
2	121	0,7475	0,731	0,0165
3	54	0,8825	0,891	0,0085
4	30	0,9575	0,964	0,0065
5	13	0,9900	0,990	0,0000
6	2	0,9950	0,997	0,0020
7	1	0,9975	0,999	0,0015
8	0	0,9975	1,000	0,0025
9	1	1,0000	1,0000	0,0000

Hinweise zu Spalte:

(2) $75/400 = 0,1875$; $(75 + 103)/400 = 0,445$ usw.;

(3) $P(X = 0|\lambda = 1,8) = 1,8^0 e^{-1,8}/0! = 1/e^{1,8}/1 = 1/6,05/1 = 0,1653$;

$P(X \leq 1|\lambda = 1,8) = 0,1653 + P(X = 1|\lambda = 1,8) = 0,1653 + 1,8^1 e^{-1,8}/1! = 0,46284$ bzw.

$P(0 + 1) = \lambda P(0)/(0 + 1) = 1,8 \cdot 0,1653/1 = 0,29754$ und $0,16530 + 0,29754 = 0,46284$ usw.

Mit $\hat{D} = 0,0225 < 0,043 = 0,86/\sqrt{400} = D$ (vgl. Tabelle 7.10) lässt sich die Nullhypothese auf dem 1%-Niveau nicht ablehnen.

7.2.6 Shapiro-Wilk Test

Der Shapiro-Wilk Test $\hat{W}$ prüft die Hypothese, dass die Beobachtungen in einer Zufallsstichprobe $x_1, x_2, \dots, x_n$ einer normalverteilten Zufallsvariablen zugeordnet werden können. Dazu wird



die Verteilung des Quotienten aus zwei Schätzungen für σ^2 betrachtet (7.29): das Quadrat einer kleinsten Fehlerquadratschätzung für die Steigung einer Regressionsgeraden im QQ-Plot (s.d.) und die Stichprobenvarianz. Für den Fall, dass eine Normalverteilung vorliegt, sollten diese beiden Schätzungen nahe zusammen liegen und somit das Verhältnis 1 ergeben. Kleine Werte von $\hat{W}$ sind evident für Abweichungen von der Annahme einer Normalverteilung. Quantile für die W -Statistik bei vorliegender Normalverteilung sind durch Pearson und Hartley ([PH72], Vol. II, Table 16) über Monte-Carlo Simulationen hergeleitet worden. Der Shapiro-Wilk Test hat gegenüber den anderen Testverfahren mit die höchste Power. Die W -Statistik wird berechnet durch (7.29)

$$\hat{W} = \frac{\left(\sum_{i=1}^n a_i x_{(i)} \right)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (7.29)$$

Darin sind die $x_{(i)}$ die Beobachtungen aus der aufsteigend geordneten Stichprobe ($x_{(1)}$ der kleinste Wert) und die a_i sind konstante Werte, die aus den Maßzahlen der Ordnungsstatistik einer normalverteilten Zufallsvariablen abhängig vom Stichprobenumfang n erzeugt oder entsprechenden Tabellen entnommen werden können ([SW65] und [PH72], Vol. I, Table 15).

In **R** wird der Shapiro-Wilk Test durch die Funktion **shapiro.test()** berechnet ([Roy82b] und [Roy82a]).

```
> shapiro.test(nblz)
      Shapiro-Wilk normality test
data:  nblz      W = 0.9801, p-value = 0.6918
> shapiro.test(chol)
      Shapiro-Wilk normality test
data:  chol      W = 0.8063, p-value = 9.187e-06
```

Für die Beispieldaten aus Tabelle 7.5 zeigt der Shapiro-Wilk Test, dass die Werte für Cholesterin nicht normalverteilt sind ($P < 0,05$), während die Werte zum Nüchternblutzucker durchaus mit dem Modell einer Normalverteilung beschrieben werden können.

7.2.7 Anderson-Darling Test

Der Anderson-Darling Test (Stephens [Ste86]) prüft die Hypothese, dass die Beobachtungen einer Zufallsstichprobe einer Zufallsvariablen mit einem speziellen Verteilungsmodell zuzuordnen sind. Er ist eine Modifikation des Kolmogoroff-Smirnoff Tests (K-S), die Abweichungen vom Verteilungsmodell an den Rändern der Verteilung eher berücksichtigt als der K-S Test (Stephens [Ste74]).

Der K-S Test ist ein verteilungsfreies Verfahren in dem Sinn, dass die kritischen Werte der Teststatistik nicht von einem speziellen Verteilungsmodell abhängen (daher die Modifikation nach Lilliefors). Der Anderson-Darling Test verwendet für die Berechnung der kritischen Werte spezielle Verteilungsmodelle. Dies hat den Vorteil, dass der Test mehr Power hat als der K-S Test, allerdings müssen kritische Werte für jedes Verteilungsmodell getrennt hergeleitet werden (Prüfung auf $N(\mu, \sigma)$: vgl. auch Sachs [Sac90], S. 164/165).

Die Teststatistik zum Anderson-Darling Test A für eine spezielle Verteilungsfunktion F wird nach (7.30) berechnet.

$$A^2 = -N - S \quad \text{mit}$$

$$S = \sum_{i=1}^N \frac{(2i-1)}{N} [\log F(Y_i) + \log(1 - Y_{N+1-i})] n\sigma^2 \quad (7.30)$$

Der Anderson-Darling Test wird üblicherweise mit einem Statistikprogramm gerechnet, welches die entsprechenden kritischen Werte bzw. P-Werte bestimmt. In **R** kann dieser Test mit der Funktion **ad.test()** aus dem Paket **nortest** (Gross [Gro05]) gerechnet werden. Das folgende Beispiel zeigt die Analyse der Daten aus Tabelle 7.5.

```
> library(nortest)
> ad.test(nblz)
      Anderson-Darling normality test
data:  nblz      A = 0.3051, p-value = 0.5525
> ad.test(chol)
      Anderson-Darling normality test
data:  chol      A = 2.761, p-value = 4.39e-07
```

7.2.8 Ausreißerproblem

Extrem hohe oder niedrige Werte innerhalb einer Reihe üblicher mäßig unterschiedlicher Messwerte, von denen fraglich ist, ob sie unter den vorgegebenen Bedingungen möglich sind, dürfen unter gewissen Umständen vernachlässigt werden. Man bezeichnet sie als Ausreißer. Messfehler, Beurteilungsfehler, Rechenfehler oder ein pathologischer Fall im Untersuchungsmaterial von Gesunden können zu Extremwerten führen, die, da sie anderen Grundgesamtheiten als die der Stichprobe entstammen, gestrichen werden müssen (sobald „zwingende sachlogische Begründungen“ dies rechtfertigen). Saubere statistische Entscheidungen über die Wertung eines Messwertes als Ausreißer sind nur selten möglich.

Ausreißer sind aus Sicht eines Beobachters überraschend extrem gelegene Beobachtungen, wobei der Beobachter ein Modell voraussetzt und den **Ausreißer als Modell-Abweichung** auffasst. Steht die Schätzung der Modellparameter im Vordergrund, dann stört er – andernfalls liegt ein „erfreulicher Fund“ vor, der wie im Fall der Entdeckung des Penizillins Neuland erschließt. **Wunder** lassen sich in diesem Kontext auch als extreme Ausreißer nichtwissenschaftlicher Art auffassen.

Gegen starke Abweichungen vom Modell der Normalverteilung (Asymmetrie, wesentlich schwächer oder stärker besetzte Verteilungsenden) werden drei Verfahren angewandt:

- (1) **robuste Verfahren,**
- (2) **Ausreißerverfahren** und
- (3) **adaptive Verfahren.**

Ein statistisches Verfahren (Schätzfunktion oder Test), das **unempfindlich** ist **gegenüber verunreinigten Daten** (z.B. durch Ausreißer), wird „**resistant**“, widerstandsfähig oder robust, genannt; $\bar{X}$ und S sind äußerst „nonresistant“, dagegen wird der Median $\tilde{X}$ erst bei mindestens 50% Verunreinigung verzerrt. Robuste Verfahren befassen sich somit hauptsächlich mit Modifikationen der Methode der kleinsten Quadrate. Als **robuste Lage-Schätzer** gelten z.B. das 10%-getrimmte arithmetische Mittel (bei asymmetrischer Kontamination ist der 25%-getrimmte Mittelwert besser geeignet). Bei kleinem Kontaminationsanteil eignen sich Ausreißerverfahren, nicht aber bei asymmetrischer Kontamination. Am bekanntesten sind Verfahren, die einen Ausreißer identifizieren und ihn dann beim Schätzen der Modellparameter nicht weiter verwenden. Adaptive Verfahren sind zweistufig. Auf der 1. Stufe wird entschieden, welches Schätz- oder Testverfahren auf der 2.

Stufe verwendet werden soll: ist z.B. die Spannweite klein, so eignet sich $(x_{\min} + x_{\max})/2$ als robusten Lageschätzer, ist sie groß, so bevorzugt man den Median.

Man unterscheide:

- (a) mutmaßliche Ausreißer,
- (b) Ziffernvertauschungen (z. B. 18,83 anstatt 18,38),
- (c) doppelte Eintragungen, spalten- oder zeilenweise, wobei die richtige Eintragung fehlt, sowie
- (d) falsche Zahlen in den entsprechenden Spalten.

Ein Teil dieser Irrtümer lässt sich erfassen, wenn für jede Variable Variationsbereiche vorgegeben werden; mitunter hilft auch ein Blick auf die Tabelle, um sonderbare Zahlen rechtzeitig aufzuspüren. Je umfangreicher der Datenkörper ist, desto schwieriger wird es, Fehler der genannten Art vollständig zu vermeiden.

Eine **allgemeine Regel** besagt, dass bei mindestens 10 Einzelwerten (besser: $n \gtrapprox 25$) dann ein Wert als Ausreißer verworfen werden darf, wenn er außerhalb des Bereiches $\bar{x} \pm 4s$ liegt, wobei Mittelwert und Standardabweichung ohne den ausreißerverdächtigen Wert berechnet werden. Der „4-Sigma-Bereich“ ($\mu \pm 4\sigma$) umfasst (große Stichprobenumfänge!) bei Normalverteilung 99,99% der Werte, bei symmetrisch-eingipfligen Verteilungen 97% und bei beliebigen Verteilungen noch 94% der Werte.

Robuste Grenzen für die Erkennung von Ausreißern für viele Verteilungstypen können auch auf der Grundlage der Quartile und der Quartildistanz abgeleitet werden:

$$Q_1 - k(Q_3 - Q_1) \leq \text{kein Ausreißer} \leq Q_3 + k(Q_3 - Q_1)$$

für eher explorative Studien wählt man $k=1,5$
konfirmative $k=3$

Beobachtungen, die außerhalb dieser Grenzen liegen, werden in Box-Plots häufig besonders markiert und einzeln dargestellt.

Ein weiteres robustes Kriterium für einen Ausreißer x_a kann aus dem Medianwert $\tilde{x}$ und der Mediandeviation $\tilde{D}$ abgeleitet werden (Hampel [Ham85]).

$$|\tilde{x} - x_a| > 5, 2\tilde{D} \quad (7.31)$$

Die Berechnung kann in **R** direkt mit den Funktionen **median()** und **mad()** erfolgen. Dabei ist zu beachten, dass in der Funktion **mad()** aus Gründen der Konsistenz ($E[\text{mad}(X)] = \sigma$) ein konstanter Faktor mitgeführt wird, der in diesem Fall auf 1 gesetzt werden muss.

```
> x <- c(2, 3, 4, 5, 6, 7, 20)
> med.x <- median(x)
> mad.x <- mad(x, constant=1)
> outlier <- (x < med.x - 5.2*mad.x) | (x > med.x + 5.2*mad.x); x[outlier]
[1] 20
```

Sind auf diese Art Ausreißer „identifiziert“ und von der Stichprobe ausgeschlossen worden, dann muss dies bei der Analyse der Daten angemerkt werden; zumindest ihre Zahl sollte nicht verschwiegen werden. Vielleicht ist es am zweckmäßigsten, wenn eine Stichprobe Ausreißer enthält, einmal die statistische Analyse **mit** und einmal **ohne** die Ausreißer vorzunehmen. Unterscheiden sich die Schlussfolgerungen aus beiden Analysen, dann ist eine außerordentlich vorsichtige und umsichtige Interpretation der Daten zu empfehlen. So kann der Ausreißer auch einmal als Ausdruck der für die Grundgesamtheit typischen Variabilität der aufschlussreichste Wert einer Stichprobe sein und Ausgangspunkt einer neuen Messreihe werden!

7.2.8.1 Grubbs' Test für Ausreißer

Der Grubbs-Test [Gru69] kann zur Identifikation von Ausreißern in normalverteilten Daten verwendet werden, d.h. ein Test auf Normalverteilung sollte dem Grubbs-Test in jedem Fall vorangehen. Der Test entdeckt jeweils einen einzelnen Ausreißer; dieser wird aus den Daten entfernt und anschließend kann der Test mit den verbleibenden Werten solange wiederholt werden, bis keine Ausreißer mehr erkannt werden. Dabei ist zu beachten, dass die wiederholte Anwendung des Tests Einfluss auf die Wahrscheinlichkeit der Entdeckung von Ausreißern nimmt (multiples Testen, Anpassung des Signifikanzniveaus). Der Test sollte nicht bei zu kleinen Stichproben ($n > 8$) angewendet werden (für $3 \leq n \leq 8$ wird der Q-Test nach Dixon bevorzugt, für $n \geq 25$ können die standardisierten Extremabweichungen verwendet werden).

Der Grubbs-Test prüft die Hypothese

H_0 : Es gibt keine Ausreißer in den Daten

H_A : Es gibt mindestens einen Ausreißer

Die Teststatistik, die größte absolute Abweichung vom Mittelwert ($\bar{x}$) der Stichprobe in Einheiten der Standardabweichung (s), wird in (7.32) definiert.

$$\hat{G} = \frac{\max(|x_i - \bar{x}|)}{s} \quad (7.32)$$

Die Nullhypothese kann abgelehnt werden, sobald der Wert der Teststatistik $\hat{G}$ größer ist als der entsprechende kritische Wert $G_{n,\alpha}$, der nach (7.33) berechnet werden kann.

$$\hat{G} > G_{n,\alpha} = \frac{(n-1)}{\sqrt{n}} \sqrt{\frac{t_{n-2,\alpha/2n}^2}{n-2 + t_{n-2,\alpha/2n}^2}} \quad (7.33)$$

Dabei ist $t_{n-2,\alpha/2n}$ der kritische Wert der t-Verteilung mit $(n-2)$ Freiheitsgraden und dem Signifikanzniveau von $\alpha/2n$.

Beispiel: Die Anzahl der Überstunden aus dem letzten Kalenderjahr für 20 Mitarbeiter einer Firma betrug:

3	4	4	5	6	6	7	8	9	10
10	11	13	15	16	17	19	19	20	50



Die Berechnung der Teststatistik zum Grubbs-Test und die Bestimmung des entsprechenden kritischen Wertes kann in **R** direkt erfolgen.

```
> x <- c(3, 4, 4, 5, 6, 6, 7, 8, 9, 10, 10, 11, 13, 15, 16, 17, 19, 19, 20, 50)
> n <- length(x); m.x <- mean(x); s.x <- sd(x);
> alpha <- 0.05; t <- qt(alpha/(2*n), n-2)
> G.hat <- max(abs(x-m.x))/s.x; G.hat
[1] 3.610448
> G.crit <- ((n-1)/sqrt(n)) * sqrt(t^2 / (n-2+t^2)); G.crit
[1] 2.708246
```

Da der Wert $\hat{G} = 3,6$ größer ist als der kritische Wert $G_{n,\alpha} = 2,71$, kann die Nullhypothese abgelehnt werden, d.h. es gibt mindestens einen Ausreißer, hier 50 Stunden, unter den beobachteten Daten.

7.2.8.2 Q-Test nach Dixon für $3 \leq n \leq 8$

Der Q-Test wird verwendet, um einen Ausreißer in einer Folge von normalverteilten Messwerten (Messwiederholungen) zu erkennen. Der Test sollte umsichtig und **niemals mehrfach** auf einen Datensatz angewendet werden.

$$\hat{Q}_n = \frac{|x_a - x_b|}{|x_{\max} - x_{\min}|} \quad (7.34)$$

Dabei bezeichnet x_a den als Ausreißer verdächtigen Wert und x_b ist der Wert, der dem vermutlichen Ausreißer am nächsten liegt. x_a kann als Ausreißer verworfen werden, wenn $\hat{Q}_n$ aus (7.34) größer ist als der entsprechenden kritische Wert aus Tabelle 7.12 (Dixon [Dix53]).

Tabelle 7.12. Kritische Werte für den Ausreißertest nach Dixon (auszugsweise aus W.J. Dixon: Processing data for outliers, Biometrics **9**, (1953), p.89

n	$\alpha = 0,10$	$\alpha = 0,05$	$\alpha = 0,01$
3	0,941	0,970	0,995
4	0,765	0,829	0,926
5	0,642	0,710	0,821
6	0,560	0,625	0,764
7	0,507	0,568	0,680
8	0,468	0,526	0,634



Beispiel: Unter den Messungen ($n = 6$)

11,67 12,23 12,42 12,44 12,45 12,48

wird der Wert 11,67 als Ausreißer verdächtigt. Die Differenz zum nächstliegenden Wert ist 0,56, die Spannweite beträgt 0,81. Daraus folgt $\hat{Q}_n = 0,69 > 0,560 = Q_{0,10}$. Der Wert kann somit auf dem 10%-Signifikanzniveau als Ausreißer entfernt werden.

7.2.8.3 Standardisierte Extremabweichungen

Bei Stichprobenumfängen über $n = 25$ lassen sich die Extremwerte mit Hilfe der Tabelle 7.13 anhand der Prüfgröße

$$T_1 = |(x_1 - \mu)/\sigma| \quad x_1 = \text{der mutmaßliche Ausreißer} \quad (7.35)$$

testen, wobei μ und σ durch $\bar{x}$ und s ersetzt werden. Erreicht oder überschreitet T_1 für die vorgegebene Irrtumswahrscheinlichkeit $\alpha = 1 - P$ die dem Stichprobenumfang n entsprechende Schranke der Tabelle 7.13, so ist anzunehmen, dass der geprüfte Extremwert einer anderen Grundgesamtheit entstammt als die übrigen Werte der Reihe. Der Extremwert darf jedoch, auch wenn er durch diese Tests als Ausreißer ausgewiesen ist, nur dann gestrichen werden, wenn wahrscheinlich ist, dass die vorliegenden Werte **angenähert normalverteilt** sind.

Für Probleme, die mit der **Qualitätsüberwachung** zusammenhängen, hat Tabelle 7.13 eine besondere Bedeutung. Angenommen, von einem Gegenstand mit $\bar{x} = 888$ und $s = 44$ werden jeweils Stichproben des Umfangs $n = 10$ geprüft. Der niedrigste Stichprobenwert sollte dann höchstens einmal in hundert Fällen kleiner sein als $888 - 44 \cdot 3,089 = 752,1$ (vgl. für $n = 10$ und $P = 99\%$ erhält man den Faktor 3,089). Durch Vorzeichenwechsel $888 + 44 \cdot 3,089 = 1023,9$ erhält man den größten Stichprobenwert, der höchstens einmal in hundert Fällen rein zufällig überschritten werden dürfte. Treten **Extremwerte** dieser Art häufiger auf, muss die Produktion des betreffenden Gegenstandes überprüft werden.

Tabelle 7.13. Obere Signifikanzschranken der standardisierten Extremabweichung $(x_{\max} - \mu)/\sigma$ oder $(\mu - x_{\min})/\sigma$ (Normalverteilung vorausgesetzt) (auszugsweise aus Pearson, E.S. and H.O. Hartley: Biometrika Tables for Statisticians, Cambridge University Press 1954, Table 24)

n	$P = 95\%$	$P = 99\%$	n	$P = 95\%$	$P = 99\%$
1	1,645	2,326	55	3,111	3,564
2	1,955	2,575	60	3,137	3,587
3	2,121	2,712	65	3,160	3,607
4	2,234	2,806	70	3,182	3,627
5	2,319	2,877	80	3,220	3,661
6	2,386	2,934	90	3,254	3,691
8	2,490	3,022	100	3,283	3,718
10	2,568	3,089	200	3,474	3,889
15	2,705	3,207	300	3,581	3,987
20	2,799	3,289	400	3,656	4,054
25	2,870	3,351	500	3,713	4,106
30	2,928	3,402	600	3,758	4,148
35	2,975	3,444	700	3,797	4,183
40	3,016	3,479	800	3,830	4,214
45	3,051	3,511	900	3,859	4,240
50	3,083	3,539	1000	3,884	4,264

7.3 Einstichprobenverfahren

- Hypothesen zu Wahrscheinlichkeiten
- Hypothesen zu Erwartungswerten
- Einstichproben-Median-Test
- Vergleich einer empirischen Varianz mit ihrem Parameter
- Prüfung der Zufallsmäßigkeit
- Prüfung der Erwartungswerte von Poisson-Verteilungen

7.3.1 Hypothesen zu Wahrscheinlichkeiten

7.3.1.1 Binomialtest

Der Binomialtest prüft Hypothesen bezüglich einer (unbekannten!) Wahrscheinlichkeit (π) an Hand des Modells einer Binomialverteilung, z.B. bei einer einseitigen Fragestellung, ob die Wahrscheinlichkeit für den Erfolg einer bestimmten Therapie größer als ein vermuteter oder angenommener Wert ($\pi_0 = 0.7$) ist.

$$H_A : \pi > \pi_0 \quad (7.36)$$

$$H_0 : \pi \leq \pi_0$$

Unter der Annahme, dass $\pi = \pi_0$ gilt, kann dann für eine feste Zahl von „Versuchen“ (z.B. $n=30$) die Verteilung durch eine binomial-verteilte Zufallsvariable beschrieben werden. Das vollständige Modell ist in Abbildung 7.11 dargestellt.

Die Entscheidung gegen die Nullhypothese aus (7.36) kann dann für das Ergebnis einer konkreten Stichprobe (X) mit Hilfe der Binomialverteilung getroffen werden. Ist z.B. $P(X \geq x) \leq 0.05$, kann die Nullhypothese bei einer Irrtumswahrscheinlichkeit von $\alpha = 0,05$ abgelehnt werden.

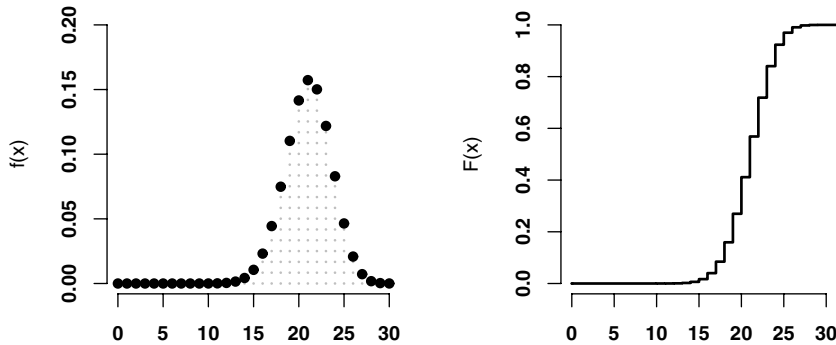


Abb. 7.11. Verteilung für die Anzahl der Erfolge unter $n = 30$ Versuchen: Annahme einer Erfolgswahrscheinlichkeit von $\pi = 0.7$

In **R** erfolgen die Berechnungen zum Binomialtest entweder direkt über die Funktionen zur Binomialverteilung (**pbinom()** und **qbinom()**), oder einfacher über die Funktion **binom.test()**, in der zusätzlich auch die Berechnung des entsprechenden Vertrauensbereiches durchgeführt wird.

```
> pbinom(25, 30, 0.7, lower.tail=FALSE)
[1] 0.03015494
> binom.test(26, 30, p=0.7, alternative="greater")

Exact binomial test

data: 26 and 30
number of successes = 26, number of trials = 30, p-value = 0.03015
alternative hypothesis: true probability of success is greater than 0.7
95 percent confidence interval: 0.7203848 1.0000000
sample estimates: probability of success 0.8666667

> qbinom(0.95, 30, 0.7)
[1] 25
```

In der Funktion **binom.test()** muss neben der Anzahl der Wiederholungen (n), der Anzahl der Erfolge (x) und der unter der Nullhypothese angenommenen Erfolgswahrscheinlichkeit (π) auch die Art der Hypothesenstellung (einseitig engl. less oder greater) bzw. zweiseitig (engl. two.sided)) angegeben werden. Werden z.B. unter $n = 30$ Behandlungen $x = 26$ Therapieerfolge beobachtet ($\hat{\pi} = 0.87$), dann kann die Nullhypothese $H_0 : \pi \leq 0.7$ mit einer Irrtumswahrscheinlichkeit von $\alpha = 0.05$ abgelehnt werden. Die Wahrscheinlichkeit für einen Erfolg ist somit größer als 0,7 ($P = 0.03015$, engl. p-value).

Um andererseits zu prüfen, ob die Wahrscheinlichkeit für das Auftreten unerwünschter Nebenwirkungen einer Therapie kleiner als ein angenommener oder vorgeschriebener Wert ist, kann in ähnlicher Weise eine einseitige Hypothesenstellung formuliert werden.

$$H_A : \pi < \pi_0 \quad (7.37)$$

$$H_0 : \pi \geq \pi_0$$

Die zweiseitige Fragestellung, z.B. bezüglich gleicher Chancen bei einem Glücksspiel (Wurf einer Münze) führt dazu, dass die entsprechende Irrtumswahrscheinlichkeit auf beiden Seiten der Verteilung (in beiden Richtungen) zu berücksichtigen ist.

$$H_A : \pi \neq \pi_0 \quad (7.38)$$

$$H_0 : \pi = \pi_0$$



Beispiel: Sind statistisch Zweifel an der „Regularität“ einer Münze ($\pi = 0,5$ für „Kopf“) berechtigt, wenn bei 20 Würfeln mit dieser Münze in 15 Fällen das Ereignis „Kopf“ beobachtet wurde?

```
> binom.test(15, 20, p=0.5, alternative="two.sided")
```

Exact **binomial** test

data: 15 and 20

number of successes = 15, number of trials = 20, p-value = 0.04139

alternative hypothesis: true probability of success is not equal to 0.5

95 percent confidence interval: 0.5089541 0.9134285

sample estimates: probability of success 0.75

Es handelt sich „vermutlich“ nicht um eine reguläre Münze. Die Nullhypothese ($H_0 : \pi = 0,5$) kann mit einer Irrtumswahrscheinlichkeit von $\alpha = 0,05$ abgelehnt werden, da $P = 0,04139$.

Die Testentscheidung für den zweiseitigen Binomialtest wird dabei exakt an Hand der Binomialverteilung getroffen. Die Nullhypothese wird verworfen, wenn gilt:

$$\sum_{i=0}^x \binom{n}{i} \pi_0^i (1 - \pi_0)^{n-i} + \sum_{i=n-x}^n \binom{n}{i} \pi_0^i (1 - \pi_0)^{n-i} < \alpha \quad (7.39)$$

Die Berechnung von (7.39) mit der Funktion **pbinom()** in **R** (Zahlen aus dem Beispiel) führt zum gleichen Ergebnis:

```
> n <- 20; x <- 15; p0 <- 0.5
> pbinom(n-x, n, p0, lower.tail=TRUE) + pbinom(x-1, n, p0, lower.tail=FALSE)
[1] 0.04138947
```

7.3.1.2 Binomialtest - Approximation durch die Normalverteilung

Gegeben sei die empirische relative Häufigkeit $x/n = \hat{p}$, der wir den Parameter $\pi = p$ zugrunde legen. Verglichen wird π mit der Wahrscheinlichkeit, dem Parameter oder Sollwert, $\pi_0 = p_0$. Für $np_0q_0 = np_0(1 - p_0) > 9$ kann anhand der Quantile der Standardnormalverteilung entschieden werden (vgl. $x_o = x_{\text{oben}}$; $x_u = x_{\text{unten}}$ sowie Tabelle 7.14).

Tabelle 7.14. Binomialtest für $np_0q_0 = np_0(1 - p_0) > 9$ und $\alpha = 0,05$

Hypothesenpaar	H_0 wird auf dem 5%-Niveau abgelehnt für
$H_{01} : p \leq p_0$ $H_{A1} : p > p_0$	$x > x_o = np_0 + \frac{1}{2} + 1,645 \cdot \sqrt{np_0(1 - p_0)}$
$H_{02} : p \geq p_0$ $H_{A2} : p < p_0$	$x < x_u = np_0 - \frac{1}{2} - 1,645 \cdot \sqrt{np_0(1 - p_0)}$
$H_{03} : p = p_0$ $H_{A3} : p \neq p_0$	$x > x_o$ oder $x < x_u$, wobei 1,960 den Wert 1,645 ersetzt

Beispiel 1: Wir prüfen für $\hat{p} = x/n = 30/100 = 0,3$ die $H_{01} : p \leq p_0 = 0,2$ gegen $H_{A1} : p > p_0 = 0,2$. Mit $100 \cdot 0,2(1 - 0,2) = 16 > 9$ und



$$x_o = 100 \cdot 0,2 + 0,5 + 1,645 \cdot \sqrt{100 \cdot 0,2 \cdot 0,8} = 27,08,$$

d. h. $x = 30 > 27,08 = x_o$, wird H_{01} bei einseitiger Fragestellung auf dem 5%-Niveau abgelehnt.

Beispiel 2: In einer Großstadt hielten $\pi = 20\%$ der Familien eine bestimmte Zeitschrift. Es besteht Grund zu der Annahme, dass die Zahl der Abonnenten jetzt unter 20% liegt. Um diese Hypothese zu überprüfen, wird eine Zufallsstichprobe, bestehend aus 100 Familien, ausgewählt und ausgewertet, wobei $\hat{p}_1 = 0,16$ (16%) gefunden wurde. Getestet wird die Nullhypothese $\pi_1 = 20\%$ gegen



die Alternativhypothese $\pi_1 < 20\%$ (Signifikanzniveau $\alpha = 0,05$). Auf die Endlichkeitskorrektur können wir verzichten, da die Grundgesamtheit im Verhältnis zur Stichprobe sehr gross ist. Da $n\pi(1 - \pi) = 16 > 9$ benutzen wir die Approximation über die Normalverteilung

$$\hat{z} = \frac{|\hat{p}_1 - \pi| - \frac{1}{2n}}{\sqrt{\frac{\pi(1 - \pi)}{n}}} = \frac{|0,16 - 0,20| - \frac{1}{2 \cdot 200}}{\sqrt{\frac{0,20 \cdot 0,08}{100}}} = 0,875$$

Ein Wert $z = 0,875$ entspricht einer Irrtumswahrscheinlichkeit $P\{\hat{p}_1 \leq 0,16 | \pi = 0,20\} = 0,19 > 0,05$. Damit weisen 19 von 100 Zufallsstichproben aus einer Grundgesamtheit mit $\pi = 0,20$ einen Abonnentenanteil $\hat{p}_1 \leq 0,16$ auf. Wir behalten daher die Nullhypothese bei.



Beispiel 3: Von 2000 Händlern entscheiden sich $\pi = 40\%$ z.B. ihren Umsatz zu erhöhen. Kurze Zeit später wird angenommen, dass sich dieser Prozentsatz wieder erhöht habe. Eine Zufallsstichprobe von 400 Händlern zeigt, dass der Prozentsatz mit $\hat{p}_1 = 46\%$ tatsächlich höher liegt. Gefragt ist, ob diese Zunahme als statistisch signifikant gelten kann. Getestet wird die Nullhypothese $\pi_1 = 0,40$ gegen die Alternativhypothese $\pi_1 > 0,40$ mit $\hat{p}_1 = 0,46$ (Signifikanzniveau $\alpha = 0,05$). Da die Stichprobe 20% der Grundgesamtheit umfasst, muss mit einer Endlichkeitskorrektur gerechnet werden:

$$\hat{z} = \frac{|\hat{p}_1 - \pi| - \frac{1}{2n}}{\sqrt{\left[\frac{\pi(1 - \pi)}{n}\right] \cdot \left[\frac{N - n}{N - 1}\right]}} = \frac{|0,46 - 0,40| - \frac{1}{2 \cdot 400}}{\sqrt{\left[\frac{0,40 \cdot 0,60}{400}\right] \cdot \left[\frac{2000 - 400}{2000 - 1}\right]}} = 2,68$$

$$P\{\hat{p}_1 \leq 0,46 | \pi = 0,40\} = 0,0037 < 0,05$$

Die Nullhypothese wird auf dem 5%-Niveau abgelehnt, da $\hat{z} > z_{0,95} = 1,64$: Es besteht eine echte Zunahme.

Eine Lösung in **R** kann entsprechend direkt formuliert werden.

```
> N <- 2000; n <- 400; x <- 184; p0 <- 0.40; p <- x/n
> z <- (abs(p - p0) - 1/(2*n)) / sqrt(((p0*(1-p0))/n)*((N-n)/(N-1))); z
[1] 2.680888
> pnorm(z, lower.tail=F)
[1] 0.003671356
```

(Der exakte P-Wert, berechnet aus der Binomialverteilung mit der Funktion **binom.test()** in **R**, beträgt 0,00854.)

7.3.1.3 Binomialtest - Fallzahlabeschätzung

Die Bestimmung der Fallzahl für den Binomialtest, hier bei einseitiger Hypothesenstellung, kann nach (7.40) erfolgen.

$$n \geq \frac{(z_{1-\alpha} + z_{1-\beta})^2}{(\pi - \pi_0)^2} \cdot (\pi_0(1 - \pi_0) + \pi(1 - \pi)) \quad (7.40)$$



Hinweis: Die Bedingung (7.40) basiert auf einer Approximation der Binomialverteilung durch die Standardnormalverteilung (vgl. den vorangehenden Abschnitt). Zur Herleitung und näheren Begründung siehe auch Fallzahlabeschätzung für den Einstichproben-t-Test.

Falls der wahre Wert von π um mindestens $\delta = \pi - \pi_0$ von dem Sollwert π_0 abweicht, so wird dieser Unterschied bei einer Stichprobe vom Umfang n mit einer Power von mindestens $(1 - \beta)$

bei einer Irrtumswahrscheinlichkeit von höchstens α aufgedeckt. Einige Stichprobenumfänge sind für $\alpha = 0,05$, $(1 - \beta) = 0,80$ und $\delta = 0,10$ in der Tabelle 7.15 zusammengestellt (vgl. die Symmetrie).

Tabelle 7.15. Stichprobenumfänge zum Binomialtest für $\alpha = 0,05$, $(1 - \beta) = 0,80$ und ausgewählte $\delta = 0,10$:

π_0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8
π	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
n	155	229	279	303	303	279	229	155

```
> alpha <- 0.05; beta <- 0.20
> p0 <- c(0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8)
> p <- c(0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9)
> ceiling(((qnorm(1-alpha) + qnorm(1-beta))^2 * (p0*(1-p0)+p*(1-p))) / (p-p0)^2)
[1] 155 229 279 303 303 279 229 155
```

In **R** steht für die Fallzahlabschätzung zum Binomialtest die Funktion **power.prop.test()** zur Verfügung.

```
> power.prop.test(n=NULL, p1=0.1, p2=0.2, sig.level=0.05, power=0.80,
  alternative="one.sided")
```

Two-sample comparison of proportions power calculation

```
      n = 156.6054
      p1 = 0.1
      p2 = 0.2
sig.level = 0.05
  power = 0.8
alternative = one.sided
```

NOTE: n is number in *each* group

Für die statistische Bewertung eines Unterschiedes von $\delta = 0,10$ zwischen $\pi_0 = 0,1$ und $\pi = 0,2$ erhält man hiermit bei einseitiger Hypothesenstellung einen erforderlichen Stichprobenumfang von $n = 157$.

7.3.1.4 Binomialtest: Likelihood-Quotienten-Test

Werden unter n Wiederholungen genau x Erfolge beobachtet ($\hat{\pi} = x/n$), dann kann für den Fall einer einseitigen Hypothesenstellung ($H_0 : \pi = \pi_0$) die logarithmierte Likelihood-Funktion für den unbekannten Parameter π

$$L(\pi) = \log(\pi^x (1 - \pi)^{n-x}) = x \log(\pi) + (n - x) \log(1 - \pi) \quad (7.41)$$

zur Konstruktion eines einfachen Likelihood-Quotiententests herangezogen werden.

$$\begin{aligned} \text{unter } H_0: \quad L_0 &= x \log(\pi_0) + (n - x) \log(1 - \pi_0) \\ \text{unter } H_A: \quad L_A &= x \log(\hat{\pi}) + (n - x) \log(1 - \hat{\pi}) \\ \text{Teststatistik:} \quad -2(L_0 - L_A) &= 2 \left(x \log \frac{x}{n\pi_0} + (n - x) \log \frac{n - x}{n - n\pi_0} \right) = \hat{\chi}^2 \end{aligned} \quad (7.42)$$

Die Teststatistik in (7.42) ist asymptotisch χ^2 -verteilt mit einem Freiheitsgrad. Die Nullhypothese kann abgelehnt werden, wenn der Wert der Teststatistik größer ist als das entsprechende Quantil der χ^2 -Verteilung: $-2(L_0 - L_A) > \chi_{1;1-\alpha}^2$.



Beispiel: Ein Würfel wird verdächtigt, gefälscht zu sein. Bei 60 Würfeln wurde nur 4mal eine Sechs beobachtet. Für den statistischen Test gegen die einseitige Alternative $H_A : \pi < 1/6$ folgt daraus:

```
> n <- 60; x <- 4; p0 <- 1/6
> minus2ll <- 2*(x*log(x/(n*p0)) + (n-x)*log((n-x)/(n-n*p0))); minus2ll
[1] 5.362487
> qchisq(0.95, 1)
[1] 3.841459
> pchisq(minus2ll, 1, lower.tail = FALSE)
[1] 0.02057441
```

Mit $\hat{\chi}^2 = 5,36 > 3,84 = \hat{\chi}_{1;0,95}^2$ kann H_0 auf dem 5%-Niveau verworfen werden ($p = 0,02057$). Der exakte P-Wert aus der Funktion **binom.test()** ist für dieses Beispiel $P=0,02019$.

7.3.2 Hypothesen zu Erwartungswerten, die sich auf einen empirischen Mittelwert beziehen

7.3.2.1 Einstichproben-t-Test

Die Frage, ob der Mittelwert $\bar{x}$ einer konkreten Stichprobe nur zufällig oder statistisch signifikant von einem vorgegebenen Erwartungswert μ_0 verschieden ist, heißt anschaulich: Schließt das mit $\bar{x}$ berechnete Konfidenzintervall für μ den vorgegebenen Erwartungswert μ_0 ein oder nicht, d. h. ist also die absolute Differenz $|\bar{x} - \mu_0|$ kleiner oder größer als die halbe Konfidenzintervallspanne $ts/\sqrt{n}$?

Eine Stichprobe habe den Umfang n und die Standardabweichung s ; dann ist der Unterschied ihres Mittelwertes $\bar{x}$ vom vorgegebenen Erwartungswert μ_0 auf dem 100 α %-Niveau statistisch signifikant, wenn

$$|\bar{x} - \mu_0| > t \frac{s}{\sqrt{n}} \quad \text{oder} \quad \frac{|\bar{x} - \mu_0|}{s} \cdot \sqrt{n} > t \quad (7.43)$$

wobei der Wert t für den Freiheitsgrad $n - 1$ und die geforderte Irrtumswahrscheinlichkeit α dem Quantil $t_{1-\alpha/2}$ (zweiseitige Fragestellung) der t-Verteilung (Tabelle 5.10 oder mit der Funktion **qt()** in **R**) entspricht. Die Grenze, bei der und oberhalb der ein Unterschied auf dem 100 α %-Niveau statistisch signifikant und unterhalb der ein Unterschied zufällig ist, liegt somit für diesen sogenannten **Einstichproben-t-Test** bei

$$t = \frac{|\bar{x} - \mu_0|}{s} \cdot \sqrt{n} \quad \text{FG} = n - 1 \quad (7.44)$$

Bei großen Stichprobenumfängen kann t durch das für die vorgegebene Irrtumswahrscheinlichkeit Quantil der Standardnormalverteilung $z_{1-\alpha/2}$ (Tabelle 5.9) ersetzt werden. Da Parameter verglichen werden – μ_0 mit dem der Stichprobe zugrundeliegenden μ – liegt ein **Parametertest** vor.



Beispiel: Eine Stichprobe vom Umfang $n = 25$ aus einer Normalverteilung habe $\bar{x} = 9$ und $s = 2$ ergeben. Gefragt wird, ob die Nullhypothese $H_0: \mu = \mu_0 = 10$ auf dem 5%-Niveau aufrechterhalten werden kann (zweiseitige Fragestellung: d. h. $H_A: \mu \neq \mu_0 = 10$).

$$\hat{t} = \frac{|9 - 10|}{2} \sqrt{25} = 2,50 > 2,06 = t_{24;0,975}.$$

Da $2,50 > 2,06$, wird H_0 auf dem 5%-Niveau abgelehnt. Die Berechnungen können leicht direkt in **R** nachvollzogen werden.

```
> m <- 9; s <- 2; n <- 25
> t.hat <- abs(m-10)/(s/sqrt(n)); t.hat
[1] 2.5
> t.krit <- qt(0.975, n-1); t.krit
[1] 2.063899
```

Tabelle 7.16. Hypothesen zum Einstichproben- t-Test und der P-Wert

Alternative	Die Nullhypothese wird abgelehnt für	P ist gleich der Fläche unter der t_{n-1} -Verteilung
$H_{A1} : \mu > \mu_0$	$\hat{t} = \frac{\bar{x} - \mu_0}{s} \cdot \sqrt{n} > t_{n-1;1-\alpha}$	rechts von $\hat{t}$
$H_{A2} : \mu < \mu_0$	$\hat{t} = \frac{\bar{x} - \mu_0}{s} \cdot \sqrt{n} < -t_{n-1;1-\alpha}$	links von $-\hat{t}$
$H_{A3} : \mu \neq \mu_0$	$\hat{t} = \frac{ \bar{x} - \mu_0 }{s} \cdot \sqrt{n} > t_{n-1;1-\alpha/2}$	rechts/links von $\pm \hat{t}$

Der P -Wert ist insbesondere dann aufschlussreich, wenn die H_0 plausibel erscheint. Außerdem wird er bei verteilungsunabhängigen Methoden (z. B. Chiquadrat-Unabhängigkeitstest und Rangsummen-Tests) angegeben sowie dann, wenn mehrere Parameter zugleich geschätzt werden, etwa in der Varianzanalyse.

Beispiel: Die Behauptung, dass der mittlere diastolische Blutdruck (DBP) bei Patienten mit Übergewicht höher als 80mmHg ist ($H_A : \mu > 80$), soll an einer Stichprobe von $n = 11$ adipösen Männern im Alter von 40-50 Jahren überprüft werden. Das Ergebnis aus einer Stichprobe ist $\bar{x} = 85\text{mmHg}$ und $s = 9\text{mmHg}$. Die Verteilung der Mittelwerte zum DBP (aus Stichproben vom Umfang $n = 11$) unter Annahme der Nullhypothese ($H_0 : \mu = 80$) enthält die linke Seite der Abbildung 7.12. Der Wert der Prüfgröße $\hat{t} = 1,84$ ist größer als das 95%-Quantil der t-Verteilung mit 10 Freiheitsgraden $t_{10;0,95} = 1.8125$ ($P = 0,048$). Andererseits kann der Bereich für eine Ablehnung von H_0 auch mit dem Wert $x_{krit} = \mu_0 + t_{n-1;1-\alpha} \cdot s/\sqrt{n} = 84,92$ bestimmt werden (Abbildung 7.12); der beobachtete Mittelwert fällt somit in den Ablehnungsbereich.



Eine Überlegung hinsichtlich der Power dieser Testentscheidung setzt die Festlegung eines Effektes $\delta = 8\text{mmHg}$, d.h. eines tatsächlichen Erwartungswertes $\mu = 88\text{mmHg}$ voraus. Unter dieser zusätzlichen Annahme kann dann die Verteilung unter der Alternativhypothese für die Berechnung der Power herangezogen werden (vgl. rechte Seite in Abbildung 7.12): $(1 - \beta) = (1 - 0,14) = 0,86$.

Die Prüfung einer einseitigen Hypothese (z.B. $H_A : \mu > \mu_0$) mit der Funktion **t.test()** in **R** soll an einem weiteren Beispiel gezeigt werden.

Beispiel: Die Behauptung, dass unter bestimmten Bedingungen die mittlere Körpertemperatur größer ist als 37°C, soll an Hand einer Stichprobe mit den folgenden Werten geprüft werden.



1	2	3	4	5	6	7	8
36,8°	37,2°	37,5°	37,0°	36,9°	37,4°	37,9°	38,0°

```
> temp <- c(36.8, 37.2, 37.5, 37.0, 36.9, 37.4, 37.9, 38.0)
> t.test(temp, alternative="greater", mu=37)
```

One Sample t-test

data: temp

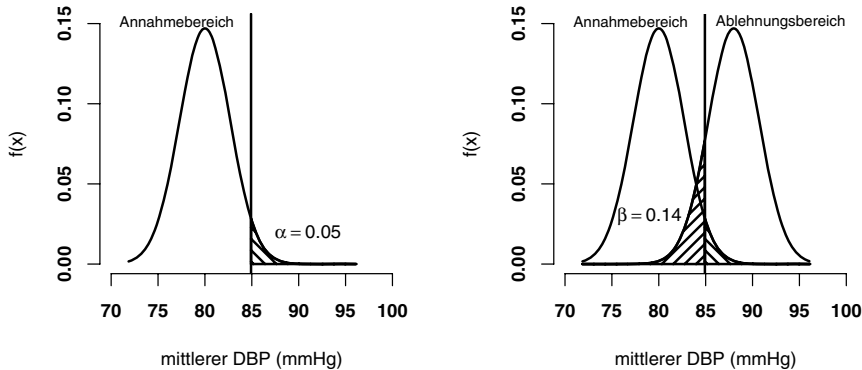


Abb. 7.12. Verteilung für den Mittelwert des diastolischen Blutdrucks (DBP) unter Null- und Alternativhypothese ($n = 11$, $s = 9\text{mmHg}$ und $\mu_0 = 80$) ($H_0 : \mu = 80\text{mmHg}$, $H_A : \mu > 80\text{mmHg}$, $\alpha = 0,05$, $\delta = 8\text{mmHg}$, $\beta = 0,14$)

```
t = 2.1355, df = 7, p-value = 0.03505
alternative hypothesis: true mean is greater than 37
95 percent confidence interval: 37.03807      Inf
sample estimates: mean of x                37.3375
```

Die Nullhypothese kann bei einseitiger Hypothesenstellung mit einer Irrtumswahrscheinlichkeit von $\alpha = 0,05$ abgelehnt werden. Die Funktion `t.test()`, die auch für den Vergleich von zwei Stichproben eingesetzt werden kann, berechnet ergänzend das Konfidenzintervall zu einer vorgegeben Wahrscheinlichkeit, hier bei einseitiger Fragestellung auch als einseitiges 95%-Konfidenzintervall, d.h. $\mu > 37,04^\circ$.



Hinweis: Eine andere Möglichkeit, die Nullhypothese ($H_0: \mu = \mu_0$ gegen $H_A: \mu \neq \mu_0$) zu prüfen, besteht darin, festzustellen, ob der konkrete Mittelwert $\bar{x}$ innerhalb des Nichtablehnungsbereiches der Nullhypothese H_0 (vgl. $\bar{K}_\alpha$ in Abschnitt [7.1.3]).

$$\mu_0 - t_{n-1;1-\alpha/2} \cdot \frac{s}{\sqrt{n}} \leq \bar{X} \leq \mu_0 + t_{n-1;1-\alpha/2} \cdot \frac{s}{\sqrt{n}} \quad (7.45)$$

liegt. Hier ist der Unterschied zwischen der Schätzfunktion $\bar{X}$ und dem konkreten Stichprobenmittelwert $\bar{x}$ wichtig! Liegt $\bar{x}$ in (7.45), so wird H_0 beibehalten. Außerhalb der beiden Annahmegrenzen liegt der **kritische Bereich, der untere und obere Ablehnungsbereich** (vgl. K_α in Abschnitt [7.1.3]). Fällt $\bar{x}$ in diesen Bereich, so wird H_0 abgelehnt. Für die einseitige Fragestellung ($H_0 : \mu \leq \mu_0$ gegen $H_A : \mu > \mu_0$) wird H_0 beibehalten, solange für den Mittelwert $\bar{x}$ einer Stichprobe des Umfanges n gilt:

$$\bar{x} \leq \mu_0 + t_{n-1;1-\alpha} \cdot \frac{s}{\sqrt{n}} \quad (7.46)$$

Bereiche dieser Art sind für die Güteüberwachung in der Industrie wichtig, sie dienen zur Überprüfung möglichst konstanter „Sollwerte“ (Parameter) wie Erwartungswerte oder Mediane, Standardabweichungen (oder Spannweiten und relativer Häufigkeiten, z. B. zulässiger Ausschussprozentsätze).

7.3.2.2 Einstichprobentest auf Äquivalenz

Die Formulierung des Äquivalenzproblems für die Situation eines Einstichprobentests in Bezug auf den Erwartungswert μ , insbesondere die Bestimmung geeigneter kritischer Grenzen für den **Äquivalenzbereich**, führt auf Probleme, da die Verteilung unter der Nullhypothese nicht bekannt ist und zusätzliche Parameter zu berücksichtigen sind (vgl. auch Abschnitt 7.1.8). Durch eine standardisierende Transformation, die unter der Hypothesenstellung invariant ist, kann das Testproblem wesentlich vereinfacht werden.

$$\begin{array}{l} H_A : \quad -\varepsilon_1 < \frac{\mu - \mu_0}{\sigma} < +\varepsilon_2 \\ H_0 : \quad \frac{\mu - \mu_0}{\sigma} \leq -\varepsilon_1 \quad \text{oder} \quad \frac{\mu - \mu_0}{\sigma} \geq +\varepsilon_2 \end{array} \quad (7.47)$$

Ein anschaulich natürliches Kriterium für die Äquivalenz ergibt sich aus der Forderung, dass die Wahrscheinlichkeit für eine positive Differenz ($(x_i - \mu_0) > 0$) genauso groß ist wie für eine negative Differenz ($(x_i - \mu_0) < 0$), d.h. jeweils „nahe“ bei $1/2$ liegt. Sind die x_i unabhängige Beobachtungen aus einer normalverteilten Grundgesamtheit, dann ist die Bedingung

$$1/2 - \tilde{\varepsilon}_1 \leq P((\bar{x} - \mu_0) > 0) = \Phi\left(\frac{\mu - \mu_0}{\sigma}\right) \leq 1/2 + \tilde{\varepsilon}_2 \quad (7.48)$$

gleichbedeutend mit

$$-\varepsilon_1 = \Phi^{-1}\left(\frac{1}{2} - \tilde{\varepsilon}_1\right) \leq \frac{\mu - \mu_0}{\sigma} \leq \Phi^{-1}\left(\frac{1}{2} + \tilde{\varepsilon}_2\right) = +\varepsilon_2 \quad (7.49)$$

Dabei bezeichnet Φ (griech. Phi) die Verteilungsfunktion der Standardnormalverteilung (Φ^{-1} bedeutet die inverse Verteilungsfunktion, Quantilfunktion). Setzt man in (7.47) nun $\varepsilon_1 = \varepsilon_2 = 0,5$, dann wird damit die maximale Abweichung zwischen der Wahrscheinlichkeit für eine positive Differenz und dem Wert 0,5 auf ca. 20% begrenzt. Diese Annahme bezüglich einer Äquivalenz ist dabei eher großzügig einzuschätzen. Für $\varepsilon_1 = \varepsilon_2 = 0,1$ sinkt der Wert auf ca. 4%.

Der Test auf Äquivalenz kann unter diesen Annahmen und Voraussetzungen sehr ähnlich dem Einstichproben t-Test gerechnet werden. Die Nullhypothese in (7.47) wird abgelehnt, wenn der Wert der Teststatistik in (7.50) kleiner ist als ein kritischer Wert, der sich aus der Wurzel des entsprechenden **Quantils der nichtzentralen Fisher-Verteilung** mit $(1, n-1)$ - Freiheitsgraden bestimmen lässt (der Nichtzentralitätsparameter wird aus $n\varepsilon^2$ berechnet).

$$\hat{t} = \frac{|\bar{x} - \mu_0|}{s} \sqrt{n} < \sqrt{F_{\alpha, 1, n-1, n\varepsilon^2}} \quad (7.50)$$

Quantile der nichtzentralen Fisher-Verteilung sind nur in wenigen Statistiklehrbüchern tabelliert und können in **R** mit einem kleinen Kunstgriff über die Funktion **uniroot()** direkt aus der Verteilungsfunktion abgeleitet werden. Das folgende Beispiel aus [Wel95] soll die Berechnung eines Einstichprobentests auf Äquivalenz in **R** zeigen.

Beispiel: In einer experimentellen Untersuchung zur medikamentösen Beeinflussung des Flows in der zerebralen Mikrozirkulation von Kaninchen ($n = 23$) soll sichergestellt werden, dass der Messparameter $[ml/min/100gGewicht]$ während einer 15-minütigen Vorbehandlungsphase ausreichend stabil ist ($\varepsilon = 0,5$). Als Signifikanzniveau wird $\alpha = 0,05$ gewählt. Die mittlere Differenz zwischen den Messzeitpunkten ist $\bar{x} = 0,16$ mit einer Standardabweichung von $s = 4,0$ ($\mu_0 = 0$).



```

> ##### Quantile zur nichtzentralen Fisher-Verteilung #####
> myqf <- function(p, df1, df2, ncp) {
+   uniroot(function(x) pf(x, df1, df2, ncp) - p, .0, 100) $ root }
> #####
>
> n <- 23; d <- 0.16; s.d <- 4.0; eps <- 0.5
>
> t.hat <- (d/s.d)*sqrt(n); t.hat # Teststatistik
[1] 0.1918333
>
> c <- sqrt(myqf(0.05, 1, n-1, ncp=n*eps^2)); c # kritischer Wert
[1] 0.7594587

```

Da der Wert der Teststatistik 0,192 hier kleiner ist als der kritische Wert 0,759, kann die Nullhypothese abgelehnt werden. Der Mikrozirkulationsflow innerhalb der Vorbehandlungsphase verändert sich nicht in einem „relevanten“ Ausmaß.

7.3.2.3 Fallzahlabschätzung zum Einstichproben-t-Test

Für die Ermittlung einer ausreichenden Fallzahl muss die Verteilung der Teststatistik unter der Alternativhypothese herangezogen werden. Einerseits ist neben dem Fehler 1. Art (Signifikanzniveau α) auch die Power ($1 - \beta$) für eine Testentscheidung vorzugeben, andererseits muss auch die zu bewertende Differenz $\mu - \mu_0$ (Effekt) festgelegt werden. Die Zusammenhänge werden in (7.51) für den Fall einer einseitigen Alternativhypothese $H_A: \mu > \mu_0$ deutlich. Die **Entscheidungsgrenze** c zwischen Ablehnungsbereich und Annahmebereich kann unter der Nullhypothese und unter der Alternativhypothese formuliert werden:

$$\begin{aligned} \text{unter } H_0 \text{ gilt: } c &= \mu_0 + t_{\nu, 1-\alpha} \frac{s}{\sqrt{n}} \\ \text{unter } H_A \text{ gilt: } c &= \mu - t_{\nu, 1-\beta} \frac{s}{\sqrt{n}} \end{aligned} \quad (7.51)$$

Fügt man die beiden Bedingungen zusammen und löst die entstehende Gleichung nach n auf, so erhält man (7.52).

$$n \geq \frac{(t_{\nu, 1-\alpha} + t_{\nu, 1-\beta})^2}{(\mu - \mu_0)^2} \cdot s^2 \quad (7.52)$$

- Hinweis:**
- Die Zahl der Freiheitsgrade ($\nu = n - 1$) für die t-Verteilung ist nicht bekannt. Man ersetzt die Quantile der t-Verteilung durch die entsprechenden Quantile der Standardnormalverteilung ($z_{1-\alpha}$ bzw. $z_{1-\beta}$ bei bekannter Varianz σ^2) und führt mit dem so berechneten Wert für n eine neue Abschätzung durch.
 - Für den Fall einer zweiseitigen Hypothesenstellung wird $\alpha/2$ anstelle von α verwendet.
 - Der Quotient aus der Differenz $\mu - \mu_0$ und der Standardabweichung σ ist ein **zentrales Maß für die Stärke des zu prüfenden Effektes**:

$$\delta = \frac{\mu - \mu_0}{\sigma}$$



Beispiel: In einer Studie soll gezeigt werden, dass der Blutzuckerwert unter einer bestimmten Therapie um 15 mg/dl gesenkt werden kann ($\mu_0 = 0$, $\mu = 15$ unter H_A). Aus Voruntersuchungen ist bekannt, dass für die Standardabweichung $\sigma \approx 30$ mg/dl angenommen werden kann. Der Test soll mit einem Signifikanzniveau $\alpha = 0,05$ und einer Teststärke (Power) von $1 - \beta = 0,80$ gerechnet werden.

```

> d <- 15; s <- 30
> effekt <- d / s
> alpha <- 0.05; beta <- 0.20
> n.1 <- ceiling((qnorm(1-alpha) + qnorm(1-beta))^2 / effekt^2); n.1
[1] 25
> n.2 <- ceiling((qt(1-alpha, n.1-1) + qt(1-beta, n.1-1))^2 / effekt^2); n.2
[1] 27
> n.3 <- ceiling((qt(1-alpha, n.2-1) + qt(1-beta, n.2-1))^2 / effekt^2); n.3
[1] 27

```

Für $n = 27$ Fälle kann mit dem Einstichproben-t-Test, bei einseitiger Fragestellung, der Rückgang des Blutzuckerwertes um 15 mg/dl geprüft werden, wenn man eine Standardabweichung von $\sigma = 30$ mg/dl voraussetzt und $\alpha = 0,05$ bzw. $1 - \beta = 0,80$ annimmt. Werden weniger Fälle für die Untersuchung herangezogen, dann sinkt die Wahrscheinlichkeit dafür, einen Effekt in dieser Höhe als statistisch signifikant zu erkennen (die Untersuchung hat eine geringere Power).

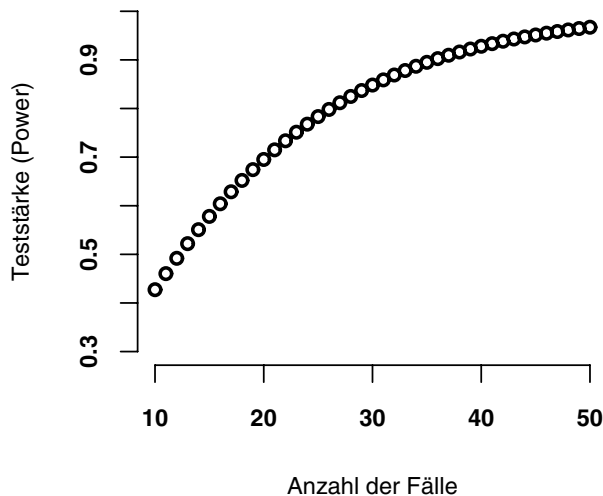


Abb. 7.13. Power für den Einstichproben-t-Test in Abhängigkeit vom Stichprobenumfang $n = 10, \dots, 50$ für einen Effekt $\delta = 15/30 = 0,5$ auf dem 5%-Signifikanzniveau

In R steht für Fallzahlberechnungen im Rahmen der t-Test-Statistiken die spezielle Funktion **power.t.test()** zur Verfügung. Mit den Bedingungen aus dem vorangehenden Beispiel folgt:

```

> power.t.test(delta=15, sd=30, sig.level=0.05, power=0.80, type="one.sample",
  alternative="one.sided")

One-sample t test power calculation

      n = 26.13751
    delta = 15
      sd = 30
sig.level = 0.05
  power = 0.8
alternative = one.sided

```

Mit Hilfe der Funktion **power.t.test()** lässt sich die Power in Abhängigkeit von der Fallzahl unter sonst gleichen Bedingungen berechnen und als Kurve im Koordinatensystem darstellen (**Powerkurve**). In Abbildung 7.13 ist die Power mit den Vorgaben aus obigem Beispiel für $n = 10, \dots, 50$ dargestellt.

7.3.3 Einstichproben-Median-Test

Der **Vorzeichen-Rangtest** für den Median nach Wilcoxon prüft die Nullhypothese, die Beobachtungen – Realisierungen unabhängiger und identisch verteilter Zufallsvariablen – entstammen einer stetigen und symmetrischen Grundgesamtheit, deren Median $\tilde{\mu}_0$ eine vorgegebene Zahl ist:

$$H_0 : \tilde{\mu} = \tilde{\mu}_0 ; \quad H_A : \tilde{\mu} \neq \tilde{\mu}_0 .$$

Aus einer Zufallsstichprobe werden die Werte n berücksichtigt, die nicht gleich $\tilde{\mu}_0$ sind. Die n Absolutbeträge $|x_i - \tilde{\mu}_0|$ werden der Größe nach aufsteigend geordnet; ihnen werden Rangzahlen zugewiesen. Bei gleichgroßen Beträgen werden mittlere Rangzahlen zugeordnet. Dann werden die Summen $\hat{R}_p$ bzw. $\hat{R}_n$ gebildet, die von positiven bzw. von negativen Werten $(x_i - \tilde{\mu}_0)$ ($i = 1, \dots, n$) stammen, und kontrolliert: $\hat{R}_p + \hat{R}_n = n(n+1)/2$.

H_0 wird auf dem 5%-Niveau abgelehnt, sobald die Summe $\hat{R}_n$ der negativen Rangzahlen oder die Summe $\hat{R}_p$ der positiven Rangzahlen außerhalb der beiden tabellierten T -Werte der Tabelle 7.17 liegt oder diese erreicht.

Tabelle 7.17. 5%-Schranken für die zweiseitige Fragestellung im Vorzeichen-Rangtest

n	T		n	T	
6	0	21	13	17	74
7	2	26	14	21	84
8	3	33	15	25	95
9	5	40	16	30	106
10	8	47	17	35	118
11	10	56	18	40	131
12	13	65	19	46	144
13	17	74	20	52	158



Hinweis: Weitere Schranken enthält Tabelle 7.29 zum Wilcoxon-Test für Paardifferenzen. Hier dient die kleinere der beiden Rangsummen, sie wird $\hat{R}$ genannt, als Prüfgröße. H_0 wird auf dem verwendeten Niveau abgelehnt, sobald $\hat{R}$ kleiner oder gleich dem kritischen Wert $R(n; \alpha)$ der Tabelle 7.29 ist. Andererseits können Schrankenwerte für ein- und zweiseitige Hypothesen in **R** auch aus den entsprechenden Quantilen der **Wilcoxon-Verteilung** ermittelt werden. Hierfür steht die Funktion **qsignrank()** zur Verfügung.

```
> ##### Quantile zur Wilcoxon-Verteilung #####
> qsignrank(0.95, 6:20, lower.tail = TRUE)
[1] 18 24 30 36 44 52 60 69 79 89 100 111 123 136 149
```

In dem Beispiel werden die 95%-Quantile, die bei einseitiger Fragestellung mit $\alpha = 0,05$ auf der rechten Seite der Verteilung verwendet werden, für $n = 6, \dots, 20$ berechnet.

Bei einseitiger Fragestellung wird nach I bzw. nach II

$$\begin{aligned} \text{(I)} \quad & H_0 : \tilde{\mu} \leq \tilde{\mu}_0 \text{ gegen } H_A : \tilde{\mu} > \tilde{\mu}_0 \\ \text{(II)} \quad & H_0 : \tilde{\mu} \geq \tilde{\mu}_0 \text{ gegen } H_A : \tilde{\mu} < \tilde{\mu}_0 \end{aligned}$$

geprüft. Für $n > 25$ kann eine **Approximation durch die Standardnormalverteilung** (7.53) verwendet werden.

$$R(n; \alpha) = \frac{n(n+1)}{4} - z_\alpha \cdot \sqrt{\frac{1}{24}n(n+1)(2n+1)} \quad (7.53)$$



Beispiel: Es liege eine Zufallsstichprobe vor. Entstemmen die Daten: 24, 12, 38, 40, 16, 26, 18, 31, geordnet: 12, 16, 18, 24, 26, 31, 38, 40 mit $\tilde{x} = 25$ einer symmetrischen Grundgesamtheit mit $\tilde{\mu}_0 = 30$ ($\alpha = 0,05$)?

x_i	Differenz $x_i - \tilde{\mu}_0$	Rangzahlen für $ x_i - \tilde{\mu}_0 $
12	-18	8
16	-14	7
18	-12	6
24	-6	3
26	-4	2
31	+1	1
38	+8	4
40	+10	5

$$\hat{R}_n = 8 + 7 + \dots + 2 = 26;$$

$$\hat{R}_p = 1 + 4 + 5 = 10; \text{Kontrolle } 26 + 10 = 36 = 8(8 + 1)/2.$$

Da 10 und 26 zwischen den T -Werten der Tabelle 7.17 ($n = 8$) 3 und 33 liegen, kann $H_0 : \tilde{\mu} = \tilde{\mu}_0$ anhand der kleinen Stichprobe auf dem 5%-Niveau nicht abgelehnt werden. Mit einer Irrtumswahrscheinlichkeit von 5% ist (1) $\tilde{x} = 25$ mit $\tilde{\mu}_0 = 30$ verträglich bzw. (2) der dem $\tilde{x}$ zugrundeliegende Parameter $\tilde{\mu}$ nicht von $\tilde{\mu}_0$ zu unterscheiden.

In **R** kann der Vorzeichen-Rangtest für den Median nach Wilcoxon mit der Funktion **wilcox.test()** gerechnet werden. Die Funktion ermittelt zusätzlich den Median mit einem entsprechenden Konfidenzintervall.

```
> x <- c(12, 16, 18, 24, 26, 31, 38, 40)
> wilcox.test(x, alternative="two.sided", mu=30, conf.int=TRUE)
```

Wilcoxon signed rank test

```
data: x
V = 10, p-value = 0.3125
alternative hypothesis: true mu is not equal to 30
95 percent confidence interval: 16.0 35.5
sample estimates: (pseudo)median 25.5
```

Hinweis: Die Funktion **wilcox.test()** berechnet exakte P-Werte, solange $n < 50$ ist und keine Bindungen durch gleiche Beobachtungswerte auftreten. Anderenfalls wird eine Approximation nach der Normalverteilung auf der Grundlage von (7.17) gerechnet. Spezielle Versionen zu den Rangtest-Verfahren, die exakte P-Werte auch für den Fall von Bindungen berechnen, stehen in **R** unter **wilcox.exact()** in einem speziellen Paket **library(exactRankTests)** (Hothorn und Hornik [HH05]) zur Verfügung.

7.3.4 Vergleich einer empirischen Varianz mit ihrem Parameter

Für normalverteilte Grundgesamtheiten gilt: Die Nullhypothese $\sigma = \sigma_0$ bzw. $\sigma^2 = \sigma_0^2$ (gegen $\sigma > \sigma_0$ bzw. $\sigma^2 > \sigma_0^2$) wird abgelehnt, sobald

$$\hat{\chi}^2 = \frac{\sum (x_i - \bar{x})^2}{\sigma_0^2} = \frac{(n-1)s^2}{\sigma_0^2} > \chi_{n-1, 1-\alpha}^2 \quad \text{wenn } \mu \text{ unbekannt ist} \quad (7.54)$$

und

$$\hat{\chi}^2 = \frac{\sum (x_i - \mu)^2}{\sigma_0^2} = \frac{ns_0^2}{\sigma_0^2} > \chi_{n, 1-\alpha}^2 \quad \text{wenn } \mu \text{ bekannt ist} \quad (7.55)$$

Wenn μ bekannt ist, dann kann die empirische Varianz in der Grundgesamtheit nach $s_0^2 = \sum (x - \mu)^2 / n$ berechnet werden.

Liegen umfangreiche Stichproben aus einer normalverteilten Grundgesamtheit vor, dann wird $H_0 : \sigma = \sigma_0$ auf dem 5%-Niveau abgelehnt und $H_A : \sigma \neq \sigma_0$ akzeptiert, sobald

$$\boxed{\frac{|s - \sigma_0|}{\sigma_0} \sqrt{2n} > 1,96} \quad \begin{array}{l} \text{zum 1\%-Niveau: ersetze} \\ 1,96 \text{ durch } 2,58 \end{array} \quad (7.56)$$



Beispiel: Sind die folgenden 8 Beobachtungen 40, 60, 60, 70, 50, 40, 50, 30 ($\bar{x} = 50$) mit der Nullhypothese $\sigma^2 = \sigma_0^2 = 60$ gegen $\sigma^2 > \sigma_0^2 = 60$ verträglich ($\alpha = 0,05$)?

$$\hat{\chi}^2 = \frac{(40 - 50)^2}{60} + \frac{(60 - 50)^2}{60} + \dots + \frac{(30 - 50)^2}{60} = 20,00$$

Da $\hat{\chi}^2 = 20,00 > 14,07 = \chi_{7;0,95}^2$ ist, muss $H_0 : \sigma^2 = \sigma_0^2$ zugunsten von $H_A : \sigma^2 > \sigma_0^2$ verworfen werden.

Für $n = 100$, $\sigma^2 = 5$ und $\sigma_0^2 = 4$ wird nach (7.54)

$$\hat{\chi}^2 = (100 - 1) \cdot 5/4 = 123,75 > 123,23 = \chi_{99;0,95}^2 \quad H_0 \text{ ebenfalls verworfen.}$$

7.3.5 Prüfung der Zufallsmäßigkeit einer Folge von Alternativdaten oder von Messwerten

7.3.5.1 Die sukzessive Differenzenstreuung

Ein einfacher **Trendtest** anhand der Dispersion zeitlich aufeinanderfolgender Stichprobenwerte $x_1, x_2, \dots, x_i, \dots, x_n$, die einer **normalverteilten Grundgesamtheit** entstammen, basiert auf der in üblicher Weise ermittelten Varianz und dem mittleren Quadrat der $n - 1$ Differenzen aufeinanderfolgender Werte, der **sukzessiven Differenzenstreuung** (engl. mean square successive difference) δ^2 (Delta-Quadrat):

$$\delta^2 = [(x_1 - x_2)^2 + (x_2 - x_3)^2 + (x_3 - x_4)^2 + \dots + (x_{n-1} - x_n)^2] / (n - 1)$$

d. h.

$$\boxed{\delta^2 = \sum (x_i - x_{i+1})^2 / (n - 1)} \quad (7.57)$$

Sind die aufeinanderfolgenden Werte unabhängig, dann gilt $\delta^2 \simeq 2s^2$ oder $\delta^2/s^2 \simeq 2$. Sobald ein Trend vorliegt, wird $\delta^2 < 2s^2$, da dann benachbarte Werte ähnlicher sind als entferntere, d. h. $\delta^2/s^2 < 2$. Die Nullhypothese, aufeinanderfolgende Werte sind unabhängig, muss zugunsten der Alternativhypothese, es besteht ein Trend, aufgegeben werden, sobald der Quotient

$$\boxed{\delta^2/s^2 = \sum (x_i - x_{i+1})^2 / \sum (x_i - \bar{x})^2} \quad (7.58)$$

die kritischen Schranken der Tabelle 7.18 erreicht oder unterschreitet.

Beispielsweise lässt sich für die Reihe: 2, 3, 5, 6 mit $\sum (x_i - \bar{x})^2 = 10$ und $\sum (x_i - x_{i+1})^2 = (2 - 3)^2 + (3 - 5)^2 + (5 - 6)^2 = 6$, d. h. $\delta^2/s^2 = 6/10 = 0,60 < 0,626$ die Nullhypothese auf dem 1%-Niveau ablehnen.

Für große Stichprobenumfänge kann man anhand der Standardnormalverteilung approximierte Schranken nach (7.59)

$$\boxed{2 - 2z \cdot \frac{1}{\sqrt{n+1}}} \quad (7.59)$$

Tabelle 7.18. Kritische Schranken für den Quotienten aus der mittleren quadratischen sukzessiven Differenzenstreuung und der Varianz (auszugsweise entnommen und mit dem Faktor $(n - 1)/n$ korrigiert aus B.I. Hart: Significance levels for the ratio of the mean square successive difference to the variance. Ann. Math. Statist. **13** (1942) 445–447)

n	0,1%	1%	5%	n	0,1%	1%	5%
4	0,5898	0,6256	0,7805	33	1,0055	1,2283	1,4434
5	0,4161	0,5379	0,8204	34	1,0180	1,2386	1,4511
6	0,3634	0,5615	0,8902	35	1,0300	1,2485	1,4585
7	0,3695	0,6140	0,9359	36	1,0416	1,2581	1,4656
8	0,4036	0,6628	0,9825	37	1,0529	1,2673	1,4726
9	0,4420	0,7088	1,0244	38	1,0639	1,2763	1,4793
10	0,4816	0,7518	1,0623	39	1,0746	1,2850	1,4858
11	0,5197	0,7915	1,0965	40	1,0850	1,2934	1,4921
12	0,5557	0,8280	1,1276	41	1,0950	1,3017	1,4982
13	0,5898	0,8618	1,1558	42	1,1048	1,3096	1,5041
14	0,6223	0,8931	1,1816	43	1,1142	1,3172	1,5098
15	0,6532	0,9221	1,2053	44	1,1233	1,3246	1,5154
16	0,6826	0,9491	1,2272	45	1,1320	1,3317	1,5206
17	0,7104	0,9743	1,2473	46	1,1404	1,3387	1,5257
18	0,7368	0,9979	1,2660	47	1,1484	1,3453	1,5305
19	0,7617	1,0199	1,2834	48	1,1561	1,3515	1,5351
20	0,7852	1,0406	1,2996	49	1,1635	1,3573	1,5395
21	0,8073	1,0601	1,3148	50	1,1705	1,3629	1,5437
22	0,8283	1,0785	1,3290	51	1,1774	1,3683	1,5477
23	0,8481	1,0958	1,3425	52	1,1843	1,3738	1,5518
24	0,8668	1,1122	1,3552	53	1,1910	1,3792	1,5557
25	0,8846	1,1278	1,3671	54	1,1976	1,3846	1,5596
26	0,9017	1,1426	1,3785	55	1,2041	1,3899	1,5634
27	0,9182	1,1567	1,3892	56	1,2104	1,3949	1,5670
28	0,9341	1,1702	1,3994	57	1,2166	1,3999	1,5707
29	0,9496	1,1830	1,4091	58	1,2227	1,4048	1,5743
30	0,9645	1,1951	1,4183	59	1,2288	1,4096	1,5779
31	0,9789	1,2067	1,4270	60	1,2349	1,4144	1,5814
32	0,9925	1,2177	1,4354	∞	2,0000	2,0000	2,0000

berechnen, wobei der Wert der Standardnormalvariablen z für die 5%-Schranke 1,645, für die 1%-Schranke 2,326 und für die 0,1%-Schranke 3,090 beträgt. Beispielsweise erhalten wir für $n = 200$ als approximierte 5%-Schranke nach (7.59)

$$2 - 2 \cdot 1,645 \cdot \frac{1}{\sqrt{200 + 1}} = 1,77.$$

7.3.5.2 Der Iterationstest für die Prüfung, ob eine Folge von Alternativdaten oder von Messwerten zufallsgemäß ist

Der Iterationstest ist wie die folgenden beiden Tests verteilungsunabhängig. Er dient zur Prüfung der Unabhängigkeit, der zufälligen Anordnung von Stichprobenwerten.

Eine **Iteration (run)** ist eine Folge identischer Symbole. So bildet die Münzwurf-Folge (W = Wappen, Z = Zahl): W, W, W, Z, W, W, Z, Z für die 8 Würfe

$$\frac{W, W, W}{1}; \frac{Z}{2}; \frac{W, W}{3}; \frac{Z, Z}{4}$$

$\hat{r} = 4$ Iterationen ($n = 8$). Iterationen erhält man nicht nur bei Alternativdaten, sondern auch bei Messwerten, die nach ihrem Medianwert in über- oder unterdurchschnittlich gruppiert werden.

Für gegebenes n weist ein kleines $\hat{r}$ auf Klumpungen ähnlicher Beobachtungen hin, ein großes $\hat{r}$ auf einen regelmäßigen Wechsel. Der Nullhypothese (H_0), die Reihenfolge ist zufällig, d. h.

es liegt eine Zufallsstichprobe vor, wird bei zweiseitiger Fragestellung die Alternativhypothese (H_A), es liegt keine Zufallsstichprobe vor, d. h. die Stichprobenwerte sind nicht unabhängig voneinander, gegenübergestellt. Bei einseitiger Fragestellung wird der H_0 entweder die H_{A1} : „**Klumpungseffekt**“ oder die H_{A2} : „**regelmäßiger Wechsel**“ gegenübergestellt. Die kritischen Schranken $r_{\text{unten}} = r_{u;\alpha/2}$ und $r_{\text{oben}} = r_{o;\alpha/2}$ für n_1 und $n_2 \leq 20$ sind der Tabelle 7.19 zu entnehmen. Abweichend von der sonst verwendeten Notation (Quantile) werden hier untere und obere kritische Werte prozentual gekennzeichnet. Für n_1 oder $n_2 > 20$ benutze man die Approximation (7.60).

$$\begin{aligned}
 &\text{Für } n_1 = n_2 = n: \\
 \hat{z} &= \frac{|\hat{r} - \mu_r|}{\sigma_r} = \frac{\left| \hat{r} - \left(\frac{2n_1n_2}{n_1 + n_2} + 1 \right) \right|}{\sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}}} \\
 &= \frac{|n(\hat{r} - 1) - 2n_1n_2|}{\sqrt{\frac{2n_1n_2(2n_1n_2 - n)}{n - 1}}} \quad (7.60) \\
 &\text{Für } n_1 = n_2 = \frac{n}{2} \text{ (d.h. } n = 2n_1 = 2n_2\text{):} \\
 \hat{z} &= \left[\left| \hat{r} - \left(\frac{n}{2} + 1 \right) \right| \right] / \sqrt{n(n-2)/[4(n-1)]}
 \end{aligned}$$

Zweiseitiger Test:

Für $r_{u;\alpha/2} < \hat{r} < r_{o;\alpha/2}$ wird H_0 beibehalten; H_0 wird abgelehnt, sobald $\hat{r} \leq r_{u;\alpha/2}$ bzw. $\hat{r} \geq r_{o;\alpha/2}$ bzw. $\hat{z} \geq z_{1-\alpha/2}$

Einseitiger Test:

H_0 wird gegen $\begin{matrix} H_{A1} \\ H_{A2} \end{matrix}$ abgelehnt, sobald $\begin{matrix} \hat{r} \leq r_{u;\alpha} \\ \hat{r} \geq r_{o;\alpha} \end{matrix}$ bzw. $\hat{z} \geq z_{1-\alpha}$.

Der Iterationstest kann auch zur Prüfung der Nullhypothese dienen, zwei Stichproben etwa gleichen Umfangs entstammen derselben Grundgesamtheit ($n_1 + n_2$ Beobachtungen der Größe nach ordnen; für kleines $\hat{r}$ wird H_0 verworfen). Dies ist der **Iterationstest von Wald und Wolfowitz** für die zweiseitige Fragestellung.



Beispiel (Prüfung von Messwerten auf **Nichtzufälligkeit** ($\alpha = 0,10$): Nacheinander erhalte man folgende 11 Beobachtungen 18, 17, 18, 19, 20, 19, 19, 21, 18, 21, 22, die größer oder gleich (G) bzw. kleiner (K) als der Median $\tilde{x} = 19$ sind. Die Folge $KKKGGGGGKGG$ ist bei $n_1 = 4$ (K), $n_2 = 7$ (G) mit $\hat{r} = 4$ auf dem 10%-Niveau (Tab. 7.19; $P = 0,10$; $r_{u;5\%} = 3$ wird nicht erreicht bzw. $3 = r_{u;5\%} < \hat{r} < r_{o;5\%} = 9$) mit der Zufälligkeitshypothese verträglich.

In **R** steht eine spezielle Funktion **runs.test()** in dem Paket für die Analyse von Zeitreihen **library(tseries)** (Trapletti [Tra05]) zur Verfügung. Dabei muss die zu untersuchende Folge den Datentyp „factor“ aufweisen. Für das Beispiel folgt:

```

> library(tseries)
> werte <- c(18, 17, 18, 19, 20, 19, 19, 21, 18, 21, 22)
> med <- median(werte)
> x <- as.factor(werte < med); x

```


(engl. under-mixing) bzw. eine besondere Regelmäßigkeit (engl. over-mixing) gegen eine zufällige Reihenfolge statistisch zu prüfen.



Beispiel (Prüfung von Beobachtungen auf **Nichtklumpungseffekt** ($\alpha = 0,05$), d. h. Prüfung von H_0 gegen H_{A1} auf dem 5%-Niveau anhand der unteren 5%-Schranken der Tab. 7.19 bzw. der Standardnormalverteilung): Anhand von Zufallsstichproben der Umfänge $n_1 = 20, n_2 = 20$ ergebe sich $\hat{r} = 15$. Da nach Tab. 7.19 $r_{u;5\%} = 15$ ist und für $\hat{r} \leq r_{u;5\%}$ H_0 abgelehnt wird, akzeptiert man die Klumpungseffekt-Hypothese auf dem 5%-Niveau. Dieses Resultat erhält man auch nach (7.60):

$$\hat{z} = [|15 - (20 + 1)|] / \sqrt{40(40 - 2) / [4(40 - 1)]} = 1,922$$

da $z_{0,95} = 1,645$ ist und H_0 für $\hat{z} \geq z_{0,95}$ abgelehnt wird.

7.3.5.3 Phasenhäufigkeitstest von Wallis und Moore

Untersucht werden die Abweichungen einer Messreihe $x_1, x_2, \dots, x_i, \dots, x_n$ ($n > 10$) von der Zufallsmäßigkeit. Die Indizes $1, 2, \dots, i, \dots, n$ bezeichnen eine *zeitliche Reihenfolge*. Ist die vorliegende Stichprobe zufälliger Art, so sollten die Vorzeichen der Differenzen $(x_{i+1} - x_i)$ ein zufälliges Bild bieten (Nullhypothese). Die Alternativhypothese wäre dann: Die Reihenfolge der Plus- und Minuszeichen weicht statistisch signifikant von der Zufallsmäßigkeit ab. Der vorliegende Test ist somit als ein **Differenzenvorzeichen-Iterationstest** aufzufassen.

Die Aufeinanderfolge gleicher Vorzeichen wird von Wallis und Moore [WM41] als „**Phase**“ bezeichnet; der Test basiert auf der Häufigkeit der Plus- und Minusphasen. Wird die Gesamtzahl der Phasen mit h bezeichnet (kleines h als Maß der Trendbeharrlichkeit), **wobei Anfangs- und Endphase weggelassen werden**, dann ist unter der Voraussetzung der Zufälligkeit einer Messreihe die Prüfgröße (7.61) für nicht zu kleines n angenähert standardnormalverteilt:

$$\hat{z} = \frac{\left| h - \frac{2n-7}{3} \right| - 0,5}{\sqrt{\frac{16n-29}{90}}} \quad \text{für } n > 10 \quad \quad \quad \hat{z} = \frac{\left| h - \frac{2n-7}{3} \right|}{\sqrt{\frac{16n-29}{90}}} \quad \text{für } n > 30 \quad (7.61)$$



Beispiel: Es liege eine aus 22 Werten bestehende Reihe von Messwerten vor:

Messwert	5	6	2	3	5	6	4	3	7	8	9	7	5	3	4	7	3	5	6	7	8	9
Vorzeichen	+	-	+	+	+	+	-	-	+	+	+	-	-	-	+	+	-	+	+	+	+	+
Nr. der Phase	1	2		3				4				5		6		7						

Für $h = 7$ wird $\hat{z} = [|7 - (2 \cdot 22 - 7)/3| - 0,5] / \sqrt{(16 \cdot 22 - 29)/90} = 4,83/1,89 = 2,56 > 1,96 = z_{0,95}$. Die Nullhypothese wird auf dem 5%-Niveau abgelehnt.

7.3.5.4 Der Vorzeichen-Trendtest von Cox und Stuart

Eine **Zeitreihe** ist eine Folge historischer Daten, sie bringt die Ausprägungen einer Veränderlichen im Zeitablauf zum Ausdruck, etwa monatliche Verkaufszahlen für Wein. Für die **Prüfung einer Zeitreihe auf Trendänderung** werden die n Werte der Reihe in drei Gruppen geteilt, so dass die erste und die letzte mit $n' = n/3$ gleich viele Messwerte enthält. Das mittlere Drittel wird bei Stichprobenumfängen n , die nicht durch 3 teilbar sind, um ein bis zwei Werte reduziert. Man vergleicht jede Beobachtung des ersten Drittels der Messreihe mit der ihr entsprechenden Beobachtung des letzten Drittels der Messreihe und markiert ein „Plus“ bei aufsteigendem Trend, ein

„Minus“ bei absteigendem Trend, also je nachdem, ob eine positive oder eine negative Differenz erscheint (Cox [CS55]).

Die Summe der Plus bzw. Minuszeichen S ist über einem Erwartungswert von $n/6$ mit einer Standardabweichung von $\sqrt{n/12}$ angenähert normalverteilt, so dass

$$\hat{z} = \frac{|S - n/6|}{\sqrt{n/12}} \quad (7.62)$$

bzw. bei kleinen Stichproben ($n < 30$) nach Yates korrigiert:

$$\hat{z} = \frac{|S - n/6| - 0,5}{\sqrt{n/12}} \quad (7.63)$$

Je nachdem, ob ein- oder zweiseitig getestet wird, gelten die Schranken $z = 1,64$ und $z = 1,96$ für $\alpha = 5\%$ bzw. $z = 2,33$ und $z = 2,58$ für $\alpha = 1\%$.

Bezeichnen wir die Anzahl der Differenzen mit n^+ , dann ist das Prüfmaß des Tests genau das gleiche wie das des Vorzeichentests mit n^+ Beobachtungen, die jeweils ungleich Null sind.

Beispiel: Wir benutzen die Werte des Beispiels aus den vorangehenden Abschnitt. Da 22 nicht durch 3 teilbar ist, bemessen wir die beiden Drittel so, als wenn $n = 24$ wäre.



Messwerte des letzten Drittels	4	7	3	5	6	7	8	9
Messwerte des ersten Drittels	5	6	2	3	5	6	4	3
Vorzeichen der Differenzen	-	+	+	+	+	+	+	+

Wir finden 7 von 8 Vorzeichen positiv. Die Prüfung auf ansteigenden Trend ergibt

$$\hat{z} = \frac{\left|7 - \frac{22}{6}\right| - 0,5}{\sqrt{11/12}} = \frac{2,83}{1,35} = 2,10$$

Einem $\hat{z} = 2,10$ entspricht bei zweiseitiger Fragestellung entsprechend der Verteilungsfunktion der Standardnormalverteilung eine Zufallswahrscheinlichkeit von $P \simeq 0,0357$. Der ansteigende Trend ist auf dem 5%-Niveau statistisch signifikant.

7.3.5.5 Variabilität der zentralen Tendenz

Beispiel: Entstammen die in der angegebenen Reihenfolge erhaltenen Werte 24, 27, 26, 28, 30, 35, 33, 37, 36, 37, 34, 32, 32, 29, 28, 28, 31, 28, 26, 25 derselben Grundgesamtheit?



Zur Beantwortung dieser Frage empfiehlt Taylor [DW58] eine andere Modifikation des Vorzeichentests zur **Erfassung der Variabilität der zentralen Tendenz** innerhalb einer Grundgesamtheit. Zunächst ermittelt man den Median der Stichprobe, dann wird ausgezählt, wie oft aufeinanderfolgende Zahlenpaare den Medianwert zwischen sich einschließen. Diesen Wert nennen wir x^* . Liegt ein Trend vor, d. h. ändert sich der Mittelwert der Grundgesamtheit, dann ist x^* klein im Verhältnis zum Stichprobenumfang n . Die Nullhypothese, das Vorliegen einer Zufallsstichprobe aus einer Grundgesamtheit ist dann auf dem 5%-Niveau abzulehnen, wenn

$$|n - 2x^* - 1| \geq 2\sqrt{n-1} \quad (7.64)$$

Der Median der obigen Stichprobe mit dem Umfang $n = 20$ ist $\tilde{x} = 29\frac{1}{2}$. An den $x^* = 4$ unterstrichenen Zahlenpaaren ändert sich der Trend. Wir erhalten $n - 2x^* - 1 = 20 - 8 - 1 = 11$

und $2\sqrt{n-1} = 2\sqrt{20-1} = 8,7$. Da $11 > 8,7$, ist anzunehmen, dass die Beobachtungen einer zeitabhängigen Grundgesamtheit entstammen.

7.3.6 Prüfung der Erwartungswerte von Poisson-Verteilungen

$$H_0 : \lambda = \lambda_0 \text{ gegen } H_A : \lambda > \lambda_0$$

Erwartet man aufgrund langer Erfahrung oder einer brauchbaren Theorie mittlere Häufigkeiten, die in einem gewählten Intervall (z. B. 1 Sekunde, 1 Jahr, 1 Quadratkilometer), etwa zwischen 1 und 100 liegen, so lassen sich anhand der Tabelle 7.20 für die einseitige Fragestellung Abweichungen von dem Erwartungswert Lambda (λ) auf dem 5%-Niveau bzw. auf dem 1%-Niveau statistisch sichern.

Tabelle 7.20. Obere 5%- und 1%-Schranken der Poisson-Verteilung für ausgewählte Erwartungswerte λ . Tabelliert sind kleinste ganze Zahlen, für die die Wahrscheinlichkeit, erreicht oder überschritten zu werden, kleiner als 0,05 bzw. kleiner als 0,01 ist; z. B. $P(X \geq 4|\lambda = 1) < 0,05$; denn $P(X \geq 4|\lambda = 1) = P(X = 4|\lambda = 1) + P(X > 4|\lambda = 1) = 0,0153 + 0,0037 = 0,019 < 0,05$

λ	Obere Schranken		λ	Obere Schranken	
	0,05	0,01		0,05	0,01
1	4	5	15	23	26
2	6	7	16	24	27
3	7	9	18	26	30
4	9	10	20	29	32
5	10	12	25	34	38
6	11	13	30	40	44
7	13	15	35	46	50
8	14	16	40	52	56
9	15	18	45	57	62
10	16	19	50	63	68
11	18	20	60	74	80
12	19	22	70	85	91
13	20	23	80	96	103
14	21	24	90	107	114
15	23	26	100	118	125



Beispiel: Man erwarte im Durchschnitt $\lambda = 10$ Ereignisse pro Intervall. Treten statt dessen 11, 12, 13, 14 oder 15 Ereignisse pro Intervall auf, so ist dies mit $H_0 : \lambda = 10$ gegen $H_A : \lambda > 10$ und $\alpha_{\text{einseitig}} = 0,05$ verträglich. Treten jedoch 16 Ereignisse pro Intervall auf, so muss H_0 auf dem 5%-Niveau zugunsten von H_A verworfen werden. Es spricht dann einiges dafür, dass Lambda größer als 10 ist.

Tabelle 7.21 gestattet die Angabe, dass bei fest vorgegebenem Wert Lambda λ einer Poisson-Verteilung mit einer Wahrscheinlichkeit von knapp $P\%$ höchstens k Ereignisse ($k \leq 10$) zu erwarten sind. So ergibt sich für $P = 5\%$ und $k = 4$ der Wert $\lambda = 9,2$, d. h. genauer: $P(X \leq 4|\lambda = 9,2) = 0,0486 < 0,05$. Damit wäre für $k = 4$ Ereignisse in einer Zufallsstichprobe aus einer nach Poisson verteilten Grundgesamtheit mit dem Erwartungswert $\lambda = 9,2$ die Nullhypothese $H_0 : \lambda = 9,2$ auf dem 5%-Niveau abzulehnen.

Für $k = 5$ Ereignisse [vgl. $P(X = 5|\lambda = 9,2) = 0,0555$] und damit $P(X \leq 5|\lambda = 9,2) = 0,0486 + 0,0555 = 0,1041 > 0,05$ hätte man H_0 auf diesem Niveau nicht ablehnen können.

Die Tabelle dient etwa zur **Kontrolle von Störfällen bzw. von nicht einwandfreien Objekten**. So lässt sich $H_0 : \lambda = 13,2$ für $k \leq 5$ auf dem 1%-Niveau ablehnen, für $k \leq 7$ auf dem 5%-Niveau.

Ist bisher mit durchschnittlich 13,2 Störfällen pro entsprechender Einheit gerechnet worden, so weisen die neuerdings festgestellten höchstens 5 (bzw. 7) Störfälle pro Einheit auf eine Prozessverbesserung hin.

Tabelle 7.21. Kumulierte Poisson-Wahrscheinlichkeiten P für k Ereignisse und Werte Lambda, die so gewählt sind, dass die links angegebenen Schranken von P gerade noch unterschritten werden

$P \backslash k$	1	2	3	4	5	6	7	8	9	10
10%	3,9	5,4	6,7	8,0	9,3	10,6	11,8	13,0	14,3	15,5
5%	4,8	6,3	7,8	9,2	10,6	11,9	13,2	14,5	15,8	17,0
1%	6,7	8,5	10,1	11,7	13,2	14,6	16,1	17,5	18,8	20,2

7.4 Zweistichprobenverfahren

- Vergleich zweier Varianzen
- Rangdispersionstest von Siegel und Tukey
- Ansari-Bradley-Test
- t-Test für unabhängige Stichproben
- t-Test für Paardifferenzen
- Wilcoxon Rangsummentest
- Wilcoxon Paardifferenzentest
- Kolmogoroff-Smirnoff-Test
- Cramér-von Mises Test
- Zweistichprobentest auf Äquivalenz

7.4.1 Vergleich zweier Varianzen (F-Test)

Ist zu untersuchen, ob zwei unabhängig gewonnene Zufallsstichproben einer **gemeinsamen normalverteilten Grundgesamtheit** entstammen, so sind zunächst ihre Varianzen (die größere Stichprobenvarianz nennen wir s_1^2) auf Gleichheit oder Homogenität zu prüfen. Die Nullhypothese (H_0): $\sigma_1^2 = \sigma_2^2$ wird abgelehnt, sobald ein aus den Stichprobenvarianzen berechneter Wert $\hat{F} = s_1^2/s_2^2$ größer ist als das zugehörige Quantil der Fisher-Verteilung F ; (vgl. Tabelle 5.12 und 5.13) dann wird die Alternativhypothese (H_A): $\sigma_1^2 \neq \sigma_2^2$ akzeptiert (zweiseitige Fragestellung). Nimmt man als Alternativhypothese an, eine der beiden Grundgesamtheiten habe eine größere Varianz als die andere, dann kann man die Stichprobe mit der nach H_A größeren Varianz als Nr. 1 mit s_1^2 und die andere als Nr. 2 mit s_2^2 bezeichnen. Für $\hat{F} > F$ wird bei dieser einseitigen Fragestellung $H_A : \sigma_1^2 > \sigma_2^2$ akzeptiert (dann sollte n_1 mindestens so groß wie n_2 sein). Wird ein Test dieser Art als Vortest einem Mittelwertvergleich (t-Test [setzt Varianzgleichheit voraus]) vorgeschaltet, dann ist das **10%-Niveau** zu bevorzugen, da der Fehler 2. Art hier der schwerwiegendere ist.

Vortests sind umstritten, da unbekannt ist, wie durch die Überprüfung der Voraussetzungen die Irrtumswahrscheinlichkeit des Haupttests verfälscht wird; außerdem bleibt der Fehler 2. Art unbekannt. Ein Vortest ist nur dann sinnvoll, wenn (1) der Haupttest hinreichend robust ist gegenüber Abweichungen von den Voraussetzungen und (2) die Stichprobenumfänge hinreichend groß sind.



Hinweis: Im Gegensatz zum zweiseitigen t -Test ist der F -Test sehr empfindlich gegenüber Abweichungen von der Normalverteilung. Man ersetze dann den F -Test durch den verteilungsunabhängigen Siegel-Tukey-Test [7.4.2].

7.4.1.1 Varianzvergleich bei kleinem bis mittlerem Stichprobenumfang

Wir bilden den Quotienten der beiden Varianzen s_1^2 und s_2^2 und erhalten als Prüfgröße

$$\hat{F} = \frac{s_1^2}{s_2^2} \quad \begin{array}{l} \text{mit } FG_1 = n_1 - 1 = \nu_1 \\ \text{mit } FG_2 = n_2 - 1 = \nu_2 \end{array} \quad (7.65)$$

Überschreitet der errechnete $\hat{F}$ -Wert den für die vorgewählte Irrtumswahrscheinlichkeit α und die Freiheitsgrade $\nu_1 = n_1 - 1$ und $\nu_2 = n_2 - 1$ tabellierten F -Wert, dann wird die Hypothese der Varianzhomogenität verworfen. Für $\hat{F} \leq F$ besteht keine Veranlassung, an dieser Hypothese zu zweifeln. Wird die Nullhypothese verworfen, dann berechne man das Konfidenzintervall (KI) für σ_1^2/σ_2^2 nach

$$\frac{s_1^2}{s_2^2} \cdot \frac{1}{F_{\nu_1, \nu_2}} \leq \frac{\sigma_1^2}{\sigma_2^2} \leq \frac{s_1^2}{s_2^2} \cdot F_{\nu_2, \nu_1} \quad \begin{array}{l} \nu_1 = n_1 - 1 \\ \nu_2 = n_2 - 1 \end{array} \quad (7.66)$$

Für den 90%-KI nehme man Tabelle 5.12, für den 95%-KI Tabelle 5.13. Die Tabellen enthalten die oberen Signifikanzschranken der F -Verteilung für die in der Varianzanalyse übliche einseitige Fragestellung. Im vorliegenden Fall sind wir im allgemeinen an Abweichungen in beiden Richtungen, also an einem zweiseitigen Test, interessiert.



Beispiel: Prüfe $H_0 : \sigma_1^2 = \sigma_2^2$ gegen $H_A : \sigma_1^2 \neq \sigma_2^2$ auf dem 10%-Niveau.

$$\begin{array}{l} \text{Gegeben: } n_1 = 41 \quad s_1^2 = 25 \\ \quad \quad \quad n_2 = 31 \quad s_2^2 = 16 \end{array} \quad \hat{F} = \frac{25}{16} = 1,56$$

Da $\hat{F} = 1,56 < 1,79 [=F_{40;30;0,95}]$, lässt sich H_0 auf dem 10%-Niveau nicht ablehnen.

In **R** kann der Varianzvergleich elementar nach (7.65) gerechnet oder mit der Funktion **var.test()** auf der Grundlage von Werten aus zwei Stichproben durchgeführt werden. Für die Zahlen aus dem obigen Beispiel folgt:

```
> n1 <- 41; sq1 <- 25;
> n2 <- 31; sq2 <- 16;
> f.hat <- sq1 / sq2;          f.hat
[1] 1.5625
> f.tab <- qf(0.95, n1-1, n2-1); f.tab
[1] 1.79179
```

Die Verwendung der Funktion **var.test()** in **R** soll an einem kleinen Zahlenbeispiel gezeigt werden:

```
> x <- round(rnorm(10, mean=90, sd=10)); x
[1] 88 105 83 94 90 91 94 90 86 90
> y <- round(rnorm(15, mean=90, sd=15)); y
[1] 109 108 97 89 64 80 77 81 81 96 67 85 104 92
> var.test(x, y, ratio=1, alternative="two.sided", conf.level=0.95)
```

F test to compare two variances

```
data: x and y F = 0.1876, num df = 9, denom df = 14, p-value =
0.01605 alternative hypothesis: true ratio of variances is not
equal to 1 95 percent confidence interval:      0.05844417
0.71236251 sample estimates: ratio of variances      0.1875649
```

Für gleichgroße Stichprobenumfänge n lässt sich H_0 auch nach

$$\hat{t} = \frac{\sqrt{n-1}(s_1^2 - s_2^2)}{2\sqrt{s_1^2 s_2^2}} \quad \text{mit} \quad \nu = n - 1 \quad (7.67)$$

prüfen (Cacoullos [Cac65]). Ein Schnelltest wird in [7.4.1.5] vorgestellt.

Beispiel: Prüfe $H_0 : \sigma_1^2 = \sigma_2^2$ gegen $H_A : \sigma_1^2 \neq \sigma_2^2$ auf dem 10%-Niveau.

Gegeben: $n_1 = n_2 = 20 = n$, $s_1^2 = 8$, $s_2^2 = 3$

$$\hat{F} = \frac{8}{3} = 2,67 > 2,12 \quad \hat{t} = \frac{\sqrt{20-1}(8-3)}{2\sqrt{8 \cdot 3}} = 2,22 > 1,729$$

Da H_0 auf dem 10%-Niveau abgelehnt wird, geben wir nach (7.66) das 90%-KI an:

$$F_{19;19;0,95} = 2,17 \quad \frac{2,67}{2,17} = 1,23 \quad 2,67 \cdot 2,17 = 5,79$$

$$90\text{-KI: } 1,23 \leq \sigma_1^2 / \sigma_2^2 \leq 5,79$$

Hinweis: Da das Ergebnis des F -Tests auch durch kleine Abweichungen von der Normalverteilung stark beeinflusst werden kann, ist ein approximatives nichtparametrisches Verfahren vorzuziehen: Man bildet in den einzelnen zu vergleichenden Messreihen jeweils die absoluten Werte $|x_i - \tilde{x}|$ und führt mit ihnen einen **Rangsummentest** durch: Bei zwei Stichproben den Wilcoxon-Rangsummentest (vgl. [7.4.6]) und bei mehr als 2 Stichproben den H -Test von Kruskal und Wallis (vgl. [7.5.4]) und prüft, ob die absoluten Abweichungen $|x_i - \tilde{x}|$ für die einzelnen Reihen als Stichproben aus Verteilungen mit gleichem Median aufgefasst werden können. Die Homogenität mehrerer (k) Varianzen lässt sich für $n_i \geq 10$ nach **Levene** [Lev60] in der **Brown-Forsythe-Version** [BF74] auch mit Hilfe der einfachen Varianzanalyse ablehnen, sobald für die insgesamt n absoluten Abweichungen der Beobachtungen von ihren k Medianwerten $\hat{F} > F_{k-1;n-k;\alpha}$ ist.

Robuster Test auf Varianzheterogenität nach Levene in der Brown-Forsythe-Version: Für k unabhängige Stichproben mit jeweils mindestens 10 Beobachtungen wird die Nullhypothese: gleiche Varianzen [$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$] geprüft. Die Alternativhypothese lautet: mindestens zwei Varianzen sind ungleich [$H_A : \sigma_i^2 \neq \sigma_j^2$]. H_0 wird auf dem $100\alpha\%$ -Niveau abgelehnt, sobald für die nach $y_{ij} = |x_{ij} - \tilde{x}_i|$ transformierten Beobachtungen ($\tilde{x}_i$ ist der Median der i -ten Stichproben), also für die jetzt vorliegenden y_{ij} -Werte, das $\hat{F}$ der Varianzanalyse größer ist als $F_{k-1;n-k;\alpha}$ (zur Berechnung von $\hat{F}$ vgl. den Abschnitt in der Varianzanalyse).

7.4.1.2 Varianzvergleich bei mittlerem bis großem Stichprobenumfang

Für nicht tabelliert vorliegende F -Werte – bei mittleren Freiheitsgraden kann man interpolieren – wird bei größeren Freiheitsgraden die Homogenität zweier Varianzen mit Hilfe des Ausdrucks (7.68) getestet, der approximativ standardnormalverteilt ist.

$$\hat{z} = \frac{\frac{1}{2} \ln F + \frac{1}{2} \left(\frac{1}{\nu_1} - \frac{1}{\nu_2} \right)}{\sqrt{\frac{1}{2} \left(\frac{1}{\nu_1} + \frac{1}{\nu_2} \right)}} = \frac{1,15129 \cdot \log F + \frac{1}{2} \left(\frac{1}{\nu_1} - \frac{1}{\nu_2} \right)}{\sqrt{\frac{1}{2} \left(\frac{1}{\nu_1} + \frac{1}{\nu_2} \right)}} \quad (7.68)$$



Beispiel: Wir wollen diese Formel anhand der Quantile der Fisher-Verteilung kontrollieren. Für $\nu_1 = \nu_2 = 60$ erhalten wir bei einer Irrtumswahrscheinlichkeit von $\alpha = 0,05$ den Wert $F = 1,53$. Nehmen wir nun an, wir hätten diesen F -Wert experimentell für $\nu_1 = \nu_2 = 60$ gefunden. Ist der gefundene F -Wert bei einseitiger Fragestellung ($\sigma_1^2 = \sigma_2^2$ gegen $\sigma_1 > \sigma_2^2$) auf dem 5%-Niveau statistisch signifikant? Für $F = 1,53$, $\nu_1 = 60$ und $\nu_2 = 60$ erhalten wir

$$\hat{z} = \frac{1,15129 \cdot \log 1,53 + \frac{1}{2} \left(\frac{1}{60} - \frac{1}{60} \right)}{\sqrt{\frac{1}{2} \left(\frac{1}{60} + \frac{1}{60} \right)}} = 1,647,$$

d. h. $\hat{z} = 1,647 > 1,645$; das einer Irrtumswahrscheinlichkeit von $P = 0,05$ entsprechende Quantil der Standardnormalverteilung $z_{0,95} = 1,6449$ wird überschritten, damit muss die Hypothese der Varianzhomogenität auf dem 5%-Niveau abgelehnt werden. Die Approximation durch die Normalverteilung ist ausgezeichnet.

7.4.1.3 Varianzvergleich bei großem bis sehr großem Stichprobenumfang ($n_1, n_2 \gtrsim 100$)

$$\hat{z} = (|s_1 - s_2|) / \sqrt{[s_1^2/(2n_1)] + [s_2^2/(2n_2)]} \quad (7.69)$$

Für $\hat{z} > z_{1-\alpha/2}$ wird $H_0 : \sigma_1^2 = \sigma_2^2$ (bzw. $\sigma_1 = \sigma_2$) auf dem $100\alpha\%$ -Niveau abgelehnt; beide gelten dann auf dem betreffenden Niveau als statistisch signifikant verschieden, d. h. als heterogen, im anderen Falle gelten sie als gleich oder homogen.



Beispiel: Gegeben seien $s_1 = 14$ $s_2 = 12$ $n_1 = n_2 = 500$;

Nullhypothese: $\sigma_1^2 = \sigma_2^2$; Alternativhypothese: $\sigma_1^2 \neq \sigma_2^2$; $\alpha = 0,05$;

$\hat{z} = (14 - 12) / \sqrt{[14^2/(2 \cdot 500)] + [12^2/(2 \cdot 500)]} = 3,430 > 1,960$; d. h.

auf dem 5%-Niveau wird $H_0 : \sigma_1^2 = \sigma_2^2$ abgelehnt und $H_A : \sigma_1^2 \neq \sigma_2^2$ akzeptiert.

7.4.1.4 Minimale Stichprobenumfänge für den F-Test

Bei jedem statistischen Test sind, wie wir wissen, α und β abzuschätzen. Tabelle 7.22 gibt einen Hinweis auf die Anzahl der Beobachtungswerte, die für den Vergleich zweier Varianzen mit dem F -Test benötigt werden. Tabelliert sind F -Werte: Man erhält z. B. für $\alpha = 0,05$, $\beta = 0,01$ und $s_{\text{Zähler}}^2/s_{\text{Nenner}}^2 = F = 4$ aus der Tafel den Hinweis, dass die Schätzung der Varianzen in beiden Stichproben auf 30 bis 40 Freiheitsgraden – entsprechend den F -Werten 4,392 und 3,579 – sagen wir, auf mindestens 35 Freiheitsgraden beruhen sollte.

7.4.1.5 Vergleich der Streuung zweier kleiner Stichproben nach Pillai und Buenaventura

Die Streuungen zweier unabhängiger Messreihen können auch über die Spannweiten (R_1, R_2) verglichen werden. Man bildet zu diesem Zweck analog dem F -Test das Verhältnis R_1/R_2 wobei $R_1 > R_2$ anzunehmen ist, und prüft, ob der Quotient R_1/R_2 die entsprechende Schranke der, Tabelle 7.23 erreicht oder überschreitet.

Wenn beispielsweise die Messreihe A mit $n_1 = 9$ und die Messreihe B mit $n_2 = 10$ die Spannweiten $R_1 = 19$ und $R_2 = 10$ aufweisen, dann ist $R_1/R_2 = 1,9$ größer als der für $\alpha = 5\%$ tabellierte Wert 1,82. Damit wird die Nullhypothese abgelehnt. Die Schranken der Tabelle 7.23 sind wie der Test für die einseitige Fragestellung eingerichtet.

Wird nach $\sigma_1^2 = \sigma_2^2$ gegenüber $\sigma_1^2 \neq \sigma_2^2$ geprüft, dann sind die 5%- und 1%-Schranken dieser Tabelle als 10%- und 2%-Niveaus des zweiseitigen Tests aufzufassen. Für kleine Stichproben ist der Test hinreichend effizient.

Tabelle 7.22. Minimale Stichprobenumfänge für den F-Test. (Auszugsweise aus Davies, O.L.: The Design and Analysis of Industrial Experiments, Oliver and Boyd, London 1956, p. 614, part of table H)

FG	$\alpha = 0,05$			
	$\beta = 0,01$	$\beta = 0,05$	$\beta = 0,1$	$\beta = 0,5$
1	654200	26070	6436	161,5
2	1881	361,0	171,0	19,00
3	273,3	86,06	50,01	9,277
4	102,1	40,81	26,24	6,388
5	55,39	25,51	17,44	5,050
6	36,27	18,35	13,09	4,284
7	26,48	14,34	10,55	3,787
8	20,73	11,82	8,902	3,438
9	17,01	10,11	7,757	3,179
10	14,44	8,870	6,917	2,978
12	11,16	7,218	5,769	2,687
15	8,466	5,777	4,740	2,404
20	6,240	4,512	3,810	2,124
24	5,275	3,935	3,376	1,984
30	4,392	3,389	2,957	1,841
40	3,579	2,866	2,549	1,693
60	2,817	2,354	2,141	1,534
120	2,072	1,828	1,710	1,352
∞	1,000	1,000	1,000	1,000

7.4.2 Rangdispersionstest von Siegel und Tukey

Der F -Test ist empfindlich gegenüber Abweichungen von dem Modell der Normalverteilung. Ist diese Annahme nicht gerechtfertigt, dann sollte stets ein robustes Rangtestverfahren für den Test unterschiedlicher Variabilität in den Stichproben eingesetzt werden. Dazu werden in den folgenden Abschnitten zwei Verfahren näher vorgestellt. Während der Siegel-Tukey-Test die Werte der geordneten kombinierten Stichprobenwerte, beginnend am unteren und oberen Ende zur Mitte hin

Tabelle 7.23. Obere Signifikanzschranken der auf den Spannweiten basierenden F' -Verteilung (aus Pillai, K.C.S. und A.R. Buenaventura: Upper percentage points of a substitute F -ratio using ranges, Biometrika **48** (1961) 195 and 196)

		oben: $\alpha = 0,05$ unten: $\alpha = 0,01$								
$n_2 \backslash n_1$		2	3	4	5	6	7	8	9	10
2	12,71	19,08	23,2	26,2	28,6	30,5	32,1	33,5	34,7	
3	3,19	4,37	5,13	5,72	6,16	6,53	6,85	7,12	7,33	
4	2,03	2,66	3,08	3,38	3,62	3,84	4,00	4,14	4,26	
5	1,60	2,05	2,35	2,57	2,75	2,89	3,00	3,11	3,19	
6	1,38	1,74	1,99	2,17	2,31	2,42	2,52	2,61	2,69	
7	1,24	1,57	1,77	1,92	2,04	2,13	2,21	2,28	2,34	
8	1,15	1,43	1,61	1,75	1,86	1,94	2,01	2,08	2,13	
9	1,09	1,33	1,49	1,62	1,72	1,79	1,86	1,92	1,96	
10	1,05	1,26	1,42	1,54	1,63	1,69	1,76	1,82	1,85	
2	63,66	95,49	116,1	131	143	153	161	168	174	
3	7,37	10,00	11,64	12,97	13,96	14,79	15,52	16,13	16,60	
4	3,73	4,79	5,50	6,01	6,44	6,80	7,09	7,31	7,51	
5	2,66	3,33	3,75	4,09	4,36	4,57	4,73	4,89	5,00	
6	2,17	2,66	2,98	3,23	3,42	3,58	3,71	3,81	3,88	
7	1,89	2,29	2,57	2,75	2,90	3,03	3,13	3,24	3,33	
8	1,70	2,05	2,27	2,44	2,55	2,67	2,76	2,84	2,91	
9	1,57	1,89	2,07	2,22	2,32	2,43	2,50	2,56	2,63	
10	1,47	1,77	1,92	2,06	2,16	2,26	2,33	2,38	2,44	

fortlaufend mit Rangzahlen erfasst, bewertet der Ansari-Bradley-Test die absoluten Abweichungen von einer mittleren Rangzahl aus allen Stichprobenwerten. Der Aufwand zur Berechnung ist in beiden Verfahren recht hoch. Für den Ansari-Bradley-Test steht in **R** die Funktion **ansari.test()** zur Verfügung.

Siegel und Tukey [ST60] haben ein verteilungsfreies Verfahren entwickelt, das auf dem Wilcoxon-Test basiert. Es gestattet die Prüfung der Nullhypothese, dass zwei unabhängige Stichproben hinsichtlich ihrer Variabilität, Streuung oder Dispersion einer gemeinsamen Grundgesamtheit angehören gegen die Alternativhypothese: Beide Stichproben entstammen keiner gemeinsamen Grundgesamtheit.

Mit zunehmendem Unterschied zwischen den Mittelwerten der Grundgesamtheiten wird allerdings die Wahrscheinlichkeit kleiner, dass die Nullhypothese beim Vorliegen echter Variabilitätsunterschiede abgelehnt wird, d. h. je größer der Mittelwertunterschied, desto größer auch die Wahrscheinlichkeit, einen Fehler zweiter Art zu begehen. Dieses gilt insbesondere dann, wenn die Dispersionen klein sind. Wenn die Grundgesamtheiten sich nicht überdecken, ist die Power gleich Null. Diesen Test, der also beim Vorliegen fast gleicher Lokalisations-Parameter gegenüber Variabilitätsunterschieden sehr empfindlich ist, hat Meyer-Bahlburg [MB70] auf k Stichproben verallgemeinert.

Zur Anwendung des Tests werden die vereinigten Stichproben ($n_1 + n_2 = n$ mit $n_1 \leq n_2$) in eine gemeinsame aufsteigende Rangordnung ($x_{(i)}, i = 1, \dots, n$) gebracht und den extremen Beobachtungswerten niedrige, den zentralen Beobachtungen hohe Rangwerte zugeteilt: Der kleinste Wert erhält den Rang 1, die beiden größten Werte bekommen die Ränge 2 und 3, 4 und 5 erhalten die nächst kleinsten Werte, 6 und 7 die nächst größten usw. Liegt eine ungerade Anzahl von Beobachtungen vor, so erhält die mittelste Beobachtung keinen Rang, damit der höchste Rang jeweils eine gerade Zahl ist. Für jede Stichprobe wird die Summe der Rangzahlen (R_1, R_2) ermittelt. Für $n_1 = n_2$ gilt unter der Nullhypothese (H_0) : $R_1 \approx R_2$; je stärker sich beide Stichproben in ihrer Variabilität unterscheiden, desto unterschiedlicher dürften die Rangsummen sein.

$$\begin{aligned}
 S_T = R_1 &= \sum_{i=1}^n g(i)V_i \quad \text{mit} \\
 V_i &= \begin{cases} 1 & \text{für } x_{(i)} \text{ in Stichprobe 1} \\ 0 & \text{für } x_{(i)} \text{ in Stichprobe 2} \end{cases} \\
 g(i) &= \begin{cases} 2i & \text{für } i \text{ gerade und } 1 < i \leq n/2 \\ 2(n-i) + 2 & \text{für } i \text{ gerade und } n/2 < i \leq n \\ 2i-1 & \text{für } i \text{ ungerade und } 1 \leq i \leq n/2 \\ 2(n-i) + 1 & \text{für } i \text{ ungerade und } n/2 < i < n \end{cases}
 \end{aligned} \tag{7.70}$$

Die formale Darstellung der Siegel-Tukey Teststatistik (für R_1) in (7.70) gibt die Verteilung der Rangzahlen ohne Berücksichtigung von Bindungen an. Diese müssen durch gemittelte Rangzahlen ausgeglichen werden, insbesondere dann, wenn Bindungen häufiger zwischen den beiden Stichproben auftreten. Als Kontrolle für die Rangsummen dient (7.71).

$$R_1 + R_2 = (n_1 + n_2)(n_1 + n_2 + 1)/2 \tag{7.71}$$

Die Verteilung der Teststatistik S_T ist unter der Nullhypothese gleich der Verteilung der Wilcoxon-Statistik aus Abschnitt 7.4.6. Kritische Werte für eine Testentscheidung können somit direkt aus der Wilcoxon-Verteilung abgeleitet werden. Für kleine Stichprobenumfänge ($n_1 \leq n_2 \leq 20$) sind einige exakte kritische Werte für R_1 (Summe der Ränge der kleineren Stichprobe) in Tabelle 7.24 aufgeführt:

Tabelle 7.24. Kritische Werte für R_1 : Siegel-Tukey-Test ($\alpha = 0,05$ zweiseitig bzw. $\alpha = 0,025$ einseitig)

n_1	4	5	6	7	8	9	10
$n_2 = n_1$	10–26	17–38	26–52	36–69	49–87	62–109	78–132
$n_2 = n_1 + 1$	11–29	18–42	27–57	38–74	51–93	65–115	81–139
$n_2 = n_1 + 2$	12–32	20–45	29–61	40–79	53–99	68–121	84–146
$n_2 = n_1 + 3$	13–35	21–49	31–65	42–84	55–105	71–127	88–152
$n_2 = n_1 + 4$	14–38	22–53	32–70	44–89	58–110	73–134	91–159
$n_2 = n_1 + 5$	14–42	23–57	34–74	46–94	60–116	76–140	94–166

H_0 wird abgelehnt, wenn R_1 für $n_1 \leq n_2$ die Schranken unterschreitet, überschreitet oder erreicht.

Für nicht zu kleine Stichprobenumfänge ($n_1 > 9; n_2 > 9$ bzw. $n_1 > 2; n_2 > 20$) lässt sich der Dispersionsunterschied mit ausreichender Genauigkeit anhand der Standardnormalvariablen beurteilen:

$$\hat{z} = \frac{2R_1 - n_1(n_1 + n_2 + 1) + 1}{\sqrt{n_1(n_1 + n_2 + 1)(n_2/3)}} \quad (7.72)$$

Wenn $2R_1 > n_1(n_1 + n_2 + 1)$, dann ersetze man in (7.72) oben) das letzte $+1$ durch -1 .

Sehr unterschiedliche Stichprobenumfänge: Beim Vorliegen sehr unterschiedlicher Stichprobenumfänge ist (7.72) zu ungenau. Man benutze die Korrektur (7.73)

$$\hat{z}_{\text{kor}} = \hat{z} + \left(\frac{1}{10n_1} - \frac{1}{10n_2} \right) \cdot (\hat{z}^3 - 3\hat{z}) \quad (7.73)$$

Viele gleichgroße Werte: Sind mehr als ein Fünftel der Beobachtungen in Gleichheiten oder Bindungen (ties) verwickelt – Bindungen innerhalb einer Stichprobe stören nicht –, so ist der Nenner der Prüfgröße (7.72) durch

$$\sqrt{n_1(n_1 + n_2 + 1)(n_2/3) - 4[n_1n_2/(n_1 + n_2)(n_1 + n_2 - 1)](S_1 - S_2)} \quad (7.74)$$

zu ersetzen. Hierbei ist S_1 die **Summe der Quadrate der Ränge gebundener Beobachtungen** und S_2 ist die **Summe der Quadrate der mittleren Ränge gebundener Beobachtungen**. Für die Folge 9,7; 9,7; 9,7; 9,7 erhalten wir beispielsweise wie üblich die Ränge 1, 2, 3, 4 oder, wenn wir mittlere Rangwerte verteilen, 2,5; 2,5; 2,5; 2,5 (vgl. $1 + 2 + 3 + 4 = 2,5 + 2,5 + 2,5 + 2,5$); entsprechend liefert die Folge 9,7; 9,7; 9,7 die Ränge 1, 2, 3 und die mittleren Ränge 2, 2, 2.

Beispiel: Gegeben: die beiden Stichproben A und B

A	10,1	7,3	12,6	2,4	6,1	8,5	8,8	9,4	10,1	9,8
B	15,3	3,6	16,5	2,9	3,3	4,2	4,9	7,3	11,7	13,1



Prüfe mögliche Dispersionsunterschiede auf dem 5%-Niveau. Da unklar ist, ob die Stichproben einer normalverteilten Grundgesamtheit entstammen, wenden wir den Siegel-Tukey-Test an. Wir ordnen die Werte und bringen sie in eine gemeinsame Rangordnung:

A	2,4	6,1	7,3	8,5	8,8	9,4	9,8	10,1	10,1	12,6
B	2,9	3,3	3,6	4,2	4,9	7,3	11,7	13,1	15,3	16,5

Wert	2,4	2,9	3,3	3,6	4,2	4,9	6,1	7,3	7,3	8,5	8,8	9,4	9,8	10,1	10,1	11,7	12,6	13,1	15,3	16,5
Stichpr.	A	B	B	B	B	B	A	A	B	A	A	A	A	A	A	B	A	B	B	B
Rang	1	4	5	8	9	12	13	16	17	20	19	18	15	14	11	10	7	6	3	2

Nach der Ermittlung der Rangsummen:

$$R_A = 1 + 13 + 16 + 20 + 19 + 18 + 15 + 14 + 11 + 7 = 134$$

$$R_B = 4 + 5 + 8 + 9 + 12 + 17 + 10 + 6 + 3 + 2 = 76$$

und ihrer Kontrolle: $134 + 76 = 210 = (10 + 10)(10 + 10 + 1)/2$ ergibt sich mit

$$n(n_1 + n_2 + 1) = 10(10 + 10 + 1) = 210:$$

$$(1) \quad 2 \cdot 134 = 268 > 210, \text{ d. h. „-1“ in (7.72)}$$

$$\sqrt{n} \text{ in (7.72): } \sqrt{10(10 + 10 + 1)(10/3)} = \sqrt{700} = 26,4575$$

$$\hat{z} = [2 \cdot 134 - 210 - 1]/26,4575 = 2,154 \text{ bzw.}$$

$$(2) \quad 2 \cdot 76 = 152 < 210, \text{ d. h. (7.72) und somit}$$

$$\hat{z} = [2 \cdot 76 - 210 + 1]/26,4575 = -2,154.$$

Einem $|\hat{z}| = 2,154$ entspricht eine Zufallswahrscheinlichkeit von $P \simeq 0,0156$. Für die zweiseitige Fragestellung erhalten wir mit $P \approx 0,03$ einen auf dem 5%-Niveau signifikanten Variabilitätsunterschied (vgl. auch Tab. 7.24: $n_1 = n_2 = 10$; $76 < 78$ und $134 > 132$): Anhand der vorliegenden Stichproben lässt sich auf dem 5%-Niveau ein Dispersionsunterschied der Grundgesamtheiten sichern.

Obwohl nur 10% der Beobachtungen in Bindungen zwischen den Stichproben verwickelt sind (7,3; 7,3; die Bindung 10,1; 10,1 stört nicht, da sie innerhalb der Stichprobe A auftritt), sei der Gebrauch der „langen Wurzel“ (7.74) demonstriert: Unter Beachtung aller Bindungen ergibt sich über

$$S_1 = 11^2 + 14^2 + 16^2 + 17^2 = 862$$

$$S_2 = 12,5^2 + 12,5^2 + 16,5^2 + 16,5^2 = 857$$

$$\text{und } \sqrt{10(10 + 10 + 1)(10/3) - 4[10 \cdot 10/(10 + 10)(10 + 10 - 1)](862 - 857)}$$

$$= \sqrt{700 - 100/19} = \sqrt{694,74} = 26,36$$

$$\hat{z} = -\frac{57}{26,36} = -2,162 \text{ ein gegenüber } \hat{z} = -2,154 \text{ minimal erhöhter } |\hat{z}|$$

Wert; denn $P(Z > 2,162) = 0,0153$, d.h. $P \approx 0,03$.

Differieren $\tilde{\mu}_A$ und $\tilde{\mu}_B$ stärker ($\tilde{\mu}_A > \tilde{\mu}_B$) dann ist es zweckmäßig, vor dem Test von allen Beobachtungen A den Wert $k = \tilde{\mu}_A - \tilde{\mu}_B$ abzuziehen.

Für die Berechnung der Teststatistik zum Siegel-Tukey Test gibt es in **R** keine spezielle Funktion. Dafür soll eine direkte Umsetzung der Berechnung in **R** skizziert und für die Daten aus dem Beispiel eingesetzt werden.

```
> siegel.tukey <- function(x, y) { # Funktion zum Siegel-Tukey Test
+ n1 <- length(x); n2 <- length(y); n <- n1+n2
+ x <- c(x, y); v <- c(rep(1, n1), rep(0, n2))
+ d <- rbind(x,v)[,order(x)] # n ungerade ?
+ if (n%%2==1) {d <- d[,c(1:trunc(n/2),(trunc(n/2)+2):n)]; n <- n - 1}
+ g <- rep(NA, n)
+ for (i in 1:n) { # Aufbau der Rangverteilung
+ if (i%%2==0 & i < n & i <= n/2) g[i] <- 2*i
+ if (i%%2==0 & n/2 < i & i <= n) g[i] <- 2*(n-i) + 2
+ if (i%%2==1 & 1 <= i & i <= n/2) g[i] <- 2*i - 1
+ if (i%%2==1 & n/2 < i & i < n) g[i] <- 2*(n-i) + 1
+ }
+ ST <- sum(g*d[2,]); ST # Berechnung der Teststatistik
+ }
>
> A <- c(10.1, 7.3, 12.6, 2.4, 6.1, 8.5, 8.8, 9.4, 10.1, 9.8)
```



```
> B <- c(15.3, 3.6, 16.5, 2.9, 3.3, 4.2, 4.9, 7.3, 11.7, 13.1)
> n1 <- length(A); n2 <- length(B)
> S <- siegel.tukey(A, B); S
[1] 134
>
# Standardnormalverteilung
> z.hat <- (2*S - n1*(n1+n2+1)-1)/sqrt(n1*(n1+n2+1)*(n2/3)); z.hat
[1] 2.154397
> pnorm(z.hat, lower.tail=FALSE)
[1] 0.01560451
```

In der Funktion **siegel.tukey()** werden zunächst die beiden Stichproben verbunden, indiziert und anschließend geordnet. Falls die Zahl aller Beobachtungen N ungerade ist, wird die mittlere Beobachtung gestrichen und anschließend die Rangaufteilung nach (7.70) bestimmt. Die berechnete Teststatistik entspricht dem R_A aus obigem Beispiel und kann entsprechend (7.72) mit der Standardnormalverteilung bewertet werden (P-Wert = 0,0156).

7.4.3 Ansari-Bradley-Test

Ein weiterer robuster, verteilungsfreier Test zur Überprüfung von Variabilitätsunterschieden zwischen zwei unabhängigen Stichproben ist der Test von Ansari und Bradley [AB60]. Die Werte aus den beiden Stichproben X und Y werden zusammengelegt $Z = X \cup Y = \{x_1, \dots, x_m, y_1, \dots, y_n\}$ und aufsteigend geordnet. Mit Hilfe einer Indikatorfunktion V_i (s.a. (7.70) im vorangehenden Abschnitt), die den Wert 1 annimmt für ein $z_i \in X$ und den Wert 0 für ein $z_i \in Y$ mit $i = 1, \dots, N = m + n$, kann die Teststatistik für den Ansari-Bradley-Test durch (7.75) beschrieben werden.

$$A_N = \sum_{i=1}^N \left(\frac{N+1}{2} - \left| i - \frac{N+1}{2} \right| \right) V_i \quad (7.75)$$

Der Test basiert somit auf einer Summe von Absolutbeträgen der Abweichungen vom Mittelwert $(N+1)/2$. Der **kleinsten und größten** Beobachtung aus Z wird durch (7.75) die Rangzahl 1, der **zweitkleinsten und zweitgrößten** die Rangzahl 2 zugewiesen usw. Je kleiner A_N ist, desto größer ist die Streuung der Werte zwischen den beiden Stichproben. Der Erwartungswert und die Varianz für A_N kann durch (7.76) angegeben werden.

$$\begin{aligned} \mu_{A_N} = E[A_N] &= \begin{cases} \frac{1}{4}m(N+2) & N \text{ ist gerade} \\ \frac{1}{4}m(N+1)^2/N & N \text{ ist ungerade} \end{cases} \\ \sigma_{A_N}^2 = Var[A_N] &= \begin{cases} mn(N^2-4)/\{48(N-1)\} & N \text{ ist gerade} \\ mn(N+1)(N^2+3)/(48N^2) & N \text{ ist ungerade} \end{cases} \end{aligned} \quad (7.76)$$

Quantile der Verteilung von A_N unter der Nullhypothese (kein Unterschied in der Streuung) sind in [AB60] für $m+n \leq 20$ tabelliert. Für größeres N kann die Standardnormalverteilung für die Prüfung der Hypothesen verwendet werden.

$$\hat{z} = \frac{A_N - \mu_{A_N}}{\sqrt{\sigma_{A_N}^2}}$$

In **R** kann der Ansari-Bradley-Test mit der Funktion **ansari.test()** berechnet werden.

```
> A <- c(10.1, 7.3, 12.6, 2.4, 6.1, 8.5, 8.8, 9.4, 10.1, 9.8)
> B <- c(15.3, 3.6, 16.5, 2.9, 3.3, 4.2, 4.9, 7.3, 11.7, 13.1)
> ansari.test(A, B, alternative="two.sided")
```

Ansari–Bradley test

data: A and B AB = 70.5, p-value = 0.01830 alternative hypothesis: true ratio of scales is not equal to 1

Warning message: Cannot compute exact p-value with ties in: ansari.test ...

Für den Fall, dass $N < 50$ und dass keine Bindungen auftreten, erfolgt in **R** die Berechnung exakter P-Werte zur Verteilung der Ansari-Bradley Teststatistik. Anderenfalls wird auch hier die Approximation mit Hilfe der Standardnormalverteilung verwendet.



Hinweis: Ein verteilungsfreier Test, der sowohl die zentrale Lage als auch die Streuung zwischen zwei Stichproben bewertet, ist der **Lepage-Test** [Lep71]. In der Teststatistik des Lepage-Tests wird neben der Ansari-Bradley-Statistik (7.75) auch die Wilcoxon-Statistik (vgl. Abschnitt [7.4.6]), hier in der Form

$$U = \sum_{i=1}^N iV_i, \quad (7.77)$$

verwendet. Die Summe 7.78 ist asymptotisch χ^2 -verteilt mit 2 Freiheitsgraden. Die exakte Verteilung ist in [Lep71] hergeleitet.

$$L = \left[\frac{U - E[U]}{\sqrt{\text{Var}[U]}} \right]^2 + \left[\frac{R - E[A_N]}{\sqrt{\text{Var}[A_N]}} \right]^2 \approx \chi^2_2 \quad (7.78)$$



Beispiel: Hendy, M.F. und Charles, J.A. [HC70] untersuchten den Silbergehalt in byzantinischen Münzen. Während der Regentschaft von Manuel I (1143-1180) gab es unterschiedliche Prägungen. In der folgenden Tabelle sollen zwei Stichproben, 9 Münzen aus der 1. Prägung und 7 Münzen aus der 4. Prägung, hinsichtlich der zentralen Lage und der Streuung verglichen werden.

1. Prägung (A)	5,9	6,0	6,4	7,0	6,6	7,7	7,2	6,9	6,2
4. Prägung (B)	5,3	5,6	5,5	5,1	6,2	5,8	5,8		

```
> A <- c(5.9, 6.0, 6.4, 7.0, 6.6, 7.7, 7.2, 6.9, 6.2)
```

```
> m <- length(A)
```

```
> B <- c(5.3, 5.6, 5.5, 5.1, 6.2, 5.8, 5.8)
```

```
> n <- length(B)
```

```
> N <- m + n
```

```
>
```

```
> W <- wilcox.test(A, B); W
```

```
W = 60.5, p-value = 0.002518
```

```
> S <- 60.5
```

```
>
```

```
> S1 <- (S - n*m/2) / sqrt(m*n*(N+1)/12); S1
```

```
[1] 3.069686
```

```
>
```

```
> A <- ansari.test(A, B); A
```

```
AB = 43.5, p-value = 0.5204
```

```
> S <- 43.5
```

```
> if (N%%2==0) {
```

```
+ S2 <- (S - (m*(N+2)/4)) / sqrt((m*n*(N^2-4)/(48*(N-1))));
```

```
+ S2 <- (S - (m*(N+1)^2/(4*N)) / sqrt(m*n*(N+1)*(3+N^2) / (48*N^2))); S2
```

```
[1] 0.6018207
```

```
>
```

```
> lepage <- S1^2 + S2^2; lepage
```

```
[1] 9.785157
> pchisq(lepage, 2, lower.tail=FALSE)
[1] 0.007502052
```

Die Berechnung der Wilcoxon-Statistik ($W = 60,5$) und der Ansari-Bradley-Statistik ($AB = 43,5$) erfolgt mit den Funktionen **wilcox.test()** und **ansari.test()** in **R**. Mit Hilfe der entsprechenden Formeln für die Erwartungswerte und die Varianzen (vgl. (7.76) und (7.106)) erfolgt die Berechnung der Lepage-Statistik elementar. Da der Wert der Lepage-Statistik 9,785 größer ist als das entsprechende Quantil der χ^2 -Verteilung $5,99 = \chi^2_{2;0,95}$, kann aus den Stichproben geschlossen werden, dass entweder die zentrale Lage (Median) oder die Variation oder aber beide Aspekte in den entsprechenden Grundgesamtheiten verschieden sind (die zuerst genannte Aussage dürfte zutreffen: $\tilde{\mu}_1 \neq \tilde{\mu}_2$).

7.4.4 t-Test für unabhängige Stichproben

7.4.4.1 Unbekannte aber gleiche Varianzen

Traditionelle Gründe führen dazu, dass in den folgenden Formeln die „**Q-Notation**“ mit aufgeführt wird. Einerseits sind die Formeln für die Prüfgrößen leichter zu lesen und zu verstehen, andererseits ist der Q-Wert bei der „manuellen“ Berechnung mehrerer Prüfgrößen häufig hilfreich. Die Summe der Abweichungsquadrate $\sum (x - \bar{x})^2$ bezeichnen wir somit im folgenden mit Q . Man berechnet sie nach

$$Q = \sum x^2 - (\sum x)^2/n \quad \text{bzw.} \quad Q = (n-1)s^2 \quad (7.79)$$

Mit einem weiteren Wert x_z , dem alten Mittelwert $\bar{x}_n$ und dem neuen Mittelwert $\bar{x}_{n+1}$ erhält man:

$$Q_{n+1} = Q_n + n(n+1)(\bar{x}_{n+1} - \bar{x}_n)^2 \quad (7.80)$$

Für den Vergleich zweier Mittelwerte ungleicher Stichprobenumfänge ($n_1 \neq n_2$) verwendet man die Prüfgröße (7.81) für den sogenannten **Zweistichproben-t-Test für unabhängige Zufallsstichproben** aus normalverteilten Grundgesamtheiten mit $n_1 + n_2 - 2$ Freiheitsgraden. Dieser Test ist bei zweiseitiger Fragestellung (d. h. $H_0: \mu_1 = \mu_2$, $H_A: \mu_1 \neq \mu_2$) und für nicht zu kleine und nicht zu unterschiedliche Stichprobenumfänge erfreulicherweise gegenüber Abweichungen von der Normalverteilung bemerkenswert robust. Für sehr unterschiedliche Stichprobenumfänge sowie $n_1 \geq 6$ und $n_2 \geq 6$ ist es angebracht, (7.81) durch (7.86) bzw. (7.87) zu ersetzen.

$$\begin{aligned} \hat{t} &= \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\left[\frac{n_1 + n_2}{n_1 \cdot n_2} \right] \cdot \left[\frac{Q_1 + Q_2}{n_1 + n_2 - 2} \right]}} \\ &= \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\left[\frac{n_1 + n_2}{n_1 n_2} \right] \cdot \left[\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \right]}}} \end{aligned} \quad (7.81)$$

Geprüft wird die Nullhypothese ($\mu_1 = \mu_2$) auf Gleichheit der den beiden Stichproben zugrunde liegenden Erwartungswerte der Grundgesamtheiten gegen $\mu_1 \neq \mu_2$ bei unbekannten aber gleichen Varianzen. Für den Fall gleicher Stichprobenumfänge ($n_1 = n_2$) ist **in der Regel vorteilhaft**, da der Fehler 2. Art minimal wird) vereinfacht sich (7.81) zu (7.82):

$$\hat{t} = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{Q_1 + Q_2}{n(n-1)}}} = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2 + s_2^2}{n}}} \quad (7.82)$$

mit $2n - 2$ Freiheitsgraden, wobei $n = n_1 = n_2$. Überschreitet der Prüfquotient die Signifikanzschranke, so gilt $\mu_1 \neq \mu_2$. Ist der Prüfquotient kleiner als die Schranke, dann kann die Nullhypothese $\mu_1 = \mu_2$ nicht abgelehnt werden.



Der t -Test ist **robust gegenüber einem α -Fehler**, nicht aber bezüglich der Power. Daher gelte für unübersichtliche Situationen: (1) $n_1 \approx n_2$; (2) $n_1 \gtrapprox 25, n_2 \gtrapprox 25$; (3) es wird zweiseitig geprüft. Dies gilt auch bei „Klumpigkeit“ (z. B. Bevorzugung stark gerundeter Daten). Abweichungen vom vorgegebenen α -Wert sind fast stets konservativ, d. h. die wirkliche Irrtumswahrscheinlichkeit ist kleiner als α , man erhält zu wenige statistisch signifikante Resultate. Dies gilt auch dann, wenn $n_1 > n_2$ und $\sigma_1^2 > \sigma_2^2$. Für $n_1 > n_2$ und $\sigma_1^2 < \sigma_2^2$ entscheidet der Test bevorzugt liberal, d. h. es treten zu viele statistisch signifikante Resultate auf; der Ansatz im folgenden Abschnitt hilft dann weiter.

Vergleich einer Beobachtung mit dem Mittelwert einer Stichprobe:

Geprüft wird H_0 : X ist mit μ verträglich. Aus (7.81) folgt:

$$\hat{t} = \frac{|x - \bar{x}|}{\sqrt{\frac{1+n}{n} \left[\frac{(n-1)s^2}{n-1} \right]}} = \frac{|x - \bar{x}|}{s\sqrt{(n+1)/n}} \text{ mit } FG = n - 1 \quad (7.83)$$

Bemerkungen zum Zweistichproben- t -Test

- (1) Der t -Test prüft die **Nullhypothese**, zwei Erwartungswerte lassen sich auf dem festgelegten Niveau nicht unterscheiden; es gibt keinen Behandlungseffekt.
- (2) Sind **mehrere Erwartungswerte** der genannten Art zu vergleichen, etwa aus Stichproben-Untergruppen, so darf der t -Test nicht mehrfach angewandt werden. Man prüft dann anhand der Varianzanalyse.
- (3) Wird die **Nullhypothese abgelehnt**, so kommen mehrere Erklärungen in Frage:
 - Beide Stichproben weisen schon vor der Behandlung unterschiedliche Erwartungswerte auf, sind also nicht vergleichbar.
 - Der t -Test darf wegen nicht erfüllter Voraussetzungen nicht benutzt werden.
 - Der Behandlungseffekt ist dem Zufall zu „verdanken“.
 - Der Behandlungseffekt existiert wirklich.



Beispiel: Prüfe $H_0 : \mu_1 = \mu_2$ gegen $H_A : \mu_1 \neq \mu_2$ auf dem 5%-Niveau. Gegeben seien $n_1, n_2; \bar{x}_1, \bar{x}_2; s_1^2, s_2^2$:

$$\begin{aligned} n_1 &= 16; \bar{x}_1 = 14,5; s_1^2 = 4 \\ n_2 &= 14; \bar{x}_2 = 13,0; s_2^2 = 3. \end{aligned}$$

Man berechne $Q_1 = (16-1) \cdot 4 = 60, Q_2 = (14-1) \cdot 3 = 39$ nach (7.79) und setze die Ergebnisse mit den anderen Größen in (7.81) ein.

$$\hat{t} = \frac{14,5 - 13,0}{\sqrt{\left[\frac{16+14}{16 \cdot 14} \right] \cdot \left[\frac{60+39}{16+14-2} \right]}} = \frac{1,5}{0,6881} = 2,180$$

Es stehen $n_1 + n_2 - 2 = 28$ Freiheitsgrade zur Verfügung, d. h. $t_{28;0,975} = 2,048$. Da $\hat{t} = 2,180 > 2,048$ ist, wird die Nullhypothese Gleichheit der Erwartungswerte auf dem 5%-Niveau abgelehnt und die Alternativhypothese $\mu_1 \neq \mu_2$ akzeptiert.

Eine elementare Berechnung des Beispiels in **R** unter Verwendung der Q-Notation:

```
> n1 <- 16; xbar1 <- 14.5; s1 <- 4
> n2 <- 14; xbar2 <- 13.0; s2 <- 3
>
> Q1 <- (n1 - 1)*s1
> Q2 <- (n2 - 1)*s2
>
> t.hat <- (xbar1 - xbar2) / sqrt(((n1+n2)/(n1*n2))*((Q1+Q2)/(n1+n2-2))); t.hat
[1] 2.179797
>
> t.krit <- qt(0.975, n1+n2-2); t.krit
[1] 2.048407
```

Beispiel: Zwei Medikamente zur Behandlung von Gerinnungsstörungen sollen hinsichtlich der Gerinnungszeiten (in Minuten) verglichen werden (Zar [Zar99]). Untersuchungen an $n_1 = 6$ und $n_2 = 7$ Patienten zeigten das folgende Ergebnis:



Gruppe	1	2	3	5	5	6	7
Medikament A	8,8	8,4	7,9	8,7	9,1	9,6	
Medikament B	9,9	9,0	11,1	9,6	8,7	10,4	9,5

Die Berechnung der Prüfgröße zum t-Test kann in **R** einfach mit der Funktion **t.test()** durchgeführt werden.

```
> x <- c(8.8, 8.4, 7.9, 8.7, 9.1, 9.6)
> y <- c(9.9, 9.0, 11.1, 9.6, 8.7, 10.4, 9.5)
>
> t.test(x, y, alternative="two.sided", var.equal=TRUE)

Two Sample t-test

data:  x and y  t = -2.4765, df = 11, p-value = 0.03076
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:  -1.8752609      -0.1104534
sample estimates: mean of x   mean of y
                8.750000    9.742857
```

Die Nullhypothese (hier zweiseitig $H_0 : \mu_1 = \mu_2$) kann verworfen werden, da der Wert der Prüfgröße $\hat{t} = 2,48$ (entsprechend einem P-Wert 0,03) größer ist als das Quantil der t-Verteilung mit 11 Freiheitsgraden $t_{11;0,975} = 2,201$.

Wichtige Hinweise zum t-Test

- (1) Das **Konfidenzintervall für die Differenz** zweier Erwartungswerte unabhängiger Stichproben aus normalverteilten Grundgesamtheiten mit gleicher Varianz ist durch (7.84) gegeben:

$$(\bar{x}_1 - \bar{x}_2) \pm t \cdot \sqrt{\quad} \quad (7.84)$$

mit $t = t_{n_1+n_2-2;1-\alpha/2}$

und $\sqrt{\quad}$ = der Nenner in (7.81) bzw. (7.82),

z. B. 95%-KI für $\mu_1 - \mu_2$ mit $t_{n_1+n_2-2;0,975}$

Wenn σ bekannt ist, wird t durch die Standardnormalvariable z ersetzt.

Ein Unterschied zwischen μ_1 , und μ_2 ist auf dem verwendeten Niveau statistisch signifikant, sobald das Konfidenzintervall den Wert $\mu_1 - \mu_2 = 0$ nicht einschließt. Statistische Testverfahren und Vertrauensbereiche führen beide zu Entscheidungen. Das Konfidenzintervall bietet darüber hinaus noch zusätzliche Informationen über den oder die Parameter!

- a) Schließt ein Konfidenzintervall die **Null** aus, so sind **Vorzeichen, Lage** und **Breite** des Konfidenzintervalles aufschlussreich.
- b) Konfidenzintervalle charakterisieren den Parameter, sind **einfacher zu verstehen als Tests und diesen praktisch gleichwertig**: eine H_0 bezüglich eines Parameters, die auf dem 5%-Niveau abgelehnt wird, wird auch abgelehnt, wenn das entsprechende 95%-KI den Nullwert-Parameter ausschließt.
- c) Im **Ein-Parameter-Fall ist das Konfidenzintervall dem Test fast stets überlegen**. Liegen zwei oder mehr Parameter vor, dann ist der P -Wert meist einfacher zu berechnen und wird dann routinemäßig angegeben.
- d) Sind **mehrere** Konfidenzintervalle zu vergleichen, so bevorzuge man **99%-Konfidenzintervalle**.

Beispiel: Wir benutzen das erste Beispiel und erhalten als 95%-KI:

$$(14,5 - 13,0) \pm 2,048 \cdot 0,6881 \text{ bzw. } 1,5 \pm 1,4.$$

Daraus folgt für das 95%-KI: $0,1 \leq \mu_1 - \mu_2 \leq 2,9$. Die Nullhypothese ($\mu_1 - \mu_2 = 0$) muss auch hier anhand der vorliegenden Stichproben auf dem 5%-Niveau verworfen werden, da das 95%-KI oberhalb der Null liegt.

- (2) Bei einseitiger Hypothesenstellung wird die Nullhypothese H_0 abgelehnt für:

$\begin{array}{ll} H_{A1} : \mu_1 > \mu_2 & \text{falls } \hat{t} > t \\ H_{A2} : \mu_1 < \mu_2 & \hat{t} < -t \end{array}$

Dabei ist $t = t_{\nu; 1-\alpha}$ das entsprechende Quantil der t -Verteilung einzusetzen.

Prüft man $H_0: \mu_1 - \mu_2 = \mu_0$ gegen $H_A: \mu_1 - \mu_2 \neq \mu_0$, so gilt für die entsprechenden Zähler [die Nenner und Freiheitsgrade bleiben unverändert] der Prüfgröße:

$H_{A1} : \dots > \mu_0$	$H_{A2} : \dots < \mu_0$	$H_{A3} : \dots \neq \mu_0$
$\bar{x}_1 - \bar{x}_2 - \mu_0$	$\bar{x}_1 - \bar{x}_2 - \mu_0$	$ \bar{x}_1 - \bar{x}_2 - \mu_0 $

- (3) **Relative Häufigkeiten** werden zur Stabilisierung der Varianz und zur Normalisierung transformiert. Häufig verwendet wird eine **Winkeltransformation** (Arcus-Sinus-Transformation, inverse Sinus-Transformation). Arcus sinus $\sqrt{p}$ (abgekürzt $\arcsin\sqrt{p}$ oder $\sin^{-1}\sqrt{p}$) bedeutet das Grad- bzw. Bogenmaß jenes Winkels, dessen Sinus gleich $\sqrt{p}$ ist. Für großes n ist $\arcsin\sqrt{p}$ normalverteilt mit dem Erwartungswert $\arcsin\sqrt{\pi}$ und der Varianz $1/4n$. Insbesondere ist die Varianz unabhängig vom Parameter π . Relative Häufigkeiten $x_i/n_i = \hat{p}_i$ (mit $n_i \simeq$ konstant und $n_i\hat{p}_i > 0,7$ sowie $n_i(1 - p_i) > 0,7$) zwischen 0 und 1 werden in Winkel von 0° bis 90° (Altgrad) umgewandelt. Es entsprechen sich somit (vgl. Tab. 7.25) z. B. relative Häufigkeit 0,25 und Altgrad 30.

Beispielsweise liegen zwei Untersuchungsreihen vor, jeweils Gruppen zu n Individuen. In jeder Gruppe weist der Anteil $\hat{p}_i$ der Individuen ein bestimmtes Merkmal auf. Sollen nun die Prozentsätze der beiden Reihen verglichen werden, so werden die auf 2 Dezimalen gerundeten $\hat{p}_i$ -Werte anhand der Transformation in x_i -Werte umgerechnet, die dann nach Berechnung der beiden Mittelwerte und Varianzen einen Vergleich der mittleren Prozentsätze beider Reihen ermöglichen.

In **R** können die Transformationen einfach mit der Funktion **asin()** durchgeführt werden. Dabei ist zu beachten, dass die Argumente für trigonometrische Funktionen im Bogenmaß ange-

Tabelle 7.25. Winkeltransformation: Werte $x = \arcsin \sqrt{p}$ (x in Altgrad) (z. B. $\arcsin \sqrt{0,25} = 30,0$; vgl. $\arcsin \sqrt{1,00} = 90,0$). [Umrechnung in Bogenmaß (Radiant): Tafelwerte durch 57,2958 teilen.]

p	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,000	5,739	8,130	9,974	11,537	12,921	14,179	15,342	16,430	17,457
0,1	18,435	19,370	20,268	21,134	21,973	22,786	23,578	24,350	25,104	25,842
0,2	26,565	27,275	27,972	28,658	29,334	30,000	30,657	31,306	31,948	32,583
0,3	33,211	33,833	34,450	35,062	35,669	36,271	36,870	37,465	38,057	38,646
0,4	39,231	39,815	40,397	40,976	41,554	42,130	42,706	43,280	43,854	44,427
0,5	45,000	45,573	46,146	46,720	47,294	47,870	48,446	49,024	49,603	50,185
0,6	50,769	51,354	51,943	52,535	53,130	53,729	54,331	54,938	55,550	56,167
0,7	56,789	57,417	58,052	58,694	59,343	60,000	60,666	61,342	62,028	62,725
0,8	63,435	64,158	64,896	65,650	66,422	67,214	68,027	68,866	69,732	70,630
0,9	71,565	72,543	73,570	74,658	75,821	77,079	78,463	80,026	81,870	84,261

geben werden, für diese konkrete Anwendung also mit dem Faktor $360^\circ/2\pi = 57,2958$ zu multiplizieren sind.

```
> asin(sqrt(c(0.1, 0.3, 0.5, 0.7, 0.9))) * (360/(2*pi))
[1] 18.43495 33.21091 45.00000 56.78909 71.56505
```

Dichotomverteilte Werte lassen sich auch durch die Logit- oder die Probit-Transformation normalisieren. Näheres ist z. B. dem Tafelwerk von Fisher und Yates [FY82] zu entnehmen, das auch eine ausführliche Tafel der Winkeltransformation enthält.

Bereinigter t-Test für k homogene Untergruppen aus zumindest angenähert normalverteilten Grundgesamtheiten mit gleichen Varianzen

Liegen zwei bezüglich einer Einfluss- oder Störgröße heterogene Zufallsstichproben vor aus zumindest angenähert normalverteilten Grundgesamtheiten mit gleichen Varianzen ($\sigma_1^2 = \sigma_2^2$) und ist ein Mittelwertvergleich geplant, so wird es sinnvoll sein, den Test für k homogene Untergruppen ($i = 1, 2, \dots, k$) (etwa von Patienten nach dem Geschlecht, dem Schweregrad der Erkrankung oder dem Alter) gemeinsam durchzuführen, vorausgesetzt die Umfänge n_{i1} in Stichprobe 1 und die Umfänge n_{i2} in Stichprobe 2 sind nicht zu klein. Für den alle k Untergruppen umfassenden und hinsichtlich der Einfluss- und Störgrößen bereinigten t-Test gilt auf dem verwendeten Signifikanzniveau α , dass sich μ_1 und μ_2 statistisch signifikant unterscheiden, sobald:

$$\hat{t} = \frac{\sum_{i=1}^k \frac{n_{1i}n_{2i}}{n_{1i} + n_{2i}} (\bar{x}_{1i} - \bar{x}_{2i})}{\sqrt{s^2 \sum_{i=1}^k \frac{n_{1i}n_{2i}}{n_{1i} + n_{2i}}}} \geq t_{\nu; 1-\alpha}$$

mit $\nu = \sum_{i=1}^k (n_{1i} + n_{2i} - 2)$ (7.85)

und $s^2 = \frac{\sum_{i=1}^k (n_{1i} - 1)s_{1i}^2 + (n_{2i} - 1)s_{2i}^2}{\sum_{i=1}^k (n_{1i} + n_{2i} - 2)}$
(aus Sachs [Sac90])

Beispiel: Ein einfaches Beispiel mit $k = 2$ homogenen Untergruppen ($H_0 : \mu_1 = \mu_2$, $H_A : \mu_1 \neq \mu_2$, $\alpha = 0,05$):



i	n_1	n_2	$\bar{x}_1$	$\bar{x}_2$	s_1^2	s_2^2
1	10	10	82	80	11	15
2	10	10	94	90	13	17

$$s^2 = \frac{[9 \cdot 11 + 9 \cdot 15] + [9 \cdot 13 + 9 \cdot 17]}{[10 + 10 - 2] + [10 + 10 - 2]} = 14$$

$$\frac{n_{11}n_{21}}{n_{11} + n_{21}} = \frac{n_{12}n_{22}}{n_{12} + n_{22}} = \frac{10 \cdot 10}{10 + 10} = 5$$

$$\hat{t} = \frac{[5(82 - 80)] + [5(94 - 90)]}{[10 + 10 - 2] + [10 + 10 - 2]} = 2,535 > 2,028 = t_{36;0,975}$$

Damit wird H_0 auf dem 5%-Niveau abgelehnt.

7.4.4.2 t-Test bei unbekannten Varianzen, die möglicherweise ungleich sind

Geprüft wird die Nullhypothese ($\mu_1 = \mu_2$) auf Gleichheit zweier Erwartungswerte bei nichtgleichen Varianzen ($\sigma_1^2 \neq \sigma_2^2$). Dies ist das sogenannte **Behrens-Fisher-Problem**. Für praktische Zwecke geeignet ist (7.86):

$$\hat{t} = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad \text{mit } \nu = n_2 - 1 \text{ Freiheitsgraden, für } n_1 \geq n_2 \quad (7.86)$$

Die Verteilung der rechten Seite von (7.86) ohne Absolutzeichen heisst bei Gültigkeit von H_0 : $\mu_1 = \mu_2$ **Behrens-Fisher-Verteilung** mit den Parametern n_1, n_2 und σ_1^2/σ_2^2 , beide Varianzen sind unbekannt. Diese Verteilung folgt keiner t -Verteilung. Näherungen, z. B. nach **Hsu** mit $\nu = \min(n_1, n_2) - 1$ (7.86) oder mit dem **Welch-Test**, für den die Abschätzung der Freiheitsgrade aufwendiger ist, sind möglich. Eine vergleichende Gegenüberstellung hinsichtlich der Einhaltung des Signifikanzniveaus befindet sich in Scheffé [Sch70].

Beispiel: Die Behauptung, dass der HDL-Wert (High-Density-Lipoprotein-Cholesterin in mg/dl) durch regelmäßigen Sport erhöht werden kann, soll an Hand zweier Stichproben von sportlich aktiven (A, $n_1 = 9$) und nicht aktiven (B, $n_2 = 11$) männlichen Studenten überprüft werden. Die Messungen ergaben:

Gruppe	1	2	3	4	5	6	7	8	9	10	11
A	29,5	44,9	54,2	55,4	58,5	59,8	60,1	84,2	97,5		
B	32,3	32,7	37,4	38,4	40,1	40,6	45,3	45,6	52,0	60,3	60,5

In **R** kann der t-Test für unabhängige Beobachtungen bei nichtgleichen Varianzen mit der Funktion **t.test()** gerechnet werden.

```
> aktiv <- c(29.5, 44.9, 54.2, 55.4, 58.5, 59.8, 60.1, 84.2, 97.5)
> inaktiv <- c(32.3, 32.7, 37.4, 38.4, 40.1, 40.6, 45.3, 45.6, 52.0, 60.3, 60.5)
>
> t.test(aktiv, inaktiv, alternative="greater", var.equal=FALSE)
```

Welch Two Sample t-test




```
data: aktiv and inaktiv t = 2.2378, df = 11.141, p-value = 0.02330
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval: 3.243236      Inf
sample estimates: mean of x mean of y      60.45556  44.10909
```

Die Annahme homogener Varianzen, die in **R** mit der Funktion **var.test()** überprüft werden kann, ist in den vorliegenden Daten nicht berechtigt. Daher wird die Funktion **t.test()** mit dem zusätzlichen Argument **var.equal=FALSE** verwendet. **R** berechnet die Prüfgröße, insbesondere die Anzahl der Freiheitsgrade nach der Welch-Statistik. Die Nullhypothese $H_0: \mu_1 \leq \mu_2$ kann auf dem 5%-Niveau abgelehnt werden, da $\hat{t} = 2,24 > 1,796 = t_{11;0,95}$ bzw. der P-Wert 0,0233 ist.

Im Falle gleicher Stichprobenumfänge ($n_1 = n_2 = n$) ergeben sich wieder folgende Vereinfachungen [Q wird nach (7.79) berechnet]

$$\hat{t} = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{Q_1 + Q_2}{n(n-1)}}} = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2 + s_2^2}{n}}} \quad \text{mit } \nu = n - 1 \text{ Freiheitsgraden} \quad (7.87)$$

Für den Vergleich mehrerer Erwartungswerte bei nicht unbedingt gleichen Varianzen existiert die sogenannte **Welch-Statistik** (vgl. Varianzanalyse), eine hervorragende Approximation (vgl. z.B. Sachs [Sac06])!

Einen weiteren Weg zur Lösung des Behrens-Fisher-Problems hat **Weir** [Wei60] vorgeschlagen. Für uns ist interessant, dass ein **Erwartungswert-Unterschied** auf dem 5%-Niveau statistisch gesichert ist, sobald für Stichprobenumfänge $n_1 \geq 3$ und $n_2 \geq 3$ die Prüfgröße

$$\frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{Q_1 + Q_2}{n_1 + n_2 - 4} \left[\frac{1}{n_1} + \frac{1}{n_2} \right]}} \quad \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 + 4} \left[\frac{1}{n_1} + \frac{1}{n_2} \right]}} \quad (7.88)$$

≥ 2 ist; unterschreitet der Quotient den Wert 2, dann lässt sich die Nullhypothese $\mu_1 = \mu_2$ auf dem 5%-Niveau nicht ablehnen. Das Biometrical Journal [28 (1986), 131–148] gibt eine Übersicht: How to Use the Two Sample *t*-Test.

Beispiel (wenn auch etwas bizarr): Vergleich zweier empirischer Mittelwerte auf dem 5%-Niveau:

$$\begin{aligned} n_1 &= 3; \quad 1,0 \quad 5,0 \quad 9,0; \quad \bar{x}_1 = 5,0; \quad Q_1 = 32; \quad s_1^2 = 16 \\ n_2 &= 3; \quad 10,9 \quad 11,0 \quad 11,1; \quad \bar{x}_2 = 11,0; \quad Q_2 = 0,02; \quad s_2^2 = 0,01 \end{aligned}$$



Q lässt sich hier schnell nach $Q = \sum (x - \bar{x})^2$ berechnen. Nach (7.88):

$$\frac{|5,0 - 11,0|}{\sqrt{\frac{32 + 0,02}{3 + 3 - 4} \left[\frac{1}{3} + \frac{1}{3} \right]}} = \frac{6}{3,27} < 2,0$$

Anhand der vorliegenden Stichproben lässt sich auf dem 5%-Niveau ein Unterschied nicht sichern. Das Standardverfahren (7.87) (beachte: n_1 und n_2 sind jetzt beide kleiner als 6)

$$\hat{t} = \frac{|5,0 - 11,0|}{\sqrt{\frac{32 + 0,02}{3(3-1)}}} = \frac{6}{3,31} < 4,303 = t_{2;0,975}$$

$$\left[\begin{array}{l} \text{Nach Hsu:} \\ \nu = 3 - 1 = 2 \quad \text{bzw. } \nu = 3 - 1 + \frac{2 \cdot 3 - 2}{\frac{32}{0,02} + \frac{0,02}{32}} \simeq 2 \end{array} \right]$$

liefert die gleiche Entscheidung.

Übersicht 20. Vergleich zweier empirischer Mittelwerte unabhängiger Stichproben aus angenähert normalverteilten Grundgesamtheiten

Stichproben- umfänge	Varianzen	
	gleich: $\sigma_1^2 = \sigma_2^2$	ungleich: $\sigma_1^2 \neq \sigma_2^2$
gleich: $n_1 = n_2 = n$	$\hat{t} = \frac{ \bar{x}_1 - \bar{x}_2 }{\sqrt{\frac{s_1^2 + s_2^2}{n}}}$ $FG = 2n - 2$	$\hat{t} = \frac{ \bar{x}_1 - \bar{x}_2 }{\sqrt{\frac{s_1^2 + s_2^2}{n}}}$ $FG = n - 1$
ungleich: $n_1 \neq n_2$	$\hat{t} = \frac{ \bar{x}_1 - \bar{x}_2 }{\sqrt{\left[\frac{n_1 + n_2}{n_1 n_2} \right] \cdot \left[\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \right]}}$ $FG = n_1 + n_2 - 2$	$\hat{t} = \frac{ \bar{x}_1 - \bar{x}_2 }{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ $FG = n_2 - 1$ für $n_1 > n_2$

Vier Bemerkungen zum Mittelwertvergleich

- (1) Stichproben, die nicht rein zufällig ausgewählt werden, sind gegenüber zufälligen Stichproben durch größere Ähnlichkeit der Stichprobenelemente untereinander und geringere Ähnlichkeit der Stichprobenmittelwerte charakterisiert. Beim nichtzufälligen Stichprobenziehen werden somit die Standardabweichungen verkleinert und die Mittelwertsunterschiede vergrößert. Beide Effekte können damit einen „signifikanten Mittelwertsunterschied“ vortäuschen! Daher müssen **knapp signifikante Resultate mit großer Vorsicht** interpretiert werden, sofern keine echten Zufallsstichproben vorgelegen haben.

- (2) Ein **Vergleich zweier Parameter aufgrund ihrer Konfidenzintervalle** ist möglich: (1) Überdecken sich die Konfidenzintervalle teilweise, so darf nicht gefolgert werden, dass sich die Parameter nicht signifikant unterscheiden. (2) Überdecken sich die Konfidenzintervalle nicht, so besteht zwischen den Parametern ein echter Unterschied: $H_0: \mu_1 = \mu_2$ lässt sich für $n_1 > 10$ und $n_2 > 10$ auf dem 5%-Niveau ablehnen, sobald sich die beiden 95%-Konfidenzintervalle nicht überlappen.

- (3) Die **Anzahl der Stichprobenwerte**, die man für den Vergleich eines Stichprobenmittelwertes mit dem Parameter der Grundgesamtheit oder für den Vergleich zweier Stichprobenmittelwerte benötigt, wird in Tabelle 7.26 für kontrollierte Fehler 1. Art ($\alpha = 0,05$ und $\alpha = 0,01$) und 2. Art ($\beta = 0,3; 0,2$ sowie $0,1$) und definierte Abweichungen gegeben.

- (4) Nach I.W. Molenaar [Mol04] kann auf eine Prüfung der Gleichheit zweier Varianzen anhand von s_1^2 und s_2^2 verzichtet werden. Gilt $0,5 \leq s_1^2/s_2^2 \leq 2$ bzw. gilt für $n_1 \approx n_2$: $0,25 \leq s_1^2/s_2^2 \leq 4$, dann ist der t-Test anwendbar, vorausgesetzt, die Zufallsstichproben sind zumindest angenähert normalverteilt.

7.4.4.3 Fallzahlabeschätzung für den t-Test für zwei unabhängige Stichproben

Die Ermittlung einer ausreichenden Fallzahl für den t-Test für zwei unabhängige Stichproben im Rahmen der Studienplanung geht von (7.89) aus. Dabei wird vorausgesetzt, dass die beiden Stichproben aus normalverteilten Grundgesamtheiten mit etwa gleicher Varianz stammen.

$$n \geq \frac{2\sigma^2}{\delta^2} (t_{\nu;1-\alpha} + t_{\nu;1-\beta})^2 \quad (7.89)$$

Damit wird der Stichprobenumfang durch vier Faktoren beeinflusst.

- (1) δ bezeichnet die kleinste Differenz, die durch den Hypothesentest bestätigt werden soll ($\delta = \mu_1 - \mu_2$). Kleine Differenzen erfordern somit eine höhere Fallzahl gegenüber großen (unter sonst gleichen Bedingungen).
- (2) Die Varianz in der Grundgesamtheit ist σ^2 . Eine hohe Variabilität in den Beobachtungen oder Messungen erfordert eine größere Fallzahl, um einen Unterschied als statistisch signifikant bestätigen zu können. σ^2 ist in der Regel unbekannt. Unter der Annahme, dass die Varianz in den Grundgesamtheiten gleich ist, kann σ^2 durch eine gemeinsame Varianzschätzung (engl. pooled variance) aus Voruntersuchungen oder auf der Grundlage einer Literaturrecherche ermittelt werden.

$$\sigma_{\bar{x}_1 - \bar{x}_2}^2 = \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} \sim \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} = s_p^2 \quad (7.90)$$

Hinweis: Die Beziehung (7.90) gilt nur dann, wenn beide Messreihen oder Stichproben stochastisch unabhängig voneinander sind.

- (3) Das Quantil der t-Verteilung $t_{\nu;1-\alpha}$ (einseitig) oder $t_{\nu;1-\alpha/2}$ (zweiseitig) wird um so größer ausfallen, je kleiner α gewählt wird, d.h. es müssen größere Stichproben untersucht werden. Mit anderen Worten, man benötigt größere Stichproben, wenn der Fehler 1. Art, das Risiko einer fälschlichen Ablehnung der Nullhypothese kleiner festgelegt wird.
- (4) Entsprechendes gilt für das Quantil der t-Verteilung $t_{\nu;1-\beta}$ hinsichtlich des Fehlers 2. Art. Eine höhere Power (Teststärke $(1 - \beta)$) fordert eine höhere Fallzahl in den Stichproben.

Die Tabelle 7.26 gibt bei einseitiger oder zweiseitiger Fragestellung für den Zweistichproben-t-Test den angenäherten Stichprobenumfang n (Zweistichprobentest: $n = n_1 = n_2$) an, der notwendig ist, um bei einer Irrtumswahrscheinlichkeit α mit der Power $1 - \beta$ eine Differenz auf dem 100 α %-Niveau als statistisch signifikant auszuweisen, wenn sich die Erwartungswerte zweier Grundgesamtheiten mit der gemeinsamen Standardabweichung σ um $(\mu_1 - \mu_2)/\sigma = \delta/\sigma$ unterscheiden. Um z. B. bei einseitiger Fragestellung auf dem 5%-Niveau eine Differenz $(\mu_1 - \mu_2)/\sigma$

$= 2/5 = 0,4$ mit einer Power von 0,7 als statistisch signifikant auszuweisen, benötigt man für den Zweistichproben- t -Test ($\alpha = 0,05$; Power = 0,7) jeweils 59 Beobachtungen; bei zweiseitiger Fragestellung auf dem 1%-Niveau und sonst gleichen Voraussetzungen werden bereits 121 Beobachtungen in jeder Stichprobe benötigt.

Tabelle 7.26. Stichprobenumfänge zum Zweistichproben- t -Test (ein- und zweiseitig)

$\frac{ \mu_1 - \mu_2 }{\sigma}$	Irrtumswahrscheinlichkeit $\alpha = 0,05$					
	einseitiger Test ($1 - \beta$)			zweiseitiger Test ($1 - \beta$)		
	Power			Power		
	0,7	0,8	0,9	0,7	0,8	0,9
0.1	942	1237	1713	1235	1570	2102
0.2	236	310	429	309	393	526
0.3	105	138	191	138	175	234
0.4	59	78	108	78	99	132
0.5	38	50	69	50	63	85
0.7	20	26	35	26	33	43
1.0	10	13	18	13	16	22
1.5	5	6	8	6	7	10

$\frac{ \mu_1 - \mu_2 }{\sigma}$	Irrtumswahrscheinlichkeit $\alpha = 0,01$					
	einseitiger Test ($1 - \beta$)			zweiseitiger Test ($1 - \beta$)		
	Power			Power		
	0,7	0,8	0,9	0,7	0,8	0,9
0.1	1626	2008	2604	1923	2336	2976
0.2	407	502	651	481	584	744
0.3	181	224	290	214	260	331
0.4	102	126	163	121	146	186
0.5	66	81	105	77	94	120
0.7	34	41	54	40	48	61
1.0	17	21	27	20	24	30
1.5	8	9	12	9	11	14

Die Abschätzung der Fallzahl kann in **R** mit der Funktion **power.t.test()** erfolgen. In dem folgenden Beispiel soll die Fallzahl für den Zweistichproben- t -Test bei einseitiger Hypothesenstellung mit $\alpha = 0,05$ und $\beta = 0,20$ bestimmt werden. Z.B. soll die Behauptung überprüft werden, dass der mittlere diastolische Blutdruck bei Patienten mit einer bestimmten Krankheit gegenüber gesunden Kontrollen „um mindestens“ 15mmHg erhöht ist. Die gemeinsame Varianz wird mit $\sigma^2 = 400$ angenommen. Für die Funktion **power.t.test()** sind neben der Teststatistik (Ein- bzw. Zweistichprobentest) und der Art der Hypothesenstellung (ein- bzw. zweiseitig) die relevanten Parameter aus (7.89) zu spezifizieren. Dabei wird die zu berechnende Größe (z.B. $n = NULL$) ausgelassen. Besonders zu beachten ist, dass der Parameter „delta=15“ in diesem Fall die wahre (zu prüfende) Differenz zwischen den Erwartungswerten angibt, die hinsichtlich der in Tabelle 7.26 verwendete Effektstärke noch durch die Standardabweichung dividiert werden muss.

```
> power.t.test(delta=15, sd=20, sig.level=0.05, power=0.80, n=NULL,
+               type="two.sample", alternative="one.sided")
```

Two-sample t test power calculation

```
      n = 22.69032
delta = 15
```

```
sd = 20
sig.level = 0.05
power = 0.8
alternative = one.sided
```

Für unser Zahlenbeispiel ist $(\mu_1 - \mu_2)/\sigma = 15/20 = 0,75$ (Effektstärke). Mit dem festen Signifikanzniveau $\alpha = 0,05$ und der Power $1 - \beta = 0,80$ liefert die Funktion **power.t.test()** in **R** einen notwendigen Stichprobenumfang von $n_1 = n_2 = 23$ Beobachtungen.

Die Formel für die Fallzahl (7.89) kann so umgeformt werden, dass insbesondere auch die Teststärke (Power) oder die „minimal aufdeckbare“ Differenz (7.91) für einen Test bei fester Fallzahl bestimmt werden können.

$$\delta \geq \sqrt{\frac{2\sigma^2}{n}}(t_{\nu;1-\alpha} + t_{\nu;1-\beta}) \quad (7.91)$$

Beispiel: Welcher Effekt kann in einem zweiseitigen Test hinsichtlich der Gerinnungszeiten (gemessen in Minuten) unter der Therapie mit zwei verschiedenen Medikamenten beurteilt werden, wenn die Untersuchung mit je 20 Fällen durchgeführt wird? Die Testentscheidung soll mit $\alpha = 0,05$ und $\beta = 0,10$ getroffen werden. aus Voruntersuchungen kann die Varianz mit $\sigma^2 = 0,5193$ angenommen werden.



$$\delta = \sqrt{\frac{2(0,5193)}{20}}(2,024 + 1,304) = 0,76$$

```
> power.t.test(n=20, sd=sqrt(0.5193), sig.level=0.05, power=0.90, delta=NULL,
+               type="two.sample", alternative="two.sided")
```

Two-sample t test power calculation

```
n = 20
delta = 0.7580757
sd = 0.7206247
sig.level = 0.05
power = 0.9
alternative = two.sided
```

Mit $n=20$ Patienten je Behandlungsgruppe kann demnach bei einer zweiseitigen Fragestellung mit einer Irrtumswahrscheinlichkeit von 5% und einer Power von 90% eine Differenz von mindestens 0,76 Minuten als statistisch signifikant erkannt werden.

Hinweis: Der Zweistichproben t-Test ist robust und hat die größte Power, wenn die Stichprobenumfänge gleich sind ($n_1 = n_2$). Ist $n_1 \neq n_2$, dann kann (7.91) mit dem harmonischen Mittel von n_1 und n_2 verwendet werden.

$$n = \frac{2n_1n_2}{n_1 + n_2} \quad (7.92)$$

7.4.5 t-Test für Paardifferenzen

7.4.5.1 Gepaarte Beobachtungen

Wenn n Personen einen Sonnenbrand haben und zwei bewährte Behandlungsmöglichkeiten verfügbar sind, wird jede Person, die keine Behandlungserfahrung hat, gut beraten sein, beide Behandlungen an vergleichbar geschädigten und symmetrisch zueinander liegenden Hautpartien anzuwenden: es liegen dann **blockinterne Vergleiche** vor.

An jeder Person lassen sich als Resultat der Behandlungen **paarweise Beobachtungen** (z. B. Heilung nach ... Stunden) feststellen. Man spricht auch von verbundenen Stichproben oder von **paarigen Stichproben**. Hätte man den Vergleich an zwei unabhängigen Stichproben durchgeführt, wären doppelt so viele Probanden nötig gewesen. Man hätte auch beachten müssen, dass sich beide Stichproben hinsichtlich Schweregrad, ..., möglichst wenig unterscheiden. Es wäre dann auch besser gewesen, jeweils **Probandenpaare** zu bilden, die sich in wesentlichen Einflussgrößen ähneln. Die Zuordnung der Partner zu den beiden Behandlungsgruppen wäre dann durch einen Zufallsprozess bestimmt worden. Auch hier hätten wir dann paarige Stichproben; denn jedes Paar wäre ein Block. Dieser Block wäre dann aber deutlich weniger homogen als in dem von uns gewählten Ansatz, bei dem wir die Streuung im Block ausgeschaltet haben (gleichwertige Hautpartien!).

Beim Übergang von unabhängigen zu paarigen Stichproben sinkt die Zahl der Freiheitsgrade auf die Hälfte. Dieser Genauigkeitsverlust wird meist durch Blockbildung mehr als ausgeglichen.

Bezeichnet man die Varianz zwischen den Differenzen und Summen der Einzelpaare mit s_d^2 und s_s^2 , dann sind paarige Stichproben (n Paare) unabhängigen Stichproben (jeweils des Umfangs n) überlegen für

$$\left[\frac{n(2n+1)}{(n+2)(2n-1)} \right] \left[\frac{(n-1)s_s^2 + ns_d^2}{(2n-1)s_d^2} \right] > 1 \quad (7.93)$$

Nehmen wir z. B. die Werte der Tabelle 7.27: $s_d^2 = [20,04 - (9,2)^2/8]/7 = 1,35$; zur Berechnung von s_s^2 sind die Spalten $(x_i + y_i)$ und $(x_i + y_i)^2$ zu ergänzen, hieraus ergibt sich

$$s_s^2 = [545,60 - (65,0)^2/8]/7 = 2,50 \text{ und damit } \left[\frac{8 \cdot 17}{10 \cdot 15} \right] \left[\frac{7 \cdot 2,50 + 8 \cdot 1,35}{15 \cdot 1,35} \right] = 1,27 > 1;$$

d. h. auch für künftige Untersuchungen sind gepaarte Beobachtungen zu bevorzugen.

Paarige Stichproben erhält man nach folgenden beiden Prinzipien. Bekannt ist der Aufbau von Versuchen mit **Testwiederholung (repeated measurements)** an einer und derselben Stichprobe von Individuen. Versuchspersonen werden z. B. einmal unter Normalbedingungen und anschließend unter Stress getestet. Hierbei ist zu beachten, dass Faktoren wie z. B. Übung oder Ermüdung ausgeschaltet werden müssen. Das zweite Prinzip bildet die **Organisierung von paarigen Stichproben mit Hilfe einer Vortestung** oder eines mess- oder schätzbaren Merkmales, das mit dem zu untersuchenden Merkmal möglichst stark korreliert (**matching**). Die Individuen werden z. B. aufgrund des Vortests in eine Rangreihe gebracht. Je zwei in dieser Rangliste aufeinanderfolgende Individuen bilden ein Paar. Durch einen Zufallsprozess etwa mit Hilfe eines Münzwurfs wird entschieden, welcher Partner zu welcher Stichprobengruppe gehören soll.

Für die Standardabweichung der Differenz zwischen den Mittelwerten zweier Messreihen oder Stichproben haben wir in Formel (7.86)

$$s_{\bar{x}_1 - \bar{x}_2} = s_{\text{Diff.}} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{s_{\bar{x}_1}^2 + s_{\bar{x}_2}^2} \quad (7.94)$$

benutzt. Diese Bezeichnung gilt aber nur dann, wenn beide Messreihen oder Stichproben **stochastisch unabhängig voneinander** sind. Sind sie miteinander **verbunden**, voneinander abhängig, d. h. besteht ein Zusammenhang zwischen den Wertepaaren, so vermindert sich die Standardabweichung der Differenz und wir erhalten jetzt $s_{\text{Diff.},r}$

$$s_{\text{Diff.},r} = \sqrt{s_{\bar{x}_1}^2 + s_{\bar{x}_2}^2 - 2rs_{\bar{x}_1}s_{\bar{x}_2}} \quad (7.95)$$

$$\left[\text{vgl. Var}(\bar{X}_1 - \bar{X}_2) = [\sigma_1^2/n_1] + [\sigma_2^2/n_2] - 2\rho\sqrt{[\sigma_1^2/n_1][\sigma_2^2/n_2]} \right]$$

Aus (7.95) folgt: $r = (s_{\bar{x}_1}^2 + s_{\bar{x}_2}^2 - s_{\text{Diff.};r}^2) / (2s_{\bar{x}_1}s_{\bar{x}_2})$

Die Größe des Subtraktionsgliedes richtet sich nach der Größe des Korrelationskoeffizienten r , der den Grad des Zusammenhangs ausdrückt. Die Standardabweichung der Differenz nimmt ab, sobald r von -1 nach 1 zunimmt.

Bei $r = 0$, d. h. in stochastisch unabhängigen Stichproben, wird das Subtraktionsglied unter der Wurzel gleich Null; bei $r = 1$, d. h. bei maximaler Korrelation oder vollständiger Abhängigkeit erreicht das Subtraktionsglied sein Maximum und die Standardabweichung der Differenz ihr Minimum. Die Power, einen Unterschied $\mu_x - \mu_y$ zu erfassen, ist dann größer, vorausgesetzt, der Stichprobenumfang der verbundenen Stichproben ist hinreichend groß.

7.4.5.2 t-Test für paarweise angeordnete Messwerte

Die Werte der beiden verbundenen Messreihen seien x_i und y_i . Für die Prüfung der Paardifferenzen $x_i - y_i = d_i$ dient der Quotient

$$\hat{t} = \frac{\bar{d}}{s_{\bar{d}}} = \frac{(\sum d_i)/n}{\sqrt{\frac{\sum d_i^2 - (\sum d_i)^2/n}{n(n-1)}}} \quad FG = n - 1 \quad (7.96)$$

aus dem Mittelwert der n Differenzen und der zugehörigen Standardabweichung mit $n - 1$ Freiheitsgraden, wobei n die Anzahl der Paardifferenzen bezeichnet. Vorausgesetzt werden **unabhängige Differenzen aus Zufallsstichproben zumindest angenähert normalverteilter Differenzen** $N(\mu_d, \sigma_d)$. Getestet wird der aus den Paardifferenzen geschätzte Erwartungswert μ_d (damit entspricht dieser Test dem Einstichproben-t-Test, der in Abschnitt [7.3.2.1] dargestellt ist). Geprüft wird $H_0: \mu_d = 0$ gegen $H_A: \mu_d > 0$ bzw. $\mu_d < 0$ oder beim zweiseitigen Test $\mu_d \neq 0$. Ergänzend wird stets auch (7.97) berechnet.

Erhält man fast ausschließlich Werte $d_i \approx 1$, dann ist [z.B. für $n = 10: \sum d_i^2 \approx 10$ und $(\sum d_i)^2/n \approx 10^2/10 = 10$] der Zähler unter der Quadratwurzel praktisch gleich Null und die Formel nicht mehr anwendbar. In diesem Fall benutze man den Wilcoxon-Test für Paardifferenzen in [7.4.7].

Beispiel: Die Tabelle 7.27 enthalte Messwerte (x_i, y_i) für ein Material, das nach zwei Verfahren behandelt wurde bzw. für unbehandeltes (x_i) und behandeltes Material (y_i). Das durchnummerierte Material sei unterschiedlicher Herkunft. Lässt sich die Nullhypothese, kein Behandlungsunterschied bzw. kein Behandlungseffekt (zweiseitige Fragestellung) auf dem 5%-Niveau sichern?

Es ist
$$\hat{t} = \frac{9,2/8}{\sqrt{\frac{20,04 - 9,2^2/8}{8(8-1)}}} = \left[\frac{\bar{d}}{s_{\bar{d}}} \right] = \frac{1,15}{0,4110} = 2,798 \text{ oder } 2,80$$

und, da $\hat{t} = 2,798 > 2,365 = t_{7;0,975}$, ist der Verfahrensunterschied bzw. der Behandlungseffekt auf dem 5%-Niveau statistisch gesichert.

In **R** kann der t-Test für Paardifferenzen direkt auch mit der Funktion **t.test()** mit dem zusätzlichen Parameter „**paired=TRUE**“ berechnet werden. Für die Werte aus dem obigen Beispiel folgt:

```
> behandelt <- c(4.0, 3.5, 4.1, 5.5, 4.6, 6.0, 5.1, 4.3)
> unbehandelt <- c(3.0, 3.0, 3.8, 2.1, 4.9, 5.3, 3.1, 2.7)
>
> t.test(behandelt, unbehandelt, alternative = c("two.sided"), paired = TRUE)
```



Tabelle 7.27. Messwerte für unterschiedlich behandeltes Material

Nr.	x_i	y_i	$d_i = (x_i - y_i)$	d_i^2
1	4,0	3,0	1,0	1,00
2	3,5	3,0	0,5	0,25
3	4,1	3,8	0,3	0,09
4	5,5	2,1	3,4	11,56
5	4,6	4,9	-0,3	0,09
6	6,0	5,3	0,7	0,49
7	5,1	3,1	2,0	4,00
8	4,3	2,7	1,6	2,56
$n = 8$			$\sum d_i = 9,2$	$\sum d_i^2 = 20,04$

Paired t-test

data: behandelt and unbehandelt $t = 2.798$, $df = 7$, $p\text{-value} = 0.0266$
 alternative hypothesis: true difference in means is not equal to 0
 95 percent confidence interval: 0.1781177 2.1218823
 sample estimates: mean of the differences 1.15

Verglichen mit dem Standardverfahren für den Vergleich der Mittelwerte zweier unabhängiger Stichproben vermeidet man beim Arbeiten mit gepaarten Beobachtungen einmal störende Streuungen. Zum anderen sind die Voraussetzungen schwächer. Es kann sein, dass die Variablen x_i und y_i von der Normalverteilung beträchtlich abweichen, die Differenzen aber recht gut normalverteilt sind!

Das Konfidenzintervall für den Erwartungswert μ_d der Paardifferenzen ist durch

$$\bar{d} \pm (t_{n-1; 1-\alpha/2}) s_{\bar{d}} \quad (7.97)$$

$$\text{gegeben mit } \bar{d} = \frac{\sum d}{n} \text{ und } s_{\bar{d}} = \frac{s_d}{\sqrt{n}} = \sqrt{\frac{\sum d_i^2 - (\sum d_i)^2/n}{n(n-1)}}$$

Für unser Beispiel erhält man das 95%-Konfidenzintervall: $1,15 \pm 2,365 \cdot 0,411$ bzw. $1,15 \pm 0,97$, d. h. 95%-KI: $0,18 \leq \mu_d \leq 2,12$, das, dem Testresultat entsprechend, die Null nicht mit einschließt.

Einseitige Vertrauensgrenzen lassen sich auch angeben. Als obere 95%-Vertrauensgrenze erhielte man auch nach (7.97) mit $t_{7; 0,0,95} = 1,895$ über $1,15 + 1,895 \cdot 0,411 = 1,15 + 0,78$ den Wert 1,93, d. h. 95%-VG_{oben}: $\mu_d = 1,93$ bzw. $\mu_d \leq 1,93$.

7.4.5.3 Prüfung der Gleichheit zweier Varianzen paariger Stichproben

Soll die Variabilität eines Merkmals vor (x_i) und nach (y_i) einem Alterungsprozess oder einer Behandlung verglichen werden, dann sind zwei Varianzen paarweise angeordneter Beobachtungen auf Gleichheit zu prüfen. Prüfgröße ist

$$\hat{t} = \frac{|(Q_x - Q_y) \cdot \sqrt{n-2}|}{2\sqrt{Q_x Q_y - (Q_{xy})^2}} \quad (7.98)$$

mit $n-2$ Freiheitsgraden. Q_x und Q_y werden nach (7.79) berechnet. Q_{xy} erhält man dementsprechend nach

$$Q_{xy} = \sum xy - \frac{\sum x \sum y}{n} \quad (7.99)$$

Beispielsweise ergibt sich für $\frac{x_i | 21 \ 18 \ 20 \ 21 | \sum x = 80}{y_i | 26 \ 33 \ 27 \ 34 | \sum y = 120}$ mit $Q_x = 6$, $Q_y = 50$ und

$$Q_{xy} = (21 \cdot 26 + 18 \cdot 33 + 20 \cdot 27 + 21 \cdot 34) - \frac{80 \cdot 120}{4} = -6$$

$$\hat{t} = \frac{|(6 - 50) \cdot \sqrt{4 - 2}|}{2 \cdot \sqrt{6 \cdot 50 - (-6)^2}} = 1,91 < 4,30 = t_{2;0,975}$$

bei zweiseitiger Fragestellung auf dem 5%Niveau, dass die Nullhypothese: Gleichheit beider Varianzen, beibehalten werden muss. Bei begründeter einseitiger Fragestellung mit $\sigma_x^2 = \sigma_y^2$ gegen $\sigma_x^2 > \sigma_y^2$ oder $\sigma_x^2 < \sigma_y^2$ wäre $t_{2;0,95} = 2,92$ die kritische Schranke.

7.4.6 Wilcoxon Rangsummentest für zwei unabhängige Stichproben

U-Test nach Wilcoxon, Mann und Whitney

Werden n Stichprobenwerte der Größe nach aufsteigend geordnet und mit $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ bezeichnet, so dass

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(i)} \leq \dots \leq x_{(n)}$$

$x_{(1)}$ = kleinste Beobachtung,

$x_{(n)}$ = größte Beobachtung

gilt, dann heißt jede der Größen $x_{(i)}$ **Ranggröße** (order statistic). Man bezeichnet die Nummer, die jedem Stichprobenwert zukommt, als **Rang**, Rangplatz, Rangwert oder **Rangzahl** (rank). Der Ranggröße $x_{(i)}$ entspricht also der Rangwert i oder die Rangzahl i . Tests, bei denen anstelle der Stichprobenwerte deren Rangzahlen verwendet werden, bilden eine besonders wichtige Gruppe verteilungsunabhängiger Tests. Beachtet sei, dass der Erwartungswert von Rängen durch $(n+1)/2$ gegeben ist, die Varianz durch $(n^2 - 1)/12$. **Rangsummentests** weisen erstaunlicherweise eine relativ hohe asymptotische Effizienz auf.



Der auf dem sogenannten Wilcoxon-Test für unabhängige Stichproben basierende Rangtest von Mann und Whitney [MW47] ist das verteilungsunabhängige Gegenstück zum parametrischen t -Test für den Vergleich zweier Erwartungswerte stetiger Verteilungen.

Hinweis: Die unter Umständen verwirrende Unterscheidung von Wilcoxon Rangsummentest und U-Test beruht lediglich auf einer Transformation in der Teststatistik. In Übereinstimmung mit älteren Auflagen wird hier der U-Test vorgestellt und auf die Umrechnung der Teststatistik, wie sie in **R** verwendet wird, besonders verwiesen.

Die Stetigkeitsannahme ist, streng genommen, in der Praxis nie erfüllt, da alle Messergebnisse gerundete Zahlen sind. Die asymptotische Effizienz des U -Tests liegt bei $100 \cdot 3/\pi \simeq 95\%$, d. h. dass die Anwendung dieses Tests bei 1000 Werten die gleiche Teststärke aufweist wie die Anwendung des t -Tests bei etwa $0,95 \cdot 1000 = 950$ Werten, wenn in Wirklichkeit Normalverteilung vorliegt. Es wird also selbst dann, wenn dies tatsächlich der Fall ist, vorteilhaft sein, den U -Test anzuwenden, z. B. bei **Überschlagsrechnungen** oder zur Kontrolle hochsignifikanter t -Test-Befunde, denen man nicht so recht traut. Vorausgesetzt wird, dass die zu vergleichenden Stichproben die **gleiche Verteilungsform** aufweisen. Wenn nicht, ist der Median-Quartile-Test anzuwenden (vgl. Hinweis am Ende dieses Abschnitts).

Voraussetzungen des U -Tests:

- Stetige Verteilungsfunktionen und
- zwei unabhängige Zufallsstichproben von Messwerten oder zumindest von Rangdaten aus Grundgesamtheiten mit ähnlicher bis gleicher Verteilungsform.

Der U -Test von Wilcoxon, Mann und Whitney prüft bei **zweiseitiger Fragestellung** die Nullhypothese: Die Wahrscheinlichkeit, dass eine Beobachtung der ersten Grundgesamtheit größer ist als eine beliebig gezogene Beobachtung der zweiten Grundgesamtheit, ist gleich $\frac{1}{2}$

$$\text{d. h. } H_0 : P(X_1 > X_2) = \frac{1}{2} \quad \text{gegen} \quad H_A : P(X_1 > X_2) \neq \frac{1}{2} \quad (7.100)$$

Beachtet sei, dass weder Parameter noch ihre Schätzwerte für die Berechnung der Prüfgröße U benötigt werden. Außerdem lässt sich auch das Hypothesenpaar ohne Parameter formulieren: **Gleichheit bzw. Ungleichheit der beiden Verteilungsfunktionen**, die vergleichbar sind und sich nicht schneiden:

$$\begin{aligned} H_0 : F_1(x) &= F_2(x) \quad \text{für alle } x \\ H_A : F_1(x) &\neq F_2(x) \quad \text{für mindestens ein } x \end{aligned} \quad (7.101)$$

Die entsprechenden **einseitigen Fragestellungen** sind ebenfalls möglich; wir schreiben sie jetzt:

$$\begin{aligned} H_{01} : P(X_1 > X_2) &\geq \frac{1}{2} \quad \text{gegen} \quad H_{A1} : P(X_1 > X_2) < \frac{1}{2} \\ &\text{bzw.} \\ H_{02} : P(X_1 > X_2) &\leq \frac{1}{2} \quad \text{gegen} \quad H_{A2} : P(X_1 > X_2) > \frac{1}{2} \end{aligned} \quad (7.102)$$

Gilt $F_1(x) = F_2(x + c)$, dann lassen sich mit U auch die Erwartungswerte prüfen, wie wir es vom t -Test her kennen: z. B. $H_0 : \mu_1 \leq \mu_2$ gegen $H_A : \mu_1 > \mu_2$.

Der Test ist empfindlich gegenüber Medianwertunterschieden, weniger empfindlich bei unterschiedlichen Schiefen und unempfindlich für Varianzunterschiede (diese werden bei Bedarf nach Siegel und Tukey geprüft, vgl. [7.4.2]). Für $n_1 = n_2$ ist der U -Test robust gegenüber Streuungsunterschieden: er prüft dann auch die Gleichheit zweier Mediane ($\tilde{\mu}_1, \tilde{\mu}_2$).

Prinzip des U -Tests: Der U -Test, ein Rangsummentest für den Vergleich zweier unabhängiger Stichproben bei nicht-normalverteilten Grundgesamtheiten, geht davon aus, dass die $n = n_1 + n_2$ Beobachtungen der Größe nach angeordnet und durchnummeriert werden, und zwar von 1 bis n . Weist die eine Stichprobe im Durchschnitt kleinere Werte als die andere Stichprobe auf, so werden sich die Rangsummen beider Stichproben unterscheiden: etwa wie in dem folgenden Beispiel:

Stichprobe A	x_i :	6, 5, 11	$(n_A = 3)$			} $n = 6$	
Stichprobe B	x_i :	17, 12, 16	$(n_B = 3)$				
Gemeinsam geordnet:		5 < 6 < 11 < 12 < 16 < 17					
Ränge:		1	2	3	4	5	6
Herkunft:		A	A	A	B	B	B
		$\underbrace{\hspace{1.5cm}}$			$\underbrace{\hspace{1.5cm}}$		
Rangsummen:		$R_A = 6$			$\ll R_B = 15$		

Hinweis: Ist $n_1 + n_2 = n$ genügend groß, so lässt sich die Verteilung der Rangsummen anhand der Standardnormalverteilung (vgl. (7.110) und (7.111)) approximieren.

Tabelle 7.28. Kritische Werte von U für den Test von Wilcoxon, Mann und Whitney für den einseitigen Test: $\alpha = 0,05$; zweiseitigen Test: $\alpha = 0,10$ (entnommen aus Milton, R.C.: An extended table of critical values for the Mann-Whitney (Wilcoxon) two-sample statistic, J. Amer. Statist. Ass. **59** (1964), 925-934)

m	n																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	-																			
2	-	-																		
3	-	-	0																	
4	-	-	0	1																
5	-	0	1	2	4															
6	-	0	2	3	5	7														
7	-	0	2	4	6	8	11													
8	-	1	3	5	8	10	13	15												
9	-	1	4	6	9	12	15	18	21											
10	-	1	4	7	11	14	17	20	24	27										
11	-	1	5	8	12	16	19	23	27	31	34									
12	-	2	5	9	13	17	21	26	30	34	38	42								
13	-	2	6	10	15	19	24	28	33	37	42	47	51							
14	-	3	7	11	16	21	26	31	36	41	46	51	56	61						
15	-	3	7	12	18	23	28	33	39	44	50	55	61	66	72					
16	-	3	8	14	19	25	30	36	42	48	54	60	65	71	77	83				
17	-	3	9	15	20	26	33	39	45	51	57	64	70	77	83	89	96			
18	-	4	9	16	22	28	35	41	48	55	61	68	75	82	88	95	102	109		
19	0	4	10	17	23	30	37	44	51	58	65	72	80	87	94	101	109	116	123	
20	0	4	11	18	25	32	39	47	54	62	69	77	84	92	100	107	115	123	130	138
21	0	5	11	19	26	34	41	49	57	65	73	81	89	97	105	113	121	130	138	146
22	0	5	12	20	28	36	44	52	60	68	77	85	94	102	111	119	128	136	145	154
23	0	5	13	21	29	37	46	54	63	72	81	90	98	107	116	125	134	143	152	161
24	0	6	13	22	30	39	48	57	66	75	85	94	103	113	122	131	141	150	160	169
25	0	6	14	23	32	41	50	60	69	79	89	98	108	118	128	137	147	157	167	177
26	0	6	15	24	33	43	53	62	72	82	92	103	113	123	133	143	154	164	174	185
27	0	7	15	25	35	45	55	65	75	86	96	107	117	128	139	149	160	171	182	192
28	0	7	16	26	36	46	57	68	78	89	100	111	122	133	144	156	167	178	189	200
29	0	7	17	27	38	48	59	70	82	93	104	116	127	138	150	162	173	185	196	208
30	0	7	17	28	39	50	61	73	85	96	108	120	132	144	156	168	180	192	204	216
31	0	8	18	29	40	52	64	76	88	100	112	124	136	149	161	174	186	199	211	224
32	0	8	19	30	42	54	66	78	91	103	116	128	141	154	167	180	193	206	218	231
33	0	8	19	31	43	56	68	81	94	107	120	133	146	159	172	186	199	212	226	239
34	0	9	20	32	45	57	70	84	97	110	124	137	151	164	178	192	206	219	233	247
35	0	9	21	33	46	59	73	86	100	114	128	141	156	170	184	198	212	226	241	255
36	0	9	21	34	48	61	75	89	103	117	131	146	160	175	189	204	219	233	248	263
37	0	10	22	35	49	63	77	91	106	121	135	150	165	180	195	210	225	240	255	271
38	0	10	23	36	50	65	79	94	109	124	139	154	170	185	201	216	232	247	263	278
39	1	10	23	38	52	67	82	97	112	128	143	159	175	190	206	222	238	254	270	286 ⁺
40	1	11	24	39	53	68	84	99	115	131	147	163	179	196	212	228	245	261	278	294 ⁺

+ anhand der Normalverteilung approximierte Werte

Zur Berechnung der Prüfgröße U bringt man die $(m + n)$ Stichprobenwerte in eine gemeinsame aufsteigende Rangfolge, wobei zu jeder Rangzahl vermerkt wird, aus welcher der beiden Stichproben der zugehörige Wert stammt. Die Summe der auf Stichprobe 1 entfallenden Rangzahlen sei R_1 , die Summe der auf Stichprobe 2 entfallenden Rangzahlen sei R_2 . Dann berechnet man (7.103) und kontrolliert die Rechnung nach (7.104)

$$U_1 = mn + \frac{m(m+1)}{2} - R_1 \quad U_2 = mn + \frac{n(n+1)}{2} - R_2 \quad (7.103)$$

$$U_1 + U_2 = mn \quad (7.104)$$

Die gesuchte Prüfgröße ist die kleinere der beiden Größen U_1 und U_2 . Die Nullhypothese wird verworfen, wenn der berechnete U -Wert kleiner oder gleich dem kritischen Wert $U(m, n; \alpha)$ aus Tabelle 7.28 ist.

Grundlage für die Berechnung kritischer Werte (Quantile) für den U -Test ist die **Wilcoxon-Verteilung**. Als Zufallsvariable wird die Summe der zu einer Stichprobe gehörenden Rangzahlen aus der gemeinsamen Rangverteilung näher betrachtet. Diese hat den kleinsten Wert $m(m+1)/2$, wenn alle Werte der ersten Stichprobe kleiner sind als der kleinste Wert aus der 2. Stichprobe und den größten Wert $N(N+1)/2 - n(n+1)/2$, wenn alle Werte der ersten Stichprobe größer sind als

der größte Wert der 2. Stichprobe (mit $N = m + n$). Für zwei Zufallsstichproben X und Y kann die Teststatistik U in kurzer Form durch (7.105) gekennzeichnet werden (hier mit einer Korrektur für mögliche gleiche Werte (Bindungen)).

$$U(X, Y) = \sum_{i=1}^m \sum_{j=1}^n (I[y_j < x_i] + \frac{1}{2} I[y_j = x_i]) \quad (7.105)$$

Darin ist $I[\dots]$ eine so genannte **Indikatorfunktion**, die den Wert 1 annimmt, wenn die Bedingung erfüllt ist und sonst den Wert 0 hat. Erwartungswert und Varianz von U sind in (7.106) angegeben.

$$\begin{aligned} E[U] &= \frac{mn}{2} \\ \text{Var}[U] &= \frac{mn(m+n+1)}{12} \end{aligned} \quad (7.106)$$

Für den Fall, dass **keine Bindungen** auftreten, kann die Verteilung der Teststatistik auf der Basis der möglichen Anordnungen, die alle zum gleichen Wert von U führen, exakt berechnet werden.

$$P(U \leq u | m, n) = \frac{A(u | N, m)}{\binom{N}{m}} \quad (7.107)$$

Darin ist $A(\dots)$ die Anzahl möglicher Anordnungen von m Werten in der Stichprobe X und n Werten in der Stichprobe Y , deren zugehöriger Wert U nicht größer ist als u ($N = n + m$). A kann rekursiv wie folgt bestimmt werden:

$$A(u | N, m) = A(u | N - 1, m) + A(u - n | N - 1, m - 1)$$

$$\text{mit } A(u | N, m) = 0 \quad \text{für } u < 0$$

$$A(u | N, m) = \binom{N}{m} \quad \text{für } u \geq mn$$

$$\text{und } A(u | N, m) = (u + 1) \quad \text{für } m = 1 \text{ oder } n = 1 \text{ und } 0 \leq u \leq mn$$

In **R** kann die Wilcoxon-Verteilung mit den Funktionen **dwilcox()** und **pwilcox()** berechnet werden. Abbildung 7.14 zeigt die Wilcoxon-Verteilung für den Fall zweier Stichproben vom Umfang $m = 3$ und $n = 5$. Dargestellt ist hier die **Zufallsvariable R_1** , Summe der Rangzahlen in der 1. Stichprobe. Diese kann Werte im Bereich von 5 bis 21 annehmen (der Erwartungswert ist 13,5).

Die Funktion **qwilcox()** in **R** berechnet **Quantile** zur Wilcoxon-Verteilung (U-Statistik), die hinsichtlich einer **Testentscheidung an Hand der Rangsummen** leicht umgeformt werden können. Im folgenden Beispiel werden die unteren und oberen Quantile für $m = 2, \dots, 10$ und $n = 10$, bezogen auf eine zweiseitige Fragestellung mit $\alpha = 0,05$ berechnet und in die entsprechenden Rangsummen übertragen.

```
> m <- 2:10; n <- 10
> utab.l <- qwilcox(0.025, m, n, lower.tail=TRUE); rtab.l <- utab.l + m*(m+1)/2
> utab.u <- qwilcox(0.975, m, n, lower.tail=TRUE); rtab.u <- utab.u + m*(m+1)/2
> utab.l; utab.u           # untere / obere Quantile zur U-Statistik
[1] 1 4 6 9 12 15 18 21 24 [1] 19 26 34 41 48 55 62 69 76
> rtab.l; rtab.u           # untere / obere Quantile zu den Rangsummen
[1] 4 10 16 24 33 43 54 66 79 [1] 22 32 44 56 69
83 98 114 131
```

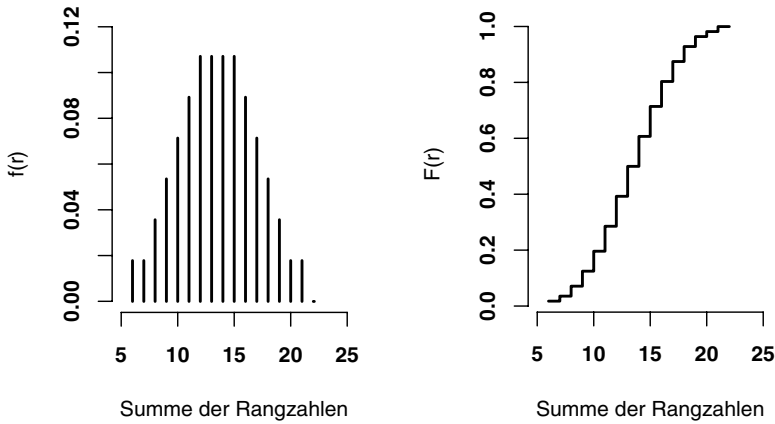


Abb. 7.14. Dichte- und Verteilungsfunktion der Wilcoxon-Verteilung für zwei Stichproben vom Umfang $m = 3$ und $n = 5$ (hier für die Rangsumme R_1)

Die Umrechnung zwischen der Teststatistik U und den entsprechenden Rangsummen R kann allgemein auch nach (7.108) erfolgen.

$$\begin{aligned} R_1 &= U_2 + \frac{m(m+1)}{2} \\ R_2 &= U_1 + \frac{n(n+1)}{2} \end{aligned} \quad (7.108)$$

Für größere Stichprobenumfänge ($m + n > 60$) gilt die ausgezeichnete Approximation

$$U(m, n; \alpha) = \frac{nm}{2} - z \cdot \sqrt{\frac{nm(n+m+1)}{12}} \quad (7.109)$$

Geeignete Werte z sind für die zwei- und die einseitige Fragestellung aus den Quantilen der Standardnormalverteilung zu bestimmen. Anstatt (7.109) benutzt man dann, wenn man ein festes α nicht vorgeben kann oder will bzw. wenn keine Tafeln der kritischen Werte $U(m, n; \alpha)$ zur Verfügung stehen und sobald die Stichprobenumfänge nicht zu klein sind ($m \geq 8, n \geq 8$; Mann und Whitney [MW47]), die Approximation (7.110).

$$\hat{z} = \frac{\left| U - \frac{mn}{2} \right|}{\sqrt{\frac{mn(m+n+1)}{12}}} \quad (7.110)$$

(7.110) lässt sich mit den Rangsummen R_1 und R_2 aus den beiden Stichproben auch (7.111) schreiben:

$$\hat{z} = \frac{\left| (R_1/m) - (R_2/n) \right|}{\sqrt{\left[\frac{(m+n)^2 - 1}{12} \right] \cdot \left[\frac{1}{m} + \frac{1}{n} \right] \cdot \left[\frac{m+n}{(m+n)-1} \right]}} = \frac{|\bar{R}_1 - \bar{R}_2|}{\sqrt{\frac{(m+n)^2(m+n+1)}{12mn}}} \quad (7.111)$$

$$U_1 = 8 \cdot 8 + \frac{8(8+1)}{2} - 89 = 11$$

$$U_2 = 8 \cdot 8 + \frac{8(8+1)}{2} - 47 = 53$$

Kontrolle: $U_1 + U_2 = 64 = 8 \cdot 8 = mn$, da $U_1 = 11 < 53 = U_2$ ist, ist U_1 die Prüfgröße. Da $11 < 15 = U(8,8; 0,05; \text{einseitiger Test})$ (Tabelle 7.28) ist, wird die Nullhypothese $\mu_A \leq \mu_B$ auf dem 5%-Niveau abgelehnt, d. h. die Alternativhypothese $\mu_A > \mu_B$ wird akzeptiert. Nach (7.110) und (7.111) ergibt sich mit

$$\hat{z} = \frac{\left| 11 - \frac{8 \cdot 8}{2} \right|}{\sqrt{\frac{8 \cdot 8(8+8+1)}{12}}} = 2,205 > 1,645 = z_{0,95}$$

$$\hat{z} = \frac{(89/8) - (47/8)}{\sqrt{\left[\frac{16^2 - 1}{12} \right] \cdot \left[\frac{1}{8} + \frac{1}{8} \right] \cdot \left[\frac{16}{16-1} \right]}} = \frac{211,125 - 5,875}{\sqrt{\frac{(8+8)^2(8+8+1)}{12 \cdot 8 \cdot 8}}} = 2,205$$

dieselbe Entscheidung.

Eine entsprechende schrittweise elementare Berechnung der Daten in **R** verdeutlicht nochmals das Prinzip des U-Tests und zeigt beispielhaft die Verwendung einiger interessanter Funktionen, wie **rank()** für die Rangzahlen, **matrix()** für den Aufbau einer Matrix, **dimnames()** zur Festlegung von Namen für Zeilen oder Spalten in einer Matrix sowie die Funktion **sum()**, mit der hier eine Summenbildung zu ausgewählten Teilgruppen durchgeführt wird.

```
> A <- c(7, 14, 22, 36, 40, 48, 49, 52); n1 <- length(A)
> B <- c(3, 5, 6, 10, 17, 18, 20, 39); n2 <- length(B)
>
> All <- c(A, B) # verbinden der Stichproben
> grp <- c(rep(1, n1), rep(2, n2)) # kennzeichnen der Gruppe
> rnk <- rank(All) # zuordnen der Rangzahlen
>
> xdata <- matrix(c(grp, All, rnk), ncol=3) # Aufbau der Matrix
> Namen <- c("Gruppe", "Wert", "Rang") # Namen der Datenspalten
> dimnames(xdata) <- list(NULL, Namen); t(xdata)
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] ...
Gruppe  1   1   1   1   1   1   1   1   2   2   2   2   2 ...
Wert    7  14  22  36  40  48  49  52  3   5   6  10  17 ...
Rang    4   6  10  11  13  14  15  16  1   2   3   5   7 ...
>
> data <- as.data.frame(xdata); attach(data)
> r1 <- sum(Rang[Gruppe==1]); r1
[1] 89
> r2 <- sum(Rang[Gruppe==2]); r2
[1] 47
>
> u1 <- r2 - n2*(n2+1)/2; u1
[1] 11
> u2 <- r1 - n1*(n1+1)/2; u2
[1] 53
```

Für die Berechnung des Wilcoxon-Rangsummentests (U-Tests) steht in **R** die spezielle Funktion **wilcox.test()** zu Verfügung, die einerseits den Wert für U (hier den größeren der beiden Werte), und andererseits auch einen exakten P-Wert aus der Wilcoxon-Verteilung angibt, solange die Stichproben weniger als 50 Werte enthalten und **keine Bindungen in den Rangzahlen** aufgrund gleicher Messwerte auftreten. Anderenfalls wird die Approximation mit der Standardnormalverteilung entsprechend (7.110) oder (7.111) gerechnet.

```
> wilcox.test(A, B, alternative="greater")

Wilcoxon rank sum test

data: A and B W = 53, p-value = 0.01406
alternative hypothesis: true mu is greater than 0
```



Beispiel 2: Gegeben $m = 6$ Beobachtungen des Typs A und $n = 5$ Beobachtungen des Typs B. Prüfung auf dem 5%-Niveau bei zweiseitiger Fragestellung: Zunächst die gemeinsam ansteigend geordneten Beobachtungen mit Herkunft und Rangzahl

63	68	70	81	91	92	95	96	97	99	104
A	A	A	A	B	B	B	B	A	B	A
1	2	3	4	5	6	7	8	9	10	11

Die Summen der Rangzahlen betrage für A: $R_1 = 30$ und für B: $R_2 = 36$.

$$U_1 = 6 \cdot 5 + \frac{6(6+1)}{2} - 30 = 21$$

$$U_2 = 6 \cdot 5 + \frac{5(5+1)}{2} - 36 = 9$$

$$30 = 6 \cdot 5 \text{ (Kontrolle)}$$

Die kleinere der beiden Prüfgrößen, d. h. $U_2 = 9$ dient als Prüfgröße U .

Für $U \leq U(m; n; 0,05; \text{zweiseitig})$ (vgl. Tab. 7.28) wird H_0 auf dem 5%-Niveau verworfen. Da $U = 9 > 3 = U(6; 5; 0,05; \text{zweiseitig})$, ist H_0 auf dem 5%-Niveau nicht abzulehnen.

```
> A <- c(63, 68, 70, 81, 97, 104)
> B <- c(91, 92, 95, 96, 99)
>
> wilcox.test(A, B, alternative="two.sided")

Wilcoxon rank sum test

data: A and B W = 9, p-value = 0.329
alternative hypothesis: true mu is not equal to 0
```

7.4.6.1 Der U -Test bei Rangaufteilung

Kommt bei zwei Stichproben, deren Elemente der Größe nach in eine Reihe gebracht werden, ein bestimmter Wert mehrfach vor – wir sprechen von einer **Bindung** –, dann erhalten die numerisch gleich großen Einzelwerte die mittlere Rangzahl. Beispielsweise [zweiseitige Fragestellung auf dem 5%-Niveau] für

Wert	3	3	4	5	5	5	5	8	8	9	10	13	13	13	15	16
Stichprobe	B	B	B	B	B	A	A	A	B	A	B	A	A	A	A	B
Rangzahl	1,5	1,5	3	5,5	5,5	5,5	5,5	8,5	8,5	10	11	13	13	13	15	16

erhalten die ersten beiden B-Werte die Rangzahl $(1 + 2)/2 = 1,5$; die 4 Fünfen jeweils den Wert $5,5 = (4 + 5 + 6 + 7)/4$; für die beiden Achten erhält man dann 8,5; der Wert 13 kommt dreimal vor und erhält die Rangzahl $\frac{12 + 13 + 14}{3} = 13$.

Bindungen beeinflussen den Wert U nur dann, wenn sie **zwischen den beiden Stichproben** auftreten, nicht aber, wenn sie innerhalb einer oder innerhalb beider Stichproben beobachtet werden. Sind in beiden Stichproben Beobachtungswerte einander gleich, dann lautet die korrigierte Formel für den U -Test bei Rangaufteilung:

$$\hat{z} = \frac{\left| U - \frac{mn}{2} \right|}{\sqrt{\left[\frac{mn}{S(S-1)} \right] \cdot \left[\frac{S^3 - S}{12} - \sum_{i=1}^{i=r} \frac{t_i^3 - t_i}{12} \right]}} \quad \text{mit } S = m + n \quad (7.112)$$

In dem Korrekturglied $\sum_{i=1}^r (t_i^3 - t_i)/12$ bezeichnet r die Anzahl der Bindungen, t_i ist die Vielfachheit der i -ten Bindung.

Für jede Gruppe ($i = 1$ bis $i = r$) ranggleicher Werte bestimmen wir, wie oft ein Wert t erscheint und bilden $(t^3 - t)/12$. Die Summe dieser r Quotienten bildet das Korrekturglied.

Für das obige Beispiel ergibt sich aus $r = 4$ Gruppen von Bindungen das Korrekturglied nach:

Gruppe 1: $t_1 = 2$: zweimal der Wert 3 mit dem Rang 1,5

Gruppe 2: $t_2 = 4$: viermal der Wert 5 mit dem Rang 5,5

Gruppe 3: $t_3 = 2$: zweimal der Wert 8 mit dem Rang 8,5

Gruppe 4: $t_4 = 3$: dreimal der Wert 13 mit dem Rang 13

$$\begin{aligned} \sum_{i=1}^{i=4} \frac{t_i^3 - t_i}{12} &= \frac{2^3 - 2}{12} + \frac{4^3 - 4}{12} + \frac{2^3 - 2}{12} + \frac{3^3 - 3}{12} \\ &= \frac{6}{12} + \frac{60}{12} + \frac{6}{12} + \frac{24}{12} = 8,00 \end{aligned}$$

$$A: m = 8, R_1 = 83,5 \quad B: n = 8, R_2 = 52,5$$

$$U_1 = 8 \cdot 8 + \frac{8(8+1)}{2} - 83,5 = 16,5 \quad U_2 = 8 \cdot 8 + \frac{8(8+1)}{2} - 52,5 = 47,5$$

$$U_1 + U_2 = 64 = mn \quad \text{d.h.} \quad \hat{z} = \frac{\left| 16,5 - \frac{8 \cdot 8}{2} \right|}{\sqrt{\left[\frac{8 \cdot 8}{16(16-1)} \right] \cdot \left[\frac{16^3 - 16}{12} - 8,00 \right]}} = 1,647$$

Da $1,65 < 1,96 = z_{0,975}$ ist, kann bei zweiseitiger Fragestellung ($\alpha = 0,05$) die Nullhypothese (etwa: $\tilde{\mu}_A = \tilde{\mu}_B$) nicht abgelehnt werden.

```
> A <- c(5, 5, 8, 9, 13, 13, 13, 15)
> B <- c(3, 3, 4, 5, 5, 8, 10, 16)
>
> wilcox.test(A, B, alternative="two.sided")
```

Wilcoxon rank sum test with continuity correction

```
data: A and B W = 47.5, p-value = 0.1109
alternative hypothesis: true mu is not equal to 0
```

```
Warning message: Cannot compute exact p-value with ties in:
wilcox.test.default(A, B, alternative = "two.sided")
```

Die Funktion **wilcox.test()** in **R** berechnet in diesem Beispiel bei Auftreten von Bindungen (und bei Stichprobenumfängen über 50) eine Approximation durch die Normalverteilung mit Kontinuitätskorrektur. Exakte Verfahren stehen in **library(exactRankTests)** [HH05], hier insbesondere mit der Funktion **wilcox.exact()** zur Verfügung.

```
> library(exactRankTests)
> wilcox.exact(A, B, alternative="two.sided")
```

Exact Wilcoxon rank sum test

data: A and B W = 47.5, p-value = 0.1071
 alternative hypothesis: true mu is not equal to 0



Der U -Test ist eines der schärfsten nichtparametrischen Prüfverfahren. Da die Teststatistik U eine ziemlich komplizierte Funktion des Mittelwertes, der Wölbung und der Schiefe ist – der U -Test also nicht lediglich μ_i oder $\tilde{\mu}_i$ sondern mittlere Ränge vergleicht –, muss betont werden, dass mit zunehmendem Verteilungsform-Unterschied der beiden Grundgesamtheiten, die Signifikanzschranken (hinsichtlich der Hypothese auf Unterschiede zweier Parameter allein) unzuverlässig werden. Die Power des U -Tests braucht für endliches n nicht kleiner zu sein als die des t -Tests und des Permutationstests, die eine fast gleiche Power aufweisen (t -Test-Voraussetzungen erfüllt).

Hinweis: Konfidenzintervall für Median-Differenzen. Mit Hilfe des U -Tests lässt sich ein Vertrauensbereich für die Differenz zweier Mediane angeben $\tilde{\mu}_1 - \tilde{\mu}_2 = \Delta$, mit $\tilde{\mu}_1 > \tilde{\mu}_2$: $k_{\min} < \Delta < k_{\max}$.

Hierzu: (1) addiert man eine Konstante k zu allen Werten der 2. Stichprobe und führt mit dieser und der 1. Stichprobe einen U -Test durch; (2) linke und rechte Schranke des Vertrauensbereiches für Δ sind der kleinste und der größte Wert k ($k_{\min}, k_{\max}$), die bei zweiseitiger Fragestellung auf dem gewählten Signifikanzniveau die Nullhypothese des U -Tests nicht abzulehnen gestatten; (3) geeignete extreme Werte k , die gerade noch zu einem nichtsignifikanten Ergebnis führen, erhält man durch geschicktes Probieren (etwa mit $k = 0, 1; k = 1; k = 10$ beginnen). Eine gründliche Übersicht gibt van der Laan [Laa70].

7.4.7 Wilcoxon-Paardifferenzentest

Optimale Tests für den Vergleich zweier verbundener Stichproben, für den Vergleich gepaarter Beobachtungen, sind der t -Test bei normalverteilten Differenzen (vgl. [7.4.5]) und der Vorzeichen-Rang-Test von Wilcoxon (**Wilcoxon matched pairs signed rank test**) bei nicht normalverteilten Differenzen. Dieser Test, als Wilcoxon-Test für Paardifferenzen bekannt, kann auch auf Rangdaten angewendet werden. Er erfordert, verglichen mit dem t -Test, wesentlich weniger Rechenarbeit und testet normalverteilte Differenzen fast ebenso scharf; seine Wirksamkeit, Effizienz, liegt für große und kleine Stichprobenumfänge bei 95%. Der Vertrauensbereich für den Median der Paardifferenzen wird nach dem Test behandelt.

Der Test gestattet die Prüfung, ob die **Differenzen paarig angeordneter Beobachtungen symmetrisch mit dem Median gleich Null verteilt** sind, d. h., unter der Nullhypothese entstammen die Paardifferenzen d_i einer Grundgesamtheit mit der Verteilungsfunktion $F(d)$ bzw. mit der Dichte $f(d)$, wobei:

$$H_0 : F(+d) + F(-d) = 1 \quad \text{bzw.} \quad f(+d) = f(-d)$$

Wird H_0 abgelehnt, so ist entweder die Grundgesamtheit nicht symmetrisch in bezug auf den Median, d. h. der Median der Differenzen ist ungleich Null ($\tilde{\mu}_d \neq 0$) oder den beiden Stichproben liegen unterschiedliche Verteilungen zugrunde. Von Paaren mit gleichen Einzelwerten abgesehen, bildet man für die restlichen n Wertepaare die Differenzen

$$d_i = x_{i1} - x_{i2} \quad (7.113)$$

und bringt die absoluten Beträge $|d_i|$ in eine ansteigende Rangordnung: Der kleinste erhält die Rangzahl 1, ..., und der größte die Rangzahl n . Bei gleichgroßen Beträgen werden mittlere Rangzahlen zugeordnet. Bei jeder Rangzahl wird vermerkt, ob die zugehörige Differenz ein positives oder ein negatives Vorzeichen aufweist. Man bildet die Summe der positiven und der negativen Rangzahlen ($\hat{R}_p$ und $\hat{R}_n$), kontrolliert sie nach

$$\hat{R}_p + \hat{R}_n = n(n+1)/2 \quad (7.114)$$

und benutzt als Testgröße **die kleinere der beiden Rangsummen** $\hat{R} = \min(\hat{R}_p, \hat{R}_n)$. Die Nullhypothese wird verworfen, wenn der berechnete $\hat{R}$ -Wert kleiner oder gleich dem kritischen Wert $R(n; \alpha)$ der Tabelle 7.29 ist. Für $n > 25$ gilt die Approximation

$$R(n; \alpha) = \frac{n(n+1)}{4} - z \cdot \sqrt{\frac{1}{24}n(n+1)(2n+1)} \quad (7.115)$$

Geeignete Werte z lassen sich aus den Quantilen für die Standardnormalverteilung ableiten. Anstatt (7.115) benutzt man dann, wenn man ein festes α nicht vorgeben kann oder will (und $n > 25$), die äquivalente Schreibweise (7.116).

$$\hat{z} = \frac{\left| \hat{R} - \frac{n(n+1)}{4} \right|}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \quad (7.116)$$

Beispiel: Ein Biochemiker prüft an 9 Probanden, ob sich die im Serum um 9.00 und um 18.00 Uhr bestimmten Konzentrationen des Metaboliten M in mmol/l bei zweiseitiger Fragestellung auf dem 5%-Niveau unterscheiden ($H_0: \tilde{\mu}_d = 0$; $H_A: \tilde{\mu}_d \neq 0$). 

Proband	1	2	3	4	5	6	7	8	9
9 ⁰⁰	0,47	1,02	0,33	0,70	0,94	0,85	0,39	0,52	0,47
18 ⁰⁰	0,41	1,00	0,46	0,61	0,84	0,87	0,36	0,52	0,51
Differenz d_i	0,06	0,02	-0,13	0,09	0,10	-0,02	0,03	0	-0,04
Rang zu $ d_i $	5	1,5	8	6	7	1,5	3		4
$\hat{R}_p$	(+)5	(+)1,5		(+)6	(+)7		(+)3		
$\hat{R}_n$			(-)8			(-)1,5			(-)4
Kontrolle	22,5 + 13,5 = 36 = 8(8+1)/2 d.h. $\hat{R} = 13,5$								

Da $13,5 > 3 = R(8; 0,05)$, kann die Nullhypothese nicht abgelehnt werden.

Hinweise:

- Treten gehäuft **Bindungen** auf, so wird in (7.115) bzw. (7.116) die Wurzel $\sqrt{A}$ durch $\sqrt{A - B/48}$ mit $B = \sum_{i=1}^{i=r} (t_i^3 - t_i)/12$ ersetzt [r = Anzahl der Bindungen, t_i = Vielfachheit der i -ten Bindung].
- Um **Dispersionsunterschiede** zu erfassen, bilde man $D_i = |x_{i1} - \bar{x}_1| - |x_{i2} - \bar{x}_2|$ und $R(|D_i|)$ sowie $\hat{R}_p$ und $\hat{R}_n$ und prüfe 2- oder 1-seitig.

In **R** wird der Wilcoxon-Test für Paardifferenzen mit der Funktion **wilcox.test()** nur dann exakt berechnet, wenn keine Bindungen auftreten. Anderenfalls, wie auch im vorangehenden Beispiel, erfolgt die Berechnung nach der Approximation mit der Standardnormalverteilung.

```
> M1 <- c(0.47, 1.02, 0.33, 0.70, 0.94, 0.85, 0.39, 0.52, 0.47)
> M2 <- c(0.41, 1.00, 0.46, 0.61, 0.84, 0.87, 0.36, 0.52, 0.51)
> D <- M1 - M2; D
[1] 0.06 0.02 -0.13 0.09 0.10 -0.02 0.03 0.00 -0.04
> wilcox.test(M1, M2, alternative="two.sided", paired=TRUE)
```

Wilcoxon signed rank test with continuity correction

data: M1 and M2 V = 22.5, p-value = 0.5749 alternative hypothesis:
true mu is not equal to 0 ...

Tabelle 7.29. Kritische Werte für den Wilcoxon-Paardifferenzen-Test (auszugsweise entnommen aus McCornack, R.L.: Extended tables of the Wilcoxon matched pair signed rank statistic. J. Amer. Statist. Assoc. **60** (1965), 864–871, 866 + 867). Beachtet sei, dass z. B. die einseitige 5%-Schranke zugleich zweiseitige 10%-Schranke ist und die zweiseitige 1%-Schranke zugleich einseitig 0,5%-Schranke ist

Test	zweiseitig			einseitig		Test	zweiseitig			einseitig	
<i>n</i>	5%	1%	0,1%	5%	1%	<i>n</i>	5%	1%	0,1%	5%	1%
6	0			2		54	514	445	368	550	473
7	2			3	0	55	536	465	385	573	493
8	3	0		5	1	56	557	484	402	595	514
9	5	1		8	3	57	579	504	420	618	535
10	8	3		10	5	58	602	525	438	642	556
11	10	5	0	13	7	59	625	546	457	666	578
12	13	7	1	17	9	60	648	567	476	690	600
13	17	9	2	21	12	61	672	589	495	715	623
14	21	12	4	25	15	62	697	611	515	741	646
15	25	15	6	30	19	63	721	634	535	767	669
16	29	19	8	35	23	64	747	657	556	793	693
17	34	23	11	41	27	65	772	681	577	820	718
18	40	27	14	47	32	66	798	705	599	847	742
19	46	32	18	53	37	67	825	729	621	875	768
20	52	37	21	60	43	68	852	754	643	903	793
21	58	42	25	67	49	69	879	779	666	931	819
22	65	48	30	75	55	70	907	805	689	960	846
23	73	54	35	83	62	71	936	831	712	990	873
24	81	61	40	91	69	72	964	858	736	1020	901
25	89	68	45	100	76	73	994	884	761	1050	928
26	98	75	51	110	84	74	1023	912	786	1081	957
27	107	83	57	119	92	75	1053	940	811	1112	986
28	116	91	64	130	101	76	1084	968	836	1144	1015
29	126	100	71	140	110	77	1115	997	862	1176	1044
30	137	109	78	151	120	78	1147	1026	889	1209	1075
31	147	118	86	163	130	79	1179	1056	916	1242	1105
32	159	128	94	175	140	80	1211	1086	943	1276	1136
33	170	138	102	187	151	81	1244	1116	971	1310	1168
34	182	148	111	200	162	82	1277	1147	999	1345	1200
35	195	159	120	213	173	83	1311	1178	1028	1380	1232
36	208	171	130	227	185	84	1345	1210	1057	1415	1265
37	221	182	140	241	198	85	1380	1242	1086	1451	1298
38	235	194	150	256	211	86	1415	1275	1116	1487	1332
39	249	207	161	271	224	87	1451	1308	1146	1524	1366
40	264	220	172	286	238	88	1487	1342	1177	1561	1400
41	279	233	183	302	252	89	1523	1376	1208	1599	1435
42	294	247	195	319	266	90	1560	1410	1240	1638	1471
43	310	261	207	336	281	91	1597	1445	1271	1676	1507
44	327	276	220	353	296	92	1635	1480	1304	1715	1543
45	343	291	233	371	312	93	1674	1516	1337	1755	1580
46	361	307	246	389	328	94	1712	1552	1370	1795	1617
47	378	322	260	407	345	95	1752	1589	1404	1836	1655
48	396	339	274	426	362	96	1791	1626	1438	1877	1693
49	415	355	289	446	379	97	1832	1664	1472	1918	1731
50	434	373	304	466	397	98	1872	1702	1507	1960	1770
51	453	390	319	486	416	99	1913	1740	1543	2003	1810
52	473	408	335	507	434	100	1955	1779	1578	2045	1850
53	494	427	351	529	454						

Eine exakte Berechnung ist mit der Funktion **wilcox.exact()** aus dem Paket **exactRankTests** [HH05] möglich:

```
> library(exactRankTests)
> wilcox.exact(M1, M2, alternative="two.sided", paired=TRUE)

Exact Wilcoxon signed rank test

data: M1 and M2 V = 22.5, p-value = 0.5703 alternative
hypothesis: true mu is not equal to 0
```

7.4.7.1 Konfidenzintervall für den Median $\tilde{\mu}_d$ der Paardifferenzen

Das 95%-KI für $\tilde{\mu}_d$ ergibt sich aus den n geordneten Differenzen d_i , indem von den $n(n+1)/2$ möglichen Differenzen $(d_{i'} + d_{i''})/2$ für alle Paare i' und i'' , einschließlich $i' = i''$ die k unteren und die k oberen gebildet werden. Den Wert k berechnet man nach (7.117) anhand der Werte $R_{n;0,05;zweiseitig}$ aus Tabelle 7.29

$$k = 1 + R_{n;0,05;zweiseitig} \quad (7.117)$$

Dann bilden die k -ten Differenzen – jeweils von unten und von oben gerechnet – das 95%-KI für $\tilde{\mu}_d$, das hier eine Vertrauenswahrscheinlichkeit von mindestens 0,95 aufweist ($P \geq 0,95$).

Für das 90%-KI mit $P \geq 0,90$ wird in (7.117) der Tabellenwert $R_{n;0,05;einseitig}$ eingesetzt.

Für $n > 100$ berechnet man k nach (7.115) mit $z = 1,960$ (95%-KI) bzw. $z = 1,645$ (90%-KI); etwa für

$$n = 100 \quad \text{und} \quad \alpha_{zweiseitig} = 0,05$$

$$\begin{aligned} R(100; 0,05) &= \frac{100(100+1)}{4} - 1,960 \cdot \sqrt{\frac{1}{24} 100(100+1)(2 \cdot 100 + 1)} \\ &= 1954,955 \text{ oder } 1955 \end{aligned}$$

wie in Tabelle 7.29.

Beispiel: 95%-KI für $\tilde{\mu}_d$: Für unsere neun geordneten Differenzen aus dem vorangehenden Beispiel $d_i: -0,13 \ -0,04 \ -0,02 \ 0 \ 0,02 \ 0,03 \ 0,06 \ 0,09 \ 0,10$ ergibt sich nach Tabelle 7.29 der Wert $k = 1 + 5 = 6$. Die geordneten mittleren Differenzen sind



<i>von unten</i>	<i>von oben</i>
$[(-0,13) + (-0,13)]/2 = -0,13$	$(0,10 + 0,10)/2 = 0,10$
$[(-0,13) + (-0,04)]/2 = -0,085$	$(0,10 + 0,09)/2 = 0,095$
$[(-0,13) + (-0,02)]/2 = -0,075$	$(0,09 + 0,09)/2 = 0,09$
$[(-0,13) + 0]/2 = -0,065$	$(0,10 + 0,06)/2 = 0,08$
$[(-0,13) + (0,02)]/2 = -0,055$	$(0,09 + 0,06)/2 = 0,075$
$[(-0,13) + (0,03)]/2 = \underline{-0,05}$	$(0,10 + 0,03)/2 = \underline{0,065}$
$[(-0,04) + (-0,04)]/2 = -0,04$	$(0,10 + 0,02)/2 = 0,06$
$[(-0,13) + (0,06)]/2 = -0,035$	$(0,09 + 0,03)/2 = 0,06$
	$(0,06 + 0,06)/2 = 0,06 \text{ usw.}$

Somit folgt: $P(-0,050 \leq \tilde{\mu}_d \leq 0,065) \geq 0,95$. Da die Null miteingeschlossen ist, wird das Testresultat bestätigt.

7.4.7.2 Der Maximum-Test für Paardifferenzen



Der Maximum-Test ist ein sehr einfacher Test für den Vergleich zweier gepaarter Messreihen. Man braucht sich nur zu merken, dass – wenn die 5 absolut größten Differenzen das gleiche Vorzeichen haben – der Unterschied auf dem 10%-Niveau statistisch gesichert ist. Bei 6 Differenzen dieser Art ist der Unterschied auf dem 5%-Niveau statistisch signifikant, bei 8 Differenzen auf dem 1%-Niveau und bei 11 Differenzen auf dem 0,1%-Niveau. Diese Zahlen 5, 6, 8 und 11 gelten bei zweiseitiger Fragestellung für Stichprobenumfänge von $n \geq 6$. Bei einseitiger Fragestellung entsprechen diesen Zahlen natürlich die 5%-, 2,5%-, 0,5%- und 0,05%-Schranken; treten zwei dem Absolutbetrag nach gleich große Differenzen mit verschiedenen Vorzeichen auf, so ordne man sie, um sicherzugehen, so ein, dass eine eventuell bestehende Folge gleicher Vorzeichen verkleinert wird (Walter 1951 [Wal51]). Der Maximum-Test dient zur **unabhängigen Kontrolle** des t -Tests, ohne ihn jedoch zu ersetzen [Wal58].



Beispiel: Die Folge der Differenzen $+3,4; +2,0; +1,6; +1,0; +0,7; +0,5; -0,3; +0,3$ – beachte die ungünstigere Anordnung von $-0,3$ – führt bei zweiseitiger Fragestellung mit 6 typischen Differenzen auf dem 5%-Niveau zur Ablehnung der $H_0: \tilde{\mu}_d = 0$.

7.4.7.3 Der Vorzeichentest von Dixon und Mood

Der Vorzeichen-Rangtest für den Median nach Wilcoxon ist in [7.4.6] näher dargestellt. Er ist dem Vorzeichentest überlegen. Die Nullhypothese des Vorzeichentests lautet:

$$P(X > Y) = P(X < Y) \quad \text{d.h.} \quad P(X > Y) = \frac{1}{2} \quad \text{und} \quad P(X < Y) = \frac{1}{2}$$

Der Name des Tests rührt daher, dass nur die Vorzeichen von Differenzen zwischen Beobachtungswerten gewertet werden. Vorausgesetzt wird die Stetigkeit der Zufallsvariablen. Der Test dient in erster Linie als **Schnelltest zur Prüfung des Unterschieds der zentralen Tendenz zweier verbundener Stichproben** [DM46]. Die einzelnen Paare brauchen – im Unterschied zum t -Test und zum Wilcoxon-Test – nicht einer gemeinsamen Grundgesamtheit zu entstammen; sie können beispielsweise hinsichtlich Alter, Geschlecht usw. verschiedenen Grundgesamtheiten angehören. Wesentlich ist, dass die Ergebnisse der einzelnen Paare unabhängig voneinander sind. Die Nullhypothese des Vorzeichentests lautet: **Die Differenzen gepaarter Beobachtungen unterscheiden sich im Durchschnitt nicht von Null**; man erwartet, dass etwa die Hälfte der Differenzen kleiner als Null ist, also ein negatives Vorzeichen aufweist und die andere Hälfte größer als Null ist, also ein positives Vorzeichen aufweist. Der Vorzeichentest prüft damit die Nullhypothese: **die Verteilung der Differenzen hat den Median Null**.

Tabelle 7.30. Wertepaare $n; h$ für den Vorzeichentest ($\alpha = 0,05$). Um zumindest auf dem 5%-Niveau eine falsche Nullhypothese ($\pi_0 = 0,5$) mit einer Power von wenigstens P entdecken zu können, dürfen von mindestens n Nicht-Null-Differenzen höchstens h das seltenere Vorzeichen aufweisen

Power P	Alternativhypothese mit $\pi =$		
	0,90	0,75	0,60
$0,70 < P \leq 0,75$	10; 1	25; 7	158; 66
$0,80 < P \leq 0,85$	14; 2	30; 9	199; 85
$0,90 < P \leq 0,95$	15; 3	42; 14	263; 115

Schranken oder Vertrauensgrenzen für den Median findet man in Tabelle 6.6. Die Nullhypothese wird abgelehnt, wenn zu wenige oder zu viele Differenzen eines Vorzeichens vorhanden sind, d.h. wenn die Schranken der Tabelle 6.6 unter- oder überschritten werden. Null-Differenzen bleiben dabei unberücksichtigt, der Stichprobenumfang vermindert sich entsprechend.

Die Wahrscheinlichkeit für das Auftreten einer bestimmten Anzahl von Plus- oder Minuszeichen ergibt sich aus der Binomialverteilung für $p = q = 1/2$ (vgl. die oben genannte H_0 des Tests). Die Tabelle 6.6 zeigt, dass mindestens 6 Paare von Beobachtungen vorliegen müssen, wenn bei zweiseitiger Fragestellung ein Ergebnis auf dem 5%-Niveau gesichert sein soll: $n = 6$, $x = 0$ oder 6. Die Power des Tests ist um so größer, je stärker p oder π von dem Wert $\pi_0 = 1/2$ abweicht, was in Tabelle 7.30 gut zum Ausdruck kommt. Die Wirksamkeit des Vorzeichentests sinkt mit zunehmendem Stichprobenumfang von 95% bei $n = 6$ auf 64% bei $n \rightarrow \text{Unendlich}$; d. h. man nutzt ihn im allgemeinen für $6 \leq n \lesssim 40$.

Beispiel: Angenommen, wir beobachten bei zweiseitiger Fragestellung auf dem 5%Niveau 15 Paare, erhalten zwei Nulldifferenzen und 13 Differenzen, von denen 11 das Plus- und 2 das Minuszeichen aufweisen. Aus Tabelle 6.6 ergeben sich für $n = 13$ die Schranken 3 und 10, die - wenn nach außen überschritten - bei zweiseitiger Fragestellung auf dem 5%-Niveau H_0 abzulehnen gestatten. Unsere Werte liegen außerhalb der Grenzen; d. h. $H_0: \tilde{\mu}_d = 0$ wird auf dem 5%-Niveau abgelehnt ($\tilde{\mu}_0 \neq 0$), was auch nach (7.118) gelingt.



Vorzeichentest (Approximation mit der Standardnormalverteilung): Die Zahl der Vorzeichen sei n , das seltenere Vorzeichen trete h mal auf. Dann wird $H_0: \tilde{\mu}_d = 0$ gegen $H_A: \tilde{\mu}_d \neq 0$ für

$$\hat{z} = (|n - 2h| - 1)/\sqrt{n} > 1,96 = z_{0,95} \quad (7.118)$$

auf dem 5%-Niveau abgelehnt, z. B. für $n = 6$, $h = 0$ mit $\hat{z} = 2,04$ für $n = 9$, $h = 1$ mit $\hat{z} = 2$; in beiden Fällen ist somit der Median der Differenzen auf dem 5%-Niveau deutlich von Null verschieden [vgl. auch das Beispiel oben: $\hat{z} = (|13 - 2 \cdot 2| - 1)/\sqrt{13} = 2,22 > 1,96$].

Der benötigte Stichprobenumfang lässt sich nach [Noe87] abschätzen.

(7.118) ohne die „1“ im Zähler: Lehne H_0 auf dem 5%-Niveau ab, sobald $\hat{T} = |n - 2h| > 1,96\sqrt{n} \approx 2\sqrt{n}$.

Diese von Duckworth und Wyatt [DW58] vorgeschlagene Modifikation ist als **Schnellschätzung** brauchbar. Prüfgröße $\hat{T}$ ist die absolut genommene Differenz der Vorzeichen (d. h. |Anzahl der Pluszeichen minus Anzahl der Minuszeichen|). Das 5%-Niveau dieser Differenz ist gegeben durch $2 \cdot \sqrt{n}$, das 10%-Niveau durch $1,6 \cdot \sqrt{n}$ mit n als Gesamtzahl der Vorzeichen gebenden Differenzen. Wenn $\hat{T} > 2 \cdot \sqrt{n}$ oder wenn $\hat{T} > 1,6 \cdot \sqrt{n}$, dann ist bei zweiseitiger Fragestellung der Unterschied als statistisch signifikant anzusehen. Das soeben gegebene Beispiel führt mit $\hat{T} = 11 - 2 = 9$ und $2 \cdot \sqrt{n} = 2 \cdot \sqrt{13} = 7,21$ und damit $9 > 7,21$ auch zur Ablehnung von H_0 auf dem 5%-Niveau [vgl. $\hat{T} = |13 - 2 \cdot 2| = 9 = |13 - 2 \cdot 11|$]. Das 10%-Niveau wird nur in besonderen Fällen genutzt.

Hinweis: Die Nullhypothese des Vorzeichentests lässt sich schreiben $H_0: P(Y > X) = 1/2$. Der Test ist auch anwendbar, wenn unter H_0 ein bestimmter Unterschied angenommen wird. Etwa, Y sei durchschnittlich 10% größer als X (beide positiv) oder Y sei durchschnittlich 5 Einheiten kleiner als X ; d. h. $H_0: P(Y > 1,10X) = 1/2$ bzw. $H_0: P(Y > [X - 5]) = 1/2$. Gezählt werden dann die Vorzeichen der Differenzen $(Y - 1,10X)$ bzw. $(Y - X + 5)$. Entsprechende sogenannte **Minimum-Effekt Nullhypothesen** sind natürlich den reinen Nullhypothesen („kein Effekt“) überlegen (vgl. Murphy, K.R. und Myers, B. [MM98]).

7.4.8 Vergleich zweier unabhängiger Stichproben nach Kolmogoroff und Smirnow

Sind zwei unabhängige Stichproben von Messwerten (oder von Häufigkeitsdaten) hinsichtlich der Frage zu vergleichen, ob sie aus derselben Grundgesamtheit stammen, dann gilt der Test von Kolmogoroff [Kol33] und Smirnow [Smi39] als schärfster Homogenitätstest. Er erfasst **Unterschiede**

der Verteilungsform aller Art: Insbesondere Unterschiede der zentralen Tendenz, der Streuung, der Schiefe und des Exzesses, d. h. Unterschiede der Verteilungsfunktion.

Als Prüfgröße dient die größte zu beobachtende Ordinatendifferenz zwischen den beiden **sich nicht überschneidenden relativierten Summenkurven**. Hierzu werden (bei gleichen Klassengrenzen für beide Stichproben) die empirischen Verteilungsfunktionen $\hat{F}_1$ und $\hat{F}_2$ sowie ihre Differenzen $\hat{F}_1 - \hat{F}_2$ berechnet. Das Maximum der Absolutbeträge dieser Differenzen ist (für die hier hauptsächlich interessierende zweiseitige Fragestellung) die gesuchte Prüfgröße $\hat{D}$ [der maximale Absolutbetrag der Abweichungen der beiden empirischen Verteilungsfunktionen]:

$$\hat{D} = \max \left| \left(\hat{F}_1 - \hat{F}_2 \right) \right| \quad (7.119)$$

Die Prüfverteilung D liegt tabelliert vor: **Kim** [Kim69] [S. 79–170 in den Tabellen von Harter und Owen, Bd. 1 [HO70]].

D kann für mittlere bis große Stichprobenumfänge ($n_1 + n_2 > 35$) durch

$$D_\alpha = K_{(\alpha)} \sqrt{(n_1 + n_2) / (n_1 \cdot n_2)} \quad (7.120)$$

approximiert werden, wobei $K_{(\alpha)}$ eine von der Irrtumswahrscheinlichkeit α abhängige Konstante darstellt:

Tabelle 7.31. Ausgewählte Konstanten Für den Kolmogoroff-Smirnoff Test

α	0,20	0,15	0,10	0,05	0,01	0,001
$K_{(\alpha)}$	1,07	1,14	1,22	1,36	1,63	1,95

Erreicht oder übersteigt ein aus zwei Stichproben ermittelter Wert $\hat{D}$ den kritischen Wert D_α , so liegt auf dem verwendeten Niveau (vgl. Tab. 7.32) hinsichtlich beider Verteilungsfunktionen ein statistisch signifikanter Unterschied vor.

Tabelle 7.32. Einige Werte $D_{n_1; n_2; \alpha}$ für die zweiseitige Fragestellung

$n_1 =$ $n_2 =$	6				9				12			
	6	12	18	24	9	12	15	18	12	16	18	20
$D_{0,10}$	0,667	0,583	0,556	0,500	0,556	0,500	0,489	0,444	0,417	0,438	0,417	0,417
$D_{0,05}$	0,667	0,583	0,611	0,583	0,556	0,556	0,533	0,500	0,500	0,479	0,472	0,467
$D_{0,01}$	0,833	0,750	0,722	0,667	0,667	0,667	0,644	0,611	0,583	0,583	0,556	0,567

$n_1 = n_2 =$	7	8	10	11	13	14	15	16	17	18	19	20
$D_{0,10}$	0,571	0,500	0,500	0,454	0,462	0,429	0,400	0,375	0,412	0,389	0,368	0,350
$D_{0,05}$	0,714	0,625	0,600	0,545	0,462	0,500	0,467	0,438	0,412	0,444	0,421	0,400
D_{001}	0,714	0,750	0,700	0,636	0,615	0,571	0,533	0,563	0,529	0,500	0,473	0,500

$$\hat{D} = \max_x |\hat{F}_{1,n_1}(x) - \hat{F}_{2,n_2}(x)| = \max_x |\hat{P}(X_1 \leq x) - \hat{P}(X_2 \leq x)| \quad (7.121)$$

$H_0 : F_1(x) = F_2(x) \text{ für alle } x ; \quad H_A : F_1(x) \neq F_2(x) \text{ für mindestens ein } x$

Der Test erfasst hauptsächlich **Verteilungsunterschiede**, die im **mittleren Bereich** auftreten. Der Test ist dem U -Test überlegen, wenn aufgrund unterschiedlicher Behandlung homogener Untersuchungseinheiten auch **unterschiedliche Verteilungsformen** erwartet werden.

Beispiel: Es sind zwei Messreihen zu vergleichen. Über mögliche Unterschiede irgendwelcher Art ist nichts bekannt. Wir prüfen die Nullhypothese: Gleichheit beider Verteilungsfunktionen gegen die Alternativhypothese: Beide Verteilungsfunktionen sind ungleich ($\alpha = 0,05$ für die zweiseitige Fragestellung).



Messreihe 1: 2,1 3,0 1,2 2,9 0,6 2,8 1,6 1,7 3,2 1,7

Messreihe 2: 3,2 3,8 2,1 7,2 2,3 3,5 3,0 3,1 4,6 3,2

Die 10 Messwerte jeder Reihe werden der Größe nach geordnet:

Meßreihe 1: 0,6 1,2 1,6 1,7 1,7 2,1 2,8 2,9 3,0 3,2

Meßreihe 2: 2,1 2,3 3,0 3,1 3,2 3,2 3,5 3,8 4,6 7,2

Aus den Häufigkeitsverteilungen (f_1 und f_2) beider Stichproben erhalten wir $\hat{F}_1$ und $\hat{F}_2$ (vgl. Tabelle 7.33).

Tabelle 7.33. Berechnung der Teststatistik zum Kolmogoroff-Smirnoff-Test anhand der Werte aus dem Beispiel

Bereich	0,0 - 0,9	1,0 - 1,9	2,0 - 2,9	3,0 - 3,9	4,0 - 4,9	5,0 - 5,9	6,0 - 6,9	7,0 - 7,9
f_1	1	4	3	2	0	0	0	0
f_2	0	0	2	6	1	0	0	1
$\hat{F}_1$	1/10	5/10	8/10	10/10	10/10	10/10	10/10	10/10
$\hat{F}_2$	0/10	0/10	2/10	8/10	9/10	9/10	9/10	10/10
$\hat{F}_1 - \hat{F}_2$	1/10	5/10	6/10	2/10	1/10	1/10	1/10	0

Als absolut größte Differenz erhalten wir mit $\hat{D} = 6/10$ einen Wert, der den kritischen Wert $D_{10;10;0,05} = 0,600$ (zweiseitige Fragestellung, Tabelle 7.8) gerade erreicht, folglich ist die Homogenitätshypothese auf dem 5%-Niveau abzulehnen: Anhand der vorliegenden Stichproben besteht keine Veranlassung, eine beiden gemeinsame Grundgesamtheit anzunehmen.

```
> m1 <- c(2.1, 3.0, 1.2, 2.9, 0.6, 2.8, 1.6, 1.7, 3.2, 1.7)
> m2 <- c(3.2, 3.8, 2.1, 7.2, 2.3, 3.5, 3.0, 3.1, 4.6, 3.2)
> ks.test(m1, m2, alternative="two.sided")
```

Two-sample Kolmogorov-Smirnov test

```
data: m1 and m2
D = 0.6, p-value = 0.05465
alternative hypothesis: two.sided
```

In **R** kann der Kolmogoroff-Smirnoff Test für die beiden Messreihen aus dem Beispiel durch die Funktion **ks.test()** berechnet werden. Das Ergebnis $D = 0,6$ stimmt mit dem in der Tabelle hergeleiteten Resultat überein. Allerdings erfolgt die Bestimmung exakter P-Werte (für $n < 1000$) in dieser Funktion nur dann, wenn keine Bindungen in den Messwerten auftreten. Anderenfalls wird eine asymptotische Approximation durchgeführt, die insbesondere bei kleinen Stichprobengrößen fehlerhaft sein kann. Besser ist dann die Bezugnahme auf die genannten Tabellenwerte.

Hinweis: Auf den einseitigen KS-Test [Formel (7.120) mit $K_{0,10} = 1,07$ bzw. $K_{0,05} = 1,22$ bzw. $K_{0,01} = 1,52$] gehen wir hier nicht näher ein, da er bei gleichen Verteilungsformen dem einseitigen U -Test von Wilcoxon, Mann und Whitney unterlegen ist.

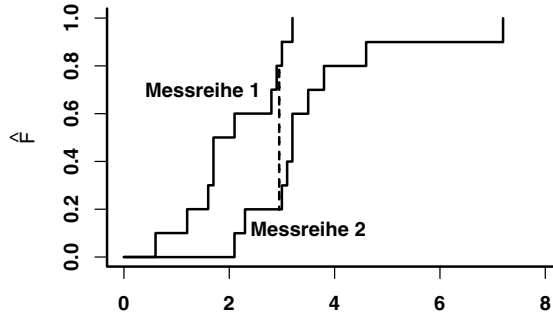


Abb. 7.15. Größter Abstand zwischen den empirischen kumulierten Häufigkeitsverteilungen $\hat{F}_1$ und $\hat{F}_2$

7.4.9 Cramér-von Mises Test

Der auf dem KS-Test aufbauende Cramér-von Mises Zweistichprobentest (kurz CM-Test) basiert auf der **Summe der quadrierten Differenzen** zwischen den beiden empirischen Verteilungsfunktionen (Voraussetzungen und Hypothesen wie für den KS-Test genannt; Überschneidungen zugelassen).

Die Prüfgröße des Cramér-von Mises Tests für den Vergleich zweier Stichproben A (x_i ; $i = 1, \dots, n_1$) und B (y_j ; $j = 1, \dots, n_2$) ist:

$$\begin{aligned} \hat{C} &= \frac{n_1 \cdot n_2}{(n_1 + n_2)^2} \sum_{i=1}^{n_1+n_2} D_i^2 \\ &= \frac{n_1 \cdot n_2}{(n_1 + n_2)^2} \left\{ \sum_{i=1}^{n_1} (F(x_i) - G(x_i))^2 + \sum_{j=1}^{n_2} (F(y_j) - G(y_j))^2 \right\} \end{aligned} \quad (7.122)$$

Obere asymptotische Schranken C_α für den CM-Test sind in Tabelle 7.34 angegeben. Diese sind nach Csörgö und Faraway [CF96] schon für kleine Stichprobenumfänge gültig. Der CM-Test erwies sich in einem Vergleich der Power (sieben unterschiedliche Verteilungen) von 11 Tests, einschließlich des KS-Tests, als überlegener Sieger (Bünig [BC99], Sachs [Sac06]).

Tabelle 7.34. Obere asymptotische Schranken C_α für den CM-Test

α	0,30	0,20	0,10	0,05	0,01	0,001
C_α	0,184	0,241	0,347	0,461	0,743	1,168



Beispiel 1: Eine vereinfachte Berechnung der Teststatistik zum CM-Test in **R** soll anhand der Daten des Beispiels aus dem vorangehenden Abschnitt (Tabelle 7.33) gezeigt werden. Die empirischen Verteilungsfunktionen zu den beiden Messreihen werden aus der Funktion **hist()** übernommen. Die absolute Differenz $KS = 0,6$ entspricht der Teststatistik des Kolmogoroff-Smirnoff Tests, die Summe der Abweichungsquadrate nach (7.122) $CM = 0,875$ führt auf die Teststatistik des CM-Tests. Da $CM = 0,875 < 0,461 = C_\alpha$ kann auch hier die Nullhypothese abgelehnt werden.

```
> m1 <- c(0.6, 1.2, 1.6, 1.7, 1.7, 2.1, 2.8, 2.9, 3.0, 3.2)
> m2 <- c(2.1, 2.3, 3.0, 3.1, 3.2, 3.2, 3.5, 3.8, 4.6, 7.2)
> n1 <- 10; n2 <- 10; x <- seq(0, 8, by=0.1)

> hm1 <- hist(m1, breaks=x, plot=F); F <- cumsum(hm1$counts)/n1
```

```
> hm2 <- hist(m2, breaks=x, plot=F); G <- cumsum(hm2$counts)/n2
> KS <- max(abs(F-G)); KS
[1] 0.6
> C <- (n1*n2)/(n1+n2)^2 * sum((hm1$counts+hm2$counts)*((F-G)^2)); C
[1] 0.875
```

Beispiel 2: Zahlenbeispiel mit zwei Stichproben X ($n_1 = 9$) und Y ($n_2 = 15$), deren empirische Verteilungsfunktionen sich schneiden.



x_i	y_j	$F(x)$	$G(y)$	$ F(x) - G(y) $	$[F(x) - G(y)]^2$
	4,3	0	0,067	0,067	0,004
	4,8	0	0,133	0,133	0,018
	5,2	0	0,200	0,200	0,040
	5,7	0	0,267	0,267	0,071
	6,0	0	0,333	0,333	0,111
	6,9	0	0,400	$\hat{D} = 0,400$	0,160
7,3		0,111	0,400	0,289	0,084
7,9		0,222	0,400	0,178	0,032
	8,0	0,222	0,467	0,245	0,060
8,7		0,333	0,467	0,134	0,018
9,0		0,444	0,467	0,023	0,001
9,4		0,555	0,467	0,088	0,008
	9,6	0,555	0,533	0,022	0,000
10,2		0,666	0,533	0,133	0,018
10,5		0,777	0,533	0,244	0,060
11,1		0,888	0,533	0,355	0,126
	11,4	0,888	0,600	0,288	0,083
12,6		1,0	0,600	$\hat{D}=0,400$	0,160
	12,8	1,0	0,667	0,333	0,111
	13,1	1,0	0,733	0,267	0,071
	13,4	1,0	0,800	0,200	0,040
	13,7	1,0	0,867	0,133	0,018
	14,5	1,0	0,933	0,067	0,004
	14,9	1,0	1,0	0	0

$\sum D_i^2 = 1,298$

Die Prüfgröße für den Cramér-von Mises-Test ist danach

$$\hat{C} = \frac{9 \cdot 15}{(9 + 15)^2} 1,298 = 0,3042$$

und damit nicht größer als der kritische Wert $C_\alpha = 0,461$ aus Tabelle 7.34, die Nullhypothese kann daher nicht abgelehnt werden.

```
> X <- c(7.3, 7.9, 8.7, 9.0, 9.4, 10.2, 10.5, 11.1, 12.6)
> Y <- c(4.3, 4.8, 5.2, 5.7, 6.0, 6.9, 8.0, 9.6, 11.8,
+      12.8, 13.1, 13.4, 13.7, 14.5, 14.9)
> n1 <- length(X); n2 <- length(Y); x <- seq(0, 15, by=0.1)
> hX <- hist(X, breaks=x, plot=F); F <- cumsum(hX$counts)/n1
> hY <- hist(Y, breaks=x, plot=F); G <- cumsum(hY$counts)/n2
> KS <- max(abs(F-G)); KS
[1] 0.4
> C <- (n1*n2)/(n1+n2)^2 * sum((hX$counts+hY$counts)*((F-G)^2)); C
[1] 0.3041667
```

7.4.10 Einige weitere verteilungsunabhängige Verfahren für den Vergleich unabhängiger Stichproben

Der einfachste verteilungsunabhängige Test für den Vergleich zweier unabhängiger Stichproben stammt von **Mosteller** [Mos48]. Vorausgesetzt wird, dass beide Stichprobenumfänge gleich groß sind ($n_1 = n_2 = n$). Die Nullhypothese, beide Stichproben entstammen Grundgesamtheiten mit gleicher Verteilung, wird für $n > 5$ mit einer Irrtumswahrscheinlichkeit von 5% verworfen, wenn für

$$\begin{aligned} n \leq 25 \text{ die } k \geq 5 \text{ größten oder kleinsten Werte} \\ n > 25 \text{ die } k \geq 6 \text{ größten oder kleinsten Werte} \end{aligned} \quad (7.123)$$

derselben Stichprobe entstammen. Conover [Con99] und andere geben interessante Weiterentwicklungen dieses Tests. Danach sind für $n_1 = n_2 \geq 20$ die kritischen Werte $k \geq 5$ ($\alpha = 0,05$) und $k \geq 7$ ($\alpha = 0,01$).

7.4.10.1 Rosenbaumsche Schnelltests

Beide Tests sind verteilungsunabhängig für unabhängige Stichproben. Wir setzen voraus, dass die Stichprobenumfänge gleich sind: $n_1 = n_2 = n$.

Lage-Test: Liegen mindestens 5 (von $n \geq 16$; $\alpha = 0,05$) bzw. mindestens 7 (von $n \geq 20$; $\alpha = 0,01$) Werte(n) einer Stichprobe, unterhalb bzw. oberhalb des Variationsbereichs der anderen Stichprobe, so ist die Nullhypothese (Gleichheit der Mediane) mit der angegebenen Irrtumswahrscheinlichkeit abzulehnen; vorausgesetzt wird, dass die **Variationsbereiche nur zufällig verschieden** sind; die Irrtumswahrscheinlichkeiten gelten für die einseitige Fragestellung, für die zweiseitige sind sie zu verdoppeln [Ros54]. Einige kritische Werte für $n_1 \neq n_2$ ($5 \leq n_1, n_2 \leq 16$) und 4α -Niveaus enthält Sachs [Sac06].

Variabilitätstest: Liegen mindestens 7 (von $n \geq 25$; $\alpha = 0,05$) bzw. mindestens 10 (von $n \geq 51$; $\alpha = 0,01$) Werte(n) einer Stichprobe (derjenigen mit dem größeren Variationsbereich; einseitige Fragestellung) außerhalb des Variationsbereichs der anderen Stichprobe, so ist die Nullhypothese (Gleichheit der Variabilität, der Streuung) mit der angegebenen Irrtumswahrscheinlichkeit abzulehnen; vorausgesetzt wird, dass die **Mediane nur zufällig verschieden** sind. Ist unbekannt, ob die beiden Grundgesamtheiten dieselbe Lage haben, so prüft dieser Test Lage **und** Variabilität beider Grundgesamtheiten. Für $7 \leq n \leq 24$ darf die 7 durch eine 6 ersetzt werden ($\alpha = 0,05$), für $21 \leq n \leq 50$ (bzw. $11 \leq n \leq 20$) die 10 durch eine 9 (bzw. eine 8) (Rosenbaum, S. [Ros53]).

Die beiden Arbeiten enthalten kritische Werte für den Fall ungleicher Stichprobenumfänge.

7.4.10.2 Permutationstest, Randomisierungstest

Die Anzahl der Möglichkeiten, $(n_1 + n_2)$ Objekte in zwei Gruppen aufzuteilen, so dass die eine Gruppe n_1 und die andere n_2 Objekte enthält, beträgt

$$\binom{n_1 + n_2}{n_1} = \frac{(n_1 + n_2)!}{n_1! n_2!} \quad (7.124)$$

Da jede dieser Permutationen die gleiche Wahrscheinlichkeit aufweist, gilt:

$$P = \frac{n_1! n_2!}{(n_1 + n_2)!} \quad (7.125)$$

Es liegen n_1 der Größe nach geordnete Beobachtungen vor. Die Wahrscheinlichkeit, dass von n_2 neuen Beobachtungen alle größer sein werden als die (Größe der) n_1 Beobachtungen, ist durch (7.125) gegeben.

Die Wahrscheinlichkeit für die **Nichtüberlappung** zweier Stichproben des Umfangs $n_1 = n_2$ ist dann (entsprechend einer zweiseitigen Hypothesenstellung):

$$P = \frac{2n_1!n_2!}{(n_1 + n_2)!} \quad (7.126)$$

Nichtüberlappung bedeutet hier: alle Beobachtungen von n_1 sind entweder kleiner als die von n_2 oder umgekehrt. Zahlreiche Ansätze für Randomisierungs- und Permutationstests sowie praktische Anwendungen in verschiedenen Disziplinen zeigen E.S. Edington [Edi95], P. Good [Goo05] und B.F.J. Manly [Man97].

Beispiel: Wie groß ist die Wahrscheinlichkeit, dass für zwei Stichproben des Umfangs $n_1 = n_2 = 3$ (a), $= 4$ (b), $= 5$ (c) die Werte sich nicht überlappen und auf dem 5%-Signifikanzniveau als statistisch signifikant (*s.s.*) ausgewiesen werden (d)?

(a) P für Nichtüberlappung $= 2 \cdot 3! \cdot 3! / 6! = 2 \cdot 3 \cdot 2 / 6 \cdot 5 \cdot 4 = 1/10$

(b) P für Nichtüberlappung $= 2 \cdot 4! \cdot 4! / 8! = 1/35$

(c) P für Nichtüberlappung $= 2 \cdot 5! \cdot 5! / 10! = 1/126$

(d) a: $1/10 = 0,1 > 0,05$ n.s.; b: $1/35 = 0,0286 < 0,05$ s.s.; c: $1/126 = 0,008 < 0,05$ s.s.

Dies ist ein Spezialfall für den allgemeineren **Randomisierungstest** (Fisher-Pitman). Dabei werden alle Möglichkeiten betrachtet, mit denen $n = n_1 + n_2$ Messwerte aus zwei Stichproben wiederum auf zwei Stichproben mit den gleichen Stichprobenumfängen verteilt werden können. Als Teststatistik kann die Summe der Werte (T) aus der kleineren Stichprobe betrachtet werden ($n_1 \leq n_2$). Unter allen möglichen Aufteilungen wird die Anzahl z_T derjenigen Aufteilungen ermittelt, die zu einer Summe führen, die **gleich der beobachteten Summe T** ist, bzw. die Anzahl z_u der Aufteilungen, die zu einer Summe führen, die **kleiner als T** ist (oder entsprechend der Fragestellung auch größer als T ist). Die einseitige Überschreitungswahrscheinlichkeit (P-Wert) für die insgesamt $z_u + z_T$ Summen, die kleiner (größer) oder gleich groß sind wie T ist durch (7.127) gegeben.

$$P_{\text{einseitig}} = \frac{z_u + z_T}{\binom{n}{n_1}} \quad (7.127)$$

Für den zweiseitigen Test sind die Summen zu berücksichtigen, die den Wert T unterschreiten und $S - T$ (S ist dabei die Gesamtsumme) überschreiten. Aus Symmetriegründen folgt:

$$P_{\text{zweiseitig}} = \frac{2 \cdot (z_u + z_T)}{\binom{n}{n_1}} \quad (7.128)$$

Die Nullhypothese, dass die beiden Stichproben aus der derselben Grundgesamtheit kommen, kann für ein fest vorgegebenes Signifikanzniveau α abgelehnt werden, wenn $P_{\text{einseitig}} \leq \alpha$ bzw. $P_{\text{zweiseitig}} \leq \alpha$ ist.

Beispiel: Gegeben sind zwei Stichproben $A = \{2, 5\}$ und $B = \{3, 7, 6\}$. Als Teststatistik wird die Summe der Beobachtungen aus der kleineren Stichprobe gewählt $T = 7$. Insgesamt gibt es 10 verschiedene Möglichkeiten, die insgesamt 5 Beobachtungen auf die beiden Stichproben zu verteilen (5 über 2).



A		B			T
2	5	3	7	6	7
3	5	2	7	6	8
7	5	3	2	6	12
6	5	3	7	2	11
2	3	5	7	6	5
2	7	3	5	6	9
2	6	3	7	5	8
3	7	2	5	6	10
3	6	2	7	5	9
7	6	3	2	5	13

Damit ist die Zahl der Summen, die gleich T sind $z_T = 1$, die kleiner als T sind $z_u = 1$ und die größer als T sind $z_o = 8$. Entsprechend ist dann $P_{\text{eins.,unten}} = 0,2$, $P_{\text{eins.,oben}} = 0,9$ und $P_{\text{zweiseitig}} = 0,4$.

Der Aufwand für eine manuelle Berechnung exakter P-Werte ist recht hoch. In **R** steht dafür eine besondere Funktion **perm.test()** im Rahmen des Zusatzpaketes **exactRankTests** [HH05] zur Verfügung, deren Anwendung an einem kleinen Zahlenbeispiel gezeigt werden soll.



Beispiel: Es soll geprüft werden, ob die Stichproben A (20, 23, 30) und B (27, 29, 35, 38, 40, 40, 45) aus derselben Grundgesamtheit stammen können.

```
> library(exactRankTests)
> x1 <- c(20, 23, 30);          n1 <- length(x1)
> x2 <- c(27, 29, 35, 38, 40, 40, 45); n2 <- length(x2)
> sum(x1)                      # Summe der Werte aus der ersten Stichprobe
[1] 73
> choose(n1 + n2, n1)          # Anzahl möglicher Summen mit 3 Summanden
[1] 120
> perm.test(x1, x2, alternative="less", exact=TRUE)

2-sample Permutation Test

data:  x1 and x2
T = 73, p-value = 0.025
alternative hypothesis: true mu is less than 0
```

Als Teststatistik wird die Summe der Werte aus der ersten (kleineren) Stichprobe (hier $T = 73$) herangezogen. Die Anzahl möglicher Summen mit 3 Summanden beträgt 120, aber nur 2 dieser Summen sind kleiner als 73. Mit der aus den Stichproben berechneten Summe folgt für den P-Wert nach (7.127) $P = 2/120 = 0,01\bar{6}$, d.h. die Nullhypothese kann bei einseitiger Fragestellung abgelehnt werden.

7.4.10.3 Der Vergleich zweier unabhängiger Stichproben: Schnelltest nach Tukey

Zwei Stichprobengruppen sind um so unterschiedlicher, je weniger sich ihre Werte überschneiden. Enthält eine Gruppe den höchsten und die andere Gruppe den niedrigsten Wert, dann sind zu zählen:

- (1) diejenigen a Werte einer Gruppe, die alle Werte der anderen Gruppe **übersteigen**,
- (2) diejenigen b Werte der anderen Gruppe, die alle Werte der Gruppe **unterschreiten**.

Beide Häufigkeiten, jede muss größer als Null sein, werden addiert. Hierdurch erhält man den Wert der Prüfgröße $T = a + b$. Wenn beide Stichprobenumfänge etwa gleich groß sind, dann betragen die kritischen Werte der Prüfgröße 7, 10 und 13:

- 7 für einen zweiseitigen Test auf dem 5%-Niveau,
- 10 für einen zweiseitigen Test auf dem 1%-Niveau und
- 13 für einen zweiseitigen Test auf dem 0,1%-Niveau (**Tukey** [Tuk59]).

Für zwei gleiche Werte ist 0,5 zu zählen. Bezeichnen wir die beiden Stichprobenumfänge mit n_1 und n_2 , wobei $n_1 \leq n_2$, dann ist der Test (H_0 : Gleichheit zweier Verteilungsfunktionen) gültig für nicht zu unterschiedliche Stichprobenumfänge, genau für

$$n_1 \leq n_2 \leq 3 + 4n_1/3 \quad (7.129)$$



Für alle anderen Fälle ist vom Wert der berechneten Prüfgröße T ein Korrekturwert abzuziehen, bevor der Wert T mit 7, 10 und 13 verglichen wird. Dieser Korrekturwert beträgt:

$$\begin{aligned} &1, \quad \text{wenn } 3 + 4n_1/3 < n_2 < 2n_1 \\ \text{die ganze Zahl in } &\frac{n_2 - n_1 + 1}{n_1}, \quad \text{wenn } 2n_1 \leq n_2 \end{aligned} \quad (7.130)$$

Beispielsweise ist für $n_1 = 7$ und $n_2 = 13$ Formel (7.129) nicht erfüllt, da $3 + \frac{4 \cdot 7}{3} = \frac{37}{3} < 13$. Formel (7.130) entspricht den Tatsachen, somit ist der Korrekturwert 1 abzuziehen. Für $n_1 = 4$ und $n_2 = 14$ ergibt (7.130) $\frac{14 - 4 + 1}{4} = \frac{11}{4} = 2,75$ den Korrekturwert 2.

Übertrifft die eine Stichprobe die andere um mindestens 9 Werte ($n_2 - n_1 \geq 9$), dann ist für das 0,1%-Niveau der kritische Wert 14 anstelle des Wertes 13 zu verwenden. Kritische Werte für den einseitigen Test (vgl. auch beide Schnelltests nach Rosenbaum), nur ein Verteilungsende interessiert und damit auch nur a oder b , gibt Westlake [Wes71]: 4 für $10 \leq n_1 = n_2 \leq 15$ und 5 für $n_1 = n_2 \geq 16$ ($\alpha = 0,05$) sowie 7 für $n_1 = n_2 \geq 20$ ($\alpha = 0,01$).

Beispiel: Es liegen die folgenden Werte vor:



A: 14,7 15,3 16,1 14,9 15,1 14,8 16,7 17,3* 14,6* 15,0

B: 13,9 14,6 14,2 15,0* 14,3 13,8* 14,7 14,4

Wir versehen die höchsten und niedrigsten Werte jeder Reihe mit einem Stern. Größer als 15,0* sind 5 Werte (unterstrichen), der Wert 15,0 der Stichprobe A wird als halber Wert gerechnet. Kleiner als 14,6* sind ebenfalls 5 $\frac{1}{2}$ Werte. Wir erhalten $T = 5\frac{1}{2} + 5\frac{1}{2} = 11$. Ein Korrekturwert entfällt, da $(n_1 \leq n_2 \leq 3 + 4n_1/3) \quad 8 < 10 < 13,7$. Da $T = 11 > 10$ ist, muss die Nullhypothese (Gleichheit der den beiden Stichproben zugrundeliegenden Verteilungsfunktionen) auf dem 1%-Niveau abgelehnt werden.

Exakte kritische Schranken für kleine Stichprobenumfänge können bei Bedarf der Originalarbeit von Tukey [Tuk59] entnommen werden. D.J. Gans (1981, *Technometrics* **23**, 193–195) gibt korrigierte und erweiterte Schranken.

7.4.10.4 Der Median Test

Der Median-Test ist ein recht einfaches, robustes Testverfahren: Man ordnet die vereinigten aus den Stichproben I und II stammenden Werte ($n_1 + n_2$) der Größe nach aufsteigend, ermittelt den Medianwert $\tilde{x}$ und ordnet die Werte jeder Stichprobe danach, ob sie kleiner oder größer als $\tilde{x}$ sind, in das Schema nach Tabelle 7.35 ein ((a, b, c, d) sind Häufigkeiten).

Die weitere Rechnung folgt den in Abschnitt [7.6] gegebenen Vorschriften und Empfehlungen. Bei statistisch signifikanten Befunden wird dann die Nullhypothese $\tilde{\mu}_1 = \tilde{\mu}_2$ auf dem verwendeten Niveau abgelehnt. Die asymptotische Effizienz des Median Tests beträgt $2/\pi = 2/3,1416 = 0,6366$ oder 64% d. h. dass die Anwendung dieses Tests bei 1000 Werten die gleiche Power aufweist wie die Anwendungen des t -Tests bei etwa $0,64 \cdot 1000 = 640$ Werten, wenn in Wirklichkeit Normalverteilung vorliegt. Bei anderen Verteilungen kann das Verhältnis ganz anders sein. Der Median-Test

Tabelle 7.35. Tabellenschema für den einfachen Mediantest

	Anzahl der Werte	
	$< \tilde{x}$	$> \tilde{x}$
Stichprobe I	a	b
Stichprobe II	c	d

wird daher auch bei Überschlagsrechnungen benutzt, außerdem dient er zur Kontrolle hochsignifikanter Befunde, denen man nicht so recht traut. Führt er zu einem anderen Ergebnis, so muss die Berechnung des fraglichen Befundes überprüft werden.

Hauptanwendungsgebiet des Median-Tests und insbesondere des Median-Quartile-Tests (vgl. Tabelle 7.36) ist der **Vergleich zweier Medianwerte bei starken Verteilungsformunterschieden**: der U -Test darf dann im Gegensatz zum Kolmogoroff-Smirnoff Test und zum Cramér-von Mises Test nicht angewandt werden.



Beispiel: Wir benutzen das Beispiel 1 zum U -Test (Abschnitt 7.4.6) und erhalten $\tilde{x} = 19$ sowie die folgende Vierfeldertafel

	Anzahl der Werte	
	$< \tilde{x}$	$> \tilde{x}$
Stichprobe I	2	6
Stichprobe II	6	2

die nach Abschnitt [7.6.2] mit $P = 0,066$ die Nullhypothese auf dem 5%-Niveau nicht abzulehnen gestattet.

Prüfen wir nicht zwei sondern k unabhängige Stichproben, so erhalten wir den erweiterten Mediantest: Die Werte der k Stichproben werden der Größe nach in eine Rangfolge gebracht, man bestimmt den Medianwert und zählt, wie viele Messwerte in jeder der k Stichproben oberhalb und wie viele unterhalb des Medianwertes liegen. Die Nullhypothese, die Stichproben entstammen einer gemeinsamen Grundgesamtheit, lässt sich unter der Voraussetzung, dass die resultierende $k \cdot 2$ -Felder-Tafel ausreichend besetzt ist (alle Erwartungshäufigkeiten müssen > 1 sein), nach den im Abschnitt 7.6 dargelegten Verfahren prüfen. Die Alternativhypothese lautet dann: Nicht alle k Stichproben entstammen einer gemeinsamen Grundgesamtheit. Das entsprechende optimale verteilungsfreie Verfahren ist der H -Test von Kruskal und Wallis.

Einen eleganten **Median-Quartile-Test**, bei dem die vereinigten Beobachtungswerte zweier unabhängiger Stichproben durch ihre drei Quartile: $Q_1, Q_2 = \tilde{x}$ und Q_3 auf die Häufigkeiten einer $2 \cdot 4$ -Feldertafel reduziert werden, beschreibt Bauer [Bau62]. Der sehr brauchbare Test prüft nicht nur Lage-, sondern auch Dispersions- und gewisse Verteilungsformunterschiede. Eine Verallgemeinerung des Tests auf mehr als zwei Stichproben ist möglich.

Tabelle 7.36. Tabellenschema für den Median-Quartile-Test

	$\leq Q_1$	$\leq Q_2$	$\leq Q_3$	$> Q_3$
Stichprobe I				
Stichprobe II				

7.4.11 Zweistichprobentest auf Äquivalenz

Es liegen zwei unabhängige Stichproben (A und B) mit normalverteilten Beobachtungen ($X_i \sim N(\mu_1; \sigma^2)$ für $i = 1, \dots, m$ und $Y_j \sim N(\mu_2; \sigma^2)$ für $j = 1, \dots, n$) vor. Die unbekannten Varianzen werden als gleich groß angenommen. Die Hypothesen zur Äquivalenz der A und B zugrunde

liegenden Verteilungen können anschaulich mit Hilfe der **standardisierten Differenz** Θ (griech. Theta):

$$\Theta = \frac{\mu_1 - \mu_2}{\sigma}$$

definiert werden:

$$\begin{aligned} H_A : \quad & -\varepsilon_1 < \Theta < +\varepsilon_2 \\ H_0 : \quad & \Theta \leq -\varepsilon_1 \quad \text{oder} \quad \Theta \geq +\varepsilon_2 \end{aligned} \quad (7.131)$$

Danach werden μ_1 und μ_2 als äquivalent betrachtet, wenn die standardisierte Differenz innerhalb fester Grenzen liegt. Für die Festlegung dieser **Äquivalenzgrenzen** ε_1 und ε_2 hilft folgende Überlegung. Die Verteilung von $(X_i - Y_j) \sim N(\mu_1 - \mu_2; 2\sigma^2)$ ermöglicht eine zu (7.131) äquivalente Formulierung des Testproblems:

$$\begin{aligned} H_A : \quad & \frac{1}{2} - \tilde{\varepsilon}_1 < P(X_i < Y_j) < \frac{1}{2} + \tilde{\varepsilon}_2 \\ H_0 : \quad & P(X_i - Y_j) \leq \frac{1}{2} - \tilde{\varepsilon}_1 \quad \text{oder} \quad P(X_i - Y_j) \geq \frac{1}{2} + \tilde{\varepsilon}_2 \end{aligned} \quad (7.132)$$

Die Alternativhypothese (hier **Äquivalenz**) ist danach gleichbedeutend mit der Annahme, dass die Wahrscheinlichkeit einen X -Wert zu erhalten, der größer ist als ein zufällig ausgewählter Y -Wert, nur unwesentlich von $1/2$ abweicht. Die Festlegung $\varepsilon_1 = \varepsilon_2 = 0,5$ führt mit $\tilde{\varepsilon}_i = \Phi(\varepsilon_i/\sqrt{2}) - 1/2$ (Φ bezeichnet hier die Verteilungsfunktion der Standardnormalverteilung) zu einer „akzeptablen“ Toleranz von ca. 15% für eine Abweichung zwischen $P(X_i > Y_j)$ und $1/2$.

Die Teststatistik für den Zweistichprobentest auf Äquivalenz lautet:

$$\hat{T} = \frac{\bar{x} - \bar{y}}{\sqrt{\sum_{i=1}^m (x_i - \bar{x})^2 + \sum_{j=1}^n (y_j - \bar{y})^2}} \sqrt{\frac{mn(m+n-2)}{m+n}} \quad (7.133)$$

Eine Entscheidung zugunsten einer Äquivalenz (H_A) von A und B wird getroffen, wenn der Wert der Teststatistik (7.133) kleiner ist als der kritische Wert, der sich für den symmetrischen Fall ($\varepsilon_1 = \varepsilon_2 = \varepsilon$) aus der „**nichtzentralen Fisher-Verteilung**“ herleiten lässt.

$$|\hat{T}| < F_{1;m+n-2;\alpha;NC} \quad \text{mit} \quad NC = \frac{mn}{m+n} \varepsilon^2 \quad (7.134)$$

Beispiel: Beobachtungen aus zwei Stichproben X und Y sollen geprüft werden, ob sie aus Normalverteilungen mit demselben Erwartungswert und derselben (unbekannten) Varianz stammen. Jeweils $m = 10$ und $n = 12$ Werte sind mit der Funktion **rnorm()** in **R** erzeugt worden und werden entsprechend (7.133) und (7.134) untersucht.



```
> ##### Quantile zur nichtzentralen Fisher-Verteilung #####
> myqf <- function(p, df1, df2, ncp) {
+   uniroot(function(x) pf(x, df1, df2, ncp) - p, ,0, 100) $ root }
> #####
> x <- c(59.3, 58.8, 62.0, 42.6, 73.3, 54.2, 50.5, 38.0, 45.3, 50.0)
> y <- c(34.9, 44.9, 52.0, 65.4, 52.5, 52.2, 68.6, 47.7, 55.9, 55.7, 53.5, 56.6)
>
> m.x <- mean(x); s.x <- sd(x); m=length(x)
> m.y <- mean(y); s.y <- sd(y); n=length(y)
> T <- ((m.x - m.y) / sqrt(sum((x-m.x)^2) + sum((y-m.y)^2))) * sqrt((m*n*(m+n-2))/(m+n)); T
[1] 0.01835230
>
> eps <- 0.5
> c <- sqrt(myqf(0.05, 1, m+n-2, ncp=(m*n/(m+n))*eps^2)); c
[1] 0.1252520
```

Der Wert der Teststatistik $T = 0,018$ ist kleiner als das entsprechende Quantil der nichtzentralen Fisher-Verteilung $C = 0,125$. Somit kann davon ausgegangen werden, dass die Beobachtungen aus einer gemeinsamen Verteilung stammen. Das Quantil der nichtzentralen Fisher-Verteilung wird in dem Beispiel über die Funktion **myqf()** aus der Verteilungsfunktion der Fisher-Verteilung **pf()** abgeleitet.

7.4.11.1 Test auf Bioäquivalenz

Die Verfügbarkeit eines Wirkstoffs (Arzneimittel) kann durch die Konzentration im Plasma in Abhängigkeit von der Zeit beschrieben werden. Die **Fläche unter der Konzentrations-Zeit-Kurve** (AUC, area under curve) ist ein Indikator für die absolute Bioverfügbarkeit einer applizierten Substanz. Der Quotient der erwarteten AUC zweier verschiedener Zubereitungen eines Arzneimittels wird **relative Bioverfügbarkeit** genannt. Zwei Zubereitungen heißen „**bioäquivalent**“, wenn das Verhältnis der beiden AUC im Bereich von 0,8 bis 1,25 erwartet werden kann. Geht man zusätzlich davon aus, dass die Verteilung der AUC mit dem Modell der **Lognormal-Verteilung** geeignet beschrieben werden kann, dann können die Hypothesen für die Bioäquivalenz durch (7.135) definiert werden. Abbildung 7.16 vermittelt eine anschauliche Vorstellung. μ_1 und μ_2 bezeichnen die Erwartungswerte der logarithmisch transformierten (hier „natürliche“ Logarithmen) Zufallsvariablen (AUC). Mit den Werten von $\delta=0,223$ wird auf diese Weise eine relative Abweichung von einem Referenzmittel um etwa 20% nach unten und um 25% nach oben als äquivalent akzeptiert (die entsprechenden Grenzen bei einer Transformation mit dekadischen Logarithmen sind etwa -0,1 und +0,1).

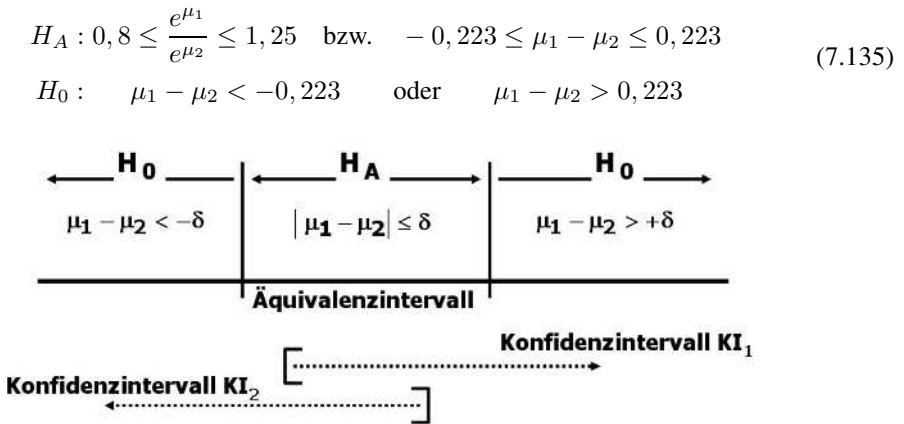


Abb. 7.16. Annahmeregionen und Ablehnungsbereich zum Test auf Äquivalenz

Die Überprüfung der Bedingung (7.135) kann sehr anschaulich anhand von einseitigen Konfidenzintervallen nach dem Intervallinklusionsprinzip erfolgen. Die Nullhypothese wird abgelehnt, wenn die Grenzen der einseitigen Konfidenzintervalle (7.136) vollständig in dem Äquivalenzbereich $[-\delta, +\delta]$ liegen (vgl. auch Abbildung 7.16).

$$\begin{aligned}
 KI_1 : & \left[\bar{d} - t_{n-1, 1-\alpha} \frac{s_d}{\sqrt{n}}; +\infty \right] \\
 KI_2 : & \left[-\infty; \bar{d} + t_{n-1, 1-\alpha} \frac{s_d}{\sqrt{n}} \right]
 \end{aligned}
 \tag{7.136}$$



Beachte: Auch wenn das Konfidenzniveau für ein zweiseitiges Konfidenzintervall unter diesen Voraussetzungen $1-2\alpha$ beträgt, hält die zugehörige Testentscheidung auf Äquivalenz das gewählte Signifikanzniveau α ein!

Hinweis: Das Prinzip der Intervallinklusion ist logisch äquivalent mit der Kombination von zwei einseitigen (Einstichproben-) Hypothesentests, also:

$$\begin{aligned} &H_{01} : \mu_1 - \mu_2 \leq -0,223 \text{ gegen } H_{A1} : \mu_1 - \mu_2 > -0,223 \\ \text{und } &H_{02} : \mu_1 - \mu_2 \geq +0,223 \text{ gegen } H_{A1} : \mu_1 - \mu_2 < +0,223 \end{aligned} \quad (7.137)$$

Beispiel: In einer Cross-Over Studie wurden an 12 männlichen Probanden in zwei durch eine (ausreichende) Washout-Phase getrennte Perioden zwei Allopurinol-Präparate (Behandlung von Gicht: durch Allopurinol wird ein Enzym gehemmt, dass beim Abbau von Purinen notwendig ist, um Harnsäure zu produzieren) appliziert und die Flächen unter den Serumspiegelkurven (AUC in $\mu\text{g/ml h}$) bestimmt. Sechs Probanden erhielten in der ersten Phase das Prüfpräparat (T) und in der zweiten das Referenzpräparat (R), die anderen in umgekehrter Reihenfolge. Die Ergebnisse sind in Tabelle 7.37 zusammengefasst.

Tabelle 7.37. Allopurinol-Daten

Nummer	Sequenz	Periode 1	Periode 2	Nummer	Sequenz	Periode 1	Periode 2
3	R/T	3.648	3.671	1	T/R	3.881	4.894
5	R/T	8.531	7.693	2	T/R	4.835	6.504
6	R/T	4.318	4.481	4	T/R	6.914	7.372
8	R/T	6.974	5.591	7	T/R	5.236	4.105
11	R/T	5.862	5.311	9	T/R	3.058	2.368
12	R/T	3.082	3.165	10	T/R	5.722	6.229

Wegen des speziellen Studienansatzes (crossover design) müssen bei der Berechnung der mittleren Differenz der logarithmierten (hier natürliche Logarithmen) AUC-Werte zwischen den beiden Zubereitungen und der zugehörigen Standardabweichung auch die Sequenzen (R/T gegen T/R) berücksichtigt werden.

$$\begin{aligned} \bar{d} &= \frac{\bar{d}_{RT} - \bar{d}_{TR}}{2} \\ s_d &= \sqrt{\frac{(n-1)s_{RT}^2 + (m-1)s_{TR}^2}{n+m-2}} \\ s_{\bar{d}} &= s_d \sqrt{\left(\frac{1}{n} + \frac{1}{m}\right)/2} \end{aligned}$$

Die Berechnung der beiden Konfidenzintervalle nach (7.136) erfolgt in **R** elementar in den folgenden Schritten.

```
> R1 <- c(3.648, 8.531, 4.318, 6.974, 5.862, 3.082)
> R2 <- c(4.894, 6.504, 7.372, 4.105, 2.368, 6.229)
> T1 <- c(3.881, 4.835, 6.914, 5.236, 3.058, 5.722)
> T2 <- c(3.671, 7.693, 4.481, 5.591, 5.311, 3.165)
>
> RT <- log(R1) - log(T2); n <- length(RT); mRT <- mean(RT); sRT <- sd(RT)
> TR <- log(R2) - log(T1); m <- length(TR); mTR <- mean(TR); sTR <- sd(TR)
>
> mD <- (mRT + mTR) / 2; mD
[1] 0.044304
> sD <- sqrt(((n-1)*sRT^2 + (m-1)*sTR^2) / (n+m-2)); sD
[1] 0.1797106
>
> alpha <- 0.05
> l.u <- md - qt(1-alpha, ntr+nrt-2) * (sD * sqrt((1/n + 1/m) * 0.5)); l.u
[1] -0.08867
> l.o <- md + qt(1-alpha, ntr+nrt-2) * (sD * sqrt((1/n + 1/m) * 0.5)); l.o
[1] 0.177278
```

Zunächst werden getrennt für die beiden Sequenzen die AUC-Werte logarithmiert, die Differenzen gebildet und Mittelwerte bzw. Standardabweichungen berechnet. Anschließend wird der gemeinsame Mittelwert $\bar{d} = 0,044$ und die gemeinsame, auf das Präparat bezogene, Standardabweichung $s_d = 0,1797$ der Differenzen berechnet und die Grenzen der Konfidenzintervalle bestimmt ($-0,089$ und $0,178$). Diese liegen vollständig im vorgegeben Äquivalenzbereich von $[-0,223; +0,223]$, somit ist von einer Äquivalenz der beiden Präparate hinsichtlich ihrer Bioverfügbarkeit auszugehen.

7.5 Mehrstichprobenverfahren, varianzanalytische Methoden

- Prüfung der Gleichheit mehrerer Varianzen
- Einfache Varianzanalyse
- Multiple Vergleiche, multiples Testproblem
- H-Test von Kruskal und Wallis
- Varianzanalyse für Messwiederholungen (Blockvarianzanalyse)
- Friedman-Test
- Zweifache Varianzanalyse
- Prinzipien der Versuchsplanung

Wesentlicher Teil einer optimalen Planung, Durchführung und Analyse von Versuchen basiert auf der Regressionsanalyse und auf der sogenannten Varianzanalyse, die R.A. Fisher (1890–1962) für die Planung und Auswertung von Experimenten, insbesondere von Feldversuchen, geschaffen hat und die es gestattet, **wesentliche von unwesentlichen Einflussgrößen zu unterscheiden**. Eine besondere Rolle spielen hierbei Vergleiche von Mittelwerten. Da die Varianzanalyse wie der t -Test Normalverteilung und Gleichheit der Varianzen voraussetzt, wollen wir zunächst dem F -Test entsprechende Verfahren kennenlernen, die zur Prüfung der Gleichheit oder der Homogenität mehrerer Varianzen dienen. Sind die Varianzen mehrerer Stichprobengruppen gleich, dann lassen sich auch die Mittelwerte mühelos vergleichen. Dies ist die einfachste Form der Varianzanalyse. Für die sichere Erfassung mehrerer wesentlicher Einflussgrößen ist es notwendig, dass die Beobachtungswerte aus speziellen Versuchsanordnungen gewonnen werden (vgl. Abschnitt [7.5.8]).

Die Varianzanalyse dient zur **quantitativen Untersuchung von Einflussgrößen auf Versuchsergebnisse**; in erster Linie interessiert der Vergleich mehrerer Mittelwerte: Geprüft wird die Gleichheit von Erwartungswerten (H_0).

Tabelle 7.38. Tests für den verteilungsunabhängigen Vergleich mehrerer Stichproben

Stichproben	geordnete Alternativen?	
	nein	ja
unabhängig	H -Test [7.5.4]	Jonckheere-Test [7.5.4.4]
verbunden	Friedman-Test [7.5.6]	Page-Test [7.5.6.2]

Unabhängige Stichprobengruppen nicht normalverteilter Daten mit nicht unbedingt gleichen Varianzen, aber angenähert gleichem Verteilungstyp lassen sich anhand des H -Tests und nach Nemenyi verglichen. Bei verbundenen Stichprobengruppen angenähert gleichen Verteilungstyps ist der Friedman-Test mit den entsprechenden multiplen Vergleichen angezeigt.

7.5.1 Prüfung der Gleichheit mehrerer Varianzen

Bei den folgenden Verfahren werden unabhängige Zufallsstichproben aus normalverteilten Grundgesamtheiten vorausgesetzt!

7.5.1.1 Prüfung der Gleichheit mehrerer Varianzen nach Hartley

Einen relativ einfachen Test zur Ablehnung der Nullhypothese auf Gleichheit oder Homogenität der Varianzen $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_i^2 = \dots = \sigma_k^2 = \sigma^2$ hat Hartley vorgeschlagen. Unter den Bedingungen **gleicher Gruppenumfänge** (n_0) kann diese Hypothese nach

$$\hat{F}_{\max} = \frac{\text{größte Stichproben-Varianz}}{\text{kleinste Stichproben-Varianz}} = \frac{s_{\max}^2}{s_{\min}^2} \quad (7.138)$$

getestet werden. Die Stichprobenverteilung der Prüfgröße $\hat{F}_{\max}$ ist Tabelle 7.39 zu entnehmen. Die Parameter dieser Verteilung sind die Anzahl k der Gruppen und die Anzahl der Freiheitsgrade $\nu = n_0 - 1$ für jede Gruppenvarianz. Wenn $\hat{F}_{\max}$ für eine vorgegebene Irrtumswahrscheinlichkeit den tabellierten Wert überschreitet, dann wird die Gleichheits- oder Homogenitätshypothese abgelehnt und die Alternativhypothese: $\sigma_i^2 \neq \sigma^2$ für bestimmte i akzeptiert [Har50].

Tabelle 7.39. Verteilung von $F_{\max}$ nach Hartley für die Prüfung mehrerer Varianzen auf Homogenität (auszugsweise entnommen aus Pearson, E.S. und H.O. Hartley: Biometrika Tables for Statisticians, vol. 1 (2nd ed.), Cambridge 1958, Table 31)

$\alpha = 0,05$

$\nu \backslash k$	2	3	4	5	6	7	8	9	10	11	12
2	39,0	87,5	142	202	266	333	403	475	550	626	704
3	15,4	27,8	39,2	50,7	62,0	72,9	83,5	93,9	104	114	124
4	9,60	15,5	20,6	25,2	29,5	33,6	37,5	41,1	44,6	48,0	51,4
5	7,15	10,8	13,7	16,3	18,7	20,8	22,9	24,7	26,5	28,2	29,9
6	5,82	8,38	10,4	12,1	13,7	15,0	16,3	17,5	18,6	19,7	20,7
7	4,99	6,94	8,44	9,70	10,8	11,8	12,7	13,5	14,3	15,1	15,8
8	4,43	6,00	7,18	8,12	9,03	9,78	10,5	11,1	11,7	12,2	12,7
9	4,03	5,34	6,31	7,11	7,80	8,41	8,95	9,45	9,91	10,3	10,7
10	3,72	4,85	5,67	6,34	6,92	7,42	7,87	8,28	8,66	9,01	9,34
12	3,28	4,16	4,79	5,30	5,72	6,09	6,42	6,72	7,00	7,25	7,48
15	2,86	3,54	4,01	4,37	4,68	4,95	5,19	5,40	5,59	5,77	5,93
20	2,46	2,95	3,29	3,54	3,76	3,94	4,10	4,24	4,37	4,49	4,59
30	2,07	2,40	2,61	2,78	2,91	3,02	3,12	3,21	3,29	3,36	3,39
60	1,67	1,85	1,96	2,04	2,11	2,17	2,22	2,26	2,30	2,33	2,36
∞	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00

$\alpha = 0,01$

$\nu \backslash k$	2	3	4	5	6	7	8	9	10	11	12
2	199	448	729	1036	1362	1705	2063	2432	2813	3204	3605
3	47,5	85	120	151	184	21(6)	24(9)	28(1)	31(0)	33(7)	36(1)
4	23,2	37	49	59	69	79	89	97	106	113	120
5	14,9	22	28	33	38	42	46	50	54	57	60
6	11,1	15,5	19,1	22	25	27	30	32	34	36	37
7	8,89	12,1	14,5	16,5	18,4	20	22	23	24	26	27
8	7,50	9,9	11,7	13,2	14,5	15,8	16,9	17,9	18,9	19,8	21
9	6,54	8,5	9,9	11,1	12,1	13,1	13,9	14,7	15,3	16,0	16,6
10	5,85	7,4	8,6	9,6	10,4	11,1	11,8	12,4	12,9	13,4	13,9
12	4,91	6,1	6,9	7,6	8,2	8,7	9,1	9,5	9,9	10,2	10,6
15	4,07	4,9	5,5	6,0	6,4	6,7	7,1	7,3	7,5	7,8	8,0
20	3,32	3,8	4,3	4,6	4,9	5,1	5,3	5,5	5,6	5,8	5,9
30	2,63	3,0	3,3	3,4	3,6	3,7	3,8	3,9	4,0	4,1	4,2
60	1,96	2,2	2,3	2,4	2,4	2,5	2,5	2,6	2,6	2,7	2,7
∞	1,00	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0

Die in Klammern gesetzten Ziffern (für $\nu = 3, 7 \leq k \leq 12$) sind unsicher, z. B. $\hat{F}_{\max}$ für $\nu = 3, k = 7$ ist etwa 216.

Beispiel: Prüfe die Homogenität der folgenden drei Stichprobengruppen mit den Umfängen $n_0 = 8$; $s_1^2 = 6,21$; $s_2^2 = 1,12$; $s_3^2 = 4,34$ ($\alpha = 0,05$). $\hat{F}_{\max} = (6,21/1,12) = 5,54 < 6,94 = F_{\max}$ {für $k = 3, \nu = n_0 - 1 = 7$ und $\alpha = 0,05$ }. Anhand der vorliegenden Stichproben lässt sich auf dem 5%-Niveau die Nullhypothese auf Homogenität der Varianzen nicht ablehnen.



7.5.1.2 Prüfung der Gleichheit mehrerer Varianzen nach Cochran

Wenn eine Gruppenvarianz ($s_{\max}^2$) wesentlich größer ist als die übrigen, bevorzuge man diesen Test [Coc41]. Prüfgröße ist

$$\hat{G}_{\max} = \frac{s_{\max}^2}{s_1^2 + s_2^2 + \dots + s_k^2} \quad (7.139)$$

Die Beurteilung von $\hat{G}_{\max}$ erfolgt mit Hilfe der Tabelle 7.40: Ist $\hat{G}_{\max}$ größer als der für $k, \nu = n_0 - 1$ und das gewählte Niveau tabellierte Wert, wobei n_0 den Umfang der einzelnen Gruppen darstellt, dann muss die Nullhypothese auf Gleichheit der Varianzen abgelehnt und die Alternativhypothese: $\sigma_{\max}^2 \neq \sigma^2$ akzeptiert werden.

Tabelle 7.40. Signifikanzschranken für den Test nach Cochran (aus Eisenhart, C., Hastay, M.W., und W.A. Wallis: Techniques of Statistical Analysis, McGraw-Hill, New York 1947)

$\alpha = 0,05$

$k \backslash \nu$	1	2	3	4	5	6	7	8	9	10	16	36	144	∞
2	0,9985	0,9750	0,9392	0,9057	0,8772	0,8534	0,8332	0,8159	0,8010	0,7880	0,7341	0,6602	0,5813	0,5000
3	0,9669	0,8709	0,7977	0,7457	0,7071	0,6771	0,6530	0,6333	0,6167	0,6025	0,5466	0,4748	0,4031	0,3333
4	0,9065	0,7679	0,6841	0,6287	0,5895	0,5598	0,5365	0,5175	0,5017	0,4884	0,4366	0,3720	0,3093	0,2500
5	0,8412	0,6838	0,5981	0,5441	0,5065	0,4783	0,4564	0,4387	0,4241	0,4118	0,3645	0,3066	0,2513	0,2000
6	0,7808	0,6161	0,5321	0,4803	0,4447	0,4184	0,3980	0,3817	0,3682	0,3568	0,3135	0,2612	0,2119	0,1667
7	0,7271	0,5612	0,4800	0,4307	0,3974	0,3726	0,3535	0,3384	0,3259	0,3154	0,2756	0,2278	0,1833	0,1429
8	0,6798	0,5157	0,4377	0,3910	0,3595	0,3362	0,3185	0,3043	0,2926	0,2829	0,2462	0,2022	0,1616	0,1250
9	0,6385	0,4775	0,4027	0,3584	0,3286	0,3067	0,2901	0,2768	0,2659	0,2568	0,2226	0,1820	0,1446	0,1111
10	0,6020	0,4450	0,3733	0,3311	0,3029	0,2823	0,2666	0,2541	0,2439	0,2353	0,2032	0,1655	0,1308	0,1000
12	0,5410	0,3924	0,3264	0,2880	0,2624	0,2439	0,2299	0,2187	0,2098	0,2020	0,1737	0,1403	0,1100	0,0833
15	0,4709	0,3346	0,2758	0,2419	0,2195	0,2034	0,1911	0,1815	0,1736	0,1671	0,1429	0,1144	0,0889	0,0667
20	0,3894	0,2705	0,2205	0,1921	0,1735	0,1602	0,1501	0,1422	0,1357	0,1303	0,1108	0,0879	0,0675	0,0500
24	0,3434	0,2354	0,1907	0,1656	0,1493	0,1374	0,1286	0,1216	0,1160	0,1113	0,0942	0,0743	0,0567	0,0417
30	0,2929	0,1980	0,1593	0,1377	0,1237	0,1137	0,1061	0,1002	0,0958	0,0921	0,0771	0,0604	0,0457	0,0333
40	0,2370	0,1576	0,1259	0,1082	0,0968	0,0887	0,0827	0,0780	0,0745	0,0713	0,0595	0,0462	0,0347	0,0250
60	0,1737	0,1131	0,0895	0,0765	0,0682	0,0623	0,0583	0,0552	0,0520	0,0497	0,0411	0,0316	0,0234	0,0167
120	0,0998	0,0632	0,0495	0,0419	0,0371	0,0337	0,0312	0,0292	0,0279	0,0266	0,0218	0,0165	0,0120	0,0083
∞	0	0	0	0	0	0	0	0	0	0	0	0	0	0

$\alpha = 0,01$

$k \backslash \nu$	1	2	3	4	5	6	7	8	9	10	16	36	144	∞
2	0,9999	0,9950	0,9794	0,9586	0,9373	0,9172	0,8988	0,8823	0,8674	0,8539	0,7949	0,7067	0,6062	0,5000
3	0,9933	0,9423	0,8831	0,8335	0,7933	0,7606	0,7335	0,7107	0,6912	0,6743	0,6059	0,5153	0,4230	0,3333
4	0,9676	0,8643	0,7814	0,7212	0,6761	0,6410	0,6129	0,5897	0,5702	0,5536	0,4884	0,4057	0,3251	0,2500
5	0,9279	0,7885	0,6957	0,6329	0,5875	0,5531	0,5259	0,5037	0,4854	0,4697	0,4094	0,3351	0,2644	0,2000
6	0,8828	0,7218	0,6258	0,5635	0,5195	0,4866	0,4608	0,4401	0,4229	0,4084	0,3529	0,2858	0,2229	0,1667
7	0,8376	0,6644	0,5685	0,5080	0,4659	0,4347	0,4105	0,3911	0,3751	0,3616	0,3105	0,2494	0,1929	0,1429
8	0,7945	0,6152	0,5209	0,4627	0,4226	0,3932	0,3704	0,3522	0,3373	0,3248	0,2779	0,2214	0,1700	0,1250
9	0,7544	0,5727	0,4810	0,4251	0,3870	0,3592	0,3378	0,3207	0,3067	0,2950	0,2514	0,1992	0,1521	0,1111
10	0,7175	0,5358	0,4469	0,3934	0,3572	0,3308	0,3106	0,2945	0,2813	0,2704	0,2297	0,1811	0,1376	0,1000
12	0,6528	0,4751	0,3919	0,3428	0,3099	0,2861	0,2680	0,2535	0,2419	0,2320	0,1961	0,1535	0,1157	0,0833
15	0,5747	0,4069	0,3317	0,2882	0,2593	0,2386	0,2228	0,2104	0,2002	0,1918	0,1612	0,1251	0,0934	0,0667
20	0,4799	0,3297	0,2654	0,2288	0,2048	0,1877	0,1748	0,1646	0,1567	0,1501	0,1248	0,0960	0,0709	0,0500
24	0,4247	0,2871	0,2295	0,1970	0,1759	0,1608	0,1495	0,1406	0,1338	0,1283	0,1060	0,0810	0,0595	0,0417
30	0,3632	0,2412	0,1913	0,1635	0,1454	0,1327	0,1232	0,1157	0,1100	0,1054	0,0867	0,0658	0,0480	0,0333
40	0,2940	0,1915	0,1508	0,1281	0,1135	0,1033	0,0957	0,0898	0,0853	0,0816	0,0668	0,0503	0,0363	0,0250
60	0,2151	0,1371	0,1069	0,0902	0,0796	0,0722	0,0668	0,0625	0,0594	0,0567	0,0461	0,0344	0,0245	0,0167
120	0,1225	0,0759	0,0585	0,0489	0,0429	0,0387	0,0357	0,0334	0,0316	0,0302	0,0242	0,0178	0,0125	0,0083
∞	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Bei nicht zu ungleichen Stichprobenumfängen [vgl. die Bemerkung in [7.5.2] unter (7.150)] berechne man ihr harmonisches Mittel $\bar{x}_H$ und interpoliere in Tabelle 7.40 für $\nu = \bar{x}_H - 1$.



Beispiel: Angenommen, es liegen die folgenden 5 Varianzen vor: $s_1^2 = 26$, $s_2^2 = 51$, $s_3^2 = 40$, $s_4^2 = 24$ und $s_5^2 = 28$, wobei jede Varianz auf 9 Freiheitsgraden basiert. Getestet werden soll auf dem 5%-Niveau. Dann ist $\hat{G}_{\max} = 51/(26 + 51 + 40 + 24 + 28) = 0,302$. Für $\alpha = 0,05$, $k = 5$, $\nu = 9$ erhalten wir den Tabellenwert 0,4241. Da $0,302 < 0,4241$, kann an der Gleichheit der vorliegenden Varianzen nicht gezweifelt werden ($P = 0,05$).

7.5.1.3 Prüfung der Gleichheit mehrerer Varianzen nach Bartlett

In den meisten Fällen führen die Tests von Hartley und Cochran zu denselben Entscheidungen. Da der Cochran-Test mehr Informationen nutzt, ist er im allgemeinen etwas empfindlicher. Weitere Empfehlungen enthält die folgende Übersicht und dieser Abschnitt [7.5.1.3].

Grundgesamtheit	Testverfahren
leicht schief verteilt	Cochran-Test
normalverteilt, $N(\mu, \sigma^2)$	$k < 10$: Hartey, Cochran; $k \geq 10$: Bartlett
flacher als $N(\mu, \sigma^2)$	Levene
höher als $N(\mu, \sigma^2)$	$k < 10$: Cochran; $k \geq 10$: Levene

Die Nullhypothese, Homogenität mehrerer Varianzen, kann beim Vorliegen gut normalverteilter Daten nach Bartlett [Bar37] geprüft werden. Bartletts Test ist die Kombination eines empfindlichen Tests auf Normalität, besser „longtailedness“ einer Verteilung, mit einem weniger empfindlichen Test auf Gleichheit der Varianzen.

$$\hat{\chi}^2 = \frac{1}{c} \left[2,3026(\nu \lg s^2 - \sum_{i=1}^k \nu_i \lg s_i^2) \right]$$

mit

$$c = \frac{\sum_{i=1}^k \frac{1}{\nu_i} - \frac{1}{\nu}}{3(k-1)} + 1$$

$$s^2 = \frac{\sum_{i=1}^k \nu_i s_i^2}{\nu} \quad \text{und} \quad FG = k - 1$$
(7.140)

$$\nu = n - k = \text{Gesamtzahl der Freiheitsgrade} = \sum_{i=1}^k \nu_i$$

k = Anzahl der Gruppen: Jede Gruppe muss mindestens 5 Beobachtungen enthalten

s^2 = Schätzung der gewogenen Varianz

ν_i = Anzahl der Freiheitsgrade in der i -ten Gruppe = $n_i - 1$

s_i^2 = Schätzung der Varianz der i -ten Gruppe

Für nicht zu kleine Freiheitsgrade ν_i ist c praktisch gleich Eins, d. h. c braucht nur berechnet zu werden, wenn der Wert der eckigen Klammer ein statistisch signifikantes $\hat{\chi}^2$ erwarten lässt.

Liegen k Stichprobengruppen gleichen Umfangs n_0 vor, wobei $n_0 \geq 5$, dann ergeben sich folgende Vereinfachungen

$$\hat{\chi}^2 = \frac{1}{c} \left[2,3026k(n_0 - 1) \left\{ \lg s^2 - \frac{1}{k} \sum_{i=1}^k \lg s_i^2 \right\} \right]$$

mit

$$c = \frac{k+1}{3k(n_0 - 1)} + 1$$

$$s^2 = \frac{1}{k} \sum_{i=1}^k s_i^2 \quad (FG = k - 1)$$
(7.141)

Übersteigt die nach (7.140) bzw. (7.141) berechnete Prüfgröße $\hat{\chi}^2$ die für die geforderte Irrtumswahrscheinlichkeit α gegebene Signifikanzschranke $((1 - \alpha)$ -Quantil der χ^2 -Verteilung), so ist die Nullhypothese (H_0) $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_i^2 = \dots = \sigma_k^2 = \sigma^2$ abzulehnen (Alternativhypothese $\sigma_i^2 \neq \sigma^2$ für bestimmte i).



Beispiel: Gegeben sind drei Stichprobengruppen mit den Umfängen $n_1 = 9$, $n_2 = 6$ und $n_3 = 5$ sowie den in der folgenden Tabelle angegebenen Varianzen. Prüfe die Gleichheit der Varianzen auf dem 5%-Niveau.

Nr.	s_i^2	$\nu_i = n_i - 1$	$\nu_i s_i^2$	$\lg(s_i^2)$	$\nu_i \lg(s_i^2)$
1	8,00	8	64,00	0,9031	7,2248
2	4,67	5	23,35	0,6693	3,3465
3	4,00	4	16,00	0,6021	2,4084
		17	103,35		12,9797

$$s^2 = \frac{103,35}{17} = 6,079, \quad \lg s^2 = 0,7838$$

$$\hat{\chi}^2 = \frac{1}{c} [2,3026(17 \cdot 0,7838 - 12,9797)] = \frac{1}{c} \cdot 0,794$$

Da $\chi_{2;0,95}^2 = 5,99$ wesentlich größer ist als 0,794, wird H_0 auf dem 5%-Niveau nicht abgelehnt. Mit c

$$c = \frac{\left[\frac{1}{8} + \frac{1}{5} + \frac{1}{4} \right] - \frac{1}{17}}{3(3-1)} + 1 = 1,086$$

ergibt sich $\hat{\chi}^2 = 0,794/1,086 = 0,731 < 5,99 = \chi_{2;0,95}^2$.

Die Funktion **bartlett.test()** in **R** berechnet den Bartlett-Test für unabhängige normalverteilte Zufallstichproben. Dazu werden die Werte der Stichprobengruppen durch die Funktion **list()** zusammengefasst als Argument der Funktion bereitgestellt.



Beispiel: Vergleich von 3 Stichprobengruppen (x, y und z) hinsichtlich der Gleichheit der Varianzen in **R**. Zunächst soll für die Beispieldaten die Teststatistik nach Bartlett (7.140) elementar berechnet ($\hat{\chi}^2 = 10,367$) und mit dem Quantil der χ^2 -Verteilung für $\alpha = 0.05$ verglichen ($\chi_{2;0,95}^2 = 5,99$) werden. Mit dem Wert der Teststatistik kann auch der P-Wert aus der χ^2 -Verteilung direkt bestimmt werden ($P = 0,0056$). Die Ergebnisse aus der Funktion **bartlett.test()** stimmen mit den so berechneten Ergebnissen überein.

```
> x <- c( 9, 11,  6, 11, 14,  7,  7, 11)
> y <- c(13, 10, 12, 16, 11, 13, 15,  9,  9, 10)
> z <- c( 7, 27,  8, 11, 17,  2, 16, 15,  9, 15, 18, 12)
>
> k <- 3
> si <- c(sd(x), sd(y), sd(z)); si
[1] 2.725541 2.440401 6.444989
> nui <- c(length(x)-1, length(y)-1, length(z)-1); nu <- sum(nui)
> c <- (sum(1/nui) - 1/nu)/(3*(k-1)) + 1
> ssqr <- sum(nui*si^2)/nu
>
> chisqr <- 1/c*(2.3026 * (nu*log10(ssqr)-sum(nui*log10(si^2))))
[1] 10.36702
> qchisq(0.95, k-1)
[1] 5.991465
> pchisq(chisqr, k-1, lower.tail=F)
[1] 0.005608289
>
> bartlett.test(list(x,y,z))
```


Bartlett test for homogeneity of variances

data: list(x, y, z)

Bartlett's K-squared = 10.367, df = 2, p-value = 0.005608

7.5.1.4 Robuster Test auf Varianzheterogenität nach Levene in der Brown-Forsythe-Version

Die Homogenität mehrerer (k) Varianzen lässt sich für $n_i \geq 10$ nach **Levene** [Lev60] in der **Brown-Forsythe-Version** [BF74] auch mit Hilfe einer einfachen Varianzanalyse ablehnen. Für k unabhängige Stichprobengruppen mit jeweils mindestens 10 Beobachtungen wird die Nullhypothese: gleiche Varianzen

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2$$

geprüft. Die Alternativhypothese lautet: mindestens zwei Varianzen sind ungleich.

$$H_A : \sigma_i^2 \neq \sigma_j^2$$

H_0 wird auf dem $100\alpha\%$ -Niveau abgelehnt, sobald für die nach $y_{ij} = |x_{ij} - \tilde{x}_i|$ transformierten Beobachtungen ($\tilde{x}_i$ ist der Median der i -ten Stichprobengruppe), also für die jetzt vorliegenden y_{ij} -Werte, das $\hat{F}$ der Varianzanalyse größer ist als $F_{k-1; n-k; 1-\alpha}$. $\hat{F}$ wird nach (7.150) oder nach (7.149) berechnet, entsprechend der Symbolik von [7.5.2].

7.5.2 Einfache Varianzanalyse

Der Vergleich zweier Erwartungswerte normalverteilter Grundgesamtheiten (Abschnitt 7.4.4) lässt sich auf den Vergleich einer beliebigen Zahl von Erwartungswerten erweitern.

Gegeben seien k Stichprobengruppen mit je n_i und insgesamt n Stichprobenelementen, also:

$$\sum_{i=1}^k n_i = n$$

Jede Stichprobengruppe entstamme einer normalverteilten Grundgesamtheit. Die k normalverteilten Grundgesamtheiten haben gleiche Varianzen; diese sind unbekannt: „Zufallsstichproben mit unbekannter gemeinsamer Varianz“.

Notation: Die Stichprobenwerte x_{ij} erhalten zwei Indizes: x_{ij} ist der j -te Wert in der i -ten Stichprobe ($1 \leq i \leq k$; $1 \leq j \leq n_i$).

Die **Gruppenmittelwerte** $\bar{x}_i$ sind gegeben durch

$$\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}$$

Der **Punkt** deutet den Index an, über den summiert worden ist; so ist z. B.

$$x_{..} = \sum_{i=1}^k \sum_{j=1}^{n_i} x_{ij}$$

 die Summe aller x -Werte

Das **Gesamtmittel** $\bar{x}$:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^{n_i} x_{ij} = \frac{1}{n} \sum_{i=1}^k n_i \bar{x}_i. \quad (7.142)$$

in vereinfachter Schreibweise:

$$\bar{x} = \frac{1}{n} \sum_{i,j} x_{ij} = \frac{1}{n} \sum_i n_i \bar{x}_i. \quad (7.143)$$

Wesentlich für die einfache „Varianzanalyse“, auch einfache „**Streuungszerlegung**“ genannt, ist, dass sich die **Summe der Abweichungsquadrate** (SAQ oder Q) der Stichprobenwerte um das Gesamtmittel („ Q insgesamt“) in zwei Anteile zerlegen lässt, in die

1. SAQ der Einzelwerte um die Gruppenmittelwerte, „ SAQ **innerhalb der Gruppen**“ genannt („ $Q_{\text{innerhalb}}$ “) und in die
2. SAQ der Gruppenmittelwerte um das Gesamtmittel, „ SAQ **zwischen den Gruppen**“ genannt („ Q_{zwischen} “), d. h.

$$Q_{\text{insgesamt}} = Q_{\text{innerhalb}} + Q_{\text{zwischen}}$$

$$\sum_{i,j} (x_{ij} - \bar{x})^2 = \sum_{i,j} (x_{ij} - \bar{x}_i)^2 + \sum_i n_i (\bar{x}_i - \bar{x})^2 \quad (7.144)$$

mit den zugehörigen Freiheitsgraden

$$(n - 1) = (n - k) + (k - 1) \quad (7.145)$$

Die Quotienten aus den SAQ und den zugehörigen FG , d. h. die Varianzen Q/ν bezeichnet man in der Varianzanalyse als „**Mittlere Quadrate**“ (MQ). Entstammen alle Gruppen derselben Grundgesamtheit, dann sollten die Varianzen, also die Mittlere Quadrate

$$s_{\text{zwischen}}^2 = MQ_{\text{zwischen}} = \frac{1}{k - 1} \sum_i n_i (\bar{x}_i - \bar{x})^2 \quad (7.146)$$

und

$$s_{\text{innerhalb}}^2 = MQ_{\text{innerhalb}} = \frac{1}{n - k} \sum_{i,j} (x_{ij} - \bar{x}_i)^2 \quad (7.147)$$

ungefähr gleich groß sein. Sind sie es nicht, d. h. ist der Quotient aus MQ_{zwischen} und $MQ_{\text{innerhalb}}$ größer als der durch $\nu_1 = k - 1$, $\nu_2 = n - k$ und α festgelegte kritische Wert der F -Verteilung, so befinden sich unter den Gruppen solche mit unterschiedlichen Erwartungswerten μ_i .

Die Nullhypothese $\mu_1 = \mu_2 = \dots = \mu_i = \dots = \mu_k = \mu$ wird anhand der Prüfgröße (7.148) [d. h. (7.149) bzw. (7.150)] abgelehnt, wenn

$$\hat{F} > F_{(k-1; n-k; 1-\alpha)}.$$

In diesem Fall sind mindestens zwei μ_i voneinander verschieden, d. h. die Alternativhypothese $\mu_i \neq \mu$ für bestimmte i wird akzeptiert.

Wenn $MQ_{\text{zwischen}} < MQ_{\text{innerhalb}}$ ist, lässt sich die Nullhypothese nicht ablehnen, dann sind (7.142) und (7.147) Schätzungen für μ sowie für σ^2 mit $n - k$ Freiheitsgraden.

Man bezeichnet MQ_{zwischen} auch als „**Stichprobenfehler**“ und $MQ_{\text{innerhalb}} = s_{\text{innerhalb}}^2$ als „**Versuchsfehler**“.

$$\hat{F} = \frac{MQ_{\text{zwischen}}}{MQ_{\text{innerhalb}}} = \frac{\frac{1}{k-1} \sum_i n_i (\bar{x}_{i.} - \bar{x})^2}{\frac{1}{n-k} \sum_{i,j} (x_{ij} - \bar{x}_{i.})^2} = \frac{\frac{1}{k-1} \sum_i n_i (\bar{x}_{i.} - \bar{x})^2}{\frac{1}{n-k} \sum_i s_i^2 (n_i - 1)} \quad (7.148)$$

$\hat{F}$ wird berechnet nach (7.149):

$$\hat{F} = \frac{\frac{1}{k-1} \left[\sum_i \frac{x_{i.}^2}{n_i} - \frac{x_{..}^2}{n} \right]}{\frac{1}{n-k} \left[\sum_{i,j} x_{ij}^2 - \sum_i \frac{x_{i.}^2}{n_i} \right]} \quad (7.149)$$

Für **Stichprobengruppen gleicher Umfänge** ($n_i = n_0$) bevorzugt man (7.150):

$$\hat{F} = \frac{\left[k \sum_i x_{i.}^2 - x_{..}^2 \right] / (k-1)}{\left[n_0 \sum_{i,j} x_{ij}^2 - \sum_i x_{i.}^2 \right] / (n_0-1)} \quad (7.150)$$

Die Wahl gleichgroßer Stichprobenumfänge pro Gruppe bietet mehrere Vorteile:

- (1) Abweichungen von der Varianzgleichheit sind nicht so schwerwiegend.
- (2) Der beim F -Test auftretende Fehler 2. Art wird minimal.
- (3) Weitere Erwartungwertvergleiche (vgl. multiple Vergleiche [7.5.3]) sind einfacher durchzuführen.

Beispiel: Ungleiche Stichprobenumfänge n_i pro Gruppe (betont einfach, um die Berechnung und die Interpretation der Teststatistik zu erklären):



Stichprobengruppe i			
Wert j	1	2	3
1	3	4	8
2	7	2	4
3		7	6
4		3	
$x_{i.}$	10	16	18
n_i	2	4	3
$\bar{x}_i$	5	4	6

$x_{..} = 44$
 $n = 9$
 $\bar{x} = 4,89$

Wenig elegant (direkt) nach (7.146; 7.147) berechnet:

$$MQ_{\text{zwischen}} = \frac{2 \left(5 - \frac{44}{9} \right)^2 + 4 \left(4 - \frac{44}{9} \right)^2 + 3 \left(6 - \frac{44}{9} \right)^2}{3-1} = \frac{6,89}{2}$$

$$MQ_{\text{innerhalb}}$$

$$= \frac{[(3-5)^2 + (7-5)^2] + [(4-4)^2 + (2-4)^2 + (7-4)^2 + (3-4)^2] + [(8-6)^2 + (4-6)^2 + (6-6)^2]}{9-3}$$

$$= \frac{30}{6}$$

Nach (7.149) und (7.150):

$$\hat{F} = \frac{\frac{1}{3-1} \left[\left(\frac{10^2}{2} + \frac{16^2}{4} + \frac{18^2}{3} \right) - \frac{44^2}{9} \right]}{\frac{1}{9-3} \left[(3^2 + 7^2 + 4^2 + 2^2 + 7^2 + 3^2 + 8^2 + 4^2 + 6^2) - \left(\frac{10^2}{2} + \frac{16^2}{4} + \frac{18^2}{3} \right) \right]}$$

$$\hat{F} = \frac{\frac{1}{2}[6,89]}{\frac{1}{6}[30]} = 0,689$$

Da $\hat{F} = 0,689 < 5,14 = F_{(2;6;0,95)}$, lässt sich die Nullhypothese, alle drei Erwartungswerte entstammen derselben Grundgesamtheit

$$\text{mit (7.142) } \bar{x} = (2 \cdot 5 + 4 \cdot 4 + 3 \cdot 6)/9 = 4,89$$

$$\text{und (7.147) } s_{\text{innerhalb}}^2 = 30/6 = 5$$

auf dem 5%-Niveau nicht ablehnen.

In **R** kann eine einfache Varianzanalyse mit der Funktion **aov()** berechnet werden. Der Fragestellung (Versuchsaufbau) entsprechend werden die Zahlenwerte in einer speziellen Datenstruktur („**data.frame**“) gespeichert. Dabei ist besonders darauf zu achten, dass die Zuordnung zu den Stichprobengruppen durch eine Variable vom Typ „**factor**“ erfolgt (einfaktorielle Varianzanalyse, oneway analysis of variances).

```
> gruppe <- c(1, 1, 2, 2, 2, 2, 3, 3, 3)
> wert <- c(3, 7, 4, 2, 7, 3, 8, 4, 6)
> daten <- data.frame(gruppe=factor(gruppe), wert); daten
```

gruppe	wert
1	3
2	7
3	4
4	2
5	7
6	3
7	8
8	4
9	6

```
> summary(aov(wert ~ gruppe, data=daten))
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
gruppe	2	6.8889	3.4444	0.6889	0.5379
Residuals	6	30.0000	5.0000		

Die Funktion **summary()** gibt in diesem Fall die klassische Ergebnistabelle für die Varianzanalyse aus. Für den Faktor „gruppe“ (Zwischeneffekt) und die „residuals“ (Abweichungen innerhalb der Gruppen, Versuchsfehler) werden die Varianzkomponenten einzeln aufgelistet. Der Wert der Teststatistik (F value) stimmt mit dem oben abgeleiteten Ergebnis überein. Eine Testentscheidung kann hier auf der Grundlage des P-Wertes (P=0,54) erfolgen.

Beispiel: Gleichgroße Stichprobenumfänge ($n_i = \text{konst.} = n_0$) pro Gruppe:



Stichprobengruppe i				
Wert j	1	2	3	
1	6	5	7	
2	7	6	8	
3	6	4	5	
4	5	5	8	
$x_{i.}$	24	20	28	$x_{..} = 72$
$n_i = n_0$	4	4	4	$n = 16$
$\bar{x}_i$	6	5	7	$\bar{x} = 6$

Nach (7.149):

$$\hat{F} = \frac{\frac{1}{3-1} \left[\frac{1}{4}(24^2 + 20^2 + 28^2) - \frac{72^2}{12} \right]}{\frac{1}{12-3} \left[(6^2 + 7^2 + \dots + 8^2) - \frac{1}{4}(24^2 + 20^2 + 28^2) \right]} = \frac{\frac{1}{2}[8]}{\frac{1}{9}[10]} = 3,60$$

Nach (7.150):

$$\hat{F} = \frac{[3(24^2 + 20^2 + 28^2) - 72^2]/(3-1)}{[4(6^2 + 7^2 + \dots + 8^2) - (24^2 + 20^2 + 28^2)]/(4-1)} = \frac{96/2}{40/3} = 3,60$$

Da $\hat{F} = 3,60 < 4,26 = F_{(2;9;0,95)}$, lässt sich die Nullhypothese, Gleichheit der 3 Erwartungswerte ($\bar{x} = 6$, $s_{\text{innerhalb}}^2 = 10/9 = 1,11$), auf dem 5%-Niveau nicht ablehnen.

```
> gruppe <- c(rep(1,4), rep(2,4), rep(3,4))
> wert <- c(6, 7, 6, 5, 5, 6, 4, 5, 7, 8, 5, 8)
> daten <- data.frame(gruppe=factor(gruppe), wert)
> summary(aov(wert ~ gruppe, daten))
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
gruppe	2	8.0000	4.0000	3.6	0.071
Residuals	9	10.0000	1.1111		

7.5.2.1 Stichprobenumfänge für den varianzanalytischen Mittelwertvergleich von maximal sechs Stichprobengruppen

Sind Mittelwerte mehrerer Stichprobengruppen zu vergleichen, so wird man zunächst $(\bar{x}_{\max} - \bar{x}_{\min})/s = \hat{\tau}$ schätzen und sich dann überlegen, wie umfangreich die k möglichst gleichgroßen Stichprobengruppen sein sollten. Tabelle 7.41 basiert auf einer Power von nur 70%; wird eine größere Power verlangt, so steigen die jeweils benötigten Stichprobenumfänge schnell an (vgl. Kastenbaum und Mitarbeiter [KHB70]). Für **rangvarianzanalytische Vergleiche vom Typ des H-Tests** wird man mit, sagen wir, etwa $1,2n$ Beobachtungen pro Stichprobengruppe rechnen müssen, um mit einer Power von 0,7 auf dem 5%-Niveau statistisch signifikante Unterschiede erfassen zu können.

Beispiel: Fünf Mittelwerte sind zu vergleichen ($\alpha = 0,05$; Power: 0,7); für $\hat{\tau} = (\bar{x}_{\max} - \bar{x}_{\min})/s = (36 - 30)/6 = 1$ sollten jeweils 20 Beobachtungen zur Verfügung stehen (Normalverteilung und homogene Varianzen vorausgesetzt). Insgesamt werden somit mindestens 100 Beobachtungen benötigt.



Tabelle 7.41. Jeweils benötigte Stichprobenumfänge für den varianzanalytischen k -Stichproben-Vergleich sind für festes k und τ der Kopfzeile zu entnehmen; unter ihnen die entsprechenden Werte einer erwarteten standardisierten Maximalabweichung $\tau = (\mu_{\max} - \mu_{\min})/\sigma$, die auf dem 5%-Niveau mit einer Power von 0,7 bei einem Vergleich von k Stichprobengruppen der Umfänge n zu erfassen ist; einige Werte aus den von Kastenbaum und Mitarbeitern (1970) gegebenen Tabellen

$k \backslash n$	6	10	20	30	40	60
3	1,8	1,3	0,90	0,73	0,63	0,51
4	1,9	1,4	0,96	0,78	0,67	0,55
5	2,0	1,5	1,01	0,82	0,70	0,57
6	2,0	1,5	1,05	0,85	0,73	0,60

7.5.3 Multiple Vergleiche, Multiples Testproblem

Prüft man auf dem 5%-Niveau, so ist bei einem Test mit 5 von 100 falsch positiven Resultaten zu rechnen, bei zwanzig Tests sind es bereits $[1 - 0,95^{20} = 0,64]$ 64 von 100; d. h. bei Mehrfachtestung steigt das Risiko, falsch positive Resultate zu erhalten, stark an. Einen Überblick gibt Tabelle 7.42.

Tabelle 7.42. Wahrscheinlichkeit für kein bzw. mindestens ein positives Testergebnis, wenn bei Gültigkeit der Nullhypothese auf dem 5%-Signifikanzniveau m -fach getestet wird (aus Sachs [Sac93])

Zahl unabh. Tests / Vergleiche m	gerundete Wahrscheinlichkeit für kein * mindestens 1 ** falsch positives Ergebnis	
1	0,95	0,05
2	0,90	0,10
3	0,86	0,14
4	0,81	0,19
5	0,77	0,23
7	0,70	0,30
10	0,60	0,40
13	0,51	0,49
14	0,49	0,51
15	0,46	0,54
20	0,36	0,64
30	0,21	0,79
40	0,13	0,87
44	0,10	0,90
50	0,08	0,92
58	0,05	0,95
89	0,01	0,99

* $P = 0,95^m$; ** $P = 1 - (1 - 0,05)^m = 1 - 0,95^m$



Beispiel: $m = 14$ Tests auf dem 5%-Niveau führen bei Gültigkeit der Nullhypothese mit der Wahrscheinlichkeit $P=0,51$ zu mindestens einem falsch positiven (auf dem 5%-Niveau statistisch signifikanten) Testresultat.

Ein Homogenitätstest für den Vergleich von k Populationen prüft die Nullhypothese

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_k \quad (\text{Globalhypothese})$$

Dabei bezeichnen die μ_i ($i = 1, \dots, k$) die unbekannten Erwartungswerte in den Populationen. Ein signifikantes Ergebnis aus der Varianzanalyse bedeutet in diesem Zusammenhang, dass sich mindestens ein μ_i von einem μ_j unterscheidet. Damit liegen jedoch insgesamt $k(k-1)/2$ mögliche sogenannte Paarhypothesen oder **Elementarhypothesen** $H_{0ij} : \mu_i \neq \mu_j$ vor. Ein multiples Testproblem liegt vor, wenn m ($m > 1$) Hypothesen zu prüfen sind. Ein **multipler Test** ist dann ein Verfahren, das zur Entscheidung über Ablehnung oder Beibehaltung dieser m Elementarhypothesen führt. Dabei unterscheidet man verschiedene Arten von Signifikanzniveaus [HV95].

- (1) Elementarhypothesen können fälschlicherweise mit bestimmten Wahrscheinlichkeiten abgelehnt werden. Lässt sich eine Schranke α angeben, die von keiner dieser Irrtumswahrscheinlichkeiten überschritten wird, dann wird bei diesem Testverfahren das **lokale Signifikanzniveau** α nicht überschritten.
- (2) Unter der Annahme, dass die globale Hypothese wahr ist, kann es mit einer bestimmten Wahrscheinlichkeit zur fälschlichen Ablehnung mindestens einer der Elementarhypothesen kommen. Lässt sich nun eine Schranke α angeben, die von dieser unbekannten Wahrscheinlichkeit nicht überschritten wird, dann wird das **globale Signifikanzniveau** α eingehalten.
- (3) Die Wahrscheinlichkeit dafür, dass wenigstens eine Elementarhypothese fälschlicherweise abgelehnt wird, hängt davon ab, wie viele und welche letztlich tatsächlich wahr sind. Für jede Kombination von wahren und unwahren Elementarhypothesen ist dabei grundsätzlich eine andere Irrtumswahrscheinlichkeit denkbar. Lässt sich eine Schranke α angeben, die von keiner dieser Wahrscheinlichkeiten überschritten wird, dann bezeichnet man α als das **multiple Signifikanzniveau**. Ziel der Verfahren in den folgenden Abschnitten ist es, in einem multiplen Vergleich einzelne Unterschiede oder Effekte aufzudecken, und dabei eine feste vorgegebene Schranke für das multiple Signifikanzniveau einzuhalten.

7.5.3.1 Simultane multiple Vergleiche (Konfidenzintervalle)

Betrachtet man zwei 95%-Konfidenzintervall, die aufgrund zweier Zufallsstichproben aus unterschiedlichen und unabhängigen Grundgesamtheiten berechnet worden sind, dann gilt für die Wahrscheinlichkeit, dass beide gemeinsam gültig sind, $P = 0,95 \cdot 0,95 = 0,9025$. Generell erhält man als untere Grenze nach Bonferroni:

Ungleichung nach Bonferroni

$$P(E_1 \cap E_2 \cap \dots \cap E_k) \geq 1 - \sum_{i=1}^k P(\bar{E}_i) \quad (7.151)$$

$$\begin{aligned} P(A \cap B) &\geq 1 - P(\bar{A}) - P(\bar{B}) \\ P(0,95 \cap 0,95) &\geq 1 - 0,05 - 0,05 = 0,90 \end{aligned} \quad (7.152)$$

d.h. die Wahrscheinlichkeit, mindestens eine fehlerhafte Aussage zu machen, ist somit sicherlich größer als die für jedes 95%-KI festgelegte Irrtumswahrscheinlichkeit von 0,05: sie kann maximal sogar $1 - 0,90 = 0,10$ betragen.

Gibt man z.B. fünf unabhängige 95%-Konfidenzintervalle an, so ist zu bedenken, dass die Wahrscheinlichkeit, wenigstens eine fehlerhafte Aussage zu machen, $P = 1 - 0,95^5 = 0,2265$ beträgt, bei 10 KIen erhält man bereits $P = 0,4013$. Um dieses zu vermeiden, wendet man sogenannte

simultane Konfidenzintervalle für k 95%-KIE an, die für festes k garantieren, dass insgesamt die Vertrauenswahrscheinlichkeit $1 - 0,05 = 0,95$ nicht unterschritten wird und dass die betreffenden Parameter bzw. Parameterdifferenzen gleichzeitig mit $P = 0,95$ überdeckt werden. Mit zunehmendem k werden die simultanen KIE breiter, die Aussagen also ungenauer.

Die an dieser Stelle interessierenden multiplen Vergleiche betreffen den simultanen Vergleich aller Paare von Mittelwerten. Prüft man k Mittelwerte μ_i paarweise simultan auf dem 5%-Niveau oder gibt man simultan 95%-Konfidenzintervalle für die Differenzen $\mu_i - \mu_j$ an, so wird in beiden Fällen die wahre Irrtumswahrscheinlichkeit für sämtliche $k(k-1)/2$ Tests bzw. 95%-KIE nicht größer sein als 0,05, d.h. mit der Wahrscheinlichkeit $P = 1 - 0,05 = 0,95$ werden bei wahrer Nullhypothese (H_0) gültige Gleichheitsentscheidungen getroffen, einmal: „die $H_0 : \mu_i = \mu_j$ wird beibehalten“, zum anderen: „das 95%-KI für $\mu_i - \mu_j$ enthält die Null“. Für den Fall ungleicher Stichprobenumfänge oder/und ungleicher Varianzen sind es unter den im Einzelfall genannten Verfahren sehr gute Approximationen für $P = 0,95$, meist gilt $P \geq 0,95$.

7.5.3.2 Mehrfacher t-Test nach Bonferroni, Simes-Hochberg-Prozedur

Zwei Stichprobengruppen, die bezüglich mehrerer (k) angenähert normalverteilter Merkmale anhand eines t-Tests verglichen werden, müssen pro Merkmal auf dem $(100\alpha/k)\%$ -Signifikanzniveau geprüft werden, sobald der gesamte simultane Vergleich zweiseitig auf dem $100\alpha\%$ -Niveau (multiples Signifikanzniveau) durchgeführt wird. Etwa für $\nu = 28 + 34 - 2 = 60$ Freiheitsgrade auf dem 5%-Niveau und $k = 17$ Merkmale: $0,05/17 = 0,00294$. Das zweiseitige $t_{60;0,002954}$ -Quantil der t-Verteilung lässt sich aus erweiterten Tabellen dieser Verteilung durch Interpolation bestimmen bzw. direkt mit der Funktion **qt()** in **R** berechnen.

```
> qt(0.00147, 60, lower.tail=FALSE)
[1] 3.100673
```

Entsprechendes gilt auch, wenn viele Stichprobengruppen oder Behandlungen vorliegen und genau $k = 17$ Mittelwertvergleiche geplant sind, um zu erkunden, welche Mittelwerte sich paarweise auf einen vorgegebenen $100\alpha\%$ -Niveau unterscheiden.

Sequentiell und simultan verwerfende Bonferroni-Prozedur.

Das Verfahren soll am Beispiel für den Vergleich von 4 Mittelwerten beschrieben werden.

$$(1, 2, 3, 4) \rightarrow (1, 4); (1, 3); (2, 4); (1, 2); (2, 3); (3, 4)$$

Nach Bonferroni muss jeder dieser 6 Tests (dem Problem (1,4) entspricht in dieser Schreibweise die Hypothese $H_0 : \mu_1 = \mu_4$) auf dem Signifikanzniveau $\alpha/6$ durchgeführt werden. Nach **Holm** [Hol79] vergleicht man die geordneten P-Werte der 6 Tests mit $\alpha/6, \alpha/5, \alpha/4, \alpha/3, \alpha/2$ und $\alpha/1$. Ist das kleinste $P > \alpha/6$, dann kann die entsprechende Nullhypothese nicht abgelehnt werden, d.h. aber auch alle anderen Nullhypothesen können nicht abgelehnt werden; gilt $P < \alpha/6$, so wird H_0 abgelehnt und das nächstgrößere P mit $\alpha/5$ verglichen, usw. Die sequentiell verwerfende Holm-Prozedur (auch Bonferroni-Holm-Test genannt) weist natürlich eine höhere Power auf als die Bonferroni-Prozedur und ist dieser in der Regel vorzuziehen.

Simes-Hochberg-Prozedur für multiple Tests

Für vorgegebenes α liegen mehrere, sagen wir m P-Werte vor, die wir der Größe nach absteigend geordnet haben: $P_{(m)} \geq P_{(m-1)} \geq \dots \geq P_{(1)}$. Für $P_{(m)} \leq \alpha$ werden alle m Hypothesen abgelehnt. Wenn nicht, dann wird $P_{(m-1)}$ mit $\alpha/2$ verglichen; ist $P_{(m-1)} \leq \alpha/2$, so werden alle H_{0i} für $i = m-1, \dots, 1$ abgelehnt. Wenn nicht, d.h. $H_{0(m-1)}$ kann nicht abgelehnt werden, dann vergleicht man $P_{(m-2)}$ mit $\alpha/3$, usw. Diese Prozedur ist der oben genannten Bonferroni-Prozedur überlegen.

Mit der Funktion **p.adjust()** stehen in **R** neben der Bonferroni-Prozedur auch die Verfahren von Holm [Hol79], Simes-Hochberg [Hoc88], Hommel [Hom88] und Benjamini-Hochberg [BH95] zur Verfügung.

Die Verfahren von Hochberg und Hommel sind zulässig, wenn die den P-Werten zugrundeliegenden Hypothesentests unabhängig sind oder wenn sie untereinander nicht negativ assoziiert sind (Sarkar [SC97]). Dabei hat der Ansatz nach Hommel eine etwas höhere Power, während die Prozedur nach Hochberg schneller berechnet werden kann.

Die genannten Verfahren verfolgen das Ziel, dass in der Gesamtheit aller durchgeführten Tests höchstens mit der Wahrscheinlichkeit α eine fälschliche Ablehnung einer Nullhypothese auftritt, unabhängig davon, welche dieser Hypothesen tatsächlich falsch oder richtig sind (familywise error rate). Dagegen kontrolliert das Verfahren von Benjamini und Hochberg nur den Anteil falscher Entscheidungen bei der Ablehnung der Nullhypothese (false discovery rate, fdr) und ist damit weniger stringent als die anderen Verfahren (höhere Power).

Die Anwendung der Funktion **p.adjust()** wird an einem Beispiel mit insgesamt 9 verschiedenen P-Werten gezeigt.

```
> p <- c(0.0000,0.0076,0.0080,0.0088,0.0088,0.0092,0.0108,0.9479,0.9672)
> p.adjust(p, method="holm")
[1] 0.0000 0.0608 0.0608 0.0608 0.0608 0.0608 1.0000 1.0000
> p.adjust(p, method="hochberg")
[1] 0.0000 0.0324 0.0324 0.0324 0.0324 0.0324 0.0324 0.9672 0.9672
```

7.5.3.3 Beurteilung linearer Kontraste nach Scheffé

Wenn die einfache Varianzanalyse zu einem statistisch signifikanten Befund führt, wird man bestrebt sein, herauszufinden, welche der Parameter $\mu_1, \mu_2, \dots, \mu_i, \dots, \mu_k$, besser, welche zwei Gruppen A und B von Parametern mit den Erwartungswerten μ_A und μ_B sich unterscheiden. Liegen z. B. Schätzungen der fünf Parameter $\mu_1, \mu_2, \mu_3, \mu_4, \mu_5$ vor, dann lassen sich unter anderen die folgenden Erwartungswerte vergleichen:

$$V_1: \mu_1 = \mu_2 = \mu_A \quad \text{mit} \quad \mu_3 = \mu_4 = \mu_5 = \mu_B \\ \mu_A = \frac{1}{2}(\mu_1 + \mu_2) \quad \text{mit} \quad \mu_B = \frac{1}{3}(\mu_3 + \mu_4 + \mu_5)$$

$$V_2: \mu_1 = \mu_A \quad \text{mit} \quad \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_B \\ \mu_A = \mu_1 \quad \text{mit} \quad \mu_B = \frac{1}{4}(\mu_2 + \mu_3 + \mu_4 + \mu_5)$$

Vergleiche dieser Art, geschrieben

$$V_1: \left[\frac{1}{2}(\mu_1 + \mu_2) - \frac{1}{3}(\mu_3 + \mu_4 + \mu_5) \right]$$

$$V_2: \left[\mu_1 - \frac{1}{4}(\mu_2 + \mu_3 + \mu_4 + \mu_5) \right]$$

heißen **lineare Kontraste**. Sie sind lineare Funktionen der k Erwartungswerte μ_i (7.153), die durch k bekannte Konstanten c_i die die Bedingung (7.154)

$$\sum_{i=1}^k c_i \mu_i \quad (7.153)$$

$$\sum_{i=1}^k c_i = 0 \quad (7.154)$$

erfüllen, festgelegt sind. Diese Konstanten sind für

$$\begin{aligned} V_1: \quad c_1 = c_2 = \frac{1}{2}; \quad c_3 = c_4 = c_5 = -\frac{1}{3}; \quad \frac{1}{2} + \frac{1}{2} - \frac{1}{3} - \frac{1}{3} - \frac{1}{3} &= 0 \\ V_2: \quad c_1 = 1; \quad c_2 = c_3 = c_4 = c_5 = -\frac{1}{4}; \quad 1 - \frac{1}{4} - \frac{1}{4} - \frac{1}{4} - \frac{1}{4} &= 0 \end{aligned}$$

Wenn

$$\hat{S} = \frac{|\bar{x}_A - \bar{x}_B|}{s_{\bar{x}_A - \bar{x}_B}} > \sqrt{(k-1)F_{(k-1; n-k; 1-\alpha)}} = S_\alpha \quad (7.155)$$

mit

$$s_{\bar{x}_A - \bar{x}_B} = \sqrt{s_{in}^2 \sum_{i=1}^k \frac{c_i^2}{n_i}} \quad s_{in}^2 = MQ_{\text{innerhalb}} \quad (7.156)$$

unterscheiden sich die den Kontrasten zugrundeliegenden Parameter (Scheffé 1953 [Sch53]).

Sind nur 2 von k Werten μ_i zu vergleichen, etwa μ_3 und μ_5 , dann setzt man, wenn z. B. $k = 6$ ist, $c_1 = c_2 = c_4 = c_6 = 0$ und lehnt $H_0: \mu_3 = \mu_5$ ab, sobald

$$\hat{S} = \frac{|\bar{x}_3 - \bar{x}_5|}{\sqrt{s_{in}^2 \left(\frac{1}{n_3} + \frac{1}{n_5} \right)}} > \sqrt{(k-1)F_{(k-1; n-k; 1-\alpha)}} = S_\alpha \quad (7.157)$$

Für den Fall markant ungleich großer Gruppen bildet man **gewichtete lineare Kontraste**, also z. B. für V_1

$$\frac{n_1\mu_1 + n_2\mu_2}{n_1 + n_2} - \frac{n_3\mu_3 + n_4\mu_4 + n_5\mu_5}{n_3 + n_4 + n_5}$$

geschätzt nach

$$\frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2} - \frac{n_3\bar{x}_3 + n_4\bar{x}_4 + n_5\bar{x}_5}{n_3 + n_4 + n_5}.$$



Beispiele: Vergleiche zwischen 5 Stichproben mit (I) gleichen und (II) ungleichen Stichprobenumfängen.

Nr. (i)	$\bar{x}_i$	s_i^2	n_i	
			I	II
1	10	10	10	15
2	9	8	10	5
3	14	12	10	15
4	13	11	10	10
5	14	7	10	5

$$\sum n_I = \sum n_{II} = 50$$

Die Mittelwerte nach (3.24) berechnet sind: $\bar{x}_I = 12,0$; $\bar{x}_{II} = 12,1$

Nach (7.149) ergibt sich für den Fall gleicher (*I*) und ungleicher (*II*) Stichprobenumfänge:

$$\hat{F}_I = \frac{10[(10-12)^2 + (9-12)^2 + (14-12)^2 + (13-12)^2 + (14-12)^2]/(5-1)}{9 \cdot 48/(50-5)}$$

$$\hat{F}_I = \frac{55}{9,6} = 5,73$$

$$\hat{F}_{II} = \frac{[15(10-12,1)^2 + 5(9-12,1)^2 + 15(14-12,1)^2 + 10(13-12,1)^2 + 5(14-12,1)^2]/(5-1)}{(10 \cdot 14 + 8 \cdot 4 + 12 \cdot 14 + 11 \cdot 9 + 7 \cdot 4)/(50-5)}$$

$$\hat{F}_{II} = \frac{48,75}{10,38} = 4,69$$

Da $5,73$ und $4,69 > 3,77 = F_{(4;45;0,99)}$, prüfen wir $\mu_1 = \mu_2 < \mu_3 = \mu_4 = \mu_5$ nach (7.155) bzw. (7.156) und bilden

für I

$$|\bar{x}_A - \bar{x}_B| = \frac{1}{2}(\bar{x}_1 + \bar{x}_2) - \frac{1}{3}(\bar{x}_3 + \bar{x}_4 + \bar{x}_5) = \frac{1}{2}(10 + 9) - \frac{1}{3}(14 + 13 + 14) = 4,17$$

$$\sqrt{s_{in}^2 \sum_{i=1}^5 c_i^2 \left(\frac{1}{n_i}\right)} = \sqrt{9,6 \left[\frac{1}{2^2} \left(\frac{1}{10} + \frac{1}{10}\right) + \frac{1}{3^2} \left(\frac{1}{10} + \frac{1}{10} + \frac{1}{10}\right) \right]} = \sqrt{0,8} = 0,894$$

für II

$$|\bar{x}_A - \bar{x}_B| = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2} - \frac{n_3 \bar{x}_3 + n_4 \bar{x}_4 + n_5 \bar{x}_5}{n_3 + n_4 + n_5}$$

$$|\bar{x}_A - \bar{x}_B| = \frac{15 \cdot 10 + 5 \cdot 9}{15 + 5} - \frac{15 \cdot 14 + 10 \cdot 13 + 5 \cdot 14}{15 + 10 + 5} = 3,92$$

und

$$\sqrt{s_{in}^2 \sum_{i=1}^5 c_i^2 \left(\frac{1}{n_i}\right)}$$

$$= \sqrt{10,38 \left[\left\{ \left(\frac{3}{4}\right)^2 \cdot \frac{1}{15} + \left(\frac{1}{4}\right)^2 \cdot \frac{1}{5} \right\} + \left\{ \left(\frac{3}{6}\right)^2 \cdot \frac{1}{15} + \left(\frac{2}{6}\right)^2 \cdot \frac{1}{10} + \left(\frac{1}{6}\right)^2 \cdot \frac{1}{5} \right\} \right]} = 0,930$$

vgl. $\frac{3}{4} = n_1/(n_1 + n_2) = 15/(15 + 5)$

und erhalten

für I

$$\frac{4,17}{0,894} = 4,66$$

für II

$$\frac{3,92}{0,930} = 4,21$$

mit $F_{(4;45;0,99)} = 3,77$ und $\sqrt{(5-1)3,77} = 3,88$ nach (7.157) in beiden Fällen ($\mathbf{I} : \hat{S}_I = 4,66 > 3,88 = S$; $\mathbf{II} : \hat{S}_{II} = 4,21 > 3,88 = S$) statistisch signifikante Unterschiede ($P = 0,01$).



Beispiel: Die Definition und Berechnung von Kontrasten in **R** wird im Kapitel Modellbildung [8] ausführlicher dargestellt. An dieser Stelle soll ein Hinweis auf die Berechnung des Standardfehlers von Kontrasten nach (7.156) mit der Funktion **se.contrast()** genügen. Eine Funktion zur erweiterten Analyse von Kontrasten ist auch mit der Funktion **fit.contrast()** aus **library(gmodels)** [War05] aus dem Paket „**gregmisc**“ möglich. In Dem Beispiel sollen drei Stichproben (x, y und z) miteinander verglichen werden.

```
> x <- c( 4,  8, 11, 14, 10,  9, 11,  6); mean(x)
[1] 9.125
> y <- c(17, 10, 11, 13, 14,  9, 11, 12, 12,  8); mean(y)
[1] 11.7
> z <- c(12, 16, 11, 12, 17, 22, 12, 16, 17, 13, 19, 12); mean(z)
[1] 14.91667
>
> grp <- c(rep(1,8), rep(2,10), rep(3,12))
> wert <- c(x, y, z)
> daten <- data.frame(grp=factor(grp), wert)
> aov.mod <- aov(wert ~ grp, daten); summary(aov.mod)
              Df Sum Sq Mean Sq F value    Pr(>F)
grp             2  166.408   83.204    8.644 0.001255 **
Residuals      27  259.892    9.626

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

>
> se.contrast(aov.mod, list(grp=="1", grp=="2", grp=="3"), coef=c(-1, 0, 1))
[1] 1.416099
>
> library(gmodels)
> fit.contrast(aov.mod, grp, c(-1, 0, 1))
              Estimate Std. Error t value    Pr(>|t|)
grp c=( -1 0 1 )  5.791667    1.416099  4.089874 0.0003487793
```

Die Mittelwerte sind: $\bar{x} = 9,1$, $\bar{y} = 11,7$ und $\bar{z} = 14,9$. Die Varianzanalyse nach (7.149) führt auf einen signifikanten Unterschied der Erwartungswerte ($\hat{F} = 8,64$ und $P = 0,001$). Der Vergleich von x und z (Gruppe 1 mit Gruppe 3) wird durch den Kontrastvektor $(-1, 0, 1)$ definiert. Die entsprechende Differenz 5,79 mit dem Standardfehler 1,42 weist auf einen hochsignifikanten Unterschied ($P < 0,001$) hin (auch nach (7.157) ergibt sich $\hat{S} = 4,09 > 2,59 = \sqrt{2 \cdot 3,354} = \sqrt{(k-1)F_{k-1;n-k;0,95}}$).

Hinweis zur Maximalzahl linearer Kontraste

Bei drei Mittelwerten gibt es bereits

$$\begin{array}{ll} \bar{x}_1 - (\bar{x}_2 + \bar{x}_3)/2 & (\bar{x}_1 + \bar{x}_2)/2 - \bar{x}_3 \\ \bar{x}_2 - (\bar{x}_1 + \bar{x}_3)/2 & (\bar{x}_1 + \bar{x}_3)/2 - \bar{x}_2 \\ \bar{x}_3 - (\bar{x}_1 + \bar{x}_2)/2 & (\bar{x}_2 + \bar{x}_3)/2 - \bar{x}_1 \end{array}$$

6 Kontraste, bei 15 Mittelwerten sind es bereits über 7 Millionen. Eine Übersicht wird in Tabelle 7.43 gegeben.

Hinweis: Bildung homogener Gruppen von Mittelwerten anhand des nach Hayter modifizierten LSD-Tests, eines Lücken-Tests für geordnete μ_i

Wenn der F -Test $H_0 (\mu_i = \mu)$ abzulehnen gestattet, ordnet man die k Mittelwerte aus Stichprobengruppen gleichen Umfangs ($n_i = \text{konst.}$, $n = \sum_i n_i$) der Größe nach absteigend ($\bar{x}_{(1)} \geq \bar{x}_{(2)} \geq \bar{x}_{(3)} \geq \dots$) und prüft, ob benachbarte Mittelwerte eine größere Differenz Δ (Delta) aufweisen als die kleinste signifikante Differenz (least significant difference, LSD , besser LSD_H nach Hayter [Hay86]):

Tabelle 7.43. Zahl maximaler Kontraste für 3 bis 15 Mittelwerte

Mittelwerte	Maximalzahl der Kontraste
3	6
4	25
5	90
6	301
7	966
8	3025
9	9330
10	28501
11	86526
12	261625
13	788970
14	2375101
15	7141686

$$LSD_H = q_{\nu;k;\alpha} \sqrt{s_{in}^2/n_i} \quad (7.158)$$

$n_i = \text{konstant}$; s_{in}^2 basiert auf ν Freiheitsgraden; $q_{\nu;k;\alpha}$ ist der Tabelle 7.44 mit $\nu = \nu_2$ zu entnehmen

Für ungleiche Stichprobenumfänge ($n_i \neq \text{konst.}$) ergibt sich

$$LSD_{H(a,b)} = \frac{q_{\nu;k;\alpha}}{\sqrt{2}} \sqrt{s_{in}^2 \left[\frac{1}{n_a} + \frac{1}{n_b} \right]} \quad (7.159)$$

Für $\Delta \leq LSD_H$ bzw. $\Delta_{(a,b)} \leq LSD_{H(a,b)}$ lässt sich H_0 (Gleichheit benachbarter Erwartungswerte) nicht ablehnen; man unterstreicht die Mittelwerte durch eine gemeinsame Linie.

Beispiel: Vergleich von $k = 6$ Stichprobengruppen:

$\bar{x}_i$	Δ
$\bar{x}_1 = 26,8$	0,5
$\bar{x}_2 = 26,3$	1,1
$\bar{x}_3 = 25,2$	5,4
$\bar{x}_4 = 19,8$	5,5
$\bar{x}_5 = 14,3$	2,5
$\bar{x}_6 = 11,8$	



$$n_i = 8; k = 6; s_{in}^2 = 10,38; \nu = 48 - 6 = 42$$

$$q_{42;6;0,05} = 4,22$$

$$LSD_H = 4,22 \sqrt{10,38/8} = 4,81 \text{ bzw.}$$

$$LSD_{H(a,b)} = 4,22 \sqrt{10,38 \cdot 0,5 \left[\frac{1}{8} + \frac{1}{8} \right]} = 4,81$$

Auf dem 5%-Niveau lassen sich drei Bereiche erkennen: $\bar{x}_{(1)}\bar{x}_{(2)}\bar{x}_{(3)}$ $\bar{x}_{(4)}$ $\bar{x}_{(5)}\bar{x}_{(6)}$

[Anwendung von (7.159): $n_4 = 6$; $n_5 = 10$; sonst unverändert;

$$LSD_{H(4,5)} = \frac{4,22}{\sqrt{2}} \sqrt{10,38 \left[\frac{1}{6} + \frac{1}{10} \right]} = 4,96 ;$$

dieser Wert ist zwar größer als 4,81, am Resultat ändert sich nichts.

7.5.3.4 Multiple Vergleiche nach Tukey-Kramer

Der Scheffé-Test für den (nicht geplanten) Vergleich von Erwartungswerten aus k -Normalverteilungen mit gleicher Varianz ist ein recht konservatives Verfahren, das ein festes multiples Signifikanzniveau α einhält, unabhängig davon, wie viele einfache Paarhypothesen oder aber auch komplexe Hypothesen aus linearen Kontrasten zu prüfen sind. Das Tukey-Kramer-Verfahren hat gegenüber der Scheffé-Prozedur eine höhere Power, wenn ohnehin nur die einfachen paarweisen Hypothesen (all pairwise) zu prüfen sind.

Im Gegensatz zum Scheffé-Test darf der Tukey-Test auch dann zum Vergleich von Erwartungswerten benutzt werden, wenn das $\hat{F}$ der Varianzanalyse (Globaltest) kleiner als der zugehörige tabellierte F -Wert ist, wenn also $H_0: \mu_i = \mu$ auf dem verwendeten Signifikanzniveau nicht abgelehnt werden kann (Ramsay [Ram81]). Der LSD_H -Test setzt Signifikanz des Globaltests voraus.

Das Tukey-Kramer-Verfahren (Tukey's HSD - honestly significant differences) basiert auf der **Verteilung der Studentisierten Extremwerte** (SR-Verteilung). Für multiple Vergleiche (mehr als zwei Gruppen) sind die Quantile der SR-Verteilung größer als die Quantile der t -Verteilung, die bei einem multiple t -Test mit anschließender Korrektur verwendet würden. Die Quantile der SR-Verteilung liegen abhängig von der Anzahl der Freiheitsgrade und der Zahl der zu vergleichenden Gruppe tabelliert vor oder sie werden mit der Funktion **qtukey()** in **R** berechnet.

Die Teststatistik für den Vergleich nach Tukey-Kramer ist in (7.160) für gleiche und ungleiche Anzahl von Beobachtungen in den Gruppen angegeben.

$$\begin{aligned} T_{ij} &= \frac{\bar{x}_i - \bar{x}_j}{s \cdot \sqrt{\frac{1}{n}}} & n = n_i = n_j \\ T_{ij} &= \frac{\bar{x}_i - \bar{x}_j}{s \cdot \sqrt{0,5 \cdot \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}} & n_i \neq n_j \\ \text{mit } s^2 &= \frac{1}{n-k} \sum_{i=1}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 \end{aligned} \quad (7.160)$$

Die Nullhypothese für den Vergleich zweier Gruppen $H_{0ij}: \mu_i = \mu_j$ ist im multiplen paarweisen Vergleich abzulehnen, wenn der Wert der Teststatistik $|T_{ij}|$ aus (7.160) größer ist als das Quantil der SR-Verteilung $q_{\nu,k,1-\alpha}$ (vgl. Tabelle 7.44) mit $\nu = \sum n_i - k$ Freiheitsgraden.

Konfidenzintervalle für die Differenzen aus allen Paaren $\mu_i - \mu_j$ lassen sich entsprechend nach (7.161) bestimmen.

$$\bar{x}_i - \bar{x}_j \pm q_{\nu,k,1-\alpha} \cdot s \cdot \sqrt{0,5 \left(\frac{1}{n_i} + \frac{1}{n_j} \right)} \quad (7.161)$$

Tabelle 7.44. Obere Signifikanzschranken des Studentisierten Extrembereiches (SR-Verteilung) $q_{\nu,k,0.95}$ ($P = 0,05$), berechnet mit der Funktion `qtukey()` in R

ν	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$	$k=8$	$k=9$	$k=10$	$k=11$	$k=12$
2	6,08	8,33	9,80	10,88	11,73	12,43	13,03	13,54	13,99	14,40	14,76
3	4,50	5,91	6,82	7,50	8,04	8,48	8,85	9,18	9,46	9,72	9,95
4	3,93	5,04	5,76	6,29	6,71	7,05	7,35	7,60	7,83	8,03	8,21
5	3,64	4,60	5,22	5,67	6,03	6,33	6,58	6,80	6,99	7,17	7,32
6	3,46	4,34	4,90	5,30	5,63	5,90	6,12	6,32	6,49	6,65	6,79
7	3,34	4,16	4,68	5,06	5,36	5,61	5,82	6,00	6,16	6,30	6,43
8	3,26	4,04	4,53	4,89	5,17	5,40	5,60	5,77	5,92	6,05	6,18
9	3,20	3,95	4,41	4,76	5,02	5,24	5,43	5,59	5,74	5,87	5,98
10	3,15	3,88	4,33	4,65	4,91	5,12	5,30	5,46	5,60	5,72	5,83
11	3,11	3,82	4,26	4,57	4,82	5,03	5,20	5,35	5,49	5,61	5,71
12	3,08	3,77	4,20	4,51	4,75	4,95	5,12	5,27	5,39	5,51	5,61
13	3,06	3,73	4,15	4,45	4,69	4,88	5,05	5,19	5,32	5,43	5,53
14	3,03	3,70	4,11	4,41	4,64	4,83	4,99	5,13	5,25	5,36	5,46
15	3,01	3,67	4,08	4,37	4,59	4,78	4,94	5,08	5,20	5,31	5,40
16	3,00	3,65	4,05	4,33	4,56	4,74	4,90	5,03	5,15	5,26	5,35
17	2,98	3,63	4,02	4,30	4,52	4,70	4,86	4,99	5,11	5,21	5,31
18	2,97	3,61	4,00	4,28	4,49	4,67	4,82	4,96	5,07	5,17	5,27
19	2,96	3,59	3,98	4,25	4,47	4,65	4,79	4,92	5,04	5,14	5,23
20	2,95	3,58	3,96	4,23	4,45	4,62	4,77	4,90	5,01	5,11	5,20
21	2,94	3,56	3,94	4,21	4,42	4,60	4,74	4,87	4,98	5,08	5,17
22	2,93	3,55	3,93	4,20	4,41	4,58	4,72	4,85	4,96	5,06	5,14
23	2,93	3,54	3,91	4,18	4,39	4,56	4,70	4,83	4,94	5,03	5,12
24	2,92	3,53	3,90	4,17	4,37	4,54	4,68	4,81	4,92	5,01	5,10
25	2,91	3,52	3,89	4,15	4,36	4,53	4,67	4,79	4,90	4,99	5,08
26	2,91	3,51	3,88	4,14	4,35	4,51	4,65	4,77	4,88	4,98	5,06
27	2,90	3,51	3,87	4,13	4,33	4,50	4,64	4,76	4,86	4,96	5,04
28	2,90	3,50	3,86	4,12	4,32	4,49	4,62	4,74	4,85	4,94	5,03
29	2,89	3,49	3,85	4,11	4,31	4,47	4,61	4,73	4,84	4,93	5,01
30	2,89	3,49	3,85	4,10	4,30	4,46	4,60	4,72	4,82	4,92	5,00
31	2,88	3,48	3,84	4,09	4,29	4,45	4,59	4,71	4,81	4,90	4,99
32	2,88	3,48	3,83	4,09	4,28	4,45	4,58	4,70	4,80	4,89	4,98
33	2,88	3,47	3,83	4,08	4,28	4,44	4,57	4,69	4,79	4,88	4,97
34	2,87	3,47	3,82	4,07	4,27	4,43	4,56	4,68	4,78	4,87	4,96
35	2,87	3,46	3,81	4,07	4,26	4,42	4,56	4,67	4,77	4,86	4,95
36	2,87	3,46	3,81	4,06	4,25	4,41	4,55	4,66	4,76	4,85	4,94
37	2,87	3,45	3,80	4,05	4,25	4,41	4,54	4,66	4,76	4,85	4,93
38	2,86	3,45	3,80	4,05	4,24	4,40	4,53	4,65	4,75	4,84	4,92
39	2,86	3,45	3,79	4,04	4,24	4,39	4,53	4,64	4,74	4,83	4,91
40	2,86	3,44	3,79	4,04	4,23	4,39	4,52	4,63	4,73	4,82	4,90
50	2,84	3,42	3,76	4,00	4,19	4,34	4,47	4,58	4,68	4,77	4,85
60	2,83	3,40	3,74	3,98	4,16	4,31	4,44	4,55	4,65	4,73	4,81
70	2,82	3,39	3,72	3,96	4,14	4,29	4,42	4,53	4,62	4,71	4,78
80	2,81	3,38	3,71	3,95	4,13	4,28	4,40	4,51	4,60	4,69	4,76
90	2,81	3,37	3,70	3,94	4,12	4,27	4,39	4,50	4,59	4,67	4,75
100	2,81	3,36	3,70	3,93	4,11	4,26	4,38	4,48	4,58	4,66	4,73
∞	2,77	3,31	3,63	3,86	4,03	4,17	4,29	4,39	4,48	4,55	4,62



Beispiel: Die Wirksamkeit von 3 Antibiotika (A, B, C) wird durch die Größe der Hemmzone (mm Durchmesser) gemessen. Das Ergebnis einer Untersuchung mit jeweils 5 Platten (Agarplatte mit *Bacillus subtilis* beimpft) ist in der folgenden Tabelle zusammengestellt.

Antibiotikum	1	2	3	4	5	$\bar{x}_i$	s_i
A	27	27	25	26	25	26,0	1,0
B	26	25	26	25	24	25,2	0,8
C	21	21	20	20	22	20,8	0,8

Die Berechnung der Teststatistik für den multiplen Vergleich nach dem Tukey-Verfahren erfolgt für das Beispiel elementar nach (7.160).

```
> A <- c(27, 27, 25, 26, 25)
> B <- c(26, 25, 26, 25, 24)
> C <- c(21, 21, 20, 20, 22)
>
> nA <- length(A); nB <- length(B); nC <- length(C)
> f <- nA + nB + nC - 3
> mA <- mean(A); mB <- mean(B); mC <- mean(C)
> s <- sqrt((sum((A-mA)^2)+sum((B-mB)^2)+sum((C-mC)^2)) / f)
>
> T.AB <- (mA - mB) / (s*sqrt(0.5*(1/nA + 1/nB))); T.AB
[1] 2
> T.AC <- (mA - mC) / (s*sqrt(0.5*(1/nA + 1/nC))); T.AC
[1] 13
> T.BC <- (mB - mC) / (s*sqrt(0.5*(1/nB + 1/nC))); T.BC
[1] 11
>
> q <- qtukey(0.95, 3, f); q
[1] 3.772929
```

Die Werte der Teststatistik für die Vergleiche A vs C und B vs C (13 bzw. 11) sind größer als das Quantil der SR-Verteilung mit $k = 3$ und $\nu = 12$ Freiheitsgraden $q_{\nu, k, 0,95} = 3,77$. Zwischen diesen Antibiotika bestehen somit signifikante Unterschiede während im Vergleich A vs B (2) im multiplen paarweisen Vergleich kein signifikanter Unterschied zu erkennen ist.

Die Funktionen **simtest()** und **simint()** aus dem Paket **multcomp** [BHW04] ermöglichen in **R** eine einfache Berechnung dieser Statistiken.

```
> grp <- c(rep("A", nA), rep("B", nB), rep("C", nC))
> d <- data.frame(Gruppe = grp, Wert = c(A, B, C))
>
> simint(Wert ~ Gruppe, data=d, type="Tukey", alternative="two.sided")
```

Simultaneous confidence intervals: Tukey **contrasts**

Call:
simint.formula(formula = Wert ~ Gruppe, data = d, type = "Tukey",
alternative = "two.sided")

95 % confidence intervals

	Estimate	2.5 %	97.5 %
GruppeB-GruppeA	-0.8	-2.309	0.709
GruppeC-GruppeA	-5.2	-6.709	-3.691
GruppeC-GruppeB	-4.4	-5.909	-2.891

Für das Beispiel werden die 95%-Konfidenzintervalle für den multiplen paarweisen Vergleich mit der Funktion **simint()** bestimmt. Die Daten werden dazu zweckmäßigerweise in einem Datenrahmen abgespeichert und die Zugehörigkeit zu der Gruppe durch den Faktor 'Gruppe' gekennzeichnet. Das Ergebnis zeigt auch hier, dass zwischen A und B kein Unterschied besteht, während C signifikant kleinere Hemmzonen gegenüber A und B aufweist.

Tabelle 7.45. Obere Schranken $[q'_{\nu,k,1-\alpha}]$ der Verteilung des „Studentized Augmented Range“ (SAR-Verteilung) mit dem Parameter k und dem Freiheitsgrad ν . Aus Stoline, M. R. (1978): Tables of the Studentized Augmented Range and applications to problems of multiple comparisons. Journal of the American Statistical Association 73, 656-660, Tables 1-4, pp. 658 and 659; mit freundlicher Erlaubnis der ASA und des Autors

$\alpha = 0,01$							
ν	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$	$k = 7$	$k = 8$
5	5,903	7,03	7,823	8,429	8,916	9,322	9,669
7	5,063	5,947	6,551	7,008	7,374	7,679	7,939
10	4,550	5,284	5,773	6,138	6,428	6,669	6,875
12	4,373	5,056	5,505	5,837	6,101	6,321	6,507
16	4,169	4,792	5,194	5,489	5,722	5,915	6,079
20	4,055	4,644	5,019	5,294	5,510	5,688	5,839
24	3,982	4,549	4,908	5,169	5,374	5,542	5,685
30	3,912	4,458	4,800	5,048	5,242	5,401	5,536
40	3,844	4,370	4,696	4,931	5,115	5,265	5,392
60	3,778	4,284	4,595	4,818	4,991	5,133	5,253
120	3,714	4,201	4,497	4,709	4,872	5,005	5,118
∞	3,653	4,121	4,403	4,603	4,757	4,882	4,987
$\alpha = 0,05$							
ν	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$	$k = 7$	$k = 8$
5	3,832	4,654	5,236	5,680	6,036	6,331	6,583
7	3,486	4,198	4,692	5,064	5,360	5,606	5,816
10	3,259	3,899	4,333	4,656	4,913	5,124	5,305
12	3,177	3,791	4,204	4,509	4,751	4,950	5,119
16	3,080	3,663	4,050	4,334	4,557	4,741	4,897
20	3,024	3,590	3,961	4,233	4,446	4,620	4,768
24	2,988	3,542	3,904	4,167	4,373	4,541	4,684
30	2,952	3,496	3,847	4,103	4,302	4,464	4,602
40	2,918	3,450	3,792	4,040	4,232	4,389	4,521
60	2,884	3,406	3,738	3,978	4,163	4,314	4,441
120	2,851	3,362	3,686	3,917	4,096	4,241	4,363
∞	2,819	3,320	3,634	3,858	4,030	4,170	4,286

Das Games-Howell-Verfahren (Varianzheterogenität zugelassen).

Besteht bei paarweisen multiplen Vergleichen von Erwartungswerten **Unklarheit darüber, ob die Varianzen der Grundgesamtheiten gleich sind**, so muss das Games-Howell-Verfahren oder das entsprechende Rangtestverfahren (H-Test) angewandt werden.

$$\bar{x}_i - \bar{x}_j \pm \frac{q'_{\nu,k,1-\alpha}}{\sqrt{2}} \cdot \sqrt{\frac{s_i^2}{n_i} + \frac{s_j^2}{n_j}} \quad (1 \leq i < j \leq k)$$

$$\nu = \frac{[s_i^2/n_i + s_j^2/n_j]^2}{s_i^4/[n_i^2(n_i - 1)] + s_j^4/[n_j^2(n_j - 1)]}$$

zur ganzen Zahl gerundet

(7.162)

Während das Tukey-Kramer-Verfahren die Schranken der SR-Verteilung (vgl. Tabelle 7.44) benötigt, werden im Games-Howell-Verfahren die Schranken der sogenannten SAR-Verteilung („**Studentized Augmented Range**“) verwendet (Stoline [Sto78]). Diese liegen über den Schranken der

SR-Verteilung und sind für $\alpha = 0,01$ und $\alpha = 0,05$ auszugsweise in Tabelle 7.45 (aus Sachs [Sac90]) angegeben.



Beispiel: Gegeben seien $\bar{x}_1 = 9,43$, $\bar{x}_2 = 6,59$ und $\bar{x}_3 = 4,06$, d.h. $k = 3$, $n_1 = n_2 = n_3 = 10$; $s_1^2 = 1,03$, $s_2^2 = 8,12$, $s_3^2 = 3,95$. Als Beispiel soll die Prüfung von $H_0 : \mu_1 = \mu_2$ gegen $H_A : \mu_1 > \mu_2$ dienen.

$$\nu = \frac{[1,03/10 + 8,12/10]^2}{1,03^2/[10^2(10-1)] + 8,12^2/[10^2(10-1)]} = 11,25 \text{ d.h. } \nu = 11$$

Aus Tabelle 7.45 folgt dann $q'_{11;3;0,05} = 3,84$ (interpoliert)

$$\sqrt{\frac{1,03}{10} + \frac{8,12}{10}} = 0,9566$$

$$3,84 \cdot 0,9566/\sqrt{2} = 2,597$$

$$9,43 - 6,59 = 2,84 > 2,597$$

Damit lässt sich auf dem 5%-Niveau die Nullhypothese ablehnen. Auf gleiche Weise lassen sich auch die anderen Nullhypothesen prüfen: einmal $H_0 : \mu_1 = \mu_3$ und zum anderen $H_0 : \mu_2 = \mu_3$.

7.5.3.5 Multiple Vergleiche nach Dunnett

Sind die Erwartungswerte von k Normalverteilungen gegen den Erwartungswert μ_0 einer Referenz (Kontrolle) zu vergleichen, wobei das multiple Signifikanzniveau α eingehalten werden soll (comparisons to control, many to one), dann ist das Verfahren von Dunnett zu verwenden ([Dun55], [Dun64]). Die Teststatistik nach Dunnett D_i (7.163) folgt dem Modell nach einer k -variaten t-Verteilung mit $\nu = \sum_{i=1}^k n_i - (k+1)$ Freiheitsgraden. Zusätzlich wird diese Verteilung durch die Korrelation r_{ij} zwischen den Gruppen bestimmt.

$$D_i = \frac{\bar{x}_i - \bar{x}_0}{s \cdot \sqrt{\frac{1}{n_i} + \frac{1}{n_0}}}$$

$$\text{mit } s^2 = \frac{\sum_{i=0}^k \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2}{\sum_{i=0}^k n_i - (k+1)}$$

$$R = r_{ij} = \sqrt{\frac{n_i}{n_0 + n_i}} \sqrt{\frac{n_j}{n_0 + n_j}} \quad (7.163)$$

Die einzelnen Hypothesen $H_{i0} : \mu_i = \mu_0$ sind abzulehnen, sobald der Wert der Teststatistik $|D_i|$ größer ist als das entsprechende Quantil der multivariaten t-Verteilung $t_{\nu,k,R,1-\alpha}$. Auf eine Tabelle zu diesen Quantilen wird verzichtet, da deren Tabellierung wegen des zusätzlichen Parameters R sehr aufwendig ist (Horn und Vollandt [HV95]). Die wichtigsten Quantile sind auch Sachs [Sac90] auf den Seiten 200-223 zu entnehmen. In **R** besteht mit der Funktion **qmvmt()** in dem Paket **mvtnorm** [GBH04] die Möglichkeit, Quantile zur multivariaten t-Verteilung direkt zu berechnen (vgl. folgendes Beispiel).

Simultane zweiseitige Konfidenzintervalle für die Differenzen der entsprechenden Erwartungswerte lassen sich nach (7.164) angeben.

$$\bar{x}_i - \bar{x}_0 \pm t_{\nu, k, R, 1-\alpha} s \cdot \sqrt{\frac{1}{n_i} + \frac{1}{n_0}} \quad (7.164)$$

Beispiel: Die folgenden Beispieldaten [Dun55] geben Messungen von Blutzellen ($10^6/mm^3$) in 3 Tiergruppen wieder. Die erste Gruppe ist eine unbehandelte Kontrollgruppe, während die beiden anderen Gruppen unterschiedliche (aktive) Medikamente erhielten. Versuchsbedingte Ausfälle führten zu unterschiedlichen Fallzahlen in den Gruppen.



Gruppe	1	2	3	4	5	6
Kontrolle	7,40	8,50	7,20	8,24	9,84	8,32
Präp. A	9,76	8,80	7,68	9,36		
Präp. B	12,80	9,68	12,16	9,20	10,55	

```
> Kontrolle <- c(7.40, 8.50, 7.20, 8.24, 9.84, 8.32)
> Praep.A <- c(9.76, 8.80, 7.68, 9.36)
> Praep.B <- c(12.80, 9.68, 12.16, 9.20, 10.55)
>
> n0 <- length(Kontrolle); nA <- length(Praep.A); nB <- length(Praep.B)
> f <- n0+nA+nB-(3+1)
> m0 <- mean(Kontrolle); mA <- mean(Praep.A); mB <- mean(Praep.B)
> s <- sqrt((sum((Kontrolle-m0)^2)+sum((Praep.A-mA)^2)+sum((Praep.B-mB)^2)) / f)
>
> D.A <- (mA - m0) / (s*sqrt(1/nA + 1/n0)); D.A
[1] 0.8205458
> D.B <- (mB - m0) / (s*sqrt(1/nB + 1/n0)); D.B
[1] 3.536499
>
> R <- sqrt(nA/(n0+nA)) * sqrt(nB/(n0+nB))
> cR <- matrix(c(1, R, R, 1), nrow=2); round(cR,2)
      [,1] [,2]
[1,] 1.00 0.43
[2,] 0.43 1.00
>
> library(mvtnorm)
> qmvt(0.95, tail="both.tail", df=f, corr=cR)$quantile
[1] 2.543489
```

Die Werte für die Teststatistik sind $D_A = 0,82$ und $D_B = 3,54$. Das Quantil der multivariaten t-Verteilung mit $\nu = 11$ Freiheitsgraden und dem Korrelationskoeffizienten $r_{AB} = 0,43$ ist $t_{11;2;R;0,955} = 2,54$. Damit zeigt nur die Therapiegruppe B auf dem 5%-Niveau signifikant veränderte Werte gegenüber der Kontrollgruppe für die Blutzellen an.

Eine einfache Lösung in R ist auch mit der Funktion `simtest()` aus dem Paket **multcomp** [BHW04] möglich. Hierzu werden die Daten mit einer Kennzeichnung der Gruppenzugehörigkeit in einen eigenen Rahmen kopiert. Die Funktion berechnet adjustierte P-Werte für die beiden Kontraste gegen die Kontrollgruppe.

```
> grp <- c(rep("Kontr", n0), rep("A", nA), rep("B", nB))
> d <- data.frame(Gruppe = grp, Wert = c(Kontrolle, Praep.A, Praep.B))
> simtest(Wert ~ Gruppe, data=d, type="Dunnett", base=3, alternative="greater")
```

Simultaneous tests: Dunnett **contrasts**

Contrast **matrix**:

	GruppeA	GruppeB	GruppeKontr
GruppeA-GruppeKontr	0	1	0
GruppeB-GruppeKontr	0	0	1

Adjusted P-Values

	p adj
GruppeB-GruppeKontr	0.003
GruppeA-GruppeKontr	0.204

Das Ergebnis stimmt mit dem zuvor elementar abgeleiteten Resultat überein ($P_{0A} = 0,204$ und $P_{0B} = 0,003$). Eine übersichtliche und informativere Bewertung der Ergebnisse liefert allerdings die Berechnung zweiseitiger Konfidenzintervalle mit der Funktion **simint()** aus demselben Paket.

```
> simint(Wert ~ Gruppe, data=d, type="Dunnnett", base=3, alternative="two.sided")
```

Simultaneous confidence intervals: Dunnnett contrasts

95 % confidence intervals

	Estimate	2.5 %	97.5 %
GruppeA-GruppeKontr	0.650	-1.256	2.556
GruppeB-GruppeKontr	2.628	0.840	4.416

Für den Vergleich Kontrolle-A ist die Differenz 0,65 (95%-KI -1,26 bis 2,56; d.h. nicht signifikant) und für den Vergleich Kontrolle-B ist diese Differenz 2,63 (95%-KI 0,84 bis 4,42; d.h. deutlicher Effekt). Zusätzlich kann damit auch eine Aussage zur Größe des Effektes einer Behandlung gegen die Kontrolle gemacht werden.

7.5.4 H-Test von Kruskal und Wallis

Der H -Test von Kruskal und Wallis [Kru52] ist eine Verallgemeinerung des U -Tests. Er prüft die Nullhypothese, die k Stichproben entstammen derselben Grundgesamtheit: die k Verteilungsfunktionen sind gleich (H_A : mindestens zwei sind ungleich).

Ähnlich wie der U -Test hat auch der H -Test, verglichen mit der bei Normalverteilung optimalen Varianzanalyse eine asymptotische Effizienz von $100 \cdot 3/\pi \simeq 95\%$.

Die $n = \sum_{i=1}^k n_i$ Beobachtungen, Zufallsstichproben von Rangordnungen oder Messwerten mit den Umfängen $n_1, n_2, \dots, n_k$ aus umfangreichen Grundgesamtheiten, werden der Größe nach aufsteigend geordnet und mit Rängen von 1 bis n versehen (wie beim U -Test). R_i sei die Summe der Ränge der i -ten Stichprobe: Unter der Nullhypothese ist die Prüfgröße

$$\hat{H} = \left[\frac{12}{n(n+1)} \right] \cdot \left[\sum_{i=1}^k \frac{R_i^2}{n_i} \right] - 3(n+1) \quad (7.165)$$

($\hat{H}$ ist die Varianz der Stichproben-Rangsummen R_i) für großes n (d.h. praktisch für $n_i \geq 5$ und $k \geq 4$) χ^2 -verteilt mit $k-1$ Freiheitsgraden; d.h. H_0 wird abgelehnt, sobald $\hat{H} > \chi_{k-1;\alpha}^2$ (vgl. Tab. 5.11). Für $n_i \leq 8$ und $k = 3$ enthält Tab. 7.47 (Kruskal [KW52] und Iman und Mitarbeiter [IQA75]) die exakten Überschreitungswahrscheinlichkeiten (H_0 wird mit P abgelehnt, wenn $\hat{H} \geq H$ mit $P \leq \alpha$). Für $k = 3$ und $n_1 = n_2 = n_3 = n$ sind die 5%-Schranken für $n = 6$ (7; 8) durch den Wert 5,8 [5,801; (5,819; 5,805)] gegeben. Für „große“ Stichprobenumfänge ($n_i \gtrapprox 25$) benutze man die Schranken der Tabelle 7.46.

Zur Kontrolle der R_i benutze man die Beziehung

$$\sum_{i=1}^k R_i = n(n+1)/2 \quad (7.166)$$

Sind die Stichproben gleich groß, ist also $n_i = \frac{n}{k}$, rechnet man bequemer nach der vereinfachten Formel:

$$\hat{H} = \left[\frac{12k}{n^2(n+1)} \right] \cdot \left[\sum_{i=1}^k R_i^2 \right] - 3(n+1) \quad (7.167)$$

Tabelle 7.46. Kritische Schranken $\chi^2_{k-1;\alpha}$ für den H -Test und den Friedman-Test für $k = 3(1)7$ und „große“ Stichprobenumfänge; H -Test: $n_i \gtrsim 25$; Friedman-Test: $n \gtrsim 25$

$k \backslash P$	0,10	0,05	0,01
3	4,605	5,991	9,210
4	6,251	7,815	11,345
5	7,779	9,488	13,277
6	9,236	11,070	15,086
7	10,645	12,592	16,812

Gehören mehr als 25% aller Werte zu **Bindungen**, d. h. zu Folgen gleicher Rangzahlen, dann muss $\hat{H}$ korrigiert werden. Die Korrekturformel für $\hat{H}$ lautet:

$$\hat{H}_{\text{kor}} = \frac{\hat{H}}{1 - \frac{\sum_{i=1}^{i=r} (t_i^3 - t_i)}{n^3 - n}} \quad (7.168)$$

wobei t_i die Anzahl der jeweils gleichen Rangplätze in der Bindung i bezeichnet. Da der korrigierte $\hat{H}$ -Wert größer als der nicht korrigierte Wert ist, braucht man bei einem statistisch signifikanten $\hat{H}$ -Wert $\hat{H}_{\text{kor}}$ nicht zu berechnen.

Beispiel: Prüfe die 4 Stichproben aus der folgenden Tabelle mit dem H -Test ($\alpha = 0,05$).



A		B		C		D	
12,1	10	18,3	15	12,7	11	7,3	3
14,8	12	49,6	21	25,1	16	1,9	1
15,3	13	10,1	6 1/2	47,0	20	5,8	2
11,4	9	35,6	19	16,3	14	10,1	6 1/2
10,8	8	26,2	17	30,4	18	9,4	5
		8,9	4				
R_i	52,0	82,5		79,0		17,5	
R_i^2	2704,00	6806,25		6241,00		306,25	
n_i	5	6		5		5	
R_i^2/n_i		540,800 + 1134,375 + 1248,200 + 61,250 = 2984,625 = $\sum_{i=1}^{k=4} \frac{R_i^2}{n_i}$					

Kontrolle zur Bestimmung der Rangzahlen:

$$52,0 + 82,5 + 79,0 + 17,5 = 231 = 21(21 + 1)/2$$

Berechnung der Teststatistik:

$$\hat{H} = \left[\frac{12}{21(21 + 1)} \right] \cdot [2984,625] - 3(21 + 1) = 11,523$$

Tabelle 7.47. Irrtumswahrscheinlichkeiten für den H -Test von Kruskal und Wallis [aus Kruskal, W.H. und W.A. Wallis: Use of ranks in one-criterion variance analysis, J. Amer. Statist. Ass. **47** (1952) 614–617, unter Berücksichtigung der Errata in J. Amer. Statist. Ass. **48** (1953) 910] sowie einiger Werte aus Iman u. Mitarb. (1975)

n_1	n_2	n_3	H	P	n_1	n_2	n_3	H	P	n_1	n_2	n_3	H	P	n_1	n_2	n_3	H	P
2	1	1	2,7000	0,500	4	3	2	6,4444	0,008	5	2	2	6,5333	0,008	5	4	4	5,6571	0,049
								6,3000	0,011				6,1333	0,013				6,5176	0,050
2	2	1	3,6000	0,200				5,4444	0,046				5,1600	0,034				4,6187	0,100
								5,4000	0,051				5,0400	0,056				4,5527	0,102
2	2	2	4,5714	0,067				4,5111	0,098				4,3733	0,090					
			3,7143	0,200				4,4444	0,102				4,2933	0,122	5	5	1	7,3091	0,009
3	1	1	3,2000	0,300														6,8364	0,011
					4	3	3	6,7455	0,010	5	3	1	6,4000	0,012				5,1273	0,046
3	2	1	4,2857	0,100				6,7091	0,013				4,9600	0,048				4,9091	0,053
			3,8571	0,133				5,7909	0,046				4,8711	0,052				4,1091	0,086
								5,7273	0,050				4,0178	0,095				4,0364	0,105
3	2	2	5,3572	0,029				4,7091	0,092				3,8400	0,123					
			4,7143	0,048				4,7000	0,101						5	5	2	7,3385	0,010
			4,5000	0,067						5	3	2	6,9091	0,009				7,2692	0,010
			4,4643	0,105	4	4	1	6,6667	0,010				6,8218	0,010				5,3385	0,047
								6,1667	0,022				5,2509	0,049				5,2462	0,051
3	3	1	5,1429	0,043				4,9667	0,048				5,1055	0,052				4,6231	0,097
			4,5714	0,100				4,8667	0,054				4,6509	0,091				4,5077	0,100
			4,0000	0,129				4,1667	0,082				4,4945	0,101					
								4,0667	0,102						5	5	3	7,5780	0,010
3	3	2	6,2500	0,011						5	3	3	7,0788	0,009				7,5429	0,010
			5,3611	0,032									6,9818	0,011				5,7055	0,046
			5,1389	0,061	4	4	2	7,0364	0,006				5,6485	0,049				5,6264	0,051
			4,5556	0,100				6,8727	0,011				5,5152	0,051				4,5451	0,100
			4,2500	0,121				5,4545	0,046				4,5333	0,097				4,5363	0,102
								5,2364	0,052				4,4121	0,109					
3	3	3	7,2000	0,004				4,5545	0,098						5	5	4	7,8229	0,010
			6,4889	0,011				4,4455	0,103	5	4	1	6,9545	0,008				7,7914	0,010
			5,6889	0,029									6,8400	0,011				5,6657	0,049
			5,6000	0,050	4	4	3	7,1439	0,010				4,9855	0,044				5,6429	0,050
			5,0667	0,086				7,1364	0,011				4,8600	0,056				4,5229	0,099
			4,6222	0,100				5,5985	0,049				3,9873	0,098				4,5200	0,101
								5,5758	0,051				3,9600	0,102					
4	1	1	3,5714	0,200				4,5455	0,099	5	4	2	7,2045	0,009	5	5	5	8,0000	0,009
								4,4773	0,102				7,1182	0,010				5,7800	0,049
4	2	1	4,8214	0,057	4	4	4	7,6538	0,008				5,2727	0,049				4,5600	0,100
			4,5000	0,076				7,5385	0,011				5,2682	0,050					
			4,0179	0,114				5,6923	0,049				4,5409	0,098					
4	2	2	6,0000	0,014				5,6538	0,054				4,5182	0,101	6	6	6	8,2222	0,010
			5,3333	0,033				4,6539	0,097									5,8011	0,049
			5,1250	0,052				4,5001	0,104	5	4	3	7,4449	0,010				4,6430	0,099
			4,4583	0,100									7,3949	0,011					
			4,1667	0,105	5	1	1	3,8571	0,143				5,6564	0,049					
													5,6308	0,050	7	7	7	8,378	0,010
4	3	1	5,8333	0,021	5	2	1	5,2500	0,036				4,5487	0,099				5,819	0,049
			5,2083	0,050				5,0000	0,048				4,5231	0,103				4,594	0,099
			5,0000	0,057				4,4500	0,071						8	8	8	8,465	0,010
			4,0556	0,093				4,2000	0,095									5,805	0,050
			3,8889	0,129				4,0500	0,119	5	4	4	7,7604	0,009				4,595	0,099
													7,7440	0,011					

Da $\hat{H} = 11,523 > 7,815 = \chi_{3;0,95}^2$ ist, ist nicht anzunehmen, dass die 4 Stichproben einer gemeinsamen Grundgesamtheit entstammen.

Für die Berechnung der Teststatistik nach dem Kruskal-Wallis-Test steht in **R** die Funktion **kruskal.test()** zur Verfügung.

```
> A <- c(12.1, 14.8, 15.3, 11.4, 10.8)
> B <- c(18.3, 49.6, 10.1, 35.6, 26.2, 8.9)
> C <- c(12.7, 25.1, 47.0, 16.3, 30.4)
> D <- c( 7.3,  1.9,  5.8, 10.1,  9.4)
>
> x <- c(A, B, C, D)
> g <- factor(rep(1:4, c(5, 6, 5, 5)), labels = c("A", "B", "C", "D"))
> kruskal.test(x, g)
```

Kruskal-Wallis rank sum test

Kruskal-Wallis chi-squared = 11.5302, df = 3, p-value = 0.009179

Die Daten aus den vier Stichproben werden in einem Vektor x zusammengelegt und die Zugehörigkeit zu den Stichproben in einer Faktorvariablen g definiert. Das Ergebnis stimmt mit dem aus der Tabelle hergeleiteten Resultat überein.

7.5.4.1 Multiple paarweise Vergleiche mittlerer Ränge

Ist $\hat{H}$ auf dem gewählten Niveau statistisch signifikant, so interessiert man sich dafür, welche Grundgesamtheiten differieren. Hierzu bildet man mittlere Ränge $\bar{R}_i$, indem die Rangsummen R_i durch ihre Stichprobenumfänge n_i dividiert werden: $\bar{R}_i = R_i/n_i$.

Die Nullhypothese, Gleichheit der erwarteten mittleren Ränge, wird auf dem 5%-Niveau abgelehnt, sobald für $n_i \geq 6$:

$$|\bar{R}_i - \bar{R}_{i'}| > \sqrt{d \cdot \chi_{k-1;0,95}^2 \cdot \left[\frac{n(n+1)}{12} \right] \left[\frac{1}{n_i} + \frac{1}{n_{i'}} \right]}$$

mit $d = 1$; treten viele Bindungen auf, dann ist

$$d = 1 - \frac{\sum_{i=1}^{i=r} (t_i^3 - t_i)}{n^3 - n}$$

(t_i die Anzahl der jeweils gleichen Rangplätze in der i -ten Bindung)

(7.169)

Die Differenz ist dann auf diesem Niveau statistisch signifikant von Null verschieden (χ^2 --Ansatz).

Sind k Stichproben zu vergleichen, so bedeutet dies :

$$\binom{k}{2} = \frac{k!}{(k-2)!2!} = \frac{k(k-1)(k-2)!}{(k-2)!2!} = k(k-1)/2$$

Einzelvergleiche jeweils zweier Stichproben. Die Nullhypothese lautet hier:

Die Stichproben i und i' entstammen derselben Grundgesamtheit.

für $i = 1, \dots, k-1$,

$i' = i+1, \dots, k$ mit $i < i'$

Beispiel: Zu den Daten aus der folgenden Tabelle sind multiple paarweise Vergleiche zwischen den Gruppen **A**, **B** und **C** zu berechnen.



A		B		C	
28	1	36	5	44	11
30	2	39	7	45	12 1/2
33	3	40	8	47	14
35	4	43	10	49	15
38	6	45	12 1/2	53	17
41	9	50	16	54	18
$\bar{x}$	34	41,5		48	
R_i	25,0	58,5		87,5	
R_i^2	625,0	3422,25		7656,25	
n_i	6	6		6	

Kontrolle: $25,0 + 58,5 + 87,5 = 171 = 18(18 + 1)/2$

Nach (7.167) ist

$$\hat{H} = \left[\frac{12 \cdot 3}{18^2(18 + 1)} \right] [625,0 + 3422,25 + 7656,25] - 3(18 + 1) = 11,44$$

Mit $\hat{H} = 11,44 > 5,801 = H$ für $n = 6$, $k = 3$ und $P = 0,05$ (aus Tabelle 7.47) wird H_0 auf dem 5%-Niveau abgelehnt. Wo liegen nun wohl die Unterschiede?

$$k = 3, \quad \text{d.h.} \quad \chi_{3-1;0,95}^2 = 5,99$$

$$\sqrt{1 \cdot 5,99 \left[\frac{18(18 + 1)}{12} \right] \left[\frac{1}{6} + \frac{1}{6} \right]} = 7,54$$

$$\text{AB: } \left| \frac{25}{6} - \frac{58,5}{6} \right| = \frac{33,5}{6} = 5,58 < 7,54$$

$$\text{AC: } \left| \frac{25}{6} - \frac{87,5}{6} \right| = \frac{62,5}{6} = 10,42 > 7,54$$

$$\text{BC: } \left| \frac{58,5}{6} - \frac{87,5}{6} \right| = \frac{29}{6} = 4,83 < 7,54$$

Damit lässt sich anhand der vorliegenden Beobachtungen nur zwischen A und C ein Unterschied auf dem 5%-Niveau feststellen.

In dem vorliegenden Beispiel liegen gleiche Stichprobenumfänge vor. Für

$$n_1 = n_2 = \dots n_k = n \geq 6$$

lässt sich ein multipler paarweiser Vergleich der Gruppen untereinander auch nach Harter [Har60] prüfen, wobei wir hier auch eine kleinere (oder größere) Irrtumswahrscheinlichkeit als $\alpha = 0,05$ wählen dürfen. Die Nullhypothese wird abgelehnt für

$$|\bar{R}_i - \bar{R}_{i'}| > q_{k;\alpha} \sqrt{k(kn+1)/12} \quad (7.170)$$

Der Wert $q_{k;\alpha}$ ist für k und α der Tabelle 7.48 zu entnehmen.

Tabelle 7.48. Einige ausgewählte Schranken nach Harter (1960) für paarweise Vergleiche von mittleren Rängen (H -Test) und Rangsummen (Friedman-Test) für hinreichend große Stichprobenumfänge (Schranken der Spannweite k unabhängiger standardnormalverteilter Zufallsvariablen)

$k \backslash \alpha$	0,10	0,05	0,01	0,001
2	2,326	2,772	3,643	4,654
3	2,902	3,314	4,120	5,063
4	3,240	3,633	4,403	5,309
5	3,478	3,858	4,603	5,484
6	3,661	4,030	4,757	5,619
7	3,808	4,170	4,882	5,730
8	3,931	4,286	4,987	5,823
9	4,037	4,387	5,078	5,903
10	4,129	4,474	5,157	5,973
11	4,211	4,552	5,227	6,036
12	4,285	4,622	5,290	6,092
13	4,351	4,685	5,348	6,144
14	4,412	4,743	5,400	6,191
15	4,468	4,796	5,448	6,234
16	4,519	4,845	5,493	6,274
17	4,568	4,891	5,535	6,312
18	4,612	4,934	5,574	6,347
19	4,654	4,974	5,611	6,380
20	4,694	5,012	5,645	6,411
22	4,767	5,081	5,709	6,469
24	4,832	5,144	5,766	6,520
26	4,892	5,201	5,818	6,568
28	4,947	5,253	5,866	6,611
30	4,997	5,301	5,911	6,651

Beispiel: Anhand der Rangsummen aus dem vorangehenden Beispiel und des Wertes $q_{3;0,05} = 3,314$ aus der Tabelle 7.48 ergibt sich: $3,314 \sqrt{3(3 \cdot 6 + 1)/12} = 7,22$



$$\text{AB: } \left| \frac{25}{6} - \frac{58,5}{6} \right| = \frac{33,5}{6} = 5,58 < 7,22$$

$$\text{AC: } \left| \frac{25}{6} - \frac{87,5}{6} \right| = \frac{62,5}{6} = 10,42 > 7,22$$

$$\text{BC: } \left| \frac{58,5}{6} - \frac{87,5}{6} \right| = \frac{29}{6} = 4,83 < 7,22$$

Damit lässt sich anhand der vorliegenden Beobachtungen nur zwischen A und C ein Unterschied auf dem 5%-Niveau feststellen.

Etwas weniger konservativ als (7.169) und das Nemenyi-Verfahren [7.5.4.2] ist der Vergleich nach Tukey und Kramer, wobei (7.169) durch (7.171) ersetzt wird, hierbei wird meist $\alpha = 0,05$ vorgegeben:

$$|\bar{R}_i - \bar{R}_{i'}| > \frac{q_{\infty;k;0,95}}{\sqrt{2}} \cdot \sqrt{\left[\frac{n(n+1)}{12} \right] \left[\frac{1}{n_i} + \frac{1}{n_{i'}} \right]} \quad (7.171)$$

$q_{\infty;k;0,95}$ wird aus Tabelle 7.44 entnommen.

Hat man vorher anhand des H -Tests die globale H_0 auf dem 5%-Niveau abgelehnt, so ist es durchaus zulässig, Unterschiede anhand der 10%-Schranken zu lokalisieren.



Beispiel: Wir nutzen die Zahlen aus dem vorangehenden Beispiel:

$k = 3$, d. h. $q_{\infty;3;0,95} = 3,31$

$$\frac{3,31}{\sqrt{2}} \cdot \sqrt{\left[\frac{18(18+1)}{12} \right] \left[\frac{1}{6} + \frac{1}{6} \right]} = 7,21$$

Das Resultat wird wieder bestätigt.

H -Test mit Stichproben-Untergruppen

Sind k heterogene Stichprobengruppen in je m einander entsprechende, homogene Untergruppen zu je n Werten unterteilbar, dann berechne man (sobald $k \geq 3$, $m \geq 2$, $n \geq 3$) die unter H_0 wie $\chi^2_{\nu=k-1}$ verteilte Prüfgröße $\hat{H}^+$ [LS67]:

$$\hat{H}^+ = \left[\frac{12}{kmn^2(kn+1)} \right] \left[\sum_{i=1}^k \dot{R}_i^2 \right] - 3m(kn+1) \quad (7.172)$$

Dieser Test eliminiert die Varianz zwischen den Untergruppen und hebt die Varianz zwischen den Stichprobengruppen deutlicher hervor. Er setzt wie der H -Test voraus, dass die k Stichprobengruppen eine angenähert gleiche Verteilungsform aufweisen; überdies dürfen keine Wechselwirkungen auftreten, d. h. die k Rangsummen in den m Untergruppen sollten etwa proportional sein. Ein Beispiel mit $k = 3$, $m = 2$, $n = 4$ mag dieses erläutern:

$\begin{smallmatrix} k \\ m \end{smallmatrix}$	k_1	k_2	k_3
m_1	22,0 12 12,5 10 11,5 9 8,0 7	8,5 8 6,5 5 6,0 4 3,5 1	13,0 11 7,5 6 5,0 3 4,5 2
m_2	6,0 12 5,0 10 4,0 8 3,0 6	4,5 9 2,0 4 1,5 3 1,0 2	5,5 11 3,5 7 2,5 5 0,5 1
$\dot{R}_i$	$R_1 = 74$	$R_2 = 36$	$R_3 = 46$

Die Proportionalitätsprobe zwischen den Rangsummen von m_1 und m_2 , d. h. $38 : 18 : 22 \approx 36 : 18 : 24$ fällt zufriedenstellend aus.

$$\hat{H}^+ = \left[\frac{12}{3 \cdot 2 \cdot 4^2(3 \cdot 4 + 1)} \right] [74^2 + 36^2 + 46^2] - 3 \cdot 2(3 \cdot 4 + 1)$$

$$\hat{H}^+ = 7,46 > 5,99 = \chi_{2;0,95}^2$$

Ohne Untergruppierung ($k = 3, n = 8$) erhält man $\hat{H} = 3,33 < 5,99$.

7.5.4.2 Multiple Vergleiche unabhängiger Stichproben nach Nemenyi

Liegen mehrere unterschiedlich behandelte Stichprobengruppen gleicher Umfänge vor und sollen alle diese Gruppen oder Behandlungen miteinander verglichen und auf mögliche Unterschiede geprüft werden, dann bietet sich für Daten mit angenähert gleicher Verteilungsform ein von Nemenyi [Nem63] vorgeschlagener Rangtest an.

Der Test im einzelnen: Es liegen k Behandlungsgruppen mit je n Elementen vor. Den $n \cdot k$ Beobachtungswerten der vereinigten Stichprobe werden Rangordnungszahlen zugeordnet; die kleinste Beobachtung erhält den Rang 1, die größte den Rang $n \cdot k$. Gleich großen Beobachtungswerten werden mittlere Ränge zugeteilt. Addiert man die Ränge der einzelnen Behandlungsgruppen und bildet alle möglichen absoluten Differenzen dieser Summen, dann lassen sich diese anhand eines kritischen Wertes D prüfen. Ist die berechnete Differenz gleich groß oder größer als der für ein gewähltes Signifikanzniveau und die Werte n und k der Tabelle 7.49 zu entnehmende kritische Wert D , dann besteht zwischen den beiden Behandlungen ein echter Unterschied. Ist sie kleiner, so kann an der Gleichheit der beiden Gruppen nicht gezweifelt werden. Bei ungleichen Stichprobenumfängen rechnet man dann, wenn keine oder nur wenige Bindungen auftreten, nach Formel (7.173). Treten viele Bindungen auf, so wird Formel (7.169) bevorzugt.

Tabelle 7.49. Kritische Differenzen D für die Einwegklassifizierung: Vergleich aller möglichen Paare von Behandlungen nach Nemenyi. $P = 0,05$ (zweiseitig) (aus Wilcoxon, F. und Roberta A. Wilcox: Some Rapid Approximate Statistical Procedures, Lederle Laboratories, Pearl River, New York 1964, pp. 29–31)

n	$k = 3$	$k = 4$	$k = 5$	$k = 6$	$k = 7$	$k = 8$	$k = 9$	$k = 10$
1	3,3	4,7	6,1	7,5	9,0	10,5	12,0	13,5
2	8,8	12,6	16,5	20,5	24,7	28,9	33,1	37,4
3	15,7	22,7	29,9	37,3	44,8	52,5	60,3	68,2
4	23,9	34,6	45,6	57,0	68,6	80,4	92,4	104,6
5	33,1	48,1	63,5	79,3	95,5	112,0	128,8	145,8
6	43,3	62,9	83,2	104,0	125,3	147,0	169,1	191,4
7	54,4	79,1	104,6	130,8	157,6	184,9	212,8	240,9
8	66,3	96,4	127,6	159,6	192,4	225,7	259,7	294,1
9	78,9	114,8	152,0	190,2	229,3	269,1	309,6	350,6
10	92,3	134,3	177,8	222,6	268,4	315,0	362,4	410,5
11	106,3	154,8	205,0	256,6	309,4	363,2	417,9	473,3
12	120,9	176,2	233,4	292,2	352,4	413,6	476,0	539,1
13	136,2	198,5	263,0	329,3	397,1	466,2	536,5	607,7
14	152,1	221,7	293,8	367,8	443,6	520,8	599,4	679,0
15	168,6	245,7	325,7	407,8	491,9	577,4	664,6	752,8
16	185,6	270,6	358,6	449,1	541,7	635,9	732,0	829,2
17	203,1	296,2	392,6	491,7	593,1	696,3	801,5	907,9
18	221,2	322,6	427,6	535,5	646,1	758,5	873,1	989,0
19	239,8	349,7	463,6	580,6	700,5	822,4	946,7	1072,4
20	258,8	377,6	500,5	626,9	756,4	888,1	1022,3	1158,1
21	278,4	406,1	538,4	674,4	813,7	955,4	1099,8	1245,9
22	298,4	435,3	577,2	723,0	872,3	1024,3	1179,1	1335,7
23	318,9	465,2	616,9	772,7	932,4	1094,8	1260,3	1427,7
24	339,8	495,8	657,4	823,5	993,7	1166,8	1343,2	1521,7
25	361,1	527,0	698,8	875,4	1056,3	1240,4	1427,9	1617,6

Tabelle 7.49. (Fortsetzung). $P = 0,01$ (zweiseitig)

n	$k = 3$	$k = 4$	$k = 5$	$k = 6$	$k = 7$	$k = 8$	$k = 9$	$k = 10$
1	4,1	5,7	7,3	8,9	10,5	12,2	13,9	15,6
2	10,9	15,3	19,7	24,3	28,9	33,6	38,3	43,1
3	19,5	27,5	35,7	44,0	52,5	61,1	69,8	78,6
4	29,7	41,9	54,5	67,3	80,3	93,6	107,0	120,6
5	41,2	58,2	75,8	93,6	111,9	130,4	149,1	168,1
6	53,9	76,3	99,3	122,8	146,7	171,0	195,7	220,6
7	67,6	95,8	124,8	154,4	184,6	215,2	246,3	277,7
8	82,4	116,8	152,2	188,4	225,2	262,6	300,6	339,0
9	98,1	139,2	181,4	224,5	268,5	313,1	358,4	404,2
10	114,7	162,8	212,2	262,7	314,2	366,5	419,5	473,1
11	132,1	187,6	244,6	302,9	362,2	422,6	483,7	545,6
12	150,4	213,5	278,5	344,9	412,5	481,2	551,0	621,4
13	169,4	240,6	313,8	388,7	464,9	542,4	621,0	700,5
14	189,1	268,7	350,5	434,2	519,4	606,0	693,8	782,6
15	209,6	297,8	388,5	481,3	575,8	671,9	769,3	867,7
16	230,7	327,9	427,9	530,1	634,2	740,0	847,3	955,7
17	252,5	359,0	468,4	580,3	694,4	810,2	927,8	1046,5
18	275,0	391,0	510,2	632,1	756,4	882,6	1010,6	1140,0
19	298,1	423,8	553,1	685,4	820,1	957,0	1095,8	1236,2
20	321,8	457,6	597,2	740,0	885,5	1033,3	1183,3	1334,9
21	346,1	492,2	642,4	796,0	952,6	1111,6	1273,0	1436,0
22	371,0	527,6	688,7	853,4	1021,3	1191,8	1364,8	1539,7
23	396,4	563,8	736,0	912,1	1091,5	1273,8	1458,8	1645,7
24	422,4	600,9	784,4	972,1	1163,4	1357,6	1554,8	1754,0
25	449,0	638,7	833,8	1033,3	1236,7	1443,2	1652,8	1864,6

Weitere Tabellenwerte D für $k > 10$ und $n = 1(1)20$ sind bei Bedarf nach $D = W\sqrt{n(nk)(nk+1)/12}$ zu berechnen, wobei W für $P = 0,05$ (0,01) der Tabelle 7.44 letzte Zeile, entnommen bzw. für anderes P in Tabelle 23 der Biometrika Tables (Pearson und Hartley 1970, S. 178/183) interpoliert wird: z. B. Tabelle 7.49; $P = 0,05$; $n = 25$; $k = 10$: 1617,6; $\sqrt{25(25 \cdot 10)(25 \cdot 10 + 1)/12} = \sqrt{\quad} = 361,5649$; (1) Tabelle 7.44 $k = 10$: $W = 4,47$ und $W\sqrt{\quad} = 1616,2$; (2) [Table 23, Seite 180; Spalte 10] $P' = 0,95$: $W = 4,4745$ und $W\sqrt{\quad} = 1617,8$.



Beispiel: Es werden in einem Vorversuch 20 Ratten auf 4 Futtergruppen verteilt. Die Gewichte nach 70 Tagen enthält die folgende Tabelle (rechts neben den Gewichten sind die Rangzahlen sowie deren Spaltensummen notiert).

I	II	III	IV
203 12	213 16	171 5	207 13
184 7,5	246 18	208 14	152 2
169 4	184 7,5	260 19	176 6
216 17	282 20	193 10	200 11
209 15	190 9	160 3	145 1
55,5	70,5	51	33

Die absoluten Differenzen der Spaltenrangsummen werden dann mit der kritischen Differenz D für $n = 5$ und $k = 4$ auf dem 5%-Niveau verglichen.

	II (70,5)	III (51)	IV (33)
I (55,5)	15	4,5	22,5
II (70,5)		19,5	37,5
III (51)			18

Tabelle 7.49 ($P = 0,05$; $k = 4$; $n = 5$) zeigt $D = 48,1$. Dieser Wert wird von keiner Differenz erreicht. Möglicherweise ließe sich bei vergrößertem Stichprobenumfang ein Unterschied zwischen den Futtergruppen II und IV sichern.

Bei **ungleichen Stichprobenumfängen** mit jeweils 6 oder mehr Beobachtungen ($n_i, n_{i'}, n_{i''}, \dots \geq 6$) und $k \geq 4$ Stichprobengruppen mit insgesamt n Beobachtungen, von denen wenigstens 75% unterschiedliche Werte aufweisen (d. h. höchstens 25% Bindungen sind zulässig), ordnet man Ränge zu, bildet die Stichprobenrangsumme $R_i, R_{i'}, R_{i''}, \dots$ und dividiert sie durch die zugehörigen Stichprobenumfänge, so dass mittlere Stichprobenränge (z. B. $\bar{R}_{i'} = R_{i'}/n_{i'}$) entstehen. Ein Behandlungsunterschied wird dann nach Nemenyi auf dem $100\alpha\%$ -Niveau akzeptiert, sobald

$$\left| \frac{R_i}{n_i} - \frac{R_{i'}}{n_{i'}} \right| > \sqrt{\chi_{k-1; 1-\alpha}^2 \left[\frac{n(n+1)}{12} \right] \left[\frac{n_i + n_{i'}}{n_i n_{i'}} \right]} \quad (7.173)$$

Für $\alpha = 0,05$ und $k = 4$ ergibt sich z. B. der Wert $\chi_{4-1; 0,95}^2 = \chi_{3; 0,95}^2 = 7,81$.

7.5.4.3 H-Test-Variante: Vergleich eines Standards mit mehreren Behandlungen

Vergleiche aller Behandlungen mit einer Kontrolle im Rahmen einer Rang-Varianzanalyse ($n \geq 5$)

Jeweils liegen mindestens 5 Beobachtungen vor. Die Kontrollgruppe wird mit den k Gruppen verglichen.

$H_0: \tilde{\mu} = \tilde{\mu}_i$ für $i = 1, 2, \dots, k$ wird bei zweiseitiger Fragestellung ($H_A: \tilde{\mu} \neq \tilde{\mu}_i$) auf dem 5%-Niveau abgelehnt, sobald

$$|\hat{z}_i| > z_{1-0,05/(2k)} \quad (7.174)$$

mit

$$\hat{z}_i = \frac{\bar{R}_{\text{Kontrolle}} - \bar{R}_i}{\sqrt{[n(n+1)/12](1/n_{\text{Kontrolle}} + 1/n_i)}} \quad (7.175)$$

Beispiel 1: Vergleich zweier Behandlungen mit einer Kontrolle, Rangzahlen in Klammern.



Ansatz		
Kontrolle	$i = 1$	$i = 2$
44 (8)	23 (1)	29 (3)
52 (10)	28 (2)	31 (4)
56 (12)	40 (7)	35 (5)
61 (14)	45 (9)	39 (6)
64 (15)	58 (13)	53 (11)
70 (16)		
$R_{\text{Kontr.}} = 75$	$R_1 = 32$	$R_2 = 29$
Kontrolle der Rangsummen: $\sum R_i = 136 = 16 \cdot 17/2$		
$n_{\text{Kontr.}} = 6$	$n_1 = 5$	$n_2 = 5$
$\bar{R}_{\text{Kontr.}} = 12,5$	$\bar{R}_1 = 6,4$	$\bar{R}_2 = 5,8$

$$z_{1-0,05/(2 \cdot 2)} = z_{0,0125}$$

$$\text{d. h. } z_{0,0125} = 2,24$$

$$\sqrt{[16(16+1)/12](1/6+1/5)} = 2,883$$

$$\hat{z}_1 = (12,5 - 6,4) / 2,883 = 2,12 < 2,24$$

$$\hat{z}_2 = (12,5 - 5,8) / 2,883 = 2,32 > 2,24$$

Für die unbekannten Median-Parameter $\tilde{\mu}_1$, $\tilde{\mu}_2$ und $\tilde{\mu}_{\text{Kontrolle}}$ lässt sich auf dem 5%-Signifikanzniveau $H_0: \tilde{\mu}_{\text{Kontrolle}} = \tilde{\mu}_1$ nicht ablehnen; dagegen wird $H_0: \tilde{\mu}_{\text{Kontrolle}} = \tilde{\mu}_2$ abgelehnt. Diese Aussagen sind vorsichtig zu interpretieren, da beide Prüfgrößen nur wenig von der kritischen Schranke der Standardnormalverteilung abweichen.



Beispiel 2: Vergleich zweier Behandlungen mit einer Kontrolle, Rangzahlen in Klammern.

Kontrolle	A	B
7,43 (2)	9,76 (12)	12,80 (16)
8,51 (6)	8,80 (7)	9,68 (11)
7,20 (1)	7,68 (3)	12,16 (15)
8,27 (4)	9,35 (10)	9,20 (9)
9,84 (13)	8,83 (8)	10,55 (14)
8,32 (5)		
$R_{\text{Kontr.}} = 31$	$R_A = 40$	$R_B = 65$
Kontrolle der Rangsummen: $\sum R = 136 = 16 \cdot 17/2$		
$n_{\text{Kontr.}} = 6$	$n_A = 5$	$n_B = 5$
$\bar{R}_{\text{Kontr.}} = 31/6 = 5,17$	$\bar{R}_A = 40/5 = 8,00$	$\bar{R}_B = 65/5 = 13,00$

$$z_{1-0,05/(2 \cdot 2)} = z_{0,0125}$$

$$\text{d. h. } z_{0,0125} = 2,24$$

$$\sqrt{[16(16+1)/12](1/6+1/5)} = 2,883$$

$$\hat{z}_A = \frac{|5,17 - 8,00|}{2,883} = 0,982 < 2,24$$

$$\hat{z}_B = \frac{|5,17 - 13,00|}{2,883} = 2,716 > 2,24$$

Für die unbekannten Median-Parameter $\tilde{\mu}_A$, $\tilde{\mu}_B$ und $\tilde{\mu}_{\text{Kontrolle}}$ lässt sich auf dem 5%-Signifikanzniveau $H_0: \tilde{\mu}_{\text{Kontrolle}} = \tilde{\mu}_A$ klar erkennbar nicht ablehnen; dagegen wird $H_0: \tilde{\mu}_{\text{Kontrolle}} = \tilde{\mu}_B$ deutlich abgelehnt.

Ein entsprechendes varianzanalytisches Verfahren für die zwei- und die einseitige Fragestellung ($H_0: \mu_i = \mu_0$ bzw. $H_0: \mu_i \leq \mu_0$, mit dem Sollwert μ_0) geben Mee u. Mitarbeiter an [MSL87].

7.5.4.4 Trendtest nach Jonckheere: Vergleich mehrerer geordneter Verteilungsfunktionen oder Mediane

Lässt sich die Alternativhypothese als **Trendhypothese** angeben, etwa: die Verteilungsfunktionen F_i oder die Mediane $\tilde{\mu}_i$ der k Grundgesamtheiten ($i = 1, \dots, k$) bilden eine bestimmte Rangordnung, sagen wir

$$H_A : F_1 \leq F_2 \leq \dots \leq F_k \quad \text{bzw.} \quad \tilde{\mu}_1 \leq \tilde{\mu}_2 \leq \dots \leq \tilde{\mu}_k, \\ \text{wobei wenigstens eine dieser } k - 1 \text{ Ungleichungen} \\ \text{gilt, dann l\"{a}\ss\text{t sich} \quad (7.176)$$

$$H_0 : F_1 = F_2 = \dots = F_k \quad \text{bzw.} \quad \tilde{\mu}_1 = \tilde{\mu}_2 = \dots = \tilde{\mu}_k$$

nach Jonckheere [Jon54] ablehnen. Die Prüfgröße E lässt sich anhand der Standardnormalverteilung beurteilen: sie gilt sogar für sehr ungleiche Stichprobenumfänge, sobald der Gesamtumfang aller Stichproben mindestens gleich 12 ist. Die Voraussetzungen des Tests entsprechen denen des H -Tests. Natürlich müssen die Alternativhypothese und die Entscheidung, diesen Test anzuwenden, vor der Datengewinnung festgelegt werden; andernfalls ist der H -Test anzuwenden.

Beispiel 1: Gegeben seien die geordneten Werte der folgende Tabelle

Gruppe							$\tilde{x}$
A	30	31	34	34	37	39	34
B	36	38	41	41	45	48	41
C	44	45	47	49	50	50	48



mit $n_A = n_B = n_C = 6$ und $\sum n_i = 18 = N$. Beachtet sei: $\sum n_i > 12$ wobei die n_i ungleich sein dürfen; jede der k Stichprobengruppen sollte mindestens 3 Beobachtungen enthalten.

Lässt sich H_0 ablehnen und $H_A : \tilde{\mu}_A \leq \tilde{\mu}_B \leq \tilde{\mu}_C$ auf dem 5%-Niveau sichern?

Zur Prüfgröße E gelangt man, indem für jedes Stichprobenpaar AB , AC , BC ausgezählt wird, wie oft sich $B < A$, $C < A$ und $C < B$ ergeben; für jede Gleichheit $B = A$; $C = A$; $C = B$ wird 1/2 gezählt; E ist die Summe dieser drei Zählungen; sollte H_A gelten, so wird E klein sein.



$$B < A: 36 < 37 \text{ und } 39, \quad \text{d. h. } 2 \\ 38 < 39 \quad 1$$

$$C < A: -$$

$$C < B: 44 < 45 \text{ und } 48 \quad 2 \\ 45 = 45 \text{ und } 45 < 48 \quad 1,5 \\ 47 < 48 \quad 1 \\ \text{und somit } E = \quad 7,5$$

H_A wird auf dem 5%-Niveau (einseitige Fragestellung) abgelehnt, sobald ($N = \sum n_i$)

$$\boxed{E \leq \mu_E - 1,645 \sigma_E} \\ \text{mit} \\ \mu_E = [N^2 - \sum n_i^2]/4 \\ \sigma_E = \sqrt{\{N^2(2N + 3) - \sum n_i^2(2n_i + 3)\}/72} \quad (7.177)$$

Für das Beispiel:

$$\mu_E = [18^2 - (6^2 + 6^2 + 6^2)]/4 = 54,000$$

$$\sigma_E = \sqrt{\{18^2(2 \cdot 18 + 3) - [6^2(2 \cdot 6 + 3) + 6^2(2 \cdot 6 + 3) + 6^2(2 \cdot 6 + 3)]\}/72} = 12,369$$

$$\text{und } E = 7,5 < 33,65 = 54 - 1,645 \cdot 12,369$$

Soll z. B. auf dem 1%-Niveau geprüft werden, so ist $z_{0,95} = 1,645$ durch $z_{0,99} = 2,326$ zu ersetzen.

Will man ein fest vorgegebenes α vermeiden, so prüfe man anhand der Standardnormalverteilung nach

$$\hat{z} = \frac{|E - \mu_E|}{\sigma_E} \quad (7.178)$$

Für das Beispiel:

$$\hat{z} = \frac{|7,5 - 54,0|}{12,369} = 3,76 \quad \text{oder} \quad P < 0,001.$$



Beispiel 2:

Gruppe						$\tilde{x}$
A	106	114	116	127	145	116
B	110	125	143	148	151	143
C	136	139	149	160	174	149

Für diese Werte mit $n_A = n_B = n_C = 5$ und $\sum n_i = 15 = N > 12$ erhält man

$$\begin{array}{ll}
 \mathbf{B} < \mathbf{A}: & 110 < 114 - 145, \quad \text{d. h.} \quad 4 \\
 & 125 < 127 - 145 \quad 2 \\
 & 143 < 145 \quad 1 \\
 \mathbf{C} < \mathbf{A}: & 136 < 145 \quad 1 \\
 & 139 < 145 \quad 1 \\
 \mathbf{C} < \mathbf{B}: & 136 < 143 - 151 \quad 3 \\
 & 139 < 143 - 151 \quad 3 \\
 & 149 < 151 \quad \underline{1} \\
 & E = 16
 \end{array}$$

$$\mu_E = [15^2 - (5^2 + 5^2 + 5^2)]/4 = 37,5$$

$$\sigma_E = \sqrt{\{15^2(2 \cdot 15 + 3) - [5^2(2 \cdot 5 + 3) + 5^2(2 \cdot 5 + 3) + 5^2(2 \cdot 5 + 3)]\}/72} = 9,465$$

und damit $\hat{z} = |16 - 37,5|/9,465 = 2,27$ oder $P \approx 0,012$.

7.5.5 Varianzanalyse für Messwiederholungen (Blockvarianzanalyse)

Für den Fall verbundenener / abhängiger Stichproben, wie sie häufig in der Form von wiederholten Messungen (engl. repeated measurements) vorliegen, kann der Ansatz einer Streuungserlegung (vgl. Kapitel [7.5.2]) weiter modifiziert werden. Neben der Streuung innerhalb (Residual) und zwischen (Faktor) den Messzeitpunkten oder Messbedingungen kann dann auch ein Streuungsanteil fall- bzw. blockweise ausgewiesen werden. Für die Messungen von n Individuen (Blöcken) zu k verschiedenen Zeitpunkten oder verschiedenen Bedingungen folgt entsprechend (7.144):

$$\begin{aligned}
 Q_{\text{insgesamt}} &= Q_{\text{innerhalb}} + Q_{\text{zwischen}} + Q_{\text{Block}} \\
 \sum_{i,j} (x_{ij} - \bar{x})^2 &= \sum_{i,j} (x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x})^2 + n \cdot \sum_{i=1}^k (\bar{x}_{i.} - \bar{x})^2 + k \cdot \sum_{j=1}^n (\bar{x}_{.j} - \bar{x})^2 \quad (7.179)
 \end{aligned}$$

Unter sonst gleichen Voraussetzungen wie in Kapitel [7.5.2] kann die F-Statistik (7.148) zur Überprüfung eines Behandlungseffektes wie folgt modifiziert werden.

$$\hat{F} = \frac{MQ_{\text{zwischen}}}{MQ_{\text{innerhalb}}} = \frac{\frac{1}{k-1} \cdot n \cdot \sum_{i=1}^k (\bar{x}_i - \bar{x})^2}{\frac{1}{(n-k)(k-1)} \cdot \sum_{i,j} (x_{ij} - \bar{x}_i - \bar{x}_{\cdot j} + \bar{x})^2} \quad (7.180)$$

Da der blockinterne, individuell zu erklärende Anteil in der Residualstreuung im Nenner herausgenommen wird, hat diese Teststatistik gegenüber (7.148) eine höhere Power.

Beispiel: Die Veränderung des Gewichtes über 4 Zeitpunkte bei 5 Patienten ist in folgender Tabelle zusammengefasst.



Patient	T1	T2	T3	T4	$\bar{x}_{\cdot j}$
P1	1,5	2,7	2,1	1,3	1,90
P2	1,4	2,9	2,2	1,0	1,88
P3	1,4	2,1	2,4	1,1	1,75
P4	1,2	3,0	2,0	1,3	1,88
P5	1,4	3,3	2,5	1,5	2,18
$\bar{x}_i$	1,38	2,80	2,24	1,24	1,91

Die Berechnung der Teststatistik soll in **R** gezeigt werden. Dazu werden die Daten in einen Rahmen eingetragen, in dem neben dem Faktor (zeit) auch der Bezug auf den einzelnen Fall (patient) zu kennzeichnen ist. Für die Varianzanalyse kann hier auch die Funktion **aov()** verwendet werden. In der Modellspezifikation wird die besondere Behandlung der Residualkomponente in der Streuungszersetzung durch den zusätzlichen Parameter **Error()** gekennzeichnet.

```
> diet <- data.frame(effect = c(1.5, 1.4, 1.4, 1.2, 1.4,
+                               2.7, 2.9, 2.1, 3.0, 3.3,
+                               2.1, 2.2, 2.4, 2.0, 2.5,
+                               1.3, 1.0, 1.1, 1.3, 1.5),
+ patient = factor(paste("pat", rep(1:5, 4), sep="")),
+ zeit = factor(paste("T", rep(c(1, 2, 3, 4), c(5, 5, 5, 5)), sep="")),
+ row.names = NULL); diet
  effect patient zeit
1    1.5   pat1  T1
2    1.4   pat2  T1
3    1.4   pat3  T1
4    1.2   pat4  T1
5    1.4   pat5  T1
6    2.7   pat1  T2
7    2.9   pat2  T2
8    2.1   pat3  T2
9    3.0   pat4  T2
10   3.3   pat5  T2
11   2.1   pat1  T3 ....

> summary(aov(effect ~ zeit + Error(patient), data=diet))
```

```
Error: patient
      Df Sum Sq Mean Sq F value Pr(>F)
Residuals  4  0.39300  0.09825

Error: Within
      Df Sum Sq Mean Sq F value    Pr(>F)
zeit    3  8.1535   2.7178  41.867 1.241e-06 ***
Residuals 12  0.7790   0.0649
```

Der individuelle Anteil der Streuung $Q_{\text{Block}} = 0,393$ wird aus dem Varianzvergleich herausgenommen. Der Wert der Teststatistik nach (7.180) ist somit $\hat{F} = 41,87$ und zeigt einen signifikanten Effekt ($P < 0,05$) über die Zeit.

7.5.6 Friedman-Test

Für den verteilungsunabhängigen Vergleich mehrerer verbundener Stichproben von Messwerten hinsichtlich ihrer zentralen Tendenz steht die von Friedman [Fri37] entwickelte Rangvarianzanalyse, eine zweifache Varianzanalyse mit Rangzahlen, zur Verfügung. Untersucht werden n Individuen, Stichprobengruppen oder Blöcke unter k Bedingungen. Wird die Gesamtstichprobe anhand eines mit dem untersuchten Merkmal möglichst hoch korrelierenden Kontrollmerkmals in Gruppen zu je k Individuen aufgeteilt, so muss beachtet werden, dass die Individuen eines Blocks bezüglich des Kontrollmerkmals gut übereinstimmen. Die k Individuen eines jeden Blocks werden dann nach Zufall auf die k Bedingungen verteilt.

Unter der Hypothese, dass die verschiedenen Bedingungen keinen Einfluss auf die Verteilung der betroffenen Messwerte nehmen, werden sich die Rangplätze der n Individuen oder Blöcke nach Zufall auf die k Bedingungen verteilen. Bildet man unter Annahme der Nullhypothese die Rangsumme für jede der k Bedingungen, so werden diese nicht oder nur zufällig voneinander abweichen. Üben einzelne Bedingungen jedoch einen systematischen Einfluss aus, so werden die k Spalten überzufällig unterschiedliche Rangsummen aufweisen. Zur Prüfung der Nullhypothese: es gibt keinen Behandlungseffekt für die k Behandlungen in n Blöcken, sämtliche Behandlungen entstammen einer gemeinsamen Grundgesamtheit, hat Friedman eine Prüfgröße $\hat{\chi}_R^2$ angegeben.

$$\hat{\chi}_R^2 = \left[\frac{12}{nk(k+1)} \sum_{i=1}^k R_i^2 \right] - 3n(k+1) \quad (7.181)$$

n = Anzahl der Zeilen (die voneinander unabhängig, aber untereinander nicht homogen zu sein brauchen): Blöcke, Individuen, Wiederholungen, Stichprobengruppen

k = Anzahl der Spalten (mit zufälliger Zuordnung der): Bedingungen, Behandlungen, Sorten, Faktoren (zu den Versuchseinheiten)

$\sum_{i=1}^k R_i^2$ = Summe der Quadrate der Spaltenrangsummen für die k zu vergleichenden Behandlungen oder Bedingungen.

Die Teststatistik $\hat{\chi}_R^2$ ist für nicht zu kleines n angenähert wie χ^2 für $k-1$ Freiheitsgrade verteilt. Für kleine Werte von n ist diese Approximation unzureichend. Tabelle 7.50 (Michaelis [Mic71] und Odeh [Ode77]) enthält 5%- und 1%-Schranken. Werden sie durch $\hat{\chi}_R^2$ erreicht oder überstiegen, dann entstammen nicht alle k Spalten einer gemeinsamen Grundgesamtheit. So ist ein $\hat{\chi}_R^2 = 9,000$ für $k = 3$ und $n = 8$ auf dem 1%-Niveau statistisch signifikant.

Bindungen innerhalb einer Zeile (d.h. gleiche Messwerte bzw. mittlere Rangplätze) sind streng genommen nicht zulässig; man berechne dann

$$\hat{\chi}_{R,B}^2 = \left\{ n / \left[n - \frac{1}{k^3 - k} \left(\sum_{i=1}^n \sum_{j=1}^{r_i} t_{ij}^3 - t_{ij} \right) \right] \right\} \cdot \hat{\chi}_R^2 \quad (7.182)$$

(7.182) mit r_i = Anzahl der Bindungen innerhalb der i -ten Zeile, des i -ten Blocks und t_{ij} = Vielfachheit der j -ten Bindung im i -ten Block.

Der Friedman-Test ist ein Homogenitäts-Test. Welche Bedingungen oder Behandlungen untereinander statistisch signifikante Unterschiede aufweisen, kann z. B. nach (7.59) oder nach Wilcoxon und Wilcox (vgl. Abschnitt [7.5.6.1]) geprüft werden.

Tabelle 7.50. 5% und 1%-Schranken für den Friedman-Test (aus Michaelis, J.: Schwellenwerte des Friedman-Tests, Biometr. Zeitschr. **13** (1971), 118–129, S. 122 mit Genehmigung des Autors und des Akademie-Verlages Berlin). Nach der F -Verteilung approximierte kritische Schranken von χ_R^2 für $P = 0,05$; links oben: exakte Werte für $P \leq 0,05$; nach Odeh (1977) korrigierte exakte Werte werden für $k = 5$ und $n = 6$ bis 8 sowie für $k = 6$ und $n = 4$ bis 6 gegeben

$n \backslash k$	3	4	5	6	7	8	9	10	11	12	13	14	15
3	6,000	7,4	8,53	9,86	11,24	12,57	13,88	15,19	16,48	17,76	19,02	20,27	21,53
4	6,500	7,8	8,8	10,29	11,63	12,99	14,34	15,67	16,98	18,3	19,6	20,9	22,1
5	6,400	7,8	8,96	10,49	11,84	13,23	14,59	15,93	17,27	18,6	19,9	21,2	22,4
6	7,000	7,6	9,07	10,57	11,97	13,38	14,76	16,12	17,4	18,8	20,1	21,4	22,7
7	7,143	7,8	9,14	10,62	12,07	13,48	14,87	16,23	17,6	18,9	20,2	21,5	22,8
8	6,250	7,65	9,20	10,68	12,14	13,56	14,95	16,32	17,7	19,0	20,3	21,6	22,9
9	6,222	7,66	9,22	10,73	12,19	13,61	15,02	16,40	17,7	19,1	20,4	21,7	23,0
10	6,200	7,67	9,25	10,76	12,23	13,66	15,07	16,44	17,8	19,2	20,5	21,8	23,1
11	6,545	7,68	9,27	10,79	12,27	13,70	15,11	16,48	17,9	19,2	20,5	21,8	23,1
12	6,167	7,70	9,29	10,81	12,29	13,73	15,15	16,53	17,9	19,3	20,6	21,9	23,2
13	6,615	7,70	9,30	10,83	12,32	13,76	15,17	16,56	17,9	19,3	20,6	21,9	23,2
14	6,143	7,71	9,32	10,85	12,34	13,78	15,19	16,58	17,9	19,3	20,6	21,9	23,2
15	6,400	7,72	9,33	10,87	12,35	13,80	15,20	16,6	18,0	19,3	20,6	21,9	23,2
16	5,99	7,73	9,34	10,88	12,37	13,81	15,23	16,6	18,0	19,3	20,7	22,0	23,2
17	5,99	7,73	9,34	10,89	12,38	13,83	15,2	16,6	18,0	19,3	20,7	22,0	23,3
18	5,99	7,73	9,36	10,90	12,39	13,83	15,2	16,6	18,0	19,4	20,7	22,0	23,3
19	5,99	7,74	9,36	10,91	12,40	13,8	15,3	16,7	18,0	19,4	20,7	22,0	23,3
20	5,99	7,74	9,37	10,92	12,41	13,8	15,3	16,7	18,0	19,4	20,7	22,0	23,3
∞	5,99	7,82	9,49	11,07	12,59	14,07	15,51	16,92	18,31	19,68	21,03	22,36	23,69

Die Methode im einzelnen:

- (1) Die Beobachtungswerte werden in eine Zweifachtafel eingetragen; horizontal k Behandlungen oder Bedingungen, vertikal: n Individuen, Blöcke, Stichprobengruppen oder Wiederholungen.
- (2) Die Werte jeder Zeile werden in eine Rangordnung gebracht; jede Zeile weist also die Rangzahlen 1 bis k auf.
- (3) Für jede Spalte wird die Rangsumme R_i (für die i -te Spalte) ermittelt; alle Rangsummen werden nach $\sum_i R_i = \frac{1}{2}nk(k+1)$ kontrolliert.
- (4) $\hat{\chi}_R^2$ wird nach (7.181) berechnet (bei Bindungen wird $\hat{\chi}_{R,B}^2$ nach (7.182) berechnet).
- (5) Beurteilung von $\hat{\chi}_R^2$ (bzw. von $\hat{\chi}_{R,B}^2$) anhand der Tabelle 7.50 bzw. für großes n anhand der χ^2 -Tabelle oder besser anhand von (7.185).

$\hat{\chi}_R^2$ ist für **nicht zu kleines n angenähert** nach χ_{k-1}^2 verteilt, wobei ein eher konservativer Test vorliegt. Strebt man einen etwas liberalen Test an, so vergleiche man

$$\hat{F} = \frac{(n-1)\hat{\chi}_R^2}{n(k-1) - \hat{\chi}_R^2} \quad \text{mit} \quad F_{(k-1);(n-1)(k-1);\alpha} \quad (7.183)$$

Tabelle 7.50. (Fortsetzung). Nach der F -Verteilung approximierte kritische Schranken χ_R^2 für $P = 0,01$; links oben: exakte Werte für $P \leq 0,01$; nach Odeh (1977) korrigierte exakte Werte werden für $k = 5$ und $n = 6$ bis 8 sowie für $k = 6$ und $n = 4$ bis 6 gegeben

$n \backslash k$	3	4	5	6	7	8	9	10	11	12	13	14	15
3	–	9,000	10,13	11,76	13,26	14,78	16,28	17,74	19,19	20,61	22,00	23,38	24,76
4	8,000	9,600	11,20	12,71	14,19	15,75	17,28	18,77	20,24	21,7	23,1	24,5	25,9
5	8,400	9,960	11,68	13,23	14,74	16,32	17,86	19,37	20,86	22,3	23,7	25,2	26,6
6	9,000	10,200	11,87	13,62	15,10	16,69	18,25	19,77	21,3	22,7	24,2	25,6	27,0
7	8,857	10,54	12,11	13,69	15,35	16,95	18,51	20,04	21,5	23,0	24,4	25,9	27,3
8	9,000	10,50	12,30	13,87	15,53	17,15	18,71	20,24	21,8	23,2	24,7	26,1	27,5
9	9,556	10,44	12,27	14,01	15,68	17,29	18,87	20,42	21,9	23,4	24,95	26,3	27,7
10	9,600	10,53	12,38	14,12	15,79	17,41	19,00	20,53	22,0	23,5	25,0	26,4	27,9
11	9,455	10,60	12,46	14,21	15,89	17,52	19,10	20,64	22,1	23,6	25,1	26,6	28,0
12	9,500	10,68	12,53	14,28	15,96	17,59	19,19	20,73	22,2	23,7	25,2	26,7	28,0
13	9,385	10,72	12,58	14,34	16,03	17,67	19,25	20,80	22,3	23,8	25,3	26,7	28,1
14	9,143	10,76	12,64	14,40	16,09	17,72	19,31	20,86	22,4	23,9	25,3	26,8	28,2
15	8,933	10,80	12,68	14,44	16,14	17,78	19,35	20,9	22,4	23,9	25,4	26,8	28,2
16	8,79	10,84	12,72	14,48	16,18	17,81	19,40	20,9	22,5	24,0	25,4	26,9	28,3
17	8,81	10,87	12,74	14,52	16,22	17,85	19,50	21,0	22,5	24,0	25,4	26,9	28,3
18	8,84	10,90	12,78	14,56	16,25	17,87	19,5	21,1	22,6	24,1	25,5	26,9	28,3
19	8,86	10,92	12,81	14,58	16,27	17,90	19,5	21,1	22,6	24,1	25,5	27,0	28,4
20	8,87	10,94	12,83	14,60	16,30	18,00	19,5	21,1	22,6	24,1	25,5	27,0	28,4
∞	9,21	11,35	13,28	15,09	16,81	18,48	20,09	21,67	23,21	24,73	26,22	27,69	29,14

Optimal erscheint eine lineare Kombination beider Methoden:

$$\hat{J} = [\hat{\chi}_R^2 + (k-1)\hat{F}]/2 = \frac{\hat{\chi}_R^2}{2} \left[1 + \frac{(n-1)(k-1)}{n(k-1) - \hat{\chi}_R^2} \right] \quad (7.184)$$

$\hat{J}$ wird mit J_α verglichen, dem Durchschnitt beider kritischen Werte:

$$J_\alpha = [\chi_{k-1;\alpha}^2 + (k-1)F_{(k-1):(n-1)(k-1);\alpha}]/2 \quad (7.185)$$

Für $k+n \geq 10$ mit $k > 3$ und $n > 3$ benutze man die Approximation $\hat{J}$



Beispiel: Drei halbbittere Schokoladensorten A, B und C wurden durch 4 Personen einer Zufallsstichprobe Erwachsener getestet. Das Ergebnis ist in der folgenden Tabelle zusammengefasst (Rang 1 gilt als beste Note; in Klammern ist der Preis in Euro angegeben, den die betreffende Person für angemessen hielt).

Block	A	B	C
1	1 (2,20)	2 (2,00)	3 (1,80)
2	1 (2,40)	2 (1,80)	3 (1,60)
3	1 (2,50)	2 (1,90)	3 (1,70)
4	3 (1,70)	1 (2,50)	2 (1,90)

Die Berechnung der Teststatistik erfolgt nach

1	2	3
1	2	3
1	2	3
3	1	2
6	7	11

$$36 + 49 + 121 = 206 = \sum_{i=1}^k R_i^2 \text{ ergibt und damit}$$

$$\hat{\chi}_R^2 = \left[\frac{12}{4 \cdot 3 \cdot (3+1)} 206 \right] - 3 \cdot 4 \cdot (3+1) = 3,5 < 6,5 = \chi_{R;0,05}^2 \quad \text{für } \begin{matrix} n=4 \\ k=3 \end{matrix}$$

d. h., H_0 : die drei Schokoladensorten sind nicht zu unterscheiden, lässt sich auf dem 5%-Niveau nicht ablehnen. Für $n = 4$ und $k = 3$ gibt es insgesamt $(k!)^n = (3!)^4 = (3 \cdot 2 \cdot 1)^4 = 6^4 = 1296$ mögliche Aufteilungen der Ränge, von denen wir noch weitere fünf herausgreifen und in der folgenden Tabelle vergleichen:

Tab.-Nr.	1	2	3	4	5	6
Ausgewählte Permutationen	1 2 3 1 2 3 3 2 1 3 2 1	1 2 3 1 2 3 2 3 1 3 1 2	1 2 3 1 2 3 1 3 2 3 2 1	1 2 3 1 2 3 1 2 3 3 1 2	1 2 3 1 2 3 1 2 3 1 3 2	1 2 3 1 2 3 1 2 3 1 2 3
$\hat{\chi}_k^2$	0	0,5	1,5	3,5	6,5	8,0

Zum Vergleich: $\chi_{R(n=4,k=3,\alpha=0,05)}^2 = 6,5$ und $\chi_{R(n=4,k=3,\alpha=0,01)}^2 = 8,0$

In **R** kann der Friedman-Test mit der Funktion **friedman.test()** berechnet werden. Die Daten aus dem Beispiel (Preise) müssen dazu entsprechend in einem Datenrahmen zusammengestellt werden. Das Ergebnis stimmt mit dem oben hergeleiteten Resultat überein.

```
> test <- data.frame(preis = c(2.20, 2.40, 2.50, 1.70,
+                             2.00, 1.80, 1.90, 1.60,
+                             1.80, 1.60, 1.70, 1.90),
+   subj = factor(paste("p", rep(1:4, 3), sep="")),
+   sorte = factor(paste("t", rep(c(1, 2, 3), c(4, 4, 4)), sep="")),
+   row.names = NULL); test
  preis subj sorte
1   2.2  p1    t1
2   2.4  p2    t1
3   2.5  p3    t1
4   1.7  p4    t1 ...

> friedman.test(preis ~ sorte | subj, data=test)

Friedman rank sum test

data:  preis and sorte and subj
Friedman chi-squared = 3.5, df = 2, p-value = 0.1738
```

Paarweise multiple Vergleiche der Rangsummen. Paarweise multiple Vergleiche der Rangsummen R_i lassen sich für nicht zu kleines n ($n \geq 5$) approximativ nach (7.186) durchführen (vgl. Tab. 7.48). Für

$$|R_i - R_{i'}| > q_{k;\alpha} \sqrt{nk(k+1)/12} \quad (7.186)$$

wird H_0 : „Gleichheit der beiden betrachteten Behandlungen“ auf dem $100\alpha\%$ -Niveau abgelehnt. Einfacher ist der in [7.5.6.1] gegebene Ansatz.

Vergleiche aller Behandlungen mit einer Kontrolle, für Experimente mit einer Kontrolle (einem Standard S) und $k \geq 2$ Behandlungen in $n \geq 15$ Blöcken. Jeder Block enthält $1 + k$ Einheiten und zwar so, dass blockintern jeweils eine Kontrolle (Standard) und k Behandlungen streng zufällig zugeordnet werden und die Zuordnungen zu den unterschiedlichen Blöcken unabhängig sind. Dem Resultat des Experiments entsprechend, erhalten in jedem Block die Einheiten Rangzahlen von 1 bis $k + 1$ zugeordnet. Die Summe aller Rangzahlen derselben Behandlung über alle Blöcke bilden die Friedman-Summen.

H_0 : alle $[(1 + k)!]^n$ Rangordnungen sind gleichwahrscheinlich, d.h. die Kontrolle (Standardbehandlung) und alle zu prüfenden Behandlungen sind gleich wirksam.
 H_A : mindestens eine Behandlung unterscheidet sich von der Kontrolle (Standardbehandlung).

Liegen mindestens $n = 15$ Blöcke vor und ist die Kontrolle, der Standard S , ihre Rangsumme sei R_s , mit den k ($i = 1, \dots, k$) Rangsummen zu vergleichen, so wird H_0 auf dem 100% -Niveau zugunsten von H_A abgelehnt, sobald

$$|R_s - R_i| > G_{k;\alpha} \cdot \sqrt{n(k+1)(k+2)/6} \quad (7.187)$$

Der Wert $G_{k;\alpha}$ ist für $k \leq 10$ der Tabelle 7.51 (aus Sachs [Sac90]; S.208, dort auch weitere Werte) zu entnehmen, etwa wenn vier Behandlungen mit einer Kontrolle, einem Standard, verglichen werden, für $\alpha = 0,05$ und $k = 4$ der Wert $G_{4;0,05} = 2,44$.

Tabelle 7.51. Kritische Werte $G_{k;\alpha}$ für den multiplen Vergleich gegen eine Kontrolle, basierend auf den zweiseitigen Quantilen der multivariaten t-Verteilung ($\nu = \infty$ und $\rho = 0, 5$); aus Sachs [Sac90], S. 208

α	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$	$k = 7$	$k = 8$	$k = 9$	$k = 10$
0,01	2,79	2,92	3,00	3,06	3,11	3,15	3,19	3,22	3,25
0,05	2,21	2,35	2,44	2,51	2,57	2,61	2,65	2,69	2,72
0,10	1,92	2,06	2,16	2,23	2,29	2,34	2,38	2,42	2,45



Beispiel: Unterscheiden sich die Kartoffelsorten A bis D bezüglich ihrer Erträge von einer Standardsorte S ($\alpha = 0,05$)? Der schlechteste Ertrag erhält die Rangzahl 1, der beste die Rangzahl 5. Berechnet und verglichen werden die Sorten-Rangsummen A bis D mit S über 15 Standorte ($n = 15$)

Ort	S	A	B	C	D
1	1	4	2	5	3
2	1	3	5	4	2
3	2	4	3	5	1
4	1	3	4	5	2
5	1	2	5	4	3
6	1	3	4	5	2
7	2	4	3	5	1
8	1	2	5	4	3
9	2	3	5	4	1
10	1	4	2	5	3
11	3	2	5	4	1
12	1	3	4	5	2
13	1	4	2	5	3
14	2	5	3	4	1
15	1	3	5	4	2
Summe	$R_S = 21$	$R_A = 49$	$R_B = 57$	$R_C = 68$	$R_D = 30$

Kontrolle: Summe aller Rangsummen = $n(k+1)(k+2)/2$
 $21 + 49 + 57 + 68 + 30 = 225 = 15(4+1)(4+2)/2$.

Mit $G_{4;0,05} = 2,44$ erhält man die kritische Schranke $2,44 \cdot \sqrt{15(4+1)(4+2)/6} = 21,13$. Die absolut genommenen Differenzen

$$\left. \begin{array}{l} |R_S - R_C| = |21 - 68| = 47 \\ |R_S - R_B| = |21 - 57| = 36 \\ |R_S - R_A| = |21 - 49| = 28 \\ |R_S - R_D| = |21 - 30| = 9 \end{array} \right\} \begin{array}{l} > 21,13 \\ > 21,13 \\ > 21,13 \\ < 21,13 \end{array}$$

zeigen auf dem 5%-Signifikanzniveau, dass sich die Sorten *A* und *B* und besonders die Sorte *C*, die auch die kleinste Streuung der Rangzahlen aufweist, von der Standardsorte deutlich unterscheiden.

7.5.6.1 Multiple paarweise Vergleiche nach Wilcoxon und Wilcoxon

Der Friedman-Test ist eine zweifache Varianzanalyse mit Rangzahlen; der entsprechende multiple Vergleich stammt von Wilcoxon und Wilcoxon [WW64]. Der Test ähnelt dem von Nemenyi [7.5.4.2] gegebenen Verfahren.

Der Vergleich im einzelnen: Verglichen werden wieder k Behandlungen mit je n Wiederholungen. Jeder Behandlung ist eine Rangzahl von 1 bis k zuzuordnen, so dass n Rangordnungen resultieren. Die Ränge der einzelnen Stichproben werden addiert; ihre Differenzen vergleicht man mit dem Wert der kritischen Differenz aus Tabelle 7.52. Wird die tabellierte kritische Differenz erreicht oder überschritten, dann entstammen die dem Vergleich zugrunde liegenden Behandlungen unterschiedlichen Grundgesamtheiten. Unterschreitet die berechnete Differenz die tabellierte D dann gilt der Unterschied noch als zufällig.

Weitere Tabellenwerte D für $k > 10$ und $n = 1(1)20$ sind bei Bedarf nach $D = W\sqrt{nk(k+1)/12}$ zu berechnen, wobei W für $P = 0,05$ (0,01) der Tabelle 7.44 letzte Zeile, entnommen bzw. für anderes P in Table 23 der Biometrika Tables (Pearson und Hartley 1970, S. 178/183) interpoliert wird (z. B. $D = 67,7$ [Tab. 187; $P = 0,05$; $n = 25$; $k = 10$], für $P' = 0,95$ ergibt sich (Table 23, S. 180, Spalte 10) $W = 4,4745$ und $4,4745\sqrt{25 \cdot 10(10+1)/12} = 67,736$; mit Tabelle 7.44 für $k = 10$; $W = 4,47$ und $D = 67,668$).

Beispiel: Sechs Personen erhalten je 6 verschiedene Diuretika (Harntreibende Mittel *A* bis *F*). Zwei Stunden nach der Behandlung wird die Natriumausscheidung (in mval) bestimmt. Gefragt wird nach den Diuretika, die sich aufgrund der Natriumausscheidung von den anderen unterscheiden. Die folgende Tabelle enthält die Daten, rechts daneben jeweils die Rangzahlen mit den Spaltenrangsummen (Quelle: Wilcoxon, F. and Roberta A. Wilcoxon: Some Approximate Statistical Procedures, Lederle Laboratories, New York 1964, pp. 11 und 12).



Person	A		B		C		D		E		F	
1	3,88	1	30,58	5	25,24	3	4,44	2	29,41	4	38,87	6
2	5,64	1	30,14	3	33,52	6	7,94	2	30,72	4	33,12	5
3	5,76	2	16,92	3	25,45	4	4,04	1	32,92	5	39,15	6
4	4,25	1	23,19	4	18,85	3	4,40	2	28,23	6	28,06	5
5	5,91	2	26,74	5	20,45	3	4,23	1	23,35	4	38,23	6
6	4,33	1	10,91	3	26,67	6	4,36	2	12,00	4	26,65	5
	8		23		25		10		27		33	

Tabelle 7.52. Kritische Differenzen für die Zweiwegklassifizierung: Vergleich aller möglichen Paare von Behandlungen. $P = 0,05$ (zweiseitig) (aus Wilcoxon, F. and Roberta A. Wilcox: Some Rapid Approximate Statistical Procedures, Lederle Laboratories, Pearl River, New York 1964, pp. 36–38)

n	k = 3	k = 4	k = 5	k = 6	k = 7	k = 8	k = 9	k = 10
1	3,3	4,7	6,1	7,5	9,0	10,5	12,0	13,5
2	4,7	6,6	8,6	10,7	12,7	14,8	17,0	19,2
3	5,7	8,1	10,6	13,1	15,6	18,2	20,8	23,5
4	6,6	9,4	12,2	15,1	18,0	21,0	24,0	27,1
5	7,4	10,5	13,6	16,9	20,1	23,5	26,9	30,3
6	8,1	11,5	14,9	18,5	22,1	25,7	29,4	33,2
7	8,8	12,4	16,1	19,9	23,9	27,8	31,8	35,8
8	9,4	13,3	17,3	21,3	25,5	29,7	34,0	38,3
9	9,9	14,1	18,3	22,6	27,0	31,5	36,0	40,6
10	10,5	14,8	19,3	23,8	28,5	33,2	38,0	42,8
11	11,0	15,6	20,2	25,0	29,9	34,8	39,8	44,9
12	11,5	16,2	21,1	26,1	31,2	36,4	41,6	46,9
13	11,9	16,9	22,0	27,2	32,5	37,9	43,3	48,8
14	12,4	17,5	22,8	28,2	33,7	39,3	45,0	50,7
15	12,8	18,2	23,6	29,2	34,9	40,7	46,5	52,5
16	13,3	18,8	24,4	30,2	36,0	42,0	48,1	54,2
17	13,7	19,3	25,2	31,1	37,1	43,3	49,5	55,9
18	14,1	19,9	25,9	32,0	38,2	44,5	51,0	57,5
19	14,4	20,4	26,6	32,9	39,3	45,8	52,4	59,0
20	14,8	21,0	27,3	33,7	40,3	47,0	53,7	60,6
21	15,2	21,5	28,0	34,6	41,3	48,1	55,1	62,1
22	15,5	22,0	28,6	35,4	42,3	49,2	56,4	63,5
23	15,9	22,5	29,3	36,2	43,2	50,3	57,6	65,0
24	16,2	23,0	29,9	36,9	44,1	51,4	58,9	66,4
25	16,6	23,5	30,5	37,7	45,0	52,5	60,1	67,7

Die absoluten Differenzen sind:

		D	B	C	E	F
		10	23	25	27	33
A	8	2	15	17	19*	25**
D	10		13	15	17	23**
B	23			2	4	10
C	25				2	8
E	27					6

Die kritische Differenz für $k = 6$ und $n = 6$ beträgt auf dem 5%-Niveau (vgl. Tabelle 7.52) 18,5, auf dem 1%-Niveau 21,8. Die auf dem 5%-Niveau statistisch signifikanten Differenzen sind mit einem Stern (*) versehen, die auf dem 1%-Niveau statistisch signifikanten Differenzen sind mit 2 Sternen (**) ausgezeichnet.

Man kann also feststellen, dass sich das Präparat F aufgrund einer starken Natriumdiurese auf dem 1%-Niveau von den Diuretika A und D unterscheidet. Das Präparat E unterscheidet sich auf dem 5%-Niveau vom Präparat A ; andere Differenzen sind auf dem 5%-Niveau nicht bedeutsam.

7.5.6.2 Page-Test für geordnete Alternativen, die auf Friedman-Rangsummen basieren

Nach Page [Pag63] lässt sich, bei entsprechendem Vorwissen, der Nullhypothese: Gleichheit der Behandlungseffekte [$H_0: E(\bar{R}_1) = E(\bar{R}_2) = \dots = E(\bar{R}_k)$] die Alternativhypothese: monoton ansteigender Behandlungseffekt (Aufwärtstrend) [$H_A: E(\bar{R}_1) \leq E(\bar{R}_2) \leq \dots \leq E(\bar{R}_k)$] mit wenigstens einer gültigen Ungleichung gegenüberstellen. H_0 wird abgelehnt, wenn auf dem zuvor

Tabelle 7.52. (Fortsetzung) $P = 0,01$ (zweiseitig)

n	k = 3	k = 4	k = 5	k = 6	k = 7	k = 8	k = 9	k = 10
1	4,1	5,7	7,3	8,9	10,5	12,2	13,9	15,6
2	5,8	8,0	10,3	12,6	14,9	17,3	19,7	22,1
3	7,1	9,8	12,6	15,4	18,3	21,2	24,1	27,0
4	8,2	11,4	14,6	17,8	21,1	24,4	27,8	31,2
5	9,2	12,7	16,3	19,9	23,6	27,3	31,1	34,9
6	10,1	13,9	17,8	21,8	25,8	29,9	34,1	38,2
7	10,9	15,0	19,3	23,5	27,9	32,3	36,8	41,3
8	11,7	16,1	20,6	25,2	29,8	34,6	39,3	44,2
9	12,4	17,1	21,8	26,7	31,6	36,6	41,7	46,8
10	13,0	18,0	23,0	28,1	33,4	38,6	44,0	49,4
11	13,7	18,9	24,1	29,5	35,0	40,5	46,1	51,8
12	14,3	19,7	25,2	30,8	36,5	42,3	48,2	54,1
13	14,9	20,5	26,2	32,1	38,0	44,0	50,1	56,3
14	15,4	21,3	27,2	33,3	39,5	45,7	52,0	58,4
15	16,0	22,0	28,2	34,5	40,8	47,3	53,9	60,5
16	16,5	22,7	29,1	35,6	42,2	48,9	55,6	62,5
17	17,0	23,4	30,0	36,7	43,5	50,4	57,3	64,4
18	17,5	24,1	30,9	37,8	44,7	51,8	59,0	66,2
19	18,0	24,8	31,7	38,8	46,0	53,2	60,6	68,1
20	18,4	25,4	32,5	39,8	47,2	54,6	62,2	69,8
21	18,9	26,0	33,4	40,9	48,3	56,0	63,7	71,6
22	19,3	26,7	34,1	41,7	49,5	57,3	65,2	73,2
23	19,8	27,3	34,9	42,7	50,6	58,6	66,7	74,9
24	20,2	27,8	35,7	43,6	51,7	59,8	68,1	76,5
25	20,6	28,4	36,4	44,5	52,7	61,1	69,5	78,1

gewählten Niveau die Summe der Produkte aus hypothetischem Rang und zugehöriger Rangsumme den entsprechenden Wert der Tabelle 7.53 erreicht oder übersteigt.

$$L = 1R_1 + 2R_2 + \dots + kR_k = \sum_{i=1}^{i=k} iR_i \quad (7.188)$$

H_0 wird für $L \geq \text{kritischer Wert aus Tab. 7.53}$ abgelehnt.

Für nicht zu kleine Stichprobenumfänge lässt sich L auch anhand der Approximation (7.189) ablehnen, sobald

$$\begin{aligned} L &\geq \mu + z\sigma + 1/2 \\ \mu &= nk(k+1)^2/4 \\ \sigma &= \sqrt{nk^2(k+1)(k^2-1)/144} \\ z_{0,95} &= 1,645 \end{aligned} \quad (7.189)$$

gilt; und zwar mit dem in (7.189) unten angegebenen Wert z auf dem 5%-Niveau. Da unter H_A : Aufwärtstrend eine einseitige Fragestellung vorliegt, sind für z auch die entsprechenden einseitigen Schranken der Standardnormalverteilung zu wählen, etwa $z_{0,99} = 2,326$ für den Test auf dem 1%-Niveau. Der Wert $1/2$ in (7.189) ist die Stetigkeitskorrektur.

Tabelle 7.53. Einige 5% und 1%-Schranken für den Page-Test Weitere Schranken für $\alpha = 0,05; 0,01; 0,001$ und $n = 2(1)50$ sowie $k = 3(1)10$ sind bei Bedarf Page (1963) zu entnehmen. Ausführliche Schranken für $\alpha \leq 0,20$, $n = 2(1)10$ und $k = 3(1)8$ gibt Odeh (1977). Exakte P -Werte berechnet man nach Wellek (1989)

P	0,05						0,01					
$k \backslash n$	3	4	5	6	7	8	3	4	5	6	7	8
3	41	84	150	244	370	532	42	87	155	252	382	549
4	54	111	197	321	487	701	55	114	204	331	501	722
5	66	137	244	397	603	869	68	141	251	409	620	893
6	79	163	291	474	719	1037	81	167	299	486	737	1063
7	91	189	338	550	835	1204	93	193	346	563	855	1232
8	104	214	384	640	950	1371	106	220	393	625	972	1401
9	116	240	431	718	1065	1537	119	246	441	701	1088	1569
10	128	266	477	793	1181	1704	131	272	487	777	1205	1737
11	141	292	523	869	1295	1868	144	298	534	852	1321	1905
12	153	317	570	946	1410	2035	156	324	581	928	1437	2072
13	165	343	615	1003	1525	2201	169	350	628	1022	1553	2240



Beispiel: 9 Gutachter (G) schlagen Rangordnungen für vier Objekte (A, B, C, D) vor. Das Ergebnis ist in der folgenden Tabelle zusammengefasst.

G	A	B	C	D
1	3	2	1	4
2	4	2	3	1
3	4	1	2	3
4	4	2	3	1
5	3	2	1	4
6	4	1	2	3
7	4	3	2	1
8	3	1	2	4
9	3	1	4	2
R_i	32	15	20	23

Das Berechnung der Teststatistik nach dem Friedman-Test führt auf:

$$\chi_R^2 = \left[\frac{12}{9 \cdot 4(4+1)} (32^2 + 15^2 + 20^2 + 23^2) \right] - 3 \cdot 9(4+1) = [(12/180)(2178)] - 135 = 10,2;$$

$$10,2 > 7,66 \text{ für } \alpha = 0,05 \text{ mit } n = 9 \text{ und } k = 4;$$

d.h. die Gutachter unterscheiden sich hinsichtlich der Beurteilung der Objekte auf dem 5%-Niveau.

Angenommen, wir hätten aufgrund unseres Vorwissens für die 4 Objekte die Reihenfolge: $R_B < R_C < R_D < R_A$ unter H_A erwartet ($\alpha = 0,05$). Mit

$$L = 1 \cdot 15 + 2 \cdot 20 + 3 \cdot 23 + 4 \cdot 32 = 252$$

$$L = 252 > 240 \quad \text{für } n = 9, k = 4 \text{ und } \alpha = 0,05 \text{ (Tab. 7.53)}$$

wird H_0 auf dem 5%-Niveau zugunsten eines Aufwärtstrends abgelehnt. Und nach (7.189):

$$\mu = 9 \cdot 4(4 + 1)^2 / 4 = 225$$

$$\sigma = \sqrt{9 \cdot 4^2(4 + 1)(4^2 - 1) / 144} = 8,660$$

$$225 + 1,645 \cdot 8,660 + 0,5 = 239,7 \quad \text{oder} \quad 240$$

$$L = 252 > 240 \quad \text{für} \quad n = 9, k = 4 \text{ und } \alpha = 0,05.$$

7.5.7 Zweifache Varianzanalyse

Bei der zweifachen Varianzanalyse (two factorial analysis of variances) wird der Einfluss zweier Faktoren (unabhängige Variablen) auf ein bestimmtes Merkmal (abhängige Variable) untersucht. Ein Beispiel dazu ist die Wirksamkeit einer bestimmten Behandlungsform (Faktor A mit drei Stufen - Placebo, einfache Dosierung, doppelte Dosierung) unter Berücksichtigung des Geschlechtes (Faktor B mit zwei Stufen - männlich, weiblich) auf die Depressivität, die durch einen bestimmten Punktwert (Score) gemessen wird [Bor99]. Vom Faktor A sollen allgemein a , vom Faktor B sollen allgemein b Stufen in der Betrachtung berücksichtigt werden. Für diesen Studienansatz gibt es somit insgesamt $a \cdot b$ verschiedene Klassen. Jede dieser Klassen soll zunächst eine feste Zahl n an Beobachtungen (balanciert) enthalten. Die beobachteten Werten können mit einer mehrfachen Indizierung beschrieben werden.

$$x_{i,j,k} \quad \text{für } i = 1, 2, \dots, a; j = 1, 2, \dots, b \text{ und } k = 1, 2, \dots, n$$

Das Schema und die verwendete Notation (Punktnotation für Summen), die zum Verständnis der folgenden Formeln notwendig ist, wird in Tabelle 7.54 dargestellt.

Tabelle 7.54. Notation zur zweifachen Varianzanalyse mit balancierten Daten

Klassen	B_1	B_2	...	B_j	...	B_b	Summen
A_1	x_{111}	x_{121}	...	x_{1j1}	...	x_{1b1}	$x_{1.1}$
	x_{112}	x_{122}	...	x_{1j2}	...	x_{1b2}	$x_{1.2}$
	...	...	...	...	...	...	...
	x_{11n}	x_{12n}	...	x_{1jn}	...	x_{1bn}	$x_{1.n}$
Summen	$x_{11.}$	$x_{12.}$	...	$x_{1j.}$	...	$x_{1b.}$	$x_{1..}$
⋮	...	...	...	...	...	...	⋮
⋮	...	...	...	...	...	...	⋮
⋮	...	...	...	...	...	...	⋮
A_a	x_{a11}	x_{a21}	...	x_{aj1}	...	x_{ab1}	$x_{a.1}$
	x_{a12}	x_{a22}	...	x_{aj2}	...	x_{ab2}	$x_{a.2}$
	...	...	...	...	...	...	...
	x_{a1n}	x_{a2n}	...	x_{ajn}	...	x_{abn}	$x_{a.n}$
Summen	$x_{a1.}$	$x_{a2.}$	...	$x_{aj.}$	...	$x_{ab.}$	$x_{a..}$
Summen	$x_{.1.}$	$x_{.2.}$	...	$x_{.j.}$	...	$x_{.b.}$	$x_{...}$

Wie bei der einfachen (einfaktoriellen) Varianzanalyse basiert die zweifache Varianzanalyse auf einer Zerlegung der Abweichungsquadrate (Quadratsummen) auf der Grundlage von

$$(x_{ijk} - \bar{x})^2 = [(\bar{x}_{i..} - \bar{x}) + (\bar{x}_{.j.} - \bar{x}) + (x_{ijk} - \bar{x}_{ij.}) + (\bar{x}_{ij.} - \bar{x}_{i..} - \bar{x}_{.j.} + \bar{x})]^2$$

Die gesamte Quadratsumme SAQ_{ges} wird in der zweifachen Varianzanalyse in die Quadratsumme des Faktors SAQ_A , die Quadratsumme des Faktors SAQ_B , die Wechselwirkungsquadratsumme $SAQ_{A \times B}$ und die Fehlerquadratsumme SAQ_{in} zerlegt.

Die Wechselwirkung oder **Interaktion** (interaction) kennzeichnet einen über die Haupteffekte (hinsichtlich der Faktoren A und B) hinausgehenden Effekt, der auf bestimmte Kombinationen der Stufen der einzelnen Faktoren zurückgeführt werden kann.



Ein Test auf einen unterschiedlichen Einfluss der Stufen der Faktoren und auf Vorliegen einer Wechselwirkung zielt somit auf verschiedene, voneinander unabhängige Nullhypothesen. Dazu ist eine Unterscheidung nach dem Studienansatz nach einem Modell mit festen (Modell I, fixed effect model), zufälligen (Modell II, random effect model) oder gemischten (Modell III, mixed effect model) Effekten notwendig.

$$\begin{aligned}
 SAQ_{ges} &= SAQ_A + SAQ_B + SAQ_{A \times B} + SAQ_{in} \quad \text{mit} \\
 SAQ_{ges} &= \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (x_{ijk} - \bar{x})^2 \quad \text{Summe der Abweichungsquadrate} \\
 SAQ_A &= \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{x}_{i..} - \bar{x})^2 \quad \text{innerhalb der Stufen von A} \\
 SAQ_B &= \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{x}_{.j.} - \bar{x})^2 \quad \text{innerhalb der Stufen von B} \\
 SAQ_{in} &= \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{x}_{ijk} - \bar{x}_{ij.})^2 \quad \text{innerhalb der Klassen (Rest)} \\
 SAQ_{A \times B} &= SAQ_{ges} - SAQ_A - SAQ_B - SAQ_{in} \quad \text{Wechselwirkung}
 \end{aligned} \tag{7.190}$$

Modell I mit festen Effekten oder systematischen Komponenten: Spezielle Behandlungen, Arzneien, Methoden, Stufen eines Faktors, Sorten, Versuchstiere, Maschinen werden bewusst ausgewählt und in den Versuch bzw. die Studie einbezogen, weil gerade sie (etwa die Dosierung A, B und C) von praktischem Interesse sind und man etwas über ihre mittleren Effekte und deren Bedeutsamkeit erfahren möchte. Vergleiche von Erwartungswerten stehen somit hier im Vordergrund! Die Nullhypothesen für diesen Studienansatz lauten:

- H_{0A} : kein unterschiedlicher Einfluss des Faktors A.
Die Beobachtungen unter den Stufen des Faktors A entstammen Grundgesamtheiten mit gleichem Erwartungswerten: $\mu_1 = \mu_2 = \dots = \mu_a$.
- H_{0B} : kein unterschiedlicher Einfluss des Faktors B.
Die Beobachtungen unter den Stufen des Faktors B entstammen Grundgesamtheiten mit gleichem Erwartungswert: $\mu_1 = \mu_2 = \dots = \mu_b$.
- $H_{0A \times B}$: keine Wechselwirkung.
Die Erwartungswerte der Kombinationen in den Faktorstufen lassen sich additiv aus den Haupteffekten ableiten: $\mu_{ij} = \mu_i + \mu_j - \mu$

Für die Prüfung dieser Hypothesen werden unter der Annahme der Normalverteilung und homogener Varianzen entsprechend der einfachen Varianzanalyse als Teststatistik ($\hat{F}$) die Quotienten der gemittelten Summe der Abweichungsquadrate - die Mittleren Quadrat-Summen (MS) - mit den Quantilen der F-Verteilung verglichen. Das allgemeine Schema zur **Varianzanalyse mit festen Effekten** ist in Tabelle 7.55 zusammengefasst.

Tabelle 7.55. Schema der zweifachen Varianzanalyse mit festen Effekten (balancierte Daten)

Faktor	SAQ	FG	MS	Test $\hat{F}$
A	$SAQ_A = \frac{1}{bn} \sum_{i=1}^a x_{i..}^2 - \frac{x_{...}^2}{abn}$	$a - 1$	$\frac{SAQ_A}{a - 1}$	$\frac{MS_A}{MS_{in}}$
B	$SAQ_B = \frac{1}{an} \sum_{j=1}^b x_{.j.}^2 - \frac{x_{...}^2}{abn}$	$b - 1$	$\frac{SAQ_B}{b - 1}$	$\frac{MS_B}{MS_{in}}$
$A \times B$	$SAQ_{A \times B} = SAQ_{ges} - SAQ_A - SAQ_B - SAQ_{in}$	$(a - 1)(b - 1)$	$\frac{SAQ_{A \times B}}{(a - 1)(b - 1)}$	$\frac{MS_{A \times B}}{MS_{in}}$
Rest	$SAQ_{in} = \sum_{ijk} x_{ijk}^2 - \frac{1}{n} \sum_{ij} x_{ij.}^2$	$ab(n - 1)$	$\frac{SAQ_{in}}{ab(n - 1)}$	
Gesamt	$SAQ_{ges} = \sum_{ijk} x_{ijk}^2 - \frac{x_{...}^2}{abn}$	$abn - 1$	$\frac{SAQ_{ges}}{abn - 1}$	

Beispiel: Die Wirksamkeit eines Antidepressivums soll in einer Studie geprüft werden. Zusätzlich soll dabei der Einfluss des Geschlechtes berücksichtigt werden (aus [Bor99]). Eine balancierte Untersuchung mit je $n = 5$ Beobachtungen pro Gruppe zeigte folgendes Ergebnis.



Geschlecht A	Therapie (B)		
	Plazebo	einfache Dosierung	doppelte Dosierung
männlich	22	16	13
	25	16	12
	22	16	12
	21	15	13
	22	15	12
weiblich	18	19	16
	19	20	14
	17	17	16
	21	16	13
	19	16	14

Die Berechnung des Beispiels soll in **R** mit der Funktion **aov()** gezeigt werden. Die einzelnen Teilsummen aus Tabelle 7.55 können zur Übung auch elementar berechnet werden.

```
> depr <- data.frame(
+   score = c(22, 25, 22, 21, 22, 16, 16, 16, 15, 15, 13, 12, 12, 13, 12,
+   18, 19, 17, 21, 19, 19, 20, 17, 16, 16, 16, 14, 16, 13, 14),
+   geschl = factor(c(rep("Mann", 15), rep("Frau", 15))),
+   therap = factor(rep(c(rep("Plazebo", 5), rep("einfach", 5), rep("doppelt", 5)), 2)))
+   score geschl therap
1      22      Mann Plazebo
2      25      Mann Plazebo
3      22      Mann Plazebo
4      21      Mann Plazebo ...

> summary(aov(score ~ therap + geschl + geschl:therap, depr))
            Df Sum Sq Mean Sq F value    Pr(>F)
therap      2  253.4    126.7   74.5294 5.061e-11 ***
geschl      1    0.3      0.3   0.1765   0.6782
therap:geschl 2   54.2    27.1  15.9412 3.938e-05 ***
Residuals  24   40.8      1.7
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Die Daten (Punktwerte) aus dem Beispiel werden zusammen mit den beiden Faktoren Therapie (A) und Geschlecht (B) in einem Datenrahmen gespeichert. Die Berechnung der zweifachen Varianzanalyse mit festen Effekten (Modell I) und balancierten Daten erfolgt dann mit der Funktion `aov()`. In der Ergebnistabelle werden die einzelne Varianzkomponenten $SAQ_A = 253,4$, $SAQ_B = 0,3$, $SAQ_{A \times B} = 54,2$ und $SAQ_{in} = 40,8$ angegeben. Die Berechnung der Teststatistik ($\hat{F}$ unter F value) und die Angabe entsprechender P-Werte erfolgt hier für die Hypothesen zu dem Modell mit festen Effekten (Typ I). Das Ergebnis zeigt einen signifikanten Haupteffekt hinsichtlich der Therapie ($\hat{F}_A = 74,53$) und eine signifikante Wechselwirkung aus der Therapie und dem Geschlecht ($\hat{F}_{A \times B} = 15,94$), während sich aus den Daten kein signifikanter Haupteffekt des Geschlechtes ableiten lässt.

Bei der Interpretation der Ergebnisse aus einer zweifachen Varianzanalyse bereitet der Wechselwirkungseffekt häufig Schwierigkeiten. In Abbildung 7.17 sind daher einige ausgewählte Kombinationen hinsichtlich der Einflussnahme zweier Faktoren auf eine abhängige Größe dargestellt.

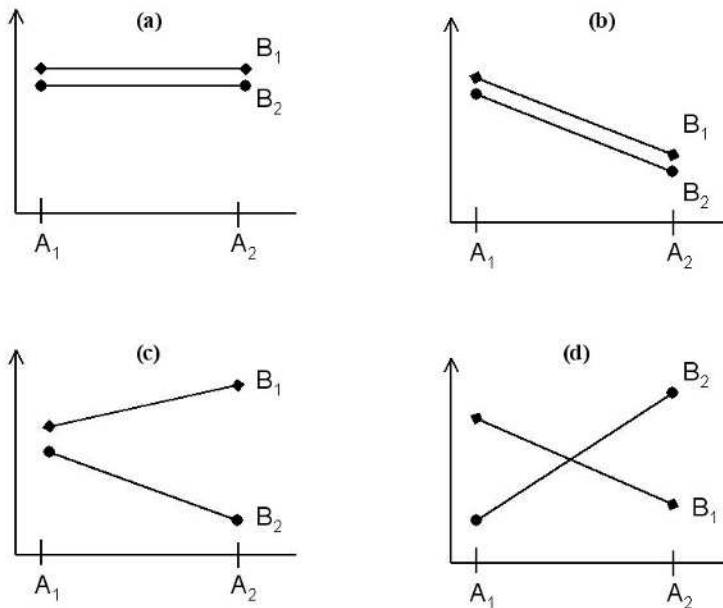


Abb. 7.17. Unterschiedlicher Einfluss zweier Faktoren im Rahmen der zweifachen Varianzanalyse unter Berücksichtigung einer Wechselwirkung. (a) Kein Effekt zu Faktor A , kein (oder nur ein geringer) Effekt zu Faktor B , keine Wechselwirkung von $A \times B$. (b) Großer Effekt von A , kein (oder nur geringer) Effekt von B , keine Wechselwirkung von $A \times B$. (c) Kein Effekt von A , großer Effekt von B und starke Wechselwirkung von $A \times B$. (d) Kein Effekt von A , kein Effekt von B , aber starke Wechselwirkung von $A \times B$

Speziell für die Daten aus dem Beispiel unter Berücksichtigung des Geschlechtes kann das Ergebnis der zweifachen Varianzanalyse in **R** auch mit Hilfe der Funktion `interaction.plot()` veranschaulicht werden (Abbildung 7.18, die an Abb. 7.17(d) erinnert).

Im Gegensatz zum Modell mit festen Effekten (Modell I) spricht man von einem **Modell mit zufälligen Effekten** oder Zufallskomponenten (Modell II, random effect model), wenn sich die Faktorstufen aus zufälligen Stichproben einer Grundgesamtheit ergeben. Hier interessieren ausschließlich die Variabilitätsanteile der einzelnen Faktoren an der Gesamtvariabilität. Die entsprechenden Hypothesen beziehen sich somit auf die Varianzen:

- H_{0A} : kein unterschiedlicher Einfluss des Faktors A .
Der Varianzanteil von Faktor A ist Null: $\sigma_\alpha^2 = 0$.

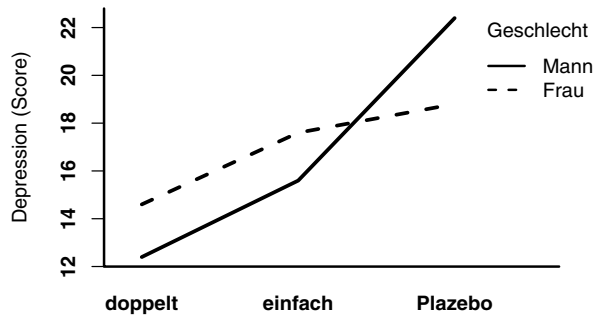


Abb. 7.18. Darstellung der Wechselwirkung zwischen Geschlecht und Therapie auf den Grad der Depression aus dem Beispiel

- H_{0B} : kein unterschiedlicher Einfluss des Faktors B .
Der Varianzanteil von Faktor B ist Null: $\sigma_\beta^2 = 0$.
- $H_{0A \times B}$: keine Wechselwirkung.
Der Varianzanteil aus der Wechselwirkung $A \times B$ ist Null: $\sigma_{\alpha\beta}^2 = 0$.

Tabelle 7.56. Erwartungswerte der gemittelten Summen aus der zweifachen Varianzanalyse unter verschiedenen Modellannahmen

Faktor	Modell I feste Effekte	Modell II zufällige Effekte	Modell III	
			A fest, B zufällig	A zufällig, B fest
A	$\sigma_e^2 + nb\sigma_\alpha^2$	$\sigma_e^2 + n\sigma_{\alpha\beta}^2 + nb\sigma_\alpha^2$	$\sigma_e^2 + n\sigma_{\alpha\beta}^2 + nb\sigma_\alpha^2$	$\sigma_e^2 + n\sigma_\alpha^2$
B	$\sigma_e^2 + na\sigma_\beta^2$	$\sigma_e^2 + n\sigma_{\alpha\beta}^2 + na\sigma_\beta^2$	$\sigma_e^2 + na\sigma_\beta^2$	$\sigma_e^2 + n\sigma_{\alpha\beta}^2 + na\sigma_\beta^2$
$A \times B$	$\sigma_e^2 + n\sigma_{\alpha\beta}^2$	$\sigma_e^2 + n\sigma_{\alpha\beta}^2$	$\sigma_e^2 + n\sigma_{\alpha\beta}^2$	$\sigma_e^2 + n\sigma_{\alpha\beta}^2$
Rest (e-error)	σ_e^2	σ_e^2	σ_e^2	σ_e^2

Die Analyse geht auch hier von einer Zerlegung der Summe der Abweichungsquadrate aus. Die Rechnung erfolgt weitgehend analog zum Modell mit festen Effekten. Allerdings dürfen bei diesem Modellansatz die Teststatistiken aus Tabelle 7.55 nicht übernommen werden, da sich die Erwartungswerte der gemittelten Summen (MS) bei den verschiedenen Modellannahmen unterscheiden. In der Tabelle 7.56 sind daher die Erwartungswerte für die gemittelten Summen für die drei verschiedenen Modellansätze (einschließlich Modell III für gemischte Effekte, mixed effects) zusammenfassend dargestellt.

Für die Herleitung geeigneter Teststatistiken werden die Varianzquotienten dann so gebildet, dass im Zähler genau ein zusätzlicher Term steht, welcher unter der Nullhypothese verschwindet (zu Null wird), z.B. für einen Haupteffekt von A in einem Modell mit zufälligen Effekten:

$$H_{0A} \quad (\text{kein Effekt von } A): \quad E(\hat{F}) = \frac{\sigma_e^2 + n\sigma_{\alpha\beta}^2 + nb\sigma_\alpha^2}{\sigma_e^2 + n\sigma_{\alpha\beta}^2} = 1$$

Tabelle 7.57 enthält eine Zusammenstellung aller Teststatistiken für die zweifache Varianzanalyse mit festen und zufälligen Effekten bei balancierten Daten. In der Regel wird in den Statistikprogrammen das Modell mit festen Effekten vollständig berechnet. Für die anderen Situationen können die Teststatistiken nach dieser Tabelle einfach berechnet werden.

Tabelle 7.57. Berechnung der Teststatistiken zur zweifachen Varianzanalyse für verschiedene Modelle nach festen, zufälligen und gemischten Effekten

Faktor	Modell I	Modell II	Modell III	
	feste Effekte	zufällige Effekte	A fest, B zufällig	A zufällig, B fest
A	$\hat{F} = \frac{MS_A}{MS_{in}}$	$\hat{F} = \frac{MS_A}{MS_{A \times B}}$	$\hat{F} = \frac{MS_A}{MS_{A \times B}}$	$\hat{F} = \frac{MS_A}{MS_{in}}$
B	$\hat{F} = \frac{MS_B}{MS_{in}}$	$\hat{F} = \frac{MS_B}{MS_{A \times B}}$	$\hat{F} = \frac{MS_B}{MS_{in}}$	$\hat{F} = \frac{MS_B}{MS_{A \times B}}$
A × B	$\hat{F} = \frac{MS_{A \times B}}{MS_{in}}$	$\hat{F} = \frac{MS_{A \times B}}{MS_{in}}$	$\hat{F} = \frac{MS_{A \times B}}{MS_{in}}$	$\hat{F} = \frac{MS_{A \times B}}{MS_{in}}$

7.5.8 Prinzipien der Versuchsplanung

Bei Experimenten wird im Gegensatz zur beobachtenden Studie ein wesentliches Detail kontrolliert: die Zuordnung der unterschiedlichen Untersuchungseinheiten zu den unterschiedlichen Behandlungen oder Zuständen.

Unverzerrte Beobachtungen und systematisches Experimentieren - d.h. Schätzwerte treffen bei kleinem Zufallsfehler im Mittel den Parameterwert - gestatten es, Fakten zu repräsentieren, Situationen und Prozesse zu beschreiben und mitunter neue Erkenntnisse und fundamentale Gesetzmäßigkeiten zu erkennen.

In einem Experiment wird an Untersuchungseinheiten mindestens eine **Einflussgröße** auf einem bestimmten Niveau, einer bestimmten Stufe vorgegeben und ihr Einfluss, ihre Wirkung auf eine **Zielgröße** bestimmt. Hierbei sollte rechtzeitig die Verallgemeinerungsfähigkeit der Resultate bedacht werden.

Die gewählten Einflussgrößen werden **Faktoren** genannt, die von ihnen angenommenen Werte werden **Stufen** genannt. Für sämtliche Faktoren werden mindestens zwei Stufen festgelegt. Die Zielgröße, das Resultat des Experimentes ist meistens ein Messwert, der z.B. die optimal Ausbeute eines Produktes oder Prozesses charakterisiert.

Gut zu bedenken ist die Wahl von z.B. zwei Faktorstufen, häufig Faktor vorhanden und Faktor nicht vorhanden, bzw. die Festlegung aufschlussreicher nicht zu extremer unterer und oberer Werte, die auch nicht zu eng beieinander liegen sollten, da sonst kaum ein unterschiedlicher Effekt bei vertretbarem Stichprobenumfang zu erwarten ist.

Nicht zu berücksichtigende Einflussgrößen werden im Experiment möglichst konstant gehalten, um die Zufallsstreuung zu minimieren, so dass auch kleinere Effekte der Faktoren sicher erfasst werden. Hierzu dienen **Blockbildung** und **Randomisierung**.

Blockbildung

Sind die Versuchseinheiten sehr unterschiedlich, dann wird die Isolierung interessierender Effekte durch die Heterogenität des Materials erschwert. In diesen Fällen ist vor der unterschiedlichen Behandlung der Versuchseinheiten zur Schaffung konstanter Bedingungen eine Zusammenfassung möglichst ähnlicher Versuchseinheiten zu empfehlen (vgl. Abb. 7.19). Man bildet Untergruppen von Versuchseinheiten, die in sich gleichförmiger sind als das gesamte Material: homogene „Versuchsblöcke“. Innerhalb eines Blocks gilt dann für die Zuordnung der Behandlungen zu den Versuchseinheiten wieder das Randomisierungsprinzip.

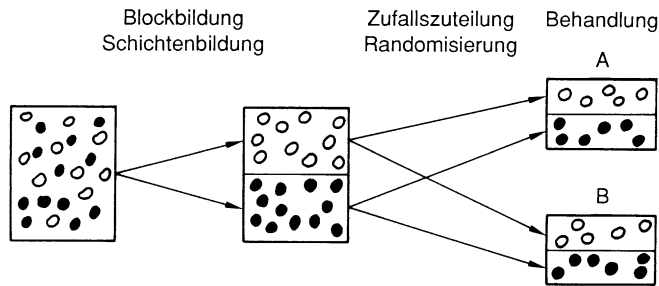


Abb. 7.19. Zur Versuchsplanung für den Vergleich zweier Behandlungen. Deutlich sich unterscheidende Untersuchungseinheiten werden durch lokale Kontrolle (Block- bzw. Schichtenbildung) getrennt erfasst und nach Zufallszuteilung (Randomisierung) zwei zu vergleichenden Einflüssen, Behandlungen (A und B), ausgesetzt. Durch lokale Kontrolle und möglichst zahlreiche Untersuchungseinheiten wird der zufällige Fehler verringert, durch Zufallszuteilung wird ein systematischer Fehler vermieden

Beispiele für Blöcke sind Versuche an demselben Patienten oder Probanden, an eineiigen Zwillingen oder an paarigen Organen oder bei Wurfgeschwistern oder an Blättern derselben Pflanze, oder die mit derselben Stammlösung durchgeführten Versuche, die nebeneinander liegenden Parzellen eines Feldes in einem landwirtschaftlichen Versuch oder andere versuchstechnische Gruppierungen, die natürliche oder künstliche Blöcke darstellen. Man vereinigt diejenigen Versuche zu einem Block, die in einem besonders wichtigen Variationsfaktor übereinstimmen. Von Block zu Block, also (zwischen den Blöcken) bestehen erhebliche Unterschiede in gerade diesem Faktor. Die einzelnen Blöcke sollten stets gleichen Umfang aufweisen. Die für das Versuchsziel wichtigen Vergleiche müssen möglichst innerhalb der Blöcke vorgenommen werden.

Randomisierung

Die Idee der zufälligen Zuordnung der Verfahren zu den Versuchseinheiten, kurz **RANDOMISIERUNG** genannt – sie stammt von R.A. Fisher –, kann als Grundlage jeder Versuchsplanung angesehen werden. Durch sie erhält man

- eine erwartungstreue Schätzung des interessierenden **Effektes**,
- eine erwartungstreue Schätzung des **Versuchsfehlers** und
- eine verbesserte **Normalität der Daten**.

Damit werden unerwünschte und unbekannte Korrelationssysteme zerstört, so dass wir unkorrelierte und **unabhängige Versuchsfehler** erhalten und unsere **Standard-Signifikanztests anwenden dürfen**.

Randomisierung ist die zufällige Zuordnung der Behandlungen zu den gegebenen Untersuchungseinheiten bei der Anlage von Versuchen. Sie soll die **Zufälligkeit der Stichprobe** im Sinne der Mathematischen Statistik garantieren, entsprechend der Auswahl einer Zufallsstichprobe. Diese zufällige Zuordnung wird z.B. mit **Tabellen von Zufallszahlen** vorgenommen. Unerwünschte Strukturen sind vor der Randomisierung zu formulieren und wenn sie auftreten zu verwerfen; dann ist erneut zu randomisieren. Mitunter muss eine weitere Einflussgröße berücksichtigt werden. Liegen die Untersuchungseinheiten in mehreren Blöcken vor, so werden die zu vergleichenden Behandlungen blockintern randomisiert, d. h. anhand von Zufallszahlen zugeordnet.

Randomisierung kann Vergleichbarkeit nicht erzwingen: die möglicherweise resultierende Nichtvergleichbarkeit ist dann aber zufallsbedingt. Wichtig ist, dass die Randomisierung im Verlauf der Studie nicht durch weitere Maßnahmen ungültig gemacht wird.

Randomisierte Blöcke

Randomisierte Blöcke ermöglichen blockinterne Vergleiche. Umfasst der Block 4 Elemente und sind die Behandlungen A und B zu vergleichen, so gibt es 6 Kombinationen von Behandlungszuordnungen: $AABB, BBAA, ABBA, BAAB, ABAB$ und $BABA$. Mit einem Würfel lässt sich eine bestimmte Anordnung auswählen, nach der die 4 Elemente zugeordnet werden. Bei 4 Behandlungen A, B, C und D ergeben sich bereits $4! = 4 \cdot 3 \cdot 2 = 24$ unterschiedliche Anordnungen. Jeweils 4 zweistellige Zufallszahlen bestimmen dann die Zuordnung. So legen z. B. 38, 93, 14, 72 die Reihenfolge C, A, D, B fest. Entsprechend werden die Zuordnungen der anderen Blöcke gewonnen. Ausgewertet werden dann paarige bzw. 4 verbundene Stichproben. Wird die Blockbildung bei der Auswertung nicht berücksichtigt, so ist die Power des Tests etwas geringer und das wahre Signifikanzniveau P kleiner als das berechnete.

Tabelle 7.58. Randomisierte Blöcke: Für k Behandlungen ordnet man den diese symbolisierenden Buchstaben die Zufallsziffern zu, so dass z. B. für $k = 3$ die Zufallsziffern 8 2 7 5 die Liste $C, A, C, B, \dots$ erzeugen

k	Buchst.	Zufallsziffer
2	A	0 - 4
	B	5 - 9
3	A	1 - 3
	B	4 - 6
	C	7 - 9
		0 wird ignoriert
4	A	1 - 2
	B	3 - 4
	C	5 - 6
	D	7 - 8
		0 und 9 werden ignoriert

Die Grundprinzipien der Versuchsplanung sind somit:

- 1. Wiederholung** (replication): Gewährt einen Einblick in die Streuung der Beobachtungen, gestattet die Schätzung des Versuchsfehlers, sorgt zugleich für seine Verkleinerung, gibt erste Hinweise auf die Verallgemeinerungsfähigkeit der Befunde und ermöglicht das Aufspüren systematischer Fehler der ursprünglichen Studie. Daher sind mehrere unabhängige Datensätze anzustreben.
- 2. Randomisierung** Zufallszuteilung (randomisation) sorgt für die:
 - Ausgewogenheit der Störgrößen in allen Behandlungsgruppen, d.h. Strukturgleichheit,
 - Ausschaltung systematischer Fehler, bei der Zuordnung der Behandlungen,
 - Möglichkeit, den Differenzen der Resultate zwischen den Gruppen gleichwirksamer Behandlungen eine Wahrscheinlichkeitsverteilung zuzuordnen.
- 3. Blockbildung** (block division, planned grouping): Erhöht die Genauigkeit blockinterner Vergleiche (vgl. Abb. 7.19).

Zusätzlich zu den drei Grundprinzipien der Versuchsplanung wird man

- Kontrollen mitlaufen lassen,
- möglichst unterschiedliche Behandlungen wählen, die zur Vermeidung subjektiver Einflüsse auch noch verschlüsselt werden, und
- die Zahl der Wiederholungen für $\sigma \neq \text{konst.}$ proportional aufteilen: $n_1/n_2 = \sigma_1/\sigma_2$.
Anderenfalls ist eine konstante Anzahl von Wiederholungen anzustreben (balanciertes De-

Tabelle 7.59. Die wichtigsten Versuchsanordnungen zur Prüfung von Unterschieden zwischen unterschiedlichen Stufen eines Faktors oder mehrerer Faktoren (verändert nach Juran, J.M. (Ed.): Quality Control Handbook, 2nd ed., New York 1962, Table 44, pp. 13–122/123)

Versuchsplan	Prinzip	Kommentar
1. Vollständige Randomisierung	Stufen eines Faktors werden nach einem Zufallsverfahren den experimentellen Einheiten zugeordnet	Zahl der Versuche kann von Stufe zu Stufe variieren; im Hinblick auf die Entdeckung signifikanter Effekte wenig empfindlich
2. Randomisierte Blöcke	Zusammenfassung möglichst ähnlicher experimenteller Einheiten zu Blöcken, denen jeweils die Stufen eines Faktors zugeordnet werden	Zahl der Versuche kann von Stufe zu Stufe variieren; empfindlicher als der vollständig randomisierte Plan
3. Lateinische Quadrate	Versuchsplan zur Prüfung von k Faktoren: aus k^2 Versuchseinheiten bestehend, die (nach zwei Merkmalen mit je k Stufen) so den Zeilen und Spalten eines Quadrates zugeordnet werden, dass jeder Faktor in jeder Zeile und jeder Spalte genau einmal auftritt	Gemeinsame Studien zweier oder mehrerer Faktoren! Vorausgesetzt wird, dass die Faktoren unabhängig voneinander wirken (keine Wechselwirkungen)
4. Faktorielle Experimente	Versuche mit beliebig vielen Faktoren, die jeweils auf beliebig vielen Stufen geprüft werden. Ein Experiment, das z.B. vier Faktoren jeweils auf 3 Stufen prüft, erfordert $3^4 = 84$ Versuchskombinationen	Exaktes Experiment; erfasst neben den Hauptfaktoren insbesondere auch alle Wechselwirkungen; werden alle Kombinationen von Faktoren und Stufen geprüft, dann kann das Experiment leicht zu unhandlich werden, außerdem erfordert es homogeneres Material als die anderen Pläne
5. Unvollständige faktorielle Experimente	Nur der zur Auswertung von Hauptfaktoren und wichtigen Wechselwirkungen notwendige Teil der gesamten Kombinationen eines faktoriellen Experimentes wird ausgewählt	Ökonomische Experimente Verglichen mit einem faktoriellen Experiment ist der Versuchsfehler größer und die Schätzung der Hauptfaktoren nicht so exakt; außerdem können einige mögliche Wechselwirkungen nicht berücksichtigt werden

sign), da dann die relative Größe des Standardfehlers der Differenz ($\sqrt{1/n_i + 1/n_j}$) für $n_i = n_j$ am kleinsten ist $[\sqrt{1/3 + 1/3} < \sqrt{1/2 + 1/4}]$.

Blindversuche sind bei Beurteilungen, etwa von Nahrungsmitteln wichtig, um nicht dem Image eines bestimmten Herstellers zu erliegen. Generell sollten bei Beurteilungen zur Ausschaltung von Autosuggestion und Suggestion seitens anderer Verschlüsselungen und Randomisierungen selbstverständlich sein!

Hinweise zu Versuchsplänen (Ergänzung zu Tabelle 7.59)

zu 1 + 2: Versuchsanordnung in Blöcken mit zufälliger Zuordnung der Verfahren zu den Versuchseinheiten.

Das Versuchsmaterial wird in möglichst homogene Blöcke aufgeteilt. Jeder Block enthält mindestens so viele Einheiten wie Faktoren (Behandlungsmethoden, Verfahren) geprüft werden sollen

(vollständige randomisierte Blöcke) bzw. ganze Vielfache dieser Zahl. Die Faktoren werden den untereinander ähnlichen Versuchseinheiten jedes Blockes mit Hilfe eines Zufallsverfahrens (Tafel der Zufallszahlen) zugeordnet. Durch Wiederholung des Versuchs mit sehr verschiedenen Blöcken wird der Vergleich zwischen den Faktoren genauer. Für die Varianzanalyse dieser verbundenen Stichproben wird das Modell der zweifachen Klassifikation ohne Wechselwirkung verwendet. Anstelle der Bezeichnungen Zeile und Spalte gelten jetzt „Block“ und „Faktor“.

Vielleicht sollten wir noch betonen, dass die Bildung von Blöcken genauso wie die Bildung paariger Beobachtungen nur dann sinnvoll ist, wenn die Streuung zwischen den Versuchseinheiten deutlich größer ist als die zwischen den Paarlingen bzw. den Blockeinheiten; denn verbundene Stichproben (paarige Beobachtungen, Blöcke) weisen weniger Freiheitsgrade auf als die entsprechenden unabhängigen Stichproben. Besteht ein deutlicher Streuungsunterschied im oben angegebenen Sinne, dann ist der Genauigkeitsgewinn durch Bildung verbundener Stichproben größer als der Genauigkeitsverlust durch die verringerte Anzahl von Freiheitsgraden.

Ist die Anzahl der Versuchseinheiten pro Block kleiner als die Anzahl der zu prüfenden Faktoren, dann spricht man von unvollständigen randomisierten Blöcken. Sie werden häufig benutzt, wenn eine natürliche Blockbildung nur wenige Elemente umfasst, z. B. bei Vergleichen an Zwillingspaaren, Rechts-Links-Vergleichen, bei technischen oder zeitlichen Beschränkungen der Durchführbarkeit von Parallelversuchen am gleichen Tag usw.

zu 3: Das Lateinische Quadrat.

Während durch die Blockbildung ein Variationsfaktor ausgeschaltet wird, dient der Versuchsplan eines sogenannten Lateinischen Quadrates zur Ausschaltung zweier Variationsfaktoren. So zeigt es sich häufig, dass ein Versuchsfeld deutlich nach zwei Richtungen Unterschiede in der Bodenbeschaffenheit aufweist. Durch geschicktes Parzellieren gelingt es mit Hilfe dieses Modells die Unterschiede nach zwei Richtungen auszuschalten. Sind k Faktoren (z. B. die Kunstdünger A und B und die Kontrolle C) zu prüfen, so benötigt man k Versuche und damit k^2 (9) Versuchseinheiten (Parzellen). Ein einfaches lateinisches Quadrat ist z. B.

A	B	C
B	C	A
C	A	B

Jeder Faktor tritt in jeder Zeile und jeder Spalte dieses Quadrates genau einmal auf. Im allgemeinen verwendet man nur Quadrate mit $k \geq 5$, da bei kleineren Quadraten für die Ermittlung des Versuchsfehlers nur wenige Freiheitsgrade zur Verfügung stehen. Erst bei $k = 5$ sind es 12. Entsprechende Versuchspläne, die natürlich nicht nur in der Landwirtschaft benutzt werden, sondern überall da, wo sich Versuchseinheiten nach zwei Richtungen oder Merkmalen randomisiert gruppieren lassen, findet man z. B. in dem Tafelwerk von Fisher und Yates [FY82]. Beim griechisch-lateinischen Quadrat erfolgt eine Randomisierung in drei Richtungen.

zu 4 + 5: Faktorielle Experimente

Sollen n Faktoren je auf 2, 3 oder k Stufen gleichzeitig verglichen werden, so benötigt man Versuchspläne mit **Kombinationsvergleichen**, sogenannte 2^n -, 3^n -, k^n -Pläne oder -Experimente. Ein faktorieller Versuchsplan ist ein Experiment zur Prüfung der Wirkung von mindestens 2 Prüffaktoren (Einflussgrößen) auf eine Zielgröße, wobei zugleich auch Wechselwirkungen erfasst werden. Sind z.B. 3 Einflussgrößen (A , B , C) auf jeweils zwei Stufen (fehlend bzw. vorhanden) zu prüfen, so liegt ein 2^3 -Plan vor. Die Untersuchungseinheiten werden randomisiert und zu gleichen Teilen auf die 8 Ansätze, z.B. Kontrolle und 7 Behandlungen, verteilt (vgl. Tabelle 7.60). Die Schätzungen der Hauptwirkungen (Haupteffekte) der 3 Faktoren und ihrer Wechselwirkungen ergeben sich dann aus dieser Tabelle (vgl. z.B. Montgomery [Mon05]).

Wechselwirkungen: Existiert eine Wechselwirkung AB , dann ist die Wirkung von A (bzw. B) unterschiedlich, je nachdem, ob B (bzw. A) vorliegt oder nicht. Entsprechend gilt für die Dreifach-

Tabelle 7.60. Schätzung der Parameter eines faktoriellen Versuchs vom Typ 2^3 : 8 Ansätze

	Ansätze:							
A:	-	+	-	-	+	+	-	+
B:	-	-	+	-	+	-	+	+
C:	-	-	-	+	-	+	+	+
Mittelwerte:	$\bar{x}$	$\bar{x}_A$	$\bar{x}_B$	$\bar{x}_C$	$\bar{x}_{AB}$	$\bar{x}_{AC}$	$\bar{x}_{BC}$	$\bar{x}_{ABC}$
A:	Hauptwirkung: $\frac{1}{4}[(\bar{x}_A - \bar{x}) + (\bar{x}_{AB} - \bar{x}_B) + (\bar{x}_{AC} - \bar{x}_C) + (\bar{x}_{ABC} - \bar{x}_{BC})]$							
AB:	Wechselwirkung: $\frac{1}{2}\{[(\bar{x}_A - \bar{x}) + (\bar{x}_{AB} - \bar{x}_B)] + [(\bar{x}_{AC} - \bar{x}_C)(\bar{x}_{ABC} - \bar{x}_{BC})]\}$							
ABC:	Wechselwirkung: $[(\bar{x}_A - \bar{x}) - (\bar{x}_{AB} - \bar{x}_B)] - [(\bar{x}_{AC} - \bar{x}_C) - (\bar{x}_{ABC} - \bar{x}_{BC})]$							

wechselwirkung ABC , dass dann, wenn sie existiert, die Wechselwirkung AB (bzw. AC , BC) unterschiedlich ist, je nachdem ob C (bzw. B , A) vorliegt oder nicht.

Schätzungen: Die Hauptwirkung ergibt sich aus dem Mittelwert aus 4 Schätzungen, die Wechselwirkung als Mittelwert aus zweien. Für die Dreifachwechselwirkung steht nur eine Schätzung zur Verfügung. Ist man somit an der Schätzung von Wechselwirkungen besonders interessiert, so muss die Zahl der Untersuchungseinheiten pro Ansatz erhöht werden.

Beispiel: In einem Feldversuch (2^3 -faktoriell) wird der Einfluss verschiedener Düngemittel (Ammoniumsulfat (a), Magnesiumsulfat (s) und Mist [10t/acre] (d), jeweils in zwei Stufen (ohne / mit einer fest bestimmten Konzentration), auf den Ertrag von Mangold untersucht. Dabei liegen 4 Wiederholungen in randomisierten Blöcken vor (Beispiel aus [SC82]). Die Beobachtungen (hier der Ertrag in t/acre) sind in der folgenden Tabelle zusammengestellt.



Block	a	s	d	1	2	3	4	Σ
-	0	0	0	19,2	15,5	17,0	11,7	63,4
a	1	0	0	20,6	16,9	19,5	21,9	78,9
s	0	1	0	18,9	20,2	16,7	20,7	76,5
as	1	1	0	25,3	27,6	29,1	25,4	107,4
d	0	0	1	20,8	18,5	20,1	19,2	78,6
ad	1	0	1	26,8	17,8	18,6	19,0	82,2
sd	0	1	1	22,2	18,6	22,3	21,1	84,2
asd	1	1	1	27,7	28,6	28,7	28,5	113,5
Σ				181,5	163,7	172,0	167,5	684,7

Die Zerlegung der Summe der Abweichungsquadrate in einer einfachen Varianzanalyse zeigt, dass von der Gesamtsumme $SAQ_{ges} = 612,7$ auf die unterschiedlichen Methoden der Düngung $SAQ_{block} = 484,2$ entfallen, somit nach der Teststatistik $\hat{F} = 12,9$ ($P < 0,001$) ein signifikanter Effekt besteht. Für die Frage, welches Düngemittel und welche Kombination von Düngemitteln zum besten Ertrag führt, ist eine weitere Aufteilung in der Summe der Abweichungsquadrate erforderlich.

Die Berechnung der Beispieldaten in **R** mit der Funktion **aov()** zeigt das Standardschema der Varianzanalyse, hier mit 3 Faktoren.

```

> a <- c(rep(0,4), rep(1,4), rep(0,4), rep(1,4), rep(0,4), rep(1,4), rep(0,4), rep(1,4))
> s <- c(rep(0,8), rep(1,8), rep(0,8), rep(1,8))
> d <- c(rep(0,16), rep(1,16))
> yield <- c(19.2, 15.5, 17.0, 11.7, 20.6, 16.9, 19.5, 21.9, 18.9, 20.2, 16.7,
+           20.7, 25.3, 27.6, 29.1, 25.4, 20.8, 18.5, 20.1, 19.2, 26.8, 17.8,
+           18.6, 19.0, 22.2, 18.6, 22.3, 21.1, 27.7, 28.6, 28.7, 28.5)
> data <- data.frame(block=gl(8,4), a=factor(a), s=factor(s), d=factor(d),
+                   yield=yield)
> yield.aov1 <- aov(yield ~ block, data)
> summary(yield.aov1)
      Df Sum Sq Mean Sq F value    Pr(>F)
block    7  484.18    69.17  12.917 8.913e-07 ***
Residuals 24  128.51     5.35

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> yield.aov2 <- aov(yield ~ a*s*d, data)
> summary(yield.aov2)
      Df Sum Sq Mean Sq F value    Pr(>F)
a         1  196.515   196.515  36.6997 2.955e-06 ***
s         1  192.570   192.570  35.9629 3.433e-06 ***
d         1   32.603    32.603   6.0886 0.021124 *
a:s        1   52.788    52.788   9.8582 0.004442 **
a:d         1    5.695     5.695   1.0636 0.312674
s:d         1    0.690     0.690   0.1289 0.722697
a:s:d        1    3.315     3.315   0.6191 0.439068
Residuals  24  128.512     5.355

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Für eine inhaltliche Interpretation dieses Ergebnisses ist die Aufteilung der Quadratsummen und der entsprechenden Anteile am Effekt (Ertrag) nach der folgenden Tabelle hilfreich:

Effekt	-	a	s	as	d	ad	sd	asd	Effekt	SAQ
	63,4	78,9	76,5	107,4	78,6	82,2	84,2	113,5	Summe	
A	-	+	-	+	-	+	-	+	***79,3	196,5
S	-	-	+	+	-	-	+	+	***78,5	192,6
A × S	+	-	-	+	+	-	-	+	**41,1	52,8
D	-	-	-	-	+	+	+	+	*32,3	32,6
A × D	+	-	+	-	-	+	-	+	-13,5	5,7
S × D	+	+	-	-	-	-	+	+	-4,7	0,7
A × S × D	-	+	+	-	+	-	-	+	10,3	3,3
Summe										484,2

Danach erhöht sich der Ertrag entsprechend der signifikanten Haupteffekte durch das Düngen mit Ammoniumsulfat *A* um $79,3/16 = 4,96$ T/acre, durch Magnesiumsulfat *S* um $78,5/16 = 4,90$ T/acre und durch die zusätzliche Gabe von Mist *D* um $32,3/16 = 2,02$ T/acre. Die in dem Beispiel auftretende signifikante positive Wechselwirkung $A \times S$ weist darauf hin, dass der Ertrag insbesondere durch die kombinierte Verwendung dieser beiden Kunstdünger gesteigert werden kann.

Generell gilt:

Um **Haupteffekte** zu erfassen, wählt man bei 2-4 Faktoren mit jeweils sorgfältig gewähltem niedrigem und hohem Niveau der Faktoren vollständige oder unvollständige faktorielle Pläne; bei mehr als 4 Faktoren kommen nur noch unvollständige faktorielle Pläne oder Plackett-Burman Ansätze infrage, 2^k -Pläne zur Erfassung der Hauptwirkungen.

Für die **Zielgrößenoptimierung** wählt man bei 2-4 Faktoren die sogenannten Box-Wilson oder Box-Behnken Ansätze. Bei mehr als 4 Faktoren erfasst man die 4 wichtigsten und wählt eines der beiden Verfahren.

Um einen Prozess störunanfälliger (robuster) zu machen, werden sowohl Haupteffekt -Erfassungs-Ansätze als auch Verfahren der Zielgrößenoptimierung eingesetzt.

Werden gute **Modellparameterschätzungen** angestrebt, so ist ein Regressionsansatz zu bevorzugen.

Näheres ist den Monographien zur Versuchsplanung mit vollständig dargestellten Fallstudien zu entnehmen, z.B. in Kleppmann [Kle01], ergänzt durch Cox und Reid [CR00], Giesbrecht und Gumpertz [GG04], Kuehl [Kue00], Montgomery [Mon05] sowie insbesondere durch Wu und Hamada [WH00] (vgl. auch Sachs [Sac86b]).

7.6 Die Analyse von Häufigkeiten

- Vergleich zweier relativer Häufigkeiten
- Die Analyse von Vierfeldertafeln
- Odds-Ratio und relatives Risiko
- Exakter Fisher-Test
- Der von McNemar modifizierte Vorzeichentest
- Der Mantel-Haenszel Test
- Der $k \cdot 2$ -Felder-Test nach Brandt und Snedecor
- Der Cochran-Armitage Test
- Die Analyse von Zweiwegtafeln des Typs $r \times c$
- Die Prüfung auf Trend
- Der Bowker Test
- Der Kappa-Koeffizient nach Cohen

Analyse kategorialer Daten

Ein kategoriales oder qualitatives Merkmal weist Ausprägungen auf, die in bestimmten, qualitativ verschiedenen Kategorien oder Klassen zusammengefasst werden. Neben dichotomen oder Alternativmerkmalen mit nur zwei Ausprägungen (z. B. einwandfrei – nicht einwandfrei) unterscheidet man nominal skalierte Merkmale, wobei Namen oder Bezeichnungen für bestimmte Kategorien existieren (z. B. ledig, verheiratet, geschieden) sowie ordinal skalierte Merkmale, bei denen die Kategorien aus Namen oder Bezeichnungen für Ränge, d. h. für Intensitätsstufen des Merkmals bestehen (z. B. die Benotungsskala in der Schule).

Werden die Ausprägungen zweier kategorialer Merkmale gemeinsam betrachtet, so liegt im einfachsten Fall eine **Vierfeldertafel** vor. Ziel der Analyse einer Vier- oder Mehrfeldertafel ist die Erfassung der Beziehungsstruktur der in der Tafel verknüpften kategorialen Merkmale.

Wird die stochastische Unabhängigkeit beider Merkmale geprüft, so liegt eine **Kontingenztafel** vor.

7.6.1 Vergleich zweier relativer Häufigkeiten

Bei nicht zu kleinen Stichprobenumfängen (mit $n\hat{p}$ sowie $n(1 - \hat{p}) > 5$) ist eine Approximation durch die Standardnormalverteilung möglich:

1. **Vergleich einer relativen Häufigkeit** $\hat{p}_1$ mit dem zugrundeliegenden Parameter π ohne (7.191) bzw. mit (7.192) Endlichkeitskorrektur.

$$\hat{z} = \frac{|\hat{p}_1 - \pi| - \frac{1}{2n}}{\sqrt{\frac{\pi(1-\pi)}{n}}} \quad (7.191)$$

$$\hat{z} = \frac{|\hat{p}_1 - \pi| - \frac{1}{2n}}{\sqrt{\left\{ \frac{\pi(1-\pi)}{n} \right\} \cdot \left\{ \frac{N-n}{N-1} \right\}}} \quad (7.192)$$

Nullhypothese: $\pi_1 = \pi$. Die Alternativhypothese lautet: $\pi_1 \neq \pi$ (oder bei einseitiger Fragestellung: $\pi_1 > \pi$ bzw. $\pi_1 < \pi$).

2. **Vergleich zweier relativer Häufigkeiten** $\hat{p}_1$ und $\hat{p}_2$ (Vergleich zweier Prozentsätze). Vorausgesetzt wird a) $n_1 \geq 50$, $n_2 \geq 50$, b) $n\hat{p} > 5$, $n(1-\hat{p}) > 5$ (vgl. auch Tabelle 7.61).

$$\hat{z} = \frac{|\hat{p}_1 - \hat{p}_2|}{\sqrt{\hat{p}(1-\hat{p})[(1/n_1) + (1/n_2)]}} \quad (7.193)$$

mit $\hat{p}_1 = x_1/n_1$, $\hat{p}_2 = x_2/n_2$, $\hat{p} = (x_1 + x_2)/(n_1 + n_2)$. Nullhypothese: $\pi_1 = \pi_2$; Alternativhypothese: $\pi_1 \neq \pi_2$ (bei einseitiger Fragestellung: $\pi_1 > \pi_2$ bzw. $\pi_1 < \pi_2$). So ergibt sich für $n_1 = n_2 = 300$, $\hat{p}_1 = 54/300 = 0,18$, $\hat{p}_2 = 30/300 = 0,10$ (beachte $n\hat{p}_2 = 300 \cdot 0,10 = 30 > 5$), $\hat{p} = (54 + 30)/(300 + 300) = 0,14$, $\hat{z} = (0,18 - 0,10)/\sqrt{0,14 \cdot 0,86(2/300)} = 2,82$, d. h. $P \approx 0,005$.

Man beachte, dass auch mit den Prozentzahlen gerechnet werden kann:

($\hat{z} = (18 - 10)/\sqrt{14 \cdot 86(2/300)} = 2,82$) und dass (für $n_1 = n_2 = n$) Differenzen größer oder gleich D (in %) nach Tabelle 7.61 auf dem 5%-Niveau bedeutsam sind.

Tabelle 7.61. Differenzen D der Prozentsätze für $n = n_1 = n_2 \geq 100$, die zumindest auf dem 5%-Niveau statistisch signifikant sind

n	100	150	200	300	500	1000	5000
D	14	11,5	10	8	6,3	4,5	2

Liegen die zu vergleichenden Prozentsätze unterhalb von 40% bzw. oberhalb von 60%, so gilt für diese D -Werte, dass ihnen ein wesentlich kleinerer P -Wert entspricht (im Beispiel für $n_1 = n_2 = 300$: $18\% - 10\% = 8\%$ mit $P \approx 0,005$).

Etwas exakter als (7.193) und nicht so strengen Voraussetzungen unterworfen ($n\hat{p}$ und $n(1-\hat{p}) \geq 1$ für n_1 und $n_2 \geq 25$) ist eine auf der Winkeltransformation basierende Approximation: $\hat{z} = (|\arcsin\sqrt{\hat{p}_1} - \arcsin\sqrt{\hat{p}_2}|)/28,648\sqrt{1/n_1 + 1/n_2}$; für das Beispiel ergibt sich $\hat{z} = (25,104 - 18,435)/28,648\sqrt{2/300} = 2,85$.

Für die Prüfung der Nullhypothese: $\pi_1 - \pi_2 = d_0$ gegen H_A : $\pi_1 - \pi_2 \neq d_0$ (bzw. $< d_0$ oder $> d_0$) verwende man ($\hat{p}_1 = x_1/n_1$, $\hat{p}_2 = x_2/n_2$, $\hat{q}_1 = 1 - \hat{p}_1$, $\hat{q}_2 = 1 - \hat{p}_2$):

$$\hat{z} = \frac{|(\hat{p}_1 - \hat{p}_2) - d_0|}{\sqrt{(\hat{p}_1 \hat{q}_1 / n_1) + (\hat{p}_2 \hat{q}_2 / n_2)}} \quad (7.194)$$

3. Vergleich zweier Häufigkeiten: die Prüfung, ob sie in einem bestimmten Verhältnis zueinander stehen. Die mitunter auftretende Frage, ob zwei beobachtete Häufigkeiten (a und b , wobei $a \leq b$ einem bestimmten Verhältnis $H_0: \beta/\alpha = \xi$ (griech. xi) entsprechen, wird mit der χ^2 -Verteilung entschieden –

für große Werte a und b ohne Kontinuitätskorrektur (7.195)

$$\hat{\chi}^2 = \frac{\{|\xi a - b| - (\xi + 1)/2\}^2}{\xi \cdot (a + b)} \quad \hat{\chi}^2 = \frac{(\xi a - b)^2}{\xi(a + b)} \quad (7.195a, 7.195)$$

– wobei ein Freiheitsgrad zur Verfügung steht. Ist das berechnete $\hat{\chi}^2$ kleiner oder gleich $\chi^2 = 3,841$, so lässt sich die Nullhypothese, die beobachteten Häufigkeiten entsprechen dem Verhältnis ξ , auf dem 5%-Niveau nicht ablehnen.

Beispiel: Entsprechen die Häufigkeiten $a = 6$ und $b = 25$ dem Verhältnis $\xi = \beta/\alpha = 5/1$ ($\alpha = 0,05$)?

$$\hat{\chi}^2 = \frac{\{|5 \cdot 6 - 25| - (5 + 1)/2\}^2}{5(6 + 25)} = 0,026 < 3,841.$$

Die Abweichung $25/6 = 4,17$ gegenüber 5,00 ist zufälliger Natur.

Lautet die zu prüfende Nullhypothese: Gleichheit zweier Häufigkeiten (Erfolgszahlen), so kann man für das Verhältnis beider: β/α , geschätzt durch b/a , ein Konfidenzintervall angeben. Schließt (7.196) den Wert 1 nicht mit ein, so wird H_0 abgelehnt.

$$\begin{array}{c} b/[(a + 1)F_1] \leq \beta/\alpha \leq [(b + 1)F_2]/a \\ \text{mit} \\ F_1 = F_{2a+2;2b;1-\alpha/2} \quad F_2 = F_{2b+2;2a;1-\alpha/2} \end{array} \quad (7.196)$$

Beispiel: Von 40 Kunden bevorzugen 25 den Typ B und 6 den Typ A eines Produktes; d. h. $b/a = 25/6 = 4,17$. Wir prüfen H_0 anhand des 95%-Konfidenzintervalles: $F_1 = F_{14;50;0,975} = 2,14$; $F_2 = F_{52;12;0,975} = 2,86$; $25/[7 \cdot 2,14] = 1,7$ $[26 \cdot 2,86]/6 = 12,4$; d. h. 95%-KI: $1,7 \leq \beta/\alpha \leq 12,4$, daß $H_0: \beta = \alpha$ erwartungsgemäß auf dem 5%-Niveau abgelehnt wird.

7.6.2 Die Analyse von Vierfeldertafeln

Beispiel: Besonders in der Medizin ist der Vergleich zweier aus Häufigkeiten ermittelter relativer Häufigkeiten wichtig (vgl. Tabelle 7.62). Es ist ein neues Heilmittel oder eine neue Operationsmethode entwickelt worden: Früher starben von 100 Patienten 15, nach der Umstellung aber von 81 Patienten nur 4 (vgl. Tabelle 7.63). Ist das neue Medikament oder die neue Operationstechnik erfolgversprechender oder liegt ein Zufallsbefund vor?

Tabelle 7.62. Vierfeldertafel für den Vergleich zweier Stichproben oder allgemeiner für den Vergleich zweier Alternativmerkmale

Merkmalspaar I	Merkmalspaar II		Summe
	Ereignis (+)	Komplement (-)	
1. Stichprobe	a	b	$a + b = n_1$
2. Stichprobe	c	d	$c + d = n_2$
Summe	$a + c$	$b + d$	$n_1 + n_2 = n$

Allgemein führt die Klassifizierung von $n_1 + n_2$ Objekten nach einem Merkmalspaar (+/−) (Tabelle 7.63) bzw. von n Objekten nach zwei Merkmalspaaren zu vier Klassen – den beobachteten Häufigkeiten a, b, c, d – und damit zu einer sogenannten Vierfeldertafel (Tabelle 7.62). Grenzfälle, die je zur Hälfte den beiden möglichen Klassen zugeordnet werden, können zu halbzahligen Werten führen. Die beiden Stichproben von Alternativdaten werden daraufhin untersucht, ob sie als Zufallsstichproben aus einer durch die Randsummen repräsentierten Grundgesamtheit aufgefasst werden können, d. h. ob die 4 Besetzungszahlen z. B. von Tabelle 7.63 sich **proportional zu den Randsummen** verteilen und Abweichungen der Verhältnisse a/n_1 und c/n_2 von dem Verhältnis $(a + c)/n$ (entsprechend der Homogenität: $a/n_1 = c/n_2 = (a + c)/n$) als Zufallsabweichungen auffassbar sind.

Das oben angedeutete Beispiel führt zum Vierfelderschema (Tabelle 7.63) mit der Fragestellung: Beruht die für die neue Behandlung ermittelte niedrigere relative Häufigkeit von Todesfällen auf einem Zufall? Die Nullhypothese lautet: **Der Heilungsprozentsatz ist stochastisch unabhängig von der angewandten Therapie**. Oder: Beide Stichproben, die Gruppe der konventionell behandelten Patienten und die mit der neuen Therapie behandelte Patientengruppe, stammen bezüglich des Therapie-Effektes aus einer gemeinsamen Grundgesamtheit, d. h. **der Therapie-Effekt ist bei beiden Behandlungen der gleiche**.

Tabelle 7.63. Vierfeldertafel zum Beispiel des Therapieerfolges

Behandlung	Patienten		Summe
	gestorben	geheilt	
übliche Therapie	15	85	100
neue Therapie	4	77	81
Summe	19	162	181

Die beiden Behandlungsgruppen sind eigentlich Stichproben zweier Binomialverteilungen. Verglichen werden somit die Grundwahrscheinlichkeiten von Binomialverteilungen, d. h. etwa (vgl. Tab. 7.63) anhand der Anteile Geheilte in beiden Therapiegruppen $85/100 = 0,85$ und $77/81 = 0,95$ sowie insgesamt $162/181 = 0,895$.

Nullhypothese $[H_0]$: Beide Stichproben entstammen einer gemeinsamen Grundgesamtheit mit der Erfolgswahrscheinlichkeit π .
 Alternativhypothese: Beide Stichproben entstammen zwei verschiedenen Grundgesamtheiten mit den Erfolgswahrscheinlichkeiten π_1 und π_2 .
 $[H_A]$

Die Nullhypothese auf Gleichheit oder Homogenität beider Parameter (π_1, π_2) [oder auf stochastische Unabhängigkeit beider Merkmalsalternativen] wird anhand des χ^2 -Tests nicht abgelehnt oder abgelehnt. Dazu ist grundsätzlich die folgende Frage zu klären:

Verteilen sich die Felderhäufigkeiten proportional zu den Randsummen? Um dies zu entscheiden, bestimmen wir die unter dieser Annahme zu erwartenden Häufigkeiten, kurz **Erwartungshäufigkeiten** E genannt. Wir multiplizieren die Zeilensumme mit der Spaltensumme des Feldes a ($100 \cdot 19 = 1900$) und dividieren das Produkt durch den Umfang n der vereinten Stichproben $1900/181 = 10,497$; $E_a = 10,50$). Entsprechend verfahren wir mit den übrigen Feldern und erhalten: $E_b = 89,50$, $E_c = 8,50$, $E_d = 72,50$.

Zur Beurteilung, ob die beobachteten Werte, a, b, c, d mit den erwarteten Werten E_a, E_b, E_c, E_d im Sinne der Nullhypothese übereinstimmen, bilden wir die Prüfgröße $\hat{\chi}^2$ (vgl. auch den χ^2 -Anpassungstest):

$$\hat{\chi}^2 = \frac{(a - E_a)^2}{E_a} + \frac{(b - E_b)^2}{E_b} + \frac{(c - E_c)^2}{E_c} + \frac{(d - E_d)^2}{E_d}$$

und erhalten hieraus nach einigen Umformungen (7.197):

$$\hat{\chi}^2 = \Delta^2 \left(\frac{1}{E_a} + \frac{1}{E_b} + \frac{1}{E_c} + \frac{1}{E_d} \right) \quad \text{mit} \quad (7.197)$$

$$|\Delta| = |a - E_a| = |b - E_b| = |c - E_c| = |d - E_d| \quad \text{und}$$

$$n = a + b + c + d$$

oder in kurzer Form auch (7.198):

$$\hat{\chi}^2 = \frac{n(ad - bc)^2}{(a + b)(c + d)(a + c)(b + d)} \quad (7.198)$$

Das Vierfelder- $\hat{\chi}^2$ besitzt nur **einen Freiheitsgrad**, da bei gegebenen Randsummen nur eine der 4 Häufigkeiten frei gewählt werden kann: H_0 wird abgelehnt, falls $\hat{\chi}^2 > \chi^2_{\nu=1;1-\alpha} = \chi^2_{1-\alpha}$.

Für kleines n ist n in (7.198) durch $(n - 1)$ zu ersetzen: $\hat{\chi}^2$ wird dann $\hat{\chi}^2_*$ (Chiquadrat-Stern)

$$\hat{\chi}^2_* = \frac{(n - 1)(ad - bc)^2}{(a + b)(c + d)(a + c)(b + d)} \quad (7.199)$$

genannt und nach (7.199) berechnet. Diese Formel ist generell anwendbar, sobald $n_1 \geq n_2 \geq 6$; günstig ist es, wenn dann auch gilt: $n_1 \approx n_2$ bzw. $n_2 \gtrsim \sqrt{n_1}$ für $n_1 > n_2$ (Sachs [Sac86a]). Bei noch kleinerem n prüfe man die Nullhypothese mit dem **exakten Fisher-Test** [7.6.4].

Beispiele:

$$\hat{\chi}^2_* = (12 - 1)(1 \cdot 1 - 5 \cdot 5)^2 / 6^4 = 4,89 > 3,84 = \chi^2_{0,95};$$

1	5	6
5	1	6
6	6	12

1	5
4	2
12	

$$\hat{\chi}^2_* = 2,83 < 3,84.$$

$\hat{\chi}^2$ -Variationsbereich: Für eine Vierfelder-Tabelle gilt, wenn das entsprechende $\hat{\chi}^2$ nach (7.199) berechnet worden ist:

$$0 \leq \hat{\chi}^2_* \leq (n - 1) \quad (7.200)$$

So ergibt sich als ideale Besetzungszahl-Anordnung für die kleinsten mit der „ $(n - 1)$ “-Formel berechenbaren Stichprobenumfänge:

6 0	6	$\hat{\chi}^2_* = \frac{(12 - 1)(6 \cdot 6 - 0 \cdot 0)^2}{6 \cdot 6 \cdot 6 \cdot 6} = 11$	Das entsprechende P ist Tabelle 7.66 zu entnehmen.
0 6	6		
6 6	12		

Für $n_1 = n_2$ gehen (7.198) und (7.199) über in:

$$\hat{\chi}^2 = \frac{n(a - c)^2}{(a + c)(b + d)} \quad \text{bzw. für} \quad \hat{\chi}^2_* = \frac{(n - 1)(a - c)^2}{(a + c)(b + d)} \quad \text{kleines } n: \quad (7.201)$$

Die Nullhypothese auf **Homogenität oder stochastische Unabhängigkeit** wird abgelehnt, sobald das nach (7.197) bis (7.201) berechnete $\hat{\chi}^2$ größer ist als der Tabellenwert $\chi^2_{1;1-\alpha}$ (vgl. Tabelle 7.64 bis 7.66).

Im allgemeinen wird der zweiseitige Test angewandt. Tabelle 7.65 gibt exakte Wahrscheinlichkeiten für $\chi^2 = 0,0$ (0,1) 10,0. Sie wird ergänzt durch Tabelle 7.66 (Kohnen und Mitarbeiter [KLB87]).

Tabelle 7.64. Schranken ($\chi^2_{1;1-\alpha}$) für den Vierfelder- χ^2 -Test auf Homogenität (Hypothesenpaare angegeben) und auf stochastische Unabhängigkeit

Irrtumswahrscheinlichkeit α	0,10	0,05	0,01	0,001
Zweiseitiger Test ($H_0: \pi_1 = \pi_2, H_A: \pi_1 \neq \pi_2$)	2,706	3,841	6,635	10,828
Einseitiger Test ($H_0: \pi_1 = \pi_2, H_A: \pi_1 > \pi_2$ od. $\pi_1 < \pi_2$)	1,642	2,706	5,412	9,550



Beispiel: Wir prüfen die Daten aus dem einführenden Beispiel (Tabelle 7.63) auf dem 5%-Niveau (einseitiger Test, Voraussetzung: neue Therapie nicht schlechter!) anhand von (7.198) und (7.199).

$$\hat{\chi}^2 = \frac{181(15 \cdot 77 - 4 \cdot 85)^2}{100 \cdot 81 \cdot 19 \cdot 162} = 4,822 \quad \text{bzw.} \quad \hat{\chi}_*^2 = 180 \cdot 0,0266417 = 4,7955$$

Da $\hat{\chi}^2 = 4,796 > 2,706 = \chi^2_{1;0,95}$ (Tabelle 7.64), wird die Homogenitätshypothese (Unabhängigkeitshypothese) anhand der vorliegenden Daten auf dem 5%-Signifikanzniveau abgelehnt. Die neue Therapie ist besser. Die Therapie-Unterschiede sind auf dem 5%-Niveau statistisch signifikant. Zwischen der neuen Behandlung und dem Absinken der Sterblichkeit besteht auf dem 5%-Niveau ein statistisch signifikanter Zusammenhang.

Tabelle 7.65. χ^2 -Tabelle für einen Freiheitsgrad (auszugsweise entnommen aus Kendall, M.G. and A. Stuart: The Advanced Theory of Statistics, Vol. II, Griffin, London 1961, pp. 629 and 630): zweiseitige Wahrscheinlichkeiten

χ^2	P	χ^2	P	χ^2	P	χ^2	P	χ^2	P
0	1,00000	2,1	0,14730	4,0	0,04550	6,0	0,01431	8,0	0,00468
0,1	0,75183	2,2	0,13801	4,1	0,04288	6,1	0,01352	8,1	0,00443
0,2	0,65472	2,3	0,12937	4,2	0,04042	6,2	0,01278	8,2	0,00419
0,3	0,58388	2,4	0,12134	4,3	0,03811	6,3	0,01207	8,3	0,00396
0,4	0,52709	2,5	0,11385	4,4	0,03594	6,4	0,01141	8,4	0,00375
0,5	0,47950	2,6	0,10686	4,5	0,03389	6,5	0,01079	8,5	0,00355
0,6	0,43858	2,7	0,10035	4,6	0,03197	6,6	0,01020	8,6	0,00336
0,7	0,40278	2,8	0,09426	4,7	0,03016	6,7	0,00964	8,7	0,00318
0,8	0,37109	2,9	0,08858	4,8	0,02846	6,8	0,00912	8,8	0,00301
0,9	0,34278	3,0	0,08326	4,9	0,02686	6,9	0,00862	8,9	0,00285
1,0	0,31731	3,1	0,07829	5,0	0,02535	7,0	0,00815	9,0	0,00270
1,1	0,29427	3,2	0,07364	5,1	0,02393	7,1	0,00771	9,1	0,00256
1,2	0,27332	3,3	0,06928	5,2	0,02259	7,2	0,00729	9,2	0,00242
1,3	0,25421	3,4	0,06520	5,3	0,02133	7,3	0,00690	9,3	0,00229
1,4	0,23672	3,5	0,06137	5,4	0,02014	7,4	0,00652	9,4	0,00217
1,5	0,22067	3,6	0,05778	5,5	0,01902	7,5	0,00617	9,5	0,00205
1,6	0,20590	3,7	0,05441	5,6	0,01796	7,6	0,00584	9,6	0,00195
1,7	0,19229	3,8	0,05125	5,7	0,01697	7,7	0,00552	9,7	0,00184
1,8	0,17971	3,9	0,04829	5,8	0,01603	7,8	0,00522	9,8	0,00174
1,9	0,16808	4,0	0,04550	5,9	0,01514	7,9	0,00494	9,9	0,00165
2,0	0,15730			6,0	0,01431	8,0	0,00468	10,0	0,00157

Hinweise: 1. Bei Vorversuchen ohne vorher spezifizierte Irrtumswahrscheinlichkeiten vergleiche man den gefundenen $\hat{\chi}^2$ -Wert mit den in Tabelle 7.65 tabellierten (zweiseitige Fragestellung). Tabelle 7.66 ergänzt Tabelle 7.65 und bietet auch einseitige Schranken.

2. Wird beachtet, dass sich der Zahlenwert des Quotienten (7.198) nicht ändert, wenn man die vier inneren Feldhäufigkeiten (a, b, c, d) und die vier Randhäufigkeiten ($a + b, c + d, a + c, b + d$) durch eine Konstante k dividiert (der Stichprobenumfang n darf nicht durch k dividiert werden), so lässt sich die Rechenarbeit merklich verringern. Für eine überschlagweise Berechnung von $\hat{\chi}^2$ kann man außerdem die durch k dividierten Häufigkeiten noch runden.

Tabelle 7.66. χ^2 -Tabelle für einen Freiheitsgrad: ein- und zweiseitige Schranken. Einige Werte aus Kohnen, Lotz und Busch (1987)

P	χ^2	
	einseitig	zweiseitig
0,0001	13,83	15,14
0,0002	12,53	13,83
0,0003	11,78	13,07
0,0004	11,24	12,53
0,0005	10,83	12,12
0,0006	10,49	11,78
0,0007	10,21	11,49
0,0008	9,96	11,24
0,0009	9,74	11,02
0,0010	9,55	10,83

P	χ^2	
	einseitig	zweiseitig
0,001	9,55	10,83
0,002	8,28	9,55
0,003	7,55	8,81
0,004	7,03	8,28
0,005	6,63	7,88
0,006	6,31	7,55
0,007	6,04	7,27
0,008	5,80	7,03
0,009	5,60	6,82
0,010	5,41	6,63

P	χ^2	
	einseitig	zweiseitig
0,01	5,41	6,63
0,02	4,22	5,41
0,03	3,54	4,71
0,04	3,06	4,22
0,05	2,71	3,84
0,06	2,42	3,54
0,07	2,18	3,28
0,08	1,97	3,06
0,09	1,80	2,87
0,10	1,64	2,71

Für großes n wird die Rechnung nach (7.198) jedoch umständlich, man bevorzuge Formel (7.193).

3. Da der Vierfelder- χ^2 -Test eine Approximation darstellt, sind von **Yates** korrigierte Formeln (7.202, 7.203) vorgeschlagen worden (die Größen $\frac{1}{2}$ bzw. $\frac{n}{2}$ werden als **Kontinuitätskorrektur** bezeichnet)

$$\hat{\chi}^2 = \left(|\Delta| - \frac{1}{2} \right)^2 \left(\frac{1}{E_a} + \frac{1}{E_b} + \frac{1}{E_c} + \frac{1}{E_d} \right) \quad (7.202)$$

$$\hat{\chi}^2 = \frac{n(|ad - bc| - n/2)^2}{(a+b)(c+d)(a+c)(b+d)} \quad (7.203)$$

Grizzle [Gri67] hat gezeigt, dass man auf (7.202, 7.203) verzichten kann (vgl. auch Haviland [Hav90] sowie Storer und Kim [SK90]). Nur wenn unbedingt die Wahrscheinlichkeiten des exakten Tests nach Fisher (vgl. [7.6.4]), eines konservativen Verfahrens, approximiert werden sollen, sind sie angebracht.

In **R** werden die Häufigkeiten zu einer Vierfeldertafel in der Struktur einer Matrix gespeichert. Hierzu kann insbesondere die Funktion **matrix()** verwendet werden. Zur besseren Lesbarkeit der Ergebnisse sollten die Zeilen (engl. rows) und Spalten (engl. columns) mit Hilfe der Funktion **dimnames()** in jedem Fall auch benannt werden. Für die Berechnung der χ^2 -Teststatistik nach (7.198) wird die Funktion **chisq.test()** verwendet. Für das Beispiel aus Tabelle 7.63 folgt:

```
> tab <- matrix(c(15, 85, 4, 77), nrow=2, ncol=2, byrow=TRUE)
> dimnames(tab) <- list(c("uebliche Therapie", "neue Therapie"),
+ c("gestorben", "geheilt")); tab
      gestorben geheilt
uebliche Therapie      15      85 neue Therapie      4 77
> chisq.test(tab, correct=FALSE)

Pearson's Chi-squared test

data:  tab X-squared = 4.8221, df = 1, p-value = 0.02810
```

Die Yates-Korrektur kann in der Funktion **chisq.test()** durch den zusätzlichen Parameter **correct=TRUE** berücksichtigt werden.

```
> chisq.test(tab, correct=TRUE)

Pearson's Chi-squared test with Yates' continuity correction
```

```
data: tab X-squared = 3.8107, df = 1, p-value = 0.05093
```

Die additive Eigenschaft von χ^2

Wiederholt durchgeführte Experimente an heterogenem Material, die sich nicht gemeinsam analysieren lassen, mögen folgende $\hat{\chi}^2$ -Werte $\hat{\chi}_1^2, \hat{\chi}_2^2, \hat{\chi}_3^2, \dots$ mit $\nu_1, \nu_2, \nu_3, \dots$ Freiheitsgraden liefern. Dann kann bei einheitlichem Zusammenhang in einer bestimmten Richtung das Ergebnis aller Versuche als äquivalent einem $\hat{\chi}^2$ -Wert aufgefasst werden, der durch $\hat{\chi}_1^2 + \hat{\chi}_2^2 + \hat{\chi}_3^2 + \dots$ mit $\nu_1 + \nu_2 + \nu_3 + \dots$ Freiheitsgraden gegeben ist.



Beispiel: Bei der Prüfung einer Nullhypothese ($\alpha = 0,05$) sei ein Experiment – sagen wir an unterschiedlichen Orten und an unterschiedlichem Material – viermal durchgeführt worden. Die entsprechenden $\hat{\chi}^2$ -Werte seien für jeweils einen Freiheitsgrad 2,30; 1,94; 3,60 und 2,92. Die Nullhypothese kann nicht abgelehnt werden. Aufgrund der additiven Eigenschaft von χ^2 lassen sich die Ergebnisse zusammenfassen:

$$\hat{\chi}^2 = 2,30 + 1,94 + 3,60 + 2,92 = 10,76 \quad \text{mit} \quad 1 + 1 + 1 + 1 = 4FG.$$

Da $\hat{\chi}_{4;0,95}^2 = 9,488$, muss für alle vier Experimente die Nullhypothese auf dem 5%-Niveau abgelehnt werden.

7.6.2.1 Fallzahl und Power zum Vierfeldertest

Die Bestimmung der benötigten Fallzahl für den Vierfeldertest ($H_0 : \pi_1 = \pi_2$ bzw. $H_0 : \pi_1 - \pi_2 = 0$), z. B. im Rahmen der Planung einer Fall- Kontroll-Studie, kann nach Fleiss [Fle81] auf der Grundlage des zentralen Grenzwertsatzes durch (7.204) erfolgen. Dabei wird zunächst von gleichen Stichprobenumfängen $n_1 = n_2 = n$ ausgegangen.

$$n' = \frac{\left(z_{1-\alpha/2} \cdot \sqrt{2\pi(1-\pi)} - z_{1-\beta} \cdot \sqrt{\pi_1(1-\pi_1) + \pi_2(1-\pi_2)} \right)^2}{(\pi_1 - \pi_2)^2} \quad (7.204)$$

Darin ist $\pi_1 - \pi_2$ ($\pi_1 > \pi_2$) die unter der Nullhypothese zu prüfende Differenz der Wahrscheinlichkeiten oder Anteile und $\pi = (\pi_1 + \pi_2)/2$ deren Mittelwert. Mit $z_{1-\alpha/2}$ bzw. $z_{1-\beta}$ sind die Quantile der Standardnormalverteilung für das Signifikanzniveau α (zweiseitig) und die Power $(1 - \beta)$ (einseitig) bezeichnet.



Beispiel: Wie viele Beobachtungen werden für die Überprüfung der Hypothese $H_0 : \pi_1 - \pi_2 = 0,38 - 0,30 = 0,08$ benötigt, wenn ein Signifikanzniveau von $\alpha = 0,05$ (zweiseitig) festgelegt ist und eine Power von $(1 - \beta) = 0,90$ eingehalten werden soll. Die Berechnung erfolgt in **R** elementar:

```
> z.alpha <- qnorm(0.975); z.beta <- qnorm(0.10)
>
> p1 <- 0.38;      q1 <- 1 - p1
> p2 <- 0.30;      q2 <- 1 - p2
> p <- (p1 + p2)/2; q <- 1 - p
>
> n <- (z.alpha * sqrt(2*p*q) - z.beta * sqrt(p1*q1+p2*q2))^2 / ((p2 - p1)^2); n
[1] 734.7537
```

Danach werden für die Prüfung der Nullhypothese mit dem χ^2 -Vierfeldertest zwei Stichproben mit dem Umfang von jeweils $n_1 = n_2 = 735$ Beobachtungen benötigt. In **R** kann die Berechnung der Fallzahl nach (7.204) auch einfach durch die Funktion **power.prop.test()** erfolgen. Insbesondere kann mit dieser Funktion auch der Zusammenhang zwischen der Power und der Fallzahl einfach dargestellt werden, unter anderem hinsichtlich eines Verlustes an Power, wenn die angestrebte Fallzahl nicht erreicht werden kann. Die Powerkurve hierzu enthält Abbildung 7.20.

```
> power.prop.test(p1=0.3,p2=0.38, sig.level =0.05, power = 0.90)
```

Two-sample comparison of proportions power calculation

```
      n = 734.7537
      p1 = 0.3
      p2 = 0.38
      sig.level = 0.05
      power = 0.9
      alternative = two.sided
```

NOTE: n is number in *each* group



Abb. 7.20. Powerkurve zu der Hypothese $H_0 : \pi_1 - \pi_2 = 0,38 - 0,30 = 0,08$; $\alpha = 0,05$ (zweiseitig)

Für die Ermittlung des Stichprobenumfangs n' nach (7.204) gibt es zur Einhaltung der Power verschiedene Möglichkeiten einer Kontinuitätskorrektur, z. B. in Casagrande et.al [CPS78]. Für den Fall, dass $n' \cdot |\pi_1 - \pi_2| \geq 4$, gibt Fleiss [Fle81] eine einfache Möglichkeit der Korrektur an.

$$n = n' + \frac{2}{|\pi_1 - \pi_2|} \quad (7.205)$$

7.6.2.2 Minimales n für den Vierfeldertest

Nach Tabelle 7.67 benötigt man für den Test $H_0 : \pi_1 = \pi_2$; $H_A : \pi_1 > \pi_2$ mit $\pi_1 = 0,7$, $\pi_2 = 0,3$ bei einer Irrtumswahrscheinlichkeit $\alpha = 0,05$ und einer Power von 0,9 $n_1 = n_2 = 31$ Beobachtungen, d. h. stehen für den Test zwei Zufallsstichproben dieser Umfänge aus Grundgesamtheiten mit $\pi_2 = 0,3$ und $\pi_1 = 0,7$ zur Verfügung, dann besteht bei einseitiger Fragestellung auf dem 5%-Niveau eine Chance von 90%, die Differenz $\delta = \pi_1 - \pi_2 = 0,7 - 0,3 = 0,4$, als statistisch signifikant auszuweisen. Nach der **Approximation**: $n_1 = n_2 \approx 5/\delta^2 = 5/0,16 = 31$

Weitere Beispiele: Gibt man eine Power von 80% vor und prüft auf dem 5%-Niveau, so benötigt man, um die extrem liegende Differenz $\pi_2 - \pi_1 = 0,9 - 0,6 = 0,3$ als statistisch signifikant auszuweisen (unteres Dreieck, untere Zahl) $n_1 = n_2 = 30$ Beobachtungen. Für die mehr in der Skalenmitte liegende Differenz $\pi_1 - \pi_2 = 0,7 - 0,4 = 0,3$ sind es jeweils bereits 41 Beobachtungen.

Zur Erfassung der Differenz $\pi_1 - \pi_2$ ($H_0 : \pi_1 = \pi_2$; $H_A : \pi_1 > \pi_2$; $\alpha = 0,05$) mit einer Power von 0,8 benötigte Stichprobenumfänge $n_1 = n_2 = n$ sind in der folgenden Übersicht angegeben. Die Werte zeigen, dass n umso kleiner wird, je größer die Differenz [vgl. (3) bis (1)] und je extremer sie liegt (d. h. je weiter beide von 0,5 entfernt sind) [vgl. (4) mit (3)]. Aus Symmetriegründen gilt natürlich auch [vgl. (4)] für $\pi_1 - \pi_2 = 0,2 - 0,1 = 0,1$ der Wert $n = 173$.



	(1)	(2)	(3)	(4)
π_1	0,6	0,6	0,6	0,9
π_2	0,1	0,3	0,5	0,8
$\pi_1 - \pi_2$	0,5	0,3	0,1	0,1
$n_1 = n_2$	13	41	321	173

Für $\pi_1 - \pi_2 = 0,95 - 0,90 = 0,05$ benötigt man unter den genannten Bedingungen schon jeweils 371 Beobachtungen; lässt man diese Differenz in Richtung auf die Skalenmitte wandern ($0,525 - 0,475 = 0,05$) und schrumpfen, etwa auf $\pi_1 - \pi_2 = 0,51 - 0,49 = 0,02$, so werden jeweils mehr als 1000 Beobachtungen benötigt, was im allgemeinen kaum möglich sein wird; dagegen ist die Erfassung der Differenz $\pi_1 - \pi_2 = 0,9 - 0,1 = 0,8$ mit $n_1 = n_2 = 6$ eher trivial.

Tabelle 7.67. Minimale Stichprobenumfänge ($n_1 = n_2$) für den Zweistichproben-Fisher-Test bei einseitiger Fragestellung; nach Haseman, J.K. (1978): Exact sample sizes for use with the Fisher-Irwin-Test for 2×2 tables. Biometrics **34**: 106–109 [der auch Stichprobenumfänge für die Power = 0,5 gibt, die im allgemeinen im Bereich 1/2 bis 1/3 der oberen Zahlen (Power = 0,9) liegen]

oberes Dreieck: $\alpha = 0,01$; obere Zahl: Power = 0,9

unteres Dreieck: $\alpha = 0,05$; untere Zahl: Power = 0,8

$\pi_2 \backslash \pi_1$	0,95	0,9	0,8	0,7	0,6	0,5	0,4	0,3	0,2	0,1	0,05
0,95		745 583	130 101	60 49	37 31	25 20	18 16	14 12	12 9	9 8	8 6
0,9	503 371		344 269	108 86	56 46	35 29	24 20	18 15	13 12	9 9	9 8
0,8	89 67	232 173		503 393	143 113	69 55	42 34	28 22	18 16	13 12	12 9
0,7	42 34	74 56	338 249		609 475	163 129	77 60	43 35	28 22	18 15	14 12
0,6	25 20	39 30	97 73	408 302		667 519	171 137	77 60	42 34	24 20	18 16
0,5	18 14	25 19	47 36	111 84	445 321		667 519	163 129	69 55	35 29	25 20
0,4	13 11	17 13	30 23	53 41	116 85	445 321		609 475	143 113	56 46	37 31
0,3	10 9	12 10	18 15	31 23	53 41	111 84	408 302		503 393	108 86	60 49
0,2	8 6	10 8	12 10	18 15	30 23	47 36	97 73	338 249		334 269	130 101
0,1	6 5	8 6	10 8	12 10	17 13	25 19	39 30	74 56	232 173		745 583
0,05	5 5	6 5	8 6	10 9	13 11	18 14	25 20	42 34	89 67	503 371	

Bemerkung: Erwartet man, dass Therapie 2 um 10% bessere Heilungschancen als Therapie 1 (z.B. Standard) aufweist, d. h. $\pi_1 - \pi_2 = 0,1$, und wird mit $\alpha = 0,05$ sowie einer Power von etwa 0,8 geprüft, so benötigt man insgesamt etwa 600 Beobachtungen, wenn man diese so aufteilt, dass $n_1 = 240$ und $n_2 = 360$ Beobachtungen umfasst; für $n_1 = n_2$ wären jeweils etwa 400 Beobachtungen notwendig (vgl. R.F. Mould 1979, Clinical Radiology **30**, 371–381).

7.6.2.3 Vorsicht vor Trugschlüssen beim Vierfeldertest

1. Sind zwei Merkmale von einem dritten Merkmal abhängig und lässt sich für die beiden erstgenannten die Unabhängigkeitshypothese ablehnen und ein stochastischer Zusammenhang sichern, so liegt ein Trugschluss vor.

2. Ein anderer Trugschluss kann bei Vierfeldertafeln auftreten, indem eine Ablehnung der Nullhypothese (Unabhängigkeit bzw. Homogenität), etwa auf dem 5%-Niveau, als Artefakt der Addition zweier der Nullhypothese entsprechender Vierfeldertafeln auftritt (vgl. auch 12, 18, 3, 9 mit 7, 3, 29, 19 sowie 19, 21, 32, 28):

$$\hat{\chi}_{*}^2 = \frac{241}{121^4} (101^2 - 20^2)^2 = 108$$

$$\begin{array}{|c|c|} \hline 1 & 10 \\ \hline 10 & 100 \\ \hline \end{array} + \begin{array}{|c|c|} \hline 100 & 10 \\ \hline 10 & 1 \\ \hline \end{array} = \begin{array}{|c|c|} \hline 101 & 20 \\ \hline 20 & 101 \\ \hline \end{array}$$

$$\hat{\chi}_{*}^2 = 0 \quad \hat{\chi}_{*}^2 = 0 \quad \hat{\chi}_{*}^2 = 108$$

Übrigens gilt dann, wenn alle 4 Randsummen gleich sind $[a + b = c + d = a + c = b + d]$:

$$\hat{\chi}_{*}^2 = (n - 1) \left[\frac{a - c}{a + c} \right]^2 \quad (7.206)$$

$$\hat{\chi}_{*}^2 = 241 \left[\frac{101 - 20}{101 + 20} \right]^2 = 108$$

7.6.3 Odds Ratio und relatives Risiko

Dieser Abschnitt vertieft den Abschnitt [4.3.1.2] über Risikomaße.

Fall-Kontrollstudie und Kohortenstudie

Vierfelder-Tafeln treten häufig in Verbindung mit zwei speziellen Studientypen auf.

Bei einer **Fall-Kontroll-Studie** wird (retrospektiv) einer Gruppe erkrankter Personen (Fallgruppe) eine gut vergleichbare Gruppe von Personen gegenübergestellt, die frei von der betreffenden Krankheit ist (Kontrollgruppe). Die Fälle müssen ebenso wie die Kontrollen bezüglich der zu untersuchenden Risikofaktoren repräsentativ für ihresgleichen in der betrachteten Grundgesamtheit sein. In beiden Gruppen werden dann Erhebungen, z. B. hinsichtlich potentieller Ursachenfaktoren vorgenommen. Studien dieser Art sind unerlässlich bei seltenen Krankheiten, z. B. der Legionärskrankheit und bei unerwünschten Arzneimittelwirkungen. Wichtig war z. B. die Erforschung der Wirkungen des Zigarettenrauchens. Fehlerquellen liegen in der möglichen Beeinflussung durch Suggestivfragen des Interviewers bei der Erhebung der zu untersuchenden Ursachenfaktoren, in der Überbewertung unbedeutender Ereignisse aus Kausalitätsbedürfnis und in der Vergesslichkeit des Patienten bei schon lange zurückliegenden Vorgängen (Wahrnehmungs-, Erinnerungs- und Wiedergabefehler). Ereignisse aus der Vergangenheit werden entweder vergessen oder näher zur Gegenwart eingeordnet. Um Strukturgleichheit zwischen der Fallgruppe und der Kontrollgruppe zu erzielen, wird häufig jedem Erkrankten eine in Alter, Geschlecht und anderen Merkmalen übereinstimmende Person gegenübergestellt, die diese Krankheit nicht aufweist (**matched pairs**). Ist die Anzahl der Erkrankten klein, so können auch jedem Patienten mehrere Kontrollfälle in dieser Weise zugeordnet werden. Ideal wäre es, wenn jedem Fall je eine Kontrolle aus Verwandtschaft, Nachbarschaft und Gesamtbevölkerung zugeordnet wäre. Nicht krankheits- sondern **expositionsorientiert ist der folgende Ansatz:**

Bei einer **Kohorten-Studie** geht man (prospektiv) von einer definierten Bevölkerung aus, „Kohorte“, die über eine längere Zeit beobachtet wird (Verlaufsstudie), am besten von einer Geburtsjahrgangskohorte, Personen, die im selben Jahr geboren wurden. Häufig handelt es sich um eine Gruppe von Personen, die einem besonderen Risiko ausgesetzt sind und die mit einer Gruppe verglichen wird, die dieses Risiko nicht hat. Nach einiger Zeit wird dann in beiden Gruppen festgestellt, wie groß die Zahl der Personen ist, die an einer bestimmten Krankheit erkrankt sind. Aus den **Neuerkrankungsraten** (Inzidenzen) wird dann das relative Risiko bestimmt. So wurde in der **Framingham-Studie** festgestellt, dass fettleibige Zigarettenraucher mit hohem Blutdruck am stärksten von Herzkrankheiten betroffen sind. Generelle Einsatzgebiete sind **risikoträchtige Situationen**, bedingt durch Lebensgewohnheiten, weitverbreitete Arznei- und Genussmittel, Arbeitsplatz (z. B. Asbest- oder Benzol-Exposition) sowie prophylaktische Maßnahmen wie Impfungen. Kohorten unterscheiden sich: fast stets sind viele Einfluss- und Störgrößen miteinander vermengt, **Confounding** genannt; Resultate von Kohortenstudien sind daher häufig nicht eindeutig zu interpretieren.

Relatives Risiko und Chancen-Verhältnis (Odds Ratio)

Die Identifizierung von Faktoren als Risikofaktoren ist eine wichtige Aufgabe der Epidemiologie. Bei einer Vierfeldertafel (vgl. Tabelle 7.68) mit Faktor und Krankheit, jeweils vorhanden oder nicht vorhanden, kann man **prospektiv** anhand einer Kohorten-Studie von **einem Faktor** ausgehen und die Neuerkrankungsraten (Inzidenzraten) bei Exponierten (mit Faktor) und Nicht-Exponierten (ohne Faktor) vergleichen. Das **Relative Risiko** wird als Quotient zweier Inzidenzraten [vgl. (7.207)] nach (7.208) bestimmt:

$$\text{Relatives Risiko} = \frac{\text{Inzidenzrate bei Exponierten}}{\text{Inzidenzrate bei Nicht-Exponierten}} = \frac{a/(a+b)}{c/(c+d)} \quad (7.207)$$

Tabelle 7.68. Vierfeldertafel für eine Kohorten-Studie

		Krankheit		Risiken	Relatives Risiko	
		+	-			
Faktor	+	a	b	Koh.-Studie prospektiv	$\frac{a}{a+b}$	$\frac{a}{a+b} : \frac{c}{c+d} = \frac{a}{a+b} \cdot \frac{c+d}{c} = \frac{ac+ad}{\underline{\underline{ac+bc}}}$
	-	c	d		$\frac{c}{c+d}$	

(7.208)

Je größer das Relative Risiko ist, um so eher lässt sich ein kausaler Zusammenhang zwischen der Exponiertheit (dem Risikofaktor) und der Krankheit postulieren.

Tabelle 7.69. Vierfeldertafel für eine Fall-Kontroll-Studie

		Krankheit	Odds	Odds Ratio
		+	-	
Faktor	+	a	b	$\frac{a}{c} : \frac{b}{d}$
	-	c	d	$\frac{b}{d} \cdot \frac{ad}{bc}$

Fall-Kontroll-Studie: retrospektiv

(7.209)

Man kann auch retrospektiv anhand einer Fall-Kontroll-Studie von einer Krankheit ausgehen (vgl. Tab. 7.69) und das Relative Risiko indirekt schätzen, indem man die Odds Ratio ad/bc bildet. Die **Odds Ratio**, auch das **Chancen-Verhältnis** („Kreuzprodukt“) genannt, ist eine gute Schätzung des Relativen Risikos, wenn bei vergleichbaren Gruppen die Inzidenzraten niedrig sind, d. h. wenn die Erkrankungshäufigkeiten a und c klein sind, d. h. das Produkt ac klein ist [vgl. (7.208) und (7.209)]. Tabelle 7.70 enthält drei Beispiele für $a + b + c + d = 200$.



Diese speziellen Maßzahlen für eine Vierfeldertafel sind nur zu berechnen, wenn (1) **unverzerrte Informationen** über Anwesenheit bzw. Abwesenheit des „Faktors“ vorliegen, (2) beide Personengruppen **repräsentativ** für ihre jeweiligen Grundgesamtheiten UND (3) auch sonst tatsächlich **vergleichbar** sind.

Tabelle 7.70. Drei Beispiele zum Relativen Risiko bzw. Odds Ratio

		Krankheit + -		Krankheit + -		Krankheit + -	
Faktor	+	40	60	20	80	2	98
	-	20	80	10	90	1	99
Inzidenzrate							
Faktor	+	0,4		0,2		0,02	
Faktor	-	0,2		0,1		0,01	
Relatives Risiko:		2,00		2,00		2,00	
Odds Ratio:		2,67		2,25		2,02	

Prospektive Studien haben die Tendenz, die Inzidenz in Gegenwart eines Risikofaktors zu überschätzen und in seiner Abwesenheit zu unterschätzen. Hierdurch wird das Relative Risiko überschätzt. Auch retrospektive Studien überschätzen im allgemeinen das Relative Risiko; außerdem werden die milderen Formen der Krankheit überrepräsentiert. Andere wichtige Gesichtspunkte bei Studien dieser Art sind geeignete Kontrollen, die Intensität des Faktors, zeitliche Abhängigkeiten sowie die Resultate ähnlicher Studien, bevor ein kausaler Zusammenhang zwischen Risikofaktor und Krankheit bedacht werden kann. Bezeichnen wir die Wahrscheinlichkeit, bei Vorliegen des Risikofaktors R an der Krankheit K zu erkranken, mit $P(K|R)$, dann ergibt sich die folgende Tabelle 7.71, in der die Wahrscheinlichkeiten $P(\bar{K}|R)$ und $P(K|\bar{R})$ irritieren; die erstere ist bei kleinen Erkrankungswahrscheinlichkeiten ebenso wie $P(\bar{K}|\bar{R})$ fast Eins, die letztere fast Null.

Tabelle 7.71. Vierfeldertafel und bedingte Wahrscheinlichkeiten

	K	$\bar{K}$	Σ
R	$P(K R)$	$P(\bar{K} R)$	$P(R)$
$\bar{R}$	$P(K \bar{R})$	$P(\bar{K} \bar{R})$	$P(\bar{R})$
Σ	$P(K)$	$P(\bar{K})$	1

Mit dem Bezug auf eine bestimmte **Krankheit** K , den **Risikofaktor** R und den Wahrscheinlichkeiten der Tabelle 7.71 bezeichnet man die Parameter:

$$\text{Delta: } \delta = P(K|R) - P(K|\bar{R}) \text{ als dem Risikofaktor } \mathbf{zuschreibbares Risiko} \quad (7.210)$$

$$\text{Psi: } \psi = \frac{P(K|R)}{P(K|\bar{R})} \text{ als } \mathbf{Relatives Risiko} \quad (7.211)$$

$$\text{Omega: } \omega = \frac{P(K|R) \cdot P(\bar{K}|\bar{R})}{P(\bar{K}|R) \cdot P(K|\bar{R})} = \frac{P(K|R) \cdot P(\bar{K}|\bar{R})}{P(\bar{K}|R) \cdot P(K|\bar{R})} = \frac{P(K|R)}{P(K|\bar{R})} \cdot \frac{P(\bar{K}|\bar{R})}{P(\bar{K}|R)} \quad (7.212)$$

als **Chancen-Verhältnis** oder **Odds Ratio**

Ein Faktor gilt als **Risikofaktor**, wenn δ größer als Null ist, dann sind auch ψ und ω , die sich bei niedrigen Erkrankungsraten in beiden Gruppen nur wenig unterscheiden, größer als Eins.

$$\delta > 0, \quad \psi > 1 \quad \text{und} \quad \omega > 1 \quad (7.213)$$

Die den Parametern δ , ψ und ω entsprechenden (mit einem „Dach“ versehenen) Schätzwerte $\hat{\delta}$, $\hat{\psi}$ und $\hat{\omega}$ sind nur zu berechnen, wenn die „Fälle“ und die „Kontrollen“ (a) vergleichbar und (b) repräsentative Zufallsstichproben aus ihren jeweiligen Grundgesamtheiten sind; d. h. man sollte die Alters- und Geschlechtsverteilung der Grundgesamtheit und die entsprechenden Sterbeziffern der Krankheit kennen.

Aus retrospektiven Studien kann $\hat{\psi}$ nicht berechnet werden; dann dient $\hat{\omega}$ als noch brauchbarer Schätzwert für $\hat{\psi}$.



Beispiele aus Kohorten-Studien:

1. Schematisches Beispiel (große Probandenzahlen vorausgesetzt)

- (1) Erkrankungsrate bei Exponierten: $0,20$ (20%)
- (2) Erkrankungsrate bei Nicht-Exponierten: $0,05$ (5%)
- (3) Dem Risikofaktor zuschreibbares Risiko: $\hat{\delta} = 0,15$ (15%)
- (4) Relatives Risiko: $\hat{\psi} = 0,20/0,05 = 4,00$ (> 1)
- (5) Chancen-Verhältnis: $\hat{\omega} = \frac{0,20}{0,80} / \frac{0,05}{0,95} = 4,75$ (> 1)

Das Erkrankungsrisiko bei Exponierten ist viermal so groß wie bei Nicht-Exponierten.

2. Ergebnisse einer Kohortenstudie in einer Vierfeldertafel:

Tabelle 7.72. Beispieldaten zu Risiko-Maßen

Exposition	Personen		Summe
	krank	nicht krank	
vorhanden	24	96	120
nicht vorhanden	48	592	640
Summe	72	688	760

- (1) Erkrankungsrate bei Exponierten: $24/120 = 0,200$ (20%)
- (2) Erkrankungsrate bei Nicht-Exponierten: $48/640 = 0,075$ (7,5%)
- (3) Dem Risikofaktor zuschreibbares Risiko: $\hat{\delta} = 0,125$ (12,5%)
- (4) Relatives Risiko: $(24/120)/(48/640) = 0,200/0,075 = \hat{\psi} = 2,667$ (> 1)
- (5) Chancen-Verhältnis: $(24/48) \cdot (592/96) = 0,5 \cdot 6,1667 = \hat{\omega} = 3,083$ (> 1)

Das Erkrankungsrisiko bei Exponierten ist 2,7mal so groß wie bei Nicht-Exponierten.

In **R** erfolgt die Berechnung dieser Maßzahlen zu Vierfeldertafeln am einfachsten direkt. Das folgende Beispiel zeigt die Zuordnung der Häufigkeiten und die Berechnung für die Ergebnisse aus Tabelle 7.72.

```

> a <- 24; b <- 96; c <- 48; d <- 592
> tab <- matrix(c(a, b, c, d), nrow=2, ncol=2, byrow=TRUE)
> dimnames(tab) <- list(c("exponiert", "nicht exponiert"),
+                        c("krank", "nicht krank")); tab
      krank nicht krank
exponiert      24      96 nicht exponiert      48
592
> IR.exp <- a / (a+b); IR.exp           # Inzidenzrate exponiert
[1] 0.2
> IR.nexp <- c / (c+d); IR.nexp         # Inzidenzrate nicht exponiert
[1] 0.075
> delta <- IR.exp - IR.nexp; delta      # zuschreibbares Risiko
[1] 0.125
> psi <- IR.exp / IR.nexp; psi          # relatives Risiko
[1] 2.666667
> omega <- (a*d) / (b*c); omega         # Odds Ratio
[1] 3.083333

```

Rein beschreibend lässt sich aus Tabelle 7.73 entnehmen, ob der Faktor und damit die Exposition unter Umständen sogar nützlich (protektiv, schützend) ist. Ob die Exposition auf dem gewählten Niveau statistisch signifikant ist, das lässt sich anhand des χ^2 -Tests nach Pearson, Mantel und Haenszel (7.214) sowie anhand des testbasierten approximierten Konfidenzintervalls für das Relative Risiko bzw. für das Chancen-Verhältnis (7.215) bestimmen.

Tabelle 7.73. Relatives Risiko und Exposition

relatives Risiko	Exposition
$\leq 0,3$	starker Nutzen
$0,4 - 0,8$	Nutzen
$0,9 - 1,1$	kein Effekt
$1,2 - 2,5$	Schaden
$\geq 2,6$	starker Schaden

Relative Risiken zwischen 1,2 und 2,5 lassen, wenn überhaupt, einen schwachen Zusammenhang zwischen Faktor und Krankheit vermuten. Je näher der Wert bei 1 liegt, umso wichtiger werden: Kontrollgruppe (a), die Abwesenheit von Störeffekten (b) und die Notwendigkeit, den zugrundeliegenden biologischen Zusammenhang zu verstehen (c).

Um eine **kausale Abhängigkeit** zwischen Risikofaktor und Krankheit wahrscheinlich zu machen, müssen mindestens 5 Bedingungen erfüllt sein (Kelsey u. Mitarb. [KTE96] nennen weitere Kriterien):



- (1) Wiederholbarkeit des Zusammenhangs, des Effektes, in unterschiedlichen Studien sowie in unterschiedlichen Untergruppen derselben Studie.
- (2) Der Effekt sollte deutlich sein, etwa in der Art einer Dosiswirkungsbeziehung.
- (3) Der Effekt sollte bezüglich der Ursache und der Wirkung spezifisch sein.
- (4) Die Ursache muss der Wirkung stets VORANGEHEN.
- (5) Der Effekt sollte biologisch plausibel und möglichst experimentell nachvollziehbar sein.

7.6.3.1 Angenäherte 95%-Konfidenzintervalle für das Relative Risiko ψ und für das Chancen-Verhältnis (die Odds Ratio) ω

χ^2 -Test nach Pearson, Mantel und Haenszel

$$\begin{array}{cc|c}
 a & b & \\
 c & d & \\
 \hline
 & & n
 \end{array}
 \quad
 \hat{\chi}_{\text{PMH}}^2 = \frac{(n-1)(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}
 \quad (7.214)$$

$$\left. \begin{matrix} H_O : \\ H_A : \end{matrix} \right\} \text{es existiert } \left\{ \begin{matrix} \text{kein} \\ \text{ein} \end{matrix} \right\} \text{Effekt: Schad- bzw. Nutzeffekt}$$

Für $\hat{\chi}_{\text{PMH}}^2 > \chi_{1;\alpha}^2$ wird H_0 auf dem $100\alpha\%$ -Niveau abgelehnt.

Mit den Zahlen aus obigem Beispiel (Tabelle 7.73) folgt ($\alpha = 0,05$):

$$\hat{\chi}_{\text{PMH}}^2 = \frac{759(24 \cdot 592 - 96 \cdot 48)^2}{120 \cdot 640 \cdot 72 \cdot 688} = 18,387$$

Damit liegt auf dem 5%-Niveau ($18,387 > 3,841 = \chi^2_{1;0,95}$) ein echter Risikofaktor vor. Für eine Fall-Kontroll-Studie hätte man die dem Risikofaktor Exponierten durch „Fälle“ und die Nicht-Exponierten durch „Kontrollen“ ersetzt.

Testbasierter 95%-Vertrauensbereich
(Approximation für großes n und $0,2 < \hat{\psi}, \omega < 6$)

$$\boxed{\hat{\psi}^{1\pm 1,96/\sqrt{\hat{\chi}_{\text{PMH}}^2}} \quad \hat{\omega}^{1\pm 1,96/\sqrt{\hat{\chi}_{\text{PMH}}^2}} \quad (7.215)}$$

Ein 95%-Konfidenzintervall, das den Wert 1 ausschließt, weist ein statistisch signifikantes Relatives Risiko bzw. Chancen-Verhältnis nach.

Fortsetzung des Beispiels: Exponent = $1 \pm 1,96/\sqrt{18,38663} = 1,45709$ bzw. $0,54291$

$$\begin{array}{cc|c} 24 & 96 & 120 \\ 48 & 592 & 640 \\ \hline 72 & 688 & 760 \end{array} \quad \hat{\psi} = \frac{24/120}{48/640} = 2,667 \quad \left[\hat{\omega} = \frac{24}{48} \Big/ \frac{96}{592} = 3,083 \right]$$

$$\begin{aligned} 95\% \text{-KI: } 2,667^{0,543} &= 1,70 \leq \psi \leq 4,18 = 2,667^{1,457} \\ [95\% \text{-KI: } 3,083^{0,543} &= 1,84 \leq \omega \leq 5,16 = 3,083^{1,457}] \end{aligned}$$

Eine andere Approximation (KatzKatz u. Mitarb., siehe Kahn und Sempos [KS89]):

$$\boxed{\text{95\%-KI für } \ln \psi = \ln \psi \pm 1,96 \sqrt{\frac{b/a}{a+b} + \frac{d/c}{c+d}}} \quad (7.216)$$

$$\ln 2,6667 \pm 1,96 \sqrt{\frac{96/24}{24+96} + \frac{592/48}{48+592}}, \text{ d. h. } 0,9808 \pm 0,5021, \text{ d. h. } 0,4787 \text{ bis } 1,4829$$

$$\text{95\%-KI: } e^{0,4787} \leq \psi \leq e^{1,4829}, \text{ d. h. } 1,61 \leq \psi \leq 4,41$$

liefert ähnliche Werte, die zwar deutlich von den exakten Werten $1,72 \leq \psi \leq 5,40$ (vgl. Mehta u. Mitarb. [MPG85]) abweichen, das Resultat „Risikofaktor“ aber bestätigen.

Mit einer Wahrscheinlichkeit von 95% wird bei wiederholter berechtigter Anwendung von (7.215) [bzw. (7.216)] ψ [bzw. ω] innerhalb der berechneten Grenzen liegen. Im vorliegenden Beispiel ist das Relative Risiko deutlich größer als 1. Damit ist der Faktor R als Risikofaktor ausgewiesen, was auch durch den oben berechneten Wert $\hat{\chi}^2 = 18,4$ zum Ausdruck kommt.

In **R** kann die Berechnung von Konfidenzintervallen der Odds Ratio indirekt durch die Funktion **fisher.test()** (näheres zum exakten Fisher-Test vgl. [7.6.4]) erfolgen. In dem Zusatzpaket **vcd** (visualizing categorical data) [MZKH05] berechnet die Funktion **oddsratio()** die Odds Ratio mit entsprechendem Konfidenzintervall.

```
> library(vcd)
> tab <- matrix(c(a, b, c, d), nrow=2, ncol=2, byrow=TRUE)
> dimnames(tab) <- list(c("exponiert", "nicht exponiert"),
+                        c("krank", "nichtkrank")); tab
```

```

exponiert      24      96
nicht exponiert 48     592
> OR <- oddsratio(tab, log=FALSE); summary(OR); confint(OR)
Odds Ratio
[1,]      3.0833
      lwr      upr
[1,] 1.812488 5.245244

```

Das Chancen-Verhältnis beträgt danach 3,08. Die Grenzen für das 95%-Konfidenzintervall sind $[1,81 - 5,25]$. Sie weichen nur geringfügig von den oben gegebenen approximativen Werten ab. Abbildung 7.21 zeigt auf der linken Seite einen Mosaikplot der Daten aus Tabelle 7.72. Die relativen Häufigkeiten werden hier durch Rechteckflächen dargestellt. Auf der rechten Seite wird das entsprechende Konfidenzintervall zur Odds Ratio wiedergegeben. Hierfür kann einfach die Funktion **plot(OR)** verwendet werden.

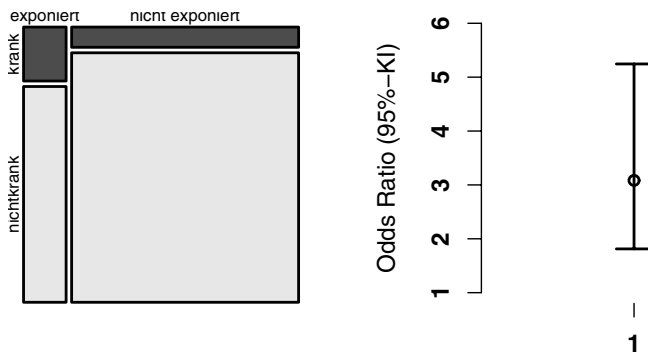


Abb. 7.21. Mosaikplot zu Tabelle 7.72 und die entsprechende Odds Ratio mit 95%-KI

7.6.3.2 Benötigte Stichprobenumfänge nach Lemeshow, Hosmer und Klar, um das Chancen-Verhältnis (die Odds Ratio) aus Fall-Kontroll-Studien und das Relative Risiko aus Kohorten-Studien zu schätzen, wenn bereits Mutmaßungen über die gesuchten Parameter vorliegen.

I. Fall-Kontroll-Studien: Die Zahl der für Fälle (n_1) und Kontrollen (n_2) benötigten Personen, um den Parameter ω mit einem Fehler von höchstens ε (z. B. $\varepsilon = 0,1$) mit einer Vertrauenswahrscheinlichkeit von 95% zu schätzen, wenn für ω ein mutmaßlicher Wert und für den Anteil der Exponierten unter den Kontrollen der Wert P_2^* vorgegeben wird, beträgt:

$$n_1 = n_2 = \frac{1,96^2 \left[\frac{1}{P_1^*(1 - P_1^*)} = \frac{1}{P_2^*(1 - P_2^*)} \right]}{[\ln(1 - \varepsilon)]^2} \quad (7.217)$$

mit $P_1^* = \frac{\omega P_2^*}{\omega P_2^* + (1 - P_2^*)}$

P_1^* ist der Anteil Exponierter in der Grundgesamtheit der Fälle (der Erkrankten), P_2^* ist der Anteil Exponierter in der Grundgesamtheit der Kontrollen.

Beispiel: $\omega \approx 2$ und $P_2^* = 0,30$

$$n_1 = n_2 = \left\{ 1,96^2 \left[\frac{1}{0,46 \cdot 0,54} + \frac{1}{0,3 \cdot 0,7} \right] \right\} / [\ln(1 - 0,1)]^2 = 33,7587 / 0,011101 = 3041$$



Wir benötigen somit 3041 Personen für die Gruppe der Fälle und 3041 Personen für die Kontrollgruppe, um die wahre Odds Ratio ω mit einer Vertrauenswahrscheinlichkeit von 95% auf 10% genau zu schätzen. Begnügen wir uns mit einer Genauigkeit von 50%, so ist im Nenner „0,1“ durch „0,5“ zu ersetzen:

$$n_1 = n_2 = \left\{ 1,96^2 \left[\frac{1}{0,46 \cdot 0,54} + \frac{1}{0,3 \cdot 0,7} \right] \right\} / [\ln(1 - 0,5)]^2 = 33,7587 / 0,480453 = 70,3$$

Man benötigt jetzt jeweils nur 71 Personen, um ω auf 50% genau zu schätzen, vorausgesetzt $\omega \approx 2$ und $P_2^* = 0,30$ bei einer Vertrauenswahrscheinlichkeit von 95% (d. h. 1,96 im Zähler).

II. Kohortenstudien: Die benötigten Stichprobenumfänge für eine Kohortenstudie (m_1 Zahl der Exponierten und m_2 Zahl der nicht Exponierten), um den Parameter ψ zu schätzen, sind dann

$$m_1 = m_2 = \frac{1,96^2 \left[\frac{(1 - P_1)}{P_1} + \frac{(1 - P_2)}{P_2} \right]}{[\ln(1 - \varepsilon)]^2} \quad (7.218)$$

mit $P_1 = \psi P_2$

P_1 ist der Anteil von Exponierten in der Grundgesamtheit, bei der sich die Krankheit entwickelt, P_2 ist der entsprechende Anteil von Nicht-Exponierten; ε ist die erstrebte Genauigkeit und für ψ liege bereits ein mutmaßlicher Wert vor.



Beispiel: Angenommen, wir planen eine Kohortenstudie. Wir erwarten, dass in der Gruppe der Nicht-Exponierten 20% Krankheitsfälle auftreten werden. Wie viele Personen brauchen wir für jede Gruppe, um das wahre Relative Risiko ψ mit einer Vertrauenswahrscheinlichkeit von 95% auf 10% genau zu schätzen, wobei wir annehmen, $\psi \approx 1,75$:

d. h. $P_2 = 0,2$ und $P_1 = \psi P_2 = 1,75 \cdot 0,2 = 0,35$

$$m_1 = m_2 = \left\{ 1,96^2 \left[\frac{(0,65)}{0,35} + \frac{(0,8)}{0,2} \right] \right\} / [\ln(1 - 0,1)]^2 = 2027$$

Damit werden jeweils 2027 Personen benötigt. Begnügen wir uns mit $\varepsilon = 0,5$, so reichen $m_1 = m_2 = 47$ Personen.

Hinweis: Breslow ([Bre82]) gibt ein Nomogramm, das es gestattet, die zur Schätzung eines mutmaßlichen Relativen Risikos benötigten minimalen Probandenzahlen abzulesen. Um ein Relatives Risiko $\psi = 2$ mit einer Power von 95% zu schätzen, benötigt man dann, wenn der dem Risikofaktor ausgesetzte Anteil der Bevölkerung (P_{RF}) zwischen 0,3 und 0,6 liegt etwa 200 Fälle und 200 Kontrollen; erwartet man $\psi = 4$ und einen Anteil Exponierter von 0,25 bis 0,45, so genügen jeweils rund 50 Probanden; nähert sich der Anteil Exponierter in der Bevölkerung den Werten 0,005 (Null) oder 0,99 (Eins), so werden sehr umfangreiche Probandenzahlen notwendig.

7.6.3.3 Der expositionsbedingte Anteil Erkrankter: Population Attributable Risk

Das sogenannte Population Attributable Risk ($\widehat{PAR}$) wird definiert anhand des Relativen Risikos $\hat{\psi}$ und des Anteils der Bevölkerung, der dem Risiko-Faktor ausgesetzt (exponiert) ist (der Prävalenz des Risikofaktors) P_{RF} :

$$\widehat{PAR} = \frac{P_{RF}(\hat{\psi} - 1)}{1 + P_{RF}(\hat{\psi} - 1)} \quad \text{für } \hat{\psi} \geq 1 \quad (7.219)$$

Bei nicht zu kleinem Produkt $P_{RF} \cdot \hat{\psi}$, sagen wir $P_{RF}(\hat{\psi} - 1) > 9$, wird $PAR > 9/10$.



Beispiel: Angenommen, in einer Bevölkerung gebe es 25% Raucher und das Relative Risiko für Lungenkrebs, bezogen auf das Rauchen sei 10, dann beträgt der Anteil auf das Rauchen zurückzuführender Lungenkrebs-Fälle in der Gesamtbevölkerung

$$\widehat{\text{PAR}} = [0,25(10 - 1)]/[1 + 0,25(10 - 1)] = 0,692$$

d. h. 69% der Fälle hätten vermieden werden können.

Tabelle 7.74. Einige $\widehat{\text{PAR}}$ -Werte [expositionsbedingte Anteile Erkrankter] in Abhängigkeit vom Anteil der Exponierten in der Bevölkerung und dem Relativen Risiko [nach (7.219)]

Relatives Risiko	Anteil Exponierter in der Bevölkerung				
	0,001	0,01	0,1	0,5	0,9
3	0,002	0,020	0,167	0,500	0,643
11	0,010	0,090	0,500	0,833	0,900

Weitere Beispiele (vgl. auch Tabelle 7.74):

(1) $\psi = 3$
 $P_{\text{RF}} = 0,05 \quad \widehat{\text{PAR}} = \frac{0,05(3 - 1)}{1 + 0,05(3 - 1)} = 0,091$

(2) $\psi = 9$
 $P_{\text{RF}} = 0,5 \quad \widehat{\text{PAR}} = \frac{0,5(9 - 1)}{1 + 0,5(9 - 1)} = 0,800$

Testen wir bei einseitiger Fragestellung $H_0: \widehat{\text{PAR}} = 0$ mit $\alpha = 0,05$ und einer Power von 0,9, dann werden für Fall-Kontroll-Studien (FK) bzw. für Kohorten-Studien (KH) jeweils zwei gleichgroße Zufallsstichproben des Umfangs n benötigt, für Querschnittstudien (QS) jeweils nur eine. Einige charakteristische Werte aus Smith und McHugh [SM88] enthält Tabelle 7.75:

Tabelle 7.75. Benötigte Stichprobenumfänge

Bevölkerungsanteil		Art der Studie					
exponiert	erkrankt	FK: $n_1 = n_2$		KH: $n_1 = n_2$		QS: n	
		PAR					
		0,30	0,60	0,30	0,60	0,30	0,60
0,2	0,001	48	11	11 259	3 050	29 057	6 842
	0,01	48	11	1 116	303	2 879	679
	0,1	40	9	102	28	261	63
0,7	0,001	370	74	81 347	16 628	200 828	42 247
	0,01	364	73	8 066	1 650	19 919	4 194
	0,1	306	62	738	152	1 828	389

Beispiel: Angenommen, es wird eine Studie geplant, die bei Frauen den möglichen Zusammenhang zwischen Zigarettenrauchen und Herzinfarkt klären soll. Erfasst werden soll mit einer Power von 0,9 zumindest ein $\widehat{\text{PAR}} \geq 0,30$ ($H_0: \widehat{\text{PAR}} = 0$). Geprüft wird bei einseitiger Fragestellung auf



dem 5%-Niveau. Wir nehmen weiter an, dass in der Bevölkerung der Frauen, aus der wir Zufallsstichproben ziehen wollen, 20% Zigaretten rauchen, das Risiko eines Herzinfarktes betrage 0,001. Die Tabelle zeigt:

- (1) Für eine Fall-Kontroll-Studie benötigt man lediglich 48 Fälle und 48 Kontrollen.
- (2) Für eine Kohortenstudie benötigt man dagegen 11 259 Exponierte und 11 259 Nichtexponierte.
- (3) Für eine Querschnittstudie sind sogar 29 057 Personen notwendig.

7.6.4 Exakter Fisher-Test

Bei Vierfeldertafeln mit sehr kleinen Besetzungszahlen geht man von dem Feld aus mit dem kleinsten Produkt der Diagonalen und dem am schwächsten besetzten Feld (Tab. 7.76: $2 \cdot 4 < 8 \cdot 10$, d. h. 2) und stellt unter Konstanz der Randsummen alle Vierfeldertafeln auf, die in dem betreffenden Feld **noch schwächer** besetzt sind. In der Gesamtheit aller dieser Vierfeldertafeln haben diejenigen mit der beobachteten oder einer noch geringeren Besetzung des am schwächsten besetzten Feldes die Wahrscheinlichkeit P . Anders ausgedrückt: Nimmt man die **Randsummen der Vierfeldertafel als gegeben** und fragt nach der Wahrscheinlichkeit dafür, dass die beobachtete Besetzung der Tafel oder eine **noch weniger wahrscheinliche** rein zufällig zustandekommt (einseitige Fragestellung), so ergibt sich diese Wahrscheinlichkeit P als eine Summe von Gliedern der **hypergeometrischen Verteilung**:

$$P = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{n!} \sum_i \frac{1}{a_i!b_i!c_i!d_i!} \quad (7.220)$$

Der Index i bedeutet, dass für jede der aufgestellten Tafeln der hinter dem Summenzeichen stehende Ausdruck zu berechnen und dann in der Summe zusammenzufassen ist.

Der „exakte“ Test nach R.A. Fisher prüft die Nullhypothese, die **Odds Ratio ist gleich Eins**. Der Fisher-Test ist „exakt“, wenn **beide Randsummenpaare fest vorgegeben** sind: dann ist auch die berechnete hypergeometrische Wahrscheinlichkeit die exakte Lösung. Nicht aber für den Fall, dass die Vierfeldertafel auf Unabhängigkeit (n fest vorgegeben) oder Homogenität (ein Randsummenpaar fest vorgegeben) geprüft wird.

Ist auf Unabhängigkeit oder Homogenität zu prüfen, so entscheidet der exakte Fisher-Test **zu konservativ**. Dies lässt sich nach J.E. Overall (1990, Statistics in Medicine **9**, 379–382) korrigieren, indem zu denjenigen beiden Besetzungszahlen, deren beobachtete Häufigkeiten größer sind als die bei Unabhängigkeit erwarteten Häufigkeiten, je ein „+1“ addiert wird („augmented 2 · 2 table“).

Tabelle 7.76 Beispiel zum exakten Test nach R.A. Fisher

2	8	10		1	9	10		0	10	10
10	4	14	→	11	3	14	→	12	2	14
12	12	24		12	12	24		12	12	24
	a				b				c	

Aus der Grundtafel (Tabelle 7.76) erhalten wir zwei Tafeln mit extremeren Verteilungen. Die Wahrscheinlichkeit, daß die in der Grundtafel vorliegende Verteilung auftritt, ist

$$P = \frac{10! \cdot 14! \cdot 12! \cdot 12!}{24!} \cdot \frac{1}{2! \cdot 8! \cdot 10! \cdot 4!}$$

Die Gesamtwahrscheinlichkeit für die beobachtete und noch extremere Verteilungen beträgt



$$P = \frac{10! 14! 12! 12!}{24!} \left(\frac{1}{2! 8! 10! 4!} + \frac{1}{1! 9! 11! 13!} + \frac{1}{0! 10! 12! 2!} \right)$$

$$P = 0,018 \text{ (einseitiger Test)}$$

In **R** kann der exakte Fisher-Test mit der Funktion **fisher.test()** berechnet werden. Neben dem P-Wert wird in dieser Funktion auch die Odds Ratio mit dem entsprechenden Konfidenzintervall (hier 95%-KI) bestimmt. Für die Beispieldaten aus Tabelle 7.76 folgt:

```
> tab <- matrix(c(2, 8, 10, 4), byrow=TRUE, nr = 2); tab
      [,1] [,2]
[1,]    2    8 [2,]   10    4
> fisher.test(tab, alternative="less", conf.level=0.95)

Fisher's Exact Test for Count Data

data:  tab
p-value = 0.01804
alternative hypothesis: true odds
ratio is less than 1
95 percent confidence interval: 0.0000000
0.6965009 sample estimates: odds ratio 0.1121872
```

Bei symmetrischer hypergeometrischer Verteilung (d. h. hier, Tab. 7.76: Zeilen- oder Spaltensummen gleich groß) gilt für die zweiseitige Fragestellung $2P$, d. h. im Beispiel $P = 0,036$. In beiden Fällen wird die Nullhypothese ($\pi_1 = \pi_2$ bzw. Unabhängigkeit) (wegen $P < 0,05$) auf dem 5%-Niveau abgelehnt.

Hinweis: Die Berechnung der Wahrscheinlichkeit (P-Wert) für den exakten Fisher-Test kann einfacher und schneller als in (7.220) durch die folgenden Rekursionsformeln erfolgen:

Für $ad - bc < 0$ Für $ad - bc > 0$

$$\boxed{P_{i+1} = \frac{a_i \cdot d_i}{b_{i+1} \cdot c_{i+1}} P_i} \quad \boxed{P_{i+1} = \frac{b_i \cdot c_i}{a_{i+1} d_{i+1}} P_i} \quad (7.221)$$

Bezeichnen wir die 3 Tabellen der Tabelle 7.76 von links nach rechts mit 1, 2, 3, so ergibt sich für die Grundtabelle (a)

$$P = \frac{10! \cdot 14! \cdot 12! \cdot 12! \cdot 1!}{24! \cdot 2! \cdot 8! \cdot 10! \cdot 4!} = 0,016659$$

für die Tabelle b $P_{1+1} = P_2 = \frac{2 \cdot 4}{9 \cdot 11} \cdot P_1 = 0,0808 \cdot 0,016659 = 0,001346$

und für die Tabelle c $P_{2+1} = P_3 = \frac{1 \cdot 3}{10 \cdot 12} \cdot P_2 = 0,0250 \cdot 0,001346 = 0,000034$

$$\text{Insgesamt: } P = P_1 + P_2 + P_3 = 0,0167 + 0,0013 + 0,0000 = 0,018.$$

P-Werte dieser Art werden heute fast stets per Programm berechnet. Früher wurden Tabellen benutzt. Eine moderne Version für $5 \leq n_1 + n_2 \leq 40$ stammt von Martin Andrés und Mitarbeitern [MALDCHT91]: einige Werte für die ein- und zweiseitige Fragestellung enthält Tabelle 7.77.

7.6.5 Der von McNemar modifizierte Vorzeichentest

Zwei Versuche an denselben Individuen: Signifikanz einer Änderung des Häufigkeitsverhältnisses zweier abhängiger Verteilungen von Alternativdaten

Wird eine Stichprobe **zweimal** – etwa in einem gewissen zeitlichen Abstand oder unter veränderten Bedingungen – auf ein bestimmtes alternatives Merkmal hin untersucht, so haben wir es im allgemeinen nicht mehr mit unabhängigen, sondern mit **abhängigen Stichproben** zu tun. Jedes Stichprobenelement liefert zwei Beobachtungsdaten, die paarweise einander zugeordnet sind.

Tabelle 7.77. Exakter Test nach R.A. Fisher: Einige Werte n_0 für $8 \leq n_1 + n_2 \leq 16$ nach Martin Andrés und Mitarbeitern

x_1	y_1	n_1	1) a_1 = kleinste der 4 Randsummen
x_2	y_2	n_2	2) x_1 so, dass $x_1/n_1 < x_2/n_2$
a_1	a_2	N	$\hat{p}_1 < \hat{p}_2$

$H_0: p_1 = p_2$; $H_{A1}: p_1 < p_2$; $H_{A2}: p_1 \neq p_2$.

Für $n_1 \geq n_0$ (tabelliert für N , a_1 und x_1) wird H_0 abgelehnt.

N	Test: einseitig					zweiseitig			Test: einseitig					zweiseitig			N
	a_1	x_1	10%	5%	1%	10%	5%	1%	a_1	x_1	10%	5%	1%	10%	5%	1%	
8	2	0	6	6	—	6	6	—	1	0	13	—	—	13	—	—	14
	3	0	4	5	—	5	5	—	2	0	10	11	—	10	11	—	
	4	0	4	4	—	4	4	—	3	0	7	9	11	8	9	11	
9	2	0	6	7	—	6	7	—	1	11	—	—	—	11	—	—	15
	3	0	5	5	—	5	5	—	4	0	6	7	9	6	8	9	
	4	0	4	4	5	4	4	5	1	9	10	—	—	9	10	—	
10	1	0	9	—	—	9	—	—	5	0	5	6	8	5	6	8	16
	2	0	7	8	—	7	8	—	1	8	9	—	—	8	9	—	
	3	0	5	6	7	6	6	7	6	0	6	6	6	6	6	6	
	4	0	4	5	6	4	5	6	1	7	8	—	—	8	8	—	
	5	0	5	5	5	5	5	5	7	0	7	7	7	7	7	7	
11	1	0	10	—	—	10	—	—	1	7	7	—	—	7	7	—	15
	2	0	8	9	—	8	9	—	1	0	14	—	—	14	—	—	
	3	0	6	7	8	6	7	8	2	0	10	12	13	10	12	13	
	4	0	5	5	7	5	6	7	3	0	8	9	11	8	9	11	
	5	0	5	5	6	5	5	6	1	12	—	—	—	12	—	—	
12	1	0	11	—	—	11	—	—	4	0	6	8	10	7	8	10	16
	2	0	8	9	—	8	9	—	1	10	11	—	—	10	11	—	
	3	0	6	7	9	7	7	9	5	0	5	6	8	6	6	8	
	4	0	5	6	8	5	7	8	1	9	9	—	—	9	10	—	
	5	0	5	5	6	5	5	7	6	0	6	6	7	6	6	7	
13	1	0	12	—	—	12	—	—	1	7	8	—	—	8	8	—	16
	2	0	9	10	—	9	10	—	7	0	7	7	7	7	7	7	
	3	0	7	8	10	7	8	10	1	7	7	8	8	7	7	—	
	4	0	5	6	8	6	7	8	1	7	7	8	8	7	7	—	
	5	0	5	5	7	5	6	7	1	0	15	—	—	15	—	—	
14	1	0	10	—	—	10	—	—	2	0	11	12	14	11	12	14	16
	2	0	8	9	—	8	9	—	3	0	8	10	12	9	10	12	
	3	0	6	7	9	7	7	9	1	13	—	—	—	13	—	—	
	4	0	5	6	8	6	7	8	4	0	7	8	10	7	9	10	
	5	0	5	5	6	5	5	7	1	11	12	—	—	11	12	—	
15	1	0	11	—	—	11	—	—	5	0	6	7	9	6	7	9	16
	2	0	9	10	—	9	10	—	1	9	10	—	—	10	10	—	
	3	0	7	8	10	7	8	10	6	0	6	6	8	6	6	8	
	4	0	5	6	8	6	7	8	1	8	9	10	—	9	9	10	
	5	0	5	5	6	5	5	7	2	10	—	—	—	—	—	—	
16	1	0	12	—	—	12	—	—	7	0	7	7	7	7	7	7	16
	2	0	9	10	—	9	10	—	1	7	8	9	7	8	9	—	
	3	0	7	8	10	7	8	10	2	9	—	—	—	—	—	—	
	4	0	5	6	8	6	7	8	8	1	8	8	8	8	8	—	
	5	0	5	5	6	5	5	7	2	8	—	—	—	—	—	—	

Beispiel:

$x_1 = 1$	9	$10 = n_1$
4	2	6
$a_1 = 5$	11	$16 = N$

$p_1 \neq p_2$ wird mit
 $P < 0,05$ akzeptiert,
da für N , a_1 und x_1
 $n_1 = n_0 = 10$ gilt

Das Häufigkeitsverhältnis der beiden Alternativen wird sich von der ersten zur zweiten Untersuchung mehr oder weniger verändern. Die Intensität dieser Änderung prüft der als χ^2 -Test von McNemar [McN47] bekannte Vorzeichentest, genauer, er schöpft die Information aus, wie viele Individuen von der ersten zur zweiten Untersuchung in eine andere Kategorie übergewechselt sind. Wir haben eine Vierfeldertafel mit einem Eingang für die erste Untersuchung und mit einem zweiten Eingang für die zweite Untersuchung vorliegen:

Tabelle 7.78. Vierfeldertafel: Schema für den McNemar-Test

I. Untersuchung	II. Untersuchung	
	+	-
+	a	b
-	c	d

Die **Nullhypothese** lautet: Die Häufigkeiten in der Grundgesamtheit sind für beide Untersuchungen nicht unterschiedlich, d. h. die Nichtübereinstimmungen anzeigenden Häufigkeiten b und c zeigen nur zufällige Stichprobenschwankungen. Da diese beiden Häufigkeiten die einzig möglichen Häufigkeiten darstellen, die sich von Untersuchung I zu Untersuchung II ändern, wobei b von + nach – und c von – nach + wechselt, konnte von McNemar gezeigt werden, dass sich Änderungen dieser Art (7.222)

$$\hat{\chi}^2 = \frac{(b - c)^2}{b + c + 1} \quad FG = 1 \quad (7.222)$$

und – wenn $8 \leq (b + c) < 30$ – mit **Kontinuitätskorrektur** nach (7.223) prüfen lassen.

$$\hat{\chi}^2 = \frac{(|b - c| - 1)^2}{b + c + 1} \quad FG = 1 \quad (7.223)$$

Man vergleicht somit die Häufigkeiten b und c und prüft, ob sie eine deutliche Abweichung vom Verhältnis 1 : 1 aufweisen. Unter der Nullhypothese gilt für beide beobachteten Häufigkeiten b und c eine Erwartungshäufigkeit $(b + c)/2$. Je mehr b und c von diesem Erwartungswert abweichen, um so weniger wird man auf die Nullhypothese vertrauen. Wenn über die Richtung der zu erwartenden Änderung bereits vor Durchführung des Versuches eine begründete Annahme gemacht werden kann, darf einseitig getestet werden, z.B anhand der Tabellen 7.64 oder 7.66.

Beispiel: An einer Stichprobe von 40 Patienten wird ein Präparat mit einem Placebo (Leer- oder Scheinpräparat) verglichen ($\alpha = 0,05$). Die Patienten beginnen je zur Hälfte mit dem einen bzw. dem anderen Präparat. Zwischen beiden Therapiephasen wird eine genügend lange therapiefreie Phase eingeschaltet. Aufgrund der Aussagen der Patienten stuft der Arzt die Wirkung als „[höchstens] schwach“ oder „stark“ ein (Tabelle 7.79).



Tabelle 7.79.

		Placebo-Wirkung (Wirkg. d. Leerpräp.)	
		stark	schwach
Wirkung des Präparates	stark	8 a	16 b
	schwach	c 5	d 11

Der Nullhypothese (gleiche Wirksamkeit beider Präparate) wird die einseitige Alternativhypothese (das Präparat ist wirksamer als das Leerpräparat) gegenübergestellt, deren Überlegenheit im Test zum Ausdruck kommt.

$$\hat{\chi}^2 = \frac{(16 - 5 - 1)^2}{16 + 5 + 1} = 4,545 > 2,706 = \chi_{1;0,90}^2$$

Hiermit wird H_0 bei einseitiger Fragestellung auf dem 5%-Niveau abgelehnt. Der Wert $\hat{\chi}^2 = 4,545$ entspricht nach Tabelle 7.65 für den vorliegenden einseitigen Test einer Wahrscheinlichkeit von $P \approx 0,0165$.

In **R** kann die Berechnung des McNemar-Tests mit der Funktion `mcnemar.test()` erfolgen. Dabei wird abweichend zu (7.222) bzw. (7.223) in der Teststatistik im Nenner nur die Summe $b + c$ verwendet. Die Kontinuitätskorrektur wird durch den zusätzlichen Parameter „`correct=TRUE`“ gewählt. Der P-Wert wird nur für die zweiseitige Fragestellung berechnet.

```
> wirk; mcnemar.test(wirk, correct=TRUE)
      placebo
verum   stark schwach
stark      8      16
schwach    5      11

McNemar's Chi-squared test with continuity correction

data:  wirk
McNemar's chi-squared = 4.7619, df = 1, p-value = 0.02910
```

Betrachten wir das Beispiel noch etwas genauer: In Tabelle 7.79 sagen uns die 11 Patienten, die auf beide Präparate schwach reagiert haben, und die 8 Patienten, die in beiden Fällen eine starke Wirkung erkennen ließen, nichts über den möglichen Unterschied zwischen Präparat und Placebo. Die wesentliche Auskunft entnimmt man den Feldern b und c mit den nicht übereinstimmenden Paaren: $16 + 5 = 21$.

Bestünde zwischen den beiden Präparaten kein echter Unterschied, dann sollten wir erwarten, dass sich die Häufigkeiten b und c wie 1 : 1 verhalten. Abweichungen von diesem Verhältnis lassen sich auch mit Hilfe der Binomialverteilung prüfen. Für die einseitige Fragestellung erhalten wir

$$P(X \leq 5 | n = 21, p = 0,5) = \sum_{x=0}^{x=5} \binom{21}{x} \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{21-x} = 0,0133$$

oder anhand der Approximation $\hat{z} = \frac{|5 + 0,5 - 21 \cdot 0,5|}{\sqrt{21 \cdot 0,5 \cdot 0,5}} = 2,182$, d. h. $P(X \leq 5) \approx 0,0146$.

Die Differenz der Anteile nicht übereinstimmender Paare wird durch

$$\hat{p}_1 - \hat{p}_2 = b/n - c/n = (b - c)/n$$

geschätzt, das entsprechende angenäherte 95%-Konfidenzintervall durch

$$(b - c)/n - 1,96\sqrt{b + c}/n < \pi_1 - \pi_2 < (b - c)/n + 1,96\sqrt{b + c}/n \quad (7.224)$$

Beispiel: (Tabelle 7.79)

$$(16 - 5)/40 - 1,96\sqrt{16 + 5}/40 < \pi_1 - \pi_2 < (16 - 5)/40 + 1,96\sqrt{16 + 5}/40$$

95%-KI: $0,05 < \pi_1 - \pi_2 < 0,50$

Dieser Bereich liegt deutlich oberhalb von Null, wodurch das Resultat des Tests bestätigt wird.

Der **angenäherte 95%-KI** für den wahren Anteil $b/(b + c)$ der die b -Version bevorzugenden nicht übereinstimmenden Paare ist gegeben durch



$$\frac{\frac{b}{b+c} \pm \frac{1}{2(b+c)} \pm 1,96 \sqrt{\frac{bc}{(b+c)^3}}}{\begin{array}{cc} \textcircled{1} & \textcircled{2} \\ \text{Für } \textcircled{1} < 0,5 \\ & \geq 0,5 \end{array}} \text{ gilt vor } \textcircled{2} \text{ das „+“} \\ \text{„-“}$$
(7.225)

Beispiel:

$$\frac{16}{16+5} - \frac{1}{2(16+5)} \pm 1,96 \sqrt{\frac{16 \cdot 5}{(16+5)^3}}$$

0,762 – 0,024 ± 0,182 , d. h. 0,556 bis 0,920.



Beispiel: An 100 Personen wird geprüft, ob zwei Urlaubsländer (A, B) gleich eingeschätzt werden: für 71 + 10 = 81 Personen (Tabelle 7.80) trifft dies zu, für 19 = 3 + 16 Personen nicht. Sind 3/19 und 16/19 als nur zufällige Abweichungen von der Nullhypothese $H_0 : \pi = 1/2$ aufzufassen?

**Tabelle 7.80.**

Land B	Land A		Summe
	+	-	
+	71	3	74
-	16	10	26
Summe	87	13	100

Bei Gültigkeit von H_0 sind 3/19 und 16/19 nur zufällige Abweichungen von $\pi = 1/2$. H_0 wird verworfen, sobald das $(1 - \alpha)$ -Konfidenzintervall den Wert 1/2 nicht mit einschließt. Mit den entsprechenden Quantilen der F-Verteilung erhält man (7.226) (aus Sachs [Sac93]).

$$\frac{1}{1 + \frac{(n-c)F_{\nu_1; \nu_2; 1-\alpha/2}}{c+1}} \leq \pi \leq \frac{1}{1 + \frac{(n-c+1)}{cF_{\nu_1; \nu_2; 1-\alpha/2}}}$$

mit mit

$$\begin{array}{ll} \nu_1 = 2(n-c) & \nu_1 = 2c \\ \nu_2 = 2(c+1) & \nu_2 = 2(n-c+1) \end{array}$$
(7.226)

In (7.226) bezeichnet n die Gesamtzahl der abweichenden Bewertungen und c die Anzahl der Abweichungen eines Types (z.B. A- und B+). Für die Berechnung eines 95%-Konfidenzintervalles anhand der Daten aus Tabelle 7.80 folgt:

$$\begin{array}{ll} n = 19, & c = 3 \\ \nu_1 = 2(19-3) = 32 & \nu_1 = 2 \cdot 3 = 6 \\ \nu_2 = 2(3+1) = 8 & \nu_2 = 2(19-3+1) = 34 \\ F_{32;8;0,975} = 3,88 & F_{6;34;0,975} = 2,81 \end{array}$$

$$\frac{1}{1 + \frac{(19-3)3,88}{3+1}} \leq \pi \leq \frac{1}{1 + \frac{19-3+1}{3 \cdot 2,81}}$$

$$\text{95\%-KI: } 0,061 \leq \pi \leq 0,331$$

H_0 kann somit auf dem 5%-Niveau verworfen werden.

In ähnlicher Weise lässt sich z.B. auch der Einfluss von Werbung oder Propaganda auf den Meinungswechsel (Ja-/Nein-Sager) untersuchen (A=vorher, B=nachher). Andere Beispiele betreffen etwa die Frage, ob zwei Aufgaben (A und B), die man Kindern gestellt hat, gleich schwer zu lösen sind oder, ob bei Bewerbungen die Beurteilung durch Test (A) und Gespräch (B) übereinstimmen oder, ob zwei Pharmaka (A und B), nacheinander an denselben Personen geprüft, hinsichtlich ihrer Nebenwirkungen (+, -) übereinstimmen.

Die **praktische Relevanz des McNemar-Tests** wird von Zimmermann [Zim84] als

$$w^2 = \hat{\chi}^2 / (a + b + c + d) \quad (7.227)$$

definiert; als Richtwerte gelten:

$$\begin{aligned} w^2 &= 0,01 \text{ für „kleine praktische Relevanz“,} \\ w^2 &= 0,10 \text{ für „mittlere praktische Relevanz“ und} \\ w^2 &= 0,25 \text{ für „große praktische Relevanz“.} \end{aligned}$$

Für die Zahlen aus dem Beispiel in Tabelle 7.79 ergibt sich so: $w^2 = 4,545/40 = 0,11$.

Power und Fallzahl für den McNemar-Test

Die Power für den McNemar-Test, d.h. die Wahrscheinlichkeit dafür, die Nullhypothese berechtigt abzulehnen, kann nach (7.228) berechnet werden [CSM87].

$$z_\beta = \frac{\sqrt{n}\sqrt{p}(\psi - 1) - z_{1-\alpha/2}\sqrt{\psi + 1}}{\sqrt{(\psi + 1) - p(\psi - 1)^2}} \quad (7.228)$$

Dabei ist $n = a + b + c + d$ die Anzahl der Beobachtungen (exakter Beobachtungspaar), $p = \min(b/n, c/n)$ der kleinere Anteil an nicht übereinstimmenden Beobachtungen und $\psi = \max(b/c, c/b) > 1$ die Stärke des Unterschieds (der Asymmetrie) in der Vierfeldertafel. z_β und $z_{1-\alpha/2}$ bezeichnen die entsprechenden Quantile der Standardnormalverteilung.

Beispiel: Für die Zahlen aus dem obigem Beispiel (Tabelle 7.79) folgt entsprechend $n = 40$, $p = 5/40 = 0,125$, $\psi = 16/5 = 3,2$ und für $\alpha = 0,05$ ist $z_{0,95} = 1,96$:

$$z_\beta = \frac{\sqrt{40}\sqrt{0,125} \cdot 2,2 - 1,96\sqrt{4,2}}{\sqrt{4,2 - 0,125(2,2)^2}} = 0,476$$

Daraus folgt $\beta = 0,317$ und die Power ist entsprechend $1 - \beta = 0,683$ oder 68,3%.

In ähnlicher Weise kann aus (7.228) auch die Abschätzung für eine erforderliche Fallzahl im Rahmen der Studienplanung abgeleitet werden.

$$n = \frac{[z_{1-\alpha/2}\sqrt{\psi + 1} + z_{1-\beta}\sqrt{(\psi + 1) - p(\psi - 1)^2}]^2}{p(\psi - 1)^2} \quad (7.229)$$

Beispiel: Soll für die Fragestellung aus dem Beispiel (Tabelle 7.79) die Zahl der Fälle bestimmt werden, die notwendig ist, um einen Unterschied von $\psi = 2$ unter sonst gleichen Voraussetzungen (ggf. durch Pilotuntersuchungen zu begründen) für $\alpha = 0,05$ und $1 - \beta = 0,90$ zu testen, dann ist:

$$n = \frac{[1,96\sqrt{3} + 1,2816\sqrt{3 - 0,125 \cdot 1^2}]^2}{0,125 \cdot 1^2} = 68,69 \quad \text{d. h.} \quad n = 69.$$

Überkreuzversuch (cross-over design)

Reihenfolge-Effekte (A, B) sind durch ausbalancierte Chancengleichheit (AB, BA) bzw.

($A, B, C \dots$) durch Randomisierung zu vermeiden.

Der Überkreuzversuch ist dadurch charakterisiert, dass zwei Behandlungen (A, B) nacheinander „über Kreuz“ auf die Untersuchungseinheiten, vor allem Probanden, angewandt werden, so dass ein intraindividueller Vergleich möglich wird. Jeder Proband erhält die beiden Behandlungen in aufeinanderfolgenden Perioden [(1) und (2)], wobei ein Zufallsverfahren die Reihenfolge entscheidet. Wir unterscheiden hier nur zwei Resultate: erfolgreiche Behandlung bzw. nicht erfolgreiche Behandlung. Hierbei ist es möglich, dass der Behandlungserfolg von der Reihenfolge abhängt: A kann B hemmen oder verstärken. Kann dies ausgeschlossen werden, so prüft man nach McNemar. Ist ein Reihenfolge-Effekt nicht auszuschließen, so interessiert zunächst nur ein Vergleich von (1) anhand des Vierfelder- χ^2 -Tests. Wird weiterer Aufschluss gewünscht, so müssen beide Behandlungsfolgen gleiche Probandenzahlen aufweisen. Dann ist auch der McNemar-Test anzuwenden.

Für „Zustand gebessert“ schreiben wir „+“:

McNemar-Test

Folge		BA	
		+	-
AB	+		
	-		

Vierfelder- χ^2 -Test

Nach (1):

	+	-
A		
B		

Ansätze dieser Art werden auch bei Therapiestudien mit schnellem Wirkungseintritt angewandt, mitunter mit einer behandlungsfreien Zwischenperiode.

7.6.6 Test nach Mantel-Haenszel

Für die kombinierte Prüfung auf Unabhängigkeit mehrerer Vierfelder-Tafeln, die nach anderen Merkmalen (z.B. Geschlecht und/oder Altersstufe) geordnet (**Stratifizierung**) vorliegen, wird der Mantel-Haenszel-Test verwendet. In der Teststatistik (7.230) geht man von der folgenden allgemeinen Notation aus:

	E	$\bar{E}$	Summe
K	a_i	b_i	m_{1i}
$\bar{K}$	c_i	d_i	m_{0i}
Summe	n_{1i}	n_{0i}	n_i

K steht für eine Erkrankung und E für eine bestimmte Exposition. Der Index $i = 1, \dots, k$ läuft dann über alle Untergruppen (Strata), d.h. entsprechende Teiltabellen. Die Teststatistik ist wie folgt definiert:

$$\hat{\chi}_{MH}^2 = \frac{(A - E[A])^2}{Var[A]} \quad \text{mit}$$

$A = \text{Anzahl aller exponierten Fälle } \sum a_i$

$$E[A] = \sum_{i=1}^k \frac{n_{1i}m_{1i}}{n_i}$$

$$Var[A] = \sum_{i=1}^k \frac{n_{1i}n_{0i}m_{1i}m_{0i}}{(n_i - 1)n_i^2}$$

(7.230)

Die Mantel-Haenszel-Teststatistik $\hat{\chi}_{MH}^2$ entspricht somit einem „gewogenen Durchschnitt“ mehrerer Vierfeldertafeln und wird mit dem entsprechenden Quantil der χ^2 -Verteilung mit einem Freiheitsgrad ($\chi_{1;1-\alpha}$) geprüft. Eine **Kontinuitätskorrektur** der Teststatistik kann dadurch erreicht werden, dass im Zähler die Differenz zum Erwartungswert um 0,5 verringert wird. (7.230) lässt sich auch anders schreiben (hier mit Kontinuitätskorrektur):

$$\hat{\chi}_{MH}^2 = \frac{\left[\left| \sum_{i=1}^k a_i - [(a_i + b_i)(a_i + c_i)/n_i] \right| - 0,5 \right]^2}{\sum_{i=1}^k \frac{(a_i + b_i)(c_i + d_i)(a_i + c_i)(b_i + d_i)}{(n_i - 1)n_i^2}} \quad (7.231)$$



Beispiel (mit $k = 2$ Gruppen nach dem Geschlecht):

gesamt			männlich $i = 1$			weiblich $i = 2$		
35	141	176	15	85	100	20	56	76
11	128	139	4	77	81	7	51	58
46	269	315	19	162	181	27	107	134
$\hat{\chi}^2 = 8,899$			$\hat{\chi}_m^2 = 4,822$			$\hat{\chi}_w^2 = 4,119$		

$$\hat{\chi}_{MH}^2 = \frac{[|(15 \cdot 77 - 85 \cdot 4)/181 + (20 \cdot 51 - 56 \cdot 7)/134| - 0,5]^2}{\left[\frac{100 \cdot 81 \cdot 19 \cdot 162}{180 \cdot 181^2} \right] + \left[\frac{76 \cdot 58 \cdot 27 \cdot 107}{133 \cdot 134^2} \right]} = \frac{75,5045}{9,5604} = 7,898$$

Mit $\hat{\chi}_{MH}^2 = 7,898 > 2,841 = \chi_{1;0,95}^2$ liegt ein signifikanter, hinsichtlich des Einflusses des Geschlechtes gewichteter, Zusammenhang vor. Der Vermengungseffekt (Confounding) durch das Geschlecht wird im Beispiel durch die hier geringen Unterschiede der χ^2 -Statistik in den Einzelstabellen im Vergleich zur Gesamttafel deutlich.

In **R** kann die Mantel-Haenszel-Statistik durch die Funktion **mantelhaen.test()** berechnet werden. Die Häufigkeiten müssen dazu in der Form einer 3-dimensionalen Tabelle durch die Funktion **array()** bereitgestellt werden. Als Standard wird in **R** die Statistik mit Kontinuitätskorrektur gerechnet.

```
> tab <- array(c(15, 4, 85, 77, 20, 7, 56, 51), dim = c(2, 2, 2),
+             dimnames = list(A = c("I", "II"), B = c("+", "-"),
+                               Geschl = c("maennl", "weibl"))); tab
Geschl = maennl

      B
A      +  -
I      15 85
II     4 77

Geschl = weibl

      B
A      +  -
I      20 56
II     7 51

> mantelhaen.test(tab, alternative="two.sided", correct=TRUE)

Mantel-Haenszel chi-squared test with continuity correction
```

```
data: tab
Mantel-Haenszel X-squared = 7.8977, df = 1, p-value = 0.00495
alternative hypothesis: true common odds ratio is not equal to 1
95 percent confidence interval: 1.410224 6.016843
sample estimates: common odds ratio 2.912919
```

Die Berechnung einer entsprechend adjustierten Odds Ratio (in der Funktion **mantelhaen.test()** integriert) und eines zugehörigen (testbasierten) Konfidenzintervalls kann auch nach (7.232) und (7.233) erfolgen.

$$\hat{\omega}_{MH} = \frac{\sum_{i=1}^k \frac{a_i d_i}{n_i}}{\sum_{i=1}^k \frac{b_i c_i}{n_i}} \quad (7.232)$$

$$(1 - \alpha)\text{-Konfidenzintervall: } \hat{\omega}_{MH}^{1 \pm z_{1-\alpha/2}/\sqrt{\hat{\chi}^2}} \quad (7.233)$$

Mit den Zahlen aus obigem Beispiel folgt für die Berechnung einer adjustierten Odds Ratio mit dem 95%-Konfidenzintervall:

$$\hat{\omega}_{MH} = \frac{15 \cdot 77/181 + 20 \cdot 51/134}{4 \cdot 85/181 + 7 \cdot 56/134} = 2,912919$$

und

$$2,913^{1 \pm 1,96/\sqrt{7,898}} \Rightarrow [95\% - KI : 1,382 \leq \hat{\omega}_{MH} \leq 6,140]$$

7.6.6.1 Die Kombination von Vierfeldertafeln

Geprüft wird die Nullhypothese der stochastischen Unabhängigkeit in k vergleichbaren Vierfeldertafeln mit fast einheitlicher Tendenz.

Liegen mehrere Vierfeldertafeln vor, die nicht als Wiederholungen aufgefasst werden können, da sich von Tafel zu Tafel die Bedingungen für die jeweils gemeinsam betrachteten Stichproben n_1 und n_2 ($n_1 + n_2 = n$) ändern, dann empfiehlt Cochran [Coc54] die beiden folgenden Verfahren als ausreichend genaue Näherungslösungen (vgl. auch den praktisch gleichwertigen Mantel-Haenszel-Test [7.6.6]).

Ansatz I. Die Stichprobenumfänge n_i der k Vierfeldertafeln ($i = 1, \dots, k$) unterscheiden sich nicht sehr stark voneinander (höchstens um den Faktor 2); die Anteile $a/(a+b)$ und $c/(c+d)$ (Tabelle 7.62) liegen für alle Tafeln im Bereich von etwa 20% bis 80%. Dann lässt sich die Frage nach der Bedeutsamkeit einer Aussage aufgrund von k kombinierten Vierfeldertafeln anhand der Standardnormalverteilung nach

$$\hat{z} = \frac{\sum \hat{\chi}}{\sqrt{k}} \quad (7.234)$$

testen. Die Prüfung im einzelnen:

- (1) Aus den nur nach (7.197) oder (7.198) für die k Vierfeldertafeln ermittelten $\hat{\chi}^2$ -Werten die Quadratwurzel ziehen.
- (2) Die Vorzeichen dieser Werte sind durch die Vorzeichen der Differenzen $a/(a+b) - c/(c+d)$ gegeben.
- (3) Die Summe der $\hat{\chi}$ -Werte bilden (Vorzeichen beachten!).
- (4) Aus der Anzahl der kombinierten Vierfeldertafeln die Quadratwurzel ziehen.
- (5) Nach obiger Formel den Quotienten $\hat{z}$ bilden.
- (6) Die Bedeutsamkeit von $\hat{z}$ anhand der Quantile der Standardnormalverteilung prüfen..

Auf ein Beispiel wird verzichtet.

Ansatz II. Hinsichtlich der Stichprobenumfänge n_i der k Vierfeldertafeln und der jeweiligen Anteile $a/(a+b)$ und $c/(c+d)$ werden keinerlei Voraussetzungen gemacht. Hier lässt sich die Frage nach der Bedeutsamkeit einer Aussage anhand der Standardnormalverteilung nach

$$\hat{z} = \frac{\sum W_i \cdot D_i}{\sqrt{\sum W_i \cdot p_i(1-p_i)}} \quad (7.235)$$

prüfen. Hierin bedeuten: W_i = das „Gewicht“ der i -ten Stichprobe mit den Häufigkeiten a_i, b_i, c_i und d_i (Tabelle 7.62), definiert als $W_i = (n_{i1} \cdot n_{i2})/n_i$,

wobei $n_{i1} = a_i + b_i$; $n_{i2} = c_i + d_i$ und $n_i = n_{i1} + n_{i2}$;

p_i = der durchschnittliche Anteil, gegeben durch $p_i = (a_i + c_i)/n_i$

und D_i = die Differenz zwischen den Anteilen: $D_i = a_i/n_{i1} - c_i/n_{i2}$.

Zur Illustration geben wir das von Cochran zitierte Beispiel.

Tabelle 7.81. Daten zur Neugeborenen-Erythroblastose; die Stichprobenumfänge variieren zwar nur von 33–60, die Anteile der Verstorbenen jedoch von 3% bis 46%, so dass die 4 Tafeln nach dem zweiten Verfahren kombiniert werden

Symptome	Geschlecht des Blut-spenders	Anzahl der Verstor-benen	Überle-benden	Ins-gesamt	Anteil der Verstorbenen in Prozent
keine	männlich	2	21	23 = n_{11}	8,7 = p_{11}
	weiblich	0	10	10 = n_{12}	0,0 = p_{12}
	Insgesamt	2	31	33 = n_1	6,1 = p_1
geringe	männlich	2	40	42 = n_{21}	4,8 = p_{21}
	weiblich	0	18	18 = n_{22}	0,0 = p_{22}
	Insgesamt	2	58	60 = n_2	3,3 = p_2
mäßige	männlich	6	33	39 = n_{31}	15,4 = p_{31}
	weiblich	0	10	10 = n_{32}	0,0 = p_{32}
	Insgesamt	6	43	49 = n_3	12,2 = p_3
starke	männlich	17	16	33 = n_{41}	51,5 = p_{41}
	weiblich	0	4	4 = n_{42}	0,0 = p_{42}
	Insgesamt	17	20	37 = n_4	45,9 = p_4



Beispiel: Die Neugeborenen-Erythroblastose beruht auf der Unverträglichkeit zwischen *rh*-negativem mütterlichen und *Rh*-positivem embryonalen Blut, die u. a. zur Zerstörung embryonaler Erythrozyten führt, ein Prozess, der nach der Geburt durch Austauschtransfusion behandelt wird: Das Blut des Kindes wird durch gruppengleiches *rh*-negatives Spenderblut ersetzt.

An 179 Neugeborenen einer Bostoner Klinik (Allen, Diamond and Watrous: The New Engl. J. Med. **241** [1949] 799–806) ist beobachtet worden, dass das Blut weiblicher Spender von den Kindern besser vertragen wird als das männlicher Spender (Tabelle 7.81). Es soll geprüft werden, ob ein Zusammenhang zwischen dem Geschlecht des Blutspenders und der Alternative Sterben oder Überleben nachweisbar ist. Die 179 Fälle konnten wegen der unterschiedlichen Symptomatik nicht als einheitlich angesehen werden. So wurden sie nach der Schwere der Symptome als einer möglicherweise intervenierenden Variablen in 4 in sich homogenere Gruppen geteilt. Die Ergebnisse sind in Tabelle 7.81 zusammengefasst.

Anhand einer Hilfstafel mit p_i in % und $H = 100$

Symptome	D_i	p_i	$p_i(H - p_i)$	$w_i = \frac{n_{i1} \cdot n_{i2}}{n_i}$	$w_i D_i$	$w_i p_i (H - p_i)$
keine	$8,7 - 0,0 = 8,7$	6,1	573	7,0	60,90	4011,0
geringe	$4,8 - 0,0 = 4,8$	3,3	319	12,6	60,48	4019,4
mäßige	$15,4 - 0,0 = 15,4$	12,2	1071	8,0	123,20	8568,0
starke	$51,5 - 0,0 = 51,5$	45,9	2483	3,6	185,40	8938,8
					429,98	25537,2
					$\sum w_i D_i$	$\sum w_i p_i (H - p_i)$

erhalten wir $\hat{z} = \frac{429,98}{\sqrt{25\,537,2}} = 2,69$.

Bei der vorliegenden zweiseitigen Fragestellung entspricht diesem $\hat{z}$ -Wert eine Irrtumswahrscheinlichkeit von 0,0072. Wir dürfen also darauf vertrauen, dass männliche Blutspender bei fetaler Erythroblastose weniger geeignet sind als weibliche – eine Tendenz, die sich vor allem bei stärker hervortretenden Symptomen auswirkt.

Nebenbei bemerkt sei, dass sich dieses Ergebnis durch andere Autoren nicht bestätigen ließ: Das Geschlecht des Blutspenders ist ohne Einfluss auf die Prognose der fetalen Erythroblastose.

7.6.7 Der $k \cdot 2$ -Felder- χ^2 -Test nach Brandt und Snedecor

Der Informationsgehalt von Häufigkeiten ist gering. Trotzdem bietet die Analyse von Vierfeldertafeln eine Reihe von Möglichkeiten. Wir können diese einfachste Zweifeldertafel auf Unabhängigkeit, Trend und Symmetrie prüfen. In diesem Kapitel werden diese und andere Prüfungen an Kontingenztafeln beschrieben, die für jedes der beiden Merkmale nicht nur eine Alternative, also 2 Klassifikationsmöglichkeiten, sondern mehrere aufweisen. Beispielsweise lassen sich Besitzer von Führerscheinen hinsichtlich der Altersgruppe und der Anzahl der Unfälle – 0, 1, 2, mehr als 2 – vergleichen. Andere Vergleichspaare, die ebenfalls zu Zweifeldertafeln führen, sind z. B. Schulbildung und Einkommen, Körperbautyp von Eheleuten sowie die Beurteilung des Eheglücks durch beide Partner. Ebenso wie eine Stichprobe nach zwei Merkmalsreihen kombiniert aufgeteilt auf Unabhängigkeit geprüft werden kann, lassen sich eine Reihe von Stichproben mit zwei oder mehr Ausprägungen auf Gleichartigkeit oder Homogenität testen.

Mit dem Vierfelder- χ^2 -Test lassen sich 2 Stichproben von Alternativdaten daraufhin untersuchen, ob sie als Zufallsstichproben aus einer durch die vier Randsummen repräsentierten Grundgesamtheit stammen. Vergleichen wir nun mehrere sagen wir k Stichproben von Alternativdaten miteinander, wobei natürlich nur die zweiseitige Fragestellung möglich ist, so erhalten wir als Ausgangsschema eine $k \cdot 2$ -Tafel der folgenden Art (siehe die Tabelle 7.82).

Dabei sei angenommen, dass x kleiner als $n - x$ ist (Tabelle 7.82, 1. Spalte „Stichprobe“). Die Nullhypothese lautet: Der Anteil des Merkmals „+“ ist in den k Grundgesamtheiten gleich. In den k Stichproben wird er durch x/n geschätzt. Unter der Nullhypothese ist für die $k \cdot 2$ -Felder der Tabelle eine zu den Randsummen weitgehend proportionale Häufigkeitsverteilung zu erwarten. Anhand des $k \cdot 2$ -Felder- χ^2 -Tests wird somit geprüft, ob die relativen Häufigkeiten in den k Klassen mit der über alle k Klassen berechneten durchschnittlichen relativen Häufigkeit übereinstimmen. Vorausgesetzt werden n unabhängige Beobachtungen sowie sich gegenseitig ausschließende und die beobachtete Mannigfaltigkeit erschöpfende Alternativen.

Tabelle 7.82 Schema zur $k \cdot 2$ -Tafel (Homogenitätstest)

Stichprobe oder [2. Merkmal]	Merkmal		Σ
	+	–	
1	x_1	$n_1 - x_1$	n_1
2	x_2	$n_2 - x_2$	n_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$
j	x_j	$n_j - x_j$	n_j
$\vdots$	$\vdots$	$\vdots$	$\vdots$
k	x_k	$n_k - x_k$	n_k
Σ	x	$n - x$	n

Zur Entscheidung über Beibehaltung oder Ablehnung der Nullhypothese der Homogenität (Gleichheit) **k binomialer Grundgesamtheiten** dient der χ^2 -Test (Karl Pearson 1857–1936). Wir benutzen dabei die Formel von Brandt und Snedecor:

$$\hat{\chi}^2 = \frac{n^2}{x(n-x)} \left[\sum_{j=1}^k \frac{x_j^2}{n_j} - \frac{x^2}{n} \right] \quad \text{mit } FG = k - 1 \quad (7.236)$$



An dieser Stelle sei noch einmal auf den Unterschied zwischen dem tabelliert vorliegenden χ^2 -Wert und dem nach einer Formel berechneten Wert der Prüfgröße $\hat{\chi}^2$ aufmerksam gemacht. Nur bei Gültigkeit der Nullhypothese sowie für großes n und große Erwartungshäufigkeiten stimmen beide überein. Die Approximation ist bei nicht zu schwach besetzten Feldern ausreichend. Als Maß der Besetzung einer $k \cdot 2$ - oder Mehrfeldertafel dienen die bei Annahme der Homogenität zu erwartenden Erwartungshäufigkeiten. Sie werden berechnet als Quotient aus dem Produkt der Randsummen und dem Gesamtstichprobenumfang (vgl. Tabelle 7.82: Die Erwartungshäufigkeit E für das Feld x_j beträgt $E(x_j) = n_j x / n$).

Für kleine $k \cdot 2$ -Feldertafeln ($k < 5$) müssen alle Erwartungshäufigkeiten mindestens gleich 2 sein; stehen wenigstens 4 Freiheitsgrade zur Verfügung ($k \geq 5$), dann sollten alle Erwartungshäufigkeiten ≥ 1 sein (Lewontin und Felsenstein [LF65]). Lassen sich diese Forderungen nicht erfüllen, dann muss die Tafel durch Zusammenfassung unterbesetzter Felder vereinfacht werden.



Beispiel: In einer Epidemie seien insgesamt 80 Personen behandelt worden. Eine Gruppe von 40 Kranken erhielt eine Standarddosis eines neuen spezifischen Mittels. Die andere Gruppe von 40 Kranken sei nur symptomatisch behandelt worden (Behandlung der Krankheitserscheinungen, nicht aber ihrer Ursachen) (Quelle: Martini [Mar53] S. 83, Tab. 14). Das Resultat der Behandlung

wird ausgedrückt in Besetzungszahlen für drei Klassen: schnell geheilt, langsam geheilt, gestorben.

Therapeutischer Erfolg	Therapie		insgesamt
	symptomatisch	spezifisch	
geheilt in x Wochen	14	22	36
geheilt in $x + y$ Wochen	18	16	34
gestorben	8	2	10
insgesamt	40	40	80

- Nullhypothese: Die therapeutischen Ergebnisse sind für beide Therapieformen gleich.
- Alternativhypothese: Die therapeutischen Ergebnisse sind $\dots$ nicht gleich.
- Signifikanzniveau: $\alpha = 0,05$ (zweiseitig).
- Testwahl: Es kommt nur der $k \cdot 2$ -Felder- χ^2 -Test in Frage (Vergleich der Erwartungshäufigkeiten in der dritten Zeile:

$$x_k = 8, n_k - x_k = 2, E(2) = \frac{10 \cdot 40}{80} = 5 > 2 \quad \text{sowie} \quad E(8) = \frac{10 \cdot 40}{80} = 5 > 2 \quad).$$

- Ergebnisse und Auswertung:

$$\hat{\chi}^2 = \frac{80^2}{40 \cdot 40} \left[\left(\frac{14^2}{36} + \frac{18^2}{34} + \frac{8^2}{10} \right) - \frac{40^2}{80} \right] = 5,495$$

- Entscheidung: Da $\hat{\chi}^2 = 5,495 < 5,99 = \chi_{2;0,95}^2$, können wir die Nullhypothese nicht ablehnen. Ist auf dem gewünschten Niveau die H_0 nicht abzulehnen, so interessiert meist der zugehörige P -Wert, das heisst hier $P \approx 0,07$.
- Interpretation: Aufgrund der vorliegenden Stichprobe lässt sich ein Unterschied zwischen den beiden Therapieformen auf dem 5%-Niveau nicht sichern.
- Bemerkung: Interessiert ein Vergleich der mittleren therapeutischen Erfolge beider Therapien, dann prüfe man nach (7.238).

In **R** kann die Berechnung der Teststatistik mit der allgemeinen Funktion **chisq.test()** erfolgen. Mit den Angaben aus dem Beispiel folgt:

```
> erfolg <- matrix(c(14, 22, 18, 16, 8, 2), nr=3, byrow=T,
+   dimnames = list(heilung=c("geheilt-x", "geheilt-x+y", "gestorben"),
+   therapie=c("symptomatisch", "spezifisch")))
> erfolg
      therapie
heilung symptomatisch spezifisch
geheilt-x           14          22
geheilt-x+y         18          16
gestorben            8           2
> chisq.test(erfolg, correct = TRUE)

      Pearson's Chi-squared test

data:  erfolg
X-squared = 5.4954, df = 2, p-value = 0.06407
```

Dieses Resultat hätte man natürlich auch nach dem Ansatz eines χ^2 -Anpassungstests (7.25) erhalten, wobei die bei Annahme der Nullhypothese auf Homogenität oder Unabhängigkeit zu erwartenden Häufigkeiten E als Quotienten des Produktes der Randsummen der Tafel und des Gesamtstichprobenumfanges ermittelt werden. So ist z. B. in der Tabelle zu unserem Beispiel links oben die beobachtete Häufigkeit $B = 14$, die zugehörige Erwartungshäufigkeit ist dann $E = 36 \cdot 40/80 = 18$. Bildet man für jedes Feld der $k \cdot 2$ -Tafel den Quotienten $(B - E)^2/E$ und addiert die einzelnen $k \cdot 2$ Quotienten, dann erhält man wieder $\hat{\chi}^2$. Zum besseren Verständnis werden die Daten aus den vorangehenden Beispiel in der folgenden Tabelle aufgeschlüsselt:

therapeutischer Erfolg Berechnung von χ^2		Therapie		Insgesamt
		symptomatisch	spezifisch	
geheilt in x Wochen				
beobachtet	B	14	22	36
erwartet	E	18,00	18,00	36,00
Abweichung	$B - E$	-4,00	4,00	0,00
quadriert	$(B - E)^2$	16,00	16,00	
Chi-Quadrat	$\frac{(B - E)^2}{E}$	0,8889	0,8889	1,7778
geheilt in $x + y$ Wochen				
beobachtet	B	18	16	34
erwartet	E	17,00	17,00	36,00
Abweichung	$B - E$	1,00	-1,00	0,00
quadriert	$(B - E)^2$	1,00	1,00	
Chi-Quadrat	$\frac{(B - E)^2}{E}$	0,0588	0,0588	0,1176
gestorben				
beobachtet	B	8	2	10
erwartet	E	5,00	5,00	10,00
Abweichung	$B - E$	3,00	-3,00	0,00
quadriert	$(B - E)^2$	9,00	9,00	
Chi-Quadrat	$\frac{(B - E)^2}{E}$	1,8000	1,8000	3,6000
Insgesamt	$B = E$	40	40	80
χ^2 -Spaltensumme:		2,7477	2,7477	5,4954

Besonders erwähnt sei noch, dass jeder Beitrag zum $\hat{\chi}^2$ -Wert relativ zur Erwartungshäufigkeit E erfolgt: Eine große Abweichung $B - E$ mit großer E steuert etwa einen gleichen Betrag zu $\hat{\chi}^2$ bei wie eine kleine Abweichung mit kleiner E :

$$\text{vgl. z. B. } \frac{(15 - 25)^2}{25} = 4 = \frac{(3 - 1)^2}{1}.$$

Angenäherte 95%-Konfidenzintervalle für $p_i - p_{i'}$

zum $k \cdot 2$ -Felder- χ^2 -Homogenitätstest; es liegen k Zufallsstichproben vor und die Beobachtungen sind Anteilswerte dichotomer Grundgesamtheiten mit den Parametern p_i und $p_{i'}$.

Für nicht zu kleine Werte n_i und $n_{i'}$ mit den Anteilen $x_i/n_i = \hat{p}_i$ und $x_{i'}/n_{i'} = \hat{p}_{i'}$ lassen sich für die Differenzen $\hat{p}_i - \hat{p}_{i'}$ angenäherte simultane 95%-Konfidenzintervalle angeben:

$$\begin{array}{c}
 (\hat{p}_i - \hat{p}_{i'}) \pm \sqrt{\chi_{k-1;0,95}^2 \left[\frac{\hat{p}_i(1 - \hat{p}_i)}{n_i} + \frac{\hat{p}_{i'}(1 - \hat{p}_{i'})}{n_{i'}} \right]} \\
 \text{bzw.} \\
 \begin{array}{c|c|c}
 \vdots & \vdots & \vdots \\
 n_{1i} & n_{2i} & n_{\cdot i} \left[\frac{n_{1i}}{n_{1\cdot}} - \frac{n_{2i}}{n_{2\cdot}} \right] \\
 \vdots & \vdots & \vdots \\
 n_{1\cdot} & n_{2\cdot} & n_{\cdot\cdot}
 \end{array} \pm \sqrt{\chi_{k-1;0,95}^2 \left[\frac{n_{1i}(n_{1\cdot} - n_{1i})}{n_{1\cdot}^3} + \frac{n_{2i}(n_{2\cdot} - n_{2i})}{n_{2\cdot}^3} \right]}
 \end{array} \quad (7.237)$$

Schließt ein 95%-Konfidenzintervall die Null mit ein, so lässt sich $H_0: p_i = p_{i'}$ auf dem 5%-Niveau nicht ablehnen, andernfalls ist $H_A: p_i \neq p_{i'}$ auf diesem Niveau zu akzeptieren.

Tabelle 7.83. Modell einer $k \cdot 3$ -Felder-Tabelle

Gruppe	1	2	...	k	Summe
Kategorie 1	n_{11}	n_{21}		n_{k1}	$n_{.1}$
Kategorie 2	n_{12}	n_{22}		n_{k2}	$n_{.2}$
Kategorie 3	n_{13}	n_{23}		n_{k3}	$n_{.3}$
Summe	$n_{1.}$	$n_{2.}$		$n_{k.}$	$n_{..}$

Für den entsprechenden $k \cdot 3$ -Felder- χ^2 -Homogenitätstest (vgl. Tabelle 7.83) und einander entsprechenden Anteilsdifferenzen $\hat{p}_i - \hat{p}_{i'}$, etwa $\hat{p}_1 = n_{12}/n_{1.}$ und $\hat{p}_2 = n_{22}/n_{2.}$, ergeben sich die entsprechenden simultanen 95%-Konfidenzintervalle nach (7.237) indem $\chi^2_{k-1;0,95}$ durch $\chi^2_{2(k-1);0,95}$ ersetzt wird. Der H_0 : „die jeweils drei Anteile in den k Gruppen entsprechen einander“ wird die H_A : „nicht alle drei Anteile in den k Gruppen entsprechen einander“ gegenübergestellt. Entschieden wird wie im Fall (7.237). Der χ^2 -Test für Tabelle 7.83 erfolgt nach [7.6.9].

Scoring I (Homogenitätstest)

Im hier vorliegenden Zweistichprobenfall wird die Nullhypothese: Gleichheit der den beiden Stichproben zugrundeliegenden Mittelwerte für $\hat{z} > z_\alpha$ auf dem $100\alpha\%$ -Niveau verworfen, sobald $n_1 \approx n_2$ und $n_1 + n_2 > 14$:

$$\hat{z} = \frac{|\left(\sum B_1 x / n_1 - \sum B_2 x / n_2\right) - n / (2n_1 n_2)|}{\sqrt{\left[\frac{\sum Bx^2 - (\sum Bx)^2 / n}{n_1 + n_2 - 1}\right] \cdot \left[\frac{n_1 + n_2}{n_1 n_2}\right]}} \quad (7.238)$$

Die Kontinuitätskorrektur $n/2n_1 n_2$ gilt für den Fall des Scoring (vgl. auch Seite 522) mit Schrittweite 1, mit Schrittweite c lautet sie $cn/2n_1 n_2$. Für das oben genannte Beispiel ergibt sich ($\alpha = 0,05$):

B_1	B_2	B	Score x	$B_1 x$	$B_2 x$	Bx	Bx^2
14	22	36	1	14	22	36	36
18	16	34	0	0	0	0	0
8	2	10	-1	-8	-2	-10	10
40	40	80	-	6	20	26	46

$$\hat{z} = \frac{|6/40 - 20/40| - 80/(2 \cdot 40 \cdot 40)}{\sqrt{\left[\frac{46 - 26^2/80}{40 + 40 - 1}\right] \left[\frac{40 + 40}{40 \cdot 40}\right]}} = 2,108 > 1,960 \quad \text{eine Ablehnung der } H_0 \quad \text{auf dem 5\%-Niveau.}$$

Der Wert für die Verteilungsfunktion der Standardnormalverteilung ist $P'(Z \leq 2,108) = 0,0175$, so dass für die vorliegende zweiseitige Fragestellung ein $P = 0,035 < 0,05$ resultiert.

Zerlegung der Freiheitsgrade einer $k \cdot 2$ -Feldertafel

Für die $k \cdot 2$ -Feldertafel bezeichnen wir die Häufigkeiten, besser die Besetzungszahlen, entsprechend dem folgenden gegenüber Tabelle 7.82 erweiterten Schema (Tabelle 7.84). Es gestattet den

Tabelle 7.84. Schema zur Zerlegung der Freiheitsgrade in einer $k \cdot 2$ -Feldertafel

Stichprobe	Merkmal		Insgesamt	Anteil
	+	-		
1	x_1	$n_1 - x_1$	n_1	$p_1 = x_1/n_1$
2	x_2	$n_2 - x_2$	n_2	$p_2 = x_2/n_2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
j	x_j	$n_j - x_j$	n_j	$p_j = x_j/n_j$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
k	x_k	$n_k - x_k$	n_k	$p_k = x_k/n_k$
Insgesamt	x	$n - x$	n	
$\hat{p} = x/n$				

direkten Vergleich der Erfolgsprozeentsätze – den Anteil der Plus-Merkmalsträger an dem jeweiligen Stichprobenumfang – für alle Stichproben. Die Formel für den χ^2 -Test nach Brandt-Snedecor lautet dann

$$\hat{\chi}^2 = \frac{\sum_{j=1}^k x_j p_j - x \hat{p}}{\hat{p}(1 - \hat{p})} \quad (7.239)$$

mit $FG = k - 1$.

Hierin bedeuten:

x = Gesamtzahl der Stichprobenelemente mit dem Merkmal „+“,

x_j = Besetzungszahl des Merkmals „+“ in der Stichprobe j ,

$\hat{p}$ = Quotient aus x und n ; der aus dem Gesamtstichprobenumfang ermittelte Anteil der Merkmalsträger („+“).

Unter der Nullhypothese: alle Stichproben entstammen Grundgesamtheiten mit π (= konstant), geschätzt durch $\hat{p} = x/n$, erwarten wir auch hier für alle Stichproben eine diesem Verhältnis entsprechende Häufigkeitsverteilung.

Die Brandt-Snedecor-Formel (7.236) gilt nicht nur für den gesamten auf Homogenität zu prüfenden Stichprobenumfang von k Stichproben, sondern natürlich auch für jeweils zwei (d. h. $FG = 1$) oder mehr – sagen wir j (mit $FG = j - 1$) – Stichproben, die als Gruppe aus den k Stichproben ausgewählt werden. Auf diese Weise gelingt es, die $k - 1$ Freiheitsgrade in Komponenten $\{1 + (j - 1) + (k - j - 1) = k - 1\}$ zu zerlegen (Tabelle 7.85).

Tabelle 7.85. Zerlegung der Freiheitsgrade nach den Komponenten von $\hat{\chi}^2$

Komponenten von $\hat{\chi}^2$	Freiheitsgrade
Unterschiede zwischen p 's zweier Stichprobengruppen mit n_1 und n_2 ($n = n_1 + n_2$)	1
Variation innerhalb der p 's in den ersten j Reihen	$j - 1$
Variation innerhalb der p 's in den letzten $k - j$ Reihen	$k - j + 1$
Gesamt- $\hat{\chi}^2$	$k - 1$

Anders ausgedrückt: Das Gesamt- $\hat{\chi}^2$ wird in Anteile zerlegt. Damit ist ein Test gegeben, der auf eine Veränderung des p -Niveaus innerhalb einer Stichprobenfolge von Alternativdaten anspricht.

**Beispiel:**

Nr.	x_j	$n_j - x_j$	n_j	$p_j = x_j/n_j$	$x_j p_j$
1	10	10	20	0,50	5,00
2	8	12	20	0,40	3,20
3	9	11	20	0,45	4,05
4	5	15	20	0,25	1,25
5	6	14	20	0,30	1,80
$\sum$	38	62	100		15,30
$\hat{p} = 38/100 = 0,38$					
Nr.	Gruppe	x_i	n_i	$p_i = \bar{p}$	$x_i p_i$
1+2+3	n_1	27	60	0,450	12,150
4+5	n_2	11	40	0,275	3,025
$\sum$	n	38	100		15,175

$$\hat{\chi}^2\text{-Gesamtabweichungen der } p\text{'s von } \hat{p} = \frac{15,300 - 38 \cdot 0,380}{0,380 \cdot 0,620} = 3,650$$

 $\hat{\chi}^2\text{-Unterschiede zwischen den } \bar{p}\text{'s von } n_1 \text{ und } n_2$

$$\frac{15,175 - 38 \cdot 0,380}{0,380 \cdot 0,620} = 3,120$$

 $\hat{\chi}^2\text{Variation zwischen den } p\text{'s innerhalb von } n_1$

$$\frac{12,250 - 27 \cdot 0,450}{0,380 \cdot 0,620} = 0,424$$

 $\hat{\chi}^2\text{Variation zwischen den } p\text{'s innerhalb von } n_2$

$$\frac{3,050 - 11 \cdot 0,275}{0,380 \cdot 0,620} = 0,106$$

Diese Komponenten werden zusammengefasst und anhand von Tabelle 7.89 beurteilt. Für $\tau = 3$ Komponenten ist auf dem 10%-Niveau die entsprechende χ^2 -Schranke für $\nu = 1$ Freiheitsgrad durch den Wert 4,529 gegeben, für $\nu = 2$ durch den Wert 6,802; beide Werte werden deutlich unterschritten, was auch das Gesamt- $\hat{\chi}^2$ erwarten ließ.

Variationskomponenten	$\hat{\chi}^2$	FG	P-Wert
Unterschiede zwischen den p 's der Stichprobengruppen n_1 (=Nr. 1-3) und n_2 (=Nr. 4+5)	3,120	1	$P > 0,10$
Variation zwischen den p 's innerhalb von n_1	0,424	2	$P > 0,10$
Variation zwischen den p 's innerhalb von n_2	0,106	1	$P > 0,10$
Gesamtabweichungen der p 's von $\hat{p}$ in $n = n_1 + n_2$	3,650	4	$0,40 < P < 0,50$

Für die Prüfung von Teilabhängigkeiten bzw. Teilhomogenitäten gibt Kimball [Kim54] einfache Formeln. Das auf $(r-1)(c-1) = \nu$ Freiheitsgraden (r Anzahl der Zeilen, c Anzahl der Spalten) basierende $\hat{\chi}^2$ der $r \cdot c$ -Tabelle wird durch die Auswahl einzelner oder Zusammenfassung benachbarter Felder in jeweils ν Vierfelder- $\hat{\chi}^2$ -Komponenten mit einem Freiheitsgrad zerlegt. Für die Prüfung der Teilhypothesen werden die Schranken der Tabelle 7.89 verwendet.

Modell I:

a_1	a_2	a_3	A
b_1	b_2	b_3	B
n_1	n_2	n_3	N

$$\hat{\chi}_1^2 = \frac{N^2[a_1b_2 - a_2b_1]^2}{ABn_1n_2(n_1 + n_2)}$$

$$\hat{\chi}_2^2 = \frac{N[b_3(a_1 + a_2) - a_3(b_1 + b_2)]^2}{ABn_3(n_1 + n_2)}$$

$$\frac{a_1}{b_1} \mid \frac{a_2}{b_2}$$

$$\frac{a_1}{b_1} \quad \frac{a_2}{b_2} \mid \frac{a_3}{b_3}$$



Beispiel:

	A	B	C	Summe
+	17	17	12	46
-	36	24	60	12
Summe	53	41	72	166

$$\hat{\chi}^2 = \frac{166^2}{46 \cdot 120} \left[\left(\frac{17^2}{53} + \frac{17^2}{41} + \frac{12^2}{72} \right) - \frac{46^2}{166} \right] = 8,759 > 5,991 = \chi_{2;0,95}^2$$

Die Zerlegung nach Modell I führt auf:

	A	B	Summe			A+B	C	Summe
+	17	17	34	und	+	34	12	46
-	36	24	60		-	60	60	120
Summe	53	41	94		Summe	94	72	166

$$\hat{\chi}_1^2 = \frac{166^2(17 \cdot 24 - 17 \cdot 36)^2}{46 \cdot 120 \cdot 53 \cdot 41(53 + 41)} = 1,017 < 5,0245 = \chi_{1;\tau=2;0,05}^2$$

$$\hat{\chi}_2^2 = \frac{166[60(17 + 17) - 12(36 + 24)]^2}{46 \cdot 120 \cdot 72(53 + 41)} = 7,742 > 5,024 = \chi_{1;\tau=2;0,05}^2$$

$$= 8,759$$

Somit sind nur für die zweite Vierfeldertafel partielle Abweichungen von der Proportionalität (Unabhängigkeit bzw. Homogenität) auf dem 5%-Niveau nachzuweisen.

Modell II:

a_1	a_2	a_3	a_4	A
b_1	b_2	b_3	b_4	B
n_1	n_2	n_3	n_4	N

$$\hat{\chi}_1^2 = \frac{N^2[a_1b_2 - a_2b_1]^2}{ABn_1n_2(n_1 + n_2)}$$

$$\hat{\chi}_2^2 = \frac{N^2[b_3(a_1 + a_2) - a_3(b_1 + b_2)]^2}{ABn_3(n_1 + n_2)(n_1 + n_2 + n_3)}$$

(aus Sachs [Sac90])

$$\hat{\chi}_3^2 = \frac{N[b_4(a_1 + a_2 + a_3) - a_4(b_1 + b_2 + b_3)]^2}{ABn_4(n_1 + n_2 + n_3)}$$

7.6.7.1 Homogenitätstest nach Ryan (Lückentest)

Die Globalhypothese der Gleichheit mehrerer (k) **geordneter Binomialparameter** ($H_0 : p_1 = p_2 = \dots = p_k$) wird anhand des $k \cdot 2$ -Felder- χ^2 -Tests geprüft. Wird H_0 nicht abgelehnt, so lassen sich für nicht zu kleine Stichprobenumfänge nach Ryan [Rya60] abweichende relative Häufigkeiten bezüglich der Gleichheit ihrer Parameter prüfen (z.B. $H_0 : p_1 = p_k$), wobei der entsprechende Vierfeldertest anstatt auf dem $\alpha\%$ -Niveau auf einem modifizierten Niveau

$$\frac{\alpha}{\binom{k}{2}} = \frac{2 \cdot \alpha}{k(k-1)}$$

geprüft wird. Bei Nichtablehnung von H_0 ist das Ryan-Verfahren abgeschlossen. Wird H_0 abgelehnt, so kommen die weniger extremen relativen Häufigkeiten zum Vergleich, d.h. $H_0 : p_1 = p_{k-1}$ und $H_0 : p_2 = p_k$, jeweils zum Niveau $2 \cdot \alpha / [k(k-2)]$. Wird für so einen Bereich relativer Häufigkeiten H_0 beibehalten, so gelten alle in diesem enthaltenen Parameter p_i als homogen, ansonsten testet man weiter $H_0 : p_1 = p_{k-2}$ und $H_0 : p_3 = p_k$, jeweils zum Niveau $2 \cdot \alpha / [k(k-3)]$, usw. bis man gegebenenfalls bis zum Test $H_0 : p_i = p_{i+1}$ mit dem Niveau $2 \cdot \alpha / k$ gelangt (aus Sachs [Sac93]).

Beispiel: Gegeben seien 5 ansteigend geordnete relative Häufigkeiten, die global und nach dem Verfahren nach Ryan auf Homogenität ihrer Parameter zu prüfen sind ($\alpha = 0,05$).



Nr.	1	2	3	4	5	
-	$\frac{18}{30}$	$\frac{17}{25}$	$\frac{21}{28}$	$\frac{24}{30}$	$\frac{27}{30}$	-
rel. Häufigkeit	0,60	0,68	0,75	0,80	0,90	Summe
	18	17	21	24	27	107
	12	8	7	6	3	36
	30	25	28	30	30	143

$$\hat{\chi}^2 = \frac{143^2}{107 \cdot 36} \left[\left(\frac{12^2}{30} + \frac{8^2}{25} + \frac{7^2}{28} + \frac{6^2}{30} + \frac{3^2}{30} \right) - \frac{36^2}{143} \right] = 8,213 < 9,488 = \chi_{4;0,95}^2$$

d.h. H_0 lässt sich auf dem 5%-Niveau nicht ablehnen. Wir prüfen nun weiter mit dem Vierfelder-Test:

18	27	45
12	3	15
30	30	60

$$\hat{\chi}^2 = \frac{59(18 \cdot 3 - 27 \cdot 12)^2}{45 \cdot 15 \cdot 30 \cdot 30} = 7,08$$

Der entsprechende Schrankenwert ist $\chi_{1;0,995}^2 = 7,88$, α -adjustiert nach $2 \cdot 0,05 / [5(5-1)] = 0,005$ z.B. mit der Funktion **qchisq()** in **R** berechnet. Da dieser Wert von $\hat{\chi}^2 = 7,08$ nicht überschritten wird, endet die Ryan-Prozedur.

Hinweis: Man hätte an dieser Stelle auch mit multiplen Vergleichen nach Tukey und Kramer prüfen können - $H_0 : p_1 = p_2 = \dots = p_k$ (Sachs [Sac93]). Die beiden Binomialparameter p_i und p_j werden auf dem $100\alpha\%$ -Niveau als ungleich aufgefasst, sobald

$$|\hat{p}_i - \hat{p}_j| > q_{\infty;k;1-\alpha} \sqrt{\frac{\hat{p}(1-\hat{p})}{2} \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}$$

mit $\hat{p} = \sum_{i=1}^k r_i / \sum_{i=1}^k n_i$

und $1 \leq i < j \leq k$, $\hat{p}_i = r_i / n_i$ mit $1 \leq i \leq k$

(7.240)

Für die Stichproben 1 und 5 aus dem Beispiel mit $n_1 = n_5 = 30$, $\hat{p}_1 = r_1/n_1 = 18/30 = 0,60$ und $\hat{p}_5 = r_5/n_5 = 27/30 = 0,90$, sowie $\hat{p} = 107/143 = 0,748$ und $\alpha = 0,05$, d.h. $q_{\infty;5;0,95} = 3,86$ (vgl. Tabelle 7.44 oder die Funktion `qtukey()` in **R**) erhalten wir über

$$3,86 \sqrt{\frac{0,748(1-0,748)}{2} \left(\frac{1}{30} + \frac{1}{30} \right)} = 0,306$$

$$|\hat{p}_1 - \hat{p}_5| = |0,60 - 0,90| = 0,30 < 0,306$$

auf dem 5%-Niveau ebenfalls keine Ablehnung der Nullhypothese.

7.6.8 Cochran-Armitage Test auf linearen Trend

Erfolgt die Zunahme der relativen Häufigkeiten in einer $k \cdot 2$ -Feldertafel regelmäßig, dann ist eine Prüfung auf „linearen“ Trend angebracht. Dabei werden die Beobachtungen in der Tafel als Realisierungen von k unabhängigen binomial-verteilten Zufallsvariablen Y_i ($i = 1, \dots, k$) aufgefasst. Cochran [Coc54] und Armitage [Arm55] haben für die Analyse eine Zerlegung der χ^2 -Statistik auf der Grundlage des Modells (7.241) vorgeschlagen.

$$\pi_i = \alpha + \beta x_i \quad (7.241)$$

Darin sind die x_i ($i = 1, \dots, k$) feste Punktwerte (**Scores**), die eine möglichst natürliche Rangfolge der k Merkmale bzw. Merkmalsausprägungen widerspiegeln, Hierfür werden häufig Zahlen verwendet, die symmetrisch zu Null liegen, wie z.B. -2, -1, 0, 1, 2. Die Abstände zwischen den Zahlenwerten müssen nicht gleich gewählt werden. Einzelne Kategorien können aufgrund herausragender Eigenschaften durchaus auch durch größere Gewichte hervorgehoben werden.

Die Nullhypothese auf Unabhängigkeit der π_i kann nach dem Modell (7.241) durch $H_0: \beta = 0$ direkt auf einen linearen Trend bezogen werden. Die Schätzung von β kann mit dem Verfahren der kleinsten Fehlerquadrate (OLS) aus den beobachteten Häufigkeiten nach (7.242) erfolgen.

$\hat{\pi}_i = p + b(x_i - \bar{x})$		
mit	$\bar{x} = \sum_{i=1}^k n_i x_i / n$	mittlerer Score
	$p_i = y_i / n_i$	einzelne Anteile
	$p = \left(\sum_{i=1}^k y_i \right) / n$	Gesamtanteil
	$b = \frac{\sum_{i=1}^k n_i (p_i - p) (x_i - \bar{x})}{\sum_{i=1}^k n_i (x_i - \bar{x})^2}$	„Regressionskoeffizient“

(7.242)

Die χ^2 -Statistik kann nun nach Cochran und Armitage in zwei Anteile zerlegt werden. Der eine entfällt auf die als linear ansteigend gedachten Häufigkeiten (Trend), der restliche Anteil entspricht den Unterschieden zwischen den beobachteten Häufigkeiten und den als linear ansteigend vorausgesetzten theoretischen Häufigkeiten (Fehler). Formal wird diese Zerlegung in (7.243) angegeben.

$$\begin{aligned}
 \hat{\chi}^2 &= \frac{1}{p(1-p)} \sum_{i=1}^k n_i (p_i - p)^2 = \hat{\chi}_{trend}^2 + \hat{\chi}_{err}^2 \\
 \text{mit } \hat{\chi}_{err}^2 &= \frac{1}{p(1-p)} \sum_{i=1}^k n_i (p_i - \hat{\pi}_i)^2 \\
 \text{und } \hat{\chi}_{trend}^2 &= \frac{b^2}{p(1-p)} \sum_{i=1}^k n_i (x_i - \bar{x})^2
 \end{aligned}
 \tag{7.243}$$

$\hat{\chi}_{err}^2$ ist asymptotisch χ^2 -verteilt mit $k - 2$ Freiheitsgraden. $\hat{\chi}_{trend}^2$ ist asymptotisch χ^2 -verteilt mit 1 Freiheitsgrad ($z = \sqrt{\hat{\chi}_{trend}^2}$ ist asymptotisch standardnormalverteilt) und kann direkt für die Prüfung der Hypothese eines linearen Trends in den Anteilen y_i/n verwendet werden.

Das $\hat{\chi}_{trend}^2$ für die „lineare Regression“ kann nach Cochran [Coc54] (vgl. auch Armitage [Arm55], Bartholomew [Bar59] sowie Bennett und Hsu [BH62]) auch durch (7.244) bestimmt werden:

$$\hat{\chi}_{trend}^2 = \frac{\left(\sum y_i x_i - \frac{\sum y_i \sum n_i x_i}{n} \right)^2}{p(1-p) \left(\sum n_i x_i^2 - \frac{(\sum n_i x_i)^2}{n} \right)} \quad \text{mit } FG = 1
 \tag{7.244}$$

Beispiel: Wenden wir den Cochran-Armitage Test auf die Werte in unserem einführenden Beispiel zum Therapieerfolg an, wobei die Punktwerte mit -1, 0 und +1 festgelegt werden, so erhalten wir



Score x_i	y_i	$n_i - y_i$	n_i	$p_i = y_i/n_i$	$y_i x_i$	$n_i x_i$	$n_i x_i^2$
+1	22	14	36	0,611	22	36	36
0	16	18	34	0,471	0	0	0
-1	2	8	10	0,200	-2	-10	10
	40	40	$n = 80$		20	26	46
$p = \sum y_i/n = 40/80 = 0,50$							

$$\hat{\chi}_{trend}^2 = \left(20 - \frac{40 \cdot 26}{80} \right)^2 \bigg/ \left[0,50 \cdot 0,50 \left(46 - \frac{26^2}{80} \right) \right] = 5,220 > 5,024 \quad \left[\begin{array}{l} = \chi^2; \text{Tab. 7.89} \\ \alpha = 0,05 \\ \tau = 2, \nu = 1 \end{array} \right]$$

Der Wert 5,22 ist auf dem 5%-Niveau statistisch signifikant. Im Beispiel war dagegen für $\hat{\chi}^2 = 5,495$ und $FG = 2$ die allgemeine Homogenitätshypothese mit einer Irrtumswahrscheinlichkeit von $\alpha = 0,05$ nicht abgelehnt worden.

Die folgende Übersicht zeigt den entscheidenden Anteil der linearen Regression an der Gesamtvariation, der schon in der Spalte der p_i Werte erkennbar ist und die Überlegenheit der spezifischen Therapie zum Ausdruck bringt.

Variationsursache	$\hat{\chi}^2$	FG	Signifikanzniveau
Lineare Regression	5,220	1	$0,01 < P < 0,05$
Abweichungen von der Regression	0,275	1	$P \approx 0,60$
Insgesamt	5,495	2	$0,05 < P < 0,10$

Die Beispieldaten zum therapeutischen Erfolg der spezifischen Behandlung lassen sich auch so umschreiben, dass ein möglicher Trend der den p_j zugrundeliegenden Wahrscheinlichkeiten π_j anhand von (7.245) beurteilt werden kann.

y_i	22	16	2
$n_i - y_i$	14	18	8
n_i	36	34	10
p_i	0,611	0,471	0,200
$p_i(1 - p_i)/n_i$	0,00660	0,00733	0,01600
x_i	+1	0	-1

$$\hat{z} = \frac{\sum x_i p_i}{\sqrt{\sum x_i^2 [p_i(1 - p_i)/n_i]}} \quad (7.245)$$

$$\hat{z} = \frac{1 \cdot 0,611 - 1 \cdot 0,200}{\sqrt{1 \cdot 0,00660 + 1 \cdot 0,01600}} = \frac{0,411}{0,15033} = 2,73$$

[Für $z_j = +2$; 0; -2 erhalte man

$$\hat{z} = \frac{2 \cdot 0,611 - 2 \cdot 0,200}{\sqrt{4 \cdot 0,00660 + 4 \cdot 0,01600}} = \frac{0,824}{0,30067} = 2,74].$$

Der Trend ist somit deutlich [$P(Z > 2,73) = 0,003$] erkennbar.

Für die Berechnung der Cochran-Armitage Statistik in **R** wird eine kleine Funktion angegeben, in der eine vollständige Zerlegung der χ^2 -Statistik zu einer $k \cdot 2$ -Feldertafel nach (7.243) in einzelnen Schritten durchgeführt wird.

```
> tabtrend <- function(tab, scores, transpose=FALSE) {
+   if (any(dim(tab)==2)) {if (transpose==TRUE) {tab <- t(tab)}}
+   if (dim(tab)[1]!=2) {stop("Cochran-Armitage nur in (2,k)-Tafel", call.=FALSE)}
+   nidot <- apply(tab,2,sum); n <- sum(nidot) # Summen und Scores
+   scri <- scores; scrq <- sum(scri*nidot)/n
+   p.i <- tab[,1,]/nidot # beobachtete Anteile
+   p <- sum(tab[,1,])/n
+   chi <- 1/(p*(1-p))*(sum(nidot*((p.i-p)^2))); chi # Chi-Quadrat gesamt
+
+   b <- sum(nidot*(p.i-p)*(scri-scrq))/sum(nidot*(scri-scrq)^2)
+   pi.h <- p + b*(scri-scrq)
+
+   chi.e <- (1/(p*(1-p)))*sum(nidot*(p.i-pi.h)^2); chi.e # Chi-Quadrat Abweichung
+   chi.t <- b^2/(p*(1-p))*sum(nidot*(scri-scrq)^2); chi.t # Chi-Quadrat Trend
+   z <- sqrt(chi.t)
+   p <- 2*pnorm(abs(z), lower.tail=FALSE) # P-Wert zweiseitig
+   out <- list(name="Cochran-Armitage Test auf trend",
+     chi.trend=chi.t, chi.err=chi.e, chi.gesamt=chi, p.wert=p)
+   return(out) }
+ }
```



Beispiel: Die Frage, in welcher Weise das Auftreten von Fehlbildungen von der Höhe des Alkoholkonsums der Mutter während der Schwangerschaft abhängt, soll an folgenden Daten überprüft werden (Agresti [Agr02]). Die Stufen hinsichtlich des Alkoholkonsums werden hier durch die Punktwerte 0, 0,5, 1,5, 4 und 7 abgebildet. Für die Berechnungen wird die zuvor definierte Funktion `tabtrend()` in **R** verwendet.

Fehlbildung	Alkoholkonsum				
	0	< 1	1 – 2	3 – 5	≥ 6
ja	48	38	5	1	1
nein	17066	14464	788	126	37
Score x_i	0	0,5	1,5	4	7

```
> malform <- matrix(c(48, 38, 5, 1, 1, 17066, 14464, 788, 126, 37),
+                   nrow=2, byrow=T,
+                   dimnames = list(fehlbildung=c("ja", "nein"),
+                   alkohol=c("0", "<1", "1-2", "3-5", ">5"))); malform
      fehlbildung alkohol
      ja         0      <1 1-2 3-5 >5
      ja         48      38   5   1   1
      nein 17066 14464 788 126 37
> tabtrend(malform, c(0,0.5,1.5,4,7), transpose=FALSE)
$name
[1] "Cochran-Armitage Test auf trend"
$chi.trend
[1] 6.570134
$chi.err
[1] 5.511921
$chi.gesamt
[1] 12.08205
$p.wert
[1] 0.01037041
```

Der χ^2 -Test auf Unabhängigkeit liefert für die vorliegende Tabelle den Wert $\hat{\chi}^2 = 12,1$ mit 4 Freiheitsgraden. Die Zerlegung nach dem Ansatz von Cochran und Armitage führt zu den Komponenten $\hat{\chi}_{err}^2 = 5,5$ und $\hat{\chi}_{trend}^2 = 6,6$. Speziell der Wert für den Trend zeigt für die gewählten Scores einen signifikanten Anstieg der Fehlbildungsrate in Abhängigkeit von der Höhe des Alkoholkonsums ($P=0,01$).

7.6.9 Die Analyse von Zweifeldern des Typs $r \cdot c$

Eine Erweiterung der Vierfeldertafel als einfachste Zweifeldertafel auf den allgemeinen Fall führt zur $r \cdot c$ -, Mehrfelder- oder **Kontingenztafel**, einer Tafel, die r Zeilen oder Reihen (rows) und c Spalten (columns) aufweist. Zwei Merkmale mit r bzw. c verschiedenen Ausprägungen werden in $r \cdot c$ verschiedenen Feldern oder Kombinationen übersichtlich dargestellt (Tabelle 7.86).

Tabelle 7.86. Schema für die zweifache Klassifikation: Eine der beiden Merkmalsreihen dieser $r \cdot c$ -Tafel ist auch als Stichprobenreihe auffassbar

1. Merkmal (r Zeilen)	2. Merkmal (c Spalten)	1 2 - j - c				Zeilensumme
		1	2	-	j - c	
1		n_{11}	n_{12}	-	$n_{1j} - n_{1c}$	$n_{1.}$
2		n_{21}	n_{22}	-	$n_{2j} - n_{2c}$	$n_{2.}$
-		-	-	-	-	-
i		n_{i1}	n_{i2}	-	$n_{ij} - n_{ic}$	$n_{i.}$
-		-	-	-	-	-
r		n_{r1}	n_{r2}	-	$n_{rj} - n_{rc}$	$n_{r.}$
Spaltensumme		$n_{.1}$	$n_{.2}$	-	$n_{.j} - n_{.c}$	$n_{..} = n$ Ecksumme

Eine Stichprobe vom Umfang n wird aus einer Verteilung zufällig entnommen. Jedes Element dieser Stichprobe wird dann nach den zwei verschiedenen diskreten Merkmalen klassifiziert. Zu

prüfen ist die Hypothese der Unabhängigkeit: Merkmal I hat keinen Einfluss auf Merkmal II . Anders ausgedrückt: Es wird getestet, ob die Verteilung qualitativer Merkmale nach einer Merkmalsreihe unabhängig ist von der Einteilung nach einer zweiten Merkmalsreihe bzw. ob eine zu den Randsummen weitgehend proportionale Häufigkeitsverteilung vorliegt.

Bei **stochastischer Unabhängigkeit** gilt für jedes Feld n_{ij} einer Kontingenztafel [vgl. Tab. 7.86 und (7.246)]: $n_{ij} = (n_{i.}n_{.j})/n$ für alle i und j .

Übersicht 21: Zum χ^2 -Test für eine $r \cdot c$ -Tafel

Einer $r \cdot c$ -Tafel können drei Modelle zugrundeliegen:

- (1) Keine Randsummen fest vorgegeben
- (2) Eine Serie von Randsummen fest vorgegeben
- (3) Beide Serien von Randsummen fest vorgegeben

In allen drei Fällen ist die Prüfgröße für großes n und großes $\hat{\chi}^2$ ($P \gtrsim 0,1$) nach χ_ν^2 verteilt, mit $\nu = (r-1)(c-1)$ Freiheitsgraden.

Für kleines n und $\hat{\chi}^2$ -Werte $< 0,1$ ist die χ_ν^2 -Verteilung nur eine grobe Approximation: für alle drei Fälle resultiert gegenüber dem exakten P -Wert ein zu kleiner Wert P , und zwar im allgemeinen für $P \simeq 0,05$ etwas zu klein, für $P \simeq 0,01$ und besonders für $P \simeq 0,001$ deutlich zu klein (vgl. Haynam und Leone [HL65]).

Es sei an dieser Stelle vermerkt, dass ein Vergleich von r verschiedenen Stichproben mit den Umfängen $n_1, n_2, \dots, n_i, \dots, n_r$, aus r verschiedenen diskreten Verteilungen auf Gleichartigkeit oder Homogenität zu demselben Testverfahren führt. Die Alternativhypothese lautet hier: mindestens zwei Verteilungsfunktionen sind ungleich. Wir haben daher genau die gleiche Testgröße, gleichgültig ob wir eine Kontingenztafel auf stochastische Unabhängigkeit testen wollen (die Ecksumme n ist fest vorgegeben), oder ob wir r Stichproben (die Stichprobenumfänge [Randsummen] sind fest vorgegeben) dahingehend vergleichen wollen, ob sie aus derselben Grundgesamtheit stammen (Vergleich der r Verteilungsfunktionen bzw. der Grundwahrscheinlichkeiten von r Multinomialverteilungen). Das ist erfreulich, da es bei vielen Problemstellungen keineswegs klar ist, welche Auffassung eher angemessen erscheint. Die Prüfgröße ist

$$\hat{\chi}^2 = \sum_{i=1}^r \sum_{j=1}^c \left[\frac{\left(n_{ij} - \frac{n_{i.}n_{.j}}{n} \right)^2}{\frac{n_{i.}n_{.j}}{n}} \right] = n \left[\sum_{i=1}^r \sum_{j=1}^c \frac{n_{ij}^2}{n_{i.}n_{.j}} - 1 \right] \quad (7.246)$$

mit $(r-1)(c-1)$ Freiheitsgraden. Hierin bedeuten:

- n = Umfang der Stichprobe bzw. Gesamtstichprobenumfang
- n_{ij} = Besetzungszahl des Feldes in der i -ten Zeile und der j -ten Spalte
- $n_{i.}$ = Summe der Besetzungszahlen der i -ten Zeile (Zeilensumme)
- $n_{.j}$ = Summe der Besetzungszahlen der j -ten Spalte (Spaltensumme)
- $n_{i.}n_{.j}$ = Produkt der Randsummen

Vertauscht man in einer $r \cdot c$ -Tabelle Zeilen und/oder Spalten, so bleibt das berechnete $\hat{\chi}^2$ unverändert.

Die Erwartungshäufigkeiten in einer Kontingenztafel berechnen sich (unter der Nullhypothese) nach $n_{i.}n_{.j}/n$. Sind die beiden Merkmale stochastisch unabhängig, so gilt $n_{ij} = n_{i.}n_{.j}/n$ und es folgt $\hat{\chi}^2 = 0$. Bei Gültigkeit der Nullhypothese auf Unabhängigkeit oder Homogenität und



für hinreichend große n ist die obige Prüfgröße $\hat{\chi}^2$ wie das tabelliert vorliegende χ^2 -verteilt mit $(r-1)(c-1)$ Freiheitsgraden. Dies ist die Zahl der Felder einer Tafel, für die man die Häufigkeiten frei wählen kann, wenn die Randsummen gegeben sind. Die Besetzungszahlen der übrigen Felder lassen sich dann durch Subtraktion ermitteln. Der Test darf angewandt werden, wenn alle Erwartungshäufigkeiten ≥ 1 sind. Treten kleinere Erwartungshäufigkeiten auf, dann ist die Tafel durch Zusammenfassung unterbesetzter Felder zu vereinfachen. Hierbei ist zu beachten, dass man ein möglichst objektives Schema anwenden sollte, um nicht durch mehr oder minder bewusste Willkür bei dieser Zusammenfassung das Ergebnis zu beeinflussen. Günstig ist es, vor der Datengewinnung festzulegen, wie eine mögliche Vereinfachung von Mehrfeldertafeln durch Zusammenfassung von Zeilen und/oder Spalten zu erfolgen hat.

Beispiel: Vergleich dreier Therapieformen an randomisierten prognostisch nahezu gleichwertigen Patienten. Versuchsplan: Drei Gruppen von je 40 Kranken wurden behandelt. Zwei Gruppen sind in dem Beispiel in vorangehenden Abschnitt verglichen worden. Die dritte Gruppe erhielt eine spezifische Therapie mit doppelter Normaldosis (Quelle: Martini [Mar53] S. 79, Tab. 13). Wir wählen auch hier den χ^2 -Test auf dem 5%-Niveau. Die H_0 kann einmal lauten „Gleichheit dreier Verteilungsfunktionen“ (den drei Therapien entsprechend), zum anderen „Unabhängigkeit von Therapie und Therapie-Erfolg“.



Tabelle 7.87. Beispieldaten zum Vergleich dreier Therapieformen

Therapeutischer Erfolg	Therapie			insgesamt
	symptomatisch	spezifisch		
		Normaldosis	2x Normaldosis	
geheilt in x Wochen	14	22	32	68
geheilt in $x + y$ Wochen	18	16	8	42
gestorben	8	2	0	10
insgesamt	40	40	40	120

- Ergebnisse und Auswertung: $\hat{\chi}^2 = 120 \left[\frac{14^2}{68 \cdot 40} + \dots + \frac{0^2}{10 \cdot 40} - 1 \right] = 21,576$.
- Entscheidung: Da $21,58 > 9,49 = \chi^2_{4;0,95}$ ist, wird die Nullhypothese abgelehnt.
- Interpretation: Der Zusammenhang zwischen dem therapeutischen Erfolg und besonders der spezifischen Therapie mit doppelter Normaldosis erscheint gesichert. Diese Therapie ist auf dem 5%-Niveau den anderen beiden Therapieformen überlegen.

Quadratische Tafeln ($r = c$) weisen bei vollständiger Abhängigkeit den Wert

$$\hat{\chi}_{max.}^2 = n(r-1) \quad (7.247)$$

auf, für unser Beispiel hier $\hat{\chi}_{max.}^2 = 120(3-1) = 240$.

Vergleicht man konservativ behandelte mit operierten Patienten, so ist die Vergleichbarkeit der Patienten vor der Zuordnung entscheidend wichtig; mitunter kommen die konservativ behandelten für eine Operation nicht (mehr) infrage; d. h. die zu operierenden Patienten haben, auch wenn sie nicht operiert werden, eine bessere Überlebenschance. Auch sind Patienten, die auf eine bestimmte Behandlung ansprechen, weniger krank als andere, die keine Wirkung zeigen. Schwierigkeiten derart, dass in Behandlungsgruppen einander widersprechende **Untergruppeneffekte** auftreten können, erschweren außerdem die erwünschten Verallgemeinerungen, deren Berechtigung sorgfältig zu prüfen ist.

Das erweiterte Beispiel soll auch in **R** mit der Funktion **chisq.test()** berechnet werden.

```

> erfolg <- matrix(c(14, 22, 32, 18, 16, 8, 8, 2, 0), nr=3, byrow=T,
+                 dimnames = list(heilung=c("geheilt-x", "geheilt-x+y", "gestorben"),
+                 therapie=c("symptomatisch", "spezifisch N1", "spezifisch N2")))
> erfolg
      therapie
heilung symptomatisch spezifisch N1 spezifisch N2
geheilt-x           14           22           32
geheilt-x+y         18           16            8
gestorben            8            2            0
> chisq.test(erfolg, correct = TRUE)

      Pearson's Chi-squared test

data:  erfolg
X-squared = 21.5765, df = 4, p-value = 0.0002433

Warning message:
Chi-squared approximation may be incorrect in: chisq.test(erfolg, correct = TRUE)

```

Das Ergebnis stimmt mit der zuvor bestimmten Lösung überein, allerdings wird ergänzend ein Warnhinweis gegeben, dass die χ^2 -Approximation möglicherweise nicht korrekt (P-Wert) ist und somit zu einem falschen Ergebnis in der Interpretation des Testergebnisses führen kann. Der Hinweis erfolgt aufgrund zu geringer Erwartungswerte (hier < 5) in der Kontingenztafel. Die Funktion **chisq.test()** bietet für diesen Fall die Möglichkeit, einen P-Wert anhand eines Randomisierungstests (Monte-Carlo-Simulation) mit einer festen Zahl von Wiederholungen (hier zufällige Auswahl von 1000 Tafeln mit gleichen Randsummen) zu bestimmen. Entsprechend der Hinweise in der Übersicht auf Seite 520 zeigt sich, dass der P -Wert auf der Grundlage der Approximation mit der χ^2 -Verteilung „deutlich“ kleiner ausfällt als der exakte, hier durch eine Simulation ermittelte, P -Wert.

```

> chisq.test(erfolg, simulate.p.value = TRUE, B = 1000)

      Pearson's Chi-squared test with simulated p-value (based on 1000
      replicates)

data:  erfolg
X-squared = 21.5765, df = NA, p-value = 0.000999

```

Scoring II (Homogenitätstest)

Etwas teststärker als (7.246) ist folgendes Verfahren. H_0 : Gleichheit der den k Stichproben zugrundeliegenden Verteilungen wird für $\hat{\chi}^2 > \chi^2_{k-1;\alpha}$ auf dem $100\alpha\%$ -Niveau verworfen. Die Umfänge der k -Stichproben sollten hierbei nicht zu unterschiedlich und hinreichend groß sein. Im vorliegenden Fall mit $k = 3$, $n_1 = n_2 = n_3 = 40$ und $n = 120$ ergibt sich:

B_s	B_e	B_d	B	Score x	$B_s x$	$B_e x$	$B_d x$	Bx	Bx^2
14	22	32	68	1	14	22	32	68	68
18	16	8	42	0	0	0	0	0	0
8	2	0	10	-1	-8	-2	0	-10	10
40	40	40	120	-	6	20	32	58	78

$$\hat{\chi}^2 = \frac{(n-1)\{[(\sum B_s x)^2/n_1 + (\sum B_e x)^2/n_2 + (\sum B_d x)^2/n_3] - (\sum Bx)^2/n\}}{\sum Bx^2 - (\sum Bx)^2/n} \quad (7.248)$$

$$\hat{\chi}^2 = \frac{119\{[6^2/40 + 20^2/40 + 32^2/40] - 58^2/120\}}{78 - 58^2/120} = 20,164$$

[Mit den Scores 1, 2, 3 hätte man dasselbe Resultat erhalten.]

Da $\hat{\chi}^2 = 20,164 > 5,9915 = \chi_{2;0,95}^2$ ist, wird H_0 auf dem 5%-Niveau ebenfalls abgelehnt.

Hinweise:

- (1) Prüft man eine **schwach besetzte Kontingenztafel** mit ν Freiheitsgraden auf dem 5%-Niveau, so sollte nach Lawal und Upton [LU90] das modifizierte

$$\hat{\chi}_{LU}^2 = \hat{\chi}^2 / [1 - (1 - 1/\sqrt{\nu})/n] \quad (7.249)$$

benutzt werden, für das 1%-Niveau verwende man

$$\hat{\chi}_{LU}^2 = \hat{\chi}^2 / [1 - (3/2n)] \quad (7.249a)$$

Die H_0 wird anhand der üblichen $\chi_{\nu;\alpha}^2$ -Schranken abgelehnt, sobald $\hat{\chi}_{LU}^2 > \chi_{\nu;\alpha}^2$. So erhält man z. B. für die Tabelle 7.87 [identisch mit der in (3) genannten]:

$$\hat{\chi}_{LU}^2 = 21,576 / [1 - (1 - 1/\sqrt{4})/120] = 21,666.$$

- (2) **$\hat{\chi}^2$ -Variationsbereich für $r \cdot c$ -Tabellen:** Allgemein gilt für das aus einer $r \cdot c$ -Tabelle mit der Ecksumme n berechnete $\hat{\chi}^2$:

$$0 \leq \hat{\chi}^2 \leq n \cdot \text{Min}[(r-1), (c-1)] \quad (7.250)$$

- (3) **95%-Konfidenzintervalle für multiple Vergleiche zweier einander entsprechender Wahrscheinlichkeiten einer Homogenitätstafel:**

Effekt E in der Gruppe G

E \ G	SY	N1	N2	Σ
++	14	22	32	68
+	18	16	8	42
-	8	2	0	10
Σ	40	40	40	120

Etwa für den Vergleich des „+“-Effektes in den Gruppen „SY“ und „N2“ anhand von $18/40 = 0,45$ und $8/40 = 0,20$: 95%-KI für $\pi_{SY;+} - \pi_{N2;+} = (0,45 - 0,20) \pm \sqrt{9,49} \cdot \sqrt{[(0,45 \cdot 0,55)/40] + [(0,2 \cdot 0,8)/40]}$ oder $0,25 \pm 0,31$: da dieses Intervall die Null enthält, lässt sich ein unterschiedlicher „+“-Effekt in den zugrundeliegenden Grundgesamtheiten auf dem 5%-Niveau nicht sichern. Näheres ist den Hinweisen (6) und (8) zu entnehmen. Entsprechend sind auch die anderen 8 Vergleiche auf dem 5%-Niveau (ohne α -Korrektur) möglich.

- (4) **Erfassung homogener Merkmalskombinationen** (Das Aufspüren von Kombinationen wichtiger Eigenschaften aus den Daten): Liegen beide Merkmale einer Kontingenztafel in jeweils mehreren geordneten Kategorien (etwa: kaum, leicht, mittel) vor und enthalten die Einzelfelder, die Randsummenfelder und das Eckfeld Anteile mit zugehörigen, in Klammern gesetzten relativen Häufigkeiten,

$$\text{etwa } \frac{7}{148} \quad (0,146)$$

dann wird man sie gern etwas vereinfachen, sagen wir eine $4 \cdot 5$ -Felder-Tabelle durch Zusammenfassen geeigneter Merkmalskombinationen in eine $3 \cdot 3$ -Felder-Tabelle umwandeln, aus der sich dann bezüglich der Anteile wenige, sagen wir 4 oder 5 in sich weitgehend homogen zusammengesetzte Gruppierungen herauschälen lassen, etwa in der Art der Tabelle 7.88, wobei dann auch bei hinreichend großem Nenner Prozente angegeben werden.

Tabelle 7.88. Übersichtsschema zur zusammenfassenden Gruppierung weitgehend homogener Merkmalskombinationen

Merkmal 1 2		Gruppe	Anteil
$\left\{ \begin{array}{ll} \text{leicht} & \text{leicht} \\ \text{leicht} & \text{mittel} \\ \text{mittel} & \text{leicht} \end{array} \right\}$ $\vdots$		A	$\left\{ \begin{array}{l} 24/531 \\ (4,5\%) \end{array} \right\}$
		B	$\vdots$
		C	$\vdots$
		D	$\left\{ \begin{array}{l} 31/118 \\ (26,3\%) \end{array} \right\}$
{stark stark}			

Ähnlich fasst auch der Mediziner bestimmte Erscheinungen einer Krankheit als „Stadium I“, „Stadium II“ usw. zusammen.

-  (5) Die **Power** eines χ^2 -Tests auf Homogenität oder Unabhängigkeit für $r \geq 2$, $c \geq 2$ Kategorien hängt in komplizierter Weise ab vom Wert $\hat{\chi}^2$, von n , von den Randsummen und vom Freiheitsgrad. Nur dieser wird für die Beurteilung von $\hat{\chi}^2$ herangezogen. Hierdurch wird der Fehler erster Art kontrolliert, nicht aber die Power. Ein größeres $\hat{\chi}^2$ bedeutet nicht unbedingt eine größere Power.
- (6) Ist im Verlauf der Analyse von Mehrfeldertafeln die Nullhypothese zugunsten der Alternativhypothese auf Abhängigkeit oder Heterogenität abzulehnen, dann besteht zuweilen das Interesse, die **Ursache der Signifikanz zu lokalisieren**. Man wiederhole dann den Test an einer Tafel, die um die betreffende Zeile oder Spalte vermindert ist; besser ist es, das Verfahren aus [7.6.9.1] anzuwenden. Andere Möglichkeiten, interessante Teilhypothesen zu prüfen, bietet die Auswahl von 4 symmetrisch zueinander gelegenen Feldern, je zwei Felder liegen in einer Zeile und einer Spalte, die dann mit einem Vierfeldertest geprüft werden. Dies sollte jedoch als „**experimentieren**“ aufgefasst werden; die Ergebnisse können lediglich als Anhaltspunkte für künftige Untersuchungen dienen. Ein echter Wert ist ihnen nur dann zuzuerkennen, wenn die entsprechenden Teilhypothesen schon vor Erhebung der Daten konzipiert worden waren. Ein anderer Hinweis sei hier angeschlossen. Erscheint die Abhängigkeit gesichert, dann ist zu bedenken, dass die Existenz eines formalen Zusammenhangs nichts aussagt über den kausalen Zusammenhang. Es ist durchaus möglich, dass indirekte Zusammenhänge einen Teil der Abhängigkeit bedingen.
- (7) Jede Kontingenztafel vom allgemeinen Typ $r \cdot c$ lässt sich in $(r - 1)(c - 1)$ **unabhängige Komponenten mit je einem Freiheitsgrad** zerlegen (vgl. Kastenbaum [Kas60], Castellan [Cas65] sowie Bresnahan und Shapiro [BS66]). Mit der Symbolik von Tabelle 7.86 ergeben sich z. B. für eine 3·3-Tafel, $2 \cdot 2 = 4$ FG stehen zur Verfügung, 4 Komponenten:

$$(1) \quad \hat{\chi}^2 = \frac{n\{n_{2.}(n_{.2}n_{11} - n_{.1}n_{12}) - n_{1.}(n_{.2}n_{21} - n_{.1}n_{22})\}^2}{n_{1.}n_{2.}n_{.1}n_{.2}(n_{1.} + n_{2.})(n_{.1} + n_{.2})} \quad (7.251a)$$

$$(2) \quad \hat{\chi}^2 = \frac{n^2\{n_{23}(n_{11} + n_{12}) - n_{13}(n_{21} + n_{22})\}^2}{n_{1.}n_{2.}n_{.3}(n_{1.} + n_{2.})(n_{.1} + n_{.2})} \quad (7.251b)$$

$$(3) \quad \hat{\chi}^2 = \frac{n^2\{n_{32}(n_{11} + n_{21}) - n_{31}(n_{12} + n_{22})\}^2}{n_{3.}n_{.1}n_{.2}(n_{1.} + n_{2.})(n_{.1} + n_{.2})} \quad (7.251c)$$

$$(4) \hat{\chi}^2 = \frac{n\{n_{33}(n_{11} + n_{12} + n_{21} + n_{22}) - (n_{13} + n_{23})(n_{31} + n_{32})\}^2}{n_{3.}n_{.3}(n_{1.} + n_{2.})(n_{.1} + n_{.2})} \quad (7.251d)$$

Für unser Beispiel, mit vereinfachten Kategorien ($A, B, C; I, II, III$), sind die folgenden 4 Vergleiche möglich:

Typ	A	B	C	Σ
I	14	22	32	68
II	18	16	8	42
III	8	2	0	10
Σ	40	40	40	120

- (1) Der Vergleich I gegen II hinsichtlich A gegen B (Symbolik: $I \times II \div A \times B$)
- (2) Der Vergleich I gegen II hinsichtlich $\{A + B\}$ gegen C ($I \times II \div \{A + B\} \times C$)
- (3) Der Vergleich $\{I + II\}$ gegen III hinsichtlich A gegen B ($\{I + II\} \times III \div A \times B$)
- (4) Der Vergleich $\{I + II\}$ gegen III hinsichtlich $\{A + B\}$ gegen C ($\{I + II\} \times III \div \{A + B\} \times C$)

Unabhängigkeit	FG	$\hat{\chi}^2$	P
(1) $I \times II \div A \times B$	1	1,0637	n.s.
(2) $I \times II \div \{A + B\} \times C$	1	9,1673	$< 0,01$
(3) $\{I + II\} \times III \div A \times B$	1	5,8909	$< 0,10$
(4) $\{I + II\} \times III \div \{A + B\} \times C$	1	5,4545	$< 0,10$
insgesamt	4	21,5764	$< 0,001$

$$(1) \hat{\chi}^2 = \frac{120\{42(40 \cdot 14 - 40 \cdot 22) - 68(40 \cdot 18 - 40 \cdot 16)\}^2}{68 \cdot 42 \cdot 40 \cdot 40 \cdot (68 + 42)(40 + 40)} = 1,0637 < 5,024$$

$$(2) \hat{\chi}^2 = \frac{120^2\{8(14 + 22) - 32(18 + 16)\}^2}{68 \cdot 42 \cdot 40 \cdot (68 + 42)(40 + 40)} = 9,1673 > 9,141$$

$$(3) \hat{\chi}^2 = \frac{120^2\{2(14 + 18) - 8(22 + 16)\}^2}{10 \cdot 40 \cdot 40 \cdot (68 + 42)(40 + 40)} = 5,8909 > 5,024$$

$$(4) \hat{\chi}^2 = \frac{120\{0(14 + 22 + 18 + 16) - (32 + 8)(8 + 2)\}^2}{10 \cdot 40 \cdot (68 + 42)(40 + 40)} = 5,4545 > 5,024$$

Wenn andere spezifische Vergleiche geprüft werden sollen, sind Zeilen oder Spalten (bzw. beide) entsprechend zu vertauschen.

(8) Mehrfache Anwendung von Tests auf denselben Datenkörper.

(a) Werden insgesamt τ (griech. tau) Tests gemacht, jeweils auf dem Signifikanzniveau α_i , so ist die Gesamtsignifikanz der τ Tests kleiner oder gleich $\sum_{i=1}^{\tau} \alpha_i$. Gewöhnlich wählt man für jeden Test $\alpha_i = \alpha/\tau$, α ist dann das nominelle Signifikanzniveau für diese Folge von Tests (Bonferroni-Verfahren).

(b) Im Rahmen einer Erhebung seien τ χ^2 -Tests geplant (Typ: $k \cdot 1$; $k \cdot 2$ und $k \geq 2$ bzw. $r \cdot c$ mit $r, c > 2$) mit jeweils ν_i Freiheitsgraden. Dann sind die kritischen Schranken der Bonferroni- χ^2 -Tabelle (Tab. 7.89) anzuwenden. Die Wahrscheinlichkeit, mindestens eine der Nullhypothesen fälschlich abzulehnen, ist dann nicht größer als das nominelle Signifikanzniveau α .



Tabelle 7.89. Obere Schranken der Bonferroni-Statistik $\chi^2(\alpha/\tau; \nu)$. Auszugsweise aus Kramer, C.Y.: A First Course in Methods of Multivariate Analysis, Virginia Polytechnic Institute and State University, Blacksburg 1972, Appendix D: G.B. Beus und D.R. Jensen, Sept. 1967, pp. 327–351 [in den drei Blöcken für $\alpha = 0, 10$, $\alpha = 0, 05$ und $\alpha = 0, 01$]; mit Genehmigung des Autors

ν	$\tau : 1$	2	3	4	5	6	7	8	9	10	11	12
1	2,706	3,841	4,529	5,024	5,412	5,731	6,002	6,239	6,447	6,635	6,805	6,960
2	4,605	5,991	6,802	7,378	7,824	8,189	8,497	8,764	9,000	9,210	9,401	9,575
3	6,251	7,815	8,715	9,348	9,837	10,236	10,571	10,861	11,117	11,345	11,551	11,739
4	7,779	9,488	10,461	11,143	11,668	12,094	12,452	12,762	13,034	13,277	13,496	13,695
5	9,236	11,070	12,108	12,833	13,388	13,839	14,217	14,544	14,831	15,086	15,317	15,527
6	10,645	12,592	13,687	14,449	15,033	15,506	15,903	16,245	16,545	16,812	17,053	17,272
8	13,362	15,507	16,705	17,535	18,168	18,680	19,109	19,478	19,802	20,090	20,350	20,586
9	14,684	16,919	18,163	19,023	19,679	20,209	20,653	21,034	21,368	21,666	21,934	22,177
10	15,987	18,307	19,594	20,483	21,161	21,707	22,165	22,558	22,903	23,209	23,485	23,736
12	18,549	21,026	22,394	23,337	24,054	24,632	25,115	25,530	25,894	26,217	26,508	26,772
14	21,064	23,685	25,127	26,119	26,873	27,480	27,987	28,422	28,803	29,141	29,446	29,722
15	22,307	24,996	26,473	27,488	28,259	28,880	29,398	29,843	30,232	30,578	30,889	31,171
16	23,542	26,296	27,808	28,845	29,633	30,267	30,796	31,250	31,647	32,000	32,317	32,605

1	3,841	5,024	5,731	6,239	6,635	6,960	7,237	7,477	7,689	7,879	8,052	8,210
2	5,991	7,378	8,189	8,764	9,210	9,575	9,883	10,150	10,386	10,597	10,787	10,961
3	7,815	9,348	10,236	10,861	11,345	11,739	12,071	12,359	12,612	12,838	13,043	13,229
4	9,488	11,143	12,094	12,762	13,277	13,695	14,048	14,358	14,621	14,860	15,076	15,273
5	11,070	12,833	13,839	14,544	15,086	15,527	15,898	16,217	16,499	16,750	16,976	17,182
6	12,592	14,449	15,506	16,245	16,812	17,272	17,659	17,993	18,286	18,548	18,783	18,998
8	15,507	17,535	18,680	19,478	20,090	20,586	21,002	21,360	21,675	21,955	22,208	22,438
9	16,919	19,023	20,209	21,034	21,666	22,177	22,607	22,976	23,301	23,589	23,850	24,086
10	18,307	20,483	21,707	22,558	23,209	23,736	24,178	24,558	24,891	25,188	25,456	25,699
12	21,026	23,337	24,632	25,530	26,217	26,772	27,237	27,637	27,987	28,300	28,581	28,836
14	23,685	26,119	27,480	28,422	29,141	29,722	30,209	30,627	30,993	31,319	31,613	31,880
15	24,996	27,488	28,880	29,843	30,578	31,171	31,668	32,095	32,469	32,801	33,101	33,373
16	26,296	28,845	30,267	31,250	32,000	32,605	33,111	33,547	33,928	34,267	34,572	34,850

1	6,635	7,879	8,615	9,141	9,550	9,885	10,169	10,415	10,633	10,828	11,004	11,165
2	9,210	10,597	11,408	11,983	12,429	12,794	13,102	13,369	13,605	13,816	14,006	14,180
3	11,345	12,838	13,706	14,320	14,796	15,183	15,510	15,794	16,043	16,266	16,468	16,652
4	13,277	14,860	15,777	16,424	16,924	17,331	17,675	17,972	18,233	18,467	18,678	18,871
5	15,086	16,750	17,710	18,386	18,907	19,332	19,690	20,000	20,272	20,515	20,735	20,935
6	16,812	18,548	19,547	20,249	20,791	21,232	21,603	21,924	22,206	22,458	22,685	22,892
8	20,090	21,955	23,024	23,774	24,352	24,821	25,216	25,557	25,857	26,124	26,366	26,586
9	21,666	23,589	24,690	25,462	26,056	26,539	26,945	27,295	27,603	27,877	28,125	28,351
10	23,209	25,188	26,320	27,112	27,722	28,216	28,633	28,991	29,307	29,588	29,842	30,073
12	26,217	28,300	29,487	30,318	30,957	31,475	31,910	32,286	32,615	32,909	33,175	33,416
14	29,141	31,319	32,559	33,426	34,091	34,631	35,084	35,475	35,818	36,123	36,399	36,650
15	30,978	32,801	34,066	34,950	35,628	36,177	36,639	37,037	37,386	37,697	37,978	38,233
16	32,000	34,267	35,556	36,456	37,146	37,706	38,175	38,580	38,936	39,252	39,538	39,798

Nach Bonferroni adjustierte Z -, t - und F -Schranken enthält das Biometrical Journal 24 (1982), 239–255; 26 (1984), 351–381 und 28 (1986), 547–576.

7.6.9.1 Lokalisation der stochastischen Abhängigkeit nach Hommel

Wird anhand des χ^2 -Tests die globale Unabhängigkeitshypothese einer Kontingenztafel abgelehnt, so lässt sich die Art der Abhängigkeit nach Hommel und Mitarb. [HLP85] näher lokalisieren, wobei wir das von den Autoren gegebene Beispiel benutzen und bezüglich der Theorie auf die Originalarbeit verweisen (aus Sachs [Sac93]). Für das Beispiel einer 3×3 -Kontingenztafel folgt

22	3	25	50
14	23	9	46
10	27	20	57
46	53	54	153

$$\hat{\chi}^2 = 153 \left\{ \left[\frac{22^2}{50 \cdot 46} + \frac{3^2}{50 \cdot 53} + \dots + \frac{20^2}{57 \cdot 54} \right] - 1 \right\} = 30,13 > 9,49 = \chi_{4;0,95}^2$$

z.B. für das Feld „Zeile 1, Spalte 1“ mit der Besetzungszahl „22“ lässt sich die Nullhypothese $H_{11} : p_{11} = p_{1.}p_{.1}$ gegen die Alternativhypothese $A_{11} : p_{11} \neq p_{1.}p_{.1}$ prüfen. [Allgemein, $H_{ij} : p_{ij} = p_{i.}p_{.j}$ gegen $A_{ij} : p_{ij} \neq p_{i.}p_{.j}$ mit $i = 1, \dots, r$ (row, Zeile) und $j = 1, \dots, c$ (column, Spalte); die $p_{i,j}$ seien die Feld-(Zell-)Wahrscheinlichkeiten, die $p_{i.}$ und $p_{.j}$ die entsprechenden Randwahrscheinlichkeiten.]

Durch Zusammenfassung von Zeilen und Spalten erhält man die folgende „kollabierte“ Vierfeldertafel, für die das entsprechende $\hat{\chi}^2$ berechnet wird.

22	28	50
24	79	103
46	107	153

$$\hat{\chi}_{11}^2 = \frac{153(22 \cdot 79 - 28 \cdot 24)^2}{50 \cdot 103 \cdot 46 \cdot 107} = 6,6589$$

sowie der (mit FG=1) zugehörige P-Wert: $P_{11} = 0,0088$.

In ähnlicher Weise lassen sich die übrigen Nullhypothesen $H_{i,j}$ prüfen und man erhält das folgende Ergebnis.

$\hat{\chi}^2$ -Werte			P-Werte		
6,68	26,91	7,03	0,0088	0,0000	0,0080
0,00	6,85	7,13	0,9479	0,0088	0,0076
6,77	6,50	0,00	0,0092	0,0108	0,9672

Man ordnet die P-Werte Aufsteigend und adjustiert diese nach der Simes-Hochberg-Prozedur (vgl. auch Funktion **p.adjust()** in **R**).

Test P-Wert	P-adjustiert
1. $P_{12} = 0,0000$	0,0000
2. $P_{23} = 0,0076$	0,0324
3. $P_{13} = 0,0080$	0,0324
4. $P_{11} = 0,0088$	0,0324
5. $P_{22} = 0,0088$	0,0324
6. $P_{31} = 0,0092$	0,0324
7. $P_{32} = 0,0108$	0,0324
8. $P_{21} = 0,9479$	0,9672
9. $P_{33} = 0,9672$	0,9672

Somit sind (wie auch die P-Werte zu erwarten ließen) 7 von 9 Nullhypothesen auf Feldunabhängigkeit mit einer Irrtumswahrscheinlichkeit von 0,05 abzulehnen.

Allgemein lassen sich, sobald H_0 für eine $r \cdot c$ -Tafel mit $r \geq 3$ und $c \geq 3$ auf dem $100\alpha\%$ -Niveau abgelehnt wird, auch die folgenden sequentiellen Schranken einer modifizierten Holm-Prozedur verwenden: $\alpha/(r \cdot c - 4)$, $\alpha/(r \cdot c - 4)$, $\alpha/(r \cdot c - 4)$, $\alpha/(r \cdot c - 4)$, $\alpha/(r \cdot c - 4)$, $\alpha/(r \cdot c - 6)$, $\alpha/(r \cdot c - 6)$, $\alpha/(r \cdot c - 7)$, $\alpha/(r \cdot c - 8)$, $\dots$, $\alpha/2$, α .

7.6.9.2 Simultane Paarvergleiche nach Royen

Vorausgesetzt werden n unabhängige Stichproben (möglichst gleicher Umfänge) mit jeweils k Kategorien aus identischen Polynomialverteilungen. Verglichen werden:

- I. jeweils zwei Stichproben,
- II. eine Stichprobe mit einer Kontrolle (Kontrollstichprobe).

Für beide Typen simultaner Paarvergleiche werden von Royen [Roy84] 10%- und 5%-Schranken gegeben. Nähere Einzelheiten und Hinweise auf ein Programm sind der Originalarbeit zu entnehmen. Die folgende Darstellung orientiert sich an Beispielen, die uns der Autor überlassen hat (aus Sachs [Sac93]).

Im allgemeinen wird man die 5%-Schranken benutzen; für eine größere Anzahl von Stichproben kann man die 10%-Schranken verwenden, um zu erkunden, ob einzelne Stichproben auffallen. Bei wenigen Kategorien sollten fast alle Besetzungszahlen ≥ 5 sein. Die Stichprobenumfänge sollten bei beiden Paarvergleichstypen - wenn überhaupt - nur wenig differieren.

Tabelle 7.90. Schranken für simultane Paarvergleiche nach Royen, Th.(1984): Multiple comparisons of polynomial distributions. Biometrical Journal 26, 319-332. Mit freundlicher Erlaubnis

Niveau	m	$\nu = 1$	$\nu = 2$	$\nu = 3$	$\nu = 4$	$\nu = 5$
$\alpha = 0.10$	2	2,706	4,605	6,251	7,779	9,236
	3	4,21	6,46	8,36	10,09	11,72
	4	5,25	7,70	9,73	11,57	13,31
	5	6,06	8,63	10,75	12,68	14,49
	6	6,70	9,37	11,58	13,57	15,42
	7	7,26	10,00	12,26	14,29	16,18
	8	7,73	10,54	12,85	14,92	16,83
$\alpha = 0,05$	2	3,841	5,991	7,815	9,488	11,071
	3	5,49	7,94	9,97	11,82	13,57
	4	6,60	9,21	11,36	13,32	15,17
	5	7,44	10,17	12,41	14,46	16,34
	6	8,12	10,94	13,25	15,33	17,25
	7	8,69	11,58	13,94	16,04	18,00
	8	9,19	12,12	14,53	16,68	18,66

I. Simultane Paarvergleiche.

Die folgenden 4 Stichproben sind paarweise auf dem 5%-Niveau auf Homogenität zu prüfen.

	Stich- probe	Kategorie			Σ
		1	2	3	
⊙	1	51	30	19	100
⊙	2	30	41	29	100
	3	33	37	30	100
	4	19	31	50	100

$$\hat{\chi}_{1,2}^2 = 200 \left\{ \left[\frac{51^2}{100 \cdot 81} + \frac{30^2}{100 \cdot 71} + \frac{19^2}{100 \cdot 48} + \frac{30^2}{100 \cdot 81} + \frac{41^2}{100 \cdot 71} + \frac{29^2}{100 \cdot 48} \right] - 1 \right\} = 9,232$$

Für die kritischen Schranken aus Tabelle 7.90 gilt die Bezeichnung $\chi_{m;\nu;\alpha}^2$ mit m = der Zahl der Stichproben, $\nu = k - 1$ = Zahl der Kategorien minus 1 und α dem gewählten Signifikanzniveau.

$$\hat{\chi}_{1,2}^2 = 9,232 > 9,21 = \chi_{4;2;0,05}^2 \text{ aus Tabelle 7.90}$$

Die insgesamt 6 möglichen paarweisen Vergleiche werden entsprechend berechnet und übersichtlich in folgender Tabelle zusammengestellt (Schranke für alle Vergleiche $\chi_{4;2;0,05}^2 = 9,21$):

Vergleich	$\hat{\chi}_{1,2}^2$	$\hat{\chi}_{1,3}^2$	$\hat{\chi}_{1,4}^2$	$\hat{\chi}_{2,3}^2$	$\hat{\chi}_{2,4}^2$	$\hat{\chi}_{3,4}^2$
Teststatistik	9,232	7,058	28,573	0,365	9,441	9,299
Signifikanz: 5%-Niveau	*	-	*	-	*	*

Tabelle 7.91. 10%-Schranken für simultane Paarvergleiche mit einer Kontrolle nach Royen (1984 und 1985, persönl. Mitteilung). Mit freundlicher Erlaubnis

m	$\nu = 1$	$\nu = 1$	$\nu = 1$	$\nu = 1$	$\nu = 1$	$\nu = 1$	$\nu = 1$	$\nu = 1$	$\nu = 1$
2	3,672	5,805	7,615	9,277	10,850	12,363	13,830	15,262	16,667
3	4,252	6,507	8,404	10,136	11,771	13,338	14,855	16,334	17,782
4	4,667	7,003	8,957	10,738	12,413	14,017	15,568	17,078	18,555
5	4,990	7,386	9,384	11,199	12,905	14,536	16,112	17,646	19,145
6	5,254	7,698	9,729	11,573	13,303	14,956	16,552	18,104	19,620
7	5,478	7,961	10,020	11,886	13,636	15,307	16,920	18,487	20,017
8	5,671	8,187	10,271	12,156	13,923	15,610	17,236	18,816	20,358
9	5,842	8,387	10,491	12,393	14,175	15,874	17,513	19,104	20,657
10	5,995	8,565	10,687	12,604	14,399	16,110	17,759	19,360	20,922
11	6,133	8,726	10,864	12,794	14,601	16,322	17,980	19,590	21,161
12	6,258	8,872	11,024	12,967	14,784	16,514	18,181	19,799	21,377
13	6,374	9,006	11,172	13,125	14,952	16,691	18,366	19,991	21,575
14	6,481	9,130	11,308	13,272	15,106	16,853	18,535	20,167	21,758
15	6,581	9,245	11,435	13,407	15,250	17,004	18,693	20,331	21,927
16	6,674	9,353	11,553	13,534	15,384	17,145	18,840	20,483	22,085
17	6,762	9,454	11,663	13,652	15,510	17,277	18,977	20,626	22,233
18	6,844	9,549	11,767	13,764	15,628	17,401	19,106	20,760	22,371
19	6,922	9,639	11,866	13,869	15,739	17,517	19,228	20,886	22,502
20	6,996	9,724	11,959	13,969	15,844	17,628	19,343	21,006	22,626

II. Simultane Paarvergleiche mit einer Kontrolle.

Eine Kontrollstichprobe ist mit 5 Stichproben zu vergleichen; der Homogenitätstest ist auf dem 5%-Niveau durchzuführen.

	Stich- probe	Kategorie			Σ
		1	2	3	
⊙	Kontrolle 0	51	30	19	100
	1	30	41	29	100
	2	30	42	28	100
⊙	3	31	42	27	100
	4	33	36	31	100
	5	37	35	28	100

$$\hat{\chi}_{0,3}^2 = 200 \left\{ \left[\frac{51^2}{100 \cdot 82} + \frac{30^2}{100 \cdot 72} + \frac{19^2}{100 \cdot 46} + \frac{31^2}{100 \cdot 82} + \frac{42^2}{100 \cdot 72} + \frac{27^2}{100 \cdot 46} \right] - 1 \right\} = 8,269$$

Für die kritischen Schranken aus Tabelle 7.91 gilt die Bezeichnung $\chi_{m;\nu;\alpha}^2$ mit m = der Zahl der Stichproben (ohne die Kontrollstichprobe), $\nu = k - 1$ = Zahl der Kategorien minus 1 und α dem gewählten Signifikanzniveau.

$$\hat{\chi}_{0,3}^2 = 8,269 > 8,887 = \chi_{5;2;0,05}^2 \text{ aus Tabelle 7.91}$$

Die insgesamt 5 paarweisen Vergleiche gegen die Kontrollstichprobe werden entsprechend berechnet und übersichtlich in folgender Tabelle zusammengestellt (Schranke für alle Vergleiche $\chi_{5;2;0,05}^2 = 8,887$):

Vergleich	$\hat{\chi}_{0,1}^2$	$\hat{\chi}_{0,2}^2$	$\hat{\chi}_{0,3}^2$	$\hat{\chi}_{0,4}^2$	$\hat{\chi}_{0,5}^2$
Teststatistik	9,232	9,168	8,269	7,283	4,335
Signifikanz: 5%-Niveau	*	*	-	-	-

Hinweis: Bezeichnet man den Umfang der Kontrollstichprobe mit n_0 und den der i -ten Stichprobe mit n_i , so sollte für $m \leq 10$ gelten: $0,9 \leq (n_i/n_0) \leq 1,2$.

Für $\frac{1}{m} \sum_{i=1}^m n_i \approx n_0$ ist sogar $|(n_i/n_0) - 1| \leq 0,5$ zulässig.

7.6.9.3 Stärke des Zusammenhangs - Kontingenzkoeffizient

Der $\hat{\chi}^2$ -Wert einer Kontingenztafel sagt nichts aus über die Stärke des Zusammenhangs zwischen zwei Klassifikationsmerkmalen. Das ist leicht einzusehen, da er bei gegebenem Verhältnis der Häufigkeiten einer Tafel der Gesamtzahl der Beobachtungen proportional ist. Für Mehrfeldertafeln wird daher, wenn die Existenz des Zusammenhangs gesichert ist, als Maß der **Straffheit des Zusammenhangs** der Pearson'sche Kontingenzkoeffizient

$$CC = \sqrt{\frac{\hat{\chi}^2}{n + \hat{\chi}^2}} \quad (7.252)$$

benutzt.

Dieses Korrelationsmaß weist bei völliger Unabhängigkeit den Wert Null auf. Im Falle völliger Abhängigkeit der beiden qualitativen Variablen ergibt CC jedoch nicht 1, sondern einen Wert, der schwankend nach der Felderzahl der Kontingenztafel kleiner als 1 ist. Damit sind verschiedene

Tabelle 7.91. Fortsetzung - 5%-Schranken für simultane Paarvergleiche mit einer Kontrolle nach Royen (1984 und 1985, persönl. Mitteilung). Mit freundlicher Erlaubnis

m	$\nu = 1$	$\nu = 1$	$\nu = 1$	$\nu = 1$	$\nu = 1$	$\nu = 1$	$\nu = 1$	$\nu = 1$	$\nu = 1$
2	4,894	7,243	9,210	11,000	12,685	14,298	15,857	17,375	18,860
3	5,518	7,973	10,016	11,869	13,609	15,271	16,875	18,435	19,959
4	5,962	8,489	10,583	12,478	14,255	15,950	17,585	19,173	20,723
5	6,307	8,887	11,019	12,946	14,751	16,471	18,128	19,737	21,308
6	6,590	9,211	11,374	13,326	15,152	16,892	18,568	20,194	21,780
7	6,828	9,485	11,672	13,645	15,489	17,245	18,936	20,576	22,175
8	7,034	9,720	11,929	13,920	15,780	17,550	19,253	20,905	22,515
9	7,216	9,928	12,155	14,161	16,034	17,816	19,531	21,193	22,813
10	7,378	10,113	12,357	14,376	16,261	18,054	19,778	21,449	23,078
11	7,525	10,280	12,538	14,570	16,466	18,268	20,001	21,680	23,316
12	7,659	10,432	12,704	14,746	16,651	18,462	20,203	21,890	23,533
13	7,782	10,571	12,855	14,908	16,822	18,640	20,389	22,082	23,731
14	7,896	10,701	12,995	15,057	16,979	18,805	20,560	22,259	23,914
15	8,002	10,820	13,125	15,196	17,125	18,957	20,718	22,423	24,083
16	8,101	10,932	13,247	15,325	17,261	19,100	20,866	22,577	24,242
17	8,193	11,037	13,361	15,446	17,389	19,233	21,005	22,720	24,390
18	8,281	11,136	13,468	15,560	17,509	19,358	21,135	22,855	24,529
19	8,364	11,230	13,569	15,668	17,622	19,477	21,258	22,982	24,660
20	8,442	11,318	13,665	15,770	17,729	19,589	21,374	23,103	24,785

CC -Werte nur dann hinsichtlich ihrer Größenordnung vergleichbar, wenn sie an gleichgroßen Kontingenztafeln berechnet werden.

Dieser Nachteil des CC wird dadurch kompensiert, dass für jede mögliche Felderordnung einer Kontingenztafel der größtmögliche Kontingenzkoeffizient $CC_{\max}$, bekannt ist, so dass der gefundene CC relativ zu diesem ausgedrückt werden kann. Der größtmögliche Kontingenzkoeffizient $CC_{\max}$ ist dabei definiert als jener Wert, den CC für eine bestimmte Felderordnung der Tafel bei völliger Abhängigkeit der Variablen erreicht. Für quadratische Kontingenztafeln (Zahl der Zeilen = Zahl der Spalten, d. h. $r = c$) hat M.G. Kendall (1907–1983) gezeigt, dass der Wert von $CC_{\max}$ lediglich von der Klassenzahl r abhängig ist, es gilt

$$CC_{\max} = \sqrt{\frac{r-1}{r}} \quad (7.253)$$

Der korrigierte Kontingenzkoeffizient nach Pearson, kurz CC_{kor} mit $0 \leq CC_{\text{kor}} \leq 1$, gestattet den Vergleich **unterschiedlicher** Kontingenztafeln. Er wird berechnet nach

$$CC_{\text{kor}} = \sqrt{\frac{m}{m-1} \cdot \frac{\hat{\chi}^2}{\hat{\chi}^2 + n}} \quad (7.254)$$

$m = \min(r-1, c-1)$
die kleinere beider Zahlen

Mit der Bezeichnung $r \leq c$ lässt sich ein Kontingenzkoeffizient nach H. Cramér

$$K = \sqrt{\hat{\chi}^2 / (n[r-1])} \quad \text{mit} \quad 0 \leq K \leq 1$$

definieren; für die Vierfeldertafel ist $K = \sqrt{\hat{\chi}^2 / n} = \sqrt{\hat{\chi}_*^2 / (n-1)}$.



Beispiel:

Tafeltyp	n	$\hat{\chi}^2$	$CC = \sqrt{\frac{\hat{\chi}^2}{n + \hat{\chi}^2}}$	CC_{korrr}	$K = \sqrt{\hat{\chi}^2 / (n[r-1])}$
3 · 4	496	24,939	0,219	0,309	0,1586
3 · 3	120	21,577	0,390	0,544	0,2998

Hinweis: Zwei $\hat{\chi}^2$ -Werte aus Tafeln (Tabellen) mit gleichem Freiheitsgrad $FG = \nu \geq 3$ lassen sich nach

$$\hat{z} = \frac{\sqrt{\hat{\chi}_1^2} - \sqrt{\hat{\chi}_2^2}}{\sqrt{1 - [1/(4\nu)]}} \quad (7.255)$$

vergleichen [Psychological Bulletin **94** (1983), 173]. So erhält man für $\nu = 6$, $\hat{\chi}_1^2 = 24,939$ und $\hat{\chi}_2^2 = 9,018$ erwartungsgemäß $\hat{z} = 2,034 > 1,96$, d. h. $P = 2 \cdot 0,021 = 0,042$.

7.6.9.4 Prüfung auf Trend: Der Anteil der linearen Regression an der Gesamtvariation

Ist die Frage, ob die Verteilung qualitativer Merkmale nach einer Merkmalsreihe von der Einteilung nach einer zweiten Merkmalsreihe abhängig ist, aufgrund des bedeutsamen $\hat{\chi}^2$ -Wertes positiv beantwortet, dann kann man weiter untersuchen, ob die **Zunahme der Häufigkeiten regelmäßig** ist; anders gesagt, ob die Häufigkeiten in Abhängigkeit von einer Merkmalsreihe linear zunehmen oder ob dieser **Trend** komplizierterer Natur ist. Das $\hat{\chi}^2$ lässt sich dann – wie für die $k \cdot 2$ -Tafel gezeigt worden ist (Cochran-Armitage Test [7.6.8]) – in zwei Anteile zerlegen: Der eine mit einem FG entfällt auf die als linear ansteigend gedachten Häufigkeiten, der sogenannte Regressionsgeraden-Anteil – der restliche Anteil entspricht den Unterschieden zwischen den beobachteten Häufigkeiten und den als linear ansteigend vorausgesetzten theoretischen Häufigkeiten. Dieser Anteil wird wieder als Differenz zwischen $\hat{\chi}^2$ und $\hat{\chi}_{\text{Regression}}^2$ berechnet.

Durch Zuordnung von Punktwerten (scores), x - und y -Werten, werden beide Merkmale einer $r \cdot c$ -Tafel in ein möglichst einfaches Koordinatensystem überführt. Nach dieser „Quantifizierung“ der Daten wird die bivariate Häufigkeitstafel auf Korrelation beider Variablen untersucht. Praktisch geht man nach Yates [Yat48] so vor, dass man die Regression einer dieser Variablen auf die andere prüft: Man ermittelt den Regressionskoeffizienten b_{yx} (bzw. b_{xy}), die zugehörige Varianz $V(b_{yx})$ [bzw. $V(b_{xy})$] und testet die Signifikanz der linearen Regression nach

$$\hat{\chi}^2 = \frac{(b_{yx})^2}{V(b_{yx})} = \frac{(b_{xy})^2}{V(b_{xy})} \quad (7.256)$$

mit 1 FG . Der Regressionskoeffizient von Y auf X ist bestimmt durch

$$b_{yx} = \frac{\sum xy}{\sum x^2} \quad \text{Beachte den Abschnitt unter Formel (7.258a)} \quad (7.257)$$

der von X auf Y durch

$$b_{xy} = \frac{\sum xy}{\sum y^2} \quad (7.257a)$$

Die Varianzen beider Regressionskoeffizienten sind bei Annahme der Nullhypothese

$$V(b_{yx}) = \frac{s_y^2}{\sum x^2} = \frac{\sum y^2}{n \sum x^2} \quad (7.258)$$

$$V(b_{xy}) = \frac{s_x^2}{\sum y^2} = \frac{\sum x^2}{n \sum y^2} \quad (7.258a)$$

In diesen Gleichungen stellen die x - und y -Werte die **Abweichungen vom Mittelwert** der jeweiligen Variablen dar, s_y^2 ist eine Schätzung der Varianz der Variablen Y , s_x^2 eine Schätzung der Varianz der Variablen X . Zur Berechnung der Ausdrücke (7.256 bis 7.258a) werden drei Häufigkeitsverteilungen – die der Größen x , y und $(x - y)$ – benötigt: Man erhält dann $\sum x^2$, $\sum y^2$, und $\sum (x - y)^2$.

Beispiel (Vergleich der Therapieformen aus Tabelle 7.87): Nach Zuordnung der Punktwerte (scores) zu den Kategorien der beiden Merkmale (s. unten) werden die Produkte gebildet aus den Randsummen und den entsprechenden Punktwerten sowie aus den Randsummen und den Quadraten der Punktwerte. Die Summen dieser Produkte sind (vgl. die Symbolik von Tabelle 7.86):

$$\begin{aligned} \sum n_{i.}y &= 58, \sum n_{i.}y^2 = 78 \\ \sum n_{.j}x &= 0, \sum n_{.j}x^2 = 80 \end{aligned}$$

Diese Produktsummen liefern $\sum x^2$ und $\sum y^2$ nach

$$\begin{aligned} \sum y^2 &= \sum n_{i.}y^2 - \frac{(\sum n_{i.}y)^2}{\sum n_{i.}} = 78 - \frac{58^2}{120} = 49,967 \\ \sum x^2 &= \sum n_{.j}x^2 - \frac{(\sum n_{.j}x)^2}{\sum n_{.j}} = 80 - \frac{0^2}{120} = 80 \end{aligned}$$

Tabelle 7.92. Punktwerte zu den Beispieldaten Therapieformen

Beide Reihen von Punktwerten müssen symmetrisch zum Wert Null liegen

!

x y \ score	-1	0	1	$n_{i.}$	$n_{i.}y$	$n_{i.}y^2$
1	14	22	32	68	68	68
0	18	16	8	42	0	0
-1	8	2	0	10	-10	10
$n_{.j}$	40	40	40	120	58	78
$n_{.j}x$	-40	0	40	0		
$n_{.j}x^2$	40	0	40	80		

Zur Berechnung von $\sum (x - y)^2$ wird die entsprechende Häufigkeitsverteilung (vgl. Tabelle 7.93) notiert. Die Spalte 2 dieser Tabelle enthält die „Diagonalsummen“ der Tabelle 7.92. Es sind die „Diagonalsummen“ von links unten nach rechts oben zu nehmen. Man erhält also $14, 18 + 22 = 40, 8 + 16 + 32 = 56, 2 + 8 = 10$ und 0 .

Spalte 1 enthält die Differenzen $x - y$ für alle Felder der Tabelle 7.92, jeweils die der „Diagonale“ zusammengefasst, da diese identische $(x - y)$ -Werte aufweisen: Beispielsweise erhält man für alle



Tabelle 7.93. Zur Berechnung von $\sum (x - y)^2$

$x - y$	$n_{\text{Diag.}}$	$n_{\text{Diag.}}(x - y)$	$n_{\text{Diag.}}(x - y)^2$
$-1 - (+1) = -2$	14	-28	56
$0 - 1 = -1 - 0 = -1$	40	-40	40
$1 - 1 = 0 - 0 = -1 - (-1) = 0$	56	0	0
$1 - 0 = 0 - (-1) = +1$	10	10	10
$1 - (-1) = +2$	0	0	0
Insgesamt	120	-58	106

Felder der Diagonalen von links unten nach rechts oben, d. h. für die Felder mit den Besetzungszahlen 8, 16, 32 für die Differenz $x - y$ den Wert Null:

$$\begin{aligned} \text{für Feld „8“, links unten: } x &= -1, \quad y = -1 \\ x - y &= -1 - (-1) = -1 + 1 = 0 \end{aligned}$$

$$\begin{aligned} \text{für Feld „16“, Tafelmitte: } x &= 0, \quad y = 0 \\ x - y &= 0 - 0 = 0 \end{aligned}$$

$$\begin{aligned} \text{für Feld „32“, rechts oben: } x &= 1, \quad y = 1 \\ x - y &= 1 - 1 = 0 \end{aligned}$$

d. h. $x - y = 0$ gilt für $8 + 16 + 32 = 56$ usw.

Aus den Summen der Produkte erhält man:

$$\begin{aligned} \sum (x - y)^2 &= \sum n_{\text{Diag.}}(x - y)^2 - \frac{[\sum n_{\text{Diag.}}(x - y)]^2}{\sum n_{\text{Diag.}}} \\ &= 106 - \frac{(-58)^2}{120} = 77,967 \end{aligned}$$

Wir erhalten dann nach (7.256, 7.257, 7.258)

$$\hat{\chi}^2 = \frac{(b_{yx})^2}{V(b_{yx})} = \frac{[(80 + 49,967 - 77,967)/(2 \cdot 80)]^2}{49,967/(120 \cdot 80)} = 20,2935$$

oder nach (7.256, 7.257a, 7.258a)

$$\hat{\chi}^2 = \frac{(b_{xy})^2}{V(b_{xy})} = \frac{[(80 + 49,967 - 77,967)/(2 \cdot 49,967)]^2}{80/(120 \cdot 49,967)} = 20,2935$$

Die ausgeprägte statistische Signifikanz beider Regressionskoeffizienten (vgl. Tab. 7.87 und Tab. 7.89: $\tau = 4$, $\alpha = 0,01$, $\nu = 1$, d. h. $\hat{\chi}^2 = 9,141$) ließe sich auch über die Standardnormalverteilung ermitteln:

$$\begin{aligned} \hat{z} &= \frac{b_{yx}}{\sqrt{V(b_{yx})}} = \frac{0,325000}{0,005205} = 4,505 \\ \hat{z} &= \frac{b_{xy}}{\sqrt{V(b_{xy})}} = \frac{0,520343}{0,013342} = 4,505 \end{aligned} \tag{7.259}$$

Auch hier ist natürlich die statistische Signifikanz gesichert.

Stellen wir die Ergebnisse in einer Übersicht zusammen, dann zeigt sich, dass die Abweichungen der Besetzungszahlen der Tabelle 7.92 von der Proportionalität fast vollständig durch die Existenz einer linearen Regression bedingt sind; mit der Therapie der doppelten Normaldosis steigt der therapeutische Erfolg markant an. Wenn diese Feststellung auch banal klingt, so darf nicht übersehen werden, dass sie erst aufgrund dieser Übersicht ihre eigentliche Bedeutung erhält (vgl. $\hat{\chi}_{1;0,001}^2 = 10,828$).

Variationsursache	$\hat{\chi}^2$	FG	Signifikanzniveau
lineare Regression	20,2935	1	$P < 0,001$
Abweichung von der Regression	1,2830	3	$P \approx 0,73$
Gesamtvariation	21,5765	4	$P < 0,001$

Besteht das Bedürfnis, Regressionslinien einander entsprechender Tafeln zu vergleichen, so prüft man anhand von (7.260), ob sich die Regressionskoeffizienten unterscheiden (Fairfield Smith [FS57]).

$$\hat{z} = \frac{|b_1 - b_2|}{\sqrt{V(b_1) + V(b_2)}} \quad (7.260)$$

Die Bedeutsamkeit des Unterschieds wird anhand der Standardnormalverteilung entschieden.

Beispiel: Angenommen, die in den Tabellen 7.87 und 7.92 dargelegten Besetzungszahlen seien an Personen eines Geschlechtes, einer Altersgruppe usw. ermittelt worden und uns stünde das Ergebnis eines entsprechenden Versuches zur Verfügung, das an Personen einer anderen Altersgruppe gewonnen wurde (wir prüfen zweiseitig):

$$\begin{aligned} b_1 &= 0,325 & b_2 &= 0,079 \\ V(b_1) &= 0,00521 & V(b_2) &= 0,00250 \end{aligned}$$

$$\text{Dann ist mit } \hat{z} = \frac{|0,325 - 0,079|}{\sqrt{0,00521 + 0,00250}} = 2,80 \quad [P(Z \geq 2,80) = 0,00256;$$

zweiseitige Fragestellung: $P = 2 \cdot 0,00256 = 0,0051$] die Nullhypothese auf Gleichheit der Regressionskoeffizienten auf dem 1%-Niveau abzulehnen.

7.6.10 Bowker-Test auf Symmetrie in quadratischen Mehrfeldertafeln

Der McNemar-Test gab uns die Möglichkeit, zu prüfen, ob eine 2·2-Tafel bezüglich ihrer Diagonalen symmetrisch ist. Ein analoger Symmetrie-Test für eine $r \cdot r$ -Tafel stammt von Bowker [Bow48]. Dieser Test prüft die Nullhypothese, dass je zur Hauptdiagonalen, der Diagonalen mit den größten Besetzungszahlen, symmetrisch gelegene Felder gleiche Häufigkeiten aufweisen. Unter der Nullhypothese (Symmetrie) erwarten wir, dass

$$B_{ij} = B_{ji}, \text{ wobei}$$

B_{ij} = beobachtete Häufigkeit in dem von der i -ten Zeile und der j -ten Spalte gebildeten Feld,

B_{ji} = beobachtete Häufigkeit in dem von der j -ten Zeile und der i -ten Spalte gebildeten Feld.



Zur Entscheidung der Frage, ob die Nullhypothese aufrechterhalten werden kann, berechnet man

$$\hat{\chi}_{\text{sym}}^2 = \sum_{j=1}^{r-1} \sum_{i>j} \frac{(B_{ij} - B_{ji})^2}{B_{ij} + B_{ji}} \quad \text{mit } FG = r(r-1)/2. \quad (7.261)$$

Man bilde alle Differenzen der symmetrisch gelegenen Besetzungszahlen, für die $i > j$ ist, quadriere sie, dividiere sie durch die Summe der Besetzungszahlen und addiere die $r(r-1)/2$ Glieder. Wenn nicht mehr als $1/5$ der $r \cdot r$ Felder Erwartungshäufigkeiten $E < 3$ aufweisen, darf man ohne Bedenken nach $\hat{\chi}_{\text{sym}}^2$ testen (vgl. auch Ireland, Ku und Kullback [IKK69], Bennett [Ben72] sowie Hettmansperger und McKean [HM73]).



Beispiel:

0	10	16	15	41
4	2	10	4	20
12	4	3	6	25
8	4	3	6	14
24	20	30	26	100

Da $(0 + 2 + 3 + 1)$ kleiner ist als $(8 + 4 + 10 + 15)$, verläuft die Hauptdiagonale von links unten nach rechts oben.

$$\hat{\chi}_{\text{sym}}^2 = \frac{(12-4)^2}{12+4} + \frac{(4-1)^2}{4+1} + \frac{(0-1)^2}{0+1} + \frac{(2-3)^2}{2+3} + \frac{(10-6)^2}{10+6} + \frac{(16-4)^2}{16+4} = 15,2$$

Die Tabelle enthält 4 Zeilen und Spalten, damit stehen $4(4-1)/2 = 6$ Freiheitsgrade zur Verfügung. Das entsprechende $\chi_{0,95}^2$ beträgt 12,59; die Nullhypothese auf Symmetrie ist daher mit einer Irrtumswahrscheinlichkeit von 5% abzulehnen.

Ein an einer größeren Personengruppe durchgeführter Vergleich der Intensität von Hand- und Fußschweiß führt ebenso wie eine Gegenüberstellung der Sehschärfen des linken und rechten Auges und ein Vergleich hinsichtlich der Schulbildung oder Hobbies von Ehepartnern zu typischen Symmetrieproblemen. Darüber hinaus bietet fast jede quadratische Mehrfeldertafel, die auf Symmetrie geprüft wird, interessante Aspekte: So zeigt Tabelle 7.87 eine deutliche Asymmetrie

$$\left[\hat{\chi}_{\text{sym}}^2 = \frac{(18-2)^2}{18+2} + \frac{(14-0)^2}{14+0} + \frac{(22-8)^2}{22+8} = 33,333 > 16,266 = \chi_{3;0,999}^2 \right],$$

sie ist bedingt durch den bei einfacher und besonders bei doppelter Normaldosis stark reduzierten Anteil gestorbener und langsam genesender Patienten.

Q-Test nach Cochran

Ein anderer Test aus der Klasse der Symmetrietests ist der Q-Test nach Cochran [Coc50]; vgl. Tab. 7.94, vgl. auch Bennett [Ben67] sowie Tate und Brown [TB70], ein **Homogenitätstest für mehrere verbundene Stichproben** (z. B. Behandlungsarten oder Zeitpunkte) von Alternativdaten (+; -). H_A : mindestens zwei der verbundenen Stichproben (v.S.) entstammen unterschiedlichen Grundgesamtheiten; H_0 (alle entstammen einer gemeinsamen Grundgesamtheit) wird, nachdem Zeilen, die nur „+“ oder nur „-“ enthalten, gestrichen werden, für die dann verbleibende Tabelle mit $ns \geq 24$ Zeichen, wobei $n \geq 4$ gelten muss, auf dem 100%-Niveau verworfen, sobald das nach (7.262) berechnete Q größer ist als der entsprechende kritische aus der χ^2 -Verteilung mit $(s-1)$ Freiheitsgraden.

Tabelle 7.94

I.	v. S.	verb. Stichpr.				Σ
		1	2	j	s	
Individuen	1					L_i
	2					
	.					
	i					
	.					
	n					
Σ		T_j				

$$Q = \frac{(s-1) \left[s \sum_{j=1}^s T_j^2 - \left(\sum_{j=1}^s T_j \right)^2 \right]}{s \sum_{i=1}^n L_i - \sum_{i=1}^n L_i^2} > \chi_{s-1; 1-\alpha}^2$$

(7.262)

L_i = Summe der Pluszeichen des Individuums i über alle s verbundenen Stichproben.

T_j = Summe der Pluszeichen der n Individuen für die Behandlung j .



Beispiel: Sind fünf Weine ($A, \dots, E$) von 6 Experten zu beurteilen oder fünf Operationsmethoden in 6 Kliniken oder interessiert an 6 Schülern bzw. Schulklassen, das für fünf Altersstufen erfasste Interesse an einer bestimmten Sportart, einem bestimmten Unterrichtsfach oder einem bestimmten Auto ($n \cdot s = 5 \cdot 6 = 30 > 24$), so resultiere z.B.

Person	Weine					Summe
	A	B	C	D	E	
1	1	0	1	1	0	3
2	1	1	1	0	1	4
3	0	0	1	1	1	3
4	1	0	1	0	0	2
5	0	0	0	1	1	2
6	1	0	1	1	0	3
Summe	4	1	5	4	3	17

$$Q = \frac{(5-1)[5(4^2 + 1^2 + 5^2 + 4^2 + 3^2) - 17^2]}{5 \cdot 17 - [3^2 + 4^2 + 3^2 + 2^2 + 2^2 + 3^2]} = 5,412 < 9,488 = \chi_{4;0,095}^2$$

Anhand der vorliegenden Alternativdaten lässt sich die Nullhypothese auf dem 5%-Niveau nicht ablehnen.

Hinweis: Multiple Vergleiche sind möglich: man prüft jeweils 2 der s Behandlungen simultan anhand des McNemar-Tests auf dem $[100\alpha/\binom{s}{2}]$ -Niveau, d. h. für $\alpha = 0,05$ und $s = 4$, $\binom{4}{2} = 6$, auf dem 0,83%- oder 1%-Niveau.

7.6.11 Cohen's Kappa-Koeffizient κ

Die Bewertung der Zuverlässigkeit (Reliabilität, reliability) bei der Bestimmung stetiger Merkmale (Messwerte) kann durch das Bland-Altman Verfahren (vgl. Abschnitt [6.15]) erfolgen. Für kategoriale Merkmale wird dagegen ein anderer Ansatz gewählt, der allgemein als Bewertung „Zufallskorrigierter Übereinstimmung“ (agreement) bezeichnet werden kann (Krummenauer [Kru99]). Diese Zufallskorrektur soll die zufällige Übereinstimmung (zufällig zu erwarten) (P_e) gegenüber der beobachteten Übereinstimmung (P_o) erfassen und kann allgemein durch (7.263) geschrieben werden (Cohen [Coh60]).

$$\kappa = \frac{P_o - P_e}{1 - P_e} \quad (7.263)$$

Für den einfachen Fall eines „dichotomen“ Merkmals, z.B. die Angaben von zwei Beobachtern (A und B) hinsichtlich einer Alternative (positiv und negativ, +/-), kann eine Schätzung für das κ (Kappa) aus (7.263) aus den Werten einer Vierfeldertafel erfolgen.

Beobachter A	Beobachter B		Summe
	+	-	
+	n_{11}	n_{12}	$n_{1.}$
-	n_{21}	n_{22}	$n_{2.}$
Summe	$n_{.1}$	$n_{.2}$	$n_{..} = n$

Die beobachtete (observed) und die erwartete (expected) Übereinstimmung lassen sich schätzen durch

$$\begin{aligned} p_o &= \frac{1}{n}(n_{11} + n_{22}) \\ p_e &= \frac{1}{n}(n_{1.}n_{.1} + n_{2.}n_{.2}) \end{aligned} \quad (7.264)$$

oder allgemeiner auch für eine quadratische Tafel ($k \times k$) mit den entsprechenden relativen Häufigkeiten:

$$\begin{aligned} p_o &= \sum_{i=1}^k p_{ii} \\ p_e &= \sum_{i=1}^k p_{i.}p_{.i} \end{aligned} \quad (7.265)$$

Mit (7.264) bzw. (7.265) kann eine Schätzung für das κ einfach hergeleitet werden.

$$\hat{\kappa} = \frac{p_o - p_e}{1 - p_e} = \frac{2(n_{11}n_{22} - n_{12}n_{21})}{n_{1.}n_{.2} + n_{.1}n_{2.}} \quad (7.266)$$

Hinweise zur Bewertung von $\hat{\kappa}$:

- Die vollständige Übereinstimmung wird durch den Wert $\kappa = 1$ beschrieben.
- $\kappa = 0$ bedeutet, dass die beobachtete Übereinstimmung (zufallsmäßig) zu erwarten war (keine Übereinstimmung).
- Der Wert für κ kann abhängig von der Verteilung der Randsummen auch kleiner werden als 0 ($-1 \leq \kappa \leq 0$) und ist dann schwer zu interpretieren.
- κ ist ein sehr allgemein definiertes Maß für die Übereinstimmung, das nicht zwischen verschiedenen Arten und Ursachen für eine fehlende Übereinstimmung unterscheidet.
- κ wird durch die Verteilung der Randsummen (Prävalenzen) beeinflusst und ist daher in der Regel nicht zwischen verschiedenen Studien/Populationen vergleichbar.
- Für die Beurteilung der Stärke einer Übereinstimmung kann die folgende Tabelle verwendet werden (Landis [LK77a]).

κ	Übereinstimmung
< 0	keine
0,10 - 0,40	schwache
0,41 - 0,60	deutliche
0,61 - 0,80	starke
0,81 - 1,00	fast vollständige

Die asymptotische Varianz für die Schätzung des Kappa-Wertes $\hat{\kappa}$ kann nach Fleiss, Cohen und Everitt [FCE69] durch (7.267) berechnet werden.

$$\begin{aligned} VAR(\hat{\kappa}) &= \left[(A + B - C) / ((1 - p_e)^2 \cdot n) \right] \\ \text{mit } A &= \sum_i p_{ii} \left[1 - (p_{i.} + p_{.i})(1 - \hat{\kappa}) \right]^2 \\ B &= (1 - \hat{\kappa})^2 \sum \sum_{i \neq j} p_{ij} (p_{i.} + p_{.j})^2 \\ C &= \left[\hat{\kappa} - p_e(1 - \hat{\kappa}) \right] \end{aligned} \quad (7.267)$$

Mit Hilfe der Varianz kann ein asymptotisches $(1 - \alpha)100\%$ -Konfidenzintervall für κ bestimmt werden.

$$\hat{\kappa} \pm z_{1-\alpha/2} \cdot \sqrt{VAR(\hat{\kappa})} \quad (7.268)$$

Für den Test der Nullhypothese $H_0: \kappa = 0$ kann nach Fleiss [Fle81] die Teststatistik (7.269) verwendet werden. Dabei vereinfacht sich die Schätzung der Varianz unter der Nullhypothese.

$$\begin{aligned} \hat{z} &= \frac{\hat{\kappa}}{\sqrt{VAR_{H_0}(\hat{\kappa})}} \\ VAR_{H_0}(\hat{\kappa}) &= \frac{p_e + p_e^2 - \sum_i p_{i.} p_{.i} (p_{i.} + p_{.i})}{(1 - p_e)^2 \cdot n} \end{aligned} \quad (7.269)$$

Beispiel: Die folgende Tabelle gibt die Beurteilung von 40 Schülern hinsichtlich ihrer Aufmerksamkeit (+; −) im Unterricht durch zwei Beobachter (A; B) wieder.



Beobachter A	Beobachter B		Summe
	+	-	
+	14	3	17
-	5	18	23
Summe	19	21	40

$$\hat{\kappa} = \frac{2(14 \cdot 18 - 3 \cdot 5)}{19 \cdot 23 + 17 \cdot 21} = 0,60$$

Es besteht erwartungsgemäß eine deutliche Übereinstimmung.

Für die Berechnung des κ -Indexes gibt es in **R** eine spezielle Funktion **Kappa()** in dem Paket **vcd** (Visualizing Categorical Data) [MZKH05]. Neben der Schätzung für κ wird in dieser Funktion auch ein asymptotischer Standardfehler berechnet und ein 95%-Konfidenzintervall angegeben.

```
> attention <- matrix(c(14, 3, 5, 18), nrow=2, ncol=2, byrow=TRUE); attention
     [,1] [,2]
[1,]  14   3
[2,]   5  18
> library(vcd)
> Kappa(attention)
      value      ASE      lwr      upr
Unweighted 0.5969773 0.1274470 0.3471859 0.8467688
```

7.6.11.1 Das gewichtete κ (Kappa)

Die Übertragung des Kappa-Koeffizienten für den Fall mehrkategorialer Merkmale (mit k Ausprägungen) ist ohne weiteres möglich (vgl. Darstellung in (7.265)). Bei der Beurteilung der Übereinstimmung wird dabei jedoch unterstellt, dass Nichtübereinstimmungen immer gleich zu bewerten sind, egal welche Kategorien davon betroffen sind. Für den Fall ordinaler Merkmale, lässt sich dann ein gewichteter Kappa-Koeffizient definieren (7.271). Dabei wird jedem Feld der Kontingenztafel ein Gewicht w_{ij} zugewiesen, mit $0 \leq w_{ij} < 1$ für $i \neq j$ und $w_{ii} = 1$ (Diagonale). Praktisch erfolgt die Festlegung der Gewichte nach den beiden folgenden Ansätzen.

$$\begin{aligned} w_{ij} &= 1 - \frac{|i - j|}{k - 1} && \text{linear} \\ w_{ij} &= 1 - \frac{(i - j)^2}{(k - 1)^2} && \text{quadratisch; Fleiss-Cohen} \end{aligned} \quad (7.270)$$

Für ein Merkmal mit $k = 4$ Kategorien resultiert daraus die folgende Tabelle mit den entsprechenden Gewichten. Es ist zu erkennen, dass die quadratische Gewichtung fehlende Übereinstimmungen stärker „ahndet“ als die lineare Gewichtung.

linear					quadratisch				
w_{ij}	$j=1$	2	3	4	w_{ij}	$j=1$	2	3	4
$i=1$	1	0,89	0,56	0	$i=1$	1	0,67	0,33	0
2	0,89	1	0,89	0,56	2	0,67	1	0,67	0,33
3	0,56	0,89	1	0,89	3	0,33	0,67	1	0,67
4	0	0,56	0,89	1	4	0	0,33	0,67	1

Die Berechnung eines gewichteten $\hat{\kappa}_w$ erfolgt dann nach:

$$\begin{aligned} \hat{\kappa}_w &= \frac{p_{o(w)} - p_{e(w)}}{1 - p_{e(w)}} \\ \text{mit } p_{o(w)} &= \sum_i \sum_j w_{ij} p_{ij} \\ p_{e(w)} &= \sum_i \sum_j w_{ij} p_{i.} p_{.j} \end{aligned} \quad (7.271)$$

Fleiss, Cohen und Everitt [FCE69] geben auch für dieses κ -Maß eine Schätzung der asymptotischen Varianz an, auf deren Grundlage entsprechende Konfidenzintervalle oder Teststatistiken (vgl. oben) abgeleitet werden können. Die Anwendung soll an einem Beispiel in **R** mit der Funktion **Kappa()** aus dem Paket **vcd** [MZKH05] gezeigt werden.



Beispiel: Botulinum A hemmt die Ausprägung mimischer Gesichtsfalten. Zur Messung des Behandlungserfolges wird häufig ein klinischer Score mit den Werten 0 (keine Falten) bis IV (maximale Faltenausprägung) verwendet. In der folgenden Tabelle sind die Bewertungen von zwei Untersuchern bei 49 Behandlungen zusammengefasst worden:

	0	I	II	III	IV
0	5	2		1	
I	1	7	2	2	
II	1	2	10	5	1
III			3	4	
IV					3

```

> botulin <- matrix(c(5,2,0,1,0, 1,7,2,2,0, 1,2,10,5,1, 0,0,3,4,0, 0,0,0,0,3),
+                    nrow=5, ncol=5, byrow=TRUE); botulin
     [,1] [,2] [,3] [,4] [,5]
[1,]    5    2    0    1    0
[2,]    1    7    2    2    0
[3,]    1    2   10    5    1
[4,]    0    0    3    4    0
[5,]    0    0    0    0    3
> library(vcd)
> Kappa(botulin, weights = "Fleiss-Cohen")
      value      ASE      lwr      upr
Unweighted 0.4650655 0.09202092 0.2847078 0.6454232
Weighted    0.6848875 0.15332909 0.3843680 0.9854070

```

Die Schätzung für das gewichtete Kappa $\hat{\kappa}_w$ ist 0,685; das 95%-Konfidenzintervall für κ_w auf der Basis des asymptotischen Standardfehlers ist $[0,384; 0,985]$.

7.6.11.2 Das κ für mehrfache Beurteilungen (Multi-Rater)

Für den Fall, dass mehr als zwei Beobachter oder Verfahren ($R > 2$) hinsichtlich einer übereinstimmenden Bewertung eines Merkmals mit k Ausprägungen zu untersuchen sind, muss ein Ansatz auf der Grundlage des folgenden Schemas gewählt werden.

Tabelle 7.95. Berechnung von $\hat{\kappa}_m$: Übereinstimmung in mehrfachen Beurteilungen

Fall (i) (Patient)	Beobachtung (j)					$p_i = \sum_{j=1}^k \frac{n_{ij}(n_{ij} - 1)}{R(R - 1)}$
	1	2	3	...	k	
1	n_{11}	n_{12}	n_{13}	...	n_{1k}	p_1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	n_{n1}	n_{n2}	n_{n3}	...	n_{nk}	p_n
Summe	$n_{.,1}$	$n_{.,2}$	$n_{.,3}$	...	$n_{.,k}$	$\bar{p} = (\sum_i p_i)/n$
$\hat{p}_j$	$\frac{n_{.,1}}{nR}$	$\frac{n_{.,2}}{nR}$	$\frac{n_{.,3}}{nR}$	...	$\frac{n_{.,k}}{nR}$	$\bar{p}_e = \sum_j \hat{p}_j^2$

In der Tabelle 7.95 bezeichnet

R - die Anzahl der Beobachter / Untersucher.

n - die Anzahl der untersuchten Fälle.

k - die Anzahl der Kategorien des untersuchten Merkmals.

n_{ij} - die Anzahl der Bewertungen / Nennungen für den i ten Fall in der j ten Kategorie; insbesondere ist $\sum_j n_{ij} = R$ für alle i .

p_i - der Anteil der beobachteten übereinstimmenden Bewertungen für den i ten Fall; insbesondere entspricht $\bar{p} = (\sum_i p_i)/n$ der „beobachteten“ (mittleren) Übereinstimmung über alle Fälle.

$\hat{p}_j$ - der Anteil der Übereinstimmungen in der j ten Kategorie; insbesondere entspricht $\bar{p}_e = \sum_j \hat{p}_j^2$ einer „geschätzten“ (mittleren) Übereinstimmung über alle Kategorien.

Auf der Grundlage von Tabelle 7.95 kann entsprechend (7.263) eine Maßzahl $\hat{\kappa}_m$ für die Übereinstimmung mehrerer Untersucher (Multi-Rater) mit einer asymptotischen Varianz geschätzt werden werden (Fleiss [Fle81] und Landis [LK77b]).

$$\hat{\kappa}_m = \frac{\bar{p} - \bar{p}_e}{1 - \bar{p}_e}$$

$$VAR(\hat{\kappa}_m) = \frac{2}{nR(R-1)} \frac{\bar{p}_e - (2R-3)\bar{p}_e^2 + 2(R-2) \sum_{j=1}^k \hat{p}_j^3}{(1 - \bar{p}_e)^2} \quad (7.272)$$



Beispiel: In der folgenden Tabelle sind die hypothetischen Daten zur Beurteilung der Bilder aus der radiologischen Diagnostik von 10 Patienten ($n = 10$) durch 5 Untersucher („Zeilensumme“: $R = 5$) in den Kategorien gutartig, verdächtig und bösartig ($k = 3$) zusammengefasst:

Patient (i)	Bewertung			p_i
	gutartig	verdächtig	bösartig	
1	1	4	0	0,60
2	2	0	3	0,40
3	0	0	5	1,00
4	4	0	1	0,60
5	3	0	2	0,40
6	1	4	0	0,60
7	5	0	0	1,00
8	0	4	1	0,60
9	1	0	4	0,60
10	3	0	2	0,40
Gesamt	20	12	18	$\bar{p} = 0,62$
$\hat{p}_j$	0,40	0,24	0,36	$\bar{p}_e = 0,35$

Die Berechnung soll mit **R** elementar anhand der Tabelle 7.95 nach (7.272) erfolgen.

```
> radiol <- matrix(c(1,4,0, 2,0,3, 0,0,5, 4,0,1, 3,0,2,
+ 1,4,0, 5,0,0, 0,4,1, 1,0,4, 3,0,2),
+ nrow=10, ncol=3, byrow=TRUE)
> n <- 10; R <- 5; k <- 3;
> p.i <- rep(NA, n);
> for (i in 1:n) p.i[i] <- sum(radiol[i,]*(radiol[i,]-1))/(R*(R-1))
> p.bar <- sum(p.i)/n; p.bar
[1] 0.62
> p.j <- rep(NA, k); for (j in 1:k) p.j[j] <- sum(radiol[,j])/(n*R)
> p.e <- sum(p.j^2); p.e
[1] 0.3472
> kappa.m <- (p.bar - p.e)/(1-p.e); kappa.m
[1] 0.4178922
> var <- (2/(n*R*(R-1))) * (p.e - (2*R-3)*p.e^2 + 2*(R-2)*sum(p.j^3))/(1-p.e)^2; var
[1] 0.005872261
> z <- kappa.m / sqrt(var); z
[1] 5.453327
> 2*pnorm(z, lower.tail=FALSE)
[1] 4.943598e-08
```

Der Kappa-Koeffizient $\hat{\kappa}_n$ hinsichtlich der Übereinstimmung der 5 Untersucher bei der Bewertung der Bilder hat den Wert 0,42. Der Standardfehler für diese Schätzung ist

$$\sqrt{VAR(\hat{\kappa}_m)} = \sqrt{0,006} = 0,077$$

Die Nullhypothese $H_0: \kappa_m = 0$ kann entsprechend (7.269) mit der Teststatistik $\hat{z} = \hat{\kappa}_m / \sqrt{VAR(\hat{\kappa}_m)} = 5,45$ anhand der Standardnormalverteilung abgelehnt werden.

Mit dem Paket **irr** [Gam05] können in **R** verschiedene Maßzahlen zur Bewertung von Reliabilität bzw. Übereinstimmung bestimmt werden. Die Funktion **kappam.fleiss()** berechnet das κ_m nach (7.272) aus den „Rohdaten“. In dem Beispiel wird die Bewertung der Bilder verschlüsselt mit 1 gutartig, 2 verdächtig und 3 bösartig.

```
> library(irr)
> data <- matrix(c(1,2,2,2,2, 1,1,3,3,3, 3,3,3,3,3, 1,1,1,1,3, 1,1,1,3,3,
+ 1,2,2,2,2, 1,1,1,1,1, 2,2,2,2,3, 1,3,3,3,3, 1,1,1,3,3),
+               nrow=10, byrow=T,
+               dimnames=list(Bild=1:10, Untersucher=c("U1", "U2", "U3", "U4", "U5"))); data
```

Bild	U1	U2	U3	U4	U5
1	1	2	2	2	2
2	1	1	3	3	3
3	3	3	3	3	3
4	1	1	1	1	3
5	1	1	1	3	3
6	1	2	2	2	2
7	1	1	1	1	1
8	2	2	2	2	3
9	1	3	3	3	3
10	1	1	1	3	3

```
> kappam.fleiss(data, exact = FALSE, detail = FALSE)
Fleiss' Kappa for m Raters

Subjects = 10
Raters = 5
Kappa = 0.418

      z = 5.45
p-value = 4.94e-08
```

Das Ergebnis stimmt mit den zuvor elementar berechneten Resultaten überein.

7.7 Hypothesentests zur Korrelation und Regression

- Prüfung des Vorhandenseins einer Korrelation
- z -Transformation nach R.A. Fisher
- Weitere Anwendungen der z -Transformation
- Der Vergleich mehrerer Korrelationskoeffizienten
- Prüfung der Linearität einer Regression
- Prüfung der Regressionsparameter
- Prüfung des Rangkorrelationskoeffizienten ρ_s

Die Korrelationsanalyse untersucht stochastische Zusammenhänge zwischen gleichwertigen Zufallsvariablen anhand einer Stichprobe. Eine Maßzahl für die Stärke und Richtung eines linearen Zusammenhangs ist der Korrelationskoeffizient. Er ist gleich Null, wenn kein linearer Zusammenhang besteht. Für den Korrelationskoeffizienten ρ (Rho) der beiden Zufallsvariablen X und Y gilt:

- (1) $-1 \leq \rho \leq +1$. Ein Korrelationskoeffizient existiert stets dann, wenn die Varianzen der beiden Zufallsvariablen existieren und verschieden von Null sind [$\rho = \sigma_{xy} / (\sigma_x \sigma_y)$].
- (2) Für $\rho = \pm 1$ besteht zwischen X und Y ein funktionaler Zusammenhang; alle Punkte (x_i, y_i) , Realisierungen der Zufallsvariablen im Rahmen einer Stichprobe, liegen auf einer Geraden.
- (3) Ist $\rho = 0$, so heißen X und Y unkorreliert; zwei Zufallsvariablen sind um so stärker korreliert, je näher $|\rho|$ bei 1 liegt.
- (4) Für zweidimensional normalverteilte Zufallsvariablen folgt aus $\rho = 0$ die stochastische Unabhängigkeit von X und Y . Die zweidimensionale Normalverteilung (vgl. auch Abbildung 5.37 im Abschnitt [5.6.4]) ist ein glockenförmiges Gebilde, das durch den Parameter ρ (und 4 weitere Parameter: $\mu_x, \mu_y, \sigma_x, \sigma_y$) charakterisiert ist. Die Schnittfigur parallel zur xy -Ebene

ist für $\rho = 0$ und $\sigma_x = \sigma_y$ ein Kreis und für $\sigma_x \neq \sigma_y$ eine Ellipse, die für $|\varrho| \rightarrow 1$ immer schmaler wird. Der Parameter ϱ wird durch den Stichprobenkorrelationskoeffizienten r geschätzt; r ist auch für nicht normalverteilte Zufallsvariablen mit angenähert linearer Regression ein Maß für die Stärke und die Richtung des stochastischen Zusammenhangs.

7.7.1 Prüfung des Vorhandenseins einer Korrelation

Vorausgesetzt werden zwei abhängige Variable aus einer zweidimensionalen Normalverteilung. Dann lässt sich nach R.A. Fisher anhand der t -Verteilung ($FG = n - 2$) prüfen, ob r Unabhängigkeit ($H_{01}: \varrho = 0$; $H_{02}: \varrho \leq 0$; $H_{03}: \varrho \geq 0$) oder Korrelation ($H_{A1}: \varrho \neq 0$; $H_{A2}: \varrho > 0$; $H_{A3}: \varrho < 0$) anzeigt.

$$\hat{t} = r \cdot \sqrt{\frac{n-2}{1-r^2}} \quad \text{siehe Übersicht 22} \quad (7.273)$$

Für $\hat{t} \geq t_{n-2;\alpha}$ wird $H_0: \varrho = 0$ abgelehnt. Einfacher ist es, Tabelle 7.96 zu benutzen. Kritische Schranken für r für $\varrho = 0,00$ (0,10) 0,90 und 0,95, zwölf Signifikanzniveaus sowie 50 Stichprobenumfänge zwischen $n = 4$ und $n = 1000$ gibt Odeh [Ode82a].

Übersicht 22. Prüfung eines Korrelationskoeffizienten

Anhand der Prüfgröße $\hat{t} = r\sqrt{(n-2)/(1-r^2)}$ wird eine der folgenden Nullhypothesen abgelehnt sobald:

- (I) $H_0: \varrho = 0$ [gegen $H_A: \varrho \neq 0$]: $|\hat{t}| \geq t_{n-2;1-\alpha/2}$
- (II) $H_0: \varrho \leq 0$ [gegen $H_A: \varrho > 0$]: $\hat{t} \geq t_{n-2;1-\alpha}$
- (III) $H_0: \varrho \geq 0$ [gegen $H_A: \varrho < 0$]: $\hat{t} \leq t_{n-2;\alpha}$

Dann liegt auf dem $100\alpha\%$ -Niveau eine Korrelation (I), positive Korrelation (II), negative Korrelation (III) vor. Für Prüfungen auf dem 5%- , 1%- oder 0,1%-Niveau wird man Tabelle 7.96 bevorzugen.

Für (7.273) wird vorausgesetzt, dass $\varrho = \text{Null}$ ist.

Denn r ist angenähert normalverteilt, wenn:

- (1) X, Y zumindest angenähert normalverteilt sind,
- (2) $|r|$ genügend weit vom Wert 1 entfernt ist sowie
- (3) n hinreichend groß ist.

Kann nicht angenommen werden, dass $\varrho = 0$ ist, d.h. $H_0: \varrho = \varrho_0$, dann muss man die Transformation der r -Werte in $\hat{z}$ -Werte (7.282) nach R. A. Fisher (vgl. Abschnitt [7.7.2]) anwenden; man prüft dann anhand von:

$$\hat{t} = \frac{\hat{z} - \hat{z}_0}{\sigma_{\hat{z}}} = 1,1513 \left| \lg \frac{1+r}{1-r} - \lg \frac{1+\varrho}{1-\varrho} \right| \sqrt{n-3} \quad (7.275)$$

mit $\nu = n - 2$ Freiheitsgraden



Beispiel: $n = 14$; $r = 0,9660$; $H_0: \varrho = 0,8$; $H_A: \varrho > 0,8$; $\alpha = 0,05$

$$\hat{t} = 1,1513 \left| \lg \frac{1+0,966}{1-0,966} - \lg \frac{1+0,8}{1-0,8} \right| \sqrt{14-3}$$

$$\hat{t} = 1,1513(|1,7621 - 0,9542|)3,3166 = 3,0849$$

Mit $\hat{t} = 3,085 > 1,782 = t_{12;0,95}$ kann auf dem 5%-Niveau angenommen werden, daß zwischen X und Y ein wesentlich stärkerer Zusammenhang als $\varrho = 0,8$ besteht. Dieses Resultat erhält man auch nach (7.278):

$$\begin{aligned}\hat{t} &= [(0,966 - 0,8)\sqrt{14-2}]/\sqrt{(1-0,966^2)(1-0,8^2)} \\ \hat{t} &= 3,707 > 1,782 = t_{12;0,95}\end{aligned}$$

Hinweise: 1. Die Nullhypothese $H_0: \varrho = 0$ kann auch anhand der F -Verteilung abgelehnt werden:

$$\begin{array}{l} \hat{F} = \frac{r^2(n-2)}{1-r^2} \\ FG_1 = 1 \quad FG_2 = n-2 \end{array} \quad (7.276)$$

Beachte: (7.273) und (7.276) sind gleichwertig, [vgl. Legende der Tab. 7.96].

$$\begin{array}{l} \hat{F} = \frac{1+r}{1-r} \\ FG_1 = FG_2 = n-2 \end{array} \quad \text{Kymn [Kym68]} \quad (7.277)$$

2. Ein Vergleich mit einem vorgegebenen Wert ϱ ist auch nach Samiuddin [Sam70] möglich:

$$\begin{array}{l} \hat{t} = \frac{(r - \varrho)\sqrt{n-2}}{\sqrt{(1-r^2)(1-\varrho^2)}} \\ FG = n-2 \end{array} \quad (7.278)$$

3. Prüfung der Differenz zweier nichtunabhängiger Korrelationskoeffizienten:

$$\begin{array}{l} \hat{t} = (r_{12} - r_{13})\sqrt{(n-1)(1+r_{23})/A} \\ \text{mit } A = 2\frac{n-1}{n-3}|R| + \frac{(r_{12} + r_{13})^2}{2}(1-r_{23})^3 \\ \text{und der Determinanten } |R| \\ |R| = (1 - r_{12}^2 - r_{13}^2 - r_{23}^2) + (2r_{12}r_{13}r_{23}) \end{array} \quad (7.279)$$

Für $\hat{t} > t_{n-3;1-\alpha}$ wird $H_0: \varrho_{12} = \varrho_{13}$ nach Williams [Wil59] [vgl. auch Neill und Dunn [ND75], Bennett [Ben78], Steiger [Ste80] sowie Wilson und Martin [WM83]] auf dem $100\alpha\%$ -Niveau abgelehnt.

Beispiel: $n = 30: \quad r_{12} = 0,85 \quad r_{13} = 0,71 \quad r_{23} = 0,80 \quad \text{für } \alpha = 0,05$

$$|R| = (1 - 0,85^2 - 0,71^2 - 0,80^2) + (2 \cdot 0,85 \cdot 0,71 \cdot 0,80) = 0,099$$

$$A = 2\frac{29}{27}0,099 + \frac{(0,85 + 0,71)^2}{2}(1 - 0,80)^3 = 0,222$$

$$\hat{t} = (0,85 - 0,71)\sqrt{29(1 + 0,80)/0,222} = 2,15 > 2,052 = t_{27;0,95}$$



Tabelle 7.96. Prüfung des Korrelationskoeffizienten r auf Signifikanz gegen Null. Die Nullhypothese ($\varrho = 0$) wird zugunsten der Alternativhypothese (zweiseitige Fragestellung: $\varrho \neq 0$, einseitige Fragestellung: $\varrho > 0$ bzw. $\varrho < 0$) abgelehnt, wenn $|r|$ den für die geeignete Fragestellung, die gewählte Irrtumswahrscheinlichkeit und den vorliegenden Freiheitsgrad ($FG = n - 2$) tabellierten Wert erreicht oder überschreitet (dann sind auch die beiden Regressionskoeffizienten β_{yx} und β_{xy} von Null verschieden). Der einseitige Test darf nur durchgeführt werden, wenn vor der Erhebung der n Datenpaare das Vorzeichen des Korrelationskoeffizienten sicher ist.

Diese Tafel ersetzt Formel (7.273): z. B. ist ein auf 60FG ($n = 62$) basierender Wert $r = 0,25$ auf dem 5%-Niveau statistisch signifikant ($\varrho \neq 0$)

FG	Zweiseitiger Test			Einseitiger Test		
	5 % #	1 %	0,1 %	5 %	1 %	0,1 %
1	0,9969	A*	B*	0,9877	0,9995	C*
2	0,9500	0,9900	0,9990	0,9000	0,9800	0,9980
3	0,8783	0,9587	0,9911	0,805	0,934	0,986
4	0,811	0,917	0,974	0,729	0,882	0,963
5	0,754	0,875	0,951	0,669	0,833	0,935
6	0,707	0,834	0,925	0,621	0,789	0,905
7	0,666	0,798	0,898	0,582	0,750	0,875
8	0,632	0,765	0,872	0,549	0,715	0,847
9	0,602	0,735	0,847	0,521	0,685	0,820
10	0,576	0,708	0,823	0,497	0,658	0,795
11	0,553	0,684	0,801	0,476	0,634	0,772
12	0,532	0,661	0,780	0,457	0,612	0,750
13	0,514	0,641	0,760	0,441	0,592	0,730
14	0,497	0,623	0,742	0,426	0,574	0,711
15	0,482	0,606	0,725	0,412	0,558	0,694
16	0,468	0,590	0,708	0,400	0,543	0,678
17	0,456	0,575	0,693	0,389	0,529	0,662
18	0,444	0,561	0,679	0,378	0,516	0,648
19	0,433	0,549	0,665	0,369	0,503	0,635
20	0,423	0,537	0,652	0,360	0,492	0,622
21	0,413	0,526	0,640	0,352	0,482	0,610
22	0,404	0,515	0,629	0,344	0,472	0,599
23	0,396	0,505	0,618	0,337	0,462	0,588
24	0,388	0,496	0,607	0,330	0,453	0,578
25	0,381	0,487	0,597	0,323	0,445	0,568
26	0,374	0,478	0,588	0,317	0,437	0,559
27	0,367	0,470	0,579	0,311	0,430	0,550
28	0,361	0,463	0,570	0,306	0,423	0,541
29	0,355	0,456	0,562	0,301	0,416	0,533
30	0,349	0,449	0,554	0,296	0,409	0,526
35	0,325	0,418	0,519	0,275	0,381	0,492
40	0,304	0,393	0,490	0,257	0,358	0,463
50	0,273	0,354	0,443	0,231	0,322	0,419
60	0,250	0,325	0,408	0,211	0,295	0,385
70	0,232	0,302	0,380	0,195	0,274	0,358
80	0,217	0,283	0,357	0,183	0,257	0,336
90	0,205	0,267	0,338	0,173	0,242	0,318
100	0,195	0,254	0,321	0,164	0,230	0,302
120	0,178	0,232	0,294	0,150	0,210	0,277
150	0,159	0,208	0,263	0,134	0,189	0,249
200	0,138	0,181	0,230	0,116	0,164	0,216
250	0,124	0,162	0,206	0,104	0,146	0,194
300	0,113	0,148	0,188	0,095	0,134	0,177
350	0,105	0,137	0,175	0,0878	0,124	0,164
400	0,0978	0,128	0,164	0,0822	0,116	0,154
500	0,0875	0,115	0,146	0,0735	0,104	0,138
700	0,0740	0,0972	0,124	0,0621	0,0878	0,116
1000	0,0619	0,0813	0,104	0,0520	0,0735	0,0975
1500	0,0505	0,0664	0,0847	0,0424	0,0600	0,0795
2000	0,0438	0,0575	0,0734	0,0368	0,0519	0,0689
A* = 0,999877 B* = 0,99999877 C* = 0,9999951						

Bei Bedarf lässt sich die 5%-Schranke durch $2/\sqrt{\nu+2}$ mit $\nu = FG$ approximieren, z. B. $\nu = 14$, $2/\sqrt{14+2} = 0,5$.

Für 3 Variablen X_i, X_j, X_k gilt, dass die Korrelation zwischen X_i und X_j im Bereich:

$$r_{ik}r_{jk} \pm \sqrt{(1 - r_{ik}^2)(1 - r_{jk}^2)}$$

liegen muss; etwa für $r_{ik} = 0,6$ und $r_{jk} = 0,9$ gilt:

$$0,6 \cdot 0,9 \pm \sqrt{(1 - 0,36)(1 - 0,81)} = 0,54 \pm 0,35, \quad \text{d. h.} \quad 0,19 \leq r_{ij} \leq 0,89.$$

Für mehr als drei Variablen gilt Entsprechendes für jede Dreiergruppe.

Mehrere nichtunabhängige r_i werden nach Meng u. Mitarb. [MRR92] verglichen, wobei auch Kontraste geprüft werden können.

4. Benötigte Stichprobenumfänge zur Schätzung des Korrelationskoeffizienten lassen sich anhand von Tabelle 7.97 abschätzen.

Beispiel: Um einen Korrelationskoeffizienten von etwa $\rho = 0,6$ auf dem 5%-Niveau mit einer Power von 0,9 zu schätzen, benötigt man nach Tabelle 7.97 (Gatsonis und Sampson [GS89]) (mindestens) $n = 24$ Beobachtungspaare, für $\rho = 0,2$ wären es $n = 258$ Beobachtungspaare.



Tabelle 7.97. Benötigte Stichprobenumfänge zur Schätzung des Korrelationskoeffizienten bei vorgegebener Power auf dem 5%-Niveau. Einige Werte aus Gatsonis und Sampson (1989)

$ \rho \backslash \text{Power:}$	0,50	0,60	0,70	0,80	0,90
0,10	384	489	616	782	1046
0,15	171	217	273	346	462
0,20	96	122	153	193	258
0,25	62	78	97	123	164
0,30	43	54	67	84	112
0,35	31	39	49	61	81
0,40	24	30	37	46	61
0,45	19	23	29	36	47
0,50	15	19	23	29	37
0,55	13	15	19	23	30
0,60	11	13	15	19	24
0,65	9	11	13	16	20
0,70	8	9	11	13	17
0,75	7	8	9	11	14
0,80	6	7	8	9	11
0,85	5	6	7	8	9
0,90	5	5	6	6	8

5. Berechnung und Prüfung eines gemeinsamen Korrelationskoeffizienten (vgl. Abschnitt [7.7.4]). Liegen mindestens zwei Zufallsstichproben vor ($k \geq 2$) und ist ein gemeinsamer Korrelationskoeffizient r_{gem} zu berechnen und zu prüfen, so ist es nicht zulässig, r_{gem} aus allen $n = \sum_{i=1}^k n_i$ Beobachtungspaaren zu berechnen. Statt dessen berechne man jeweils r_i und

$$r_{\text{gem}} = \frac{\sum_{i=1}^k (n_i - 1) r_i}{\sum_{i=1}^k (n_i - 1)} \quad (7.280)$$

Tabelle 7.98. Umrechnung des Korrelationskoeffizienten $\hat{z} = \frac{1}{2} \ln \frac{1+r}{1-r}$ (auszugsweise entnommen aus Fisher, R.A. und F. Yates: Statistical Tables for Biological, Agricultural and Medical Research, published by Oliver and Boyd Ltd., Edinburgh, 1963, p. 63)

$\hat{z}$	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0100	0,0200	0,0300	0,0400	0,0500	0,0599	0,0699	0,0798	0,0898
0,1	0,0997	0,1096	0,1194	0,1293	0,1391	0,1489	0,1586	0,1684	0,1781	0,1877
0,2	0,1974	0,2070	0,2165	0,2260	0,2355	0,2449	0,2543	0,2636	0,2729	0,2821
0,3	0,2913	0,3004	0,3095	0,3185	0,3275	0,3364	0,3452	0,3540	0,3627	0,3714
0,4	0,3800	0,3885	0,3969	0,4053	0,4136	0,4219	0,4301	0,4382	0,4462	0,4542
0,5	0,4621	0,4699	0,4777	0,4854	0,4930	0,5005	0,5080	0,5154	0,5227	0,5299
0,6	0,5370	0,5441	0,5511	0,5580	0,5649	0,5717	0,5784	0,5850	0,5915	0,5980
0,7	0,6044	0,6107	0,6169	0,6231	0,6291	0,6351	0,6411	0,6469	0,6527	0,6584
0,8	0,6640	0,6696	0,6751	0,6805	0,6858	0,6911	0,6963	0,7014	0,7064	0,7114
0,9	0,7163	0,7211	0,7259	0,7306	0,7352	0,7398	0,7443	0,7487	0,7531	0,7574
1,0	0,7616	0,7658	0,7699	0,7739	0,7779	0,7818	0,7857	0,7895	0,7932	0,7969
1,1	0,8005	0,8041	0,8076	0,8110	0,8144	0,8178	0,8210	0,8243	0,8275	0,8306
1,2	0,8337	0,8367	0,8397	0,8426	0,8455	0,8483	0,8511	0,8538	0,8565	0,8591
1,3	0,8617	0,8643	0,8668	0,8692	0,8717	0,8741	0,8764	0,8787	0,8810	0,8832
1,4	0,8854	0,8875	0,8896	0,8917	0,8937	0,8957	0,8977	0,8996	0,9015	0,9033
1,5	0,9051	0,9069	0,9087	0,9104	0,9121	0,9138	0,9154	0,9170	0,9186	0,9201
1,6	0,9217	0,9232	0,9246	0,9261	0,9275	0,9289	0,9302	0,9316	0,9329	0,9341
1,7	0,9354	0,9366	0,9379	0,9391	0,9402	0,9414	0,9425	0,9436	0,9447	0,9458
1,8	0,94681	0,94783	0,94884	0,94983	0,95080	0,95175	0,95268	0,95359	0,95449	0,95537
1,9	0,95624	0,95709	0,95792	0,95873	0,95953	0,96032	0,96109	0,96185	0,96259	0,96331
2,0	0,96403	0,96473	0,96541	0,96609	0,96675	0,96739	0,96803	0,96865	0,96926	0,96986
2,1	0,97045	0,97103	0,97159	0,97215	0,97269	0,97323	0,97375	0,97426	0,97477	0,97526
2,2	0,97574	0,97622	0,97668	0,97714	0,97759	0,97803	0,97846	0,97888	0,97929	0,97970
2,3	0,98010	0,98049	0,98087	0,98124	0,98161	0,98197	0,98233	0,98267	0,98301	0,98335
2,4	0,98367	0,98399	0,98431	0,98462	0,98492	0,98522	0,98551	0,98579	0,98607	0,98635
2,5	0,98661	0,98688	0,98714	0,98739	0,98764	0,98788	0,98812	0,98835	0,98858	0,98881
2,6	0,98903	0,98924	0,98945	0,98966	0,98987	0,99007	0,99026	0,99045	0,99064	0,99083
2,7	0,99101	0,99118	0,99136	0,99153	0,99170	0,99186	0,99202	0,99218	0,99233	0,99248
2,8	0,99263	0,99278	0,99292	0,99306	0,99320	0,99333	0,99346	0,99359	0,99372	0,99384
2,9	0,99396	0,99408	0,99420	0,99431	0,99443	0,99454	0,99464	0,99475	0,99485	0,99495
3,3	0,99505	0,99595	0,99668	0,99728	0,99777	0,99818	0,99851	0,99878	0,99900	0,99918
4,4	0,99933	0,99945	0,99955	0,99963	0,99970	0,99975	0,99980	0,99983	0,99986	0,99989

und prüfe nach

$$\hat{t} = r_{\text{gem}} \cdot \sqrt{\frac{n-k-1}{1-r_{\text{gem}}^2}} \quad (7.281)$$

Beispielsweise erhält man für $n_1 = n_2 = n_3 = 30$ und $r_1 = 0,422$, $r_2 = 0,388$ und $r_3 = 0,569$; d. h. hier ($n_i = \text{konstant}$)

$$r_{\text{gem}} = \sum_{i=1}^3 r_i / 3$$

$r_{\text{gem}} = (0,422 + 0,388 + 0,569) / 3 = 0,460$ sowie $\hat{t} = 0,460 \sqrt{(90 - 3 - 1) / (1 - 0,460^2)} = 4,80$ für $FG = n - k - 1 = 90 - 3 - 1 = 86$ den deutlichen Hinweis, dass ρ positiv ist (vgl. $t_{86;0,99} = 2,37$).

7.7.2 $\hat{z}$ -Transformation nach R.A. Fisher

Wenn sich der Korrelationskoeffizient statistisch signifikant von Null unterscheidet, weicht seine Verteilung um so stärker von der zweidimensionalen Normalverteilung ab, je kleiner die Anzahl der Beobachtungspaare n und je größer sein Absolutwert ist. Durch die $\hat{z}$ -Transformation nach

R. A. Fisher wird die Verteilung des Korrelationskoeffizienten **approximativ normalisiert**. Diese Approximation ist umso besser, je kleiner $|\varrho|$ ist und je größer n ist.

Die Transformation („z-Punkt“-Transformation, vgl. Tabelle 7.98) basiert auf einer Reihenentwicklung

$$\dot{z} = r + \frac{1}{3}r^3 + \frac{1}{5}r^5 + \frac{1}{7}r^7 + \dots,$$

die nach den folgenden Formeln berechnet werden kann (vgl. auch (6.118) und (6.119) auf Seite 297):

$$\dot{z} = 0,5 \ln[(1+r)/(1-r)] = 1,1513 \lg[(1+r)/(1-r)] \text{ mit } s_{\dot{z}} = 1/\sqrt{n-3} \quad (7.282)$$

$$r = (e^{2\dot{z}} - 1)/(e^{2\dot{z}} + 1) = (10^{\dot{z}/1,1513} - 1)/(10^{\dot{z}/1,1513} + 1) \quad (7.283)$$

[für $r < 0,3$ erhalten wir $\dot{z} \approx r$]

Die Umrechnung von r in $\dot{z}$ und umgekehrt kann auch mit Hilfe der Tabellen 7.98 und 7.99 erfolgen: In der ersten Spalte der Tabelle 7.98 stehen die $\dot{z}$ -Werte mit der ersten Dezimalstelle, während die zweite Dezimalstelle in der obersten Zeile zu finden ist.

Die Umwandlung kleiner Werte r ($0 < r < 0,20$) in $\dot{z} = \tanh^{-1} r$ erfolgt ausreichend genau nach $\dot{z} = r + (r^3/3)$ (z. B. $\dot{z} = 0,100$ für $r = 0,10$); $\dot{z}$ -Werte für r gleich $0,00(0,01)0,99$ sind der Tabelle 7.99 zu entnehmen (für $r = 1$ wird $\dot{z} = \infty$, für $r = 0,999$ wird $\dot{z} = 3,80020$). Die wissenschaftlichen Tabellen der Ciba-Geigy AG [CG80] enthalten auf den Seiten 64, 66 und 67 $\dot{z}$ -Werte für $r = 0,000(0,001)0,999$ und r -Werte für $\dot{z} = 0,000(0,001)1,999$ und für $\dot{z} = 0,0(0,1)9,9$.

Tabelle 7.99 Umrechnung der nach Fisher transformierten Werte für den Korrelationskoeffizienten $r = (e^{2\dot{z}} - 1)/(e^{2\dot{z}} + 1)$

r	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,00000	0,01000	0,02000	0,03001	0,04002	0,05004	0,06007	0,07011	0,08017	0,09024
0,1	0,10034	0,11045	0,12058	0,13074	0,14093	0,15114	0,16139	0,17167	0,18198	0,19234
0,2	0,20273	0,21317	0,22366	0,23419	0,24477	0,25541	0,26611	0,27686	0,28768	0,29857
0,3	0,30952	0,32055	0,33165	0,34283	0,35409	0,36544	0,37689	0,38842	0,40060	0,41180
0,4	0,42365	0,43561	0,44769	0,45990	0,47223	0,48470	0,49731	0,51007	0,52298	0,53606
0,5	0,54931	0,56273	0,57634	0,59015	0,60416	0,61838	0,63283	0,64752	0,66246	0,67767
0,6	0,69315	0,70892	0,72501	0,74142	0,75817	0,77530	0,79281	0,81074	0,82911	0,84796
0,7	0,86730	0,88718	0,90764	0,92873	0,95048	0,97296	0,99622	1,02033	1,04537	1,07143
0,8	1,09861	1,12703	1,15682	1,18814	1,22117	1,25615	1,29334	1,33308	1,37577	1,42193
0,9	1,47222	1,52752	1,58903	1,65839	1,73805	1,83178	1,94591	2,09230	2,29756	2,64665

7.7.3 Weitere Anwendungen der $\dot{z}$ -Transformation

1. Die **Prüfung der Differenz** zwischen einem geschätzten Korrelationskoeffizienten r_1 und einem hypothetischen oder theoretischen Wert, dem Parameter ϱ , erfolgt auch [vgl. (7.275) und (7.278)] anhand der Standardnormalvariablen z nach (7.284).

$$\hat{z} = |\dot{z}_1 - \dot{z}| \sqrt{n_1 - 3} \quad (7.284)$$

Ist das Prüfprodukt kleiner als das entsprechende Quantil der Standardnormalverteilung $z_{1-\alpha}$ (einseitig) bzw. $z_{1-\alpha/2}$ (zweiseitig), so kann angenommen werden, dass $\varrho_1 = \varrho$ ist.

2. Der **Vergleich zweier geschätzter Korrelationskoeffizienten** r_1 und r_2 erfolgt nach

$$\hat{z} = \frac{|\hat{z}_1 - \hat{z}_2|}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}} \quad (7.285)$$

Ist der Prüfquotient kleiner als die Signifikanzschranke, so kann angenommen werden, dass die zugrundeliegenden Parameter gleich sind ($\varrho_1 = \varrho_2$). Die Schätzung des gemeinsamen Korrelationskoeffizienten $\bar{r}$ erfolgt dann über $\hat{\hat{z}}$:

$$\hat{\hat{z}} = \frac{\hat{z}_1(n_1 - 3) + \hat{z}_2(n_2 - 3)}{n_1 + n_2 - 6} \quad (7.286)$$

mit

$$s_z^\Delta = \frac{1}{\sqrt{n_1 + n_2 - 6}} \quad (7.287)$$

Das 95%-Konfidenzintervall für ϱ ergibt sich aus:

$$\hat{\hat{z}} \pm 1,960 s_z^\Delta$$



Beispiel 1: Gegeben $r_1 = 0,3$; $n_1 = 40$; $\varrho = 0,4$. Kann angenommen werden, dass $\varrho_1 = \varrho$ ($\alpha = 0,05$ mit $H_A: \varrho_1 \neq \varrho$)? Nach (7.284) gilt (Tab. 7.99):

$$\hat{z} = (|0,30952 - 0,423651|)\sqrt{40 - 3} = 0,694 < 1,96.$$

Mit $\hat{z} = 0,694 < 1,96 = z_{0,975}$ lässt sich $H_0: \varrho_1 = \varrho$ anhand der vorliegenden Stichprobe auf dem 5%-Niveau nicht ablehnen.



Beispiel 2: Gegeben $r_1 = 0,97$; $n_1 = 14$; $\varrho = 0,8$. Kann angenommen werden, dass $\varrho_1 \leq \varrho$ ($\alpha = 0,05$ mit $H_A: \varrho_1 > \varrho = 0,8$)? Nach (7.284):

$$\hat{z} = |2,09230 - 1,09861|\sqrt{14 - 3} = 3,296$$

mit $\hat{z} = 3,296 > 1,645 = z_{0,95}$ kann auf dem 5%-Niveau angenommen werden, dass zwischen X und Y ein wesentlich stärkerer Zusammenhang als $\varrho = 0,8$ besteht. Für kleines n ist (7.275) oder (7.278) zu bevorzugen.



Beispiel 3: Gegeben $r_1 = 0,6$; $n_1 = 28$ und $r_2 = 0,8$; $n_2 = 23$. Kann angenommen werden, dass $\varrho_1 = \varrho_2$ ($\alpha = 0,05$ mit $H_A: \varrho_1 \neq \varrho_2$)? Nach (7.285) gilt:

$$\hat{z} = \frac{|0,6932 - 1,0986|}{\sqrt{\frac{1}{28 - 3} + \frac{1}{23 - 3}}} = 1,35 < 1,96.$$

Da $\hat{z} = 1,35 < 1,96$ ist, kann die Nullhypothese $\varrho_1 = \varrho_2$ auf dem 5%-Niveau nicht abgelehnt werden. Das 95%-Konfidenzintervall für ϱ ist dann nach (7.286) und (7.287):

$$\hat{z} = \frac{17,330 + 21,972}{28 + 23 - 6} = 0,8734$$

$$s_{\hat{z}} = \frac{1}{\sqrt{28 + 23 - 6}} = 0,1491$$

$$\begin{aligned} &= 0,8734 \pm 1,96 \cdot 0,1491 \\ \hat{z} \pm 1,960 s_{\hat{z}} &= 0,8734 \pm 0,2922 \end{aligned}$$

$$0,5812 \leq \hat{z} \leq 1,1656$$

$$95\text{-KI: } 0,5235 \leq \varrho \leq 0,8223 \quad \text{oder} \quad 0,52 \leq \varrho \leq 0,82$$

7.7.4 Der Vergleich mehrerer Korrelationskoeffizienten

Gegeben seien k Schätzungen $r_1, r_2, \dots, r_i, \dots, r_k$ mit den Stichprobenumfängen $n_1, n_2, \dots, n_i, \dots, n_k$. Die Prüfung auf Homogenität der Korrelationskoeffizienten (Nullhypothese: $\varrho_1 = \varrho_2 = \dots = \varrho_i = \dots = \varrho_k = \varrho$, wobei ϱ ein rein hypothetischer Wert ist) erfolgt nach

$$\hat{\chi}^2 = \sum_{i=1}^k (n_i - 3)(\dot{z}_i - \hat{z})^2 \quad (7.288)$$

mit k Freiheitsgraden. Ist die Prüfgröße gleich dem Tabellenwert $\chi_{k;1-\alpha}^2$ oder kleiner – wenn beispielsweise $k = 4$ Korrelationskoeffizienten verglichen werden, dann ist mit einer Irrtumswahrscheinlichkeit $\alpha = 0,05$ die Schranke durch den Wert $\chi_{0,95}^2$ für $FG = k = 4$ gleich 9,49 gegeben –, dann weisen die Korrelationskoeffizienten nur zufällige Abweichungen vom theoretischen Wert ϱ auf, die Nullhypothese kann nicht abgelehnt werden.

Ist der hypothetische Wert nicht bekannt, dann wird er nach

$$\hat{\hat{z}} = \frac{\sum_{i=1}^k \dot{z}_i (n_i - 3)}{\sum_{i=1}^k (n_i - 3)} \quad (7.289)$$

geschätzt; die zugehörige Standardabweichung ist

$$s_{\hat{\hat{z}}} = \frac{1}{\sqrt{\sum_{i=1}^k (n_i - 3)}} \quad (7.290)$$

Die Prüfung der Nullhypothese $\varrho_1 = \varrho_2 = \dots = \varrho_k = \varrho$ [ϱ unbekannt] erfolgt dann nach

$$\hat{\chi}^2 = \sum_{i=1}^k (n_i - 3)(\dot{z}_i - \hat{\hat{z}})^2 \quad (7.291)$$

mit $FG = k - 1$. Ist $\hat{\chi}^2$ kleiner oder gleich dem Tabellenwert $\chi_{k-1;1-\alpha}^2$, so darf die Nullhypothese beibehalten und ein **durchschnittlicher Korrelationskoeffizient** $\bar{r}$ geschätzt werden. Die Vertrauensgrenzen für den gemeinsamen Korrelationskoeffizienten, für den Parameter ϱ , erhält man in bekannter Weise über den entsprechenden $\hat{z}$ -Wert und seine Standardabweichung $s_{\hat{z}}$

$$95\text{-KI: } \hat{z} \pm 1,960 s_{\hat{z}} \quad \text{bzw.} \quad 99\text{-KI: } \hat{z} \pm 2,576 s_{\hat{z}}$$

indem man die oberen und unteren Grenzen in die entsprechenden r -Werte transformiert.

Beispiel: in der folgenden Tabelle werden die Schätzungen r_1, r_2 und r_3 verglichen.

r_i	$\hat{z}_i$	n_i	$n_i - 3$	$\hat{z}_i (n_i - 3)$	$\hat{z}_i - \hat{z}$	$(\hat{z}_i - \hat{z})^2$	$(n_i - 3)(\hat{z}_i - \hat{z})^2$
0,60	0,6932	28	25	17,330	0,1777	0,03158	0,7895
0,70	0,8673	33	30	26,019	0,0036	0,00001	0,0003
0,80	1,0986	23	20	21,972	0,2277	0,05185	1,0369
$\sum (n_i - 3) = 75$				65,321	$\hat{\chi}^2 = 1,8268$		

Da $\hat{\chi}^2 = 1,83$ wesentlich kleiner ist als $\chi_{2;0,95}^2 = 5,99$, darf ein mittlerer Korrelationskoeffizient geschätzt werden

$$\hat{z} = 65,321/75 = 0,8709; \quad \bar{r} = 0,702$$

$$s_{\hat{z}} = 1/\sqrt{75} = 0,115; \quad \hat{z} \pm 1,96 \cdot 0,115 = \hat{z} \pm 0,2254 \quad \text{bzw.} \quad 0,6455 \text{ bis } 1,0963;$$

$$95\text{-KI für } \varrho: \quad 0,5686 \leq \varrho \leq 0,7992 \quad \text{oder} \quad 0,57 \leq \varrho \leq 0,80$$

Mit den durchschnittlichen Korrelationskoeffizienten lassen sich dann wieder Vergleiche zwischen zwei Schätzwerten $\bar{r}_1$ und $\bar{r}_2$ bzw. Vergleiche zwischen einem Schätzwert $\bar{r}_1$ und einem hypothetischen Korrelationskoeffizienten ϱ durchführen.

7.7.5 Prüfung der Linearität einer Regression

Die Prüfung der Nullhypothese, es liegt eine lineare Regression vor, ist möglich, wenn die Gesamtzahl n der y -Werte größer ist als die Anzahl k der x -Werte: Zu jedem Wert x_i der k x -Werte liegen also n_i y -Werte vor. [Wenn der Punkteschwarm die Linearität oder Nichtlinearität deutlich zum Ausdruck bringt, kann man auf den Linearitätstest verzichten.]

Beim Vorliegen einer linearen Regression müssen die Gruppenmittelwerte $\bar{y}_i$ angenähert auf einer Geraden liegen, d. h. ihre Abweichung von der Regressionsgeraden darf nicht zu groß sein im Verhältnis zur Abweichung der Werte einer Gruppe von ihrem zugehörigen Mittelwert. Erreicht oder übersteigt somit das Verhältnis

$$\frac{\text{Abweichung der Mittelwerte von der Regressionsgeraden}}{\text{Abweichung der } y\text{-Werte von ihrem Gruppenmittelwert}}$$

d.h. die Prüfgröße

$$\hat{F} = \frac{\frac{1}{k-2} \sum_{i=1}^k n_i (\bar{y}_i - \hat{y}_i)^2}{\frac{1}{n-k} \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2} \quad \begin{matrix} \nu_1 = k - 2 \\ \nu_2 = n - k \end{matrix} \quad (7.292)$$

mit $(k-2, n-k)$ Freiheitsgraden den Tabellenwert $F_{k-2; n-k; \alpha}$, so muss die Linearitätshypothese verworfen werden. Die Summen in (7.292) sind die beiden Komponenten der Gesamtstreuung der Werte y_{ij} um die Regressionsgerade.

$$\sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \hat{y}_i)^2 = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 + \sum_{i=1}^k n_i (\bar{y}_i - \hat{y}_i)^2$$

Beispiel: Die Linearität der Beobachtungen aus folgender Tabelle mit $n=8$ Beobachtungen in $k=4$ Gruppen auf dem 5%-Niveau soll überprüft werden.



x_i		1	5	9	13
y_{ij}	$j=1$	1	2	4	5
	$j=2$	2	3		6
	$j=3$		3		
n_i		2	3	1	2

Das folgende kleine R-Programm zeigt die Berechnung der Prüfgröße (unter ausführlicher Verwendung der Vektor- bzw. Matrixindizes) in einzelnen Schritten. Zunächst werden die Mittelwerte $\bar{y}_i$ bestimmt. Anschließend schätzen wir die Regressionsgerade und berechnen für die vier x_i -Werte die entsprechenden Schätzungen $\hat{y}_i$. Die Abweichungen von der Regression (Zählerterm in (7.292)) und die Abweichungen der einzelnen Beobachtungen von den Gruppenmittelwerten (Nenner) können dann einfach bestimmt werden.

```
> xi <- c(1, 5, 9, 13); k <- length(xi)
> ni <- c(2, 3, 1, 2); n <- sum(ni)
> yij <- matrix(c(1,2,NA, 2,3,3, 4,NA,NA, 5,6,NA), ncol=k, byrow=FALSE)
>
> yisum <- rep(0, k) # Gruppenmittelwerte
> for (j in 1:k) {for (i in 1:ni[j]) yisum[j] <- yisum[j] + yij[i,j]}
> yibar <- yisum / ni
>
> # lineare Regression (x, y)
> x <- NULL; for (j in 1:k) x <- c(x, rep(xi[j], ni[j]))
> y <- NULL; for (j in 1:k) {for (i in 1:ni[j]) y <- c(y, yij[i,j])}
> linreg <- lm(y~x); a <- linreg$coeff[1]; b <- linreg$coeff[2]
> yihat <- a + b*xi # Schätzung aus lin. Regression
>
> ZF <- (1/(k-2))*sum(ni*(yibar-yihat)^2) # Abweichung von der Regression
> sn <- 0 # Abweichung vom Gruppenmittelwert
> for (j in 1:k) {for (i in 1:ni[j]) sn <- sn + (yij[i,j] - yibar[j])^2}
> NF <- (1/(n-k))*sn
>
> F <- ZF/NF; F # Teststatistik F
[1] 0.06582278
```

Als Prüfgröße ergibt sich dann $\hat{F} = 0,066$. Da $\hat{F} < 6,94 = F_{2;4;0,95}$ ist, kann die Linearitätshypothese beibehalten werden.

7.7.6 Prüfung der Regressionsparameter

Prüfung des Regressionskoeffizienten gegen Null

Spricht nichts gegen die Linearität einer Regression, so prüft man $H_0: \beta_{yx} = 0$ gegen $H_A: \beta_{yx} \neq 0$ bzw. gegen eine der beiden einseitigen Alternativhypothesen anhand des entsprechenden t -Tests (7.293) mit $FG = n - 2$.



Beispiel: Gegeben $b_{yx} = 0,426$; $s_{b_{yx}} = 0,190$; $n = 80$, $\alpha = 0,05$ bei zweiseitiger Fragestellung:
 $\hat{t} = \frac{0,426}{0,109} = 2,24 > 1,99 = t_{78;0,975}$. $H_0: \beta_{yx} = 0$ wird auf dem 5%-Niveau verworfen, d. h. der zugrunde liegende Parameter β_{yx} unterscheidet sich statistisch signifikant von Null.

Beachte: $s_{b_{yx}} = \frac{s_{y \cdot x}}{s_x} \sqrt{n-1}$ mit $s_{y \cdot x}^2 = \sum (x_i - \bar{x})(y_i - \bar{y}) / n - 2$.

Ist r berechnet worden, so gilt dann, wenn $\varrho = 0$, auch β_{yx} (und β_{xy}) = 0.

H_0	H_A	H_0 wird abgelehnt für
$\beta_{yx} = 0$	$\beta_{yx} \neq 0$	$\hat{t} = b_{yx} / s_{b_{yx}} \geq t_{n-2; 1-\alpha/2}$
$\beta_{yx} \leq 0$	$\beta_{yx} > 0$	$\hat{t} = b_{yx} / s_{b_{yx}} \geq t_{n-2; 1-\alpha}$
$\beta_{yx} \geq 0$	$\beta_{yx} < 0$	$\hat{t} = b_{yx} / s_{b_{yx}} \leq t_{n-2; \alpha}$

(7.293)

Prüfung der Differenz zwischen einem geschätzten und einem hypothetischen Regressionskoeffizienten.

Für die Prüfung, ob ein geschätzter Regressionskoeffizient b_{yx} mit einem theoretischen Parameterwert β_{yx} verträglich ist, benutzt man – entsprechend der vorliegenden Fragestellung – den t -Test (7.294) mit $FG = n - 2$.

H_0	H_A	H_0 wird abgelehnt für
$\beta_{0;yx} = 0$	$\beta_{0;yx} \neq 0$	$\hat{t} = b_{yx} - \beta_{yx} / s_{b_{yx}} \geq t_{n-2; 1-\alpha/2}$
$\beta_{0;yx} \leq 0$	$\beta_{0;yx} > 0$	$\hat{t} = (b_{yx} - \beta_{yx}) / s_{b_{yx}} \geq t_{n-2; 1-\alpha}$
$\beta_{0;yx} \geq 0$	$\beta_{0;yx} < 0$	$\hat{t} = (b_{yx} - \beta_{yx}) / s_{b_{yx}} \leq t_{n-2; \alpha}$

(7.294)

Verträglichkeit heißt hier und weiter unten, dass der unter H_0 zum Schätzwert (z. B. b_{yx}) gehörige Parameter (d. h. hier $\beta_{0;yx}$) mit dem theoretischen Parameter (d. h. hier β_{yx}) identisch ist; d. h. z. B. $H_0: \beta_{0;yx} = \beta_{yx}$ [sowie $H_A: \beta_{0;yx} \neq \beta_{yx}$ (Nichtverträglichkeit)]. Bemerkt sei, dass für die zweiseitige Fragestellung (7.294) als (7.295) geschrieben werden kann.

$$\hat{t} = \frac{|b_{yx} - \beta_{yx}|}{s_{y \cdot x} / s_x} \cdot \sqrt{n-1} = \frac{|b_{yx} - \beta_{yx}|}{\sqrt{1-r^2}} \cdot \frac{s_x}{s_y} \cdot \sqrt{n-2} = \frac{|b_{yx} - \beta_{yx}|}{s_{b_{yx}}} \quad \text{mit } FG = n-2 \quad (7.295)$$



Beispiel:

Gegeben: $b_{yx} = 0,426$; $\beta_{yx} = 0,500$; $s_{b_{yx}} = 0,190$; $n = 80$; $\alpha = 0,05$ bei zweiseitiger

$$\text{Fragestellung: } \hat{t} = \frac{|0,426 - 0,500|}{0,190} = 0,39 < 1,99 = t_{78;0,975}$$

Die Nullhypothese wird auf dem 5%-Niveau nicht abgelehnt.

Prüfung der Differenz zwischen einem geschätzten und einem hypothetischen Achsenabschnitt.

Für die Prüfung der Nullhypothese: a_{yx} ist mit α_{yx} verträglich, benutzt man (7.296).

$\hat{t} = \frac{ a_{yx} - \alpha_{yx} }{s_{a_{yx}}} \quad \text{mit } FG = n - 2 \text{ Freiheitsgraden}$	(7.296)

Beachte: $s_{a_{yx}} = s_{y \cdot x} \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{s_x^2} (n-1)}$ mit $s_{y \cdot x}^2 = \sum (x_i - \bar{x})(y_i - \bar{y}) / (n-2)$.

Beispiel:

Gegeben: $a_{yx} = 7,729$; $\alpha_{yx} = 15,292$; $s_{a_{yx}} = 2,862$; $n = 80$; $\alpha = 0,05$ bei zweiseitiger

$$\text{Fragestellung: } \hat{t} = \frac{|7,729 - 15,292|}{2,862} = 2,64 > 1,99 = t_{78;0,975}$$

Beide Achsenabschnitte und damit beide Regressionsgeraden unterscheiden sich auf dem 5%-Niveau statistisch signifikant.

Vergleich zweier Regressionskoeffizienten

Zwei Regressionskoeffizienten b_1 und b_2 lassen sich nach (7.297) vergleichen.

$$\hat{t} = \frac{|b_1 - b_2|}{\sqrt{\frac{s_{y_1 \cdot x_1}^2 (n_1 - 2) + s_{y_2 \cdot x_2}^2 (n_2 - 2)}{n_1 + n_2 - 4} \left[\frac{1}{Q_{x_1}} + \frac{1}{Q_{x_2}} \right]}} = \frac{\text{Differenz}}{\text{Standardfehler der Differenz}} \quad (7.297)$$

$FG = n_1 + n_2 - 4$

Unter der Quadratwurzel ist das Produkt aus der gemeinsamen empirischen Restvarianz und der Summe beider reziprok genommener Abweichungssummen.

Nullhypothese: $\beta_1 = \beta_2$ (Parallelität). Vorausgesetzt werden unabhängige Stichproben (n_1, n_2) aus Grundgesamtheiten mit gleicher Restvarianz ($\sigma_{y_1 \cdot x_1}^2 = \sigma_{y_2 \cdot x_2}^2$) und gleichen Varianzen: $\sigma_{x_1}^2 = \sigma_{x_2}^2$.

Hinweis zum Standardfehler der Differenz:

Ist der Variationsbereich für x klein, so heisst dies:
groß
kaum
 b ist schätzbar,
gut
d. h. der Standardfehler der Differenz ist groß,
klein
d. h. b_1 und b_2 sind kaum zu unterscheiden.
leicht

Beispiel:

$$n_1 = 40; s_{y_1 \cdot x_1}^2 = 0,14; Q_{x_1} = 163; b_1 = 0,40$$

$$n_2 = 50; s_{y_2 \cdot x_2}^2 = 0,16; Q_{x_2} = 104; b_2 = 0,31$$

$$\text{Nullhypothese: a) } \beta_1 \leq \beta_2; \quad \text{b) } \beta_1 = \beta_2$$

a) Einseitige Fragestellung ($\alpha = 0,05$): Alternativhypothese: $\beta_1 > \beta_2$

b) Zweiseitige Fragestellung ($\alpha = 0,05$): Alternativhypothese: $\beta_1 \neq \beta_2$

$$\hat{t} = \frac{|0,40 - 0,31|}{\sqrt{\frac{0,14(40-2) + 0,16(50-2)}{40+50-4} \left(\frac{1}{163} + \frac{1}{104} \right)}} = 1,85$$

Zu a: Da $\hat{t} = 1,85 > 1,66 = t_{86;0,95}$ ist, wird H_0 auf dem 5%-Niveau abgelehnt.

Zu b: Da $\hat{t} = 1,85 < 1,99 = t_{86;0,975}$ ist, wird H_0 auf dem 5% Niveau nicht abgelehnt.

Für den Fall ungleicher Restvarianzen (zur Benennung: die größere dient als Zähler), d. h. wenn

$$\frac{s_{y_1 \cdot x_1}^2}{s_{y_2 \cdot x_2}^2} > F_{(n_1-2; n_2-2; 0,10)} \quad (7.298)$$

ist, lässt sich der Vergleich approximativ nach

$$\hat{z} = \frac{|b_1 - b_2|}{\sqrt{\frac{s_{y_1 \cdot x_1}}{Q_{x_1}} + \frac{s_{y_2 \cdot x_2}}{Q_{x_2}}}} \quad (7.299)$$

durchführen, sobald beide Stichprobenumfänge > 20 sind. Ist ein Stichprobenumfang kleiner, dann kann die Verteilung der Prüfgröße durch die t -Verteilung mit ν Freiheitsgraden approximiert werden, wobei ν

$$\nu = \frac{1}{\frac{c^2}{n_1-2} + \frac{(1-c)^2}{n_2-2}} \quad \text{mit} \quad c = \frac{\frac{s_{y_1 \cdot x_1}^2}{Q_{x_1}}}{\frac{s_{y_1 \cdot x_1}^2}{Q_{x_1}} + \frac{s_{y_2 \cdot x_2}^2}{Q_{x_2}}} \quad n_1 \leq n_2 \quad (7.300)$$

stets zwischen dem kleineren Wert von $(n_1 - 2)$ und $(n_2 - 2)$ sowie $(n_1 + n_2 - 4)$ liegt (vgl. auch Potthoff 1965 [Pot65]).

Bei Bedarf lässt sich ein **Konfidenzintervall** für $\beta_1 - \beta_2$ angeben:

$$b_1 - b_2 \pm t\sqrt{\quad} \quad \text{mit} \quad \sqrt{\quad} \text{ aus (5.104) oder (5.106)} \quad (7.301)$$

Wird $H_0: \beta_1 = \beta_2$ abgelehnt, so kann es interessant sein, den **Schnittpunkt beider Geraden** zu bestimmen:

$$\begin{aligned} x_1 &= (a_2 - a_1)/(b_1 - b_2) \\ y_1 &= a_1 + b_1 x_1 = a_2 + b_2 x_1 \end{aligned} \quad (7.302)$$

Wird H_0 nicht abgelehnt, d. h. kann man für beide einen **gemeinsamen Regressionskoeffizienten** β_{yx} annehmen, so lässt er sich durch

$$b_{yx} = (Q_{x_1 y_1} + Q_{x_2 y_2}) / (Q_{x_1} + Q_{x_2}) \quad (7.303)$$

schätzen, seine Standardabweichung, sein Standardfehler durch

$$s_{b_{yx}} = \sqrt{\frac{s_{y_1 \cdot x_1}^2(n_1 - 2) + s_{y_2 \cdot x_2}^2(n_2 - 2)}{n_1 + n_2 - 4}} \bigg/ (Q_{x_1} + Q_{x_2}) \quad (7.304)$$

Der Vergleich zweier Achsenabschnitte

Für den Vergleich zweier Achsenabschnitte a_1 und a_2 ($H_0: \alpha_1 = \alpha_2$; $H_A: \alpha_1 \neq \alpha_2$) dient

$$\hat{t} = \frac{|a_1 - a_2|}{\sqrt{\frac{s_{y_1 \cdot x_1}^2(n_1 - 2) + s_{y_2 \cdot x_2}^2(n_2 - 2)}{n_1 + n_2 - 4} \left[\frac{\sum x_1^2}{n_1 Q_{x_1}} + \frac{\sum x_2^2}{n_2 Q_{x_2}} \right]}} \quad (7.305)$$

Liegen für jeden Wert x_i n_i y -Werte vor, so sind im eckig geklammerten zweiten Term des Nenners beide Summen der Quadrate, d. h. $\sum x_1^2$ und $\sum x_2^2$ durch $\sum n_{i1} x_{i1}^2$ und $\sum n_{i2} x_{i2}^2$ zu ersetzen ($n_1 = \sum n_{i1}$; $n_2 = \sum n_{i2}$).

Einen Test auf Äquivalenz („Homogenität“) zweier Regressionsgeraden stellt Eva Bofinger vor [Australian and New Zealand Journal of Statistics **41** (1999), 481–491].

Mehrere Regressionsgeraden lassen sich z. B. nach E: 442 (vgl. S. 2) vergleichen. Näheres hierzu ist z. B. Bosch [Bos98], Fleiss [Fle86], Hewett und Lababidi [HL82], Rogosa [Rog80] sowie Wilcox [Wil87] zu entnehmen [vgl. auch die Übersicht in Psychological Methods **1** (1996), 261–277 sowie Biometrical Journal **44** (2002), 801–812].

7.7.7 Prüfung des Rang-Korrelationskoeffizienten ρ_S

Sind Zusammenhänge zwischen nicht normalverteilten Reihen zu ermitteln, entstammt also die zweidimensionale Stichprobe (x_i, y_i) einer beliebigen stetigen Verteilung, dann lässt sich die Abhängigkeit von Y und X durch den Spearmanschen Rang-Korrelationskoeffizienten r_S beurteilen:

$$r_S = 1 - \frac{6 \sum D^2}{n(n^2 - 1)} \quad (7.306)$$

Die Signifikanz von r_S wird für $n \leq 30$ Wertepaare der Tabelle 7.100 (Zar [Zar99]) entnommen. Anhand dieser Tabelle wird H_0 für den zweiseitigen und den einseitigen Test auf dem $100\alpha\%$ -Niveau abgelehnt, sobald ein beobachteter absoluter r_S -Wert den Tabellenwert r_S^* erreicht oder übersteigt:

$$|r_S| \geq \text{Schranke} = r_S^* \quad \text{für} \quad \begin{bmatrix} 1) \text{ Seitigkeit} \\ 2) \alpha \\ 3) n: 6 \text{ bis } 30 \end{bmatrix} \quad (7.307)$$

Für $n > 30$ wird r_S anhand der Approximation

$$\hat{J}_S = \frac{r_S}{2} \left[\sqrt{(n-1)} + \sqrt{(n-2)/(1-r_S^2)} \right] \quad (7.308)$$

geprüft.

Tabelle 7.100 Einige besonders wichtige Schranken für den Rangkorrelationskoeffizienten r_S nach Spearman aus Zar (1999)

n	Test:					n
	$\alpha: 0,001$	0,01	0,05	0,10	0,05	0,01
6	–	1,000	0,886	0,829	0,943	6
7	1,000	0,929	0,786	0,714	0,893	7
8	0,976	0,881	0,738	0,643	0,833	8
9	0,933	0,833	0,700	0,600	0,783	9
10	0,903	0,794	0,648	0,564	0,745	10
11	0,873	0,755	0,618	0,536	0,709	11
12	0,846	0,727	0,587	0,503	0,678	12
13	0,824	0,703	0,560	0,484	0,648	13
14	0,802	0,679	0,538	0,464	0,626	14
15	0,779	0,654	0,521	0,446	0,604	15
16	0,762	0,635	0,503	0,429	0,582	16
17	0,748	0,615	0,485	0,414	0,566	17
18	0,728	0,600	0,472	0,401	0,550	18
19	0,712	0,584	0,460	0,391	0,535	19
20	0,696	0,570	0,447	0,380	0,520	20
21	0,681	0,556	0,435	0,370	0,508	21
22	0,667	0,544	0,425	0,361	0,496	22
23	0,654	0,532	0,415	0,353	0,486	23
24	0,642	0,521	0,406	0,344	0,476	24
25	0,630	0,511	0,398	0,337	0,466	25
26	0,619	0,501	0,390	0,331	0,457	26
27	0,608	0,491	0,382	0,324	0,448	27
28	0,598	0,483	0,375	0,317	0,440	28
29	0,589	0,475	0,368	0,312	0,433	29
30	0,580	0,467	0,362	0,306	0,425	30

Dieser Wert wird mit $J_{S;\alpha} = [z_\alpha + t_{n-2;\alpha}]/2$ verglichen.

Beispiel:

$$n = 30 \text{ und } r_S = 0,3061$$

$$H_0: \varrho_S = 0 \text{ gegen } H_A: \varrho_S \neq 0; \alpha = 0,05$$

$$\hat{J}_S = \frac{0,3061}{2} \left[\sqrt{30-1} + \sqrt{(30-2)/(1-0,3061^2)} \right]$$

$$\hat{J}_S = 1,675 < 2,004 = (1,960 + 2,048)/2 = J_{S;0,05}$$

H_0 kann auf dem 5%-Niveau nicht abgelehnt werden.

Auch nach Tabelle 7.100 gilt dieses Resultat: $r_S = 0,3061 < 0,362$. Mit $r_S = 0,3061 = 0,306$ erhalte man für den einseitigen Test, $H_0: \varrho_S = 0$ gegen $H_A: \varrho_S > 0$, gerade noch ein auf dem 5%-Niveau statistisch signifikantes Resultat.



Zwei Bemerkungen zu ϱ_S und ϱ

- (1) Im Vergleich zu r schätzt r_S für sehr großes n und beim Vorliegen einer binormalen Grundgesamtheit mit $\varrho = 0$ den Parameter ϱ mit einer asymptotischen Effizienz von $9/\pi^2$ oder 91,2%.
- (2) Für wachsendes n und binormalverteilte Zufallsvariable ist $2 \sin\left(\frac{\pi}{6} r_S\right)$ asymptotisch gleich r . Für $n \geq 100$ kann man daher neben r_S auch r angeben. So erhält man für
 $r_S = 0,840$ mit $\pi/6 = 0,5236$ ein
 $r = 2 \sin(0,5236 \cdot 0,840) = 2 \sin 0,4398 = 2 \cdot 0,426 = 0,852$.
- (3) Allgemein kann der Ansatz nach Fisher's $\hat{z}$ -Transformation auch für den Rangkorrelationskoeffizienten ϱ_S verwendet werden, wenn $n \geq 10$ und $\varrho_s < 0,9$ (Zar [Zar99]). Somit können die Verfahren dieses Abschnittes zum Hypothesentest, zu Konfidenzintervallen und zur Powerberechnung für ϱ auch für ϱ_S eingesetzt werden, allerdings sollte zur Korrektur nach Zar in den entsprechenden Formeln für den Term $1/(n-3)$ (beachte, dass $\sqrt{1/(n-3)}$ der Standardfehler von $\hat{z}$ ist) generell der Wert $1.06/(n-3)$ verwendet werden.

Statistische Modellbildung

8.1 Einführung

In zahlreichen wissenschaftlichen Studien (in der Medizin, der Industrie, der Ökonometrie) ist es erforderlich, den Zusammenhang zwischen mindestens zwei Variablen in mathematischen Modellen darzustellen. Diese Modelle führen zu

- einem besseren Verständnis dieser Zusammenhänge,
- ermöglichen Vorhersagen oder
- unterstützen Entscheidungsprozesse.

Dabei handelt es sich nicht um deterministische (vollständig reproduzierbare), sondern um **'stochastische' Zusammenhänge**, in denen eine Zufallskomponente zu berücksichtigen ist.

Die Suche nach einem geeigneten Modell geht dabei von Zielgrößen aus, die durch Einflussgrößen beeinflusst werden. **Ziel der Modellbildung ist es, die Eigenschaften einer Zielgröße durch eine Funktion von Einflussgrößen zu beschreiben (mathematisches Modell)**. Multivariate Verfahren, für den Fall mehrerer Zielgrößen, werden hier nicht näher behandelt. Eine Einführung geben Backhaus [BEPW03] und Timm [Tim02].

Unter einer Zielgröße (response variable) verstehen wir die Messungen oder Beobachtungen einer Zufallsvariable unter der Einwirkung von Einflussgrößen, Faktoren oder erklärenden Variablen (explanatory variables). Die in diesem Zusammenhang auch verwendete Bezeichnung als abhängige Variable und als unabhängige Variablen im Sinne einer funktionalen Zuordnung ist missverständlich. Entscheidend für das Verständnis der folgenden Methodenansätze ist, dass die Messungen oder Beobachtungen der Einflussgrößen im Rahmen der Modellbildung nicht als zufällig behandelt werden. Sie können durch das Design einer Studie durchaus als fest angenommen werden.

Zielgrößen (Y) können stetig gemessen oder in diskreten Werten beobachtet werden. Ihre Verteilung kann unter bestimmten Annahmen auch ohne Berücksichtigung von anderen Einflüssen durch spezielle Verteilungsmodelle beschrieben werden.

- Ist Y der systolische **Blutdruck** eines zufällig aus dem Patientengut einer Klinik ausgewählten Patienten, dann ist unter Umständen das Modell einer Normalverteilung mit dem Erwartungswert μ und der Varianz σ^2 angemessen (vgl. [5.4.2]):

$$Y \sim normal(\mu, \sigma^2) \quad (8.1)$$

- Ist Y die Anzahl der **Masernerkrankungen** bei Kindern im Alter unter 10 Jahren, dann kann das Modell der Poisson-Verteilung mit dem Erwartungswert λ für die Modellbildung verwendet werden (vgl. [5.3.4]):

$$Y \sim poisson(\lambda) \quad (8.2)$$

- Ist Y die Anzahl der beobachteten **Behandlungen** mit Nebenwirkungen unter insgesamt n Behandlungen, dann ist für die Modellbildung die Binomialverteilung mit der „Erfolgswahrscheinlichkeit“ π geeignet (vgl. [5.3.3]):

$$Y \sim binomial(n, \pi) \quad (8.3)$$

Die Wahl eines adäquaten Modells orientiert sich somit an der **Verteilung der Zielgrößen**. Ein falsches oder nicht geeignetes Modell führt zu einem systematischen Fehler in der Beschreibung und der statistischen Bewertung der Beobachtungen.

Die Einflussgrößen werden durch eine funktionale Beziehung mit dem Parameter (oder den Parametern) der Verteilung in das Modell eingeführt. So kann in dem Modell (8.1) mit

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 x_i + \epsilon_i & i = 1, \dots, n \\ \mu &= E[Y] = \beta_0 + \beta_1 x \end{aligned} \quad (8.4)$$

die Zielgröße linear anhand der Koeffizienten β_0 und β_1 mit dem Wert einer Einflussgröße (x) und einer normalverteilten **Zufallskomponente** ϵ_i in Zusammenhang gebracht werden. Die gewählte Funktion, ohne eine ausdrückliche stochastische Komponente, beschreibt die **systematische** oder **deterministische** Komponente des Modells. Mit Bezug auf die vorliegenden Beobachtungen im Rahmen einer Zufallsstichprobe ist diese Funktion um die **zufällige** Komponente zu ergänzen.



Im Modell (8.2) führt ein analoger Ansatz zu dem Problem, dass der Parameter λ nicht negative Werte annehmen kann. Daher erfolgt hier zweckmäßigerweise eine Transformation

$$\log(\lambda) = \log(E[Y]) = \beta_0 + \beta_1 x \quad (8.5)$$

Dieser Ansatz führt zum **log-linearen** Modell, in dem die Logarithmusfunktion als sogenannte „**Linkfunktion**“ verwendet wird.

Auch für den Parameter aus dem Modell der Binomialverteilung (8.3) gilt $0 < \pi < 1$. Zu seiner Modellierung dient der folgende Ansatz:

$$\text{logit}(\pi) = \log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x \quad (8.6)$$

Die logistische Transformation führt hier als Linkfunktion zum **logistischen** Regressionsmodell.

In analoger Weise können mehrere Einflussgrößen durch lineare oder auch nichtlineare (quadratische, exponentielle) Funktionen mit einer oder mehreren Veränderlichen in die Modellbildung eingebracht werden. **Die Modellbildung selbst besteht somit aus folgenden Stufen:**

- Wahl eines adäquaten Modellansatzes nach der Verteilung der Zielgröße (z.B. nach (8.1), (8.2) oder (8.3)).
- Suche nach einem geeigneten Modell für den Zusammenhang zwischen Einfluss- und Zielgröße, speziell Auswahl der Einflussgrößen und einer speziellen Linkfunktion.
- Schätzen der Koeffizienten β_0 und β_1 anhand der Beobachtungsdaten nach dem **Verfahren der kleinsten Abweichungsquadrate** (OLS, ordinary least squares) oder nach dem **Maximum-Likelihood-Verfahren** (MLE, maximum likelihood estimation).
- Überprüfen der Modellannahmen und bewerten der Güte des Modells anhand
 - (1) der Verteilung der Abweichungen ϵ_i ($i = 1, \dots, n$) zwischen geschätzten und tatsächlich beobachteten Werten in der Zielgröße sowie
 - (2) den statistischen Eigenschaften der geschätzten Modellparameter.

Der Prozess der Modellierung und die dabei verwendete Schreibweise soll am Beispiel der linearen Regression im folgenden Abschnitt verdeutlicht werden.

8.2 Regressionsmodelle

- Die einfache lineare Regression
- Die multiple lineare Regression
- Verfahren der Variablenauswahl
- Nominalskalierte Einflussgrößen

8.2.1 Die einfache lineare Regression

Im Modell der einfachen linearen Regression (nach Modellansatz 8.1) können die Beobachtungen Y_i durch die Werte einer Einflussgröße x_i , ergänzt um eine stochastische Komponente ϵ_i ($i = 1, \dots, n$) im Rahmen einer **Zufallsstichprobe** wie folgt dargestellt werden:

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i \quad (8.7)$$

Insbesondere wird angenommen, dass die Fehler ϵ_i normalverteilt sind nach $\epsilon \sim N(0, \sigma^2)$, d.h. der Erwartungswert für die Abweichungen ist 0 und die Varianz σ^2 resultiert aus der Variabilität in den beobachteten Daten. Unter diesen Voraussetzungen lassen sich die Koeffizienten β_0 und β_1 des Modells nach der Methode der kleinsten Fehlerquadrate schätzen.

$$SAQ_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2 \quad SAQ_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

$$\hat{\beta}_1 = b = \frac{SAQ_{xy}}{SAQ_{xx}} \quad \hat{\beta}_0 = a = \bar{y} - b\bar{x}$$

Hinweis: Eine ausführliche Darstellung zur einfachen linearen Regression, insbesondere auch zur Schätzung der Koeffizienten nach der Methode der kleinsten Fehlerquadrate, findet sich im Abschnitt [3.6.6].



Beispiel: Bei 24 Patienten mit einer Hyperlipoproteinämie wurde der Cholesterinwert bestimmt. Unter der Annahme, dass der Cholesterinwert altersbedingt steigt, stellt sich die Frage: Wie gut lässt sich die Höhe des Cholesterinwertes (Zielgröße) in einem linearen Modell aus dem Alter der Patienten (Einflussgröße) schätzen?

Alter (x)	46	20	52	30	57	25	28	36	22	43	57	33
Cholesterin (y)	3,5	1,9	4,0	2,6	4,5	3,0	2,9	3,8	2,1	3,8	4,1	3,0
Alter (x)	22	63	40	48	28	49	52	58	29	34	24	50
Cholesterin (y)	2,5	4,6	3,2	4,2	2,3	4,0	4,3	3,9	3,3	3,2	2,5	3,3

Die elementare Berechnung kann in **R** in wenigen Schritten erfolgen:

```
> ss.xx <- sum((Alter - m.x)^2);          ss.xx
[1] 4139.833
> ss.xy <- sum((Alter - m.x)*(Chol - m.y)); ss.xy
[1] 217.8583
>                                     # Schätzung der Koeffizienten
> betal <- ss.xy / ss.xx;              betal
[1] 0.0526249
> beta0 <- m.y - betal * m.x;          beta0
[1] 1.279868
```

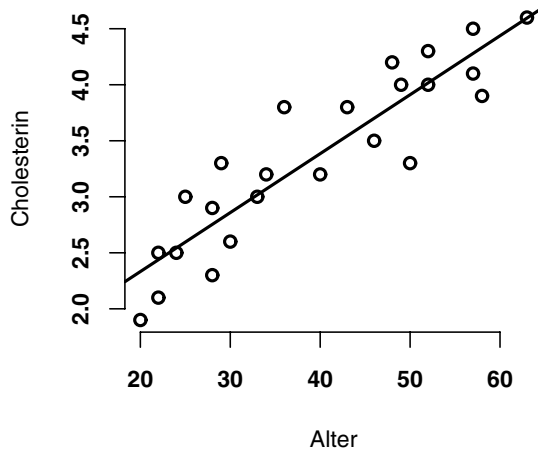


Abb. 8.1 Cholesterin in Abhängigkeit vom Alter bei 24 Patienten mit Hyperlipoproteinämie

Das Ergebnis der Schätzung $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = 1,28 + 0,053x$, nach dem für das Alter (x) ein entsprechender Cholesterinwert ($\hat{y}$) berechnet werden kann, ist in Abbildung 8.1 dargestellt.

Der **Modellcharakter** wird besonders deutlich in der **Matrixschreibweise** zu (8.7). Insbesondere erleichtert diese Darstellung die Verallgemeinerung auf mehrere Einflussgrößen.

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (8.8)$$

mit

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

Eine Schätzung $\hat{\boldsymbol{\beta}}$ der Koeffizienten, in der der Fehler $\boldsymbol{\epsilon}$ möglichst klein ist, ergibt sich aus der Lösung des entsprechenden Gleichungssystems

$$\begin{aligned} (\mathbf{Y} - E(\mathbf{Y}))'(\mathbf{Y} - E(\mathbf{Y})) &= (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \\ &= \boldsymbol{\epsilon}'\boldsymbol{\epsilon} = \text{minimal} \end{aligned}$$

mit

$$\boxed{\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}} \quad (8.9)$$

Die Güte des gewählten Modells, insbesondere die Verlässlichkeit der daraus abgeleiteten Schätzwerte $\hat{y}_i$, wird einerseits an den Abweichungen (**Residuen**) untersucht, andererseits werden die geschätzten **Koeffizienten** mit Verfahren der Inferenzstatistik bewertet.

Die Schätzung der Residuen $\hat{\epsilon}$ wird aus der Differenz zwischen den geschätzten und beobachteten Werten der Zielgröße berechnet:

$$\boxed{\hat{\epsilon} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}} \quad (8.10)$$

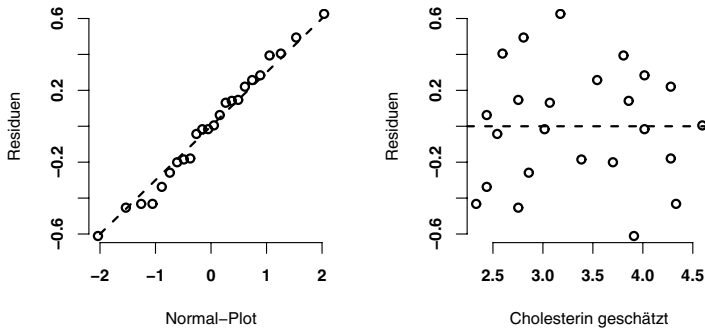


Abb. 8.2 Graphische Residuenanalyse für die Cholesterindaten aus dem Beispiel

Die Summe der Abweichungsquadrate ($RSS = \hat{\epsilon}'\hat{\epsilon}$) ist für die weiteren Rechnungen von zentraler Bedeutung. Unter den genannten Modellannahmen sind die Residuen normalverteilt nach $\epsilon \sim N(0, \sigma^2)$ (vgl. 8.7). Somit können die wichtigsten Modellannahmen durch eine Analyse der Verteilung der Residuen, zum Beispiel graphisch in einem **Normal-Plot** und durch eine **Punktwolke** überprüft werden. Abbildung 8.2 zeigt, dass die Residuen aus dem Beispiel modellkonform zufällig verteilt sind. Ein Fehler in der Modellwahl käme zum Beispiel in einer nichtzufälligen Verteilung der Residuen zum Ausdruck. So kann eine (funktionale) Abhängigkeit zwischen den Residuen und der Zielgröße bestehen oder die Streuung der Residuen eine Abhängigkeit von der Zielgröße zeigen. Durch geeignete Transformationen oder Gewichtungen in der Einflussgröße kann ein Fehler bei der Modellierung vermieden werden.

Auffällig streuende **Einzelwerte** (Ausreißer oder Extremwerte), die nicht durch das Modell erfasst werden, verfälschen die Schätzung. Die OLS-Schätzung ist in diesem Sinne keine robuste Schätzung der Modellparameter. Extremwerte können ausgeschlossen werden, wenn sie aus Messfehlern oder Beobachtungsfehlern resultieren. Anderenfalls müssen Verfahren der **robusten Regressionsrechnung** eingesetzt werden, in denen zur Schätzung der Parameter die Summe der absoluten Abweichungen vom Mittelwert oder vom Median minimal ausfällt (MAD-Schätzer).

Die statistischen Eigenschaften der geschätzten Koeffizienten sind in den folgenden Schritten nachzuvollziehen. Dabei soll auch für den Fall der einfachen linearen Regression die Matrixschreibweise verwendet werden, die im Rahmen der multiplen linearen Regression notwendig ist (Abschnitt [8.2.2]).

- $\hat{\beta}$ ist eine unverzerrte Schätzung für β , das heißt $E[\hat{\beta}] = \beta$. Die Güte der Schätzung wird durch die Varianz der Schätzung bestimmt, die sich aus dem Produkt der geschätzten Residualvarianz und dem i ten Diagonalelement der Inversen der Matrix $X'X$ bestimmen lässt, das heißt $\widehat{Var}[\hat{\beta}_i] = \hat{\sigma}^2((X'X)^{-1})_{ii}$.
- Eine unverzerrte Schätzung für σ^2 wird als **Residualvarianz** aus den Residuen abgeleitet:

$$\hat{\sigma}^2 = s_{y.x}^2 = \frac{\sum (y_i - \hat{y}_i)^2}{n - 2} = \frac{\hat{\epsilon}'\hat{\epsilon}}{n - 2} \quad (8.11)$$

Die Residualvarianz kennzeichnet somit den mittleren Schätzfehler (im Modell) und bildet die Grundlage für alle weiteren Rechnungen.

- Der **Standardfehler** der geschätzten Regressionskoeffizienten wird aus $s_{y.x}^2$ und den entsprechenden Diagonalelementen (hier mit den Indizes 11 und 22) der Varianz-Kovarianzmatrix $(X'X)^{-1}$ bestimmt:

$$\begin{aligned} se(\hat{\beta}_0) &= \sqrt{s_{y.x}^2 \{(X'X)^{-1}\}_{11}} \\ se(\hat{\beta}_1) &= \sqrt{s_{y.x}^2 \{(X'X)^{-1}\}_{22}} \end{aligned} \quad (8.12)$$

- Das $(1 - \alpha)$ -**Konfidenzintervall** für β_i wird dann bestimmt nach:

$$\hat{\beta}_i \pm se(\hat{\beta}_i) t_{n-2, 1-\alpha/2} \quad i = 0, 1 \quad (8.13)$$

- Ein **Hypothesentest** zur Prüfung der einzelnen Koeffizienten im Modell $H_0 : \beta_i = 0$ ist über die folgende Teststatistik möglich:

$$\hat{t} = \frac{\hat{\beta}_i}{se(\hat{\beta}_i)} \sim t_{n-2} \quad i = 0, 1 \quad (8.14)$$

- Eine zusammenfassende Prüfung aller Koeffizienten im Regressionsmodell kann durch eine Varianzanalyse (vgl. [7.5.2]) erfolgen. Die Zerlegung der Abweichungsquadrate in der Zielgröße nach

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$$SSY = MSS + RSS$$

ist auch in der Matrixschreibweise darstellbar mit

$$\begin{aligned} SSY &= \mathbf{y}'\mathbf{y} - n\bar{y}^2 \\ MSS &= \hat{\beta}'X'\hat{\mathbf{y}} - n\bar{y}^2 \\ RSS &= \mathbf{y}'\mathbf{y} - \hat{\beta}'X'\mathbf{y} = \hat{\epsilon}'\hat{\epsilon} \end{aligned} \quad (8.15)$$

Für die Prüfung der Nullhypothese, dass alle Koeffizienten 0 sind,

$$H_0 : \beta_0 = 0 \quad \text{und} \quad \beta_1 = 0$$

wird der folgende Varianzquotient berechnet

$$\hat{F} = \frac{MSS}{RSS/(n-2)} \sim F_{k, n-(k+1)} \quad (8.16)$$

- Die Güte (Qualität) des vollständigen Modells, insbesondere unter Berücksichtigung der Regressionskonstanten β_0 , kann unter optimalen Bedingungen durch das Bestimmtheitsmaß zusammenfassend bewertet werden:

$$R^2 = \frac{SSY - RSS}{SSY} = 1 - \frac{RSS}{SSY} \quad (8.17)$$

$0 \leq R^2 \leq 1$ beschreibt den Anteil der durch das Modell erklärten Varianz der Zielgröße. Je größer R^2 ist, desto besser kann die Zielgröße in dem Modell beschrieben werden.

Die Berechnung dieser Statistiken erfolgt in **R** am einfachsten mit der Funktion **lm()**, über die auch die Modellierung multipler Zusammenhänge, wie sie im folgenden Abschnitt näher dargestellt wird, erfolgen kann:

```
> lin.model <- lm(Chol ~ Alter)
> summary(lin.model)

Call:
lm(formula = Chol ~ Alter)

...

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.279868    0.215699   5.934 5.69e-06 ***
Alter       0.052625    0.005192  10.136 9.43e-10 ***

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

...

Multiple R-Squared:  0.8236,
F-statistic: 102.7 on 1 and 22 DF,  p-value: 9.428e-10
```

Das Ergebnis der Schätzung für die Koeffizienten $\hat{\beta}_0 = 1,28$ (Intercept) und $\hat{\beta}_1 = 0,053$ (Alter) stimmt mit dem oben elementar berechneten Resultat überein. Zusätzlich werden in der Funktion **lm()** die Standardfehler $se(\hat{\beta}_0) = 0,216$ und $se(\hat{\beta}_1) = 0,0052$ nach (8.12), die Hypothesentests hinsichtlich der Koeffizienten nach (8.14) und die Varianzanalyse nach (8.16) mit $\hat{F} = 102,7$ ($P < 0,001$) berechnet. Das R^2 aus (8.17) hat den Wert 0,82. Somit können etwa 82% der Varianz der Zielgröße (hier das Cholesterin) durch das Modell erklärt werden.

8.2.2 Die multiple lineare Regression

Die Betrachtung von p Einflussgrößen $x_1, x_2, \dots, x_p$, z.B. multiple Risiken oder unterschiedliche Aspekte aus der Krankengeschichte in Beobachtungsstudien, führt zu dem Modell der multiplen linearen Regression. Jede Einflussgröße (Variable) hat einen eigenen Effekt auf die Zielgröße y . Dieser partielle Effekt resultiert aus einer Erhöhung von x_i um eine Einheit während alle anderen x_j ($i \neq j$) konstant gehalten werden; er wird durch den Regressionskoeffizienten β_i beschrieben. Das vollständige Modell für insgesamt n Beobachtungen kann dann wie folgt beschrieben werden.

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

$$\mathbf{y} = \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\epsilon}$$

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \epsilon_i$$

Das Regressionsproblem besteht darin, die Koeffizienten so zu bestimmen, dass eine Schätzung $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ möglichst nahe an $\mathbf{y}$ liegt, d.h. die aus der Schätzung resultierenden Abweichungen (Residuen) $\hat{\boldsymbol{\epsilon}} = \mathbf{y} - \hat{\mathbf{y}}$ sollen minimal sein. Geometrisch bedeutet dies, für die beobachteten Werte $\mathbf{y} \in \mathbb{R}^n$ eine optimale Schätzung mit Hilfe von $\boldsymbol{\beta} \in \mathbb{R}^p$ zu finden.

Die Schätzung resultiert entsprechend Abbildung 8.3 aus einer orthogonalen Projektion von $\mathbf{y}$ auf die durch die Beobachtungen $\mathbf{X}$ aufgespannte (Modell-)Ebene. Algebraisch bedeutet dies eine kleinste Fehlerquadratschätzung in $\boldsymbol{\beta}$ analog zum vorangehenden Abschnitt.

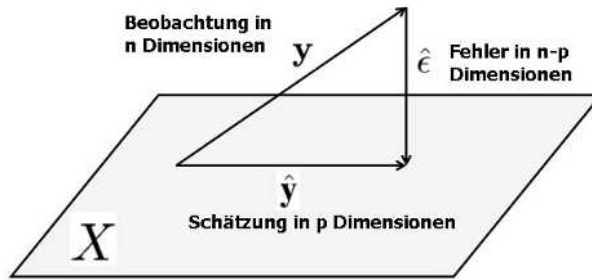


Abb. 8.3 Geometrische Darstellung der Schätzung von $\mathbf{y}$ durch $\hat{\mathbf{y}}$ und des damit verbundenen Fehlers $\hat{\epsilon}$ durch eine orthogonale Projektion von $\mathbf{y}$ auf die Ebene X

$$\sum_{i=1}^n \epsilon_i^2 = \epsilon' \epsilon = (\mathbf{y} - X\beta)'(\mathbf{y} - X\beta) = \text{minimal} \quad (8.18)$$

Die Lösung von (8.18) führt über die Ableitung nach β auf:

$$\begin{aligned} \hat{\beta} &= (X'X)^{-1}X'\mathbf{y} \\ \text{Damit folgt: } \hat{\mathbf{y}} &= X\hat{\beta} = \underbrace{X(X'X)^{-1}X'}_H \mathbf{y} \\ &= H \mathbf{y} \end{aligned}$$

H bezeichnet die so genannte **'Hut'-Matrix** (hat-matrix). Sie beschreibt numerisch die Projektion von $\mathbf{y}$ auf die durch X definierte Ebene. Das Modell der multiplen linearen Regression kann somit in Matrixschreibweise sehr übersichtlich zusammengefasst werden:

$\hat{\beta} = (X'X)^{-1}X'\mathbf{y}$	Schätzung für β	(8.19)
$\hat{\mathbf{y}} = H\mathbf{y} = X\hat{\beta}$	Schätzung für $\mathbf{y}$	
$\hat{\epsilon} = \mathbf{y} - X\hat{\beta} = \mathbf{y} - \hat{\mathbf{y}} = (I - H)\mathbf{y}$	Schätzfehler	
$\hat{\epsilon}'\hat{\epsilon} = \mathbf{y}'(I - H)\mathbf{y}$	Summe der Abweichungssquadrate, RSS (residual sum of squares)	

$\hat{\beta}$ ist somit eine geometrisch sinnvolle und nachvollziehbare Schätzung für β . Es lässt sich zeigen, dass unter den zusätzlichen Annahmen **unabhängiger und normalverteilter Fehler** (8.20) $\hat{\beta}$ auch der beste, linear unverzernte (best linear unbiased) Schätzer ist (Gauss-Markov-Theorem), der unter diesen Bedingungen auch gleich der entsprechenden Maximum-Likelihood-Schätzung ist.

$$E[\epsilon] = 0; \quad \text{Var}[\epsilon] = \sigma^2 I \quad \Rightarrow \quad \epsilon \sim N(0, \sigma^2 I) \quad (8.20)$$

Die Schätzung der Varianz erfolgt nach

$$\hat{\sigma}^2 = \frac{RSS}{n - p - 1} = \frac{\hat{\epsilon}'\hat{\epsilon}}{n - p - 1} \quad (8.21)$$

Für den Standardfehler der geschätzten Regressionskoeffizienten folgt daraus:

$$se(\hat{\beta}_i) = \sqrt{(X'X)^{-1}_{ii} \hat{\sigma}^2} \quad (8.22)$$

Die Güte der Schätzung kann zusammenfassend durch das Bestimmtheitsmaß R^2 beschrieben werden, welches den Anteil der im Modell erklärten Varianz von y ausdrückt.

$$R = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} = 1 - \frac{RSS}{SSY} \quad (8.23)$$

Beispiel: Der Zusammenhang zwischen der Größe eines Wurfes bei Mäusen (lsize, litter size) sowie dem Körpergewicht (bodywt, body weight) und Gehirngewicht (brainwt, brain weight) des Muttertiers soll auf der Grundlage von 20 Würfen bei Mäusen untersucht werden (Beispiel litters aus dem Zusatzpaket **library(DAAG)** in **R** (Maindonald [MB04])). Die beobachteten Daten sind in der folgenden Tabelle aufgeführt; eine übersichtliche graphische Darstellung der Abhängigkeiten zwischen diesen drei Variablen erfolgt durch die Punktwolken in Abbildung 8.4.

lsize	bodywt	brainwt
3	9.447	0.444
4	9.155	0.417
5	8.850	0.425
6	8.298	0.404
7	7.400	0.409
8	7.040	0.414
9	6.600	0.387
10	6.305	0.410
11	7.183	0.435
12	5.450	0.368

lsize	bodywt	brainwt
3	9.780	0.436
4	9.613	0.429
5	9.610	0.434
6	8.543	0.439
7	8.335	0.429
8	7.253	0.409
9	7.260	0.433
10	6.655	0.405
11	6.133	0.407
12	6.050	0.401

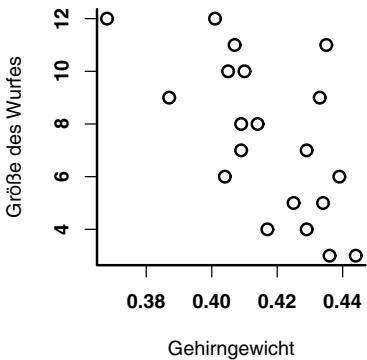
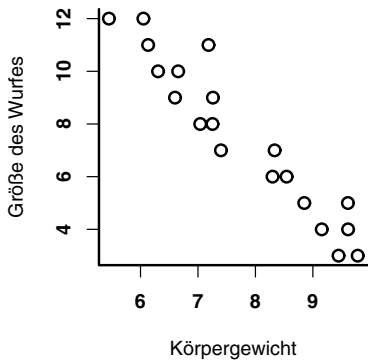


Abb. 8.4 Punktwolken zur Abhängigkeit der Wurfgröße von dem Körpergewicht und dem Gehirngewicht in einer Mäusepopulation

Die Rechnung soll anhand der Formeln (8.19) bis (8.23) ausführlich in einzelnen Schritten mit **R** gezeigt werden. Besonders wird auf den Abschnitt [2.4] hingewiesen, in dem die wesentlichen Operationen mit Matrizen näher beschrieben sind.

Zunächst werden die beobachteten Werte in einer Matrix (X) bzw. in einem Vektor (y) gespeichert.

```
> lsize <- c(3,3,4,4,5,5,6,6,7,7,8,8,9,9,10,10,11,11,12,12)
> bodywt <- c(9.447,9.780,9.155,9.613,8.850,9.610,8.298,8.543,7.400,8.335,
+ 7.040,7.253,6.600,7.260,6.305,6.655,7.183,6.133,5.450,6.050)
> brainwt <- c(0.444,0.436,0.417,0.429,0.425,0.434,0.404,0.439,0.409,0.429,
+ 0.414,0.409,0.387,0.433,0.410,0.405,0.435,0.407,0.368,0.401)
> y <- lsize;
> X <- matrix(c(rep(1,20),bodywt,brainwt),nrow=20); p <- 2
> data.frame(cbind(rep(" ", 20), y, rep(" ", 20), X))
```

	y	X
1	3	1 9.447 0.444
2	3	1 9.78 0.436
3	4	1 9.155 0.417
4	4	1 9.613 0.429
5	5	1 8.85 0.425
6	5	1 9.61 0.434
7	6	1 8.298 0.404
8	6	1 8.543 0.439
9	7	1 7.4 0.409
10	7	1 8.335 0.429
11	8	1 7.04 0.414
12	8	1 7.253 0.409
13	9	1 6.6 0.387
14	9	1 7.26 0.433
15	10	1 6.305 0.41
16	10	1 6.655 0.405
17	11	1 7.183 0.435
18	11	1 6.133 0.407
19	12	1 5.45 0.368
20	12	1 6.05 0.401

Für die Berechnung der transponierten Matrix (X') wird die Funktion **t()** verwendet. Das Produkt der Matrizen wird mit **%*%** gebildet und die inverse Matrix mit der Funktion **solve()** berechnet - $(X'X)^{-1}$.

```
> t(X) %*% X
      [,1]      [,2]      [,3]
[1,] 20.000 154.96000  8.335000
[2,] 154.960 1235.85592 64.948762
[3,]  8.335  64.94876  3.480561
>
> xtxi <- solve(t(X) %*% X); xtxi
      [,1]      [,2]      [,3]
[1,] 38.3034161  0.9216196 -108.924114
[2,]  0.9216196  0.0640439 -3.402116
[3,] -108.9241143 -3.4021156 324.615971
```

Mit Hilfe der Matrix $(X'X)^{-1}$ kann eine Schätzung der Parameter $\hat{\beta}_0 = 12,9$ (Achsenabschnitt), $\hat{\beta}_1 = -2,4$ und $\hat{\beta}_2 = 31,6$ direkt berechnet werden.

```
> b.h <- xtxi %*% t(X) %*% y; b.h # Schaetzen der Parameter
      [,1]
[1,] 12.898778
[2,] -2.398031
[3,] 31.628479
```

Mit Hilfe der Hut-Matrix $X(X'X)^{-1}X'$ können die Schätzungen $\hat{y}$ bzw. die Residuen $\hat{e}$ bestimmt werden.

```
> H <- X %*% xtxi %*% t(X) # Berechnung der Hut-Matrix
>
> y.h <- H %*% y; # Schaetzen der Werte (x %*% b.h)
> e.h <- y - y.h; # Schaetzfehler - Residuen
> cbind(y, y.h, e.h)
```

	y	y.h	e.h
[1,]	3	4.287621	-1.28762073
[2,]	3	3.236048	-0.23604844

```
[3,] 4 4.133877 -0.13387695
[4,] 4 3.415120 0.58487967
[5,] 5 5.118304 -0.11830436
[6,] 5 3.580457 1.41954318
[7,] 6 5.777820 0.22218038
[8,] 6 6.297299 -0.29729871
[9,] 7 8.089394 -1.08939421
[10,] 7 6.479804 0.52019555
[11,] 8 9.110828 -1.11082791
[12,] 8 8.441905 -0.44190482
[13,] 9 9.311993 -0.31199277
[14,] 9 9.184202 -0.18420211
[15,] 10 10.746867 -0.74686706
[16,] 10 9.749414 0.25058632
[17,] 11 9.432107 1.56789251
[18,] 11 11.064443 -0.06444302
[19,] 12 11.468788 0.53121224
[20,] 12 11.073709 0.92629125
```

Die Schätzung der Standardabweichung $\hat{\sigma}$ erfolgt über die Summe der Abweichungsquadrate $RSS = \hat{\epsilon}'\hat{\epsilon} = 11,5$. Der Standardfehler für die Regressionskoeffizienten wird dann mit Hilfe der Funktion **diag()** aus Diagonalelementen der Matrix $(X'X)^{-1}$ berechnet. Das Bestimmtheitsmaß ist $R = 0,93$.

```
> RSS <- t(e.h) %*% e.h; RSS # Summe über die quadrierten Abweichungen
[1,]
[1,] 11.48166
> s.h <- sqrt(RSS/(n-p-1)) # Schätzung der Standardabweichung
>
> se.b <- sqrt(diag(xtxi))*s.h; se.b # Standardfehler der Schätzung
[1] 5.0862375 0.2079777 14.8068549
>
> R <- 1 - RSS/sum((y-mean(y))^2);R # Berechnung von R
[1,]
[1,] 0.9304142
```

Die an dem Beispiel gezeigten Berechnungen werden in **R** in der Funktion **lm()** zusammengefasst. Das wichtigste Argument zu dieser Funktion ist die Spezifikation des Modells in der allgemeinen Form

Zielgröße $\sim$ Einflussgröße(n).

Für die Aufnahme mehrerer Einflussgrößen in die **Modellgleichung** ist dabei eine spezielle Syntax zu beachten, die auch Konstanten und Interaktionen im Rahmen der Modellbildung zulässt (vgl. auch Tabelle 9.7 in Kapitel [9]).

Das Ergebnis der Funktion **lm()** stellt in **R** ein eigenes Objekt vom Typ eines linearen Modells dar, dessen Eigenschaften mit der Funktion **summary()** übersichtlich dargestellt werden können.

```
> library(DAAG)
Loading required package: leaps
Loading required package: oz
> data(litters)
> fit <- lm(lsize ~ bodywt + brainwt, data=litters)
> summary(fit)

Call:
lm(formula = lsize ~ bodywt + brainwt, data = litters)

Residuals:
    Min       1Q   Median       3Q      Max
-1.2876 -0.3445 -0.1261  0.5229  1.5679

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   12.899      5.086   2.536  0.0213 *
bodywt        -2.398      0.208  -11.530 1.85e-09 ***
brainwt       31.628     14.807   2.136  0.0475 *
---
Signif. codes:  0.001 '***' 0.01 '**' 0.05 '*' 0.1 '.' 0.2 ' ' 1
```

```

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8218 on 17 degrees of freedom
Multiple R-Squared:  0.9304,
F-statistic: 113.7 on 2 and 17 DF,  p-value: 1.450e-10

```

Die Ergebnisse stimmen mit den oben „elementar“ berechneten Ergebnissen überein. Darüber hinaus werden hier einige Statistiken berechnet, die im nächsten Abschnitt näher beschrieben sind.

Die Lösung des dem Modell der multiplen linearen Regression zugrundeliegenden Gleichungssystems ist nur möglich, wenn $X'X$ nicht-singulär ist und somit eine (generalisierte) Inverse berechnet werden kann. Eine **Singularität** in $X'X$ wird insbesondere durch lineare Abhängigkeiten der Einflussgrößen untereinander bedingt; praktisch können auch schon quasilineare Abhängigkeiten dazu führen, dass die Matrix $X'X$ „fast singulär“ ist. Diese Form der Abhängigkeit unter den Einflussgrößen wird **Kollinearität** genannt (Interkorrelation, nicht orthogonale Einflussgrößen). Eine direkte Folge der Kollinearität ist, dass die Standardfehler der geschätzten partiellen Regressionskoeffizienten sehr groß werden können oder nur geringfügige Modifikationen im Modell zu sehr unterschiedlichen Ergebnissen führen können (instabile Modelle). Eine gute (sichere) Schätzung der Zielgröße ist somit nicht möglich, obwohl bei einer bivariaten Betrachtungsweise ein deutlicher Zusammenhang vorliegt. Die Analyse der Varianz- bzw. Kovarianzmatrix der geschätzten partiellen Regressionskoeffizienten zeigt eine mögliche Kollinearität auf und ermöglicht den gezielten, theoretisch begründeten Ausschluss einzelner Einflussgrößen im Rahmen der Modellbildung. Zur Vermeidung der Kollinearität kann aber auch die Zahl der Fälle erhöht werden oder die Zahl der Einflussgrößen reduziert werden, unter anderem durch das Zusammenfassen hochkorrelierter Variablen zu einem Faktor.

Korrelationen zwischen den unabhängigen Variablen im linearen Regressionsmodell treten häufig bei **Umfragedaten** auf. So werden einerseits bestimmte Kombinationen von Merkmalsausprägungen nicht beobachtet, oder es gibt andererseits unmögliche / nicht plausible Kombinationen, wie zum Beispiel die Tätigkeit als ungelernter Arbeiter und der Ausbildung mit Hochschulabschluss. Enge Beziehungen im Sinne der Kollinearität können so zum Beispiel zwischen dem Schulabschluss, dem Beruf, dem monatlichen Einkommen und der politischen Meinung bestehen.

8.2.2.1 Hypothesentest und Konfidenzintervalle zum linearen Modell

Aus der Annahme, dass die Residuen unabhängig und identisch normalverteilt sind ($\epsilon \sim N(0, \sigma^2 I)$) resultiert, dass die Schätzung der Koeffizienten $\hat{\beta}$ ebenfalls normalverteilt ist mit dem Erwartungswert β und der Varianz $(X'X)^{-1}\sigma^2$.

$$\boxed{\hat{\beta} = (X'X)^{-1}X'y \sim N(\beta, (X'X)^{-1}\sigma^2)} \quad (8.24)$$

Mit Hilfe dieser Verteilung lassen sich Teststatistiken für die Überprüfung von Hypothesen hinsichtlich der Koeffizienten im linearen Modell ableiten. Die generelle Nullhypothese, dass es unter den Einflussgrößen mindestens einen „signifikanten Prädiktor“ gibt, d.h. dass die Zielgröße nicht genauso gut allein aus dem Erwartungswert geschätzt werden kann, ist in (8.25) angegeben.

$$H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0 \quad (8.25)$$

Die Teststatistik (8.26) zu dieser Hypothese ist F-verteilt mit $n, (n - p - 1)$ -Freiheitsgraden und gestattet eine Aussage zum Erklärungswert des vollständigen Modells.

$$\hat{F} = \frac{\frac{1}{p} \cdot (SSY - RSS)}{\frac{1}{(n-p-1)} \cdot RSS} \quad (8.26)$$

mit $SSY = (\mathbf{y} - \bar{y})'(\mathbf{y} - \bar{y})$
 und $RSS = (\mathbf{y} - \hat{\mathbf{y}})'(\mathbf{y} - \hat{\mathbf{y}}) = (\mathbf{y} - X\hat{\beta})'(\mathbf{y} - X\hat{\beta}) = \hat{\epsilon}'\hat{\epsilon}$

Die Berechnung von (8.26) kann in **R** auch explizit durchgeführt werden. Das Ergebnis stimmt mit den in der Funktion **summary(fit)** im vorangegangenen Abschnitt berechneten Resultaten überein.

```
> fit <- lm(lsize ~ bodywt + brainwt, data=litters)
> RSS <- sum(fit$res^2)
> SSY <- sum((litters$lsize - mean(litters$lsize))^2)
> p <- 2; n <- 20
> F <- ((SSY-RSS)/p)/(RSS/(n-p-1)); F
[1] 113.6513
> p <- 1-pf(F, p, n-p-1); p
[1] 1.450194e-10
```

Einzelne Koeffizienten lassen sich nach (8.27) gezielt überprüfen. Die entsprechende Teststatistik ist t-verteilt mit $(n-p-1)$ Freiheitsgraden. Die Ergebnisse können mit der Funktion **summary(fit)** (s.o.) übersichtlich angezeigt werden. Auf eine explizite Berechnung der Teststatistik, mit dem Koeffizienten aus $(\text{fit}\$coef[i])$ und dem entsprechenden Standardfehler nach $\sqrt{(X'X)^{-1}_{ii}\hat{\sigma}^2}$, wird an dieser Stelle verzichtet.

$$\begin{aligned} H_0 : & \quad \beta_i = 0 \quad \text{für ein } i \in \{1, 2, \dots, p\} \\ H_A : & \quad \beta_i \neq 0 \\ \text{Teststatistik:} & \quad \hat{t}_i = \frac{\hat{\beta}_i}{se(\hat{\beta}_i)} \end{aligned} \quad (8.27)$$

Für die Bewertung der Unsicherheit einer Schätzung können besser auch **Konfidenzintervalle** verwendet werden. Entsprechend der Verteilung der geschätzten Koeffizienten (8.24) kann ein $(1 - \alpha)100\%$ -Konfidenzintervall nach (8.28) angegeben werden.

$$\hat{\beta}_i \pm t_{(n-p-1), 1-\alpha/2} \hat{\sigma} \sqrt{(X'X)^{-1}_{ii}} \quad (8.28)$$

Für die Unsicherheit einer **Vorhersage** (prediction) $\hat{y}_0$ auf der Grundlage beobachteter oder hypothetischer Werte der Einflussgrößen $\mathbf{x}_0$ ($x_{0i}, i = 1, 2, \dots, p$) sind zwei Situationen zu unterscheiden.

- Die Unsicherheit der Vorhersage **eines einzelnen zukünftigen Wertes** wird einerseits durch die Unsicherheit der Schätzung des Mittelwertes $\hat{y}_0$ bestimmt. Andererseits kommt noch die Abweichung eines einzelnen Wertes vom Mittelwert ϵ_i hinzu. Es spielen somit die folgenden Varianzen eine Rolle:

$$Var(\hat{y}_0) = Var(\mathbf{x}'_0 \hat{\beta}) = \mathbf{x}'_0 (X'X)^{-1} \mathbf{x}_0 \hat{\sigma}^2$$

$$Var(\hat{\epsilon}) = \hat{\sigma}^2 I$$

Daraus kann ein $(1 - \alpha)100\%$ -**Prädiktionsintervall** wie folgt abgeleitet werden:

$$\hat{y}_0 \pm t_{(n-p-1), 1-\alpha/2} \hat{\sigma} \sqrt{1 + \mathbf{x}_0'(X'X)^{-1}\mathbf{x}_0} \quad (8.29)$$

- Die Vorhersage für **einen „mittleren“ zukünftigen Wert** kann dagegen mit einem geringeren Fehler (ohne den Einfluss von $Var(\hat{\epsilon})$) durch das folgende Prädiktionsintervall beschrieben werden:

$$\hat{y}_0 \pm t_{(n-p-1), 1-\alpha/2} \hat{\sigma} \sqrt{\mathbf{x}_0'(X'X)^{-1}\mathbf{x}_0} \quad (8.30)$$

Für das Beispiel soll mit **R** im Einzelfall die Wurfgröße für eine Maus mit dem Körpergewicht 8 mg und dem Gehirngewicht 0,4 mg geschätzt und das zugehörige 95%-Prädiktionsintervall (8.29) bestimmt werden:

```
> fit <- lm(lsize ~ bodywt + brainwt, data=litters)
> p <- 2; n <- 20
> x0 <- c(1.0, 8.0, 0.4) # einzelne Beobachtung
> y0 <- sum(x0*fit$coef) # Schätzung der Wurfgrösse
> X <- cbind(1, bodywt, brainwt) # Aufbau der X-Matrix
> xtxi <- solve(t(X) %*% X) # Varianz-Matrix
>
> t <- qt(0.975, n-p-1) # Quantil der t-Verteilung
> sigma <- sqrt(sum(fit$res^2)/(n-p-1)) # Schätzung Standardabweichung
> W <- sqrt(1 + x0 %*% xtxi %*% x0) # Wurzelterm
> round(c(y0-t*sigma*W, y0, y0+t*sigma*W), 2)
[1] 4.49 6.37 8.24
```

Die geschätzte Wurfgröße ist $\hat{y}_0 = 6,4$, das 95%-Prädiktionsintervall umfasst die Werte von 4,4 bis 8,3 (jeweils nach „außen“ gerundet).

8.2.3 Verfahren der Variablenauswahl

Von großer Bedeutung im Rahmen der multiplen statistischen Modellierung ist die Frage, welche Einflussgrößen zu berücksichtigen sind und welchen Stellenwert einzelne Größen im Modell haben. Ein übergeordnetes Zielkriterium für die Modellbildung kann an den Residuen bzw. an dem Bestimmtheitsmaß formuliert werden. Während die Summe der Abweichungsquadrate RSS (residual sum of squares) möglichst klein sein soll, ist ein möglichst großer Wert in dem Bestimmtheitsmaß anzustreben:

$$RSS \longrightarrow 0 \quad (\text{minimal}) \qquad R^2 \longrightarrow 1 \quad (\text{maximal})$$

Hinweis: Das Bestimmtheitsmaß R^2 steigt mit der Anzahl unabhängiger Variablen. Um Modelle mit unterschiedlicher Anzahl von Variablen vergleichen zu können, muss R^2 entsprechend (8.31) angepasst (adjusted) werden.

$$R_a^2 = 1 - \frac{RSS/(n - (p + 1))}{SSY/(n - 1)} = 1 - \frac{(n - 1)}{n - (p + 1)}(1 - R^2) \quad (8.31)$$

Für den Prozess der Modellbildung gibt es verschiedene Strategien. Neben der Teilmengenanalyse werden häufig **schrittweise Verfahren** (stepwise regression modelling) eingesetzt, wobei einerseits eine gute Anpassung an die vorliegenden Daten, andererseits eine gute Vorhersage „zukünftiger“ Beobachtungen mit möglichst wenigen Einflussgrößen angestrebt wird.

Mit den p Einflussgrößen, die in der Modellbildung berücksichtigt werden sollen, werden im Rahmen einer **Teilmengenanalyse** alle Teilmengen von Einflussgrößen gebildet, aus denen dann jeweils ein Regressionsmodell abgeleitet wird. Unter diesen insgesamt 2^p Modellen wird das Modell gewählt, welches den höchsten Wert für das Bestimmtheitsmaß bzw. den kleinsten Wert für die

Restvarianz liefert (best subset regression). Abgesehen von dem Rechenaufwand, der unter Verwendung leistungsfähiger Computer / Programme bewältigt werden kann, ist auch die Sichtung und Bewertung der Einzelergebnisse sehr aufwendig.

Rückwärts-Elimination und **Vorwärts-Einschluss**: Aus einem vollständigen Modell, welches alle zu berücksichtigenden Einflussgrößen einschließt, werden bei der **Rückwärts-Elimination** nacheinander die Einflussgrößen herausgenommen, die keinen signifikanten ($H_0 : \beta_j = 0$) Einfluss auf die Zielgröße haben, die somit nur geringfügig zur Erhöhung des multiplen Bestimmtheitsmaßes beitragen. Diese Entscheidung kann auch mit einer speziellen F-Statistik begründet werden: Betrachtet wird das Modell mit p Einflussgrößen. Ausgeschlossen werden soll die Variable mit dem kleinsten F-Wert nach:

$$\hat{F} = \frac{RSS_{(p-1)} - RSS_{(p)}}{RSS_{(p)} / (n - (p + 1))} < F_{out} \quad (8.32)$$

Die Verteilung dieser Statistik variiert mit der Anzahl der Beobachtungen und der Anzahl der im Modell berücksichtigten Einflussgrößen. In der Regel wird daher ein konstanter (konservativer) **Schwellenwert**, z.B. $F_{out} = 4$, gewählt, um den schrittweisen Ausschluss der Einflussgrößen an einer geeigneten Stelle abzubrechen.

In **R** kann ein vergleichbarer Prozess in einzelnen Schritten mit der Funktion **update()** bearbeitet werden. Ausgehend von einem vollständigen Modell

```
fit <- lm(y ~ ., data)
```

mit allen Einflussgrößen wird im ersten Schritt die Variable (z.B. A) entfernt, deren Koeffizient den größten P-Wert nach der t-Statistik aufweist und der größer als ein fest vorgegebener Wert α (z.B. 0,05) ist.

```
fit.neu <- update(fit, . ~ . - A)
```

Mit dem neuen Modell verfährt man entsprechend solange, bis alle P-Werte kleiner als α sind.

Das Verfahren der Rückwärts-Elimination steht in **R** mit der Funktion **drop1()** zur Verfügung. Ausgehend von einem aktuellen (vollständigen) Modell wird für jede einzelne Modellkomponente der Verlust in der Güte der Anpassung ohne diese Komponente berechnet. Neben den Einflussgrößen können auch komplexere Komponenten (z.B. Wechselwirkungen) in dem Prozess der Modellbildung verwendet werden. Die Entscheidung über den Verbleib einzelner Komponenten wird anhand der F-Statistik (8.32) getroffen.

```
> library(DAAG)
> data(litters)
> fit <- lm(lsize ~ ., data=litters)
> drop1(fit, test="F")
Single term deletions

Model:
lsize ~ bodywt + brainwt
Df Sum of Sq  RSS    AIC  F value    Pr(F)
<none>                 11.482   -5.100
bodywt  1      89.791 101.272  36.442 132.9463 1.849e-09 ***
brainwt 1       3.082  14.563  -2.345  4.5628  0.04751 *
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Nach dem Kriterium der F-Statistik kann in dem Beispiel keine der beiden Einflussgrößen aus dem Modell ausgeschlossen werden. Ergänzend wird hier auch ein Wert für RSS und AIC (s.u.) berechnet. Ein Vergleich dieser Werte ermöglicht eine Einschätzung der Bedeutung des Körpergewichtes gegenüber dem Gehirngewicht im Rahmen der Modellbildung.

Bei dem **Vorwärts-Einschluss** werden zu einem Ausgangsmodell, z.B. nur mit der Konstanten β_0 (Nullmodell), nacheinander die Einflussgrößen einbezogen, die einen signifikanten ($H_0 : \beta_j = 0$) Einfluss auf die Zielgröße haben, die somit einen wesentlichen Beitrag zur Verbesserung des Modells, d.h. der Erhöhung des multiplen Bestimmtheitsmaßes, leisten. Auch hier kann die jeweilige Entscheidung mit einer speziellen F-Statistik begründet werden. Eingeschlossen werden soll die Variable mit dem größten F-Wert nach:

$$\hat{F} = \frac{RSS_{(p)} - RSS_{(p+1)}}{RSS_{(p+1)}/(n - (p + 2))} > F_{in} \quad (8.33)$$

Auch hier wird ein konstanter **Schwellenwert**, z.B. $F_{in} = 4$, gewählt, mit dem der schrittweise Einschluss der Einflussgrößen an einer geeigneten Stelle beendet werden kann. In **R** kann das Verfahren des Vorwärts-Einschlusses mit der Funktion **add1()** bearbeitet werden.

Das AIC Kriterium: Die Suche nach einem optimalen Modell (unter insgesamt 2^p möglichen Modellen) kann in **R** auch mit Hilfe des *AIC* Kriteriums (Akaike: an information criterion) erfolgen. Akaike [Aka73] definiert allgemein eine Maßzahl für die „Distanz“ zwischen einem unbekannten (wahren) Mechanismus, der die beobachteten Daten erzeugt haben könnte, und einem den Daten angepassten Modell.

$$\begin{aligned} AIC &= -2 \log(\text{likelihood}) + 2K \\ &= -2 \log(P(\text{Daten}|\hat{\beta})) + 2K \end{aligned} \quad (8.34)$$

Darin bezeichnet K die Anzahl der in dem Modell zu schätzenden Parameter ($K = p + 1$ incl. der Konstanten im linearen Modellansatz), d.h. es erfolgt eine „Bestrafung“ für zu viele Einflussgrößen in einem Modell.

Gesucht ist somit eine adäquate Beschreibung der beobachteten Daten durch ein Modell mit möglichst wenigen Parametern. Wenige Parameter erhöhen das Risiko, wichtige Effekte oder Zusammenhänge zu übergehen (**underfit**). Zu viele Parameter führen zu Pseudoeffekten oder Artefakten (**overfit**). Das *AIC*-Kriterium bietet einerseits eine Balance zwischen diesen beiden Fehlermöglichkeiten in der Modellbildung [BA02], andererseits ist dieses empirische Maß ohne theoretische Rechtfertigung. In der Anwendung hat sich das *AIC*-Kriterium häufig als praktikabel erwiesen.

Speziell für den Fall einer Modellanpassung nach der Methode der kleinsten Fehlerquadrate, unter der Annahme identisch normalverteilter Fehler mit konstanter Varianz (s.o.), kann der Wert für das *AIC* direkt angegeben werden.

$$\begin{aligned} AIC &= n \log(\hat{\sigma}^2) + 2(p + 1) \\ \text{mit } \hat{\sigma}^2 &= \frac{\hat{\epsilon}'\hat{\epsilon}}{n - p - 1} = \frac{RSS}{n - p - 1} \end{aligned} \quad (8.35)$$

Der Wert für *AIC* kann somit in linearen Modellen leicht berechnet werden. Dabei ist auf die korrekte Festlegung von K , z.B. mit oder ohne konstanten Term im Regressionsmodell, zu achten. **Ziel der Modellbildung ist es, einen möglichst kleinen Wert für das *AIC* zu erhalten.**

Bei kleiner Fallzahl und vergleichbar großer Anzahl von Parametern ($n/K < 40$) sollte der Wert für das *AIC* nach (8.36) korrigiert werden (empirische Korrektur).

$$AIC_c = AIC + 2 \frac{K(K + 1)}{n - K - 1} \quad (8.36)$$

In **R** kann dieser Prozess einer Modellbildung mit Hilfe der Funktion **step()** erfolgen. Ausgehend von dem vollständigen Modell wird in einer Rückwärtselimination versucht, ein Modell mit kleinerem *AIC* und weniger Parametern zu finden. Für die vorliegenden Daten zeigt das vollständige Modell optimale Eigenschaften hinsichtlich *RSS* bzw. *AIC*.

```
> data(litters)
> fit <- lm(lsize ~ ., data=litters)
> step(fit)
Start: AIC= -5.1
      lsize ~ bodywt + brainwt

            Df Sum of Sq      RSS       AIC
<none>                 11.482    -5.100
- brainwt    1         3.082    14.563    -2.345
- bodywt     1        89.791   101.272    36.442

Call:
lm(formula = lsize ~ bodywt + brainwt, data = litters)

Coefficients:
(Intercept)      bodywt      brainwt
      12.899       -2.398       31.628
```



Hinweis: Der schrittweise Modellierungsprozess (stepwise regression modelling) des Ein- und Ausschließens einzelner Einflussgrößen führt nicht zwangsläufig zu einem optimalen Modell! Der Ausschluss weniger signifikanter Größen führt einerseits zu einer **Überschätzung** des Einflusses der verbleibenden Einflussgrößen (Prädiktoren). Andererseits können auch die ausgeschlossenen Größen sehr wohl einen hohen Erklärungswert für die Zielgröße haben (hohe Korrelation); sie liefern lediglich neben den bereits in dem Modell erfassten Größen keinen zusätzlichen Beitrag für eine verbesserte Modellanpassung. Der Ausgang diese Modellierung ist somit nicht eindeutig und muss insbesondere **sachlogisch** (inhaltlich) sehr kritisch bewertet werden.

8.2.4 Nominalskalierte Einflussgrößen

Das Modell der linearen Regression setzt messbare Einflussgrößen (mindestens intervallskaliert) voraus. Die Berücksichtigung qualitativer Einflussgrößen (nominal skaliert) ist unter dem Aspekt der **Varianz-** bzw. **Kovarianzanalyse** möglich (vgl. folgenden Abschnitt). Dazu kann die qualitative Einflussgröße durch die sogenannte **Dummy-Codierung** [8.3.1.2] in mehrere zweiwertige Indikator- oder Dummy-Variablen (*D*) abgebildet werden. Für dichotome Einflussgrößen, wie zum Beispiel das Geschlecht, ist diese Abbildung naheliegend durch die Werte 0 (männlich) und 1 (weiblich) möglich.

Kategorielle Größen, zum Beispiel die Blutgruppe mit den Werten A, B, AB und 0, müssen dagegen in mehrere Dummy-Variablen abgebildet werden.

Blutgruppe (x_j)	D_{j1}	D_{j2}	D_{j3}
A	0	0	0
B	1	0	0
AB	0	1	0
0	0	0	1

Für jede der so definierten Variablen D_{ji} muss im linearen Modell ein eigener Regressionskoeffizient β_{ji} geschätzt werden, der den Einfluss der zugehörigen Kategorie auf die untersuchte Zielgröße erfasst.

$$Y = \beta_0 + \beta_1 x_1 + \dots + \underbrace{\beta_{j1} D_{j1} + \beta_{j2} D_{j2} + \beta_{j3} D_{j3}} + \dots + \beta_p x_p + \epsilon$$

Hinweis: Die Blutgruppe A wird im vorliegenden Beispiel nur indirekt durch **nicht B und nicht AB und nicht 0** im Modell erfasst.

8.3 Varianzanalyse im linearen Modell

- Einfaktorielle Analyse
- Zweifaktorielle Analyse

8.3.1 Einfaktorielle Varianzanalyse

In faktoriellen Versuchen wird der Einfluss von Faktoren auf eine Zielgröße in definierten, festen Stufen (Versuchbedingungen) untersucht.

Beispiel: Die Untersuchung der Größe des Hemmhofes (in mm) beim Bakterienwachstum im Vergleich von 3 unterschiedlichen Antibiotika führt zum Beispiel zu folgenden Daten:



A	13.2	14.1	7.8	11.7	
B	15.9	16.2	19.3	18.0	17.3
C	6.8	9.2	12.4		

Die Frage, wie die Evidenz für den Unterschied in der Wirksamkeit zwischen den Antibiotika statistisch bewertet werden kann, lässt sich durch ein lineares Modell formulieren:

$$\begin{aligned}
 Y_{ij} &\sim N(\mu_i, \sigma^2) \\
 \mu_i &= \mu + \alpha_i \\
 Y_{ij} &= \mu + \alpha_i + \epsilon_{ij} \\
 \epsilon_{ij} &\sim N(0, \sigma^2)
 \end{aligned} \tag{8.37}$$

mit $i = 1, \dots, k$ (k : Anzahl der Stufen eines Faktors) und $j = 1, \dots, n_i$ (n_i : Anzahl der Beobachtungen je Faktorstufe). Der Parameter μ bezeichnet in diesem Modell den Erwartungswert der Zielgröße (über alle Stufen) und die $\alpha_i = \mu_i - \mu$ kennzeichnen die Einflüsse (**Effekte**) jeder Faktorstufe. Das Modell kann auch über eine Matrix X (**Design-Matrix**) in Matrixschreibweise $Y = X\beta + \epsilon$ beschrieben werden, mit

$$X = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \quad \beta = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} .$$

Allerdings ist das Modell in der vorliegenden Form überparametrisiert, d.h. X hat den Rang k anstatt $k+1$ (es gibt mehr Parameter als Freiheitsgrade). Für eine Lösung nach $\hat{\beta}$ müssen demnach Beschränkungen (Randbedingungen) eingeführt werden, so dass nur noch k Parameter vorliegen und X den Rang k hat. Dann kann ein kleinster Quadrate Schätzer für β nach (8.9) bzw. (8.19) ermittelt werden. Drei Ansätze werden im Folgenden kurz vorgestellt.

8.3.1.1 Erwartungswert-Parametrisierung

Die Annahme $\mu = 0$ führt dazu, dass in (8.37) die Faktoreffekte α_i identisch mit den Erwartungswerten μ_i sind, also ohne einen gemeinsamen Bezug geschätzt werden können.

$$X = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \beta = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}$$

In **R** kann dieses Modell mit der Funktion **lm()** berechnet werden. Dazu werden die Antibiotika in einem Vektor vom Typ Faktor (Antibiotika) und die zugehörigen Messwerte in einem Vektor (Wert) gespeichert. Das Programm erzeugt anhand der Modellgleichung, in der durch „-1“ ein gemeinsamer Erwartungswert ausgeschlossen wird, automatisch die entsprechende *X*-Matrix.

```
> Antibiotikum <- as.factor(c(rep("A",4), rep("B",5), rep("C",3))); Antibiotikum
[1] A A A A B B B B C C C
Levels: A B C
> Wert <- c(13.2,14.1,7.8,11.7,15.9,16.2,19.3,18.0,17.3,6.8,9.2,12.4); Wert
[1] 13.2 14.1 7.8 11.7 15.9 16.2 19.3 18.0 17.3 6.8 9.2 12.4
> fit <- lm(Wert ~ Antibiotikum - 1)
> summary(fit)
```

Call:

```
lm(formula = Wert ~ Antibiotikum - 1)
```

Residuals:

Min	1Q	Median	3Q	Max
-3.900	-1.215	-0.020	1.615	2.933

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
AntibiotikumA	11.700	1.138	10.277	2.85e-06 ***
AntibiotikumB	17.340	1.018	17.029	3.73e-08 ***
AntibiotikumC	9.467	1.315	7.201	5.08e-05 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.277 on 9 degrees of freedom

Multiple R-Squared: 0.9803, Adjusted R-squared: 0.9737

F-statistic: 149.2 on 3 and 9 DF, p-value: 5.445e-08

Die Koeffizienten, in dem Beispiel $\hat{\alpha}_1 = 11,7$, $\hat{\alpha}_2 = 17,3$ und $\hat{\alpha}_3 = 9,5$, sind identisch mit den Mittelwerten der Gruppen. Die Teststatistiken nach (8.14) beziehen sich somit auf einen Vergleich der entsprechenden Erwartungswerte mit 0 und sind für einen Vergleich der Gruppen untereinander nicht zu verwenden. Insbesondere die F-Statistik ist hier irreführend! Eine sinnvolle Interpretation dieser Statistiken ist nur möglich, wenn die Konstante (intercept) in das Modell mit aufgenommen wird.

8.3.1.2 Effekt-Parametrisierung: Dummy-Codierung

Die Annahme $\alpha_1 = 0$ führt dazu, dass der mittlere Effekt μ in der ersten Faktorstufe zusammengefasst wird und die Effekte der anderen Faktorstufen stets auf die 1. Stufe bezogen werden.

$$\mu_1 = \mu$$

$$\mu_2 = \mu + \alpha_2$$

$$\mu_3 = \mu + \alpha_3$$

Diese Parametrisierung entspricht der Einführung einer Design-Matrix, wie sie bereits in Abschnitt [8.2.4] kurz erläutert wurde.

$$X = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \quad \beta = \begin{bmatrix} \mu_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}$$

```
> summary(lm(Wert ~ Antibiotikum))
...

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    11.700      1.138   10.277 2.85e-06 ***
AntibiotikumB     5.640      1.527    3.693 0.00498 **
AntibiotikumC    -2.233      1.739   -1.284 0.23113

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.277 on 9 degrees of freedom
Multiple R-Squared: 0.7438,    Adjusted R-squared: 0.6869
F-statistic: 13.07 on 2 and 9 DF,  p-value: 0.002179
```

Neben der Schätzung für den Erwartungswert in der ersten Gruppe ($\hat{\mu}_1 = 11,7$, intercept) werden nach diesem Modell die Effekte gegen die erste Gruppe $\alpha_2 = 5,6 = \hat{\mu}_2 - \hat{\mu}_1$ und $\alpha_3 = -2,2 = \hat{\mu}_3 - \hat{\mu}_1$ geschätzt und es wird nach der Statistik (8.14) geprüft, ob diese von Null verschieden sind. Die F-Statistik $\hat{F} = 13,07$ entspricht bei dieser Parametrisierung dem Ergebnis einer einfachen Varianzanalyse, auf die im nächsten Abschnitt [8.3.1.4] näher eingegangen wird.

Die Codierung hinsichtlich der Effekte kann natürlich auch auf eine andere Faktorstufe, in dem Beispiel auch mit Bezug auf das Antibiotikum B oder C, bezogen werden. In **R** können die entsprechenden X -Matrizen mit der Funktion **contr.treatment()** erzeugt werden.

```
> contr.treatment(3, base = 1, contrasts = TRUE)
 2 3
1 0 0
2 1 0
3 0 1
```

8.3.1.3 Effekt-Parametrisierung: Effekt-Codierung

Die Annahme $\sum \alpha_i = 0$ führt dazu, dass die Effekte der einzelnen Faktoren gegenüber einem mittleren Erwartungswert unabhängig von der Anordnung der einzelnen Faktorstufen betrachtet werden können.

$$\gamma = \frac{1}{3}(\mu_1 + \mu_2 + \mu_3)$$

$$\alpha_i = \mu_i - \gamma \quad (i = 1, 2, 3)$$

$$0 = \alpha_1 + \alpha_2 + \alpha_3$$

Die Modellspezifikation zu dieser Parametrisierung ist

$$X = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \end{bmatrix} \quad \beta = \begin{bmatrix} \gamma \\ \alpha_1 \\ \alpha_2 \end{bmatrix}$$

und kann in **R** ebenfalls mit der Funktion **lm()** berechnet werden. Dazu ist eine geeignete Designmatrix X durch das zusätzliche Argument „contrasts=” auszuwählen.

```
> fit <- lm(Wert ~ Antibiotikum, contrasts = list(Antibiotikum="contr.sum"))
> summary(fit)
....
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   12.8356     0.6717   19.108 1.36e-08 ***
Antibiotikum1  -1.1356     0.9398   -1.208 0.257729
Antibiotikum2   4.5044     0.8927    5.046 0.000694 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
....
```

Aus den geschätzten Modellparametern lassen sich die entsprechenden Erwartungswerte in den Gruppen mit Bezug auf $\hat{\gamma} = 12,8$ (intercept) und $\hat{\alpha}_3 = -\hat{\alpha}_1 - \hat{\alpha}_2 = 1,14 - 4,5 = -3,37$ wie folgt ableiten:

$$\hat{\gamma} + \hat{\alpha}_1 = \hat{\mu}_1 = 11,70$$

$$\hat{\gamma} + \hat{\alpha}_2 = \hat{\mu}_2 = 17,34$$

$$\hat{\gamma} + \hat{\alpha}_3 = \hat{\mu}_3 = 9,47$$

Die in diesem Modellansatz verwendete Design-Matrix kann in **R** auch mit der Funktion **contr.sum()** angezeigt werden.

```
> contr.sum(3, contrasts = TRUE)
[,1] [,2]
1    1    0
2    0    1
3   -1   -1
```

8.3.1.4 Varianzkomponenten - ANOVA

Das lineare Modell ermöglicht eine eindeutige Varianzzerlegung (ANOVA). Insbesondere kann ohne Beschränkung der Allgemeingültigkeit für den Fall einer Dummy-Codierung oder Effekt-Codierung eine OLS-Schätzung der Modellparameter hergeleitet und ein genereller Faktoreffekt anhand einer F-Statistik geprüft werden. Dieses Verfahren ist inhaltlich konsistent und gebräuchlicher als die in den vorangehenden Abschnitten dargestellten Parametrisierungen. Die algebraische Schreibweise soll zudem besonders auf die Äquivalenz zu den im Abschnitt [7.5.2] zur Varianzanalyse dargestellten Ansätzen hinweisen!

$$S(\mu, \alpha_1, \dots, \alpha_n) = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \mu - \alpha_i)^2 \rightarrow \min \quad (8.38)$$

führt zu den Schätzungen: $\hat{\mu} = \bar{y}_{..}$ und $\hat{\alpha}_i = \bar{y}_{i.} - \bar{y}_{..}$.

Die Ergebnisse werden nach Tabelle 8.1 (ANOVA) zusammengefasst

Die Nullhypothese $H_0: \alpha_1 = \alpha_2 = \dots = \alpha_k = 0$ wird mit der Fisher-Verteilung geprüft.

$$\hat{F} = \frac{MS_F}{MS_E} \sim F_{k-1, n-k, \alpha} \quad (8.39)$$

Die Lösung zu den Beispieldaten in **R** mit den Funktionen **anova(lm())** ist:

```
> anova(lm(Wert ~ Antibiotikum))
Analysis of Variance Table

Response: Wert
      Df Sum Sq Mean Sq F value    Pr(>F)    
Antibiotikum  2  135.490   67.745   13.067 0.002179 **
Residuals    9   46.659    5.184
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Tabelle 8.1 ANOVA-Tabelle zum linearen Modell mit einem Faktor

Quelle	SS	FG	MS
Faktor	$SS_F = \sum_i n_i \cdot (\bar{y}_{i.} - \bar{y}_{..})^2$	$k - 1$	$SS_F / (k - 1)$
Fehler	$SS_E = \sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2$	$n - k$	$SS_E / (n - k)$
gesamt	$SS_T = \sum_i \sum_j (y_{ij} - \bar{y}_{..})^2$	$n - 1$	

Auf der Grundlage der vorangehenden Modellüberlegungen lassen sich **multiple paarweise Vergleiche** der einzelnen Gruppen untereinander in **R** sehr anschaulich mit den Funktionen **simtest()** und **simint()** aus dem Paket **library(multcomp)** durchführen (Bretz, Hothorn und Westfall [BHW04]).

```
> library(multcomp)
> summary(simtest(Wert ~ Antibiotikum, type=c("Tukey")))
```

Simultaneous tests: Tukey **contrasts**

Call:
simtest.formula(formula = Wert ~ Antibiotikum, type = c("Tukey"))

Tukey **contrasts for factor** Antibiotikum

Contrast **matrix:**

	AntibiotikumA	AntibiotikumB	AntibiotikumC
AntibiotikumB-AntibiotikumA	0	-1	1
AntibiotikumC-AntibiotikumA	0	-1	0
AntibiotikumC-AntibiotikumB	0	0	-1

Absolute Error Tolerance: 0.001

Coefficients:

	Estimate	t value	Std. Err.	p raw	p Bonf	p adj
AntibiotikumC-AntibiotikumB	-7.873	-4.735	1.527	0.001	0.003	0.003
AntibiotikumB-AntibiotikumA	5.640	-3.693	1.739	0.005	0.010	0.009
AntibiotikumC-AntibiotikumA	-2.233	-1.284	1.663	0.231	0.231	0.231

Für den Vergleich der Gruppen untereinander (all pairwise) nach Tukey (vgl. auch [7.5.3.4]) bestimmt das Programm die verwendete Kontrastmatrix und berechnet die Schätzung bzw. statistische Prüfung der paarweisen Effekte. Anschaulicher ist in jedem Fall die Berechnung und graphische Darstellung entsprechender Konfidenzintervalle. In Abbildung 8.5 sind die simultanen 95%-Konfidenzintervalle nach Tukey für die Beispieldaten dargestellt (C-A ohne Effekt).

8.3.2 Zweifaktorielle Varianzanalyse

Das Beispiel aus dem vorigen Abschnitt kann auf zwei Faktoren erweitert werden. Neben der Art des Antibiotikums ($k = 3$) sollen zusätzliche zwei unterschiedliche Konzentrationen ($l = 2$) untersucht werden.

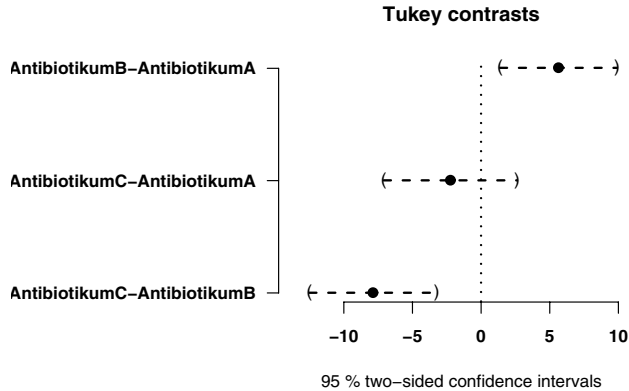


Abb. 8.5 Simultane Konfidenzintervalle nach Tukey zu den Antibiotikadaten: nur C-A ist auf dem 5%-Niveau ohne Effekt



Beispiel: Der Hemmhofdurchmesser soll für 3 verschiedene Antibiotika (A, B, und C) mit jeweils 2 unterschiedlichen Konzentrationen (hoch und niedrig) verglichen werden.

hoch	A	13.2	14.1	7.8	11.7
	B	15.9	16.2	19.3	18.0
	C	6.8	9.2	12.4	
niedrig	A	10.4	12.6	6.3	
	B	11.5	13.7	10.9	15.1
	C	12.3	14.5	16.7	10.3

Dieser Versuchsaufbau allgemein kann durch das folgende lineare Modell beschrieben werden:

$$y_{iju} = \mu + \alpha_i + \delta_j + \epsilon_{iju} \quad (8.40)$$

mit $i = 1, \dots, k$ (k : Anzahl der Stufen des ersten Faktors) und $j = 1, \dots, l$ (l : Anzahl der Stufen des zweiten Faktors) und $u = 1, \dots, n_{ij}$ (n_{ij} : Anzahl der Beobachtungen zu der entsprechenden Faktorstufenkombination).

Die vollständige, wiederum überparametrisierte, Designmatrix für das Modell zu dem Beispiel der Antibiotika hat die Form

$$X = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 \end{bmatrix} \quad \beta = \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \delta_1 \\ \delta_2 \end{bmatrix}$$

und muss durch eine geeignete Parametrisierung, z.B. eine Effektparametrisierung (vgl. [8.3.1.3]) mit $\sum_i \alpha_i = 0$ und $\sum_j \delta_j = 0$ beschränkt werden. Die Designmatrix

$$X = \begin{bmatrix} 1 & 1 & 0 & -1 \\ 1 & 0 & 1 & -1 \\ 1 & -1 & -1 & -1 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$$

ermöglicht eine OLS-Schätzung der Parameter, die in **R** mit der Funktion **lm()** berechnet werden kann.

```
> fit <- lm(Wert ~ Antibiotikum + Konz,
  contrasts=list(Antibiotikum="contr.sum", Konz="contr.sum"))

> summary(fit)
...

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  12.6240     0.6344  19.900 3.49e-14 ***
Antibiotikum1 -1.8356     0.9167  -2.002  0.05973 .
Antibiotikum2  2.6336     0.8612   3.058  0.00647 **
Konzl        -0.5817     0.6350  -0.916  0.37109

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

...
```

Mit den aus dem Modell geschätzten Werten für einen gemeinsamen Erwartungswert (intercept) und den gruppenspezifischen Effekten $\hat{\alpha}_i$ bzw. $\hat{\delta}_j$ können Schätzungen für die einzelnen Erwartungswerte direkt bestimmt werden:

$$\begin{aligned}\hat{\eta} &= 1/5(\hat{\mu}_{1.} + \hat{\mu}_{2.} + \hat{\mu}_{3.} + \hat{\mu}_{.1} + \hat{\mu}_{.2}) = 12,624 \\ \hat{\mu}_{1.} &= \hat{\eta} + \hat{\alpha}_1 = 10,79 \\ \hat{\mu}_{2.} &= \hat{\eta} + \hat{\alpha}_2 = 15,26 \\ \hat{\mu}_{3.} &= \hat{\eta} + \hat{\alpha}_3 = 11,83 \\ \hat{\mu}_{.1} &= \hat{\eta} + \hat{\delta}_1 = 12,04 \\ \hat{\mu}_{.2} &= \hat{\eta} + \hat{\delta}_2 = 13,20\end{aligned}$$

Die Analyse der Daten in einer Zerlegung nach Varianzkomponenten (ANOVA) für die Situation zweier Faktoren ist in der folgenden Tabelle dargestellt. Der Einfluss beider Faktoren wird jeweils mit einem F-Test nach (8.39) geprüft.

Tabelle 8.2 Varianzkomponenten zum linearen Modell mit zwei Faktoren (ANOVA-Tabelle)

Quelle	SS	FG
Faktor 1	$\sum_i n_{i.} (\bar{y}_{i.} - \bar{y}_{...})^2$	$k - 1$
Faktor 2	$\sum_j n_{.j} (\bar{y}_{.j} - \bar{y}_{...})^2$	$l - 1$
Fehler	$\sum_i \sum_j \sum_u (y_{iju} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{...})^2$	$n - k - l + 1$
gesamt	$\sum_i \sum_j \sum_u (y_{iju} - \bar{y}_{...})^2$	$n - 1$

In **R** kann dieses Schema mit der Funktion **anova()** zu dem aktuellen Modell berechnet werden.

```
> anova(fit)
Analysis of Variance Table

Response: Wert
      Df Sum Sq Mean Sq F value Pr(>F)
Antibiotikum  2  90.972   45.486   4.9905 0.01812 *
Konz          1   7.649    7.649   0.8393 0.37109
Residuals    19 173.178    9.115
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Das Ergebnis weist nur auf einen signifikanten Effekt hinsichtlich der untersuchten Antibiotika hin ($P = 0,018$), während die unterschiedliche Konzentration die Größe des Hemmhofes nicht signifikant verändert ($P = 0,371$).



Grundsätzlich ist jedoch nicht davon auszugehen, dass die beiden Faktoren unabhängig voneinander auf die Größe des Hemmhofes wirken. Einen Einblick in mögliche **Wechselwirkungen (Interaktionen)** zwischen den Faktoren erhält man, indem die Mittelwerte der Zielgröße getrennt für die einzelnen Faktorkombinationen graphisch dargestellt werden (Profildiagramm (interaction plot) in Abbildung 8.6). Ein paralleler Verlauf der Linien wäre ein Hinweis darauf, dass keine Wechselwirkungen vorliegen. Für die Daten aus unserem Beispiel muss das lineare Modell offensichtlich durch einen zusätzlichen **Interaktionsterm** erweitert werden.

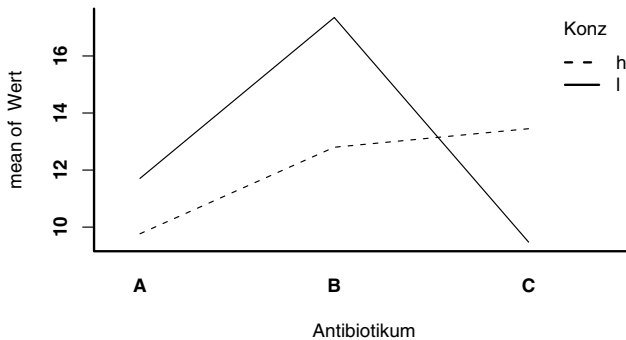


Abb. 8.6 Interaction-Plot für die Größe des Hemmhofes der drei Antibiotika und der Konzentration (h-high, l-low)

Zur Klärung des Begriffes einer Interaktion sollen zwei zweistufige Faktoren betrachtet werden, die jeweils mit den Werten 0 und 1 codiert sind. Die Berücksichtigung einer Wechselwirkung in der Modellbildung wird für diesen Fall in Tabelle 8.3 gezeigt (nach Caliebe [CFK05]).

Ohne Berücksichtigung einer Interaktion verhalten sich die Erwartungswerte einer Zielgröße additiv und sind für die vier möglichen Kombinationen durch drei Parameter β_0 , β_1 und β_2 in dem Modell vollständig erklärt. Liegt eine Interaktion zwischen den beiden Faktoren x_1 und x_2 vor, dann muss ein vierter Parameter γ_{12} in das Modell aufgenommen werden.

$$Y_{iju} = \mu + \alpha_i + \delta_j + \gamma_{ij} + \epsilon_{iju} \quad (8.41)$$

Ein Interaktionseffekt kann durch einen direkten Vergleich mit dem Modell (8.40) über die Residuen statistisch geprüft werden. In **R** ist dies mit Hilfe der **update()** Funktion möglich, wobei in der Modellspezifikation ein Term „Antibiotikum:Konz“ für die Interaktion zusätzlich mit aufgenommen wird (Näheres zur Spezifikation der Modellgleichung enthält Tabelle 9.7 in Kapitel [9]).

Tabelle 8.3 Erwartungswerte einer Zielgröße Y in einer zweifaktoriellen Varianzanalyse mit jeweils zweistufigen Einflussgrößen: x_1 und x_2

	ohne Interaktion		mit Interaktion	
	$E[Y] = \beta_0 + \beta_1 x_1 + \beta_2 x_2$		$E[Y] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \gamma_{12} x_1 x_2$	
	$x_2 = 0$	$x_2 = 1$	$x_2 = 0$	$x_2 = 1$
$x_1 = 0$	β_0	$\beta_0 + \beta_2$	β_0	$\beta_0 + \beta_2$
$x_1 = 1$	$\beta_0 + \beta_1$	$\beta_0 + \beta_1 + \beta_2$	$\beta_0 + \beta_1$	$\beta_0 + \beta_1 + \beta_2 + \gamma_{12}$

```
> fit1 <- update(fit, . ~ . + Antibiotikum:Konz)
> anova(fit, fit1)
Analysis of Variance Table
```

```
Model 1: Wert ~ Antibiotikum + Konz
Model 2: Wert ~ Antibiotikum + Konz + Antibiotikum:Konz
Res.Df    RSS Df Sum of Sq  F    Pr(>F)
1      19 173.178 2      17 101.415 2    71.762 6.0147 0.01059 *
...
```

Eine vollständige ANOVA-Tabelle für das Modell einer zweifaktoriellen Varianzanalyse mit Berücksichtigung der Wechselwirkung für die Beispieldaten aus diesem Abschnitt wird in **R** mit der Funktion **lm()** wie folgt berechnet

```
> fit <- lm(Wert ~ Antibiotikum + Konz + Antibiotikum:Konz)
> anova(fit)
Analysis of Variance Table
```

```
Response: Wert
          Df Sum Sq Mean Sq F value    Pr(>F)
Antibiotikum  2  90.972   45.486    7.6247 0.004329 **
Konz          1   7.649    7.649    1.2823 0.273200
Antibiotikum:Konz  2  71.762   35.881    6.0147 0.010585 *
Residuals    17 101.415    5.966
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Im Vergleich mit der vorangehenden Analyse der Daten ohne Berücksichtigung einer Wechselwirkung zeigt sich, dass sich die Wirksamkeit des Antibiotikums C bei unterschiedlichen Konzentrationen im Vergleich zu den Antibiotika A und B „gegensinnig“ verändert. Für den Fall einer solchen Interaktion dürfen die Haupteffekte, hier Konzentration und Antibiotikum, nicht unabhängig voneinander bewertet werden.

8.4 Logistische Regression

- Hypothesentest im logistischen Regressionsmodell
- Multiple logistische Regression
- Interpretation der Regressionskoeffizienten
- Variablenauswahl im Rahmen der Modellbildung
- Residuenanalyse

Die Regressionsmodelle aus dem vorangehenden Abschnitt [8.2] gehen davon aus, dass die Zielgröße Y intervallskaliert, quantitativ messbar ist und mit dem Modell einer Normalverteilung $Y \sim N(\mu, \sigma^2)$ beschrieben werden kann. Die **Verteilung einer dichotomen (zweiwertigen,**

binären) Zielgröße Y ($n = n_0 + n_1$, mit n_0 Anzahl von Misserfolgen und n_1 Anzahl von Erfolgen; Erfolg: $y_i = 1$ und Misserfolg: $y_i = 0$, $i = 1, \dots, n$) kann dagegen mit dem Modell der Binomialverteilung beschrieben werden. Gegenstand der Modellierung ist in diesem Fall die unbekannte Erfolgswahrscheinlichkeit π , die von verschiedenen Faktoren (unabhängige Variablen) abhängen kann. So schließt man zum Beispiel aus der Erfahrung, dass es einen Zusammenhang zwischen der Wahrscheinlichkeit für das Auftreten einer bestimmten Erkrankung ($P(Y = 1) = \pi$) und dem Alter X gibt.

Ein Modellansatz nach $\pi = \beta_0 + \beta_1 x$ führt zu dem Problem, dass die Wahrscheinlichkeit nur zwischen 0 und 1 liegen kann ($0 \leq \pi \leq 1$). Mit Hilfe der „**logistischen**“ **Transformation** (8.43) können die Funktionswerte auf diesen Bereich begrenzt werden.

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \quad (8.42)$$

Hinweis: Die Wahl dieses speziellen Ansatzes wird durch die Eigenschaften der logistischen Funktion (Abbildung 8.7) gerechtfertigt. In der einfachsten Form durch $y = 1/(1 + e^x)$ gegeben, wird die logistische Kurve nach unten durch einen Basiswert (0-Wert, Ausgangswert) und nach oben durch einen Maximalwert (**Sättigung**) begrenzt. Die Zunahme der untersuchten Größe, z.B. in Abhängigkeit von der Zeit (Wachstum) erfolgt zunächst exponentiell bis zu einem Wendepunkt, z.B. bei der Hälfte der erreichbaren Sättigung (Halbwertszeit). Anschließend verlangsamt sich das Wachstum, so dass der Wert der Sättigung nicht überstiegen werden kann.

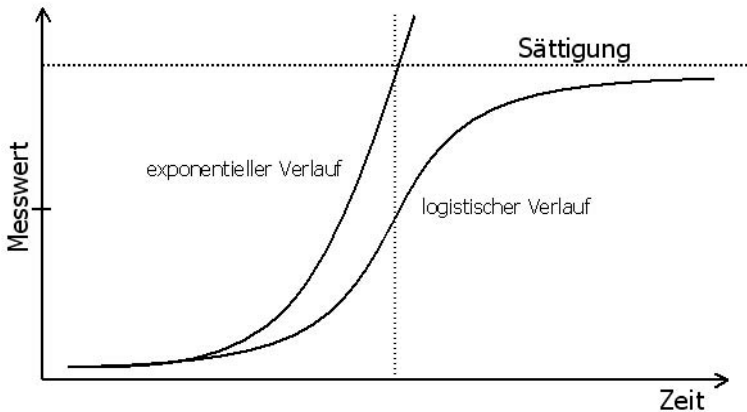


Abb. 8.7 Eigenschaften der logistischen Funktion

Für $\beta_1 > 0$ steigt die Funktion mit wachsendem x monoton an (Risiko zunehmend), ist $\beta_1 < 0$, dann fällt die Funktion mit wachsendem x monoton (Risiko abnehmend, Schutz oder Protektion). Ist $\beta_1 = 0$, dann hat X keinen Einfluss auf das Auftreten des Ereignisses Y . Die „**logit**“-**Transformation** nach (8.43), die hier als **Linkfunktion** verwendet wird, führt zu einer linearen Darstellung des Modells.

$$\text{logit}(\pi(x)) = \log\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \log(\text{odds}(\pi)) = \beta_0 + \beta_1 x \quad (8.43)$$

Diese Art der Transformation verdeutlicht einerseits den engen Zusammenhang mit dem Begriff der Chancen (odds) in dem Modell, andererseits können damit geeignete Schätzungen für die Parameter β_0 und β_1 nach der Maximum-Likelihood-Methode einfach hergeleitet werden.



Beispiel: Am 28.1.1986 explodierte die Raumfähre Challenger beim Start. Der Grund lag in einer Materialermüdung von Dichtungsringen an den Triebwerken. Ein Zusammenhang zwischen dem Versagen der Dichtungsringe und niedriger Aussentemperaturen liegt auf der Hand. Aus vorangehenden Starts lagen folgende Daten (Tabelle 8.4) vor [SFH89].

Tabelle 8.4 Temperatur (°F) und Ausfall (0-nein, 1-ja) von Dichtungsringen in den Triebwerken beim Start der Raumfähre Challenger

Start	Temperatur (°F)	Ausfall	Start	Temperatur (°F)	Ausfall
1	66	0	13	67	0
2	67	0	14	67	0
3	68	0	15	69	0
4	70	0	16	70	0
5	72	0	17	73	0
6	75	0	18	76	0
7	76	0	19	78	0
8	79	0	20	81	0
9	53	1	21	57	1
10	58	1	22	63	1
11	70	1	23	70	1
12	75	1			

Die Box-Plot Darstellung dieser Daten in Abbildung 8.8 (linke Seite) macht den Zusammenhang besonders deutlich. Die Wahrscheinlichkeit für eine Fehlfunktion ($P(Y = 1)$) in Abhängigkeit von der Temperatur (X) wird in dem Ansatz einer logistischen Regression in (8.44) modelliert.

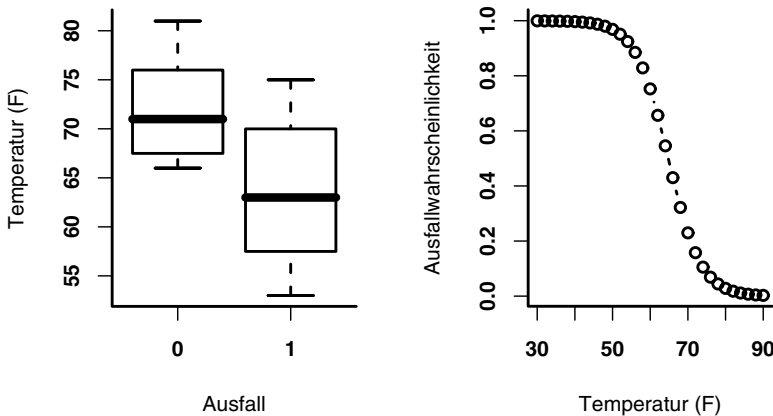


Abb. 8.8 Außentemperatur (°F) und das Versagen der Dichtungsringe beim Unglück der Challenger, links Box-Plot-Darstellung, rechts die Kurve zur logistischen Funktion aus der Modellrechnung nach (8.44)

Die Wahrscheinlichkeit für den i ten Ausgang eines Experimentes, formal beschrieben durch $y_i = 1$ für einen Erfolg und $y_i = 0$ für einen Misserfolg, kann in dem logistischen Modell allgemein durch (8.44) beschrieben werden.

$$\begin{aligned}
P(y_i; \beta_0, \beta_1) &= [\pi(x_i)]^{y_i} [1 - \pi(x_i)]^{1-y_i} \\
&= \left[\frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \right]^{y_i} \left[\frac{1}{1 + e^{\beta_0 + \beta_1 x_i}} \right]^{1-y_i} \\
&= \frac{(e^{\beta_0 + \beta_1 x_i})^{y_i}}{1 + e^{\beta_0 + \beta_1 x_i}}
\end{aligned} \tag{8.44}$$

Die Likelihood-Funktion zu (8.44) ist

$$L(\beta_0, \beta_1; X) = \prod_{i=1}^n P(y_i; \beta_0, \beta_1) \tag{8.45}$$

bzw. nach der Transformation mit der Logarithmusfunktion (log-Likelihood)

$$\log(L) = \sum_{i=1}^n [y_i(\beta_0 + \beta_1 x_i) - \log(1 + e^{\beta_0 + \beta_1 x_i})] \tag{8.46}$$

Nach dem Prinzip der Maximum-Likelihood Schätzung werden die partiellen Ableitungen nach β_0 bzw. β_1 gebildet.

$$\begin{aligned}
\frac{\partial \log(L)}{\partial \beta_0} &= \sum_{i=1}^n y_i - \sum_{i=1}^n \frac{e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}} \\
\frac{\partial \log(L)}{\partial \beta_1} &= \sum_{i=1}^n x_i y_i - \sum_{i=1}^n \frac{x_i e^{\beta_0 + \beta_1 x_i}}{1 + e^{\beta_0 + \beta_1 x_i}}
\end{aligned}$$

Übersichtlicher ist die Darstellung in der Matrix-Notation. Diese ermöglicht insbesondere auch eine übersichtliche Behandlung mehrerer Einflussgrößen im Rahmen einer multiplen logistischen Regression. Mit

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

kann der ML-Ansatz auch durch (8.47) beschrieben werden.

$$\boxed{
\begin{aligned}
\frac{\partial \log(L)}{\partial \boldsymbol{\beta}} &= X'(\mathbf{y} - \boldsymbol{\pi}) \\
X'(\mathbf{y} - \hat{\boldsymbol{\pi}}) &= 0 \\
\hat{\boldsymbol{\beta}} &= (X'VX)^{-1} X'V\mathbf{z}
\end{aligned}
} \tag{8.47}$$

Darin ist $V = \text{diag}(\hat{\pi}_i(1 - \hat{\pi}_i))$ eine $n \times n$ Diagonalmatrix mit den Wahrscheinlichkeiten, die aus dem Modell zu schätzen sind, und $\mathbf{z}$ hat die Bedeutung von $\mathbf{y}$ im Rahmen einer iterativen Lösung des Gleichungssystems (iteratively reweighted least squares (IRLS), Fisher scoring). Das Verfahren konvergiert in der Regel nach wenigen Schritten. Die Herleitung einer geeigneten Ausgangslösung wird hier nicht näher beschrieben (vgl. Hosmer [HL89]).

$$\begin{aligned}\hat{\beta}^{(t+1)} &= \hat{\beta}^{(t)} + (X'VX)^{-1}X'(\mathbf{y} - \hat{\pi}^{(t)}) \\ \text{mit } \hat{\pi}_i^{(t)} &= \frac{1}{1 + e^{-x_i' \hat{\beta}^{(t)}}} \\ \text{und } V &= \text{diag}(\hat{\pi}_i^{(t)}(1 - \hat{\pi}_i^{(t)}))\end{aligned}\quad (8.48)$$

In **R** erfolgt die Schätzung durch die Funktion **glm()** - verallgemeinerte lineare Modelle (generalized linear models) - mit dem speziellen Parameter **family=binomial**, der den Typ der verwendeten Linkfunktion auswählt. Für die Daten zum Unglück der Challenger folgt:

```
> t <- c(66,67,68,70,72,75,76,79,53,58,70,75,67,67,69,70,73,76,78,81,57,63,70)
> d <- c( 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1)
>
> fit <- glm(d ~ t, family=binomial)
> summary(fit)
```

Call:

glm(formula = d ~ t, family = binomial)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.0611	-0.7613	-0.3783	0.4524	2.2175

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	15.0429	7.3786	2.039	0.0415 *
t	-0.2322	0.1082	-2.145	0.0320 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 28.267 on 22 degrees of freedom
Residual deviance: 20.315 on 21 degrees of freedom
AIC: 24.315

Number of Fisher Scoring iterations: 5

Das Ergebnis der Schätzung mit **R** ist $\hat{\beta}_0 = 15,04$ und $\hat{\beta}_1 = -0,232$. Die entsprechende logistische Funktion ist in Abbildung 8.8 auf der rechten Seite dargestellt. Für eine Aussentemperatur von 31°F, wie sie am Tag des Unglücks vorlag, ergibt sich aus dieser Modellrechnung eine Wahrscheinlichkeit von 99,96%, d.h. die Dichtungsringe mussten mit hoher Wahrscheinlichkeit ausfallen. Nähere Hinweise zu den in der Funktion **glm()** berechneten Hypothesentests sowie zur Bedeutung der Devianz werden im folgenden Abschnitt gegeben.

Neben der Schätzung der beiden Parameter werden auch die zugehörigen Standardfehler (*se*) berechnet, aus denen sich unter der Annahme einer asymptotischen Normalverteilung auch die entsprechenden $(1 - \alpha)100\%$ -Konfidenzintervalle herleiten lassen.

$$\hat{\beta}_i \pm z_{1-\alpha/2} se(\hat{\beta}_i) \quad \text{für } i = 0, 1 \quad (8.49)$$

8.4.1 Hypothesentest im logistischen Regressionsmodell

Ein Test für die Überprüfung der Nullhypothese $H_0 : \beta_i = 0$ ergibt sich aus der **Wald Statistik**.

$$\hat{W} = \frac{\hat{\beta}_i}{se(\hat{\beta}_i)} \quad \text{für } i = 0, 1 \quad (8.50)$$

$\hat{W}$ ist asymptotisch standardnormalverteilt mit $\hat{W} \sim N(0, 1)$ (oder auch im Quadrat asymptotisch χ^2 -verteilt mit einem Freiheitsgrad). Dieser Test wird standardmäßig in vielen Statistikprogrammen zu jedem geschätzten Parameter berechnet. Die Bezeichnungen zu dieser Statistik sind allerdings nicht einheitlich und die Interpretation im Rahmen der Modellbildung kann zu Problemen und Widersprüchen führen (s. in Hauck und Donner [HD77], Jennings [Jen86]).

Informativer ist ein Likelihood-Quotienten Test, der auf der Basis der Likelihood-Schätzung durchgeführt wird. Zum besseren Verständnis dieser Teststatistik ist der Bezug auf ein **gesättigtes Modell** M_{sat} (saturated) hilfreich, in dem genauso viele Parameter wie Beobachtungspaare auftreten. Das saturierte Modell ist das allgemeinste Modell und hat somit die höchste Likelihood.

$$D = -2 \log \left[\frac{\text{likelihood } M_{mod}}{\text{likelihood } M_{sat}} \right]$$

Mit der Likelihoodfunktion (8.46) folgt daraus

$$\hat{D} = -2 \sum_{i=1}^n \left[y_i \log \left(\frac{\hat{\pi}(x_i)}{y_i} \right) + (1 - y_i) \log \left(\frac{1 - \hat{\pi}(x_i)}{1 - y_i} \right) \right] \quad (8.51)$$

Der Ausdruck $\hat{D}$ in (8.51) wird **Devianz** genannt und ist ein zentrales Maß für die Bewertung der Güte von Schätzungen im linearen Modell. Der Begriff Devianz kommt aus der Soziologie und bezeichnet die Abweichung (frz. "dévier") von allgemeinen Normen und Wertvorstellungen. Im Zusammenhang mit der Modellbildung entspricht die Devianz der Summe der Abweichungsquadrate (*SAQ* bzw. *RSS*) bei linearen Regressionsmodellen. Die Devianz für das gesättigte Modell ist $\hat{D} = 0$.

Für den Signifikanztest eines einzelnen Parameters (z.B. $H_0 : \beta_1 = 0$ zu einer Einflussgröße x) kann die Differenz $\hat{G}$ der Devianzen herangezogen werden.

$$\begin{aligned} \hat{G} &= \hat{D}_{\text{ohne } x} - \hat{D}_{\text{mit } x} \\ &= 2 \left(\sum_{i=1}^n \left[y_i \log(\hat{\pi}_i) + (1 - y_i) \log(1 - \hat{\pi}_i) \right] - \right. \\ &\quad \left. \left[n_1 \log(n_1) + n_0 \log(n_0) - n \log(n) \right] \right) \\ &\text{mit } n_1 = \sum y_i \text{ und } n_0 = \sum (1 - y_i) \end{aligned} \quad (8.52)$$

Die Statistik $\hat{G}$ ist unter der Nullhypothese $H_0 : \beta_1 = 0$ wie χ^2 verteilt mit einem Freiheitsgrad und kann in **R** zu dem Modell aus obigem Beispiel mit der Funktion **anova()** wie folgt berechnet werden.

```
> anova(fit, test="Chi")
Analysis of Deviance Table

Model: binomial, link: logit

Response: d

Terms added sequentially (first to last)
```

	Df	Deviance	Resid.	Df	Resid.	Dev	P(> Chi)
NULL			22		28.2672		
t	1	7.9520	21		20.3152		0.0048

Die Devianz im „Nullmodell“ (gesamt Devianz in den Beobachtungsdaten) beträgt hier $D=28,27$. Die Devianz im Modell unter Beachtung der Temperatur ist $D=20,32$. Daraus ergibt sich für $G=7,95$, d.h. die Aussentemperatur ist eine signifikante Komponente für die Erklärung des Ausfalls der Dichtungsringe ($P = 0,005$).

Insbesondere lassen sich nach diesem Ansatz auch multiple logistische Modelle in einzelnen Variablen oder Variablenkombinationen im Rahmen der Modellbildung miteinander vergleichen.

8.4.2 Multiple logistische Regression

In einer multiplen logistischen Regression werden mehrere Variablen $X' = (X_1, X_2, \dots, X_p)$ in dem Modell als Einflussgrößen berücksichtigt. Für die abhängige Zielgröße Y gilt die bedingte Wahrscheinlichkeit

$$P(Y = 1|X = x) = \pi(x) \quad .$$

Als **Linkfunktion** wird die Logit-Transformation verwendet.

$$\begin{aligned} g(x) &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p \\ \pi(x) &= \frac{e^{g(x)}}{1 + e^{g(x)}} \quad 0 \leq \pi(x) \leq 1 \end{aligned} \quad (8.53)$$

Grundsätzlich werden in diesem Modell die Einflussgrößen als intervallskalierte, quantitativ erfassbare Merkmale angenommen. Auch dichotome Variablen können in die Modellgleichung aufgenommen werden, wenn diese 0/1-skaliert sind. Nominalskalierte Variablen mit mehr als 2 Kategorien (kategoriiell unterteilt) müssen dagegen mittels **Indikatorvariablen** (Designvariablen) in das Modell aufgenommen werden (siehe auch die Abschnitte [8.2.4] und [8.3.1.2]). Mit diesen können spezielle Kontraste in dem Modell formuliert werden. Hat eine nominalskalierte Variable X_j zum Beispiel k Kategorien, dann werden für diese $k - 1$ neue Variablen D_{ju} definiert, die jeweils dichotom (0/1-skaliert) sind. Für diese müssen auch die entsprechenden Parameter im Modell berücksichtigt und geschätzt werden.

$$g(x) = \beta_0 + \beta_1 x_1 + \dots + \sum_{u=1}^{k-1} \beta_{ju} D_{ju} + \dots + \beta_p x_p$$

Beispiel: Das Auftreten einer Kyphose, einer rückenwärts gerichteten Verkrümmung der Wirbelsäule, wird bei 81 Kindern nach einer Wirbelsäulenoperation untersucht. Als Einflussgrößen sollen das Alter (Age) in Monaten, die Anzahl der Wirbel (Number) und der Beginn des operierten Wirbelsäulenabschnittes (Start) untersucht werden [CH92]. Die Daten zu diesem Beispiel (Tabelle 8.5) sind auch in dem Paket **rpart** (Therneau und Atkinson [TB05]) unter dem Namen **data(kyphosis)** enthalten. Eine Übersicht geben die Box-Plots in Abbildung 8.9.



Die Modellbildung zu den Beispieldaten in **R** durch die Funktion **glm()** führt zu dem Ergebnis:

```
> library(rpart)
> attach(kyphosis)
> fit <- glm(Kyphosis ~ Age + Number + Start, family="binomial", data=kyphosis)
> summary(fit)
```

```
Call:
glm(formula = Kyphosis ~ Age + Number + Start, family = "binomial",
    data = kyphosis)
```

```
Deviance Residuals:
```

Tabelle 8.5 Beispieldaten zum Auftreten einer Kyphose (Wirbelsäulenverkrümmung) nach Operation an der Wirbelsäule bei 81 Kindern

Kyphosis	Age	Number	Start	Kyphosis	Age	Number	Start
absent	71	3	5	absent	35	3	13
absent	158	3	14	absent	143	9	3
present	128	4	5	absent	61	4	1
absent	2	5	1	absent	97	3	16
absent	1	4	15	present	139	3	10
absent	1	2	16	absent	136	4	15
absent	61	2	17	absent	131	5	13
absent	37	3	16	present	121	3	3
absent	113	2	16	absent	177	2	14
present	59	6	12	absent	68	5	10
present	82	5	14	absent	9	2	17
absent	148	3	16	present	139	10	6
absent	18	5	2	absent	2	2	17
absent	1	4	12	absent	140	4	15
absent	168	3	18	absent	72	5	15
absent	1	3	16	absent	2	3	13
absent	78	6	15	present	120	5	8
absent	175	5	13	absent	51	7	9
absent	80	5	16	absent	102	3	13
absent	27	4	9	present	130	4	1
absent	22	2	16	present	114	7	8
present	105	6	5	absent	81	4	1
present	96	3	12	absent	118	3	16
absent	131	2	3	absent	118	4	16
present	15	7	2	absent	17	4	10
absent	9	5	13	absent	195	2	17
absent	8	3	6	absent	159	4	13
absent	100	3	14	absent	18	4	11
absent	4	3	16	absent	15	5	16
absent	151	2	16	absent	158	5	14
absent	31	3	16	absent	127	4	12
absent	125	2	11	absent	87	4	16
absent	130	5	13	absent	206	4	10
absent	112	3	16	absent	11	3	15
absent	140	5	11	absent	178	4	15
absent	93	3	16	present	157	3	13
absent	1	3	9	absent	26	7	13
present	52	5	6	absent	120	2	13
absent	20	6	9	present	42	7	6
present	91	5	12	absent	36	4	13
present	73	5	1				

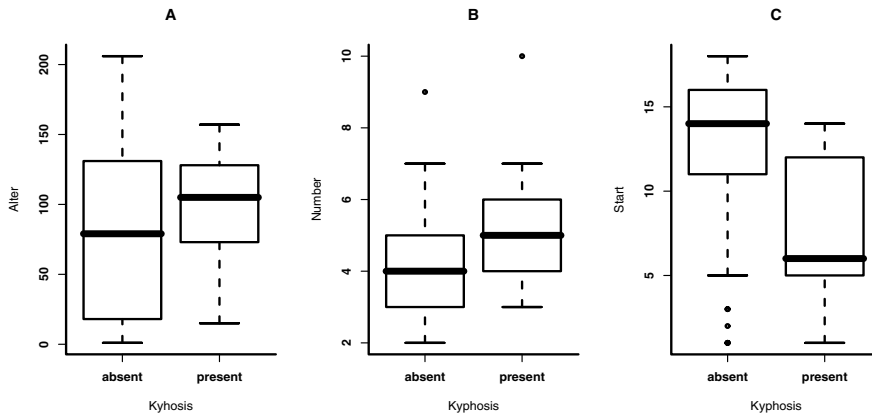


Abb. 8.9 Auftreten einer Kyphose (Wirbelsäulenverkrümmung) in Abhängigkeit vom Alter (A), der Zahl der operierten Wirbel (B) und der Lokalisation (C) bei 81 Kindern nach Operation an der Wirbelsäule

	Min	1Q	Median	3Q	Max
	-2.3124	-0.5484	-0.3632	-0.1659	2.1613

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-2.036934	1.449575	-1.405	0.15996
Age	0.010930	0.006446	1.696	0.08996 .
Number	0.410601	0.224861	1.826	0.06785 .
Start	-0.206510	0.067699	-3.050	0.00229 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 83.234 on 80 degrees of freedom
 Residual deviance: 61.380 on 77 degrees of freedom
 AIC: 69.38

Number of Fisher Scoring iterations: 5

Die Schätzung der Parameter $\hat{\beta}_0 = -2,037$, $\hat{\beta}_1 = 0,011$ (Alter), $\hat{\beta}_2 = 0,411$ (Anzahl) und $\hat{\beta}_3 = -0,207$ (Startlokalisierung) erfolgt nach dem im vorangehenden Abschnitt näher beschriebenen Verfahren. Die Parameterschätzungen lassen sich nach der Wald-Statistik prüfen. Das Ergebnis zeigt (vgl. Abbildung 8.9) nur die Lokalisierung (Start) als eine signifikante Einflussgröße ($P < 0,05$) in dem Modell an. Entsprechend der Zielsetzung, in dem Prozess der Modellbildung die beste mögliche Erklärung für die Zielgröße (mit möglichst wenigen Variablen) zu finden, hat die Bewertung der Koeffizienten mit Hilfe der **Devianz-Differenzen** (G-Statistik, vgl. (8.52)) mehr Aussagekraft. Mit der **anova()** Funktion können in **R** die zugehörigen Devianzanteile bewertet werden. Insbesondere lassen sich auch Modelle mit unterschiedlichen Variablenkombinationen hinsichtlich der Güte ihrer Anpassung direkt miteinander vergleichen.

```
> anova(fit, test="Chi")
Analysis of Deviance Table

Model: binomial, link: logit

Response: Kyphosis

Terms added sequentially (first to last)

      Df Deviance Resid. Df Resid. Dev P(>|Chi|)
NULL                80      83.234
```

```
Age      1      1.302      79      81.932      0.254
Number   1     10.306      78      71.627      0.001
Start    1     10.247      77      61.380      0.001
> fit1 <- update(fit, . ~ .- Age)
> anova(fit, fit1, test="Chi")
Analysis of Deviance Table

Model 1: Kyphosis ~ Age + Number + Start
Model 2: Kyphosis ~ Number + Start
  Resid. Df Resid. Dev Df Deviance P(>|Chi|)
1      77      61.380
2      78      64.536 -1      -3.157      0.076
```

Das Null-Modell (Beobachtungsdaten) weist eine Devianz von $D = 83,234$ (gegenüber dem gesättigten Modell) auf. Die Anzahl der operierten Wirbel und die Lokalisation erklären gleiche Anteile (10,31 bzw. 10,25) von etwa 12% der Devianz aus dem Null-Modell und sollten bei der Modellbildung berücksichtigt werden ($P < 0,01$). Gegenüber einem Modell, das die Anzahl der Wirbel und die Lokalisation erfasst, trägt das Alter zusätzlich nur wenig zur Erklärung bei ($D = -3,16$, mit $P = 0,076$).

Mit den geschätzten Parametern können Modellrechnungen mit hypothetischen Werten im Sinne einer Vorhersage (Prädiktion) durchgeführt werden. Dazu werden in **R** diese Werte in einem neuen Datensatz definiert und mit der Funktion **predict()** über das aktuelle Modell (fit) für eine Vorhersage verwendet.

```
> new.d <- data.frame(Age=c(12,24,60), Number=c(2, 4, 6), Start=c(15, 10, 5))
> new.p <- round(predict(fit, new.d, type = "response"), 4)
> cbind(new.d, new.p)
  Age Number Start new.p
1  12      2    15 0.0150
2  24      4    10 0.1000
3  60      6      5 0.5125
```

Danach ist die Wahrscheinlichkeit einer Kyphose für ein Kind, das zum Beispiel im Alter von 60 Monaten an 6 Wirbeln, beginnend am 5. Wirbel, operiert wird, mit $P = 0,51$ anzusetzen.

8.4.3 Interpretation der Regressionskoeffizienten (odds)

Auf den engen Zusammenhang zwischen den Chancen (odds) für das Eintreten des Ereignisses und den Parametern im Modell der logistischen Regression haben wir in den einleitenden Bemerkungen zu diesem Abschnitt bereits hingewiesen. Für den Fall einer einzelnen dichotomen Einflussgröße kann dies besonders deutlich dargestellt werden.

Tabelle 8.6 Das Modell der logistischen Regression für eine einzelne dichotome Einflussgröße

	$X = 1$	$X = 0$
$Y = 1$	$\pi(1) = \frac{e^{\beta_0 + \beta_1}}{1 + e^{\beta_0 + \beta_1}}$	$\pi(0) = \frac{e^{\beta_0}}{1 + e^{\beta_0}}$
$Y = 0$	$1 - \pi(1) = \frac{1}{1 + e^{\beta_0 + \beta_1}}$	$1 - \pi(0) = \frac{1}{1 + e^{\beta_0}}$

Das Chancenverhältnis (odds ratio) kann für diesen einfachen Fall aus Tabelle 8.6 direkt abgeleitet werden.

$$\psi = \frac{\pi(1)/(1 - \pi(1))}{\pi(0)/(1 - \pi(0))} = e^{\beta_1} \quad (8.54)$$

$$\log(\psi) = \beta_1$$

Damit kann eine Schätzung für das Chancenverhältnis aus der Schätzung für den Parameter im logistischen Regressionsmodell entnommen werden. Dieser Zusammenhang erklärt die große Bedeutung des Modellansatzes der logistischen Regression in epidemiologischen Studien.

Die in (8.54) aufgezeigte Sichtweise kann auf intervallskalierte Einflussgrößen übertragen werden. Dazu betrachtet man die Veränderung im Chancenverhältnis in festen Intervallen $[a, b]$.

$$\begin{aligned}\log \psi(a, b) &= g(x = b) - g(x = a) \\ &= \beta_0 + \beta_1 b - \beta_0 - \beta_1 a \\ &= \beta_1 (b - a)\end{aligned}$$

Danach kann aus der Veränderung der Einflussgröße um $(b - a)$ Einheiten auf die entsprechende Veränderung des Risikos geschlossen werden. Vorausgesetzt wird hierbei allerdings ein linearer Zusammenhang, welcher besonders zu begründen und unter dem Aspekt der Güte der Modellanpassung kritisch zu bewerten ist!



$$\psi(a, b) = e^{\beta_1(b-a)} \quad (8.55)$$

Der Wert e^{β_1} ist ein Maß, dass sich auf eine Einheit der betrachteten Einflussgröße bezieht. Aus den Daten zum Challenger-Unglück wurde der Parameter $\hat{\beta}_1 = -0,2322$ geschätzt (Seite 589). Das Chancenverhältnis für die Temperaturdifferenz von $+10^\circ\text{F}$ ist danach $\hat{\psi}(10) = e^{-2,322} = 0,098$. Anschaulicher ist hier die Bedeutung eines Abfalls der Temperatur um 10°F ($\hat{\psi}(-10) = e^{2,322} = 10,2$), die das Risiko für eine Fehlfunktion der Dichtungsringe um das 10,2fache erhöht.

Werden im Rahmen einer multiplen logistischen Regression mehrere unabhängige Einflussgrößen betrachtet, dann spiegelt sich in den Koeffizienten des Modellansatzes der Einfluss oder Effekt wieder, den die jeweiligen Variablen sich gegenseitig ergänzend zur Erklärung der abhängigen Zielgröße beitragen können. In diesem Sinn kann aus den Koeffizienten auf ein **adjustiertes Chancenverhältnis** geschlossen werden. Diese Form der Adjustierung ist jedoch nur dann angemessen, wenn zwischen den Einflussgrößen keine Wechselwirkungen (Interaktionen) auftreten. Für diesen Fall müssen zusätzliche Terme in die Modellgleichung aufgenommen werden.

8.4.4 Variablenauswahl im Rahmen der Modellbildung

Ziel der Modellbildung ist es, ein Modell mit möglichst wenigen Einflussgrößen (relevante Erklärungsvariablen) zu finden, das möglichst viel über die beobachtete Zielgröße aussagt (sparsame Modellierung, parsimonious modeling). Dazu wird in der Regel ein schrittweises Verfahren (**stepwise logistic regression**) eingesetzt.

Ausgehend von p Einflussgrößen X_i ($i = 1, \dots, p$), die alle einen **sachlogisch begründeten Bezug** zur Zielgröße Y aufweisen, wird in einem ersten Schritt das Nullmodell geschätzt, in welchem nur der konstante Term $\hat{\beta}_0$ (intercept) enthalten ist. Dieses Modell hat den Likelihoodwert L_0 (vgl. (8.45)). Für jedes X_i wird anschließend eine univariate logistische Regression berechnet, die zu den entsprechenden Likelihoodwerten $L_j^{(0)}$ führt. Ein Vergleich dieser Modelle untereinander erfolgt, bezogen auf das Nullmodell, mit der G-Statistik (Likelihood-Ratio-Statistik).

$$G_j^{(0)} = 2(L_j^{(0)} - L_0) \quad j = 1, \dots, p \quad (8.56)$$

Die Statistik G ist χ^2 -verteilt mit ν Freiheitsgraden ($\nu = 1$ für kontinuierliche (intervallskalierte) Einflussgrößen und $\nu = k - 1$ für kategorielle (nominalskalierte) Einflussgrößen). Ein Kriterium für die Entscheidung, welche der Variablen in das Modell aufgenommen werden soll, kann aus dem P-Wert dieser Statistik abgeleitet werden.

$$P_{e_1}^{(0)} = \min\{P(\chi_\nu^2 > G_j^{(0)})\} < \alpha_{in} \quad (8.57)$$

Die Regel besagt, dass die Variable X_{e_1} mit dem kleinsten P-Wert, der kleiner ist als eine vorgegebene Wahrscheinlichkeit α_{in} , als Kandidat für das Modell auszuwählen ist. Die Festlegung eines geeigneten Signifikanzniveaus erfolgt dabei relativ willkürlich. Der Wert $\alpha_{in} = 0,05$ ist im Rahmen der Modellbildung sehr stringend, er wird häufig auf $\alpha_{in} = 0,10$ erhöht. Das Verfahren endet, wenn keine der Einflussgrößen das Kriterium (8.57) erfüllt.

Im nächsten Schritt werden alle Modelle betrachtet, die neben X_{e_1} eine weitere Einflussgröße X_{e_j} enthalten:

$$\begin{aligned} G_j^{(1)} &= 2(L_{e_1 j}^{(1)} - L_{e_1}^{(1)}) \\ P_{e_2}^{(1)} &= \min\{P(\chi_\nu^2 > G_j^{(1)})\} < \alpha_{in} \end{aligned}$$

Die Einflussgröße mit dem kleinsten P-Wert ($< \alpha_{in}$) wird als weiterer Kandidat in das Modell aufgenommen (forward selection) und der Likelihoodwert $L_{e_1 e_2}$ bestimmt. Bevor nun nach weiteren relevanten Einflussgrößen gesucht wird, prüft man, ob nicht auf eine der bereits im Modell befindlichen Variablen verzichtet werden kann (backward elimination). Ein Entscheidungskriterium wird analog zu (8.57) auf der Grundlage der G-Statistik formuliert.

$$G_{-e_j}^{(2)} = 2(L_{e_1 e_2}^{(2)} - L_{-e_j}^{(2)}) \quad j = \{1, 2\}$$

Auch die Festlegung eines Signifikanzniveaus α_{out} für den Ausschluss von Einflussgrößen ist recht willkürlich und richtet sich nach der Bereitschaft, einmal erfasste Variablen wieder aufzugeben. Auf diesen Schritt folgt die Suche nach weiteren Variablen in den verbliebenen $p - 2$ Einflussgrößen entsprechend den beschriebenen Kriterien. Das Verfahren endet, wenn keine gefunden werden. Prinzipiell erfolgt die Variablenauswahl genauso wie im linearen Modell. Der F-test wird durch den G-Test ersetzt. Forward-Selection und Backward-Elimination sind möglich.

Die Festlegung der Signifikanzgrenzen α_{in} und α_{out} in dem beschriebenen Verfahren kann zu unterschiedlichen und widersprüchlichen Ergebnissen in der Modellbildung führen. Vorzuziehen ist ein vergleichbarer Algorithmus auf der Basis des *AIC*-Kriteriums.



$$AIC = -2 \log L + 2K \quad (8.58)$$

Darin bezeichnet K die Anzahl der Freiheitsgrade, in der Regel entsprechend der Zahl der Parameter in dem jeweiligen Modell. Die Modellsuche beginnt auch hier mit dem Nullmodell, in dem nur $\hat{\beta}_0$ (intercept) geschätzt wird. Für jede Einflussgröße X_j ($j = 1, \dots, p$) wird dann eine univariate logistische Regression gerechnet und diejenige Variable ausgewählt, für die das *AIC* nach (8.58) am kleinsten ist. Dieses Modell hat die kleinste Devianz unter allen univariaten Modellen, somit den größten Erklärungswert entsprechend dem *RSS*-Kriterium bei der multiplen linearen Regression. Anschließend wird versucht, weitere Variablen zu finden, die zu einer Reduktion des *AIC* in dem Modell beitragen können. Dabei wird auch geprüft, ob nicht auch der Ausschluss von Variablen aus dem Modell zu einem kleineren *AIC* führen kann. Das Verfahren wird solange fortgesetzt, bis keine weiteren Ein- oder Ausschlüsse von Variablen zu einer Verbesserung des Modells im Sinne des *AIC*-Kriterium führen.

In **R** berechnet die Funktion **extractAIC()** das *AIC* zu einem Modell. Die Funktion **step()**, oder besser **stepAIC()** aus dem Paket **library(MASS)** (Venables und Ripley [VB02]), führt den Prozess der schrittweisen Variablenselektion auf der Grundlage des *AIC*-Kriterium durch. Für die Kyphose-Daten zeigt das Verfahren, dass ausgehend vom Nullmodell (hier „initial model“) durch den schrittweisen Einschluss der Einflussgrößen der Wert für *AIC* von 85,23 auf 69,38 reduziert werden kann.

```
> model.step <- stepAIC(model, Kyphosis ~ Age + Number + Start,
  trace = FALSE, direction="both")
> model.step$anova
Stepwise Model Path
Analysis of Deviance Table

Initial Model:
Kyphosis ~ 1

Final Model:
Kyphosis ~ Start + Number + Age
```

	Step	Df	Deviance	Resid. Df	Resid. Dev	AIC
1				80	83.23447	85.23447
2	+ Start	1	15.162295	79	68.07218	72.07218
3	+ Number	1	3.535712	78	64.53647	70.53647
4	+ Age	1	3.156541	77	61.37993	69.37993

8.4.5 Residuenanalyse

Die Analyse von Residuen im Modell der logistischen Regression erfolgt nach zwei verschiedenen Ansätzen. Ähnlich zu den standardisierten Residuen im linearen Regressionsmodell werden die **Pearson-Residuen** definiert.

$$r_i = \frac{y_i - n_i \hat{\pi}_i}{\sqrt{n_i \hat{\pi}_i (1 - \hat{\pi}_i)}} \quad i = 1, \dots, n \quad (8.59)$$

In (8.59) ist n_i die Anzahl der Beobachtungen, die mit der Kovariatenkombination $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ des i -ten Falles übereinstimmen, und $\hat{\pi}_i$ ist die Schätzung für die Wahrscheinlichkeit aus dem Modell ($\text{logit}(\hat{\pi}_i) = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_p x_{ip}$).

Eine andere Sicht auf die Residuen ergibt sich aus einer Zerlegung der **Modelldevianz** D . Diese kann auch aus einzelnen Komponenten, die sich jeweils aus der Differenz zum gesättigten Modell in den Einzelwahrscheinlichkeiten ergeben, nach (8.60) berechnet werden (**Devianz-Residuen**).

$$D = \sum_{i=1}^n d_i^2 \quad (8.60)$$

mit
$$d_i = \pm \sqrt{-2(y_i \log \frac{y_i}{n_i \hat{\pi}_i}) + (n_i - y_i) \log \frac{(n_i - y_i)}{n_i(1 - \hat{\pi}_i)}}$$

Die Summe der quadrierten Pearson-Residuen nach (8.59) ist angenähert χ^2 -verteilt mit $(n-p-1)$ Freiheitsgraden. Man kann zeigen, dass die Statistik aus den Pearson-Residuen asymptotisch äquivalent zu der Devianzstatistik (D) ist und somit für einen Test der Güte der Modellanpassung (goodness of fit) verwendet werden kann. Die Residuen nach beiden Ansätzen können in **R** mit der Funktion **residuals()** berechnet werden. In Abbildung 8.10 sind diese für die Kyphose-Daten auf der linken Seite dargestellt. Residuen sollten unter korrekten Modellannahmen annähernd normalverteilt sein. Für die vorliegenden Daten ist dies offenbar nicht der Fall!

Von besonderer Bedeutung im Rahmen der Modellbildung ist die Erkennung von Einzelbeobachtungen, die einen **wesentlichen Einfluss** auf das jeweilige Modell haben (influential points). Dazu kann eine Statistik herangezogen werden, die sowohl die Devianzanteile nach (8.60) als auch die Pearson-Residuen nach (8.59) berücksichtigt.

$$\Delta D_i = d_i^2 + r_i^2 \frac{h_{ii}}{1 - h_{ii}} \quad (8.61)$$

Darin sind die h_{ii} die Diagonalelemente der Hut-Matrix aus der iterativen Lösung im Rahmen der Maximum-Likelihood Schätzung (vgl. (8.47)). Auf diese kann in **R** mit der Funktion **influence()** zugegriffen werden. Abbildung 8.10 zeigt auf der rechten Seite die Werte für ΔD_i für alle Beobachtungen. Einzelbeobachtungen können damit identifiziert und von der Modellbildung ausgeschlossen werden.

```
> deviance.resid <- residuals(fit)
> pearson.resid <- residuals(fit, type="pearson")
> hats <- influence(fit)$hat
> idev <- deviance.resid^2 + pearson.resid^2 * hats/(1-hats)
```

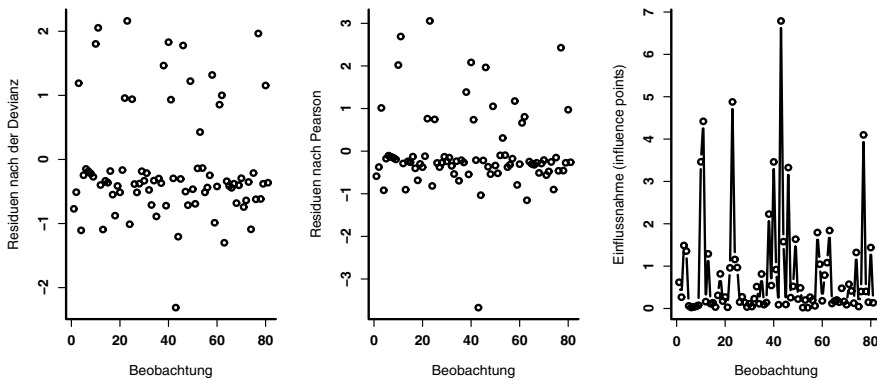


Abb. 8.10 Residuenanalyse für das Modell der logistischen Regression anhand der Kyphose-Daten; linke Seite Devianz-Residuen und Pearson-Residuen; rechte Seite Einzelbeobachtungen mit wesentlichem Einfluss (influential points)

8.5 Log-lineare Modelle

- Kontingenztafeln
- Log-lineares Modell am Beispiel von 2 Faktoren
- Drei-dimensionale Kontingenztafeln
- Allgemeines log-lineares Modell

8.5.1 Kontingenztafeln

Die statistische Analyse nominalskaliert (kategorialer) Merkmale basiert auf Häufigkeiten, bzw. auf den aus den Häufigkeiten geschätzten Wahrscheinlichkeiten. Allgemein werden p Faktoren mit jeweils k_i ($i = 1, \dots, p$) Ausprägungen (Faktorstufen) betrachtet. Dabei wird zunächst noch keine Einteilung oder Zuordnung in Zielgröße und Einflussgrößen, wie in den Regressionsmodellen gefordert, vorgenommen.



Beispiel A: Für $p = 2$ Faktoren können die Beobachtungen in einer **2-dimensionalen Kontingenztafel** mit insgesamt $k_1 k_2 = m$ Feldern (Zellen) zusammengefasst werden.

Bei einer Befragung von 447 männlichen deutschen Arbeitslosen, die vom Deutschen Institut für Wirtschaftsforschung (DIW) durchgeführt wurde, sind unter anderem die beiden folgenden Faktoren (Merkmale) erhoben worden (Fahrmeier [FKPT00]):

- (1) Das Ausbildungsniveau (X) mit 4 Ausprägungen ($k_1 = 4$): keine Ausbildung (k), Lehre (l), fachspezifische Ausbildung (f) und Hochschulabschluss (h).
- (2) Die Dauer der Arbeitslosigkeit (Y) mit 3 Ausprägungen ($k_2 = 3$): Kurzzeitarbeitslosigkeit (≤ 6 Monate), mittelfristige Arbeitslosigkeit (7 – 12 Monate) und Langzeitarbeitslosigkeit (> 12 Monate).

Ausbildung	Dauer der Arbeitslosigkeit			Summe
	≤ 6 Monate	7 – 12 Monate	> 12 Monate	
keine	86	19	18	123
Lehre	170	43	20	233
fachspezifisch	40	11	5	56
Hochschule	28	4	3	35
Summe	324	77	46	447

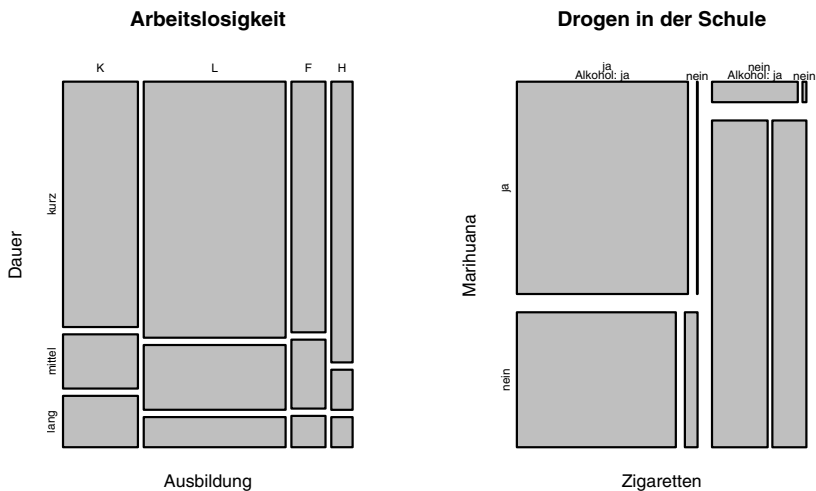


Abb. 8.11 Mosaikplot zu den Beispieldaten zur Arbeitslosigkeit (Beispiel A) und zum Gebrauch von Drogen (Beispiel B)

Beispiel B: Für $p = 3$ Faktoren können die Beobachtungen in einer **3-dimensionalen Kontingenztafel** (Kontingenzwürfel) mit insgesamt $k_1 k_2 k_3 = m$ Feldern zusammengefasst werden. In einer Umfrage der Wright State University School of Medicine und des United Health Services in Dayton (Ohio) unter 2276 Studenten, die sich im letzten Jahr der High School befanden, wird gefragt, ob sie jemals Alkohol (A), Zigaretten (Z) oder Marihuana (M) konsumiert hätten (Agresti [Agr02]).



Alkohol	Zigaretten	Marihuana	
		ja	nein
ja	ja	911	538
	nein	44	456
nein	ja	3	43
	nein	2	279

Allgemein werden im Rahmen der Beobachtung oder Untersuchung n Fälle auf die m Felder einer Kontingenztafel verteilt. Die Wahrscheinlichkeit dafür, dass ein zufällig ausgewähltes Individuum in das i te Feld gelangt ist π_i , mit

$$\pi_1 + \pi_2 + \dots + \pi_m = 1 \quad .$$

Die Verteilung der Anzahl der Fälle in dem i ten Feld einer Kontingenztafel Y_i ($\sum Y_i = n$) kann durch eine **Multinomialverteilung** beschrieben werden kann. Die Wahrscheinlichkeit für ein beobachtetes Ergebnis ist danach:

$$P(y_1, \dots, y_m) = \frac{n!}{y_1! y_2! \dots y_m!} \pi_1^{y_1} \dots \pi_m^{y_m} \quad (8.62)$$

Aus der Likelihood-Funktion zu (8.62) kann eine Maximum-Likelihood-Schätzung abgeleitet werden.

$$P = L(\pi_1, \dots, \pi_m | y_1, \dots, y_m)$$

$$-2 \log L = -2 \sum_{i=1}^m y_i \log \pi_i + \text{Konstante} \quad (8.63)$$

Der kleinste Likelihood-Wert nach (8.63) wird für das saturierte (vollständige) Modell erreicht, für das keine Randbedingungen hinsichtlich der π_i vorliegen (ausser $\sum \pi = 1$). Dann ist die Schätzung

$$\hat{\pi}_i = \frac{y_i}{n} \quad (8.64)$$

und

$$-2 \log L_{sat} = -2 \sum_{i=1}^n y_i \log \frac{y_i}{n} + K = -2 \sum_{i=1}^n y_i \log y_i + 2n \log n + K \quad (8.65)$$

Einschränkungen (constraints - C) führen zu anderen ML-Schätzungen. Der Test von Hypothesen, die sich durch Einschränkungen in den Modellvoraussetzungen formulieren lassen, kann durch die Differenz im Likelihood-Wert gegenüber dem saturierten Modell durchgeführt werden (**Likelihood-Quotiententest**). Die **Devianz-Statistik** D in (8.66), die auch schon im Abschnitt zur logistischen Regression ausführlicher dargestellt wurde, ist angenähert χ^2 -verteilt mit $m - c - 1$ Freiheitsgraden, wobei c die Anzahl freier Parameter in dem eingeschränkten Modell ist.

$$D = (-2 \log L_C) - (-2 \log L_{sat}) \sim \chi_{m-c-1}^2 \quad (8.66)$$

So lässt sich die **Hypothese der Unabhängigkeit** zweier Faktoren X und Y in einer 2-dimensionalen Kontingenztafel (Beispiel A) durch eine **Restriktion in den Randwahrscheinlichkeiten** der Kontingenztafel formulieren.

$$\pi_{ij} = P(X \cap Y) = P(X = x_i \cap Y = y_j)$$

$$\pi_{ij} = P(X \cap Y) = P(X) \cdot P(Y) = \pi_{i+} \cdot \pi_{+j} \quad \text{mit}$$

$$\pi_{i+} = \sum_{j=1}^{k_2} \pi_{ij} \quad \text{und} \quad \pi_{+j} = \sum_{i=1}^{k_1} \pi_{ij}$$

Die ML-Schätzung ist dann

$$\hat{\pi}_{ij} = \frac{y_{i+} \cdot y_{+j}}{n} \quad (8.67)$$

und der Likelihood-Wert für das so eingeschränkte Modell ist

$$-2 \log L_C = -2 \sum_{i=1}^{k_1} \sum_{j=1}^{k_2} y_{ij} \log \left(\frac{y_{i+} \cdot y_{+j}}{n^2} \right) \quad (8.68)$$

Die Statistik nach (8.66) kann dann für den Test auf Unabhängigkeit ($H_0 : \pi_{ij} = \pi_{i+} \cdot \pi_{+j}$) verwendet werden. Die Rechnung soll in **R** schrittweise an den DIW-Daten (Beispiel A) gezeigt werden.

```
> y <- c(86, 19, 18, 170, 43, 20, 40, 11, 5, 28, 4, 3)
> n <- sum(y)
> tab <- matrix(y, byrow=TRUE, nrow = 4) # Tabelle zu Beispiel A
> dimnames(tab) <- list(ausbildung=c("K","L","F","H"), zeit=c("k","m","l"))
> tab
      zeit
ausbildung k  m  l
K      86 19 18
L     170 43 20
F      40 11  5
H      28  4  3
> zeit.sum <- apply(tab, 2, sum) # Randsummen
> ausb.sum <- apply(tab, 1, sum)
> L.sat <- -2*sum(y*log(y/n)); L.sat # saturiertes Modell
[1] 1715.890
> L.c <- c(0)
> for (i in 1:4) { for (j in 1:3) { # feste Randsummen
+   L.c <- L.c + tab[i, j] * log(ausb.sum[i]*zeit.sum[j]/n^2) }}
> L.c <- -2*L.c; L.c
1720.577
> devianz <- L.c - L.sat; devianz
4.687199
> 1-pchisq(devianz, 6)
0.5845111
```

Der Likelihood-Wert für das saturierte Modell ist $-2 \log L_{sat} = 1715,89$. Unter der Nullhypothese der Unabhängigkeit von Ausbildungsniveau X und Dauer der Arbeitslosigkeit Y ist der Wert für die Likelihood $-2 \log L_C = 1720,58$. Die Devianz $D = 4,69$ ist unter der Nullhypothese angenähert χ^2 -verteilt mit 6 Freiheitsgraden, das bedeutet $P = 0,585$. Die Nullhypothese kann somit nicht verworfen werden.

Hinweis: Die Statistik (8.66) entspricht dem Likelihood-Quotiententest, der asymptotisch äquivalent zur χ^2 -Statistik nach Pearson ist (vgl. Abschnitt [7.6]). Unser Resultat kann somit in **R** auch einfach durch die Funktion **chisq.test()** bestätigt werden. Abweichungen in den Ergebnissen erklären sich aus der Tatsache, das die beiden Ansätze nur asymptotisch äquivalent sind.

```
> chisq.test(tab)

Pearson's Chi-squared test

data:  tab X-squared = 4.8195, df = 6, p-value = 0.5672
```

8.5.2 Log-lineares Modell am Beispiel von 2 Faktoren

Die log-lineare Methode zur Analyse von Zusammenhängen in mehrdimensionalen Häufigkeitsverteilungen (Kontingenztafeln) wurde von L.A. Goodman entwickelt. Sie kann als ein Spezialfall der verallgemeinerten linearen Modelle für Poisson-verteilte Daten betrachtet werden. Auch wenn **log-lineare Modelle** grundsätzlich für die Analyse von Zusammenhängen zwischen zwei kategorialen Variablen (2-dimensionale Kontingenztafel) verwendet werden können (zur Analyse von Kontingenztafeln siehe Abschnitt [7.6]), werden sie vorwiegend für die Bewertung **mehrdimensionaler Kontingenztafeln** mit drei oder mehr Variablen eingesetzt. Dabei werden die untersuchten Merkmale alle gemeinsam als Zielgrößen behandelt, d.h. es wird insbesondere hinsichtlich der Zusammenhänge keine Unterscheidung zwischen Ziel- und Einflussgrößen gemacht. Sollte eines der Merkmale explizit als Zielgröße, und die anderen als Einflussgrößen gesehen werden, dann ist besser ein logistisches Modell zu wählen.

Die Modellbildung im log-linearen Modell erfolgt hinsichtlich der logarithmierten Häufigkeiten $\log n\pi_{ij} = \eta_{ij}$ in einem der 2-faktoriellen Varianzanalyse vergleichbaren Ansatz (s.a. Abschnitt [8.3]). Für den Fall einer 2-dimensionalen Kontingenztafel mit den Faktoren X und Y wird das Modell in (8.69) angegeben.

$$\begin{aligned}
 \eta_{ij} &= \log n\pi_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij} \\
 \text{mit } \mu &= \eta_{++} \\
 \alpha_i &= \eta_{i+} - \eta_{++} \\
 \beta_j &= \eta_{+j} - \eta_{++} \\
 \gamma_{ij} &= \eta_{ij} - \eta_{i+} - \eta_{+j} + \eta_{++}
 \end{aligned}
 \tag{8.69}$$

Das Gleichungssystem ist überbestimmt. Die Lösung nach den Parametern α_i , β_j und γ_{ij} erfordert wie bei der 2-faktoriellen Varianzanalyse Restriktionen in den Parametern, die zu einer geeigneten Modellmatrix führen, z.B. durch

$$\begin{aligned}
 \sum_{i=1}^{k_1} \alpha_i &= 0 & \sum_{j=1}^{k_2} \beta_j &= 0 \\
 \sum_{i=1}^{k_1} \gamma_{ij} &= 0 \quad \text{für alle } j = 1, \dots, k_2 \\
 \sum_{j=1}^{k_2} \gamma_{ij} &= 0 \quad \text{für alle } i = 1, \dots, k_1
 \end{aligned}$$

Die zusätzliche Randbedingung $\sum \sum \pi_{ij} = 0$ führt in diesem Zusammenhang dazu, dass auch der unbekannte Parameter μ durch die anderen Parameter mitbestimmt werden kann. Somit können die Wahrscheinlichkeiten im Modell durch $k_1 k_2 - 1$ unabhängige Parameter beschrieben und geschätzt werden.

Das Modell (8.69) entspricht dem saturierten Modell im vorangegangenen Abschnitt. Unter der Annahme, dass die Häufigkeiten y_{ij} der Kontingenztafel durch Poisson-Verteilungen mit den Erwartungswerten λ_{ij} beschrieben werden können, lässt sich das Modell mit der **Logarithmustransformation als Linkfunktion** als ein **verallgemeinertes lineares Modell** auffassen. Für die Beschreibung der Modelle soll daher im Folgenden die Symbolik nach (8.70) verwendet werden.

$$\log n\pi_{ij} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY} \quad (8.70)$$

Die Parameter λ_i^X und λ_j^Y kennzeichnen die Haupteffekte in den Merkmalen X und Y , die Parameter λ_{ij}^{XY} einen Wechselwirkungseffekt (Interaktion) zwischen den Merkmalen X und Y . Für die Überprüfung von Hypothesen hinsichtlich der Unabhängigkeit der Faktoren X und Y ist die Darstellung der Schätzung der einzelnen Modell-Parameter und deren Interpretation nicht notwendig. Eine ausführliche Statistik kann mit der Funktion **summary()** zu einem konkreten Modellansatz bei Bedarf berechnet werden. An dieser Stelle interessiert vielmehr die Frage, ob auf bestimmte Komponenten in der Modellgleichung (8.70) verzichtet werden kann, ohne dabei die Güte der Anpassung an die beobachteten Daten wesentlich zu verringern.

Die Güte der Anpassung (goodness of fit) eines Modells an die beobachteten Daten kann durch die χ^2 -Statistik nach (8.71) beschrieben werden (**Pearson-Residuen**),

$$\chi^2 = \sum_{i,j} \frac{(y_{ij} - n\hat{\pi}_{ij})^2}{n\hat{\pi}_{ij}} \quad (8.71)$$

mit $\sum y_{ij} = n$ und $\hat{\pi}_{ij}$ den im Rahmen eines bestimmten Modelles geschätzten Wahrscheinlichkeiten. Bevorzugt wird jedoch häufig die Likelihood-Quotienten-Statistik (8.72) verwendet, deren Minimum aus einer Maximum-Likelihood-Schätzung der Wahrscheinlichkeiten aus den Daten resultiert (**Devianz-Residuen**).

$$G^2 = 2 \sum_{i,j} y_{ij} \log \frac{y_{ij}}{n\hat{\pi}_{ij}} \quad (8.72)$$

Beide Statistiken, (8.71) und (8.72), sind asymptotisch χ^2 -verteilt und ermöglichen somit eine auf Hypothesen basierende statistische Bewertung einzelner Modelle (P-Werte). Dagegen ist das im Rahmen der Modellbildung, insbesondere bei den schrittweisen Verfahren, ebenfalls häufig verwendete AIC-Kriterium eine relative Maßzahl für die Güte der Modellanpassung. Diese ist maximal für die beobachteten Daten selbst (Nullmodell) und wird durch ein „bestes“ Modell minimiert.

Für einen Test auf Unabhängigkeit im Fall der 2-dimensionalen Kontingenztafel wird z.B. das Modell (8.73) berechnet. Die verbleibende Devianz (residual deviance) ist χ^2 -verteilt mit $(k_1 - 1)(k_2 - 1)$ Freiheitsgraden und kann gegenüber dem saturierten Modell bewertet werden.

$$\log n\pi_{ij} = \mu + \lambda_i^X + \lambda_j^Y \quad (8.73)$$

Für das Beispiel A (Arbeitslosigkeit) können die Modelle in **R** durch die Funktion **glm()** spezifiziert und mit der Funktion **anova()** gegenübergestellt werden.

```
> y <- c(86, 19, 18, 170, 43, 20, 40, 11, 5, 28, 4, 3)
> mit.ausbildung <- c(rep("K",3), rep("L",3), rep("F",3), rep("H",3))
> zeit <- rep(c("k","m","l"),4)
> tab <- data.frame(ausbildung, zeit, y)
> fit.sat <- glm(y ~ zeit+ausbildung+zeit:ausbildung, family=poisson, data=tab)
> fit.c <- update(fit.sat, . ~ . - zeit:ausbildung)
> anova(fit.sat, fit.c)
Analysis of Deviance Table

Model 1: y ~ zeit + ausbildung + zeit:ausbildung
Model 2: y ~ zeit + ausbildung
  Resid. Df Resid. Dev Df Deviance
1         0  -2.065e-14
2         6   4.6872  -6   -4.6872
```

Die Differenz in den Devianzen $D = 4,69$ ist χ^2 -verteilt mit 6 Freiheitsgraden ($P=0,585$). Auf einen Interaktionsterm λ_{ij}^{XY} kann somit in der Modellgleichung verzichtet werden. Das Ergebnis spricht damit nicht gegen die Annahme einer Unabhängigkeit der beiden Faktoren X und Y .

8.5.3 Drei-dimensionale Kontingenztafeln

8.5.3.1 Modellbildung unter verschiedenen Restriktionen

Die Analyse mehrerer Faktoren im loglinearen Modell geht von der Untersuchung aller Parameter im saturierten Modell aus. Dabei lässt sich von den signifikanten Modellparametern auf die Notwendigkeit schließen, die entsprechenden Effekte bei der Modellierung zu berücksichtigen. Das vollständige loglineare Modell für eine 3-dimensionale Kontingenztafel (in den Faktoren X , Y und Z) ist in (8.74) angegeben.

$$\log n\pi_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{jk}^{YZ} + \lambda_{ik}^{XZ} + \lambda_{ijk}^{XYZ} \quad (8.74)$$

Der entsprechende Likelihood-Wert kann nach (8.75) berechnet werden.

$$-2 \log L_{sat} = -2 \left(\sum_{i=1}^{k_1} \sum_{j=1}^{k_2} \sum_{k=1}^{k_3} y_{ijk} \log y_{ijk} - n \log n \right) + K \quad (8.75)$$

Für die Überprüfung von Nullhypothesen in 3-dimensionalen Kontingenztafeln müssen unterschiedliche Formen der Unabhängigkeit betrachtet werden. Eine Übersicht wird in Tabelle 8.7 gegeben.

(A) Das Modell der **totalen Unabhängigkeit** verlangt, dass alle Faktoren unabhängig sind. Daraus folgt insbesondere eine gegen- oder **wechselseitige Unabhängigkeit** (mutual independence) der drei Faktoren untereinander. Die **Randbedingung** für dieses Modell, in dem keinerlei Wechselwirkungen berücksichtigt werden, wird durch (8.76) gegeben.

$$\lambda^{XY} = \lambda^{YZ} = \lambda^{XZ} = \lambda^{XYZ} = 0$$

$$H_0 : \pi_{ijk} = \pi_{i++}\pi_{+j+}\pi_{++k} \Rightarrow \log n\pi_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z \quad (8.76)$$

(B) Die gemeinsame oder **blockweise Unabhängigkeit** (joint independence) eines Faktors von den jeweils zwei anderen Faktoren resultiert aus der Randbedingung (8.77) für die Variante Y von XZ . Im Vergleich zu dem Modell (8.76) wird hier die Unabhängigkeit von X und Z nicht angenommen, d.h. eine Wechselwirkung zugelassen.

$$H_0 : \pi_{ijk} = \pi_{+j+}\pi_{i+k} \Rightarrow \log n\pi_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ik}^{XZ} \quad (8.77)$$

Die entsprechenden Möglichkeiten für die anderen Variablenkombinationen sind in Tabelle 8.7 aufgeführt. Aus einer vollständigen Unabhängigkeit kann auf die blockweise Unabhängigkeit geschlossen werden.

(C) Zwei Faktoren, z.B. X und Y , heißen **bedingt unabhängig bei gegebenem dritten Faktor**, z.B. Z , wenn gilt

$$\begin{aligned} \pi_{ij|k} &= P(X=i, Y=j|Z=k) \\ &= P(X=i|Z=k) \cdot P(Y=j|Z=k) \\ &= \pi_{i+|k} \pi_{+j|k} \end{aligned}$$

Mit dieser Definition kann eine weitere Restriktion für die **bedingte Unabhängigkeit** (conditional independence) im loglinearen Modell formuliert werden (Beispiel $X - Y|Z$), in dem nur die Interaktionsterme für XZ und YZ berücksichtigt werden (8.78), d.h. nur X und Y sind unabhängig.

$$H_0 : \pi_{ijk} = \pi_{i+k}\pi_{+jk}/\pi_{++k} \Rightarrow \log n\pi_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} \quad (8.78)$$

Die anderen Kombinationen sind in Tabelle 8.7 vollständig aufgeführt. Die Randbedingungen zur bedingten Unabhängigkeit sind „schwächer“ als die Randbedingungen zur blockweisen Unabhängigkeit, oder anders, bei blockweiser Unabhängigkeit kann auch auf eine bedingte Unabhängigkeit geschlossen werden (vgl. auch Abbildung 8.12).

(D) Das Modell **ohne Wechselwirkung 3.Ordnung** (ohne den Term λ_{ijk}^{XYZ}) kennzeichnet die Situation einer paarweisen Abhängigkeit zwischen den Faktoren, ohne dass zwischen allen drei Faktoren eine Wechselwirkung in dem Modell zugelassen wird.

$$H_0 : \log n\pi_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} \quad (8.79)$$

Grundsätzlich kann die Nullhypothese der Unabhängigkeit auch als **Chancenverhältnis** (odds ratio) formuliert werden. So sind für den Fall einer 4-Feldertafel (2 Faktoren mit je 2 Kategorien) die beiden Faktoren unabhängig, wenn gilt:

$$H_0 : \theta = \frac{\pi_{11}/\pi_{21}}{\pi_{12}/\pi_{22}} = \frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}} = 1$$

Für 3-dimensionale Kontingenztafeln können **bedingte Chancenverhältnisse** zwischen den Kategorien i und j in den Faktoren X und Y betrachtet werden (8.80), die unter der Annahme des Fehlens einer Wechselwirkung 3. Ordnung für alle Kategorien k des dritten Faktors Z konstant sind.

$$\theta_{ij(k)} = \frac{\pi_{ijk}\pi_{IJk}}{\pi_{iJk}\pi_{Ijk}} = \text{const} \quad \text{mit } I = i + 1 \text{ und } J = j + 1 \text{ und für alle } k \quad (8.80)$$

Eine entsprechende Darstellung ist auch hinsichtlich der Chancenverhältnisse $\theta_{i(j)k}$ und $\theta_{(i)jk}$ möglich. Insbesondere können die Modellparameter aus (8.79) direkt auch zur Schätzung der Chancenverhältnisse herangezogen werden. Aus

$$\log \theta_{ij(k)} = \lambda_{ij}^{XY} + \lambda_{IJ}^{XY} - \lambda_{iJ}^{XY} - \lambda_{Ij}^{XY}$$

folgt unter der Restriktion des Fehlens einer Wechselwirkung 3. Ordnung:

$$\log \hat{\theta}_{ij(k)} = \hat{\lambda}_{ij}^{XY} \quad (8.81)$$

8.5.3.2 Modellauswahl im log-linearen Ansatz

Die Suche nach einem geeigneten loglinearen Modell zur Beschreibung der beobachteten Häufigkeiten beginnt nach Goodman [Goo71] mit dem Modell der vollständigen Unabhängigkeit (A). Zeigt die Devianz-Statistik eine signifikante Abweichung gegenüber dem saturierten Modell, dann wird aus den Modellen mit blockweiser Abhängigkeit (B) das Modell mit der kleinsten Devianz betrachtet. Wird auch dieses abgelehnt, dann wird die Suche unter den Modellen mit bedingter Unabhängigkeit (C) und ggf. mit dem Modell ohne Wechselwirkung 3. Ordnung (D) fortgesetzt. Diese Form einer hierarchischen Modellbildung wird ergänzend zur Tabelle 8.7 anschaulich auch in Abbildung 8.12 dargestellt.

Für jedes Modell kann die Güte der Anpassung an die beobachteten Daten durch die Devianzstatistik (Likelihood-Quotiententest) bzw. durch das entsprechende AIC-Kriterium bestimmt werden. In Tabelle 8.8 sind die Ergebnisse aller Modelle zu den Daten aus dem Drogen-Beispiel (vgl. Seite 599, Abbildung 8.11) zusammengestellt, die in **R** direkt aus der Funktion **glm()** abgeleitet werden können, z.B. für das Modell (A):

Tabelle 8.7 Hypothesen zur Unabhängigkeit der Faktoren X , Y und Z im loglinearen Modell (dreidimensionale Kontingenztafeln)

Typ	Symbol	Nullhypothese (H_0)	Modell
A	$X - Y - Z$	$\pi_{ijk} = \pi_{i++} + \pi_{+j+} + \pi_{++k}$	$\mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z$
B1	$Y - XZ$	$\pi_{ijk} = \pi_{+j+} + \pi_{i+k}$	$\mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ik}^{XZ}$
B2	$X - YZ$	$\pi_{ijk} = \pi_{i++} + \pi_{+jk}$	$\mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{jk}^{YZ}$
B3	$Z - XY$	$\pi_{ijk} = \pi_{++k} + \pi_{ij+}$	$\mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY}$
C1	$X - Y Z$	$\pi_{ijk} = \pi_{i+k} + \pi_{+jk} / \pi_{++k}$	$\mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}$
C2	$X - Z Y$	$\pi_{ijk} = \pi_{ij+} + \pi_{+jk} / \pi_{+j+}$	$\mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{jk}^{YZ}$
C3	$Y - Z X$	$\pi_{ijk} = \pi_{ij+} + \pi_{i+k} / \pi_{i++}$	$\mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ}$
D	XY, XZ, YZ		$\mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}$
	XYZ	vollständiges (saturiertes) Modell nach (8.74)	

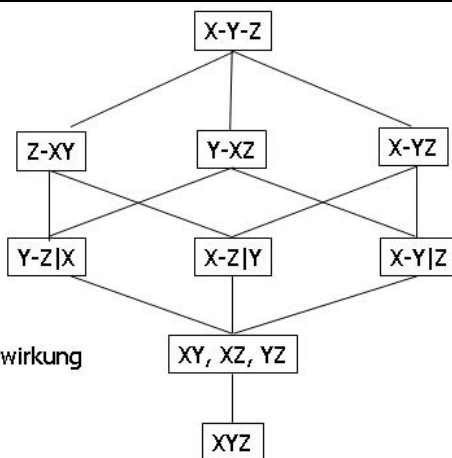
(A) totale Unabhängigkeit

(B) blockweise Unabh.

(C) bedingte Unabh.

(D) keine 3-Faktorwechselwirkung

saturiertes Modell

**Abb. 8.12** Hierarchisch geordnete Hypothesen zur Unabhängigkeit im loglinearen Modell 3-dimensionaler Kontingenztafeln

```

> fit.a <- glm(y ~ marihuana + zigarette + alkohol, family=poisson, data=tab)
> val[,1] <- round(fitted.values(fit.a), 1)
> stats[,1,1] <- round(fit.a$deviance, 1); stats[,1,2] <- round(fit.a$aic, 1)
> stats[,1,3] <- fit.a$df.residual

```

Das Ergebnis zeigt die beste Anpassung im Modell D, d.h. der Beitrag einer Wechselwirkung 3. Ordnung ist nicht signifikant. In **R** kann die Suche nach dem geeigneten Modell nach dem beschriebenen Verfahren auch mit der Funktion **stepAIC()** erfolgen.

```

> model.step <- stepAIC(model, list(upper = ~.^3,
                                   lower = formula(model)), trace=FALSE)
> model.step$anova
Stepwise Model Path
Analysis of Deviance Table

Initial Model:
y ~ marihuana + zigarette + alkohol

Final Model:
y ~ marihuana + zigarette + alkohol + marihuana:zigarette + zigarette:alkohol +
  marihuana:alkohol

```

Tabelle 8.8 Test zur Güte der Anpassung der Modelle zu den Unabhängigkeitshypothesen mit den Daten zum Drogenkonsum; der P-Wert basiert auf der Devianz-Statistik

	Devianz	AIC	Freiheitsgrad	P-Wert
A	1286.00	1343.10	4	< 0.001
B1	843.80	902.90	3	< 0.001
B2	939.60	998.60	3	< 0.001
B3	534.20	593.30	3	< 0.001
C1	187.80	248.80	2	< 0.001
C2	497.40	558.40	2	< 0.001
C3	92.00	153.10	2	< 0.001
D	0.40	63.40	1	0.53

	Step	Df	Deviance	Resid. Df	Resid. Dev	AIC
1				4	1286.0199544	1343.06338
2	+ marihuana: zigarette	1	751.80828	3	534.2116714	593.25510
3	+ zigarette: alkohol	1	442.19331	2	92.0183606	153.06179
4	+ marihuana: alkohol	1	91.64437	1	0.3739859	63.41741

Tabelle 8.9 Anpassung der Daten aus dem Drogen-Beispiel unter verschiedenen log-linearen Modellen (Tabelle 8.7). In der letzten Spalte sind die beobachteten Häufigkeiten (saturiertes Modell) aufgeführt

	A	B1	B2	B3	C1	C2	C3	D	beobachtet
1	540.0	611.2	627.3	782.7	909.2	710.0	885.9	910.4	911
2	740.2	837.8	652.9	497.5	438.8	739.0	563.1	538.6	538
3	282.1	210.9	327.7	39.4	45.8	245.0	29.4	44.6	44
4	386.7	289.1	341.1	629.4	555.2	255.	470.6	455.4	456
5	90.6	19.4	3.3	131.3	4.8	0.7	28.1	3.60	3
6	124.2	26.6	211.5	83.5	142.2	45.3	17.9	42.4	43
7	47.3	118.5	1.7	6.6	0.2	4.3	16.6	1.4	2
8	64.9	162.5	110.5	105.6	179.8	276.7	264.4	279.6	279
$\hat{\chi}^2$	1410.98	704.80	824.10	505.59	181.03	443.83	80.80	0.38	sat.
FG	4	3	3	3	2	2	2	1	Modell

Im Kontext der Analyse von Kontingenztafeln können die Abweichungen zwischen den beobachteten und den unter den Modellannahmen erwarteten Häufigkeiten durch die **Pearson-Residuen** beurteilt werden:

$$r_{ijk} = \frac{y_{ijk} - n\hat{\pi}_{ijk}}{\sqrt{n\hat{\pi}_{ijk}}}$$

Die quadrierten Pearson-Residuen sind somit die Grundlage für die χ^2 -Statistik (8.82), mit der die Güte der verschiedenen Modelle beurteilt werden kann (Tabelle 8.9).

$$\hat{\chi}^2 = \sum_{ijk} \frac{(y_{ijk} - n\hat{\pi}_{ijk})^2}{n\hat{\pi}_{ijk}} \quad (8.82)$$

Die Anzahl der Freiheitsgrade resultiert aus den im Modell frei verfügbaren, d.h. nicht durch Restriktionen betroffenen, Parameter. In der Tabelle 8.9 sind die Schätzungen, die in **R** aus der

Funktion **glm()** mit dem Wert „**fitted.values**“ übernommen werden können, zu den verschiedenen Modellen (entsprechend Tabelle 8.7) zusammengestellt. Das Ergebnis der χ^2 -Statistik in der untersten Zeile stimmt mit den oben genannten Resultaten zur Modellbildung überein.

Interpretation der Modellparameter

Die Berechnung des nach den vorangehenden Überlegungen besten Modells zur Beschreibung der Daten können in **R** mit der Funktion **summary()** übersichtlich ausgegeben werden. Dabei werden die Faktoren aus dem Beispiel B hier durch m (Marihuana), a (Alkohol) und z (Zigaretten) abgekürzt.

```
> fit <- glm(y ~ m * a * z - m:z:a, family=poisson, data=tab, x=T)
> summary(fit)
```

Call:
glm(formula = y ~ m * a * z - m:z:a, family = poisson, data = tab, x = T)

...

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	6.81387	0.03313	205.699	< 2e-16 ***
m nein	-0.52486	0.05428	-9.669	< 2e-16 ***
a nein	-5.52827	0.45221	-12.225	< 2e-16 ***
z nein	-3.01575	0.15162	-19.891	< 2e-16 ***
m nein:a nein	2.98601	0.46468	6.426	1.31e-10 ***
m nein:z nein	2.84789	0.16384	17.382	< 2e-16 ***
a nein:z nein	2.05453	0.17406	11.803	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 2851.46098 on 7 degrees of freedom
 Residual deviance: 0.37399 on 1 degrees of freedom
 AIC: 63.417

Number of Fisher Scoring iterations: 4

Die Modellparameter des Modells ohne Wechselwirkung sind alle signifikant von Null verschieden. Für die Interpretation der Modellparameter wird von den Termen mit höchster Ordnung ausgegangen. So lässt sich aus $\hat{\lambda}_{00}^{ma} = 2,98$ auf ein bedingtes Chancenverhältnis $\hat{\theta}_{00}^{ma} = e^{\hat{\lambda}_{00}^{ma}} = 19,8 \sim 20$ zwischen Alkohol und Marihuana schließen, welches sowohl für die Raucher als auch für die Nichtraucher zutrifft. Das bedeutet, dass die Chance, einen Studenten mit Marihuana-Erfahrung anzutreffen, unter den Studenten mit Alkohol-Erfahrung um das 20fache höher ist als bei den Studenten ohne Alkohol-Erfahrung. Mit Hilfe des Standardfehlers kann hierfür auch ein 95%-Konfidenzintervall aus $e^{2,98 \pm 1,96 \cdot 0,465}$ mit $[8, 0; 49, 3]$ bestimmt werden. Die Interpretation der beiden anderen Modellparameter erfolgt analog.

8.5.3.3 Einschränkungen und Hinweise zum log-linearen Modell

- (1) Der Einschluss mehrerer Faktoren in das loglineare Modell erschwert sehr schnell die Interpretation der Teilmodelle unter den verschiedenen Unabhängigkeitshypothesen.
- (2) In log-linearen Modellen wird Abhängigkeit (Assoziation) zwischen verschiedenen Faktoren untersucht. Eine „Rollverteilung“ nach Ziel- und Einflussgrößen wie beim linearen Modell ist nicht gegeben, muss aber ggf. im Rahmen einer inhaltlichen, sachlogisch begründeten, Interpretation der Modelle berücksichtigt werden.
- (3) Die Modellbildung im allgemeinen log-linearen Modell ist nur bei einer ausreichenden Fallzahl n möglich. Diese sollte mindestens das 5fache der in dem Modell betrachteten Zellen (Felder) N betragen.

$$n = \sum_{i=1}^N n_i > 5 \cdot N$$

Ist die Fallzahl nach dieser Regel nicht ausreichend, dann können Zellen (Kategorien in den Faktoren) zusammengelegt werden, oder es sind weniger Faktoren in dem Modell zu berücksichtigen.

- (4) Die unter den Modellannahmen erwarteten Häufigkeiten $\hat{\mu}_i$ sollten größer sein als 1.

$$\hat{\mu}_i > 1 \quad \text{für alle } i$$

und nicht mehr als 20% der Zellen sollten Erwartungswerte aufweisen, die kleiner als 5 sind.

- (5) Das Paket **exactLoglinTest** in **R** [Caf05] bietet die Möglichkeit einer Prüfung log-linearer Modelle mit einer Monte-Carlo-Simulation durch die Funktion **mcexact()**.

8.6 Analyse von Überlebenszeiten

- Kaplan-Meier Schätzung der Überlebensfunktion
- Der Logrank-Test
- Parametrische Modelle zu Überlebenszeiten
- Das Cox-Regressionsmodell

Unter dem Begriff „**Überlebenszeitanalyse**“ (survival analysis) wird eine Reihe statistischer Verfahren zur Modellierung und Bewertung von **Ereigniszeiten** zusammengefasst. Das Auftreten von definierten Ereignissen, z.B. der Tod eines Patienten oder das Auftreten einer Komplikation nach Behandlung, wird in Abhängigkeit von der Zeit und von anderen Faktoren (Kovariate) untersucht. Der Terminus „Überleben“ wird dabei in diesem Abschnitt als Synonym für das Ausbleiben eines Ereignisses verwendet. Die Verfahren sind jedoch auf zahlreiche vergleichbare Anwendungen übertragbar.

Ziel der Überlebenszeitanalyse sind

- Schätzen und Interpretation von Überlebens- und/oder Risikofunktionen.
- Vergleiche von Überlebensfunktionen zwischen verschiedenen Gruppen (Strata) oder verschiedenen Behandlungen.
- Nachweis und Aufstellen von Beziehungen zwischen der Überlebenszeit (Zielgröße) und ausgewählten erklärenden Faktoren (Einflussgrößen) im Rahmen eines prognostischen Modells.

Die Ereigniszeiten werden grundsätzlich auf einen festen Zeitpunkt (t_0) bezogen, z.B. das Datum der Diagnosestellung (Histologie) oder das Datum der Behandlung (Operation). Das Prinzip einer solchen zeitlichen Synchronisation wird durch Abbildung 8.13 deutlich.

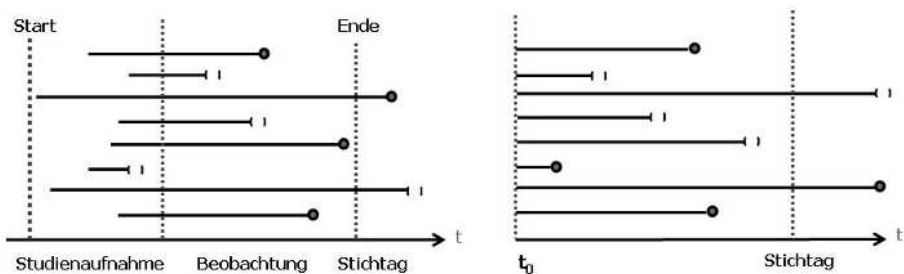


Abb. 8.13 Synchronisation von Ereigniszeiten von der kalendarischen Zeit auf einen festen Zeitpunkt (t_0). Ereignisse sind durch ausgefüllte Kreise markiert

Während in der kalendarischen Zeit die Aufnahme in die Studie oder Untersuchung (Rekrutierungsphase) und die Beobachtung der Fälle jeweils über einen festen Zeitraum erfolgen (linke

Seite in Abbildung 8.13), wird für die Datenanalyse die Beobachtungszeit auf einen festen Zeitpunkt bezogen (rechte Seite in Abbildung 8.13). Dabei können je nach Fragestellung auch Fälle von der Analyse ausgeschlossen werden, für die eine minimale Beobachtungszeit nicht eingehalten werden kann, z.B. bei intraoperativem Exitus oder bei einem Exitus während des stationären Aufenthaltes (Krankenhaus-Letalität).

Aus mathematischer Sicht wird eine Zufallsvariable T in einem Wertebereich $T > t_0 = 0$ betrachtet. In Übereinstimmung mit der Notation aus dem Abschnitt über Zufallsvariablen bezeichnet t somit eine spezielle Realisierung (Beobachtung) der Zufallsvariablen T . Die Verteilungsfunktion $F(t)$ dieser Zufallsvariablen (8.83) wird hier als „**Sterbefunktion**“ bezeichnet (mit der Dichtefunktion $f(t)$). Das entsprechende Komplement ist die entscheidende „**Überlebensfunktion**“ $S(t)$ (Survival function). Sie beschreibt die Wahrscheinlichkeit dafür, mindestens bis zum Zeitpunkt t zu überleben, bzw. dass der Tod (das Ereignis) erst nach dem Zeitpunkt t eintreten wird.

$$\begin{aligned} F(t) &= P(T \leq t) = \int_0^t f(x) dx \\ S(t) &= P(T > t) = 1 - F(t) = \int_t^\infty f(x) dx \end{aligned} \quad (8.83)$$

Dem **Erwartungswert** für T entspricht die Fläche unter der Überlebensfunktion $S(t)$, er wird als **mittlere Überlebenszeit** bezeichnet.

$$E[T] = \int_0^\infty S(t) dt \quad (8.84)$$

Entsprechend kann auch eine **mittlere verbleibende Überlebenszeit** für einen Zeitpunkt u in (8.85) durch die Fläche unter $S(t)$ für $t > u$ gekennzeichnet werden (normiert mit $S(u)$).

$$E[T - u | T > u] = \int_u^\infty S(t) dt / S(u) \quad (8.85)$$

Von Interesse ist auch die Wahrscheinlichkeit dafür, dass das Ereignis zu einem bestimmten Zeitpunkt auftritt, gegeben $T \geq t$. Die Dichte dieser Funktion wird in (8.86) definiert und heißt **Risikofunktion** $h(t)$ (hazard function).

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t < T \leq t + \Delta t | T \geq t)}{\Delta t} = \frac{f(t)}{S(t)} \quad (8.86)$$

Die Verwendung der bedingten Wahrscheinlichkeit in (8.86) bringt zum Ausdruck, dass das Ereignis nicht bereits früher eingetreten ist. Die Verteilungsfunktion für $h(t)$ aus (8.86) heißt **kumulierte Hazardfunktion** und wird mit $H(t)$ bezeichnet. Zwischen der Risikofunktion und der Überlebensfunktion kann die Beziehung (8.87) nachgewiesen werden, die insbesondere für die Betrachtungen im Rahmen der Modellbildung von zentraler Bedeutung ist.

$$S(t) = e^{-H(t)} \quad (8.87)$$

Typische Verläufe dieser drei Funktionen sind in Abbildung 8.14 skizziert. Die Sterbefunktion steigt von 0 auf den Wert 1 an (bei ausreichend langer Beobachtungszeit), entsprechend fällt die Überlebensfunktion von 1 auf den Wert 0 ab. Das Risiko für ein Ereignis kann über die Beobachtungszeit als konstant angenommen werden, unterliegt jedoch in der Regel zeitlichen Einflüssen. So kann häufig ein höheres Risiko zu Beginn und am Ende der Beobachtungszeit vorliegen, so

dass die Hazardfunktion durch einen typischen „wannenförmigen“ Verlauf charakterisiert wird (in der Technik Produktionsfehler zu Beginn und Materialermüdung am Ende der Beobachtung).

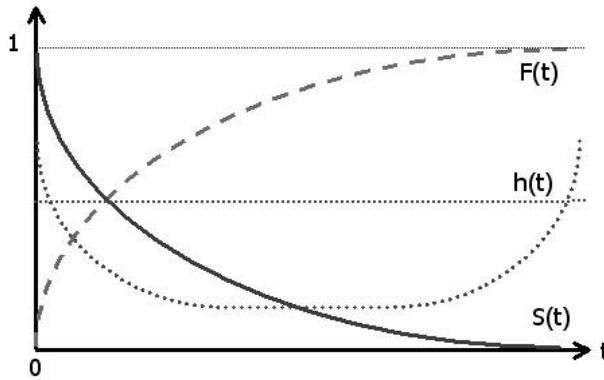


Abb. 8.14 Typische Verläufe der Sterbefunktion $F(t)$, Überlebensfunktion $S(t)$ und der Hazardfunktion $h(t)$ mit konstantem Risiko bzw. mit zeitabhängigem Risiko (Badewanne)

Diese Darstellung wird verständlich, wenn wir eine diskrete Schreibweise einführen. Auf der Basis von n beobachteten Ereignissen, die jeweils zum Zeitpunkt t_i ($i = 1, \dots, n$) eingetreten sind, folgt:

$$\begin{aligned}
 P(T = t_i) &= p_i \quad (i = 1, \dots, n) \\
 h_i &= P(T = t_i | T \geq t_i) = \frac{p_i}{\sum_{j=i}^n p_j} \\
 &= \frac{p_i}{1 - \sum_{j=1}^{i-1} p_j} \\
 \text{oder: } p_i &= h_i \left(1 - \sum_{j=1}^{i-1} p_j \right)
 \end{aligned}$$

Analog zu (8.83) und (8.86) können dann kumulierte Hazardfunktion und Überlebensfunktion für $t \geq 0$ diskret beschrieben werden.

$$\boxed{
 \begin{aligned}
 H(t) &= \sum_{i:t_i \leq t} h_i \\
 S(t) &= \prod_{i:t_i \leq t} (1 - h_i)
 \end{aligned}
 } \quad (8.88)$$

8.6.1 Kaplan-Meier Schätzung der Überlebensfunktion

Eine Schätzung der Überlebensfunktion kann mit der Zahl der Fälle, die eine bestimmte Zeitspanne überleben, berechnet werden.

$$\hat{S}(t) = \frac{\text{Anzahl der Fälle, die die Zeit } t \text{ überleben}}{\text{Gesamtzahl aller Fälle}}$$

Allerdings können im Rahmen einer Untersuchung oder Studie in der Regel nicht alle Ereignisse beobachtet werden. In diesem Fall spricht man von **Zensierungen**. Gründe für eine Zensierung sind:

- Die Studie endet, bevor das zu untersuchende Ereignis eintreten konnte. Die Beobachtungszeit wurde zu kurz gewählt.
- Ein Fall geht aus der Studie verloren, z.B. durch Umzug (lost to follow up).
- Ein Fall fällt aus der Studie aus anderen Gründen heraus, z.B. durch Tod bei Verkehrsunfall im Rahmen einer Therapiestudie oder bei Therapieabbruch wegen unerwünschter Nebenwirkungen.

Man unterscheidet die folgenden Arten einer Zensierung:

- (1) **Rechts-Zensierung**; ein Ereignis tritt zu einem unbekannten Zeitpunkt nach der Beobachtungszeit (Stichtag) ein.
- (2) **Links-Zensierung**; der Zeitpunkt eines bereits vor der Beobachtungszeit eingetretenen Ereignisses ist nicht bekannt.
- (3) **Intervall-Zensierung**; Informationen über das Eintreten des Ereignisses liegen nur für feste Zeitintervalle vor.

Die Frage an einen Schüler, 'Wann hast Du zum ersten Mal geraucht', wird mit Bezug auf das Datum der Geburt (t_0)

- nicht zensiert durch die Angabe des Alters beantwortet.
- rechts zensiert gewertet, wenn der Schüler noch nie geraucht hat.
- links zensiert beantwortet, wenn der Schüler schon einmal geraucht hat, aber unbekannt ist, wann zum ersten Mal geraucht wurde.

Das Problem der Zensierung wird auch am Beispiel der Analyse von Krankheitsdauern deutlich. Von einer Linkszensierung spricht man, wenn der Beginn einer Erkrankung nicht beobachtet werden kann, da dieser vor dem ersten Untersuchungszeitpunkt liegt. Rechtszensierung bedeutet, dass das Ende einer kontinuierlichen Krankheitsperiode nicht beobachtet werden kann, da es erst nach Ende des Untersuchungszeitraums auftritt.



Die folgenden Ausführungen beziehen sich grundsätzlich auf **rechts zensierte Ereigniszeiten**.



Beispiel: In einem klinischen Versuch wird Patienten mit einer Tumorerkrankung eine von zwei Chemotherapien C_1 oder C_2 zufällig zugeteilt. Es soll geprüft werden, ob das intensivierte Therapiekonzept C_2 gegenüber der Therapie C_1 zu einer verlängerten Überlebenszeit (in Tagen) führt. Die Ergebnisse sind in der Tabelle 8.10 zusammengestellt. Zensierte Angaben (hier Zeiträume, in denen das Ereignis sicher nicht eingetreten ist) sind mit einem + gekennzeichnet [HHR92].

Tabelle 8.10 Überlebenszeiten in Tagen bei zwei unterschiedlichen Tumorthapien ; zensierte Beobachtungen sind mit einem + gekennzeichnet

Chemotherapie C_1									
26 ⁺	50 ⁺	51 ⁺	57 ⁺	70 ⁺	93	105	108	135	193 ⁺
229 ⁺	241 ⁺	242	263	455 ⁺	489 ⁺	518	566 ⁺	582	595
Chemotherapie C_2									
4 ⁺	8 ⁺	10 ⁺	18 ⁺	30	55	56 ⁺	71 ⁺	89	90
101	148	155	207 ⁺	233	266 ⁺	283	441 ⁺		

Die Schätzung der Überlebenszeit nach dem Kaplan-Meier Verfahren erfolgt durch (8.89).

$$\hat{S}(t) = \prod_{i:t_{(i)} \leq t} \left(\frac{n_i - d_i}{n_i} \right) \quad (8.89)$$

Dabei bezeichnet $t_{(i)}$ die aufsteigend geordneten Ereigniszeiten, d_i die Zahl der Ereignisse zum Zeitpunkt $t_{(i)}$ und n_i die Zahl der bis zu diesem Zeitpunkt noch nicht eingetretenen Ereignisse, d.h. die Zahl der Fälle, die unter dem betrachteten Risiko zu diesem Zeitpunkt stehen. Nach diesem Ansatz fallen rechts zensierte Beobachtungen sukzessive heraus, werden also zum Zeitpunkt des nächst folgenden Ereignisses nicht mehr berücksichtigt.

In **R** kann die Kaplan-Meier Schätzung mit der Funktion **survfit()** aus **library(survival)** (Therneau [TT05]) berechnet werden. Dazu werden die Daten aus Tabelle 8.10 zunächst in einer Datenstruktur unter dem Namen „chemo“ gespeichert. Neben den Ereigniszeiten (in Tagen) muss der Status, hier Ereignis (=1) und Zensierung (=0), in einem zusätzlichen Merkmal gekennzeichnet werden. Diese Konvention in der Codierung sollte zur Vermeidung von Missverständnissen und Fehlern konsequent beibehalten werden. Aus den Daten wird mit der Funktion **Surv()** ein neues Objekt erzeugt, dass die Grundlage für einen Teil der folgenden Auswertungsschritte darstellt. Das Ergebnis der Funktion **survfit()** ist in Tabelle 8.11 zusammengefasst.

```
> t1 <- c(26, 50, 51, 57, 70, 93, 105, 108, 135, 193,
+        229, 241, 242, 263, 455, 489, 518, 566, 582, 595)
> z1 <- c(0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 0,
+        0, 0, 1, 1, 0, 0, 1, 0, 1, 1); c1 <- rep(1, length(t1))
> t2 <- c(4, 8, 10, 18, 30, 55, 56, 71, 89, 90,
+        101, 148, 155, 207, 233, 266, 283, 441)
> z2 <- c(0, 0, 0, 0, 1, 1, 0, 0, 1, 1,
+        1, 1, 1, 0, 1, 0, 1, 0); c2 <- rep(2, length(t2))
> gruppe <- c(c1, c2); zeit <- c(t1, t2); status <- c(z1, z2);
> chemo <- data.frame(gruppe, zeit, status); chemo
```

	gruppe	zeit	status
1	1	26	0
2	1	50	0
3	1	51	0
4	1	57	0
5	1	70	0
6	1	93	1
7	1	105	1
			

```
> library(survival)
> tab <- summary(survfit(Surv(zeit, status) ~ gruppe, data=chemo))
```

Die Varianz von $\hat{S}(t)$ kann klassisch auf der Basis der Formel von Greenwood [Gre26] nach (8.90) geschätzt werden.

$$\widehat{var}(\hat{S}(t)) = \hat{S}^2(t) \sum_{i:t_{(i)} \leq t} \frac{d_i}{n_i(n_i - d_i)} \quad (8.90)$$

Die Wurzel aus der Varianz führt zum Standardfehler der Schätzung ($\sqrt{\widehat{var}(\hat{S}(t))} = se(\hat{S}(t))$), mit dem auch ein allgemeines $(1 - \alpha)100\%$ -Konfidenzintervall nach (8.91) angegeben werden kann.

$$\hat{S}(t) \pm z_{1-\alpha/2} \cdot se(\hat{S}(t)) \quad (8.91)$$

Diese Rechnungen können in **R** ebenfalls mit der Funktion **survfit()** durchgeführt werden. Die Ergebnisse zu den Daten aus Tabelle 8.10 sind vollständig in Tabelle 8.11 wiedergegeben. Allerdings werden die Konfidenzintervalle hier nach einem von Kalbfleisch und Prentice [KP02] mo-

Tabelle 8.11 Schätzen der Überlebensfunktion nach Kaplan-Meier mit den Daten aus dem Beispiel in Tabelle 8.10

Chemotherapie C_1						
Zeit: t_i	unter Risiko: n_i	Ereignis: d_i	Überleben: $\hat{S}(t_i)$	Standardfehler	95% KI	
93	15	1	0,933	0,064	0,815	1,000
105	14	1	0,867	0,088	0,711	1,000
108	13	1	0,800	0,103	0,621	1,000
135	12	1	0,733	0,114	0,541	0,995
242	8	1	0,642	0,132	0,429	0,959
263	7	1	0,550	0,141	0,333	0,910
518	4	1	0,413	0,159	0,194	0,880
582	2	1	0,206	0,166	0,043	1,000
595	1	1	0,000			

Chemotherapie C_2						
Zeit: t_i	unter Risiko: n_i	Ereignis: d_i	Überleben: $\hat{S}(t_i)$	Standardfehler	95% KI	
30	14	1	0,929	0,069	0,803	1,000
55	13	1	0,857	0,094	0,692	1,000
89	10	1	0,771	0,117	0,573	1,000
90	9	1	0,686	0,132	0,471	0,999
101	8	1	0,600	0,140	0,379	0,949
148	7	1	0,514	0,144	0,297	0,891
155	6	1	0,429	0,143	0,223	0,826
233	4	1	0,321	0,142	0,135	0,764
283	2	1	0,161	0,134	0,031	0,824

difizierten Verfahren auf der Grundlage einer Logarithmustransformation bestimmt, welche das Konfidenzintervall auf den Bereich zwischen 0 und 1 begrenzt:

$$\exp(\log \hat{S}(t) \pm 1,96 \cdot se(\hat{H}(t)))$$

Hinweis: Mit dem zusätzlichen Argument **conf.type="plain"** können mit der Funktion **survfit()** die Konfidenzgrenzen auch nach (8.90) berechnet werden.

Die Schätzung der kumulierten Hazardfunktion $H(t)$ kann nach dem Kaplan-Meier Verfahren analog zu (8.89) durch (8.92) erfolgen.

$$\boxed{\begin{aligned} \hat{H}(t) &= -\log \prod_{i:t_i \leq t} \left(\frac{n_i - d_i}{n_i} \right) \\ \widehat{var}(\hat{H}(t)) &= \sum_{i:t(i) \leq t} \frac{d_i}{n_i(n_i - d_i)} \end{aligned}} \quad (8.92)$$

Graphisch wird die Schätzung der Überlebensfunktion in einer Treppenfunktion dargestellt. Beginnend mit $\hat{S}(t_0) = 1$ zeigt diese Funktion eine Treppenstufe zu jedem Zeitpunkt eines Ereignisses. Zwischenzeitliche Zensierungen werden markiert, zeigen allerdings keinen Einfluss auf den

Verlauf der Funktion. In Abbildung 8.15 sind die Überlebensfunktionen für die Daten aus Tabelle 8.10 dargestellt, einschließlich des 95%-Konfidenzbereichs. (Dazu kann in **R** einfach die Funktion `plot()` mit dem durch `survfit()` erzeugten Objekt verwendet werden.)

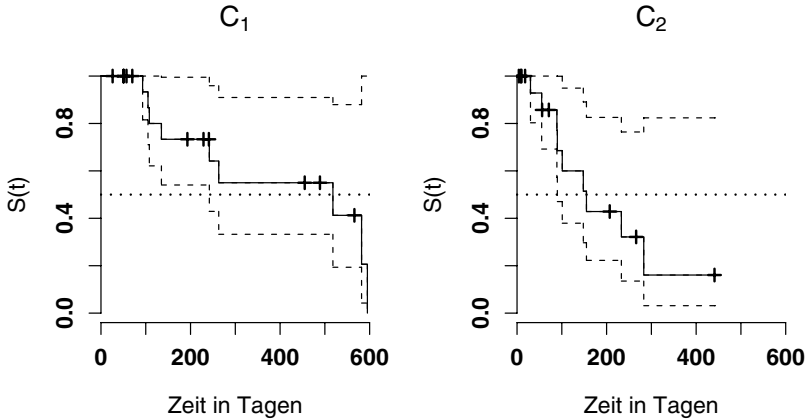


Abb. 8.15 Schätzung der Überlebensfunktionen nach dem Kaplan-Meier Verfahren für die Daten aus Tabelle 8.10

Auf der Grundlage der Kaplan-Meier Schätzung (8.89) können auch Quantile $\hat{t}_q$ für die Überlebenszeit nach (8.93) bestimmt werden, speziell der Wert für das „**mediane Überleben**“ - Median der Überlebenszeiten - $\hat{t}_{0,5}$ (median survival). Dieser Wert kann auch aus den Kurven in Abbildung 8.15 durch den Schnittpunkt einer horizontalen, zur Zeitachse parallelen, Linie zum Wert $\hat{S}(t) = 0,5$ abgelesen werden.

$$\hat{t}_q = \min\{t_i : \hat{S}(t_i) \leq 1 - q\} \quad (8.93)$$

Der Medianwert für die Überlebenszeit ist größer als die maximale Beobachtungszeit, wenn die Überlebensfunktion nicht geschnitten wird ($\hat{S}(t) > 0,5$ für alle t). Die Schnittpunkte mit den Konfidenzgrenzen der Überlebensfunktion können entsprechend als Konfidenzintervall für die mediane Überlebenszeit interpretiert werden.

Die Schätzung für die „**mittlere Überlebenszeit**“ (mean survival) $\hat{T}$ setzt im Prinzip voraus, dass die letzte Beobachtung $t_{(m)}$ nicht zensiert ist. Nur dann ist die Fläche unter der Kurve begrenzt und kann nach (8.94) bestimmt werden.

$$\hat{T} = \sum_{i=1}^m \hat{S}(t_{(i-1)})(t_{(i)} - t_{(i-1)})$$

mit $\hat{S}(t_{(0)}) = 1$ und den beobachteten Ereignissen
 $t_{(1)} \leq t_{(2)} \leq \dots \leq t_{(m)}$

(8.94)

Ist die letzte Beobachtung zensiert ($t_n^+ > t_{(m)}$), dann kann die mittlere Überlebenszeit beschränkt auf die Zeit vor der letzten Zensierung (restricted mean survival) geschätzt werden.

$$\hat{T}_r = \hat{T} + \hat{S}(t_{(m)})(t_n^+ - t_{(m)}) \quad (8.95)$$

Zur Schätzung der Varianzen dieser Größen vgl. auch Andersen, Borgan, Gill und Keiding [ABGK93]. In **R** kann die Berechnung der medianen und der mittleren Überlebenszeit durch die

Funktion **print()** angezeigt werden. Dazu muss allerdings der zusätzliche Parameter „**show.rmean=TRUE**“ angegeben werden.

```
> print(fit, show.rmean=TRUE)
Call: survfit(formula = Surv(zeit, status) ~ gruppe, data = chemo)

      n events rmean se(rmean) median 0.95LCL 0.95UCL
gruppe=1 20      9   390    57.0    518    242     Inf
gruppe=2 18      9   197    40.2    155     90     Inf
```

Anmerkungen zum Kaplan-Meier Verfahren

- (1) Für den Fall, dass zensierte und nicht zensierte Beobachtungen zusammen auftreten, wird bei der Kaplan-Meier Schätzung unterstellt, dass die zensierten Beobachtungen den tatsächlich beobachteten Ereignissen zeitlich folgen.
- (2) Ist die längste Beobachtungszeit zensiert, dann verbleibt der Kaplan-Meier Schätzer auf dem Niveau der Schätzung zur letzten nicht zensierten Beobachtung, insbesondere erreicht die Überlebensfunktion dann nicht den Wert 0. Konsequenter Weise sollte die Überlebenskurve dann auch nur bis zum letzten beobachteten Ereignis diskutiert werden.
- (3) Der Verlauf von $\log(\hat{S}(t))$ nach der Zeit im Koordinatensystem ist für exponentiell verteilte Überlebenszeiten linear. Der Verlauf von $\log(-\log(\hat{S}(t)))$ nach $\log(t)$ ist näherungsweise linear, wenn für die Überlebenszeiten das Modell einer Weibull-Verteilung angenommen werden kann (vgl. auch Abschnitt [8.6.3]).
- (4) Auch wenn die Schätzung für das kumulierte Risiko $\hat{H}(t) = -\log(\hat{S}(t))$ nach dem Kaplan-Meier Schätzer direkt übernommen werden kann, wird hierfür häufig eine alternative Schätzung nach dem **Nelson-Aalen Schätzer** (Nelson [Nel72]) empfohlen.

$$\hat{H}^*(t) = \sum_{t_{(i)} \leq t} \frac{d_i}{r_i} \quad (8.96)$$

Die beiden Schätzungen unterscheiden sich in der Regel nur sehr wenig. Während der Kaplan-Meier Schätzer bevorzugt für die Überlebensfunktion verwendet wird, wird der Nelson-Aalen Schätzer häufig für die Schätzung des kumulierten Risikos eingesetzt. Eine daraus abgeleitete Schätzung für die Überlebenszeiten ist $\hat{S}^*(t) = \exp(-\hat{H}^*(t))$ (**Fleming-Harrington** Schätzer).

8.6.2 Der Logrank-Test

Ein Vergleich der Überlebenszeiten zweier unabhängiger Stichproben kann mit dem Logrank-Test erfolgen. Dabei werden die Überlebenszeiten als Realisierungen zweier Zufallsvariablen T_1 und T_2 aufgefasst, deren Verteilungsfunktionen (Überlebensfunktionen) zu vergleichen sind. Ausgehend von den in beiden Stichproben gemeinsam beobachteten Ereigniszeiten

$$t_{(1)} \leq t_{(2)} \leq \dots \leq t_{(m)} \quad ,$$

die aufsteigend sortiert vorliegen, kann zu jedem Zeitpunkt, zu dem mindestens ein Ereignis auftrat, die folgende Tabelle erstellt werden.

Gruppe	Ereignisse $t_{(i)}$	unter Risiko
1	d_{1i}	n_{1i}
2	d_{2i}	n_{2i}
Summe	d_i	n_i

Dabei steht d_i für die Zahl der Ereignisse zum Zeitpunkt $t_{(i)}$, und n_i für die entsprechende Anzahl der Personen oder Fälle, die zu diesem Zeitpunkt dem Risiko des Ereignisses unterliegen (getrennt nach den beiden Stichproben 1 und 2).

Unter der Annahme, dass es keinen Unterschied in den Überlebenszeiten zwischen den beiden Stichproben zugrundeliegenden Gesamtheiten gibt, kann dann eine erwartete Anzahl von Ereignissen zum Zeitpunkt $t_{(i)}$ nach (8.97) geschätzt werden.

$$\hat{e}_{1i} = \frac{d_i}{n_i} n_{1i} \quad \text{bzw.} \quad \hat{e}_{2i} = \frac{d_i}{n_i} n_{2i} \quad (8.97)$$

Die Varianz unter der Nullhypothese kann nach dem Modell einer hypergeometrischen Verteilung nach (8.98) geschätzt werden.

$$\hat{v}_{1i} = \frac{n_{1i} n_{2i} d_i (n_i - d_i)}{n_i^2 (n_i - 1)} = \hat{v}_{2i} \quad (8.98)$$

Für die Berechnung einer einfachen Teststatistik zu dieser Hypothesenstellung können dann die Unterschiede zwischen den erwarteten (8.97) und beobachteten Häufigkeiten benutzt werden.

$$\hat{\chi}^2 = \frac{\sum \hat{e}_{1i} - \sum d_{1i}}{\sum \hat{e}_{1i}} + \frac{\sum \hat{e}_{2i} - \sum d_{2i}}{\sum \hat{e}_{2i}} \quad (8.99)$$

Unter der Nullhypothese H_0 , dass die Verteilung der Überlebenszeiten in den beiden Gesamtheiten gleich ist, sowie unter der Annahme, dass der Mechanismus für die Zensierungen in beiden Gruppen gleich wirkt, ist die Teststatistik (8.99) angenähert χ^2 -verteilt mit einem Freiheitsgrad und kann entsprechend mit den Quantilen der χ^2 -Verteilung geprüft werden.

Praktisch handelt es sich jedoch um die Prüfung von m Vierfeldertafeln, für die nach Mantel und Haenszel [Man66] eine geeignete Teststatistik (8.100) angegeben werden kann (vgl. auch Hosmer und Lemeshow [HL99]). Der Wert der Teststatistik (Logrank-Test nach Mantel und Haenszel) ist unter der Nullhypothese asymptotisch χ^2 -verteilt mit einem Freiheitsgrad.

$$\hat{\chi}^2 = \frac{\left[\sum_{i=1}^m (d_{1i} - \hat{e}_{1i}) \right]^2}{\sum_{i=1}^m \hat{v}_{1i}} \quad (8.100)$$

In **R** kann der Vergleich zwischen den Überlebenskurven mit der Funktion **survdifff()** durchgeführt werden.

```
> survdiff(Surv(zeit, status) ~ gruppe, data=chemo, rho=0) # log-rank-Test
```

Call:

```
survdifff(formula = Surv(zeit, status) ~ gruppe, data = chemo,
rho = 0)
```

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
gruppe=1	20	9	12.73	1.09	4.12
gruppe=2	18	9	5.27	2.64	4.12

Chisq= 4.1 on 1 degrees of freedom, p= 0.0424

Der Anzahl von jeweils 9 in den Gruppen aus dem Beispiel beobachteten Ereignissen stehen unter der Nullhypothese 12,7 bzw. 5,3 „erwartete“ Ereignisse gegenüber. Die Funktion **survdifff()** berechnet die Statistiken nach (8.99) und (8.100) und gibt den P-Wert nach der Mantel-Haenszel

Statistik an ($P=0,04$). Die beiden zugehörigen Kurven sind in Abbildung 8.16 zusammen dargestellt.

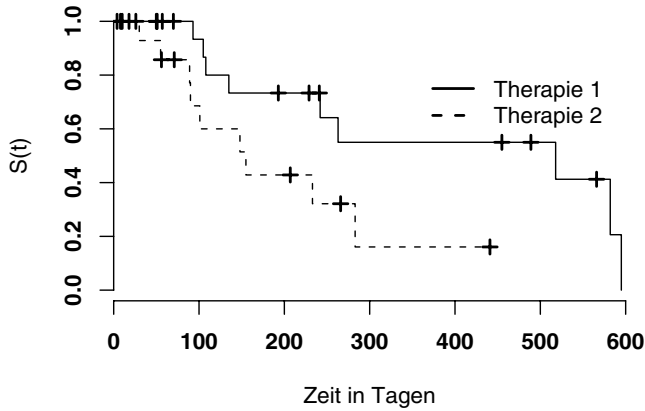


Abb. 8.16 Überlebensfunktionen nach dem Kaplan-Meier Verfahren für die Daten der Tabelle 8.10



Für einen Vergleich von Überlebenszeiten nach dem Logrank-Test dürfen sich die Kurven zu den Überlebensfunktionen $\hat{S}(t)$ (vgl. Abbildung 8.16) **nicht überschneiden**. Die Frage, in welchem Maß die Form der Kurven und unterschiedliche Zeitpunkte für die Zensierungen das Ergebnis der Teststatistik beeinflussen, ist von zahlreichen Autoren untersucht worden. Eine Zusammenfassung verschiedener Möglichkeiten einer Gewichtung der Teststatistik (8.100) geben Andersen, Borgan, Gill und Keiding [ABGK93] sowie auch Hosmer und Lemeshow [HL99]. So führt zum Beispiel eine Gewichtung mit der Zahl der Beobachtungen n_i in der Teststatistik dazu, dass früher auftretenden Ereignissen gegenüber späteren Ereignissen eine größere Bedeutung zukommt (Breslow-Test). Eine Verallgemeinerung des Ansatzes ermöglicht auch den Vergleich mehrerer ($k > 2$) Stichproben.

8.6.3 Parametrische Modelle für Überlebenszeiten

Die Beschreibung und statistische Bewertung von Überlebenszeiten kann auch auf der Grundlage „parametrischer“ Modelle erfolgen. Einen einfachen Ansatz bietet die **Exponentialverteilung** mit nur einem Parameter λ (vgl. auch Kapitel [5.4.4]).

$$\begin{aligned} f(t) &= \lambda \exp(-\lambda t) \\ S(t) &= \exp(-\lambda t) \end{aligned} \quad (8.101)$$

Der Parameter λ kennzeichnet die mittlere Überlebensdauer, also insbesondere die Steilheit der Überlebenskurve. Erwartungswert und Varianz der Zufallsvariablen T (Zeitspanne des Überlebens) sind durch λ nach (8.102) bestimmt.

$$\mu = E[T] = \frac{1}{\lambda} \quad \text{und} \quad \text{Var}[T] = \frac{1}{\lambda^2} \quad (8.102)$$

Das exponentielle Überlebenszeit-Modell geht (vereinfachend) von einem **konstanten Risiko** aus. Die Risikofunktion kann, unabhängig von der Zeit, nach (8.103) hergeleitet werden.

$$h(t) = \lambda \quad \text{bzw.} \quad H(T) = \log(\lambda) + \log(t) \quad (8.103)$$

Der Parameter λ kann durch einen Mittelwert auch aus zensierten Daten geschätzt werden, wobei über die Verteilung der Zensierungszeiten keine zusätzlichen Annahmen gemacht werden. Ausgehend von den ansteigend geordneten Zeiten zu m Ereignissen

$$t_{(1)} \leq t_{(2)} \leq \dots \leq t_{(m)}$$

und insgesamt $n - m$ zensierten Beobachtungen

$$t_1^+, t_2^+, \dots, t_{n-m}^+$$

kann die Likelihood-Funktion nach (8.104) geschrieben werden (Lee [LW03]).

$$L(\lambda) = \prod_{i=1}^m \lambda \exp(-\lambda t_{(i)}) \prod_{j=1}^{n-m} \exp(-\lambda t_j^+) \quad (8.104)$$

Eine Lösung nach dem unbekannten Parameter λ mit dem Maximum-Likelihood-Verfahren führt auf (8.105).

$$\hat{\mu} = \frac{1}{\hat{\lambda}} = \frac{1}{m} \left[\sum_{i=1}^m t_{(i)} + \sum_{j=1}^{n-m} t_j^+ \right] \quad (8.105)$$

Eine entsprechende Schätzung für die Varianz von $\hat{\mu}$ bzw. $\hat{\lambda}$ kann nach (8.106) bestimmt werden.

$$\text{Var}[\hat{\mu}] = \frac{1}{m} \hat{\mu}^2 \quad \text{bzw.} \quad \text{Var}[\hat{\lambda}] = \frac{1}{m} \hat{\lambda}^2 \quad (8.106)$$

Auf der Grundlage von (8.105) und (8.106) kann auch ein angenähertes $(1 - \alpha)$ **100%-Konfidenzintervall** angegeben werden.

$$\hat{\mu} - z_{1-\alpha/2} \sqrt{\frac{1}{m} \hat{\mu}^2} \leq \mu \leq \hat{\mu} + z_{1-\alpha/2} \sqrt{\frac{1}{m} \hat{\mu}^2} \quad (8.107)$$

Für die Chemotherapie-Gruppe C_1 aus dem einführenden Beispiel wird für $m = 9$ beobachtete Ereignisse der Erwartungswert (mittlere Überlebenszeit) mit $\hat{\mu} = 562,1$ Tagen geschätzt ($\hat{\lambda} = 0,0018$). Die Überlebensfunktion ist in Abbildung 8.17 dargestellt. Der Standardfehler zu dieser Schätzung ist nach (8.106) 187,7 Tage (0,0006). Auch eine **Schätzung von Quantilen** ist in diesem Modell relativ einfach möglich.

$$\hat{t}_p = -\frac{1}{\hat{\lambda}} \log(1 - p) \quad (8.108)$$

So folgt zum Beispiel aus (8.108) der **Medianwert** für das Überleben nach der Chemotherapie-Gruppe C_1 :

$$\hat{t}_{0,5} = -\frac{\log(0,5)}{\log(0,0018)} = 385,1 \quad \text{Tage.}$$

Die Annahme einer konstanten Ausfallrate mit der Exponentialverteilung über die Beobachtungszeit ist unrealistisch. In dem Modell der **Weibull-Verteilung** (8.109) wird ein zusätzlicher Parameter α eingeführt, der ein mit der Zeit (monoton) wachsendes oder abnehmendes Risiko kennzeichnet (vgl. auch Kapitel [5.4.5]).



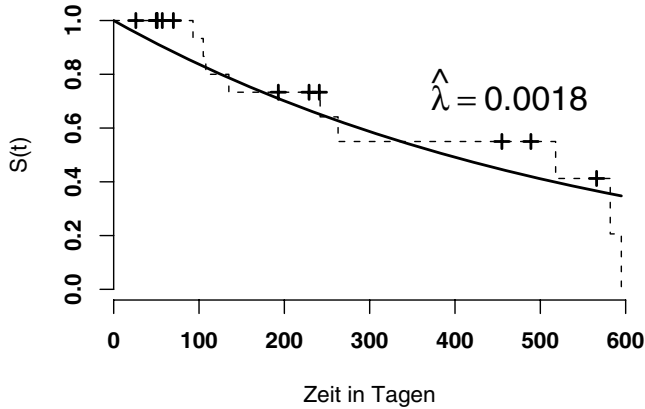


Abb. 8.17 Schätzung einer exponentiellen Überlebensfunktion zu den Beispieldaten der Therapiegruppe 1; die Kaplan-Meier Schätzung ist gestrichelt dargestellt

$$\begin{aligned} f(t) &= \alpha t^{\alpha-1} \lambda^\alpha \exp((-\lambda t)^\alpha) \\ S(t) &= \exp((-\lambda t)^\alpha) \\ h(t) &= \alpha \lambda^\alpha (t)^{\alpha-1} \end{aligned} \quad (8.109)$$

Neben dem Skalenparameter (scale) λ bezeichnet $1/\alpha$ den sogenannten „Formparameter“ (shape) der Weibull-Verteilung. Dieser legt die „Ausfallsteilheit der Risikofunktion“ $h(t)$ fest. Für $\alpha = 1$ erhält man wiederum das Modell der Exponentialverteilung, d.h. (8.109) und (8.101) sind identisch. Typische Werte für α liegen im Bereich von 0,25 bis 5. Die charakteristische Lebensdauer wird im Modell der Weibull-Verteilung auch durch den Parameter λ bestimmt.

Näheres zur Definition und Schätzung der Parameter der Weibull-Verteilung ist im Abschnitt [5.4.5] dargestellt. In **R** kann die Schätzung der Parameter zu den Modellen mit der Funktion **survreg()** durchgeführt werden. Mit den Daten aus Tabelle 8.10 (hier nur für Chemotherapie 1) folgt:

```
> summary(survreg(Surv(zeit, status) ~ 1, dist="exponential", data=chemol))
...
      Value Std. Error z      p
(Intercept)  6.33      0.333 19 1.69e-80
...

> summary(survreg(Surv(zeit, status) ~ 1, dist="weibull", data=chemol))
...
      Value Std. Error z      p
(Intercept)  6.192      0.211 29.34 3.31e-189
Log(scale)  -0.473      0.260 -1.82 6.93e-02
```

Die Schätzung für das exponentielle Modell führt hier zu dem Erwartungswert $\hat{\mu} = \exp(6,33) = 561,2$ Tagen. Für das Weibull-Modell sind die Schätzungen $\hat{\mu} = \exp(6,192) = 488,8$ Tage mit $1/\hat{\alpha} = \exp(-0,473) = 0,623$ ($\hat{\alpha} = 1,605$). Das Ergebnis der Schätzung nach dem Weibull-Modell ist in Abbildung 8.18 dargestellt.

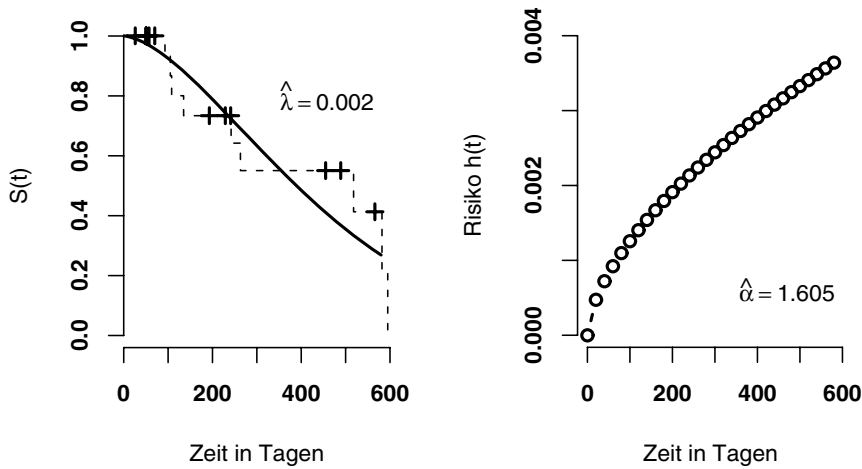


Abb. 8.18 Schätzung der Überlebensfunktion (links) und der Risikofunktion (rechts) nach dem Modell einer Weibull-Verteilung für die Daten der Chemotherapie C_1

8.6.4 Das Cox-Regressionsmodell

8.6.4.1 Regressionsmodelle zu Überlebenszeiten

Die Methode der multiplen Regression ist ein Verfahren zur Analyse des Zusammenhangs zwischen der Überlebenszeit und Faktoren, die möglicherweise prognostisch relevant sind, d.h. die Einfluss auf die Überlebenszeit haben könnten.

Sind $T_1, T_2, \dots, T_n$ Zufallsvariablen, die für die Überlebenszeiten (allgemeiner: Ereigniszeiten) von n Individuen stehen, und ist $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})'$ $i = 1, \dots, n$ ein Vektor mit den beobachteten Werten zu insgesamt p Einflussgrößen, dann kann ein allgemeines Regressionsmodell nach (8.110) formuliert werden.

$$\log(T_i) = \mathbf{x}_i' \boldsymbol{\beta} + \epsilon_i \quad (8.110)$$

Darin ist $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$ ein Vektor mit Regressionskoeffizienten, der zusammen mit $\mathbf{x}_i$ den systematischen Teil des Modells bestimmt, während die zufällige Komponente durch ϵ_i erfasst wird. Eine Transformation mit der Exponentialfunktion führt auf

$$T_i = \exp(\mathbf{x}_i' \boldsymbol{\beta}) \eta_i \quad (8.111)$$

Unter der Annahme, dass die ϵ_i unabhängig und identisch normalverteilt sind, folgt, dass $\eta_i = \exp(\epsilon_i) = T_i \exp(\mathbf{x}_i' \boldsymbol{\beta})$ lognormal-verteilt sind (häufig wird hierfür auch das Modell einer Weibull-Verteilung angenommen).

Die Verteilung der η_i ist die Basis für ein Regressionsmodell zu den Überlebenszeiten (baseline survival distribution). Mit Bezug auf diese Basisverteilung, die durch die Dichte f_0 , die Verteilungsfunktion F_0 (Überleben $S_0 = 1 - F_0$) und das Risiko h_0 gekennzeichnet ist, kann für jedes Individuum das Überleben in Abhängigkeit von den Einflussgrößen nach (8.112) beschrieben werden.

$$\begin{aligned}
f(t_i) &= \exp(-\mathbf{x}'_i \boldsymbol{\beta}) \cdot f_0(\exp[-\mathbf{x}'_i \boldsymbol{\beta}] t_i) \\
S(t_i) &= S_0(\exp[-\mathbf{x}'_i \boldsymbol{\beta}] t_i) \\
h(t_i) &= \exp(-\mathbf{x}'_i \boldsymbol{\beta}) \cdot h_0(\exp[-\mathbf{x}'_i \boldsymbol{\beta}] t_i)
\end{aligned}
\tag{8.112}$$

Die Idee dieses Modellansatzes ist am einfachsten noch hinsichtlich der Überlebensfunktion einsehbar, in der der Term $\exp[-\mathbf{x}'_i \boldsymbol{\beta}]$ als Faktor in einer Exponentialfunktion (S_0) auftritt, mit dem ein Effekt der Einflussgrößen zusammengefasst wird.

Hinweis: Unter der Annahme, dass S_0 mit dem Modell einer Weibull-Verteilung und dem Parameter $\lambda = 1$ beschrieben werden kann, lässt sich zeigen, dass auch die Verteilung der T_i nach Weibull verteilt sind, wobei dann der Parameter λ durch die Komponente $\exp[-\mathbf{x}'_i \boldsymbol{\beta}]$ ersetzt wird (Weibull accelerated life model).

8.6.4.2 Das Proportional-Hazards Modell von Cox

Der Modellansatz nach Cox [Cox72] basiert auf der Annahme proportionaler Risikofunktionen:

$$h(t_i) = \exp(-\mathbf{x}'_i \boldsymbol{\beta}) \cdot h_0(t_i)$$

Ist das **Basisrisiko** $h_0(t)$ zum Beispiel Weibull-verteilt (ohne Beschränkung der Allgemeingültigkeit dieser Überlegung zur Modellbildung), dann lässt sich unter der Annahme proportionaler Risiken zeigen, dass auch die Risikofunktion $h(t)$ durch eine Weibull-Verteilung mit einem modifizierten Parameter $\tilde{\lambda}$ beschrieben werden.

$$\begin{aligned}
h(t_i) &= \exp(-\mathbf{x}'_i \boldsymbol{\beta}) \cdot \alpha \lambda^\alpha t_i^{\alpha-1} \\
&= \alpha \underbrace{(\lambda (\exp(-\mathbf{x}'_i \boldsymbol{\beta}))^{1/\alpha})^\alpha}_{\tilde{\lambda}^\alpha} t_i^{\alpha-1} \\
&= \alpha \tilde{\lambda}^\alpha t_i^{\alpha-1}
\end{aligned}$$

Insbesondere folgt daraus für die Überlebensfunktion $S(t)$ und für die kumulierte Risikofunktion $H(t)$:

$$\begin{aligned}
S(t_i) &= \exp(-(\tilde{\lambda} t_i)^\alpha) \\
\log(H(t_i)) &= \alpha \log(\tilde{\lambda}) + \alpha \log(t_i) \\
&= \underbrace{\alpha \log(\lambda) + \alpha \log(t_i)}_{\log(H_0(t_i))} + \mathbf{x}'_i \boldsymbol{\beta} \\
&= \log(H_0(t_i)) + \mathbf{x}'_i \boldsymbol{\beta}
\end{aligned}
\tag{8.113}$$

Der Effekt, den die Einflussgrößen auf die Überlebenszeiten haben, kann somit nach (8.113) durch eine spezielle Linkfunktion (8.114) in der Terminologie der verallgemeinerten linearen Modelle ausgedrückt werden. $H_0(t)$ kennzeichnet darin ein nicht näher spezifiziertes Basisrisiko, in dem die Einflussgrößen nicht berücksichtigt sind.

$$\log\left(\frac{H(t_i)}{H_0(t_i)}\right) = \mathbf{x}'_i \boldsymbol{\beta} \tag{8.114}$$

8.6.4.3 Schätzen der Parameter im PH-Modell

Das Proportional-Hazard Modell (PH-Modell) zielt auf die Schätzung der Regressionskoeffizienten β , ohne dass zusätzliche Annahmen zu einem Basisrisiko gemacht werden müssen. Cox [Cox75] schlägt dazu ein modifiziertes Maximum-Likelihood-Verfahren vor, in dem die Likelihoodfunktion hinsichtlich des Risikos auf bedingte Wahrscheinlichkeiten für die Ereignisse zu den entsprechenden Zeitpunkten bezogen wird (**partial likelihood estimation**).

Sind $t_{(1)} \leq t_{(2)} \leq \dots \leq t_{(m)}$ die geordneten Ereigniszeiten, dann kann die Wahrscheinlichkeit für das Ereignis für ein einzelnes Individuum zum Zeitpunkt $t_{(i)}$, bedingt auf alle zu diesem Zeitpunkt noch unter dem Risiko stehenden Individuen, durch den Quotienten (8.115) ausgedrückt werden.

$$\frac{\exp(\mathbf{x}'_i \beta)}{\sum_{t_{(j)} \geq t_i} \exp(\mathbf{x}'_j \beta)} \quad (8.115)$$

Jedes Ereignis trägt somit einen bestimmten Faktor zur Likelihood bei. Die partielle Likelihoodfunktion für die Regressionskoeffizienten (8.116) ist insbesondere unabhängig von dem Basisrisiko $h_0(t)$.

$$L(\beta) = \prod_{i=1}^m \frac{\exp(\mathbf{x}'_i \beta)}{\sum_{t_{(j)} \geq t_i} \exp(\mathbf{x}'_j \beta)} \quad (8.116)$$

Eine Maximum-Likelihood Lösung nach β ist analytisch nicht möglich. Für die numerische Lösung stehen zahlreiche Statistikprogramme zur Verfügung. Dabei sind besondere Korrekturen erforderlich, wenn Bindungen in den beobachteten Ereigniszeiten auftreten, z.B. nach Breslow [Bre75] oder nach Efron [Efr77]. In **R** erfolgt die Berechnung des PH-Modells nach Cox mit der Funktion **coxph()** aus **library(survival)**.

Beispiel (Ovarial-Karzinom): In einer Studie zur Behandlung des Ovarial-Karzinoms wurden 26 Patientinnen nach zwei unterschiedlichen Therapien (rx) behandelt (vgl. Tabelle 8.12). Neben der Behandlung sollen auch das Alter (age), eine nach der Therapie bestehende Resterkrankung (resid.ds) und die Beurteilung der Leistungsfähigkeit der Patientinnen auf der Grundlage des ECOG-Scores [Oke82] als weitere Einflussgrößen auf die Überlebenszeit untersucht werden.

```
> library(survival)
Loading required package: splines
> data(ovarian)
> fit <- coxph( Surv(futime, fustat) ~ age + rx + resid.ds + ecog.ps, ovarian )
> summary(fit)
Call:
coxph(formula = Surv(futime, fustat) ~ age + rx + resid.ds
      + ecog.ps, data = ovarian)

n = 26
```

	coef	exp(coef)	se(coef)	z	p
age	0.125	1.133	0.0469	2.662	0.0078
rx	-0.914	0.401	0.6533	-1.400	0.1600
resid.ds	0.826	2.285	0.7896	1.046	0.3000
ecog.ps	0.336	1.400	0.6439	0.522	0.6000
...					

Die statistische Prüfung und Interpretation der aus diesem Modell geschätzten Regressionskoeffizienten wird in den folgenden Abschnitten ausführlich dargestellt.



Tabelle 8.12 Daten zur Therapie des Ovarial-Karzinoms; Überlebenszeit (time), Status zur Zensierung (status), Alter (age), Resterkrankung (resid.ds), Behandlung (rx), und ECOG-Status (ecog.ps)

nr	time	status	age	resid.ds	rx	ecog.ps
1	59	1	72,33	2	1	1
2	115	1	74,49	2	1	1
3	156	1	66,47	2	1	2
4	421	0	53,36	2	2	1
5	431	1	50,34	2	1	1
6	448	0	56,43	1	1	2
7	464	1	56,94	2	2	2
8	475	1	59,85	2	2	2
9	477	0	64,18	2	1	1
10	563	1	55,18	1	2	2
11	638	1	56,76	1	1	2
12	744	0	50,11	1	2	1
13	769	0	59,63	2	2	2
14	770	0	57,05	2	2	1
15	803	0	39,27	1	1	1
16	855	0	43,12	1	1	2
17	1040	0	38,89	2	1	2
18	1106	0	44,60	1	1	1
19	1129	0	53,91	1	2	1
20	1206	0	44,21	2	2	1
21	1227	0	59,59	1	2	2
22	268	1	74,50	2	1	2
23	329	1	43,14	2	1	1
24	353	1	63,22	1	2	2
25	365	1	64,42	2	2	1
26	377	0	58,31	1	2	1

Hinweise zur Schätzung im Modell der Cox-Regression:

- Die Schätzung der Koeffizienten im Modell der Cox-Regression „maximiert“ den Logarithmus der partiellen Likelihoodfunktion (8.116).
- Die Schätzung basiert letztlich auf einer Rangfolge der Ereigniszeiten, nicht auf den beobachteten Zeiten selbst; in diesem Sinn wird das Verfahren der Cox-Regression auch als „nicht parametrisch“ bezeichnet.
- Die Bewertung des Einflusses von Kovariaten (Faktoren) auf das beobachtete Ereignis kann im Modell der Cox-Regression ohne die Kenntnis des zugrundeliegenden „Basisrisikos“ ($h_0(t)$) erfolgen. Allerdings wird angenommen, dass sich dieser Effekt proportional zu dem Basisrisiko auswirkt.
- Zensierte Beobachtungen leisten keinen Beitrag im Zähler der Likelihoodfunktion (8.116). Sie werden im Nenner durch die Bildung der Summe über alle Fälle unter Risiko berücksichtigt (risk set; z.B. Individuen, die einen bestimmten Zeitpunkt erlebt haben und später versterben).

8.6.4.4 Interpretation der Parameter

Eine besondere Bedeutung kommt dem Modellansatz nach Cox hinsichtlich der Interpretation der Regressionskoeffizienten zu. Für zwei unterschiedliche Beobachtungen in den Einflussgrößen $\mathbf{x} = (x_1, \dots, x_p)'$ und $\mathbf{y} = (y_1, \dots, y_p)'$ kann nach (8.117) ein Risikoverhältnis (hazard ratio) definiert werden.

$$HR(t, \mathbf{x}, \mathbf{y}) = \frac{h(t, \mathbf{x})}{h(t, \mathbf{y})} = \frac{\exp(\mathbf{x}'\boldsymbol{\beta})}{\exp(\mathbf{y}'\boldsymbol{\beta})} = \exp((\mathbf{x} - \mathbf{y})\boldsymbol{\beta}) \quad (8.117)$$

Das Risikoverhältnis wird in dem PH-Modell somit als unabhängig von der Zeit angenommen. Die Interpretation der Modellparameter erfolgt mit direktem Bezug auf das Risikoverhältnis.

(a) Dichotome Einflussgrößen

Dichotome (zweiwertige) Einflussgrößen sollten 0/1-skaliert sein (anderenfalls erfolgt in den meisten Programmen eine entsprechende Recodierung automatisch). Dabei steht 1 für das Vorliegen des Faktors. Der Einfluss eines dichotomen Faktors x_k , für den ein Koeffizient $\hat{\beta}_k$ geschätzt wurde, kann durch das Risikoverhältnis (8.118) beschrieben werden.

$$HR(t, 0, 1, \hat{\beta}_k) = \exp((1 - 0)\hat{\beta}_k) = \exp(\hat{\beta}_k) \quad (8.118)$$

Dabei entspricht das HR dem Chancenverhältnis (odds ratio) aus der logistischen Regression. Der Unterschied ist, dass sich das HR auf Raten und nicht auf auf Chancen bezieht.

```
> summary(fit)
...
      exp(coef) exp(-coef) lower .95 upper .95
age           1.133      0.883   1.033   1.24
rx             0.401      2.496   0.111   1.44
resid.ds      2.285      0.438   0.486  10.74
ecog.ps       1.400      0.714   0.396   4.94
...
```

Das Risikoverhältnis der Therapie 2 (rx) gegen die Therapie 1 ist für das Beispiel (Tabelle 8.12) 0,4. Damit ist das Risiko für einen Exitus in Therapiegruppe 1 um den Faktor 2,5 (=1/0,4) höher als in der Therapiegruppe 1. Ein $(1 - \alpha)100\%$ -Konfidenzintervall kann näherungsweise nach (8.119) angegeben werden.

$$\exp\left(\hat{\beta} \pm z_{1-\alpha/2} \sqrt{se(\hat{\beta})}\right) \quad (8.119)$$

Das 95%-Konfidenzintervall für das Risikoverhältnis der Therapiegruppen (1 vs. 2) ist danach 0,11 - 1,44 und somit nicht signifikant, da es die 1 enthält.

(b) Nominalskalierte Einflussgrößen

Nominalskalierte Einflussgrößen können im Modell der Cox-Regression durch die Einführung von Dummy- bzw. Designvariablen berücksichtigt werden (vergleiche Abschnitt [8.2.4] und [8.3.1.2]). Das Vorgehen soll hier am Beispiel eines fiktiven Faktors mit 4 Stufen, z.B Stadium: I, II, III und IV, aufgezeigt werden.

Stadium	S_2	S_3	S_4
I	0	0	0
II	1	0	0
III	0	1	0
IV	0	0	1

D	HR
I	1
II- S_2	$HR(2, 1) = \exp(\hat{\beta}_1)$
III- S_3	$HR(3, 1) = \exp(\hat{\beta}_2)$
IV- S_4	$HR(4, 1) = \exp(\hat{\beta}_3)$

Der Einfluss des Faktors „Stadium“ wird somit durch drei dichotome Hilfsvariablen erfasst. Die zugehörigen Regressionskoeffizienten drücken jeweils das Verhältnis der Risiken zum Stadium I aus.

(c) **Stetige Einflussgrößen**

Für stetige Einflussgrößen erfolgt die Interpretation der Regressionskoeffizienten über konstante Intervalle. Dabei wird in dem Modell von einem linearen Zuwachs oder einer linearen Abnahme des Risikos in Abhängigkeit von der Einflussgröße ausgegangen. Ist x_k eine stetige Einflussgröße und $\hat{\beta}_k$ der entsprechende geschätzte Regressionskoeffizient, dann kann das Risikoverhältnis durch (8.120) beschrieben werden.

$$HR(t, x_k, x_k + c, \hat{\beta}_k) = \exp(c\hat{\beta}_k) \quad (8.120)$$

Verändert sich die Einflussgröße x_k um c Einheiten, dann verändert sich das Risikoverhältnis um das $e^{c\hat{\beta}_k}$ -fache. Für die Daten aus dem Beispiel (Tabelle 8.12) entspricht diese Veränderung, auf eine Altersdifferenz von 10 Jahren bezogen, dem 3,5fachen ($e^{10 \cdot 0,125}$). Ein $(1 - \alpha)100\%$ -Konfidenzintervall kann nach (8.121) bestimmt werden.

$$\exp\left(c\hat{\beta}_k \pm z_{1-\alpha/2}|c|\sqrt{se(\hat{\beta}_k)}\right) \quad (8.121)$$

(d) **Interaktionen zwischen**

• **nominalskalierten Einflussgrößen:**

Die Behandlung von zwei nominalskalierten Einflussgrößen in der Modellbildung ohne und mit Berücksichtigung einer Interaktion soll am Beispiel Rauchen (ja/nein) und Tumorgroße (klein/mittel/groß) gezeigt werden.

Rauchen (α)	ohne Interaktion Tumorgroße (β)			mit Interaktion Tumorgroße (β)		
	klein	mittel	groß	klein	mittel	groß
nein	1	e^{β_1}	e^{β_2}	1	e^{β_1}	e^{β_2}
ja	e^{α_1}	$e^{\alpha_1 + \beta_1}$	$e^{\alpha_1 + \beta_2}$	e^{α_1}	$e^{\alpha_1 + \beta_1 + \gamma_{11}}$	$e^{\alpha_1 + \beta_2 + \gamma_{12}}$

• **nominalskalierter und stetiger Einflussgröße:**

Der (kombinierte) Einfluss einer nominalskalierten und einer stetigen Einflussgröße soll am Beispiel Rauchen (ja/nein) und Alter (Jahre) untersucht werden.

Alter (δ)	ohne Interaktion Rauchen (α)		mit Interaktion Rauchen (α)	
	nein	ja	nein	ja
x	$e^{\delta x}$	$e^{\alpha_1 + \delta x}$	$e^{\delta x}$	$e^{\alpha_1 + (\delta + \gamma_{12})x}$

• **stetigen Einflussgrößen:**

Für zwei stetige Einflussgrößen, Alter (Jahre) und Tumorgroße (mm) kann die Modellbildung nach dem folgenden Schema erfolgen.

Alter x	ohne Interaktion Tumorgröße y (δ_y)	mit Interaktion Tumorgröße y (δ_y)
(δ_x)	$e^{\delta_x x + \delta_y y}$	$e^{\delta_x x + \delta_y y + \gamma_{xy} xy}$

(e) **Modellrechnung (Prognose) aus der Schätzung:** Mit den geschätzten Regressionskoeffizienten lassen sich Modellrechnungen zur „erwarteten“ Überlebenszeit durchführen.

```
> data(ovarian)
> fit <- coxph( Surv(futime, fustat) ~ age + rx, ovarian)
> fit
...
      coef exp(coef) se(coef)      z      p
age  0.147    1.159   0.0461   3.19 0.0014
rx  -0.804    0.448   0.6320  -1.27 0.2000
...
```

Für das Beispiel aus Tabelle 8.12 mit den Einflussgrößen Alter (age) und Therapie (rx) werden die Koeffizienten $\hat{\beta}_1 = 0,147$ und $\hat{\beta}_2 = -0,804$ geschätzt. Mit festgelegtem Alter, z.B. 40 Jahre bzw. 60 Jahre, und unter Vorgabe einer Therapieart kann dann die erwartete Überlebenszeit geschätzt und graphisch dargestellt werden (vgl. Abbildung 8.19). Die geschätzte Basis-Überlebensfunktion $\hat{S}(t)$ ist zum Vergleich mit angegeben (gestrichelt).

```
> plot( survfit( fit ), conf.int=FALSE, lty=2,
        xlim=c(0,700), xlab="Überlebenszeit (Tage)")
> lines( survfit( fit, newdata=data.frame( age=40, rx=2 ) ), col="blue", lwd=3)
> lines( survfit( fit, newdata=data.frame( age=60, rx=1 ) ), col="red", lwd=3)
> legend(350, 0.95, "Alter 40 Jahre, Therapie 2", bty="n", cex=0.9)
> legend(100, 0.4, "Alter 60 Jahre, Therapie 1", bty="n", cex=0.9)
```

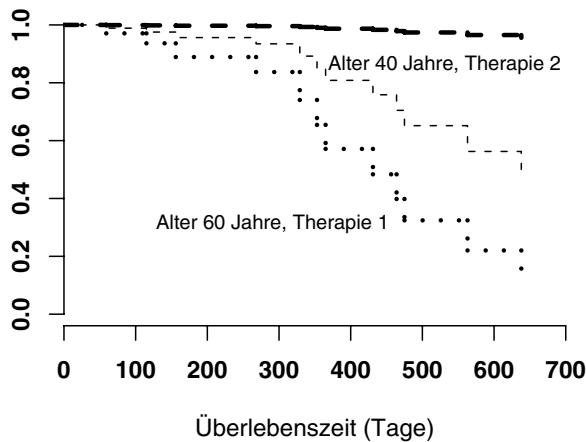


Abb. 8.19 Schätzung der Überlebensfunktion nach der Behandlung eines Ovarial-Karzinoms (Tabelle 8.12) für Patientinnen im Alter von 40 bzw. von 60 Jahren

8.6.4.5 Modellbildung; zur Auswahl geeigneter Einflussgrößen

Für die Entscheidung, ob eine einzelne Einflussgröße in das Modell der Cox-Regression aufgenommen werden soll, kann wiederum die Wald-Statistik (8.122) verwendet werden. In **R** wird diese standardmäßig mit der Schätzung für die Regressionskoeffizienten in der Funktion **coxph()** berechnet.

$$\hat{z} = \frac{\hat{\beta}}{se(\hat{\beta})} \quad (8.122)$$

Bei der statistischen Bewertung eines Modells, insbesondere hinsichtlich der Berücksichtigung einzelner Einflussgrößen im Prozess der Modellbildung, ist der Likelihood-Quotienten-Test (8.123) wesentlich informativer. Mit dieser Statistik können einzelne Faktoren oder Faktorkombinationen mit dem „Nullmodell“, in dem ausser der Konstanten $\hat{\beta}_0$ keine Faktoren berücksichtigt sind, verglichen werden.

$$\hat{G} = 2[\log(L(\hat{\beta})) - \log(L(0))] \quad (8.123)$$

Die Statistik $\hat{G}$ ist χ^2 -verteilt, wobei die Anzahl der Freiheitsgrade durch die Zahl der berücksichtigten Einflussgrößen bestimmt wird. In **R** kann der „Log-Likelihood“-Wert aus der Funktion **coxph()** explizit verwendet werden. Die Betrachtung unterschiedlicher Modelle mit dem Likelihood-Quotienten-Test (8.123) wird an den Daten aus Tabelle 8.12 gezeigt.

```
> fitm <- coxph(Surv(futime, fustat) ~ age + rx + resid.ds + ecog.ps, ovarian)
> fitm$loglik[1]; fitm$loglik[2]
[1] -34.98494
[1] -26.46329
> fit1 <- update(fitm, . ~ . -ecog.ps)
> gm <- 2*(fitm$loglik[2]-fitm$loglik[1]); gm; pchisq(gm, 4, lower.tail=F)
[1] 17.04329
[1] 0.001895867
> g1 <- 2*(fit1$loglik[2]-fit1$loglik[1]); g1; pchisq(g1, 3, lower.tail=F)
[1] 16.76757
[1] 0.0007889437
> fit2 <- update(fit1, . ~ . - resid.ds)
> g2 <- 2*(fit2$loglik[2]-fit2$loglik[1]); g2; pchisq(g2, 2, lower.tail=F)
[1] 15.88608
[1] 0.0003551247
> fit3 <- update(fit2, . ~ . - age)
> g3 <- 2*(fit3$loglik[2]-fit3$loglik[1]); g3; pchisq(g3, 2, lower.tail=F)
[1] 1.051453
[1] 0.5911257
```

Die Log-Likelihood für das Nullmodell ist -34,99. Für das „volle“ Modell (fitm), mit allen zu berücksichtigenden Einflussgrößen, ist die log-Likelihood -26,46, der Wert für die G -Statistik ist somit $\hat{G} = 17,04$ ($P=0,002$). Der erste Schritt (fit1) zeigt, dass in dem Modell auf den ECOG-Score verzichtet werden kann ($\hat{G} = 16,77$; die Differenz zum vollständigen Modell beträgt nur 0,27). Entsprechend führt auch ein Modell ohne ECOG-Score und ohne den Faktor Resterkrankung (fit2) zu einem Wert $\hat{G} = 15,89$ (Differenz 1,15). Im letzten Schritt (fit3) ist zu erkennen, dass auf das Alter in dem Modell nicht verzichtet werden kann ($\hat{G} = 1,05$), d.h. dass das Alter den wesentlichen Anteil am „Erklärungswert“ des Modells ausmacht.

Vergleiche zwischen verschiedenen Modellen können auch auf der Grundlage des Informationskriteriums nach Akaike (8.124) durchgeführt werden.

$$AIC = -2(\log(\text{likelihood})) + 2p \quad (8.124)$$

Je kleiner der Wert für dieses empirische Maß AIC nach (8.124) ist, desto besser beschreibt das Modell die vorliegenden Beobachtungen. Der Wert p entspricht der Anzahl der in dem Modell berücksichtigten Einflussgrößen. Eine automatisierte Prozedur zur Modellfindung ist in **R** mit der Funktion **stepAIC()** aus dem Paket **library(MASS)** verfügbar.

```
> stepAIC(fit, upper = ~ age + rx + resid.ds + ecog.ps, trace=TRUE)
Start: AIC= 60.93
Surv(futime, fustat) ~ age + rx + resid.ds + ecog.ps
```

```
...
Step: AIC= 59.20 ... ohne ecog.ps
Step: AIC= 58.08 ... resid.ps
Step: AIC= 57.68 ... ohne rx

Surv(futime, fustat) ~ age

Call:
coxph(formula = Surv(futime, fustat) ~ age, data = ovarian)

      coef exp(coef) se(coef)      z      p
age 0.162      1.18    0.0497  3.25 0.0012

Likelihood ratio test=14.3 on 1 df, p=0.000156 n= 26
```

Die Prozedur zeigt, dass das Alter der Patientinnen den wesentlichen Einfluss in einem Modell zur Schätzung von Überlebenszeiten ausmacht.

8.6.4.6 Güte der Modellanpassung, Residuen

Die Überprüfung von Modellannahmen und die Bewertung der Güte der Anpassung des Modells an die beobachteten Daten sind auch bei der Cox-Regression zentraler Bestandteil der Modellbildung. Neben der Aufdeckung systematischer Fehler muss dabei besonders auch auf einzelne Beobachtungen geachtet werden, die das Modell verfälschen, bzw. die nicht durch das Modell zu erklären sind.

„Cox-Snell“-Residuen

Cox-Snell-Residuen aus dem Modell der Cox-Regression für das i te Individuum (zum Zeitpunkt t_i) sind nach (8.125) definiert. Dabei wird in der Regel der Nelson-Aalen Schätzer für die kumulierte Risikofunktion verwendet.

$$r_{Ci} = \exp(\mathbf{x}'\hat{\beta})\hat{H}'_0(t_i) = \hat{H}'(t_i) = -\log(\hat{S}'(t_i)) \quad (8.125)$$

Eine anschauliche Bewertung der Cox-Snell-Residuen basiert auf der folgenden Überlegung: Ist $S(t)$ die Überlebensfunktion, die die Verteilung einer Zufallsvariablen T von Überlebenszeiten beschreibt, dann lässt sich nach Collett [Col03] allgemein zeigen, dass die Zufallsvariable $Y = -\log(T)$ exponential verteilt ist mit dem Erwartungswert 1. Nach entsprechender Transformation lässt sich das kumulierte Risiko dann als lineare Funktion durch den Ursprung mit der Steigung 1 im Koordinatensystem darstellen (Winkelhalbierende): Zu dem Kaplan-Meier Schätzer der Cox-Snell-Residuen (an Stelle der beobachteten Überlebenszeiten) $\hat{S}(r_{Ci})$ wird demnach das geschätzte **kumulierte Risiko**

$$\hat{H}(r_{Ci}) = -\log(\hat{S}(r_{Ci}))$$

berechnet und graphisch dargestellt. Abweichungen zwischen beobachteten und nach dem vorliegende Modell erwarteten Überlebenszeiten sind durch Abweichungen von der Winkelhalbierenden deutlich zu erkennen. Abbildung 8.20 zeigt das kumulierte Risiko (logarithmiert) der Cox-Snell-Residuen für die Beispieldaten aus Tabelle 8.12.

```
> library(survival)
> data(ovarian)
> fit0 <- coxph(Surv(futime, fustat) ~ 1, ovarian)
> fitm <- coxph(Surv(futime, fustat) ~ age + rx + resid.ds + ecog.ps, ovarian)
> m.resid <- resid(fitm)
> cs.resid <- ovarian$fustat - m.resid
```

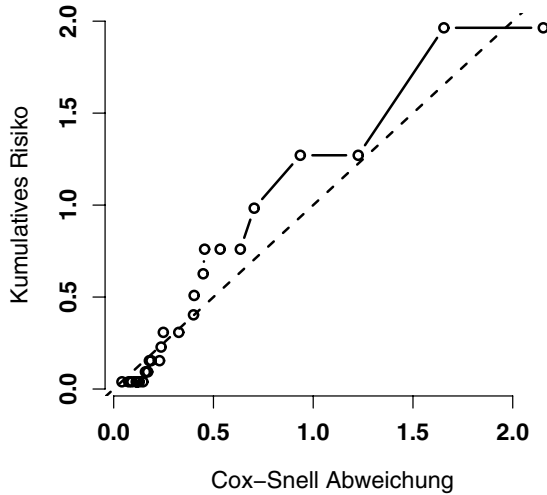



Abb. 8.20 Log-kumuliertes Risiko der Cox-Snell-Residuen für die Daten der Tabelle 8.12

```
> km.cs <- survfit(Surv(cs.resid, ovarian$fustat))
> cs.times <- km.cs$time
> cs.S <- km.cs$surv
> cs.exp <- -log(cs.S)
> plot(cs.times, cs.exp, type="b", xlab="Cox-Snell Abweichung",
+       ylab="Kumulatives Risiko")
> abline(0, 1, lty=2)
```

Hinweise:

- Die Cox-Snell-Residuen sind am ehesten dazu geeignet, die Güte der Modellanpassung einer Cox-Regression insgesamt graphisch zu beurteilen.
- Abweichungen von der Diagonalen können statistisch auch nach Grambsch und Therneau [GT94] bewertet werden. In **R** steht hierfür auch die Funktion **cox.zph()** zur Verfügung.
- Bei kleinen Stichproben können die aufgezeigten Abweichungen auch aus der Unsicherheit der Schätzung (Schätzfehler) der Regressionskoeffizienten resultieren. Diese nimmt auf der „rechten“ Seite der Verteilung wegen der Ausfälle (Zensierungen) zu.

„Martingal“-Residuen

In der Wahrscheinlichkeitstheorie ist ein **Martingal** ein stochastischer Prozess, in dem der Erwartungswert einer „neuen“ Beobachtung gleich dem Wert der vorigen Beobachtung ist. Die Residuen nach (8.126) werden Martingal-Residuen genannt, da diese sich auch aus Zählprozessen ableiten lassen. Eine ausführliche Darstellung geben Fleming und Harrington [FH91].

$$r_{Mi} = \delta_i - r_{Ci} \quad (8.126)$$

δ_i ist ein Indikator, der angibt, ob eine Beobachtung t_i zensiert ist ($\delta_i = 0$) oder ob ein Ereignis vorliegt ($\delta_i = 1$). Die Cox-Snell-Residuen r_{Ci} werden nach (8.125) berechnet. Martingal-Residuen kennzeichnen danach die Abweichung zwischen beobachteten und nach dem Modell erwarteten Ereignissen:

- $r_{Mi} \gg 0$, die Überlebenszeit ist kürzer als erwartet.
- $r_{Mi} \ll 0$, die Überlebenszeit ist länger als erwartet.

Die Aufzeichnung von Martingal-Residuen gegen die beobachtete Überlebenszeit oder gegen die Werte der in dem Modell berücksichtigten Einflussgrößen gibt Aufschluss über systematische Fehler. Eine gute Modellanpassung ist an einem konstanten Verlauf, möglichst nahe an der Nulllinie, zu erkennen.

```
> fit0 <- coxph(Surv(futime, fustat) ~ 1, ovarian)
> scatter.smooth(ovarian$age, resid(fit0), xlab="Alter", ylim=c(-1,1),
+ ylab="Martingal-Residuen (Null-Modell)"); abline(h=0, lty=2)
> scatter.smooth(ovarian$rx, resid(fit0), xlab="Resterkrankung", ylim=c(-1,1),
+ ylab="Martingal-Residuen (Null-Modell)"); abline(h=0, lty=2)
```

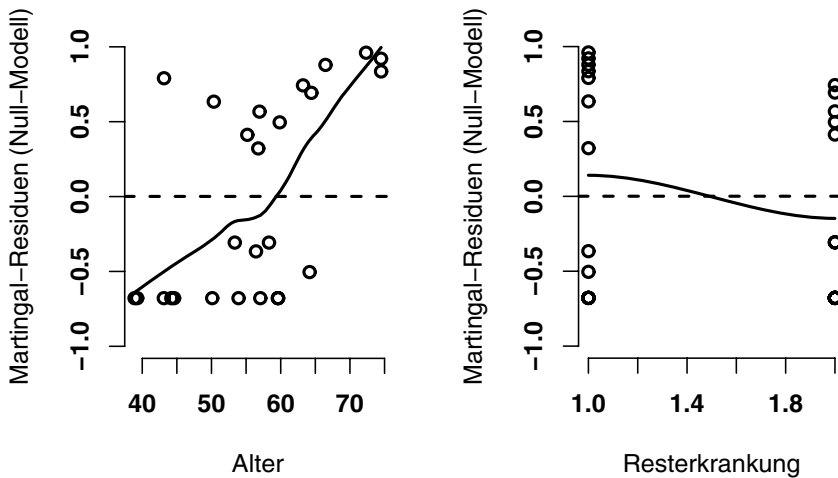


Abb. 8.21 Martingal-Residuen aus dem Nullmodell nach dem Alter der Patientinnen und dem Bestehen einer Resterkrankung für die Daten der Tabelle 8.12

Der Verlauf der Martingal-Residuen aus dem Nullmodell (ohne Berücksichtigung der Einflussgrößen) in Abbildung 8.21 zeigt deutlich, dass das Alter in das Modell aufgenommen werden muss [unterhalb von 60 Jahren ist erwartungsgemäß die Überlebenszeit länger als erwartet], während der Resterkrankung in der Modellbildung keine entscheidende Bedeutung zukommt. Die eingezeichneten Linien resultieren aus einer Kurvenanpassung mit einem Glättungsverfahren (smoothing) in R.

„Schoenfeld“-Residuen

Eine Überprüfung der zentralen Annahme eines proportionalen Risikos im Modell der Cox-Regression kann auch mit Hilfe der Schoenfeld-Residuen (8.127) erfolgen.

$$r_{Sji} = x_{ji} - \frac{\sum_{t_{(l)} \geq t_{(i)}} x_{jl} \exp(\mathbf{x}'_l \hat{\beta})}{\sum_{t_{(l)} \geq t_{(i)}} \exp(\mathbf{x}'_l \hat{\beta})} \quad (8.127)$$

$$= x_{ji} - E[x_{ji} | R_i, \delta_i = 1, \hat{\beta}]$$

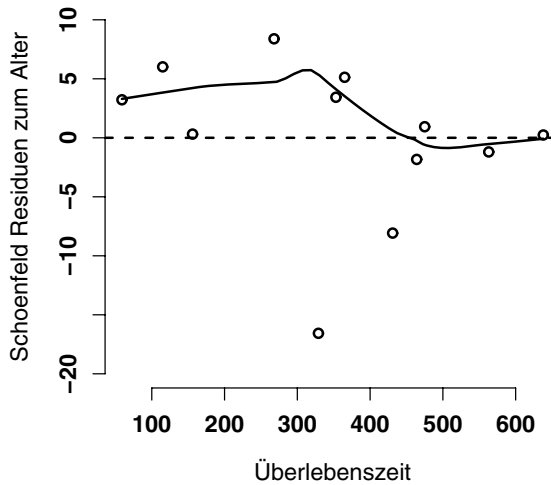


Abb. 8.22 Schoenfeld-Residuen für das Alter zur Überprüfung der Annahme des proportionalen Risikos in den Daten aus Tabelle 8.12

Schoenfeld-Residuen beziehen sich danach auf die Abweichungen zwischen den beobachteten Werten x_{ji} der j ten Einflussgröße (beim i ten Individuum) und einem nach dem Modell erwarteten Wert. Berücksichtigt werden nur die Fälle, zu denen nicht zensierte Beobachtungen vorliegen. Für jede Einflussgröße kann danach eine Menge von entsprechenden Schoenfeld-Residuen berechnet werden und nach der Überlebenszeit in das Koordinatensystem eingezeichnet werden. Ein horizontaler Verlauf (möglichst nahe der Nulllinie) zeigt an, dass die Annahme eines proportionalen Risikos für das Cox-Modell berechtigt ist. In **R** können die Schoenfeld-Residuen mit der Funktion **resid()** berechnet werden. Die entsprechenden Überlebenszeiten (nicht zensiert) werden mit der Funktion **coxph.detail()** aus dem Modell übernommen. Für die Daten aus Tabelle 8.12 sind die Schoenfeld-Residuen zum Alter in Abbildung 8.22 dargestellt.

```
> fit.age <- coxph(Surv(futime, fustat)~age, data=ovarian)
> detail <- coxph.detail(fit.age)
> time <- detail$y[,2]
> stat <- detail$y[,3]
> res <- resid(fit.age, type="schoenfeld")
> par(mfcol=c(1,1),lwd=2, font.axis=2, bty="n", ps=14)
> scatter.smooth(time[stat==1], res, ylim=c(-20,10),
+ xlab="Überlebenszeit", ylab="Schoenfeld Residuen zum Alter")
> abline(h=0, lty=2)
```

Sofern die Modellannahme zum proportionalen Risiko zutrifft, sollten die Schoenfeld-Residuen nach den geordneten Ereigniszeiten zufällig um den Wert Null streuen; systematische Verschiebungen oder sehr große Abweichungen zu einzelnen Zeitpunkten weisen auf eine Verletzung der Modellannahme hin (vgl. den [bzw. beide] Ausreißer in Abbildung 8.22).

Einführung in R

R ist in erster Linie eine Programmiersprache und Programmierumgebung für die statistische Analyse von Daten. **R** kann einerseits elementare mathematische Rechenoperationen ausführen, berechnet andererseits aber auch anspruchsvolle komplexe statistische Funktionen.

R wurde ursprünglich von **Ross Ihaka** und **Robert Gentleman** am Statistics Department of the University of Auckland entwickelt [IG96]. Aktuell wird das Programm durch eine internationale Arbeitsgruppe, das „R Development Core Team“ gepflegt und weiterentwickelt [R D05].

R wird unter der „GNU general public license“ entwickelt und kann somit aus dem Internet unter der Adresse

<http://cran.r-project.org>

frei herunter geladen werden. „**CRAN**“ steht dabei für „Comprehensive **R** Archive Network“ und ist ein weltweites Netz, durch das die Programme im Quellcode und als Binärdatei für verschiedene Rechnerplattformen (incl. Windows und Unix) bereitgestellt werden. Unter der gleichen Adresse ist eine ausführliche Einführung (Introduction to **R**) und ein vollständiges Handbuch (R Reference) im Adobe-Format (PDF) einzusehen bzw. herunter zu laden. Zahlreiche Fragen, die sich bei dem Einstieg in die Benutzung von **R** ergeben, finden unter Umständen in der FAQ-Liste (frequently asked questions)

<http://cran.r-project.org/doc/FAQ/R-FAQ.html>

eine Antwort. Eine kompakte Einführung in das Programmieren mit **R** gibt U. Ligges [Lig05]. Vielseitige Möglichkeiten der Anwendung von **R** zeigt auch P. Dalgaard [Dal05] auf.

Einige entscheidende Gründe, die bei der statistischen Datenanalyse für das Programm **R** sprechen, sind:

- Die Benutzung von **R** ist einfach und weitgehend intuitiv. Dabei zeigt **R** eine große Flexibilität im Umgang mit statistischen Funktionen und ist für spezielle Fragen erweiterbar.
- Das Programm **R** bietet vielseitige Möglichkeiten der graphischen Aufbereitung von Daten und Ergebnissen.
- **R** ist frei erhältlich und kann unter den weit verbreiteten Betriebssystemen, speziell Windows und Linux, installiert werden.

9.1 Das Konsolfenster

Nach dem Start des Programmes erscheint ein Fenster mit der **R**-Konsole (Abbildung 9.1). Über die im Kopf des Fensters angezeigten Menüs lassen sich einige wichtige Funktionen in **R** ausführen, z.B. unter

- **Datei** das Speichern und Einlesen von Befehlsdateien (Skript) und Arbeitsumgebungen (Workspace) sowie die Festlegung einer Verzeichnismgebung.
- **Bearbeiten** das Einfügen und Auswählen von Befehlen oder Befehlssequenzen (Skript), das Löschen des Konsolfensters, die Datenerfassung oder Korrektur über einen internen Dateneditor und die Konfigurierung der Benutzeroberfläche (GUI).

- **Verschiedenes** das Abbrechen laufender Berechnungen und die Anzeige der Objekte in der aktuellen Arbeitsumgebung.
- **Pakete** die Installation und Aktualisierung von Zusatzpaketen aus dem Internet.
- **Windows** die Fensterverwaltung, insbesondere bei der Verwendung mehrerer Fenster zur Anzeige von Graphiken.
- **Hilfe** der Einstieg in die umfangreichen Materialien zur Hilfestellung bei der Benutzung von R (vgl. auch den folgenden Abschnitt).



Abb. 9.1 Fenster mit der **R**-Konsole

Die zentrale Aufgabe der **R**-Konsole besteht darin, **Befehle** (Operationen, Funktionen) durch den Benutzer entgegengenommen und die Ergebnisse der Berechnungen anzuzeigen.

Hinweis zur Schreibweise: Befehle, Funktionen und Beispiele in **R** werden im Buch einheitlich wie folgt dargestellt:

```
> mean(c(4,6,8,9))           # Mittelwertberechnung in R
[1] 6.75
```

Namen von Funktionen in **R** (z.B. hier für die Berechnung des Mittelwertes **mean()**) werden im Text einheitlich durch eine fette Darstellung hervorgehoben.

Befehle werden im Dialog interpretiert und ausgeführt. Dazu gibt es eine festgelegte **Syntax** (feste Schreib- und Zeichenregeln), die während der Arbeit mit dem Programm genau eingehalten werden muss, damit **R** die gewünschten Operationen auch ausführen kann. Besonders zu beachten ist, dass **R** grundsätzlich **Groß- und Kleinschreibung** unterscheidet! Einige wichtige Zeichen der Syntax von **R** sind in Tabelle 9.1 zusammengefasst.

Nach der **Eingabeaufforderung** (**>**) ist stets ein vollständiger Befehl der Syntax entsprechend einzugeben und mit der **Return-Taste** abzuschließen (z.B. **sqr(5)**; square root) für die Berechnung der Wurzel aus 5.

```
> sqrt(5)                     # Wurzelfunktion in R
[1] 2.236068
```

Wird die Return-Taste vor dem Ende des Befehls betätigt, d.h. der Befehl ist unvollständig, dann zeigt das Programm am Beginn der folgenden Zeile durch ein Pluszeichen (+) an, dass der Befehl fortgesetzt werden kann.

Tabelle 9.1 Wichtige Zeichen, die in der Syntax von **R** zu verwenden sind

Symbol	Funktion
>	Zeichen für die Eingabeaufforderung (prompt); das Zeichen zu Beginn der Eingabezeile zeigt an, dass ein neuer Befehl eingegeben werden kann.
<-	Zeichen für die Zuordnung von Werten (in neueren Versionen von R kann auch das Gleichheitszeichen „=“ verwendet werden).
[]	(feste) Positionen in den Objekten (Index, Adresse) werden in eckigen Klammern angegeben.
„a“	Texte (Zeichenketten) werden in R in Hochkommata eingeschlossen.
:	Der Doppelpunkt wird für die Erzeugung von Zahlenfolgen eingesetzt, z.B. 1:5 erzeugt die Folge 1,2,3,4,5.
;	Das Semikolon trennt mehrere Befehle in einer Eingabezeile.
.	Dezimalzahlen werden in R generell mit Punkt geschrieben. Das Komma dient als Trennzeichen in Aufzählungen (Listen).
+, -, *, /	Für die Grundrechenarten Addition, Subtraktion, Multiplikation und Division werden die üblichen Zeichen verwendet.
^	Zeichen für das Potenzieren.
Logische Operationen (Vergleiche)	
==	für die Gleichheit,
!=	für die Ungleichheit,
> >=	für größer bzw. größer oder gleich,
< <=	für kleiner bzw. kleiner oder gleich.

Das Programm **R** arbeitet somit befehlsorientiert. Ein fehlerhaft eingegebener Befehl kann korrigiert werden. Dazu wird mit der Taste ↑ (up) der letzte Befehl aus dem internen **Befehlsspeicher** wieder in die **R**-Konsole geschrieben, kann einfach geändert und erneut ausgeführt werden.

```
> round(5.23454, digits=3)           # Rundung auf 3 Dezimalziffern
Fehler in round(5.23454, digits = 3) : unbenutzte(s) Argument(e) (digits ...)
> round(5.23454, digits=3)
[1] 5.235
```

In dem Beispiel wurde ein Fehler in der Schreibweise des Argumentes „*digits*“ zu der Funktion **round()** gemacht. Mit den Tasten ↑ (up) und ↓ (down) kann auf diese Weise auch der gesamte Befehlsspeicher durchsucht werden.

Die Auswertung von Daten erfolgt in der Regel durch eine Folge von Befehlen (**Programme**), die häufig wiederholt oder modifiziert werden müssen. Daher ist es sinnvoll, diese Befehlssequenzen über einen externen **Texteditor** zu bearbeiten. Grundsätzlich kann jeder Texteditor hierfür verwendet werden, allerdings sind für einige Editoren spezielle Schnittstellen (interfaces) verfügbar, mit denen der Umgang mit **R** wesentlich vereinfacht werden kann, z.B. WinEdt (<http://www.winedt.com/>) unter Windows oder EMacs unter Linux (<http://www.gnu.org/software/emacs/>).

9.2 Objekte in R

Das Programm **R** arbeitet mit **Objekten**. Das wichtigste Objekt in **R** ist der **Vektor**. Unter einem Vektor versteht man eine geordnete Menge von einzelnen Elementen. Die Anzahl der Elemente legt die **Länge des Vektors** fest. Einzelne Zahlenwerte werden übrigens als Vektoren der Länge 1 behandelt. Alle Elemente eines Vektors haben den gleichen **Datentyp**. In **R** werden die Datentypen **numeric**, **logical** und **character** unterschieden (vgl. Tabelle 9.2).

Tabelle 9.2 Datentypen in R

numeric	Zahlenwerte; alle Zahlen werden in R mit der doppelten Genauigkeit behandelt; z.B. 5 oder 3.467 oder $2.46e5 = 2.46 \cdot 10^5 = 246000$
character	freie Texte (Zeichenketten) werden in Hochkommata eingeschlossen, z.B. "hoch" oder "Hannover"
logical	logische Marker, speziell T (TRUE) und F (FALSE)

Die Länge eines Objektes kann in **R** mit der Funktion **length()** und der Datentyp mit der Funktion **mode()** abgefragt werden.

```
> x <- 1:10; length(x); mode(x)
[1] 10
[1] "numeric"
> name <- c("Statistik", "Mathematik"); length(name); mode(name)
[1] 2
[1] "character"
```

Objekte in **R** erhalten in der Regel **Namen**. Diese werden beliebig aus großen und kleinen Buchstaben, Zahlen und Punkten gebildet, wobei das erste Zeichen immer ein Buchstabe ist. Besonders zu beachten ist, dass in **R** Groß- und Kleinschreibung unterschieden wird. Konflikte mit den Namen von bestehenden Funktionen in **R** müssen möglichst vermieden werden! Grundsätzlich sollten die Namen von Objekten informativ und nicht zu kurz festgelegt werden.

Tabelle 9.3 Wichtige Objekttypen in R

vector	Ein Vektor ist eine geordnete Sammlung von Elementen des gleichen Typs. Die Elemente eines Vektors können mit der Funktion c (für combine oder concatenate) zusammengefügt werden, z.B. <code>werte <- c(1, 4, 5, 15)</code>
matrix	Eine Matrix besteht aus einer Anzahl von Vektoren (Spalten gleichen Typs und gleicher Länge). Die Anzahl der Spalten und Zeilen legt die Dimension der Matrix fest. Vektoren können mit den Funktionen rbind() zeilenweise und cbind() spaltenweise zu Matrizen zusammengefügt werden.
list	Eine Liste ist eine geordnete Sammlung von Objekten in R . Im Gegensatz zum Vektor können in einer Liste auch Objekte unterschiedlichen Typs geführt werden. Listen werden mit der Funktion list() erzeugt.
data.frame	Ein Datenrahmen (data.frame) ist eine Kombination aus Liste und Matrix. Er enthält Vektoren unterschiedlichen Typs (Merkmale, Variablen) gleicher Länge, wie sie im Rahmen von Erhebungen und Experimenten häufig auftreten. Diese Datenrahmen können durch die Funktion data.frame() aus Vektoren erzeugt oder aus externen Dateien mit der Funktion read.table() in R eingelesen werden.

Den Objekten werden Daten oder Werte entsprechend ihres Typs zugewiesen. Als **Zuweisungs-symbol** wird einheitlich in diesem Buch „<-“ verwendet.


```
> wurzel.12 <- sqrt(12)           # Wurzel aus 12
> wurzel.12
[1] 3.464102
```

R quittiert eine Zuweisung lediglich durch ein neues Prompt (**>**) in der nachfolgenden Eingabezeile. Der Inhalt eines Objektes kann dann durch die Eingabe des Namens angezeigt werden!

Die Objekte in **R** haben eine feste Struktur. Die wichtigsten verwendeten **Objekttypen** sind in Tabelle 9.3 zusammengestellt.

Eine Übersicht zu den Objekten, die aktuell im Speicher definiert sind und mit denen gearbeitet werden kann, wird durch die Funktion **ls()** (list objects) angezeigt. Andererseits können Objekte mit der Funktion **rm()** (remove objects) jederzeit aus dem Speicher wieder gelöscht werden. Die Art der Datenstruktur oder der Typ der Daten in den Objekten können durch eine Reihe spezieller „**is**“-Funktionen abgefragt werden, z.B. **is.numeric()** oder **is.matrix()**, die jeweils die logischen Werte „TRUE“ oder „FALSE“ liefern. Das Beispiel zeigt die Definition einer Matrix (Vierfeldertafel) und die Abfrage auf den Objekttyp und den Datentyp.

```
> m <- matrix(c("a","b","c","d"), nrow=2); m
      [,1] [,2]
[1,]  "a"  "c"
[2,]  "b"  "d"
> is.matrix(m)
[1] TRUE
> is.numeric(m)
[1] FALSE
```

9.3 Hilfestellung in R

Für den Einstieg in **R** ist die Lektüre der Kurzeinführung **Introduction to R** (als PDF-Dokument aus dem Internet erhältlich) dringend zu empfehlen. Informationen zur Syntax und zur Verwendung der zahlreichen Funktionen können auch während der Arbeit mit **R** auf unterschiedlichen Wegen angezeigt werden:

- Ist der Name der Funktion bekannt, wird durch ein voran gestelltes Fragezeichen eine vollständige Beschreibung dieser Funktion (Syntax, Funktionsbeschreibung und einfache Beispiele) in einem gesonderten „Hilfefenster“ angezeigt.
- Ist nur ein Teil des Namens bekannt, dann kann über die Funktion **apropos()** eine Liste aller Funktionen angezeigt werden, in denen dieser Text auftritt (z.B. **apropos(mean)**).

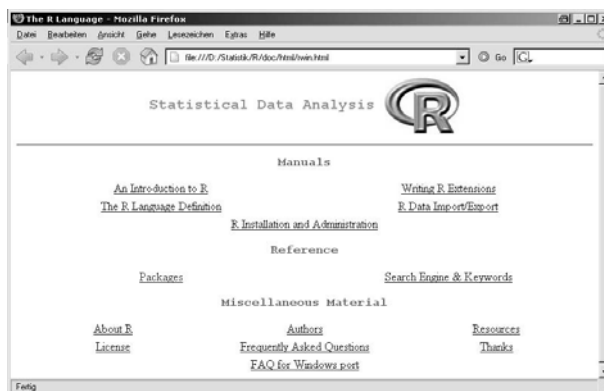


Abb. 9.2 Allgemeine Hilfestellung (HTML) in **R** durch den Befehl **help.start()**

- Eine sehr umfassende und übersichtliche Hilfestellung ist im HTML-Format über das jeweilige Browser-Programm (z.B. *Internet Explorer* oder *Mozilla Firefox*) einfach zugänglich. Mit dem Befehl **help.start()** wird der Browser gestartet, mit dem dann die gewünschten Informationen durch interne *Links* schnell aufzufinden sind (Abbildung 9.2).

Ergänzend zu diesen Hilfestellungen muss insbesondere auf die Hilfe im Internet (CRAN) hingewiesen werden. Häufig gestellte Fragen (**FAQ - frequently asked questions**) finden hier in umfangreichen Listen eine Antwort. Daneben werden Suchfunktionen (u.a. *Google Search*) angeboten, mit denen auch für spezielle Probleme Lösungswege gefunden werden können. Letztlich besteht eine sehr engagierte Liste (**r-help@stat.math.ethz.ch**), über die Erfahrungen und Probleme bei der Benutzung von **R** ausgetauscht werden können.

9.4 Erzeugen von Daten in R mittels Funktionen

Mit der Funktion `'.'` lässt sich in **R** einfach eine fortlaufende Zahlenreihe erzeugen und in einem Vektor ablegen. So erzeugt der Befehl **1:10** die Zahlenreihe von 1 bis 10 in aufsteigender oder **20:15** die Zahlenreihe von 20 bis 15 in absteigender Folge.

```
> 1:10
[1] 1 2 3 4 5 6 7 8 9 10
> 20:15
[1] 20 19 18 17 16 15
```

Andere wichtige Funktionen zur Erzeugung von Zahlenfolgen in **R** sind **seq()** (sequence) und **rep()** (repeat). Die Syntax der Funktion **seq()** ist:

seq(*from*, *to*, *by* =) oder
seq(*from*, *to*, *length* =)

Die Zahlenfolge wird durch die erste und die letzte Zahl sowie durch die Schrittweite (*by*) oder alternativ auch durch die Anzahl (*length*) der zu erzeugenden Zahlen bestimmt, zum Beispiel:

```
> seq(1, 5, by=0.5)
[1] 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0
```

oder

```
> seq(1, 5, length=11)
[1] 1.0 1.4 1.8 2.2 2.6 3.0 3.4 3.8 4.2 4.6 5.0
```

Die Syntax der Funktion **rep()** ist:

rep(*x*, *times*)

Die Anzahl der Wiederholungen von *x* wird durch das Argument *times* festgelegt, z.B.

```
> rep(5, 20)
[1] 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5
```

Eine der wichtigsten Funktionen in **R** ist die Funktion **c()** für das Verbinden bzw. Zusammenfügen (**concatenate**) von Objekten, also insbesondere auch von Zahlen oder Vektoren.

```
> c(1,7:9)
[1] 1 7 8 9
```



Besonders zu beachten ist, dass **R** in der Funktion **c()** den Datentyp der Werte automatisch anpasst, wenn unterschiedliche Typen unter den Werten auftreten! In dem folgenden Beispiel werden Zahlenwerte (**numeric**) automatisch in Zeichenfolgen (**character**) umgewandelt.

```
> c(1:5, 10.5, "next")
[1] "1" "2" "3" "4" "5" "10.5" "next"
```

9.5 Dateneingabe: „Daten in Rahmen“ (data.frame)

Ein **Datenrahmen** (data frame) besteht aus Vektoren, die vom Typ her unterschiedlich sein können, aber alle die gleiche Länge haben. Dieser Objekttyp entspricht damit der in der statistischen Datenanalyse üblichen **Tabellenstruktur**, d.h. in den Zeilen werden die Fälle (Beobachtungseinheiten) und in den Spalten die Variablen (Merkmale, Beobachtungen) aufgeführt. Datenrahmen können mit der Funktion **data.frame()** erzeugt werden oder aus externen Dateien (z.B. Excel-Tabellen) importiert werden. Das folgende Beispiel zeigt die Erzeugung von 3 Datensätzen zu den Merkmalen Alter, Geschlecht und Körpergröße.

```
> alter      <- c(19, 22, 24)
> geschlecht <- c("maennlich", "weiblich", "maennlich")
> groesse    <- c(170, 165, 181)
> studenten <- data.frame(alter, geschlecht, groesse)
> studenten                                     # Ausgabe des erzeugten Datenrahmens
  alter geschlecht groesse
1   19  maennlich    170
2   22   weiblich    165
3   24  maennlich    181
```

Der **Import von Daten** aus externen Tabellen erfolgt am einfachsten mit der Funktion **read.csv()** über das **CSV-Format** (CSV, character separated values, ist ein Format zur Speicherung oder zum Austausch einfach strukturierter Daten). Hierfür sind die Beobachtungen für jeden Fall in einer Zeile durch Semikolon getrennt (*sep=";"*) angeordnet. In der ersten Zeile stehen die Variablennamen. Dezimalzahlen werden mit Kommata geschrieben (*dec=","*).

```
> infarkt <- read.csv("C:/Eigene Dateien/infarkt.CSV", sep=";", dec=",")
> edit(infarkt)
```

Die Daten eines Übungsbeispiels (aus Werner [Wer92]) zu den Risiken für einen Herzinfarkt werden aus einer Datei „*infarkt.csv*“ in einen Datenrahmen unter dem Namen „*infarkt*“ importiert. Das Ergebnis wird mit der Funktion **edit()** angezeigt (Abbildung 9.3).

Dateneditor										
	Nummer	Gruppe	Geschlecht	Alter	BldGese	Bldzeit	Bldtag	Diabet	Chol	Trigl
1	1	Infarkt	m	39	128	155	95	100	FALSE	195 146
2	2	Infarkt	m	43	106	145	95	140	TRUE	205 138
3	3	Infarkt	m	45	94	160	90	95	FALSE	245 200
4	4	Infarkt	m	50	98	200	110	80	FALSE	190 101
5	5	Infarkt	m	51	132	190	110	90	FALSE	260 202
6	6	Infarkt	m	52	108	195	110	90	FALSE	190 164
7	7	Infarkt	m	53	165	220	105	70	FALSE	340 150
8	8	Infarkt	m	53	116	140	85	85	FALSE	195 93
9	9	Infarkt	m	56	138	180	100	200	TRUE	285 135
10	10	Infarkt	m	56	140	175	95	90	FALSE	380 127
11	11	Infarkt	m	57	93	160	90	85	FALSE	220 131
12	12	Infarkt	m	57	115	185	105	85	FALSE	240 90
13	13	Infarkt	m	59	120	185	100	100	FALSE	235 119
14	14	Infarkt	m	59	114	190	120	110	FALSE	215 104
15	15	Infarkt	m	61	99	175	95	95	FALSE	190 122
16	16	Infarkt	m	61	136	140	80	130	TRUE	275 140
17	17	Infarkt	m	62	90	220	105	90	FALSE	205 161
18	18	Infarkt	m	62	102	160	95	90	FALSE	290 203
19	19	Infarkt	m	67	99	150	95	100	FALSE	200 110

Abb. 9.3 Anzeige der Daten aus einem Datenrahmen mit der Funktion **edit()**

Auf die Daten in den Spalten eines Datenrahmens kann mit Hilfe der Funktion **attach()** auch direkt über den Spaltennamen Bezug genommen werden, z.B. für den Cholesterinwert:

```
> attach(infarkt)
> Chol
[1] 195 205 245 190 260 190 340 195 285 380 220 ...
[20] 210 220 265 235 200 350 220 800 230 185 295 ...
[39] 190 210 220 200 185 220 215 135 220 180 220 ...
[58] 180 160 200 205 230 125 195 100 185 180 205 ...
[77] 180 190 175 200
> mean(Chol)
[1] 219.75
```

9.6 Auswahl und Sortierung von Daten

Die Elemente in den Objekten von **R**, insbesondere in Vektoren, Matrizen oder Datenrahmen, sind grundsätzlich geordnet. Somit kann auf einzelne Werte direkt über den **Index**, der in eckigen Klammern eingeschlossen wird, Bezug genommen werden.

```
> Zahl1bis20 <- 1:20
> Zahl1bis20[6:10]           # Auswahl des 6. Wertes aus einem Vektor
[1] 6 7 8 9 10
> blut <- c("A", "B", "AB", "0")
> blut[3]                   # Auswahl des 3. Wertes aus dem Vektor blut
[1] "AB"
> attach(infarkt)
> Chol[5]                   # Auswahl des 5. Wertes des Vektors Chol
[1] 260
```

Die Größe eines Datenrahmens wird durch die Anzahl der Zeilen und die Anzahl der Spalten bestimmt (zweidimensional). Der erste Index zeigt die Zeile, der zweite Index die Spalte an. Die Indices werden durch Komma getrennt. Wird bei einem **indizierten Zugriff** auf die Daten ein Index nicht angegeben, dann werden alle Elemente der entsprechenden Zeile oder Spalte ausgewählt. Für den im vorangehenden Abschnitt erzeugten Datenrahmen „*studenten*“ folgt zum Beispiel:

```
> studenten[,3]
[1] 170 165 181
> studenten[2,]
  alter geschlecht groesse
2    22    weiblich    165
```

Dabei können in den eckigen Klammern auch logische Ausdrücke zur Auswahl (**Selektion**) von Teilmengen eines Objektes spezifiziert werden, z.B.

```
> Zahl1bis20 <- 1:20
> Zahl1bis20[Zahl1bis20 > 13]
[1] 14 15 16 17 18 19 20
```

für die Auswahl der Werte, die größer als 13 sind, oder

```
> studenten [geschlecht == "maennlich"]
  alter geschlecht groesse
1    19    maennlich    170
3    24    maennlich    181
```

für die Auswahl aller männlichen Studenten aus dem Datenrahmen *studenten*, oder

```
> infarkt [Gruppe=="Infarkt" & Blutz > 100, ]
  Gruppe Sex  Alter  RRsys RRdias  Blutz  Diabet  Chol  Trigl  HbdH  Got  ...
2  Infarkt 1   43   145    95   140    1   205   138  380  19.0  ...
9  Infarkt 1   56   180   100   200    1   285   135  277  18.2  ...
14 Infarkt 1   59   190   120   110    2   215   104  285  18.6  ...
16 Infarkt 1   61   140    80   130    1   275   140  325  23.8  ...
20 Infarkt 1   68   180   105   105    2   210    95  236  15.3  ...
31 Infarkt 2   61   165   105   160    1   380   134  449  16.3  ...
37 Infarkt 2   70   165    95   130    1   395   125  482  21.6  ...
38 Infarkt 2   72   160    95   110    2   290   148  436  19.0  ...
```

für die Auswahl aller Personen aus der Infarktgruppe, deren Blutzuckerwert höher als 100 ist.

Eine **Sortierung** (aufsteigend oder fallend) der Elemente in Vektoren kann mit Hilfe der Funktion **sort()** vorgenommen werden.

```
> a <- c(3, 7, 2, 8, 5, 10, 4)
> a
[1] 3 7 2 8 5 10 4
> sort(a)
[1] 2 3 4 5 7 8 10
> sort(a, decreasing=TRUE)
[1] 10 8 7 5 4 3 2
```

Die in der Statistik häufig verwendete Bestimmung von **Rangzahlen** (Ordnungszahlen) kann mit der Funktion **rank()** erfolgen. Dabei werden den Werten eines Vektors Zahlen zugeordnet, die sich aus der Anordnung vom kleinsten (1) bis zum größten Wert (Anzahl der Werte) ergeben. Für den Fall, dass gleiche Werte auftreten (Bindungen), können gemittelte Rangzahlen zugewiesen werden, oder es werden wie im Sport gleiche Ränge gewählt.

```
> a <- c(3, 7, 2, 8, 5, 10, 4)
> rank(a)
[1] 2 5 1 6 4 7 3
> b <- c(3, 5, 7, 3, 6, 5)
> rank(b)
[1] 1.5 3.5 6.0 1.5 5.0 3.5
> rank(b, ties.method = "min")
[1] 1 3 6 1 5 3
```

Für das **Sortieren von Matrizen** oder „Daten in Rahmen“ (Tabellen) nach einer ausgewählten Spalte ist der Befehl **sort()** nicht geeignet. Hierfür steht in **R** die Funktion **order()** zur Verfügung, die einen „geordneten“ Index bestimmt, nach dem auch andere Spalten (Variablen) sortiert werden können.

```
> o <- order(a); a; o; a[o]
[1] 3 7 2 8 5 10 4
[1] 3 1 7 5 2 4 6
[1] 2 3 4 5 7 8 10
```

Eine Sortierung der Daten zu dem Rahmen *Studenten* aus dem vorangehenden Abschnitt nach der Körpergröße kann danach wie folgt durchgeführt werden.

```
> studenten[order(studenten[,3]),]
  alter geschlecht groesse
2    22    weiblich    165
1    19    maennlich    170
3    24    maennlich    181
```

9.7 Ablaufsteuerung: logische Bedingungen und Funktionen in R

Eine Folge von Befehlen (**Programm**) kann in **R** durch spezielle Funktionen gesteuert werden. Insbesondere können wiederholte Arbeitsschritte auch in **Schleifen** festgelegt werden.

Beispiel: In einem Vektor **a** mit 10 Elementen sollen die ersten 5 Elemente den Wert *'unten'* und die nächsten 5 Elemente den Wert *'oben'* erhalten. Der Vektor **a** wird zunächst mit fehlenden Angaben „**NA**“ (in **R** allgemein das Synonym für **fehlende oder unbekannte Werte**) initialisiert. Anschließend wird in einer **for()** - Schleife elementweise nach dem entsprechenden Indexwert entschieden, ob der Wert *'unten'* oder *'oben'* eingetragen werden soll.

```
> a <- rep(NA, 10)
> a
[1] NA NA NA NA NA NA NA NA
> for (i in 1:10) if (i < 6) a[i] <- "unten" else a[i] <- "oben"
> a
[1] "unten" "unten" "unten" "unten" "unten"
[6] "oben"  "oben"  "oben"  "oben"  "oben"
```

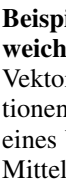
Die wichtigsten Funktionen zur Ablaufsteuerung in **R** sind in der Tabelle 9.4 zusammengestellt. Dort kann ein „Ausdruck“ (*expr*) aus einem einzelnen Befehl bestehen oder durch eine Folge von Befehlen zusammengesetzt werden, die dann durch geschweifte Klammern { } zusammengefasst werden müssen.

Tabelle 9.4 Ablaufsteuerung und Funktionen in **R**

if (<i>cond</i>) <i>expr1</i> else <i>expr2</i>	Der Ausdruck <i>expr1</i> wird ausgeführt, wenn die Bedingung <i>cond</i> wahr ist. Sonst wird alternativ der Ausdruck <i>expr2</i> ausgeführt
for (<i>var in seq</i>) <i>expr</i>	Der Ausdruck <i>expr</i> wird für jeden Wert von <i>var</i> in der Folge von <i>seq</i> ausgeführt.
while (<i>cond</i>) <i>expr</i>	Der Ausdruck <i>expr</i> wird ausgeführt, solange die Bedingung <i>cond</i> wahr ist.
break	Beenden bzw. Verlassen einer Befehlsfolge innerhalb der Ausdrücke von for- oder while-Schleifen.
next	Abbrechen bzw. Überspringen des Restes einer Befehlsfolge innerhalb der Ausdrücke von for- oder while-Schleifen.
function (<i>arglist</i>) { <i>expr</i> }	Definition eigener neuer Funktionen. Die Argumente werden in einer Liste (<i>arglist</i>), die Ausdrücke zur Berechnung der Funktion (<i>expr</i>) werden in geschweiften Klammern zusammengefasst.
return (<i>value</i>)	Liefert den Wert <i>value</i> in der Berechnung einer Funktion. Standardmäßig wird der Wert des letzten Befehls einer Funktion als Ergebnis zurück gegeben. Mehrere Ergebnisse können in einer Liste mit dem Befehl list() zusammengefasst werden.

 **Beispiel:** Die Summe der natürlichen Zahlen von 1 bis 10 kann mit einer **while()**-Schleife wie folgt berechnet werden. Einfacher ist jedoch die Verwendung des Befehls **sum()**.

```
> i <- 0; summe <- 0
> while (i < 10) {i <- i+1; summe <- summe +i }
> summe
[1] 55
> sum(1:10)
[1] 55
```

 **Beispiel:** In dem folgenden Beispiel wird eine eigene Funktion zur Berechnung der **Standardabweichung** definiert. Die neue Funktion erhält den Namen **stdabw()** und hat als Argument nur den Vektor *x*. Die Berechnung erfolgt in einzelnen Schritten, in denen hier beispielhaft auch die Funktionen **length()** (Anzahl der Elemente in einem Vektor) und **sum()** (Summe über die Elemente eines Vektors) verwendet werden (natürlich gibt es in **R** bereits Funktionen zur Berechnung von Mittelwert und Standardabweichung).

```
> stdabw <- function (x) {
+   anzahl <- length (x)
+   summe <- sum (x)
+   mittel <- summe / anzahl
+   saq <- sum( (x-mittel)^2)
+   return (sqrt (saq/(anzahl -1)))
+ }
> x <- c (2,3,4,5,6,7)
> stdabw(x) # neue Funktion
[1] 1.870829
> mean(x) # Mittelwertfunktion in R
[1] 4.5
> sd(x) # Standardabweichung in R
[1] 1.870829
```



Auch wenn die Definition von Schleifen in **R** mit diesen Befehlen recht einfach einsehbar ist, sollte stets die Möglichkeit einer „vektorwertigen Programmierung“ genutzt werden. Diese ist wesentlich übersichtlicher und effizienter. Hierfür stehen in **R** spezielle Funktionen zur Verfügung, die im Rahmen dieser kurzen Einführung nicht ausführlicher dargestellt werden können. Ein einfaches Beispiel ist die Funktion **apply()**, die die Anwendung von Funktionen auf Vektor- oder Matrixelemente unterstützt.

Die Berechnung von Zeilen- und Spaltensummen in einer Matrix kann mit Hilfe der Funktion **apply()** einfach wie folgt durchgeführt werden.

```
> x <- matrix(c(2, 6, 4, 8), nrow=2); x
     [,1] [,2]
[1,]    2    4
[2,]    6    8
> apply(x, 1, sum)      # Zeilensummen
[1]  6 14
> apply(x, 2, sum)      # Spaltensummen
[1]  8 12
```

9.8 Einige mathematische und statistische Funktionen

Die folgenden Tabellen geben einen kurzen Einblick in die vielseitigen Möglichkeiten der mathematischen und statistischen Berechnungen in **R**. Die gesamte Funktionalität von **R** erschließt sich aber erst bei der Nutzung in der jeweiligen Problemstellung. Für die Einführung sind nur einige Beispiele ausgewählt. Als Argumente werden in diesen Funktionen in der Regel Vektoren (*vect*) verwendet. Die ausführliche Syntax- und Funktionsbeschreibung muss in **R** nachgelesen werden.

Tabelle 9.5 Einige mathematische Funktionen in **R**

abs(vect)	bildet die Absolutbeträge zum Vektor <i>vect</i> .
round(vect, digits=d)	rundet Dezimalzahlen auf <i>d</i> Nachkommastellen.
ceiling(vect), floor(vect), trunc(vect)	bestimmt die nächsthöhere ganze Zahl, die nächstniedrigere ganze Zahl bzw. die nächste ganze Zahl, die näher zur Null liegt, aus den Werten eines Vektors.
exp(vect), log(vect), log10(vect)	berechnet die Exponentialfunktion (zur Basis <i>e</i>), den natürlichen Logarithmus und den Logarithmus zur Basis 10.
max(vect), min(vect)	bestimmt den größten bzw. kleinsten Wert.
sign(vect)	liefert einen Vektor mit der Kennzeichnung der Vorzeichen mit -1 bei negativen Werten, +1 bei positiven Werten und 0 bei Nullwerten.
sin(vect), cos(vect), tan(vect)	berechnet die trigonometrischen Funktionen.
sqrt(vect)	berechnet die Wurzel.

Die Anwendung der Funktionen aus Tabelle 9.5 soll an einigen einfachen Beispielen gezeigt werden.

```
> vect <- c(1.42, 4.84, -2.55, -1.24)
> abs(vect)
[1] 1.42 4.84 2.55 1.24
> round(vect, digits=1)
[1] 1.4 4.8 -2.5 -1.2
> ceiling(vect)
[1] 2 5 -2 -1
> floor(vect)
```

```

[1] 1 4 -3 -2
> trunc(vect)
[1] 1 4 -2 -1
> max(vect)
[1] 4.84
> min(vect)
[1] -2.55
> exp(5)
[1] 148.4132
> round(sin(seq(0, 2*pi, by=(pi/4))), digits=3)
[1] 0.000 0.707 1.000 0.707 0.000 -0.707 -1.000 -0.707 0.000
> sqrt(7)
[1] 2.645751

```

Tabelle 9.6 Einige statistische Funktionen in R

sum(vect)	berechnet die Summe über alle Werte.
prod(vect)	berechnet das Produkt über alle Werte.
mean(vect)	berechnet den arithmetischen Mittelwert.
median(vect)	berechnet den Medianwert.
cumsum(vect) cumprod(vect)	bildet einen Vektor mit der kumulierten Summe bzw. dem kumulierten Produkt über alle Elemente eines Vektors.
sort(vect)	bildet einen Vektor mit den sortierten Werten.
rank(vect)	bildet einen Vektor mit den Rangzahlen zu den Werten.
range(vect)	bildet einen Vektor mit der Spannweite (min, max) zu den Werten.
quantile(vect, ..., probs=)	bestimmt die Quantile zu den Werten; die entsprechenden Anteile werden optional in der Liste <i>probs=</i> festgelegt.
sd(vect)	berechnet die Standardabweichung.
var(vect)	berechnet die Varianz.
summary(object)	erstellt eine beschreibende Statistik zu den Elementen eines Objektes; insbesondere die Häufigkeitsverteilung zu Faktoren und beschreibende Maßzahlen zu numerischen Vektoren.
aov(formular, ...)	Varianzanalyse (-modelle) in R; <i>formular</i> dient der Modelldefinition nach einer eigenen Syntax (vgl. Tabelle 9.7).
lm(formular,...)	lineare Modelle (multiple lineare Regression)
glm(formular, ...)	verallgemeinerte lineare Modelle (loglineare Modelle, logistische Regression)

Die Anwendung statistischer Funktionen soll in einigen einfachen Beispielen mit den Daten aus dem Datenrahmen *infarkt* (aus dem obigen Beispiel) gezeigt werde.

```

> infarkt <- read.csv("G:/AS12/Rprogramm/infarkt.CSV", sep=";", dec=",")
> attach(infarkt)
> mean(Alter)
[1] 58.4875

> sd(Alter)
[1] 10.70549

> max(Blutz)
[1] 350

```



```
> quantile(RRsyst, prob=c(0.10,0.25,0.50,0.75,0.90))
10%    25%    50%    75%    90%
140.00 148.75 160.00 175.00 190.00

> summary(Chol)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  85.0   180.0   200.0   219.8   222.5   800.0
```

9.8.1 Formulierung von Modellgleichungen

Im Rahmen einer einführenden Lektüre kann dieser Abschnitt übersprungen werden. Er beschreibt einige Besonderheiten der Formulierung von Modellgleichungen, die im Kapitel zur Modellbildung [8] auftreten.

Die Formulierung einer Modellgleichung (formular) im Rahmen der Modellbildung, speziell in den statistischen Funktionen **lm()** und **glm()**, erfolgt in **R** mit einer speziellen Syntax. Der Operator „~“ trennt die Zielgröße y (response) von der Modellgleichung in der Form „ $y \sim model$ “. Dabei besteht „ $model$ “ aus Termen mit den Namen der Einflussgrößen. Durch den Operator „+“ werden einzelne Komponenten in das Modell eingeschlossen. Andererseits können aus einem bestehenden Modell durch den Operator „-“ auch einzelne Komponenten ausgeschlossen werden. Der Ausschluss des konstanten Terms (intercept) in einem statistischen Modell muss in **R** explizit durch „- 1“ oder „+ 0“ erfolgen. Wechselwirkungen (Interaktionen) können in das Modell durch den Operator „:“ eingeführt werden.

In der Modellgleichung können auch transformierte Einflussgrößen (z.B. mit der Logarithmusfunktion **log()** oder der Exponentialfunktion **exp()**) auftreten. Arithmetische Ausdrücke müssen dagegen durch die spezielle Funktion **I()** gekennzeichnet werden, um Verwechslungen mit den Operationen zur Modellbildung zu vermeiden.

Tabelle 9.7 Syntaxelemente zur Modellspezifikation in **R**

Modellterm	Erklärung
$Y \sim A + B$	Symbolischer Operator für die Berücksichtigung der Einflussgrößen A und B (Haupteffekte) im Modell; Y bezeichnet die Zielgröße im Modell.
$. \sim .$	Einschluss aller Größen (Haupteffekte) in das Modell
$. \sim . - 1$	Ausschluss einer Regressionskonstanten (β_0); Spezialfall der Regression durch den „Ursprung“.
$. \sim \pm A$	Einschluss/Ausschluss einzelner Größen; häufig in Verbindung mit der update() Funktion im Rahmen der Modellbildung.
$. \sim A : B$	Symbolischer Operator für die Wechselwirkung zwischen A und B .
$. \sim A * B$	Symbolischer Operator für das Modell mit beiden Haupteffekten und der Wechselwirkung.
$. \sim \text{offset}(A)$	Die Einflussgröße A wird mit konstantem Koeffizienten (1) in das Modell mit einbezogen.
$. \sim I(A + B)$	Die (arithmetische) Summe der Größen A und B wird als ein gemeinsamer Term in das Modell aufgenommen. Eine entsprechende Bezeichnung ist auch für andere mathematische Operationen oder Funktionen möglich.

9.9 Einfache graphische Funktionen und Hilfsmittel

Das Programm **R** bietet zahlreiche Möglichkeiten der graphischen Darstellung von Beobachtungen und Messungen. Insbesondere können auch die statistischen Eigenarten der Daten, z.B. Verteilungen, sehr übersichtlich dargestellt werden. Die sogenannten **High-Level-Plot-Befehle** (Tabelle 9.8) öffnen ein Graphikausgabefenster, in dem bestimmte Graphiktypen aufgebaut werden, die anschließend mit weiteren (**Low-Level**)-Funktionen (Tabelle 9.9) modifiziert bzw. ergänzt werden können. Die Graphikfunktionen sind in der Grundausstattung von **R** in dem Paket (Bibliothek) **graphics** zusammengefasst.



Beispiel: Mit den Beispieldaten aus der Studie zum Herzinfarkt wird in Abbildung 9.4 die Verteilung zum Cholesterin (Histogramm), ein Box-Plot für den Vergleich der Studiengruppen hinsichtlich des Nüchternblutzuckers und eine Punktwolke zum systolischen und diastolischen Blutdruck (von links nach rechts) gezeigt.

Mit dem Befehl **hist()** kann ein Histogramm zu den Cholesterinwerten gezeichnet werden. Das Hauptargument dieser Funktion ist ein Vektor mit den Werten, deren Verteilung im Rahmen der Statistik näher untersucht werden soll. Die Wahl einer geeigneten Klasseneinteilung (**breaks=**), die Skalierung der Achsen (**xlim=** und **ylim=**) und die Beschriftung der Graphik (**xlab=**, **ylab=** und **main=**) kann in zusätzlichen Argumenten festgelegt werden.

Tabelle 9.8 Einige Graphikfunktionen (high-level) in R

plot (<i>vect</i>), plot (<i>vect1</i> , <i>vect2</i> , <i>type=</i>)	erstellt eine Punktwolke der Werte eines Vektors. Werden zwei Vektoren als Argumente verwendet, dann werden diese in einem kartesischen Koordinatensystem als Punktwolke eingetragen. Die Punkte können mit dem zusätzlichen Argument <i>type=l</i> auch durch eine Linie miteinander verbunden werden (Kurven-, Liniendiagramm).
barplot (<i>vect</i>)	erstellt ein Balkendiagramm zu den Werten eines Vektors z.B. für die Darstellung von Häufigkeiten.
piechart (<i>vect</i>)	erstellt ein Tortendiagramm.
boxplot (<i>vect1</i> , <i>vect2</i> ,...)	erstellt ein Boxplot-Diagramm zur Verteilung der Werte in einem Vektor; werden mehrere Vektoren als Argumente angegeben, dann werden die Boxplots für den Vergleich nebeneinander gezeichnet.
dotchar (<i>vect</i>)	erstellt ein Punktdiagramm zu Messwerten.
hist (<i>vect</i>)	erstellt ein Histogramm zu den Messwerten in einem Vektor.

Die Funktion **boxplot()** erstellt Boxplots auf der Grundlage von Medianwerten und Quartilen. Die Werte können einerseits in der Form einzelner Vektoren (hier die Werte zum Blutzucker für die beiden Studiengruppen), andererseits aber auch durch die Angabe einer Modellgleichung der Form „*y ~ faktor*“ spezifiziert werden.

```
> hist (Chol)
> boxplot (Blutz [ Gruppe==" Infarkt "], Blutz [ Gruppe==" Kontrolle "])
> plot (RRsyst, RRdias)
```

Die Funktion **plot()** hat in **R** eine zentrale Bedeutung. Sie ermöglicht die graphische Darstellung zahlreicher Objekte (generic function). In der einfachsten Form können die Werte zweier Vektoren (gleicher Länge) im Koordinatensystem als Funktionsverlauf (*type="l"*) oder als Punktwolke (*type="p"*) dargestellt werden. In dem Beispiel in Abbildung 9.4 wird ein Zusammenhang zwischen

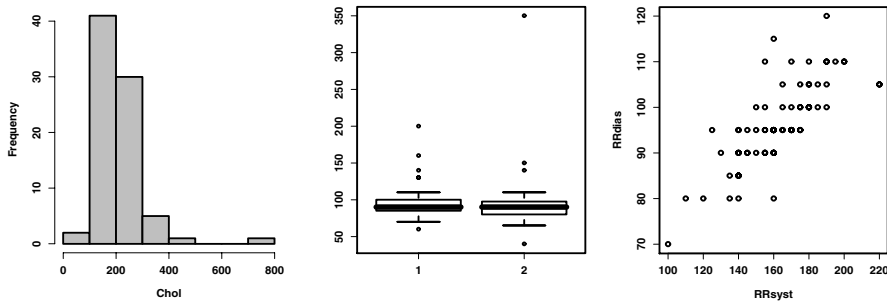


Abb. 9.4 Beispiele für statistische Graphiken in **R**; Histogramm, Boxplot und Punktwolke (von links nach rechts)

systolischem und diastolischem Blutdruck untersucht. Die Plotsymbole (`pch`) und die Linienart (`lty`) können durch zusätzliche Argumente gewählt werden (vgl. Abbildung 9.5).

Symbole (<code>pch</code>)										Linien (<code>lty</code>)									
1	○	5	◇	9	⊕	13	⊗	17	▲	7									
2	△	6	▽	10	⊕	14	⊗	18	◆	6									
3	+	7	⊗	11	⊗	15	■	19	●	5									
4	×	8	*	12	⊗	16	●	20	●	4									
										3									
										2									
										1									

Abb. 9.5 Unterschiedliche Plotsymbole (`pch`) und Linienarten (`lty`) in den Graphikfunktionen

Eine ausführliche Beschreibung der Argumente zu den Graphikfunktionen ist in der Dokumentation zu diesen Funktionen nachzulesen. Ergänzungen zu einer aktuell erstellten Graphik können mit speziellen Hilfsfunktionen (low level) nachträglich eingezeichnet werden. Eine Übersicht zu ausgewählten Aspekten befindet sich in Tabelle 9.9.

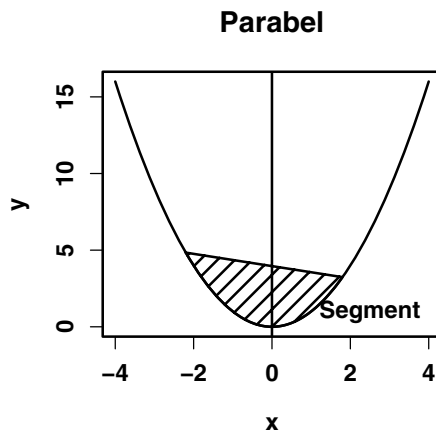


Abb. 9.6 Parabelsegment mit Ergänzungen aus Graphik-Funktionen (low level) in **R**

Tabelle 9.9 Einige Graphikfunktionen (low-level) in R

lines (<i>vect1</i> , <i>vect2</i>)	verbindet die Punkte, deren Koordinaten über die Vektoren <i>vect1</i> und <i>vect2</i> festgelegt sind, durch eine Linie.
points (<i>vect1</i> , <i>vect2</i>)	fügt einzelne Punkte mit den Koordinaten in <i>vect1</i> und <i>vect2</i> in eine Graphik ein.
abline (<i>a</i> , <i>b</i>)	zeichnet eine Linie mit der Steigung <i>b</i> und dem y-Achsenabschnitt <i>a</i> (lineare Funktion).
abline (<i>h</i> = <i>y</i>)	zeichnet eine horizontale Linie bei <i>y</i> .
abline (<i>v</i> = <i>x</i>)	zeichnet eine vertikale Linie bei <i>x</i> .
text (<i>vect1</i> , <i>vect2</i> , <i>labels</i> =)	fügt Texte <i>labels</i> an den durch die Vektoren <i>vect1</i> und <i>vect2</i> festgelegten Koordinaten in die Graphik ein.
polygon (<i>vect1</i> , <i>vect2</i> , <i>density</i> =)	zeichnet einen Polygonzug (geschlossene Linie) mit den Koordinaten aus <i>vect1</i> und <i>vect2</i> ; die eingeschlossene Fläche kann durch ein zusätzliches Argument <i>density</i> farbig markiert oder schraffiert werden.
legend (<i>vect1</i> , <i>vect2</i>)	fügt Legenden (Erklärungstexte) an den festgelegten Koordinaten ein.
title (<i>text</i>)	fügt nachträglich eine Überschrift ein.

Mit den Funktionen **lines()** und **points()** können zum Beispiel nachträglich Linien und Punkte in eine aktuelle Graphik eingezeichnet werden. Durch **legend()** und **title()** werden Legenden und Überschriften zum besseren Verständnis einer Graphik eingefügt.



Beispiel: Mit den Graphikfunktionen aus Tabelle 9.8 und Tabelle 9.9 soll eine Parabel gezeichnet werden, in der ein spezielles Segment zu markieren ist. Das Ergebnis ist in Abbildung 9.6 dargestellt.

```
> par(ps=14, font=2, font.axis=2, font.lab=2, font.main=2, font.sub=2, lwd=2)
> x <- seq(-4, +4, by=0.2)
> y <- x^2
> plot(x, y, type="l")
> abline(v=0)
> polygon(x[10:30], y[10:30], density=10)
> text(2.5, 1, "Segment")
> title("Parabel")
```

Für die weitere individuelle Gestaltung von Graphiken, unter anderem hinsichtlich der Skalierung und Beschriftung, können zahlreiche zusätzliche Argumente in den entsprechenden Funktionen spezifiziert werden. Eine Auswahl ist in Tabelle 9.10 zusammengestellt. Dabei können Aspekte, die für alle möglichen graphischen Darstellungen eine gleiche oder ähnliche Bedeutung haben, zum Beispiel hinsichtlich der Schriftart, Schriftgröße oder Farben für Linien und Hintergrund, mit einer zusätzlichen Funktion **par()** fest eingestellt werden. Mit dem Befehl **?par** wird eine Übersicht zu allen Argumenten dieser Funktion angezeigt; einzelne zur Zeit gültige Einstellungen können mit **par("arg")** nachgesehen werden.

Sollen mehrere Graphiken in einer Darstellung, z.B. wie in Abbildung 9.5, zusammengefasst werden, dann kann ein „Zeichenblatt“ (Graphikfenster) in mehrere Felder aufgeteilt werden. Zum Beispiel wird mit dem Befehl **par(mfrow=c(1,3))** das Blatt in eine Zeile mit drei Spalten geteilt. Somit können drei Graphiken nebeneinander auf einem Blatt erstellt werden (flexibler sind die Gestaltungsmöglichkeiten mit der Funktion **layout()**).

Tabelle 9.10 Spezielle Argumente in den Graphikfunktionen und in der Funktion **par()**

xlab ="label" ylab ="label"	fügt eine Beschriftung (<i>label</i>) an den Achsen ein.
xlim =c(<i>min</i> , <i>max</i>) ylim =c(<i>min</i> , <i>max</i>)	skaliert die Achsen in dem durch die Werte <i>min</i> und <i>max</i> begrenzten Bereich.
type ="p"	Typ der Darstellung: p-Punkte, l-Linie, etc.
lty = <i>n</i> lwd = <i>n</i>	Linientyp (Nummer von 1 - 7, siehe auch Abb. 9.5). Linienstärke als positive ganze Zahl.
pch = <i>n</i> pch = "c"	Typ der Plotsymbole (Nummer von 1 - 20, siehe auch Abb. 9.5). Zeichen oder Namen können auch frei zugeordnet werden.
weitere Parameter zur Gestaltung von Graphiken sind zum Beispiel:	
font = <i>n</i> ps = <i>n</i> col = "col" bg = "col" cex = <i>x</i> bty ="l" mfrow = c(<i>nr</i> , <i>nc</i>)	Auswahl einer Schriftart. Größe von Texten und Symbolen (in Punkten). Farbe, die aus 657 Möglichkeiten ausgewählt werden kann (eine Übersicht gibt die Funktion colors()). Farbe für den Hintergrund. Skalierungsfaktor (character expansion), um den Texte und Symbole in der Graphik vergrößert oder verkleinert werden. Umrahmungen (Koordinaten-Achsen), zum Beispiel auf 2 oder 3 Seiten ("l", "7", "c") der Darstellung. Mit "n" können die Umrahmungen bzw. Achsen auch ganz unterdrückt werden. Zahl der Graphiken, die in einer Darstellung kombiniert zusammengefasst werden sollen.

Die Erzeugung einer Graphik erfolgt in **R** stets mit Bezug auf ein aktuelles **Ausgabegerät** (device). Als Standard verwendet **R** ein getrenntes Graphikfenster. Zusätzliche Ausgabefenster können unter dem Betriebssystem Windows mit der Funktion **windows()** geöffnet und mit der Funktion **dev.off()** auch wieder geschlossen werden. Mit der Funktion **dev.set()** kann ein bestimmtes Ausgabefenster ausgewählt werden. An Stelle eines Fensters kann die Ausgabe auch in unterschiedlichen Graphikformaten (PDF-, BMP-, JPEG-, PNG- oder Postscript-Format) in eine externe Datei geschrieben werden. Die verschiedenen Funktionen zur Steuerung der Graphikausgabe sind in dem Paket **grDevices** zusammengefasst.

Übungsaufgaben zu ausgewählten Themen

Wahrscheinlichkeitsrechnung

- (1) Zwei Würfel werden geworfen. Wie groß ist die Wahrscheinlichkeit, dass die geworfene Augensumme 7 oder 11 beträgt?
- (2) Drei Geschütze schießen je einmal. Sie treffen mit einer Wahrscheinlichkeit von 0,1, 0,2 und 0,3. Gefragt ist nach der Trefferwahrscheinlichkeit insgesamt.
- (3) Die Verteilung der Geschlechter unter den Neugeborenen (Knaben : Mädchen) ist nach langjährigen Beobachtungen 514 : 486. Das Auftreten blonder Haare habe bei uns die relative Häufigkeit 0,15. Geschlecht und Haarfarbe seien stochastisch unabhängig. Wie groß ist die relative Häufigkeit eines blonden Knaben?
- (4) Wie groß ist die Wahrscheinlichkeit, mit einem Würfel in 4 Würfen wenigstens einmal die 6 zu werfen?
- (5) In wie vielen Würfen ist mit 50%-iger Wahrscheinlichkeit die 6 wenigstens einmal zu erwarten?
- (6) Wie groß ist die Wahrscheinlichkeit, mit einer Münze 5-, 6-, 7-, 10-mal hintereinander Wappen zu werfen?

Mittelwert und Standardabweichung

- (7) Berechne Mittelwert und Standardabweichung der Häufigkeitsverteilung

x	5	6	7	8	9	10	11	12	13	14	15	16
n	10	9	94	318	253	153	92	40	26	4	0	1

- (8) Berechne den Medianwert, den Mittelwert, die Standardabweichung, die Maßzahlen für die Schiefe der Stichprobenverteilung:
 62, 49, 63, 80, 48, 67, 53, 70, 57, 55, 39, 60, 65, 56, 61, 37
 63, 58, 37, 74, 53, 27, 94, 61, 46, 63, 62, 58, 75, 69, 47, 71,
 38, 61, 74, 62, 58, 64, 76, 56, 67, 45, 41, 38, 35, 40.
- (9) Zeichne die Häufigkeitsverteilung und berechne Mittelwert, Standardabweichung, Schiefe und Wölbung nach dem Momentenverfahren anhand der folgenden klassierten Daten

Klassengrenzen	Häufigkeiten
71,0 - 73,9	7
74,0 - 75,9	31
76,0 - 77,9	42
78,0 - 79,9	54
80,0 - 81,9	33
82,0 - 83,9	24
84,0 - 85,9	22
86,0 - 87,9	8
88,0 - 89,9	4
Insgesamt	225

Binomialkoeffizient

- (10) Angenommen, 8 Insektizide sind jeweils paarweise in ihrer Wirkung auf Mücken zu testen. Wie viele Versuche müssen durchgeführt werden?
- (11) Durchschnittlich sterben 10% der von einer bestimmten Krankheit befallenen Patienten. Wie groß ist die Wahrscheinlichkeit, dass von 5 Patienten, die an dieser Krankheit leiden, (a) alle geheilt werden, (b) genau 3 sterben werden, (c) mindestens 3 sterben werden?
- (12) Wie groß ist die Wahrscheinlichkeit, dass 5 einem gut gemischten Kartenspiel (52 Karten) entnommene Spielkarten vom Karo-Typ sind?
- (13) Ein Würfel wird 12mal geworfen. Wie groß ist die Wahrscheinlichkeit, dass die Augenzahl 4 genau zweimal erscheint?
- (14) Ein Seminar werde von 13 Studentinnen und 18 Studenten besucht. Wie viele Möglichkeiten gibt es für die Auswahl eines Komitees, bestehend aus 2 Studentinnen und 3 Studenten?

Binomialverteilung

- (15) Wie groß ist die Wahrscheinlichkeit, in 10 Münzwürfen fünfmal Wappen zu erzielen?
- (16) Die Wahrscheinlichkeit für einen Dreißigjährigen, das kommende Jahr zu überleben, betrage laut Sterbetafel $p = 0,99$. Wie groß ist die Wahrscheinlichkeit, dass von 10 Dreißigjährigen 9 das kommende Jahr überleben werden?
- (17) Wie groß ist die Wahrscheinlichkeit dafür, dass unter 100 Würfeln mit einem Würfel sich genau 25mal eine 6 befindet?
- (18) Zwanzig Wochentage werden in einem Zufallsprozess ausgewählt. Wie groß ist die Wahrscheinlichkeit, dass 5 von ihnen auf einen bestimmten Tag in der Woche – sagen wir auf einen Sonntag – fallen?
- (19) Angenommen, dass im Durchschnitt 33% der im Krieg eingesetzten Schiffe versenkt werden. Wie groß ist die Wahrscheinlichkeit, dass von 6 Schiffen (a) genau 4, (b) wenigstens 4 wieder zurückkehren?
- (20) Hundert Münzen werden geworfen. Wie groß ist die Wahrscheinlichkeit, dass genau 50 auf die Wappenseite fallen?

- (21) Eine Urne enthalte 2 weiße und 3 schwarze Bälle. Wie groß ist die Wahrscheinlichkeit, dass in 50 Zügen mit Zurücklegen genau 20 weiße Bälle gezogen werden?

Poisson-Verteilung

- (22) Ein hungriger Frosch fange im Durchschnitt 3 Fliegen pro Stunde. Wie groß ist die Wahrscheinlichkeit, dass er in einer Stunde keine Fliege erwischt?
- (23) Angenommen, die Wahrscheinlichkeit, das Ziel zu treffen, sei bei jedem Schuss $p = 0,002$. Wie groß ist die Wahrscheinlichkeit, genau 5 Treffer zu erzielen, wenn insgesamt $n = 1000$ Schüsse abgegeben werden? Dieses Beispiel ist etwas verschroben aber numerisch einfach.
- (24) Die Wahrscheinlichkeit der Produktion eines fehlerhaften Artikels in einem Industriebetrieb sei $p = 0,005$. Dieser Artikel werde in Kisten zu je 200 Stück verpackt. Wie groß ist die Wahrscheinlichkeit, dass in einer Kiste genau 4 fehlerhafte Artikel vorhanden sind?
- (25) In einem Warenhaus wird ein Artikel sehr selten verlangt, beispielsweise im Mittel in einer Woche nur 5mal. Wie groß ist die Wahrscheinlichkeit, daß der Artikel in einer bestimmten Woche k mal verlangt wird?
- (26) Angenommen, 5% aller Schulkinder seien Brillenträger. Wie groß ist die Wahrscheinlichkeit, daß in einer Schulklasse von 30 Kindern keines, 1 Kind, 2 bzw. 3 Kinder eine Brille tragen?

Testverfahren

- (27) Mit Hilfe eines Zufallsprozesses werden einer normalverteilten Grundgesamtheit 16 Stichprobenelemente mit $\bar{x} = 41,5$ und $s = 2,795$ entnommen. Gibt es Gründe für die Ablehnung der Hypothese, daß der Erwartungswert der Grundgesamtheit 43 sei ($\alpha = 0,05$)?
- (28) Prüfe die Gleichheit der Varianzen der beiden Stichproben A und B auf dem 5%-Niveau mit Hilfe des F -Tests.

$A:$ 2,33 4,64 3,59 3,45 3,64 3,00 3,41 2,03 2,80 3,04

$B:$ 2,08 1,72 0,71 1,65 2,56 3,27 1,21 1,58 2,13 2,92

- (29) Prüfe auf dem 5%-Niveau die Gleichheit der zentralen Tendenz (H_0) zweier unabhängiger Stichproben A und B (a) mit Hilfe des Schnelltests von Tukey, (b) mit Hilfe des U -Tests.

$A:$ 2,33 4,64 3,59 3,45 3,64 3,00 3,41 2,03 2,80 3,04

$B:$ 2,08 1,72 0,71 1,65 2,56 3,27 1,21 1,58 2,13 2,92

- (30) Zwei Schlafmittel A und B wurden jeweils an denselben 10 an Schlaflosigkeit leidenden Patienten getestet (Student 1908) [Stu08]. Dabei ergaben sich für die Schlafverlängerung in Stunden die folgenden Werte:

Patient	1	2	3	4	5	6	7	8	9	10
A	1,9	0,8	1,1	0,1	-0,1	4,4	5,5	1,6	4,6	3,4
B	0,7	-1,6	-0,2	-1,2	-0,1	3,4	3,7	0,8	0,0	2,0
Differenz	1,2	2,4	1,3	1,3	0,0	1,0	1,8	0,8	4,6	1,4

Besteht zwischen A und B auf dem 1%-Niveau ein Unterschied? Formuliere die Nullhypothese und prüfe sie (a) mit dem t -Test für Paardifferenzen und (b) mit dem Maximum-Test.

- (31) Prüfe die Gleichheit der zentralen Tendenz (H_0) zweier verbundener Stichproben A und B auf dem 5%-Niveau anhand der folgenden Tests für Paardifferenzen: (a) t -Test, (b) Wilcoxon-Test, (c) Maximum-Test.

Nummer	1	2	3	4	5	6	7	8	9
A	34	48	33	37	4	36	35	43	33
B	47	57	28	37	18	48	38	36	42

- (32) Gregor Mendel erhielt bei einem Erbsenversuch 315 runde gelbe, 108 runde grüne, 101 kantige gelbe und 32 kantige grüne Erbsen. Stehen diese Zahlen im Einklang mit der Theorie, nach der sich die vier Häufigkeiten wie 9 : 3 : 3 : 1 verhalten ($\alpha = 0,05$)?
- (33) Stellt die folgende Häufigkeitsverteilung eine zufällige Stichprobe dar, die einer Poisson-Grundgesamtheit mit dem Parameter $\lambda = 10,44$ entstammen könnte? Prüfe die Anpassung auf dem 5%-Niveau mit Hilfe des χ^2 -Tests.

Anzahl der $\underline{E}$ reignisse:	0	1	2	3	4	5	6	7	8					
Beobachtete $\underline{H}$ äufigkeiten:	0	5	14	24	57	111	197	278	378					
E :	9	10	11	12	13	14	15	16	17	18	19	20	21	22
H :	418	461	433	413	358	219	145	109	57	43	16	7	8	3

Korrelation, Regression

- (34) Prüfe die statistische Signifikanz von $r = 0,5$ auf dem 5%-Niveau ($n = 16$).
- (35) Wie groß muss r sein, damit er für $n = 16$ auf dem 5%-Niveau statistisch signifikant ist?
- (36) Schätze die Regressionsgeraden und den Korrelationskoeffizienten für die folgenden Wertepaare:

x	22	24	26	26	27	27	28	28	29	30	30	30	31	32	33	34	35	35	36	37
y	10	20	20	24	22	24	27	24	21	25	29	32	27	27	30	27	30	31	30	32

Unterscheidet sich der Korrelationskoeffizient auf den 0,1%-Niveau statistisch signifikant von Null?

- (37) Ein auf 19 Beobachtungspaaren basierender Korrelationskoeffizient weise den Wert 0,65 auf.
(a) Kann diese Stichprobe einer Grundgesamtheit mit dem Parameter $\rho = 0,35$ entstammen ($\alpha = 0,05$)? (b) Schätze aufgrund der Stichprobe den 95%-Vertrauensbereich für ρ . (c) Wenn eine zweite Stichprobe, die ebenfalls aus 19 Beobachtungspaaren besteht, einen Korrelationskoeffizienten $r = 0,30$ aufweist, können dann beide Stichproben einer gemeinsamen Grundgesamtheit entstammen ($\alpha = 0,05$)?
- (38) Passe den Werten

x	0	1	2	3	4	5	6
y	125	209	340	561	924	1525	2512

eine Funktion vom Typ $y = ab^x$ an.

(39) Passe den Werten

x	273	283	288	293	313	333	353	373
y	29,4	33,3	35,2	37,2	45,8	55,2	65,6	77,3

eine Funktion vom Typ $y = ab^x$ an.

(40) Passe den folgenden Werten eine Parabel zweiten Grades an:

x	7,5	10,0	12,5	15,0	17,5	20,0	22,5
y	1,9	4,5	10,1	17,6	27,8	40,8	56,9

(41) Passe den folgenden Werten eine Parabel zweiten Grades an:

x	1,0	1,5	2,0	2,5	3,0	3,5	4,0
y	1,1	1,3	1,6	2,1	2,7	3,4	4,1

Test auf Unabhängigkeit oder Homogenität

- (42) Die Häufigkeiten einer Vierfeldertafel seien: $a = 140$, $b = 60$, $c = 85$, $d = 90$. Prüfe die Unabhängigkeit auf dem 0,1%-Niveau.
- (43) Die Häufigkeiten einer Vierfeldertafel seien: $a = 605$, $b = 135$, $c = 195$, $d = 65$. Prüfe die Unabhängigkeit auf dem 5%-Niveau.
- (44) Die Häufigkeiten einer Vierfeldertafel seien: $a = 620$, $b = 380$, $c = 550$, $d = 450$. Prüfe die Unabhängigkeit auf dem 1%-Niveau.
- (45) Prüfe die $2 \cdot 6$ -Feldertafel

13	10	10	5	7	0
2	4	9	8	14	7

auf Homogenität ($\alpha = 0,01$).

(46) Prüfe die Unabhängigkeit und Symmetrie der Kontingenztafel

102	41	57
126	38	36
161	28	11

auf dem 1%-Niveau.

- (47) Prüfe, ob die beiden Stichprobenverteilungen I und II derselben Grundgesamtheit entstammen können ($\alpha = 0,05$). Verwende die Formel von Brandt-Snedecor zur Prüfung der Homogenität zweier Stichproben.

Kategorie	Häufigkeiten		Insgesamt
	I	II	
1	160	150	310
2	137	142	279
3	106	125	231
4	74	89	163
5	35	39	74
6	29	30	59
7	28	35	63
8	29	41	70
9	19	22	41
10	6	11	17
11	8	11	19
12	13	4	17
Insgesamt	644	699	1343

(48) Prüfe die Homogenität der folgenden Tafel auf dem 5%-Niveau.

23	5	12
20	13	10
22	20	17
26	26	29

Varianzanalyse

(49) Prüfe die drei unabhängigen Stichproben A , B , C auf Gleichheit der Erwartungswerte ($\alpha = 0,05$) (a) varianzanalytisch, (b) anhand der H -Tests.

A : 40, 34, 84, 46, 47, 60

B : 59, 92, 117, 86, 60, 67, 95, 40, 98, 108

C : 92, 93, 40, 100, 92

(50) Gegeben

B							Summe
A	B_1	B_2	B_3	B_4	B_5	B_6	
A_1	9,5	11,5	11,0	12,0	9,3	11,5	64,8
A_2	9,6	12,0	11,1	10,8	9,7	11,4	64,6
A_3	12,4	12,5	11,4	13,2	10,4	13,1	73,0
A_4	11,5	14,0	12,3	14,0	9,5	14,0	75,3
A_5	13,7	14,2	14,3	14,6	12,0	13,2	82,0
Summe	56,7	64,2	60,1	64,6	50,9	63,2	359,7

Prüfe mögliche Spalten und Zeileneffekte auf dem 1 %-Niveau.

- (51) Drei Bestimmungsmethoden werden an 10 Proben verglichen. Prüfe mit Hilfe des Friedman-Tests (a) die Gleichheit der Methoden ($\alpha = 0,001$), (b) die Gleichheit der Proben ($\alpha = 0,05$).

Probe	Bestimmungsmethode		
	A	B	C
1	15	18	9
2	22	25	20
3	44	43	25
4	75	80	58
5	34	33	31
6	15	16	11
7	66	64	45
8	56	57	40
9	39	40	27
10	30	34	31

Lösungen der Übungsaufgaben

Wahrscheinlichkeitsrechnung

1. Die Summe 7 lässt sich auf sechs verschiedenen Wegen erhalten, die Summe 11 auf nur zwei, damit wird

$$P = \frac{6}{36} + \frac{2}{36} = \frac{2}{9} = 0,222$$

2. Die Trefferwahrscheinlichkeit insgesamt beträgt knapp 50%.

$$P(A+B+C)=P(A)+P(B)+P(C) - P(AB) - P(AC) - P(BC) + P(ABC)$$

$$P(A+B+C)=0,1 + 0,2 + 0,3 -0,02 -0,03 -0,06 + 0,006 = 0,496$$

3. $P = 0,514 \cdot 0,15 = 0,0771$

In etwa 8% aller Geburten sind blonde Knaben zu erwarten.

4. $1 - (5/6)^4 = 0,5177$

In einer langen Reihe von Würfeln ist in etwa 52% aller Fälle mit diesem Ereignis zu rechnen.

$$5. P = \left(\frac{5}{6}\right)^n = \frac{1}{2}; \quad n = \frac{\lg 2}{\lg 6 - \lg 5} = \frac{0,3010}{0,7782 - 0,6990} \simeq 4$$

6. Die Wahrscheinlichkeiten sind $(1/2)^5, (1/2)^6, (1/2)^7, (1/2)^{10}$, gerundet 0,031, 0,016, 0,008, 0,001.

Mittelwert und Standardabweichung

7. $\bar{x} = 9,015 \quad s = 1,543$

```
> x <- c( 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16)
> n <- c(10, 9, 94, 318, 253, 153, 92, 40, 26, 4, 0, 1)
> summe <- sum(x*n); N <- sum(n)
> mittelwert <- summe/N; mittelwert
[1] 9.015
> stdabw <- sqrt(sum(n*(x-mittelwert)^2)/(N-1)); stdabw
[1] 1.543748
```

8. Statistiken

$$\bar{x} = 57,3 \quad \text{Schiefe I} = -0,37$$

$$s = 13,8 \quad \text{Schiefe II} = -0,18$$

$$\tilde{x} = 59,0 \quad \text{Schiefe III} = -0,39$$

$$\text{Wölbung} = 0,25$$

```
> x <- c(62, 49, 63, 80, 48, 67, 53, 70, 57, 55, 39, 60, 65, 56, 61, 37,
+      63, 58, 37, 74, 53, 27, 94, 61, 46, 63, 62, 58, 75, 69, 47, 71,
+      38, 61, 74, 62, 58, 64, 76, 56, 67, 45, 41, 38, 35, 40)
> mean(x); sd(x); median(x)
[1] 57.28261
[1] 13.78028
[1] 59
```

```
> mittelwert <- mean(x); stdabw = sd(x); medianwert <- median(x)
>
> schiefel <- (3*(mittelwert-medianwert))/stdabw; as.numeric(schiefel)
[1] -0.3738802
>
> dezile <- quantile(x, probs = seq(0, 1, 0.10), names = TRUE, type = 4)
> dz1 <- dezile[2]; dz9 <- dezile[10]
> schiefelI <- (dz9 + dz1 - 2*medianwert)/(dz9-dz1); as.numeric(schiefelI)
[1] -0.1758242
>
> quartile <- quantile(x, probs = seq(0, 1, 0.25), names = TRUE, type = 4)
> Q1 <- quartile[2]; Q3 <- quartile[4]
> schiefelII <- (Q3 + Q1 - 2*medianwert)/(Q3 - Q1); as.numeric(schiefelII)
[1] -0.3888889
>
> woelbung <- (Q3 - Q1)/(2*(dz9 - dz1)); as.numeric(woelbung)
[1] 0.2472527
```

9. $\bar{x} = 79,658$ Schiefe = 0,426
 $s^2 = 13,505$ Wölbung = -0,437

```
> x <- seq(73, 89, by =2)
> d <- 81; b <- 2
> f <- c( 7, 31, 42, 54, 33, 24, 22, 8, 4)
> z <- (x - d) / b
> n <- sum(f)
>
> m1 <- sum(f*z) / n; m1
[1] -0.6711111
> m2 <- sum(f*z^2) / n; m2
[1] 3.826667
> m3 <- sum(f*z^3) / n; m3
[1] -4.457778
> m4 <- sum(f*z^4) / n; m4
[1] 31.45333
>
> mittelwert <- d + b*m1; mittelwert
[1] 79.65778
> varianz <- b^2*(m2 - m1^2); varianz
[1] 13.50511
> schiefe <- (b^3*(m3 - 3*m1*m2 + 2*m1^3))/varianz^(3/2); schiefe
[1] 0.4258775
> woelbung <- (b^4*(m4 - 4*m1*m3 + 6*m1^2*m2 - 3*m1^4))/varianz^2 - 3; woelbung
[1] -0.4367527
>
> library(e1071)
> x1 <- c(rep(73, 7), rep(75,31), rep(77,42), rep(79,54), rep(81,33),
+ rep(83,24), rep(85,22), rep(87, 8), rep(89, 4))
> mean(x1); var(x1); skewness(x1); kurtosis(x1)
[1] 79.65778
[1] 13.56540
[1] 0.4230415
[1] -0.4594865
```

Binomialkoeffizient

$$10. P = {}_8C_2 = \frac{(8)}{(2)} = \frac{8!}{6! \cdot 2!} = \frac{8 \cdot 7}{2} 28$$

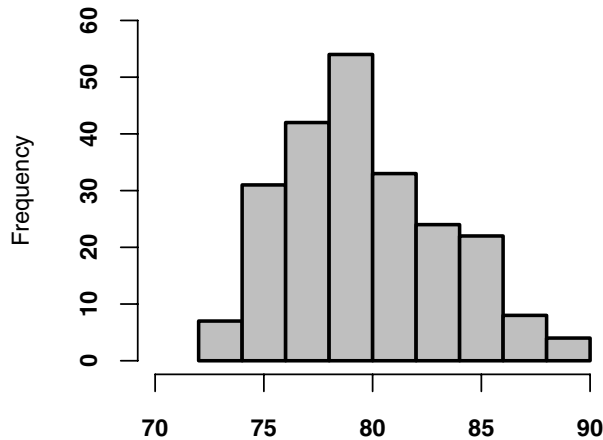


Abb. 10.1 Histogramm zu Beispieldaten aus Aufgabe 9

11. Zu a: $P = 0,90^5 = 0,59049$

Zu b: vgl. ${}_5C_3 = 5!/(3! \cdot 2!) = 5 \cdot 4/2 \cdot 1 = 10$
 $P = 10 \cdot 0,90^2 \cdot 0,10^3 = 0,00810$

Zu c: vgl. ${}_5C_3 = 10$, ${}_5C_4 = 5$
 $P = 10 \cdot 0,90^2 \cdot 0,10^3 + 5 \cdot 0,90 \cdot 0,10^4 + 0,10^5$
 $P = 0,00810 + 0,00045 + 0,00001 = 0,00856$

```
> p <- 0.10
> (1-p)^5                                # zu a)
[1] 0.59049
> choose(5, 3) * (1-p)^2 * p^3           # zu b)
[1] 0.0081
> 1 - pbinom(2, 5, p)                    # zu c)
[1] 0.00856
```

$$12. P = \frac{{}_{13}C_5}{{}_{52}C_5} = \frac{13! \cdot 47! \cdot 5!}{8! \cdot 5! \cdot 52!} = \frac{13 \cdot 12 \cdot 11 \cdot 10 \cdot 9}{52 \cdot 51 \cdot 50 \cdot 49 \cdot 48}$$

$$P \simeq \frac{11 \cdot 3}{17 \cdot 5 \cdot 49 \cdot 16} = \frac{33}{66\,640} = 0,0004952$$

$$P \simeq 0,0005 \quad \text{oder} \quad 1:2000.$$

13. Für die Auswahl zweier aus insgesamt zwölf Objekten bieten sich ${}_{12}C_2 = 12!/(10! \cdot 2!) = 12 \cdot 11/(2 \cdot 1)$ Möglichkeiten. Die Wahrscheinlichkeit, 2 Vieren und 10 Nicht-Vieren zu würfeln, beträgt $(1/6)^2 \cdot (5/6)^{10} = 5^{10}/6^{12}$. Die Wahrscheinlichkeit, daß die Augenzahl 4 in 12 Würfeln genau zweimal erscheint, beträgt damit

$$P = \frac{12 \cdot 11 \cdot 5^{10}}{2 \cdot 1 \cdot 6^{12}} = \frac{11 \cdot 5^{10}}{6^{11}} = 0,296.$$

In einer langen Serie von Zwölferwürfen mit intaktem Würfel ist in etwa 30% der Fälle mit dem jeweils zweimaligen Erscheinen der Augenzahl 4 zu rechnen.

14. Die Antwort ist das Produkt der Möglichkeiten, die Vertreter der beiden Geschlechter auszuwählen, d. h.

$$P = {}_{13}C_2 \cdot {}_{18}C_3 = \frac{13!}{11! \cdot 2!} \cdot \frac{18!}{15! \cdot 3!} = \frac{13 \cdot 12}{2 \cdot 1} \cdot \frac{18 \cdot 17 \cdot 16}{3 \cdot 2 \cdot 1}$$

$$P = 13 \cdot 18 \cdot 17 \cdot 16 = 63\,648$$

Binomialverteilung

$$15. P = {}_{10}C_5 \left(\frac{1}{2}\right)^5 \left(\frac{1}{2}\right)^5 = \frac{10!}{11! \cdot 2!} \cdot \frac{1}{2^{10}} = \frac{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} \cdot \frac{1}{1024} = \frac{252}{1024}$$

$$P = 0,2461$$

In einer langen Serie von jeweils 10 Würfeln ist in knapp 25% der Fälle mit diesem Ereignis zu rechnen.

```
> dbinom(5, 10, 0.5)
[1] 0.2460938
```

$$16. P = {}_{10}C_9 \cdot 0,99^9 \cdot 0,01^1 = 10 \cdot 0,9135 \cdot 0,01 = 0,09135$$

$$17. P = \binom{100}{25} \left(\frac{1}{6}\right)^{25} \left(\frac{5}{6}\right)^{75} = 0,0098. \text{ Bei einer großen Anzahl von Würfeln ist in etwa 1\%}$$

der Fälle mit diesem Ereignis zu rechnen.

$$18. P(X=5) = \frac{20!}{15! \cdot 5!} \left(\frac{6}{7}\right)^{15} \left(\frac{1}{7}\right)^5 = \frac{20 \cdot 19 \cdot 18 \cdot 17 \cdot 16}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} \cdot \frac{6^{15}}{7^{20}}$$

$$P = 0,0914$$

$$19. \text{ Zu a: } P = {}_6C_4 \cdot 0,67^4 \cdot 0,33^2 = 15 \cdot 0,2015 \cdot 0,1089 = 0,3292$$

$$\text{Zu b: } P = \sum_{x=4}^6 {}_6C_x \cdot 0,67^x \cdot 0,33^{6-x} = 0,3292 + 6 \cdot 0,1350 \cdot 0,33 + 0,0905$$

$$P = 0,6804$$

```
> p <- 1/3
> dbinom(4, 6, 1-p) # zu a)
[1] 0.3292181
> pbinom(3, 6, 1-p, lower.tail=FALSE) # zu b)
[1] 0.6803841
```

$$20. P = \frac{100!}{50! \cdot 50!} \cdot \left(\frac{1}{2}\right)^{50} \left(\frac{1}{2}\right)^{50} = 0,0796$$

$$21. P = {}_{50}C_{20} \left(\frac{2}{5}\right)^{20} \left(\frac{3}{5}\right)^{30} = \frac{50!}{20! \cdot 30!} \left(\frac{2}{5}\right)^{20} \left(\frac{3}{5}\right)^{30} = 0,0364$$

Poisson-Verteilung

$$22. P = \frac{\lambda^x \cdot e^{-\lambda}}{x!} = \frac{3^0 \cdot e^{-3}}{0!} = \frac{1 \cdot e^{-3}}{1} = \frac{1}{e^3} = \frac{1}{20,086} \simeq 0,05$$

$$23. \lambda = n \cdot \hat{p} = 1000 \cdot 0,002 = 2$$

$$P = \frac{\lambda^x \cdot e^{-\lambda}}{x!} = \frac{2^5 \cdot e^{-2}}{5!} = 0,0361$$

$$24. \lambda = n \cdot \hat{p} = 200 \cdot 0,005 = 1$$

$$P = \frac{\lambda^x \cdot e^{-\lambda}}{x!} = \frac{1^4 \cdot e^{-1}}{4!} = \frac{0,3679}{24} = 0,0153$$

$$25. P(k, 5) = \frac{5^k \cdot e^{-5}}{k!}$$

$$26. \lambda = n \cdot \hat{p} = 30 \cdot 0,05 = 1,5 \quad P = \frac{\lambda^x \cdot e^{-\lambda}}{x!}$$

$$\text{Kein Kind:} \quad P = \frac{1,5^0 \cdot e^{-1,5}}{0!} = 0,2231$$

$$\text{Ein Kind:} \quad P = \frac{1,5^1 \cdot e^{-1,5}}{1!} = 0,3346$$

$$\text{Zwei Kinder:} \quad P = \frac{1,5^2 \cdot e^{-1,5}}{2!} = 0,2509$$

$$\text{Drei Kinder:} \quad P = \frac{1,5^3 \cdot e^{-1,5}}{3!} = 0,1254$$

```
> n <- 30; p <- 0.05; l <- n*p
>
> dpois(0:3, l)
[1] 0.2231302 0.3346952 0.2510214 0.1255107
```

Testverfahren

$$27. \text{Ja: } \hat{t} = \frac{|41,5 - 43|}{2,795} \cdot \sqrt{16} = 2,15 > t_{15;0,975} = 2,13$$

$$28. \hat{F} = \frac{s_B^2}{s_A^2} = \frac{0,607}{0,542} = 1,12 < F_{9;9;0,95} = 3,18$$

```
> A <- c(2.33, 4.64, 3.59, 3.45, 3.64, 3.00, 3.41, 2.03, 2.80, 3.04)
> B <- c(2.08, 1.72, 0.71, 1.65, 2.56, 3.27, 1.21, 1.58, 2.13, 2.92)
> var.test(B, A, alternative = "two.sided", conf.level = 0.95)

F test to compare two variances

data: B and A
F = 1.1193, num df = 9, denom df = 9, p-value = 0.8694
alternative hypothesis: true ratio of variances is not equal to 1
...
> qf(0.95, 9, 9) # einseitig
[1] 3.178893
```

29. Zu a: $\hat{T} = 10 > 7$; H_0 wird auf dem 5%-Niveau abgelehnt.

Zu b: $\hat{U} = 12 < U_{10;10;0,05} = 27$; H_0 wird gleichfalls abgelehnt.

```
> A <- c(2.33, 4.64, 3.59, 3.45, 3.64, 3.00, 3.41, 2.03, 2.80, 3.04); m <- 10
> B <- c(2.08, 1.72, 0.71, 1.65, 2.56, 3.27, 1.21, 1.58, 2.13, 2.92); n <- 10
> test <- wilcox.test(A, B, alternative = "two.sided"); test

      Wilcoxon rank sum test

data:  A and B
W = 88, p-value = 0.002879
...
> U <- m*n - test$statistic; U
12
```

30. Zu a: $\hat{t} = 4,06 > t_{9;0,995} = 3,25$

Die Nullhypothese – gleiche Wirksamkeit beider Schlafmittel A und B – wird abgelehnt; es ist anzunehmen, daß A wirksamer ist als B .

Zu b: Entscheidung wie bei a.

```
> A <- c(1.9, 0.8, 1.1, 0.1, -0.1, 4.4, 5.5, 1.6, 4.6, 3.4)
> B <- c(0.7, -1.6, -0.2, -1.2, -0.1, 3.4, 3.7, 0.8, 0.0, 2.0)
> diff <- A - B; n <- length(diff); diff
[1] 1.2 2.4 1.3 1.3 0.0 1.0 1.8 0.8 4.6 1.4
> t.test(A, B, alternative = "two.sided", paired = TRUE)

      Paired t-test

data:  A and B
t = 4.0621, df = 9, p-value = 0.002833
alternative hypothesis: true difference in means is not equal to 0
...
> qt(0.995, n-1)
[1] 3.249836
>
```

31. Zu a: $\hat{t} = 2,03 < t_{8;0,975} = 2,31$

Zu b: $\hat{R}_p = 5 > R_{8;0,10} = 6$

Zu c: Der Unterschied ist lediglich auf dem 10%-Niveau gesichert. In allen drei Fällen wird H_0 nicht abgelehnt.

```
> A <- c(34, 48, 33, 37, 4, 36, 35, 43, 33)
> B <- c(47, 57, 28, 37, 18, 48, 38, 36, 42)
> t.test(A, B, paired=TRUE)

      Paired t-test

data:  A and B
t = -2.0279, df = 8, p-value = 0.0771
alternative hypothesis: true difference in means is not equal to 0
...
> wilcox.test(A, B, paired=TRUE)

      Wilcoxon signed rank test with continuity correction

data:  A and B
V = 5, p-value = 0.07969
alternative hypothesis: true mu is not equal to 0
...
```

32. Ja: $\hat{\chi}^2 = 0,47 < \chi_{3;0,05}^2 = 7,815$

```
> obs <- c(315, 108, 101, 32); sum_o <- sum(obs)
> mod <- c(9, 3, 3, 1); sum_m <- sum(mod)
> exp <- mod/sum_m * sum_o
> chi <- sum((obs-exp)^2/exp); chi
[1] 0.470024
> qchisq(0.95, 3)
[1] 7.814728
```

33. Nein: $\hat{\chi}^2 = 43,43 > \chi_{20;0,05}^2 = 31,4$

```
> lambda <- 10.44
> obs <- c(0, 5, 14, 24, 57, 111, 197, 278, 378, 418, 461, 433,
           413, 358, 219, 145, 109, 57, 43, 16, 7, 8, 3)
> exp <- dpois(0:22, lambda) * sum(obs)
> chi <- sum((obs-exp)^2/exp); chi
[1] 45.07203
> qchisq(0.95, 20)
[1] 31.41043
```

Korrelation und Regression

34. $\hat{t} = 2,16 > t_{14;0,975} = 2,14$

$\hat{F} = 4,67 > F_{1;14;0,95} = 4,60$

35. $r^2 \cdot \frac{16-2}{1-r^2} = 4,60; |r| \geq 0,497$

36. $\hat{y} = 1,083x - 6,90$ mit $s_{y,x}^2 = 8,70$
 $\hat{x} = 0,654y + 13,26$ mit $s_{x,y}^2 = 5,25$
 $r = 0,842$
 $\hat{t} = 6,62 > t_{18;0,9995} = 3,92$

```
> x <- c(22, 24, 26, 26, 27, 27, 28, 28, 29, 30, 30, 30, 31, 32, 33, 34, 35, 35, 36, 37)
> y <- c(10, 20, 20, 24, 22, 24, 27, 24, 21, 25, 29, 32, 27, 27, 30, 27, 30, 31, 30, 32)
> mod <- lm(y~x); summary(mod)
```

```
Call:
lm(formula = y ~ x)
...
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -6.9000    4.9592  -1.391    0.181
x              1.0833    0.1638   6.612  3.3e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
...
> cor(x,y)
[1] 0.8416439
> cor.test(x, y)
```

Pearson's product-moment correlation

```
data: x and y
t = 6.6122, df = 18, p-value = 3.297e-06
alternative hypothesis: true correlation is not equal to 0
...
sample estimates: cor 0.8416439
```

37. Zu a: $\hat{z} = 1,639 < 1,96$, ja

Zu b: $0,278 \leq \varrho \leq 0,852$

Zu c: $\hat{z} = 1,159 < 1,96$, ja

38. $\hat{y} = 125 \cdot 1,649^x$

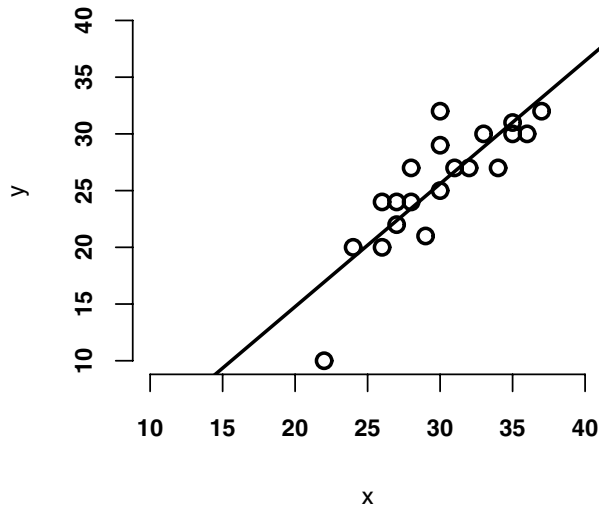


Abb. 10.2 Punktwolke und lineare regression zu den Beispieldaten aus Aufgabe 36

```
> x <- c( 0, 1, 2, 3, 4, 5, 6)
> y <- c(125, 209, 340, 561, 924, 1525, 2512)
> nls(y ~ a * (b^x), start = list(a = 1, b = 1))
Nonlinear regression model
  model: y ~ a * (b^x)
  data: parent.frame()
      a      b
125.411063  1.647970
residual sum-of-squares:  7.414154
```

39. $\hat{y} = 2,4 \cdot 1,009^x$

40. $\hat{y} = 0,2093 \cdot x^2 - 2,633x + 10$

41. $\hat{y} = 0,9500 - 0,0976x + 0,2238x^2$ mit $s_{y \cdot x}^2 = 0,002$

```
> x <- c( 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0)
> y <- c( 1.1, 1.3, 1.6, 2.1, 2.7, 3.4, 4.1)
> nls(y ~ a + b*x + c*x^2, start = list(a = 1, b = 1, c = 1))
Nonlinear regression model
  model: y ~ a + b * x + c * x^2
  data: parent.frame()
      a      b      c
0.95000000 -0.09761905  0.22380952
residual sum-of-squares:  0.008095238
```

Test auf Unabhängigkeit oder Homogenität

42. Da $\hat{\chi}^2 = 17,86 > \chi_{1;0,001}^2 = 10,83$, ist die Unabhängigkeitshypothese abzulehnen.

```
> tab <- matrix(c(140, 60, 85, 90), nrow=2, byrow=TRUE); tab
  [,1] [,2]
[1,] 140  60
[2,]  85  90
> chisq.test(tab, correct=FALSE)
```

```
Pearson's Chi-squared test

data:  tab
X-squared = 17.8571, df = 1, p-value = 2.381e-05

> qchisq(0.999, 1)
[1] 10.82757
```

43. Da $\hat{\chi}^2 = 5,49 > \chi_{1;0,05}^2 = 3,84$, ist die Unabhängigkeitshypothese abzulehnen.

44. Da $\hat{\chi}^2 = 10,09 > \chi_{1;0,01}^2 = 6,635$, ist die Unabhängigkeitshypothese abzulehnen.

45. Da $\hat{\chi}^2 = 20,7082$ größer ist als $\chi_{5;0,01}^2 = 15,086$, wird die Homogenitätshypothese abgelehnt.

```
> tab <- matrix(c(13, 10, 5, 7, 0, 2, 4, 9, 8, 14, 7), nrow=2, byrow=TRUE); tab
      [,1] [,2] [,3] [,4] [,5] [,6]
[1,]  13   10   10   5   7   0
[2,]   2    4    9    8   14   7
> chisq.test(tab, correct=FALSE)
```

```
Pearson's Chi-squared test

data:  tab
X-squared = 20.7077, df = 5, p-value = 0.0009198

Warning message:
Chi-squared approximation may be incorrect in: chisq.test(tab, correct = FALSE)
> qchisq(0.99, 5)
[1] 15.08627
```

46. Da $\hat{\chi}_{\text{unabh.}}^2 = 48,8 > \chi_{4;0,01}^2 = 13,3$, muß die Unabhängigkeitshypothese abgelehnt werden.
Da $\hat{\chi}_{\text{sym}}^2 = 135,97 > \chi_{3;0,01}^2 = 11,345$, ist auch die Symmetrie-Hypothese abzulehnen.

47. $\hat{\chi}^2 = 11,12$

$\chi_{11;0,05}^2 = 19,675$ wird nicht erreicht. Es besteht somit keine Veranlassung, an der Homogenitätshypothese zu zweifeln.

```
> xi <- c(160, 137, 106, 74, 35, 29, 28, 29, 19, 6, 8, 13); x <- sum(xi)
> ni <- c(310, 279, 231, 163, 74, 59, 63, 70, 41, 17, 19, 17); n <- sum(ni)
> stat <- (n^2 / (x*(n-x))) * (sum(xi^2/ni) - x^2/n)
> stat; qchisq(0.95, length(xi)-1)
[1] 11.11813
[1] 19.67514
```

48. Da $\hat{\chi}^2 = 10,88 < \chi_{6;0,05}^2 = 12,59$, ist die Homogenitätshypothese nicht abzulehnen.

Varianzanalyse

49. Zu a: $\hat{F} = 3,86 > F_{2;18;0,05} = 3,55$

Zu b: $\hat{H} = 6,05 > \chi_{2;0,05}^2 = 5,99$

```
> A <- c(40, 34, 84, 46, 47, 60)
> B <- c(59, 92, 117, 86, 60, 67, 95, 40, 98, 108)
> C <- c(92, 93, 40, 100, 92)
> grp <- as.factor(c(rep("A",6), rep("B",10), rep("C",5)))
> val <- c(A, B, C)
> summary(aov(val ~ grp))
              Df Sum Sq Mean Sq F value Pr(>F)
grp           2 4061.6   2030.8   3.8643 0.04015 *
Residuals    18 9459.6    525.5
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
> qf(0.95, 2, 18)
[1] 3.554557
> kruskal.test(list(A, B, C))

Kruskal-Wallis rank sum test

data: list(A, B, C)
Kruskal-Wallis chi-squared = 6.0509, df = 2, p-value = 0.04853

> qchisq(0.95, 2)
[1] 5.991465
```

50.

Variabilität	Summe der Abweichungsquadrate	FG	Mittleres Quadrat	$\hat{F}$	$F_{0.01}$
zwischen den A's	36,41	4	9,102	19,12	4,43
zwischen den B's	28,55	5	5,710	12,00	4,10
Versuchsfehler	9,53	20	0,476		
Gesamtvariabilität	74,49	29			

Multiple Vergleiche der Zeilen- sowie der Spalten-Mittelwerte auf dem 1%-Niveau sind zu empfehlen (vgl. $D_{I, \text{Zeilenmittelwerte}} = 1,80$ und $D_{I, \text{Spaltenmittelwerte}} = 1,84$).

```
> val <- c(9.5, 11.5, 11.0, 12.0, 9.3, 11.5,
+         9.6, 12.0, 11.1, 10.8, 9.7, 11.4,
+         12.4, 12.5, 11.4, 13.2, 10.4, 13.1,
+         11.5, 14.0, 12.3, 14.0, 9.5, 14.0,
+         13.7, 14.2, 14.3, 14.6, 12.0, 13.2)
> A <- as.factor(c(rep("A1", 6), rep("A2", 6), rep("A3", 6), rep("A4", 6), rep("A5", 6)))
> B <- as.factor(rep(c("B1", "B2", "B3", "B4", "B5", "B6"), 5))
> data <- as.data.frame(cbind(A, B, val))
> summary(aov(val ~ A + B))
      Df Sum Sq Mean Sq F value    Pr(>F)
A      4  36.412    9.103   19.108 1.315e-06 ***
B      5  28.547    5.709   11.985 1.855e-05 ***
Residuals 20   9.528    0.476

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

51. Zu a: $\hat{\chi}_R^2 = 13,4 > \chi_{2;0,001}^2 = 13,82$

Zu b: $\hat{\chi}_R^2 = 25,5 > \chi_{9;0,05}^2 = 16,92$; beide Homogenitätshypothesen sind auf den verwendeten Niveaus abzulehnen.

```
> val.b <- matrix(c(15, 22, 44, 75, 34, 15, 66, 56, 39, 30,
+                 18, 25, 43, 80, 33, 16, 64, 57, 40, 34,
+                 9, 20, 25, 58, 31, 11, 45, 40, 27, 31),
+               nr = 3,
+               byrow = TRUE,
+               dimnames = list(1 : 3, c("1", "2", "3", "4", "5", "6", "7", "8", "9", "10")))
> friedman.test(val.b)

Friedman rank sum test

data: val.b
Friedman chi-squared = 25.4665, df = 9, p-value = 0.002496
```

Literaturverzeichnis

Im Text nicht erwähnt sind einige allgemeine und weiterführende Darstellungen, wie z.B. Y. Dodge (2003), D. Rasch (1995) und H. Rinne (2003). Weitere Arbeiten hier genannter und anderer Autoren enthält die 11. Auflage dieses Buches sowie der Guide von L. Sachs [Sac86b].

- [AB60] ANSARI, A.R. ; BRADLEY, R.A.: Rank-sum tests for dispersion. In: *Ann. Math. Statist.* 31 (1960), S. 1174–1189
- [ABGK93] ANDERSEN, P.K. ; BORGAN, O. ; GILL, R.D. ; KEIDING, N.: *Statistical Models Based on Counting Processes*. Springer-Verlag, 1993
- [Act59] ACTON, F.S.: *Analysis of Straight-Line Data*. New York : Dover, 1959
- [Agr02] AGRESTI, A.: *Categorical Data Analysis*. 2nd edition. Wiley; pp. 710, 2002
- [Ait87] AITCHISON, J.: *The Statistical Analysis of Compositional Data*. London and New York : Chapman and Hall; pp. 416, 1987
- [Aka73] AKAIKE, H.: Information theory as an extension of the maximum likelihood principle. In: PETROV, B.N. (Hrsg.) ; CSAKSI, F. (Hrsg.): *2nd International Symposium on Information Theory*. Akademiai Kiado, Budapest, Hungary, 1973, S. 267–281
- [Arm55] ARMITAGE, P.: Tests for linear trends in proportions and frequencies. In: *Biometrics* 11 (1955), S. 375–386
- [BA86] BLAND, J.M. ; ALTMAN, D.G.: Statistical methods for assessing agreement between two methods of clinical measurement. In: *Lancet* i: (1986), S. 307–310
- [BA02] BURNHAM, K.P. ; ANDERSON, D.R.: *Model Selection and Multimodel Inference: a Practical Information-Theoretic Approach*. 2nd edition. New York : Springer; pp. 488, 2002
- [Bar37] BARTLETT, M.S.: Properties of sufficiency and statistical tests. In: *Proceedings of the Royal Statistical Society Series A* 160 (1937), S. 268–282
- [Bar49] BARTLETT, M.S.: Fitting a straight line when both variables are subject to error. In: *Biometrics* 5 (1949), S. 207–212
- [Bar59] BARTHOLOMEW, D.J.: A test of homogeneity for ordered alternatives, I and II. In: *Biometrika* 46 (1959), S. 36–48 and 328–335
- [Bau62] BAUER, R.K.: Der „Median-Quartile-Test“: Ein Verfahren zur nichtparametrischen Prüfung zweier unabhängiger Stichproben auf unspezifische Verteilungsunterschiede. In: *Metrika* 5 (1962), S. 1–16
- [Bau72] BAUER, F.D.: Constructing confidence sets using rank statistics. In: *Journal of the American Statistical Association* 67 (1972), S. 687–690
- [BC99] BÜNING, H. ; CHAKRABORTI, S.: Power comparison of several two-sample tests for general alternatives. In: *Allgemeines Statistisches Archiv* 83 (1999), S. 190–210
- [BD80] BRESLOW, N.E. ; DAY, N.E.: *Statistical Methods in Cancer Research. Vol. I: The Analysis of Case-Control Studies*. Lyon : IARC Scientific Publ. No. 32; International Agency for Research on Cancer; pp. 338, 1980
- [BD87] BRESLOW, N.E. ; DAY, N.E.: *Statistical Methods in Cancer Research. Vol. II: The Design and Analysis of Cohort Studies*. Lyon : IARC Scientific Publ. No. 82; International Agency for Research on Cancer; pp. 406, 1987
- [Bel02] BELLE, G. van: *Statistical Rules of Thumb*. New York : John Wiley; pp. 248, 2002
- [Ben67] BENNETT, B.M.: Tests of hypotheses concerning matched samples. In: *J. Roy. Statist. Soc.* 29 (1967), S. 468–474
- [Ben72] BENNETT, B.M.: Tests for marginal symmetry in contingency tables. In: *Metrika* 19 (1972), S. 23–26
- [Ben78] BENNETT, B.M.: On a test for equality of dependent correlation coefficients. In: *Statistische Hefte* 19 (1978), S. 71–76

- [BEPW03] BACKHAUS, K. ; ERICHSON, B. ; PLINKE, W. ; WEIBER, R.: *Multivariate Analysemethoden: eine anwendungsorientierte Einführung*. 10. Auflage. Berlin : Springer Verlag; 818 S., 2003
- [BF74] BROWN, M.B. ; FORSYTHE, A.B.: Robust tests for the equality of variances. In: *J. Amer. Statist. Assoc.* 69 (1974), S. 364–367
- [BH62] BENNETT, B.M. ; HSU, P.: Sampling studies on a test against trend in binomial data. In: *Metrika* 5 (1962), S. 96–104
- [BH95] BENJAMINI, Y. ; HOCHBERG, Y.: Controlling the false discovery rate: a practical and powerful approach to multiple testing. In: *Journal of the Royal Statistical Society* 57 (1995), S. 289–300
- [BHW04] BRETZ, F. ; HOTHORN, T. ; WESTFALL, P.: *multcomp: Multiple Tests and Simultaneous Confidence Intervals.*, 2004. – R package version 0.4-8
- [Bla00] BLAND, M.: *An Introduction to Medical Statistics*. 3rd edition. Oxford, New York : Oxford University Press; pp. 422, 2000
- [Blo58] BLOM, G.: *Statistical Estimates and Transformed Beta Variables*. New York : John Wiley, 1958
- [Bor98] BORTKIEWICZ, L. von: *Das Gesetz der kleinen Zahlen*. Leipzig : Teubner, 1898
- [Bor99] BORTZ, J.: *Statistik für Sozialwissenschaftler*. 5. Auflage. Berlin : Springer Verlag, 836 S., 1999
- [Bor05] BORTZ, J.: *Statistik*. 6. Auflage. Berlin, Heidelberg, New York : Springer; 882 S., 2005
- [Bos98] BOSCH, K.: *Statistik-Taschenbuch*. 3. verbesserte Auflage. München, Wien : R. Oldenbourg; 840 S., 1998
- [Bow48] BOWKER, A.H.: A test for symmetry in contingency tables. In: *J. Amer. Statist. Assoc.* 43 (1948), S. 572–574
- [Bre75] BRESLOW, N.E.: Analysis of survival data under the proportional hazards model. In: *International Statistics Review* 43 (1975), S. 45–48
- [Bre82] BRESLOW, N.E.: Design and analysis of case-control studies. In: *Annual Review of Health* 3 (1982), S. 29–54
- [Bro54] BROSS, I.: A confidence interval for a percentage increase. In: *Biometrics* 10 (1954), S. 245–250
- [BS66] BRESNAHAN, J.I. ; SHAPIRO, M.M.: A general equation and technique for the exact partitioning of chi-square contingency tables. In: *Psychol. Bull.* 66 (1966), S. 252–262
- [BS03] BONETT, D.G. ; SEIER, Edith: Confidence intervals for mean absolute deviations. In: *The American Statistician* 57 (2003), S. 233–236
- [BW06] BACKHAUS, K. ; WEIBER, R.: *Multivariate Analysemethoden*. 11. Auflage. Berlin, Heidelberg, New York : Springer; 559 S., 2006
- [Cac65] CACOULOS, T.: A relation between t and F-distributions. In: *J. Amer. Statist. Assoc.* 60 (1965), S. 528–531
- [Caf05] CAFFO, Brian: *exactLoglinTest: Monte Carlo Exact Tests for Log-linear models*, 2005. – R package version 1.3.2
- [Cas65] CASTELLAN, N.J.: On the partitioning of contingency tables. In: *Psychol. Bull.* 64 (1965), S. 330–338
- [CB02] CASELLA, G. ; BERGER, R.L.: *Statistical Inference*. 2nd edition. Pacific Grove : Wadsworth and Brooks; pp. 660, 2002
- [CF96] CSÖRGÖ, S. ; FARAWAY, J.J.: The exact and asymptotic distribution of Cramér -von Mises statistics. In: *Journal of the Royal Statistical Society* 58 (1996), S. 221–234
- [CFK05] CALIEBE, Amke ; FREITAG, Sandra ; KRAWCZAK, M.: Stochastische Modelle für Interaktion und Effektmodifikation. In: *medgen* 17 (2005), S. 14–19
- [CG59] CROW, E.L. ; GARDNER, R.S.: Confidence intervals for the expectation of a POISSON variable. In: *Biometrika* 46 (1959), S. 441–453
- [CG80] CIBA-GEIGY, AG: *Wissenschaftliche Tabellen Geigy, Teilband Statistik*. Basel : Ciba-Geigy Ltd.; 241 S., 1980
- [CH92] CHAMBERS, J.M. ; HASTIE, T.J.: *Statistical Models in S*. Pacific Grove, CA : Wadsworth and Brooks, 1992

- [Cha71] CHAKRAVARTI, I.M.: Confidence set for the ratio of means of two normal distributions when the ratio of variances is known. In: *Biometrische Zeitschrift* 13 (1971), Nr. 12, S. 89–94
- [Cha02] CHASALOW, Scott: *combinat: combinatorics utilities.*, 2002. – R package version 0.0-5
- [Che99] CHERNICK, M.R.: *Bootstrap Methods*. New York : John Wiley; pp. 264, 1999
- [Chi70] CHISSOM, B.S.: Interpretation of the kurtosis statistic. In: *The American Statistician* 24 (1970), S. 19–22
- [CK84] CANAVOS, G.C. ; KOUTROUVELIS, I.A.: The robustness of two-sided tolerance limits for normal distributions. In: *Journal of Quality Technology* 16 (1984), S. 144–149
- [CO79] CAMPBELL, D.B. ; OPRIAN, C.A.: On the Kolmogorov-Smirnov test for the Poisson distribution with unknown mean. In: *Biometrical Journal* 21 (1979), S. 17–24
- [Coc41] COCHRAN, W.G.: The distribution of the largest of a set of estimated variances as a fraction of their total. In: *Ann. Eugen. (Lond.)* 11 (1941), S. 47–61
- [Coc50] COCHRAN, W.G.: The comparison of percentages in matched samples. In: *Biometrika* 37 (1950), S. 256–266
- [Coc54] COCHRAN, W.G.: Some methods for strengthening the common chi-square tests. In: *Biometrics* 10 (1954), S. 417–451
- [Coc63] COCHRAN, W.G.: *Sampling Techniques*. 2nd edition. New York : J. Wiley, 1963
- [Coh60] COHEN, J.: A coefficient of agreement for nominal scales. In: *Educational and Psychological Bulletin* 20 (1960), S. 37–46
- [Col03] COLLETT, D.: *Modelling Survival Data in Medical Research*. 2nd edition. London : Chapman and Hall; pp. 408, 2003
- [Con99] CONOVER, W.J.: *Practical Nonparametric Statistics*. 3rd edition. London : Wiley; pp. 584, 1999
- [Cox72] COX, D.R.: Regression models and life tables. In: *J.R. Statist. Soc. B*, 34 (1972), S. 187–220
- [Cox75] COX, D.R.: Partial Likelihood. In: *Biometrika* 62 (1975), S. 269–276
- [CP34] CLOPPER, C.J. ; PEARSON, E.S.: The use of confidence or fiducial limits illustrated in the case of the binomial. In: *Biometrika* 26 (1934), S. 404–413
- [CPS78] CASAGRANDE, J.T. ; PIKE, M.C. ; SMITH, P.G.: An improved approximate formula for calculating sample sizes for comparing two binomial distributions. In: *Biometrics* 34 (1978), S. 483–486
- [CR00] COX, D.R. ; REID, N.: *The Theory of the Design of Experiments*. London : Chapman + Hall; pp. 323, 2000
- [CS55] COX, D.R. ; STUART, A.: Quick sign test for trend in location and dispersion. In: *Biometrika* 42 (1955), S. 80–95
- [CSM87] CONNETT, J.E. ; SMITH, J.A. ; MCHUGH, R.H.: Sample size and power for pair-matched case-control studies. In: *Statist. Med.* 6 (1987), S. 53–59
- [CSW66] CARLSON, F.D. ; SOBEL, E. ; WATSON, G.S.: Linear relationships between variables affected by errors. In: *Biometrics* 22 (1966), S. 252–267
- [CSW03] CHOW, SC ; SHAO, J. ; WANG, H.: *Sample Size Calculations in Clinical Research*. New York : M. Dekker; pp. 358, 2003
- [Cur66] CURETON, E.E.: Quick fits for the lines $y = bx$ and $y = a + bx$ when errors of observation are present in both variables. In: *The American Statistician* 20 (1966), S. 49
- [Dal05] DALGAARD, P.: *Introductory Statistics with R*. New York : Springer; pp. 267, 2005
- [Dar70] DARLINGTON, R.B.: Is kurtosis really „peakedness“? In: *The American Statistician* 24 (1970), S. 19–22
- [Dav38] DAVID, F.N.: *Tables of the Ordinates and Probability Integral of the Distribution of the Correlation Coefficient in Small Samples*. London : The Biometrika Office, 1938
- [Dav63] DAVIES, O.L.: *The Design and Analysis of Industrial Experiments*. London : Oliver and Boyd, 1956 [3rd edition 1963]
- [dBd90] D’AGOSTINO, R.B. ; BELANGER, A. ; D’AGOSTINO, R.B.Jr.: A suggestion for using powerful and informative tests of normality. In: *The American Statistician* 44 (1990), S. 316–321

- [DHL⁺05] DIMITRIADOU, Evgenia ; HORNIK, Kurt ; LEISCH, Friedrich ; MEYER, David ; WEINGESSEL, Andreas: *e1071: Misc Functions of the Department of Statistics (e1071)*, TU Wien., 2005. – R package version 1.5-8
- [DHP54] DAVID, H.A. ; HARTLEY, H.O. ; PERASON, E.S.: The distribution of the ratio, in a single normal sample, of range to standard deviation. In: *Biometrika* 41 (1954), S. 482–493
- [Dix53] DIXON, W.J.: Processing data for outliers. In: *Biometrics* 9 (1953), S. 74–89
- [DM46] DIXON, W.J. ; MOOD, A.M.: The statistical sign test. In: *J. Amer. Statist. Assoc.* 41 (1946), S. 557–566
- [DM83] DIXON, W.J. ; MASSEY, F.J. J.: *Introduction to Statistical Analysis*. 4th edition. New York, Hamburg, London : McGraw Hill, 1983
- [DN03] DAVID, H.D. ; NAGARAJA, H.N.: *Order Statistics*. 3rd edition. New York : John Wiley; pp. 488, 2003
- [Dod03] DODGE, Yadolah: *The Oxford Dictionary of Statistical Terms*. 6th edition. New York : Oxford University Press; pp. 498, 2003
- [DR04] DESU, M.M. ; RAGHAVARAO, D.: *Nonparametric Statistical Methods for Complete and Censored Data*. Boca Raton, Florida : Chapman and Hall, 2004
- [Dun55] DUNNETT, C.W.: A multiple comparison procedure for comparing several treatments with a control. In: *J. Amer. Statist. Assoc.* 50 (1955), S. 1096–1121
- [Dun64] DUNNETT, C.W.: New tables for multiple comparisons with a control. In: *Biometrics* 20 (1964), S. 482–491
- [DW58] DUCKWORTH, W.E. ; WYATT, J.K.: Rapid statistical techniques for operations research workers. In: *Oper. Res. Quarterly* 9 (1958), S. 218–233
- [DW86] DALLAL, G.E. ; WILKINSON, L.: An analytic approximation to the distribution of Lilliefors's test statistic for normality. In: *The American Statistician* 40 (1986), S. 294–295
- [Edi95] EDINGTON, E.S.: *Randomization Tests*. 3rd edition. New York : M. Dekker; pp. 409, 1995
- [Efr77] EFRON, B.: The efficiency of Cox's likelihood function for censored data. In: *Journal of the American Statistical Association* 72 (1977), S. 557 – 565
- [Ehr86] EHRENBERG, Andrew S.: *Statistik oder der Umgang mit Daten*. Weinheim : VCH Verlagsgesellschaft; 344 S., 1986
- [ET93] EFRON, B. ; TIBSHIRANI, R.: *An Introduction to the Bootstrap*. New York, London : Chapman and Hall, 1993
- [Fag75] FAGAN, T.J.: Letter: Nomogram for Bayes's theorem. In: *New Engl. J. Med.* 293 (1975), S. 257
- [FC91] FOUNTAIN, R.L. ; CHOU, Y.-M.: Minimum sample sizes for two-sided tolerance intervals for finite populations. In: *Journal of Quality Technology* 23 (1991), S. 90–95
- [FCE69] FLEISS, J.L. ; COHEN, J. ; EVERITT, B.S.: Large sample standard errors of kappa and weighted kappa. In: *Psychological Bulletin* 72 (1969), S. 323–327
- [Fer85] FERSCHL, F.: *Deskriptive Statistik*. Würzburg und Wien : Physica-Vlg.; 308 S., 1985
- [FH91] FLEMING, T.R. ; HARRINGTON, D.P.: *Counting Processes and Survival Analysis*. New York : Wiley, 1991
- [Fie92] FIENBERG, S.: A brief history of statistics in three and one-half chapters: A review essay. In: *Statistical Science* 7 (1992), S. 208–225
- [Fin64] FINUCAN, H.M.: A note on kurtosis. In: *J. Roy. Statist. Soc. Ser. B* 26 (1964), S. 111 + 112
- [Fis60] FISHER, R.A.: *The Design of Experiments*. 7th edition. Edinburgh : Oliver and Boyd, 1960
- [Fis73] FISHER, R.A.: *Statistical Methods and Scientific Inference*. 3rd edition. Macmillan, Hafner; pp. 180, 1973
- [FKPT00] FAHRMEIR, L. ; KÜNSTLER, R. ; PIGEOT, I. ; TUTZ, G.: *Statistik*. Berlin : Springer, 2000
- [Fle81] FLEISS, J.L.: *Statistical Methods for Rates and Proportions*. 2nd edition. New York : Wiley; pp.432, 1981
- [Fle86] FLEISS, J.L.: *The Design and Analysis of Clinical Experiments*. New York : Wiley; pp. 432, 1986
- [Fri37] FRIEDMAN, M.: The use of ranks to avoid the assumption of normality implicit in the analysis of variance. In: *J. Amer. Statist. Assoc.* 32 (1937), S. 675–701

- [FS57] FAIRFIELD SMITH, H.: On comparing contingency tables. In: *The Philippine Statistician* 6 (1957), S. 71–81
- [FY82] FISHER, R.A. ; YATES, F.: *Statistical Tables for Biological, Agricultural and Medical Research*. 6th edition. Harlow : Longman; pp.146, 1982
- [Gam05] GAMER, M.: *irr: Various Coefficients of Interrater Reliability and Agreement.*, 2005. – R package version 0.5
- [GBH04] GENZ, A. ; BRETZ, F. ; HOTHORN, T.: *mvtnorm: Multivariate Normal and T Distribution.*, 2004. – R package version 0.7-1
- [GG04] GIESBRECHT, F.G. ; GUMPERTZ, M.L.: *Planning, Construction, and Statistical Analysis of Comparative Experiments*. New York : Wiley; pp. 693, 2004
- [GHW74] GRAF, U. ; HENNING, H.J. ; WILRICH, P.T.: *Statistische Methoden bei textilen Untersuchungen*. Berlin : Springer, 1974
- [GJ57] GIBSON, Wendy M. ; JOWETT, G.H.: "Three-group" regression analysis. Part I. Simple regression analysis. Part II. Multiple regression analysis. In: *Applied Statistics* 6 (1957), S. 114–122 and 189–197
- [GN96] GREENWOOD, Priscilla E. ; NIKULIN, M.S.: *A Guide to Chi-Squared Testing*. New York : Wiley; pp. 280, 1996
- [Goo71] GOODMAN, L.A.: The analysis of multidimensional contingency tables: stepwise procedures and direct estimation methods for building models for multiple classifications. In: *Technometrics* 13 (1971), S. 33–61
- [Goo05] GOOD, P.: *Permutation, Parametric, and Bootstrap Tests of Hypotheses*. 3rd edition. New York : Springer; pp. 315, 2005
- [Gre26] GREENWOOD, M.: The natural duration of cancer. In: *Reports on Public Health and Medical Subjects, Vol. 33*. London : Her Majesty's Stationery Office, 1926, S. 1 – 26
- [Gri67] GRIZZLE, J.E.: Continuity correction in the χ^2 -test for 2x2 tables. In: *The American Statistician* 21 (1967), S. 28–32
- [Gro05] GROSS, J.: *nortest: Tests for Normality*, 2005. – R package version 1.0
- [Gru69] GRUBBS, F.: Procedures for detecting outlying observations in samples. In: *Technometrics* 11 (1969), S. 1–21
- [GS89] GATSONIS, C. ; SAMPSON, A.R.: Multiple correlation: exact power and sample size calculations. In: *Psychological Bulletin* 106 (1989), S. 516–524
- [GT94] GRAMBSCH, P. ; THERNEAU, T.M.: Proportional hazards tests and diagnostics based on weighted residuals. In: *Biometrika* 81 (1994), S. 515–526
- [GY20] GREENWOOD, M. ; YULE, G.U.: An inquiry into the nature of frequency distributions representative of multiple happenings, with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. In: *J. Roy. Statist. Soc.* 83 (1920), S. 255
- [Ham85] HAMPEL, F.R.: The breakdown points of the mean combined with some rejection rules. In: *Technometrics* 27 (1985), S. 95–107
- [Har42] HART, B.I.: Significance levels for the ratio of the mean square successive difference to the variance. In: *Ann. Math.* 13 (1942), S. 445–447
- [Har50] HARTLEY, H.O.: The maximum F-ratio as a short cut test for heterogeneity of variance. In: *Biometrika* 37 (1950), S. 308–312
- [Har60] HARTER, H.L.: Tables of range and Studentized range. In: *Ann. Math. Statist.* 31 (1960), S. 1122–1147
- [Hav90] HAVILAND, M.G.: Yates' correction for continuity and the analysis of 2x2 contingency tables. With discussion. In: *Statistics in Medicine* 9 (1990), S. 363–283
- [Hay86] HAYTER, A.J.: The maximum familywise error rate of Fisher's least significant difference. In: *Journal of the American Statistical Association* 81 (1986), S. 1001–1004
- [HC70] HENDY, M.F ; CHARLES, J.A ..: The production techniques, silver content, and circulation history of the twelfth-century byzantine trachy. In: *Archaeometry* 12 (1970), S. 13–21
- [HD77] HAUCK, W.W. ; DONNER, A.: Wald's Test as applied to hypothesis in logit analysis. In: *Journal of the American Statistical Association* 72 (1977), S. 851–853

- [HH71] HAHN, G.J. ; HENDRICKSON, R.W.: A table of percentage points of the distribution of the largest absolute value of k Student t variates and its applications. In: *Biometrika* 58 (1971), S. 323–332
- [HH05] HOTHORN, Torsten ; HORNIK., Kurt: *exactRankTests: Exact Distributions for Rank and Permutation Tests.*, 2005. – R package version 0.8-10
- [HHR92] HEINECKE, A. ; HULTSCH, E. ; REPGES, R.: *Medizinische Biometrie: Biomathematik und Statistik.* Berlin : Springer; 287 S., 1992
- [HL65] HAYNAM, G.E. ; LEONE, F.C.: Analysis of categorical data. In: *Biometrika* 52 (1965), S. 654–660
- [HL82] HEWETT, J.E. ; LABABIDI, Z.: Comparison of three regression lines over a finite interval. In: *Biometrics* 38 (1982), S. 837–841
- [HL89] HOSMER, D.W. ; LEMESHOW, S.: *Applied Logistic Regression.* New York : Wiley; pp. 307, 1989
- [HL99] HOSMER, D.W. ; LEMESHOW, S.: *Applied Survival Analysis: regression modeling of time to event.* New York : Wiley; pp. 386, 1999
- [HLP85] HOMMEL, G. ; LEHMACHER, W. ; PERLI, H.-G.: Residuenanalyse des Unabhängigkeitsmodells zweier kategorischer Variablen. In: JESDINSKY, H.J. (Hrsg.) ; TRAMPISCH, H.J. (Hrsg.): *Prognose und Entscheidungsfindung in der Medizin.* Bd. 62. Berlin, Heidelberg, New York : Springer; 524 S., 1985, S. 494–503
- [HM73] HETTMANSPERGER, T.P. ; MCKEAN, J.W.: On testing for significant change in $c \times c$ tables. In: *Commun. Statist.* 2 (1973), S. 551–560
- [HO70] HARTER, H.L. ; OWEN, D.B.: *Selected Tables in Mathematical Statistics. Vol. I.* Chicago : Markham, pp. 405, 1970
- [Hoc88] HOCHBERG, Y.: A sharper Bonferroni procedure for multiple tests of significance. In: *Biometrika* 75 (1988), S. 800–802
- [Hoc91] HOCHSTÄDTER, D.: *Einführung in die statistische Methodenlehre.* 7. Auflage. Frankfurt / M., Thun : H. Deutscher; 744 S., 1991
- [Hol79] HOLM, S.: A simple sequentially rejective multiple test procedure. In: *Scandinavian Journal of Statistics* 6 (1979), S. 65–70
- [Hom88] HOMMEL, G.: A stagewise rejective multiple test procedure. In: *Biometrika* 75 (1988), S. 383–386
- [Hot53] HOTELLING, H.: New light on the correlation coefficient and its transforms. In: *J. Roy. Statist. Soc. B* 15 (1953), S. 193–232
- [HV95] HORN, M. ; VOLLANDT, R.: *Multiple Tests und Auswahlverfahren.* Stuttgart : Gustav Fischer Verlag, 1995
- [IG96] IHAKA, R. ; GENTLEMAN, R.: R: A language for data analysis and graphics. In: *Journal of Computational and Graphical Statistics.* 5 (1996), Nr. 3, S. 299 – 314
- [IKK69] IRELAND, C.T. ; KU, H.H. ; KULLBACK, S.: Symmetry and marginal homogeneity of an $r \cdot r$ contingency table. In: *J. Amer. Statist. Assoc.* 64 (1969), S. 1323–1341
- [IQA75] IMAN, R.L. ; QUADE, D. ; ALEXANDER, D.A.: *Selected Tables in Mathematical Statistics..* Bd. III: *Exact Probability Levels for the Kruskal-Wallis Test.* Providence, Rhode Island : Institute of Mathematical Statistics and American Mathematical Society; pp. 329–384, 1975
- [Jen86] JENNINGS, D.E.: Judging inference adequacy in logistic regression. In: *Journal of the American Statistical Association* 81 (1986), S. 471–476
- [Jon54] JONCKHEERE, A.R.: A distribution-free k -sample test against ordered alternatives. In: *Biometrika* 41 (1954), S. 133–145
- [JW40] JOHNSON, N.L. ; WELCH, B.L.: Applications of the noncentral t -distribution. In: *Biometrika* 31 (1940), S. 362–389
- [Kal68] KALTON, G.: Standardization: a technique to control for extraneous variables. In: *Applied Statistics* 17 (1968), S. 118–136
- [Kas60] KASTENBAUM, M.A.: An note on the additive partitioning of chi-square in contingency tables. In: *Biometrics* 16 (1960), S. 416–422

- [KCFT97] KUCZMARSKI, R.J. ; CAROL, M.D. ; FLEGAL, K.M. ; TROJANO, R.P.: Varying body mass index cutoff points to describe overweight prevalence among U.S. adults: NHANES III (1988 to 1994). In: *Obesity Research* 5 (1997), S. 542–548
- [Ker66] KERRICH, J.E.: Fitting the line $y = ax$ when errors of observation are present in both variables. In: *The American Statistician* 20 (1966), S. 24
- [KHB70] KASTENBAUM, M.A. ; HOEL, D.G. ; BOWMAN, K.O.: Sample size requirements: one-way analysis of variance. In: *Biometrika* 57 (1970), S. 421–430
- [Kim54] KIMBALL, A.W.: Short-cut formulae for the exact partition of χ^2 in contingency tables. In: *Biometrics* 10 (1954), S. 452–458
- [Kim69] KIM, P.J.: On the exact and approximate sampling distribution of the two sample Kolmogorov-Smirnov criterion D_{mn} , $m \geq n$. In: *J. Amer. Statist. Assoc.* 64 (1969), S. 1625–1637
- [Kit64] KITAGAWA, Evelyn M.: Standardized comparisons in population research. In: *Demography* 1 (1964), S. 296–315
- [KLB87] KOHNEN, R. ; LOTZ, R. ; BUSCH, H.: A table of one- and two-tailed fourfold chisquare limits for unconventional alphas. In: *EDV in Medizin und Biologie* 18 (1987), S. 50–54
- [Kle01] KLEPPMANN, W.: *Taschenbuch Versuchsplanung. Produkte und Prozesse optimieren.* 2. Auflage. München und Wien : C. Hanser; 281 S., 2001
- [KM03] KLEIN, John P. ; MOESCHBERGER, Melvin L.: *Survival Analysis: Techniques for Censored and Truncated Data.* 2nd edition. New York : Springer; pp. 536, 2003
- [Koc00] KOCKELKORN, U.: *Lineare statistische Methoden.* München, Wien : R. Oldenbourg; 728 S., 2000
- [Koe05] KOENKER, Roger: quantreg: Quantile Regression. (2005). – R package version 3.76, initial R port from Splus by Brian Ripley
- [Kol33] KOLMOGOROFF, A.N.: *Grundbegriffe der Wahrscheinlichkeitsrechnung.* Berlin : Springer, 1933
- [Kol63] KOLLER, S.: Typisierung korrelativer Zusammenhänge. In: *Metrika* 6 (1963), S. 65–75
- [KP02] KALBFLEISCH, J.D. ; PRENTICE, R.I.: *The Statistical Analysis of Failure Time Data.* 2nd edition. New York : John Wiley; pp. 462, 2002
- [Kru52] KRUSKAL, W.H.: A nonparametric test for the several sampling problem. In: *Ann. Math. Statist.* 23 (1952), S. 525–540
- [Kru99] KRUMMENAUER, F.: Erweiterungen von Cohen's kappa-Maß für Multi-Rater-Studien: Eine Übersicht. In: *Informatik, Biometrie und Epidemiologie in Medizin und Biologie* 30 (1999), S. 3–20
- [KS89] KAHN, H.A. ; SEMPOS, Ch.T.: *Statistical Methods in Epidemiology.* Oxford and New York : Oxford University Press; pp. 292, 1989
- [KTE96] KELSEY, Jennifer L. ; THOMPSON, W.D. ; EVANS, A.S.: *Methods in Observational Epidemiology.* 2nd edition. New York and Oxford : Oxford Univ. Press; pp. 448, 1996
- [Kue00] KUEHL, R.O.: *Design of Experiments. Statistical Principles of Research Design and Analysis.* 2nd edition. Pacific Grove/USA : Duxbury Press; pp. 666, 2000
- [KW52] KRUSKAL, W.H. ; WALLIS, W.A.: Use of ranks in one-criterion variance analysis. In: *J. Amer. Statist. Ass.* 47 (1952), S. 614–617
- [Kym68] KYMN, K.O.: The distribution of the sample correlation coefficient under the null hypothesis. In: *Econometrica* 36 (1968), S. 187–189
- [Laa70] LAAN, P. van d.: Simple distribution-free confidence intervals for a difference in location. In: *Philips Res. Repts. Suppl.* 5 (1970), S. 158
- [Lep71] LEPAGE, Y.: A combination of Wilcoxon's and Ansari-Bradley's statistics. In: *Biometrika* 58 (1971), S. 213–217
- [Lev60] LEVENE, H.: Robust Tests for equality of variances. In: *I. Olkin et al. (Eds): Contributions to Probability and Statistics. Essays in Honor of Harold Hotelling.* Stanford, 1960, S. 278–292
- [LF65] LEWONTIN, R.C. ; FELSENSTEIN, J.: The robustness of homogeneity tests in $2 \times n$ tables. In: *Biometrics* 21 (1965), S. 19–33

- [Lig05] LIGGES, Uwe: *Programmieren in R*. Berlin : Springer Verlag, 237 S., 2005
- [Lil67] LILLIEFORS, H.W.: On the Kolmogorov-Smirnov test for normality with mean and variance unknown. In: *J. Amer. Statist. Assoc.* 62 (1967), S. 399–402, Corrigenda **64** (1969) 1702
- [LK77a] LANDIS, J.R. ; KOCH, G.G.: The measurement of observer agreement for categorical data. In: *Biometrics* (1977), Nr. 159–174
- [LK77b] LANDIS, J.R. ; KOCH, G.G.: A one-way components of the variance model for categorical data. In: *Biometrics* 33 (1977), S. 671–679
- [LR05] LEHMANN, E.L. ; ROMANO, J.P.: *Testing Statistical Hypotheses*. 3rd edition. New York : Springer; pp. 786, 2005
- [LS67] LIENERT, G.A. ; SCHULZ, H.: Zum Nachweis von Behandlungswirkungen bei heterogenen Patientenstichproben. In: *Ärztliche Forschung* 21 (1967), S. 448–455
- [LU90] LAWAL, H.B. ; UPTON, G.J.G.: Comparisons of some chi-squared tests for the test of independence in sparse two-way contingency tables. In: *Biometrical Journal* 32 (1990), S. 59–72
- [LW03] LEE, Elisa T. ; WANG, John W.: *Statistical Methods for Survival Data Analysis*. 3rd edition. Hoboken, New Jersey : Wiley; pp. 513, 2003
- [Mad59] MADANSKY, A.: The fitting of straight lines when both variables are subject to error. In: *J. Amer. Statist. Assoc.* 54 (1959), S. 173–205
- [MALDCHT91] MARTIN ANDREŠ, A. ; LUNA DEL CASTILLO, J.D. ; HERRANZ TEJEDOR, I.: New critical tables for Fisher's exact test. In: *Journal of Applied Statistics* 18 (1991), S. 233–254
- [Man66] MANTEL, N.: Evaluation of survival data and two new rank order statistics arising in its consideration. In: *Cancer Chemotherapy Reports* 50 (1966), S. 163–170
- [Man97] MANLY, B.F.J.: *Randomization, Bootstrap and Monte Carlo Methods in Biology*. London, New York : Chapman and Hall; pp. 399, 1997
- [Mar53] MARTINI, P.: *Methodenlehre der therapeutisch-klinischen Forschung*. Berlin-Göttingen-Heidelberg : Springer Verlag, 1953
- [MB70] MEYER-BAHLBURG, H.F.L.: A nonparametric test for relative spread in k unpaired samples. In: *Metrika* 15 (1970), S. 23–29
- [MB86] MASON, A.L. ; BELL, C.B.: New Lilliefors and Srinivasan tables with applications. In: *Communications in Statistics - Simulation and Computation* 15 (1986), S. 451–477
- [MB04] MAINDONALD, John ; BRAUN, W. J.: *DAAG: Data Analysis And Graphics*, 2004. – R package version 0.37. <http://www.stats.uwo.ca/DAAG>
- [MC81] MIETTINEN, O.S. ; COOK, E.F.: Confounding: essence and detection. In: *American Journal of Epidemiology* 114 (1981), S. 593–603
- [McN47] MCNEMAR, Q.: Note on sampling error of the differences between correlated proportions or percentages. In: *Psychometrika* 12 (1947), S. 153–154
- [McN69] MCNEMAR, Q.: *Psychological Statistics*. 4th edition. New York : Wiley; pp. 529, 1969
- [Mei87] MEIS, T.: Brauchen wir eine Hochgenauigkeitsarithmetik? In: *Praxis der Informationsverarbeitung und Kommunikation* 10 (1987), S. 19–23
- [Mic71] MICHAELIS, J.: Schwellenwerte des Friedman-Tests. In: *Biometr. Zeitschr.* 13 (1971), S. 118–129
- [Mil56] MILLER, L.H.: Table of percentage points of Kolmogorov statistics. In: *J. Amer. Statist. Assoc.* 51 (1956), S. 111–121
- [MM98] MURPHY, K.R. ; MYORS, B.: *Statistical Power Analysis*. Amsterdam : L. Erlbaum, 1998
- [Mol70] MOLENAAR, W.: *Approximations to the Poisson, Binomial, and Hypergeometric Distribution Functions*. Amsterdam : Math. Centr., pp. 160, 1970
- [Mol04] MOLENAAR, I.W.: About handy, handmade and handsome models. In: *Statistica Neerlandica* 58 (2004), S. 1–20
- [Mon05] MONTGOMERY, D.C.: *Design and Analysis of Experiments*. 6th edition. New York : Wiley; pp. 643, 2005
- [Moo88] MOORS, J.J.A.: A quantile alternative for kurtosis. In: *Statistician* 37 (1988), S. 25–32
- [Mos48] MOSTELLER, F.: A k-sample slippage test for an extreme population. In: *Ann. Math. Stat.* 19 (1948), S. 58–65

- [MOS71] MENDENHALL, W. ; OTT, L. ; SCHEAFFER, R.L.: *Elementary Survey Sampling*. Belmont, California : Wadsworth, 1971
- [MPG85] MEHTA, C.R. ; PATEL, N.R. ; GRAY, R.: Computing an exact confidence interval for the common odds ratio in several 2x2 contingency tables. In: *Journal of the American Statistical Association* 80 (1985), S. 969–973
- [MRR92] MENG, X.-L. ; ROSENTHAL, R. ; RUBIN, D.B.: Comparing correlation coefficients. In: *Psychological Bulletin* 111 (1992), S. 172–175
- [MSL87] MEE, R.W. ; SHAH, A.K. ; LEFANTE, J.J.: Comparing k independent sample means with a known standard. In: *Journal of Quality Technology* 19 (1987), S. 75–81
- [MW47] MANN, H.B. ; WHITNEY, D.R.: On a test of whether one of two random variables is stochastically larger than the other. In: *Ann. Math. Statist.* 18 (1947), S. 50–60
- [MZKH05] MEYER, D. ; ZEILEIS, A. ; KARATZOGLOU, A. ; HORNIK, K.: *vcd: Visualizing Categorical Data.*, 2005. – R package version 0.1-3.5
- [Nat63] NATRELLA, Mary G.: *Experimental Statistics. NSB Handbook 91.* Washington : U.S.Gvt.Print.Office, 1963
- [ND75] NEILL, J.J. ; DUNN, O.J.: Equality of dependent correlation coefficients. In: *Biometrics* 31 (1975), S. 531–543
- [Nel72] NELSON, W.B.: Theory and applications of hazard plotting for censored failure data. In: *Technometrics* 14 (1972), S. 945 – 965
- [Nem63] NEMENYI, P.: *Distribution-Free Multiple Comparisons.* New York : State University of New York. Downstate Medical Center, 1963
- [Ney50] NEYMAN, J.: *First Course in Probability and Statistics.* New York : Holt, 1950
- [Noe87] NOETHER, G.E.: Sample size determination for some common nonparametric tests. In: *Journal of the American Statistical Association* 82 (1987), S. 645–647
- [NP33] NEYMAN, J. ; PEARSON, E.S.: On the problem of the most efficient type of statistical hypothesis. In: *Philosophical Transactions of the Royal Society A* 231 (1933), S. 289–337
- [OCO87] ODEH, R.E. ; CHOU, Y.-M. ; OWEN, D.B.: The precision for coverages and sample size requirements for normal tolerance intervals. In: *Communications in Statistics - Simulation and Computation* 16 (1987), S. 969–985
- [OCO89] ODEH, R.E. ; CHOU, Y.-M. ; OWEN, D.B.: Sample-size determination for two-sided β -expectation tolerance intervals for a normal distribution. In: *Technometrics* 31 (1989), S. 461–468
- [Ode77] ODEH, R.E.: Extended tables of the distribution of Friedman's S-statistic in the two-way layout. In: *Communications in Statistics - Simulation and Computation* B6 (1977), S. 29–48
- [Ode78] ODEH, R.E.: Tables of two-sided tolerance factors for a normal distribution. In: *Communications in Statistics - Simulation and Computation* 7 (1978), S. 183–201
- [Ode82a] ODEH, R.E.: Critical values of the sample product-moment correlation coefficient in the bivariate distribution. In: *Communications in Statistics - Simulation and Computation* 11 (1982), S. 1–26
- [Ode82b] ODEH, R.E.: Tables of percentage points of the distribution of the maximum absolute value of equally correlated normal random variables. In: *Communications in Statistics - Simulation and Computation* 11 (1982), S. 65–87
- [Oke82] OKEN, M.M. et al.: Toxicity and response criteria of the eastern cooperative oncology group. In: *Am. J. Clin. Oncol.* 5 (1982), S. 649 – 655
- [Pag63] PAGE, E.B.: Ordered hypotheses for multiple treatments: A significance test for linear ranks. In: *J. Amer. Statist. Assoc.* 58 (1963), S. 216–230
- [Paw01] PAWITAN, Y.: *In All Likelihood.* Oxford, New York : Oxford Univ. Press; pp. 528, 2001
- [PB61] PILLAI, K.C.S. ; BUENAVENTURA, A.R.: Upper percentage points of a substitute F-ratio using ranges. In: *Biometrika* 48 (1961), S. 195–196
- [PH72] PEARSON, E.S. ; HARTLEY, H.O. Eds.: *Biometrika Tables for Statisticians (Vol. I and II).* Cambridge : Cambridge Univ. Press; pp. 270 and pp. 385, 1970 , 1972
- [Pit49] PITMAN, E.J.G.: *Lecture Notes on Nonparametric Statistics.* New York : Columbia University, 1949

- [Pit93] PITMAN, J.: *Probability*. New York, Heidelberg, Berlin : Springer; pp. 559, 1993
- [Pot65] POTTHOFF, R.F.: Some Scheffe-type tests for some Behrens-Fisher type regression problems. In: *J. Amer. Statist. Assoc.* 60 (1965), S. 1163–1190
- [PS64] PEARSON, E.S. ; STEPHENS, M.A.: The ratio of range to standard deviation in the same normal sample. In: *Biometrika* 51 (1964), S. 484–487
- [R D05] R DEVELOPMENT CORE TEAM: *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, 2005. – 3-900051-07-0
- [Ram81] RAMSAY, P.H.: Power of univariate pairwise multiple comparison procedures. In: *Psychological Bulletin* 90 (1981), S. 352–366
- [Ras95] RASCH, D.: *Mathematische Statistik*. Heidelberg : J.A. Barth; 851 S., 1995
- [RG98] ROTHMAN, K.J. ; GREENLAND, S.: *Modern Epidemiology*. 2nd edition. Philadelphia : Lippincott Williams and Wilkins, 1998
- [RHB⁺96] RASCH, D. ; HERRENDÖRFER, G. ; BOCK, J. ; VICTOR, N. ; GUIARD, V. (Hrsg.): *Versuchsbibliothek; Versuchsplanung und -auswertung (Band I und II)*. München, Wien : R. Oldenbourg, 1996
- [RHW89] REIBNEGGER, G. ; HAUSEN, A. ; WACHTER, H.: Null results when diagnostic tests are imperfect. In: *The Lancet* 2 (1989)
- [Rin03] RINNE, H.: *Taschenbuch der Statistik*. 3. Auflage. Frankfurt am Main : Harri Deutsch; 849 S., 2003
- [Rog80] ROGOSA, D.: Comparing nonparallel regression lines. In: *Psychological Bulletin* 88 (1980), S. 307–321
- [Ros53] ROSENBAUM, S.: Tables for a nonparametric test of dispersion. In: *Ann. Math. Stat.* 24 (1953), S. 663–668
- [Ros54] ROSENBAUM, S.: Tables for a nonparametric test of location. In: *Ann. Math. Stat.* 25 (1954), S. 146–150
- [Roy82a] ROYSTON, P.: Algorithm AS 181: The W test for normality. In: *Applied Statistics* 31 (1982), S. 176–180
- [Roy82b] ROYSTON, P.: An extension of Shapiro and Wilk's W test for normality to large samples. In: *Applied Statistics* 31 (1982), S. 115–124
- [Roy84] ROYEN, Th.: Multiple comparisons of polynomial distributions. In: *Biometrical Journal* 26 (1984), S. 319–332
- [Rum86] RUMP, S.M.: Sichere Ergebnisse auf Rechenanlagen. In: *Informatik-Spektrum* 9 (1986), S. 174–183
- [Rya60] RYAN, T.A.: Significance tests for multiple comparisons of proportions, variances and other statistics. In: *Psychological Bulletin* 57 (1960), S. 318–328
- [SA87] SHIFFLER, R.E. ; ADAMS, A.J.: A correction for biasing effects of pilot sample size on sample size determination. In: *Journal of Marketing Research* 24 (1987), S. 319–321
- [SA00] SAHAI, H. ; AGEEL, M.I.: *The Analysis of Variances; Fixed, Random and Mixed Models*. Boston, Basel, Berlin : Birkhäuser; pp. 742, 2000
- [Sac84] SACHS, L.: *Applied Statistics. A Handbook of Techniques*. 2nd edition. New York, Heidelberg, Berlin : Springer; pp. 707, 1984
- [Sac86a] SACHS, L.: Alternatives to the chi-square test of homogeneity in 2×2 tables and to Fisher's exact test. In: *Biometrical Journal* 28 (1986), S. 975–979
- [Sac86b] SACHS, L.: *A Guide to Statistical Methods and to the Pertinent Literature. Literatur zur Angewandten Statistik*. Berlin, Heidelberg, New York : Springer; 212 S., 1986
- [Sac90] SACHS, L.: *Statistische Methoden 2: Planung und Auswertung*. Berlin, Heidelberg, New York : Springer; 273 S., 1990
- [Sac93] SACHS, L.: *Statistische Methoden: Planung und Auswertung*. 7. Auflage. Berlin, Heidelberg, New York : Springer; 312 S., 1993
- [Sac06] SACHS, L.: *Einführung in die Stochastik und das stochastische Denken*. Frankfurt am Main : H. Deutsch; 182 S., 2006
- [Sam70] SAMIUDDIN, M.: On a test for an assigned value of correlation in a bivariate normal distribution. In: *Biometrika* 57 (1970), S. 461–464

- [SC82] SNEDECOR, G.W. ; COCHRAN, W.G.: *Statistical Methods*. 7th edition. Ames, Iowa, USA : The Iowa State University Press; pp. 507, 1982
- [SC97] SARKAR, S. ; CHANG, C.K.: Simes' method for multiple hypothesis testing with positively dependent test statistics. In: *Journal of the American Statistical Association* 92 (1997), S. 1601–1608
- [Sch53] SCHEFFÉ, H.: A method for judging all contrasts in the analysis of variance. In: *Biometrika* 40 (1953), S. 87–104
- [Sch70] SCHEFFÉ, H.: Practical solutions of the Behrens-Fisher problem. In: *J. Amer. Statist. Assoc.* 65 (1970), S. 1501–1508
- [SE43] SWED, Frida S. ; EISENHART, C.: Tables for testing randomness of grouping in a sequence of alternatives. In: *Ann. Math. Statist.* 14 (1943), S. 83–86
- [Sev00] SEVERINI, T.A.: *Likelihood Methods in Statistics*. Oxford, New York : Oxford Univ. Press; pp. 392, 2000
- [SFH89] SIDDHARTA, R.D. ; FOWLKES, E.B. ; HANDLEY, B.: Risk analysis of the space shuttle: Pre-challenger prediction of failure. In: *Journal of the American Statistical Association* 84 (1989), S. 945–957
- [She04] SHESKIN, D. J.: *Handbok of Parametric and Nonparametric Statistical Procedures*. 3rd edition. Boca Raton, Florida : Chapman and Hall; pp. 1193, 2004
- [She05] SHEYNIN, O.: *Theory of Probability. A Historical Essay*. Berlin (Tel.: 030/4442460; Fax.: 030/44739165) : NG Verlag; pp. 278, 2005
- [Sim49] SIMPSON, E.H.: Measurement of diversity. In: *Nature* 163 (1949), S. 688
- [SK90] STORER, B.E. ; KIM, Ch.: Exact properties of some exact statistics for comparing two binomial proportions. In: *Journal of the American Statistical Association* 85 (1990), S. 146–155
- [SM88] SMITH, Judith A. ; MCHUGH, R.B.: Sample size requirements in studies of the etiologic fraction. In: *Biometrical Journal* 30 (1988), S. 187–201
- [Smi39] SMIRNOFF, N.W.: On the estimation of the discrepancy between empirical curves of distribution for two independent samples. In: *Bull. Universite Moskov. Ser. Internat. Sect. A2* (1939), S. 3–19
- [SO94] STUART, A. ; ORD, J.K.: *Kendall's Advanced Theory of Statistics. Vol. I: Distribution Theory*. 6th edition. Sevenoaks, Kent : E. Arnold, 1994
- [Spe04] SPEARMAN, C.: The proof and measurement of association between two things. In: *Amer. J. Psychol.* 15 (1904), S. 72–101
- [ST60] SIEGEL, S. ; TUKEY, J.W.: A nonparametric sum of ranks procedure for relative spread in unpaired samples. In: *J. Amer. Statist. Assoc.* 55 (1960), S. 429–445 [Errata 56 (1961), 1005
- [Ste74] STEPHENS, M.A.: EDF statistics for goodness of fit and some comparisons. In: *Journal of the American Statistical Association* 69 (1974), S. 730–737
- [Ste80] STEIGER, J.H.: Tests for comparing elements of a correlation matrix. In: *Psychological Bulletin* 87 (1980), S. 245–251
- [Ste86] STEPHENS, M.A.: Tests based on EDF statistics. In: D'AGOSTINO, R.B. (Hrsg.) ; STEPHENS, M.A. (Hrsg.): *Goodness-of-fit Techniques*. New York : Marcel Dekker, 1986
- [Sto78] STOLINE, M.R.: Tables of the Studentized augmented range and applications to problems of multiple comparisons. In: *Journal of the American Statistical Association* 73 (1978), S. 656–660
- [Stu08] STUDENT: The probable error of a mean. In: *Biometrika* 6 (1908), S. 1–25
- [SW65] SHAPIRO, S.S. ; WILK, M.B.: An analysis of variance test for normality (complete samples). In: *Biometrika* 52 (1965), S. 591–611
- [TB70] TATE, M.W. ; BROWN, Sara M.: Note on the Cochran Q-test. In: *J. Amer. Statist. Assoc.* (1970), Nr. 155–160
- [TB05] THERNEAU, T.M. ; B., Atkinson: *rpart: Recursive Partitioning (R port by Brian Ripley)*., 2005. – R package version 3.1-22
- [TD01] TODMAN, J.B. ; DUGARD, P.: *Single-Case and Small-n Experimental Designs; A Practical Guide to Randomization Tests*. Mahawah, New Jersey : Lawrence Erlbaum Associates; pp. 245, 2001

- [TF05] TIBSHIRANI, R. ; F., Leisch: *bootstrap: Functions for the Book „An Introduction to the Bootstrap” (S original Rob Tibshirani, R port by Friedrich Leisch)*, 2005. – R package version 1.0-18
- [Tie87] TIEDE, M.: *Statistik. Regressions- und Korrelationsanalyse*. München und Wien : R. Oldenbourg; 171 S., 1987
- [Tim02] TIMM, Neil H.: *Applied Multivariate Analysis*. New York : Springer Verlag; pp. 693, 2002
- [Tra05] TRAPLETTI, A.: *tseries: Time series analysis and computational finance.*, 2005. – R package version 0.9-26
- [TT05] THERNEAU, T. ; T., Lumley: *survival: Survival analysis, including penalised likelihood.*, 2005. – R package version 2.17
- [Tuk51] TUKEY, J.W.: Components in regression. In: *Biometrics* 7 (1951), S. 33–70
- [Tuk59] TUKEY, J.W.: A quick, compact, two-sample test to Duckworth's specifications. In: *Technometrics* 1 (1959), Nr. 31-48
- [Tuk60] TUKEY, J.W.: Conclusions vs. decisions. In: *Technometrics* 2 (1960), S. 423–433
- [VB02] VENABLES, W.N. ; B.D., Ripley: *Modern Applied Statistics with S*. 4th edition. Springer, 2002
- [Ver05] VERZANI, J.: *Using R for Introductory Statistics*. Boca Raton, Florida : Chapman and Hall, 2005
- [VGSM05] VITTINGHOFF, E. ; GLIDDEN, D.V. ; SHIBOSKI, S.C. ; MCCULLOCH, C.E.: *Linear, Logistic, Survival, and Repeated Measures Models*. New York : Springer; pp. 344, 2005
- [Wal51] WALTER, E.: Über einige nichtparametrische Testverfahren (I, II). In: *Mathemat. Statistik* 3 (1951), S. 31–44, 73–92
- [Wal58] WALTER, E.: Einige einfache nichtparametrische überall wirksame Tests zur Prüfung der Zweistichprobenhypothese mit paarigen Beobachtungen. In: *Metrika* 1 (1958), S. 81–88
- [Wal64] WALTER, E.: Rezension des Buches „Verteilungsfreie Methoden in der Biostatistik” von G. Lienert. In: *Biometrische Zeitschrift* 6 (1964), S. 61–62
- [War05] WARNES, Gregory R.: *gmodels: Greg's Miscellaneous Functions.*, 2005. – R package version 2.0.6
- [Wei60] WEIR, J.B. de V.: Significance of the difference between two means when the population variances may be unequal. In: *Nature* 187 (1960), S. 438
- [Wei05] WEISS, C.: *Basiswissen Medizinische Statistik*. 3. Auflage. Berlin, Heidelberg, New York : Springer; 324 S., 2005
- [Wel89] WELLEK, S.: Computing exact p-values in Page's nonparametric test against trend. In: *Biometrie und Informatik in Medizin und Biologie* 20 (1989), S. 163–170
- [Wel95] WELLEK, S.: Einführung in die statistische Methodik von Äquivalenzstudien. In: *Informatik, Biometrie und Epidemiologie in Medizin und Biologie* 26 (1995), Nr. 2, S. 81–106
- [Wer92] WERNER, J.: *Biomathematik und Medizinische Statistik*. 2. Auflage. München : Urban und Schwarzenberg, 1992
- [Wes71] WESTLAKE, W.J.: A one-sided version of Tukey-Duckworth test. In: *Technometrics* 13 (1971), S. 901–903
- [WH00] WU, C.F.J. ; HAMADA, M.: *Experiments. Planning, Analysis, and Parameter Design Optimization*. New York : Wiley; pp. 630, 2000
- [Wil40] WILLIAMS, C.B.: A note on the statistical analysis of sentence length as a criterion of literary style. In: *Biometrika* 31 (1940), S. 356–361
- [Wil41] WILKS, S.S.: Determination of sample sizes for setting tolerance limits. In: *Ann. Math. Statist.* 12 (1941), S. 91–96
- [Wil42] WILKS, S.S.: Statistical prediction with special reference to the problem of tolerance limits. In: *Ann. Math. Statist.* 13 (1942), S. 400–409
- [Wil59] WILLIAMS, E.J.: The comparison of regression variables. In: *Journal of the Royal Statistical Society (Series B)* 21 (1959), S. 396–399
- [Wil87] WILCOX, R.R.: Pairwise comparisons of J independent regression lines over a finite interval, simultaneous pairwise comparisons of their parameters, and the Johnson-Neyman procedure. In: *British Journal of Mathematical and Statistical Psychology* 40 (1987), S. 80–93

- [WM41] WALLIS, W.A. ; MOORE, G.H.: A significance test for time series analysis. In: *J. Amer. Statist. Assoc.* 36 (1941), S. 401–409
- [WM83] WILSON, G.A. ; MARTIN, S.A.: An empirical comparison of two methods for testing the significance of a correlation matrix. In: *Educational and Psychological Measurement* 43 (1983), S. 11–14
- [Wri90] WRIGHT, T.: When zero defectives appear in a sample: upper bounds on confidence coefficients of upper bounds. In: *The American Statistician* 44 (1990), S. 40–41
- [WW64] WILCOXON, F. ; WILCOX, Roberta A.: Some Rapid Approximate Statistical Procedures. In: *Lederle Laboratories, Pearl River, New York* (1964), S. 29–31
- [Yat48] YATES, F.: The analysis of contingency tables with groupings based on quantitative characters. In: *Biometrika* 35 (1948), S. 176–181
- [Zar99] ZAR, J.H.: *Biostatistical Analysis*. 4th edition. Englewood Cliffs : Prentice-Hall; pp. 928, 1999
- [Zim84] ZIMMERMANN, H.: Die praktische Relevanz des McNemar-Tests. In: *Biom. J.* 26 (1984), S. 219–220
- [ZR71] ZAHN, D.A. ; ROBERTS, Gail C.: Exact χ^2 criterion tables with cell expectations one: an application to Coleman's measure of consensus. In: *Journal of the American Statistical Association* 66 (1971), S. 145–148

Namensverzeichnis

- Abbe, E., 215
Acton, F.S., 93
Adams, A.J., 271
Agresti, A., 518, 599
Aitchison, J., 91
Akaike, H., 575
Allen, F.R., Jr., 507
Altman, D.G., 302
Andersen, P.K., 615, 618
Ansari, A.R., 375
Armitage, P., 516
Atkinson, B., 591
- Backhaus, K., 560
Bartholomew, D.J., 517
Bartlett, M.S., 93, 421
Bauer, F.D., 275
Bayes, Th., 108, 307
Bell, C.B., 339
Benjamini, Y., 431
Bennett, B.M., 517, 536, 545
Berger, R.L., 96
Bernoulli, J., 108, 112, 125, 162
Beus, G.B., 526
Bienaymé, I.-J., 215
Bland, J.M., 302
Blom, G., 332
Bofinger, Eva, 557
Bonett, D.G., 272
Borgan, O., 615, 618
Bortkiewicz, L. von, 179
Bosch, K., 557
Bowker, A.H., 535
Bradley, R.A., 375
Brandt, A.E., 507
Breslow, N.E., 494, 623
Bresnahan, J.I., 524
Bretz, F., 581
Briggs, Henry, 30
Bross, I., 261
Brown, Sara M., 536
Buenaventura, A.R., 371
Busch, H., 483
Büning, H., 408
- Cacoullos, T., 369
Caliebe, Amke, 584
- Canavos, G.C., 299
Cantelli, F.P., 243
Cantor, Georg, 21
Carlson, F.D., 93
Casagrande, J.T., 485
Casella, G., 96
Castellan, N.J., 524
Cauchy, A.L., 80
Chakravarti, I.M., 269
Charles, J.A., 376
Chernick, M.R., 276
Chissom, B.S., 156
Chou, Y.-M., 302
Ciba-Geigy, AG, 258
Clopper, C.J., 257, 260
Cochran, W.G., 257, 420, 516, 536
Cohen, J., 537, 539, 540
Collett, D., 629
Conover, W.J., 299, 410
Cox, D.R., 477, 622, 623
Cramér, H., 531
Crow, E.L., 280, 281
Csörgö, S., 408
Cureton, E.E., 94
- d'Agostino, A., 329
d'Agostino, R.B., 330
Dalgaard, P., 19, 633
Dallal, G.E., 339
Darlington, R.B., 156
David, F.N., 170, 297, 298
David, H.A., 327
Davies, O.L., 371
Descartes, René, 24
Diamond, L.K., 507
Dimitriadou, E., 156
Dixon, W.J., 288, 346
Dodge, Y., 667
Donner, A., 590
Duckworth, W.E., 405
Dunn, O.J., 545
Dunnett, C.W., 440
- Edington, E.S., 411
Efron, B., 275, 623
Ehrenberg, A.S.C., 184
Eisenhart, C., 300, 363, 420

- Endriss, J., 280
 Euler, Leonhard, 24, 28, 49
 Everitt, B.S., 539, 540

 Fagan, T.J., 138
 Fahrmeier, L., 599
 Faraway, J.J., 408
 Felsenstein, J., 508
 Fermat, P. de, 108, 125
 Ferschl, F., 141
 Fienberg, S.E., 108
 Finucan, H.M., 156
 Fisher, R.A., 108, 139, 185, 244, 305, 318,
 322, 367, 370, 381, 418, 496, 498, 544,
 548
 Fleiss, J.L., 141, 484, 539–541, 557
 Fleming, T.R., 630
 Fountain, R.L., 302
 Fowlkes, E.B., 587
 Friedman, M., 456

 Gans, D.J., 413
 Gardner, R.S., 280, 281
 Gatsonis, C., 547
 Gauß, C.F., 41, 92, 203, 251
 Gebhardt, F., 330
 Gibson, Wendy M., 94
 Giesbrecht, F.G., 477
 Gill, R.D., 615, 618
 Gini, C., 72
 Good, P., 411
 Goodman, L.A., 605
 Gosset, W.S., 211
 Graf, U., 284
 Grambsch, P., 630
 Greenwood, M., 183, 613
 Greenwood, P., 335
 Grizzle, J.E., 483
 Gross, J., 336, 343
 Grubbs, F., 345
 Gumpertz, M.L., 477

 Haenszel, W., 491, 617
 Hahn, G.J., 292, 293
 Hamada, M., 477
 Hampel, F.R., 344
 Handley, B., 587
 Harrington, D.P., 630
 Hart, B.I., 361
 Harter, H.L., 406, 447
 Hartley, H.O., 330, 342, 347, 419, 450, 461
 Haseman, J.K., 486

 Hastay, M.W., 300, 420
 Hauck, W.W., 590
 Haviland, M.G., 483
 Haynam, G.E., 520
 Hayter, A.J., 434
 Helmert, F.R., 215
 Hendrickson, R.W., 293
 Hendy, M.F., 376
 Henning, H.J., 284
 Hettmansperger, T.P., 536
 Hewett, J.E., 557
 Hochberg, Y., 431
 Hochstädter, D., 141
 Holm, S., 430, 431
 Hommel, G., 431, 526
 Horn, M., 440
 Hornik, K., 359
 Hosmer, D.W., 493, 588, 617, 618
 Hotelling, H., 297
 Hothorn, T., 275, 359, 581
 Hsu, P., 517

 Iman, R.L., 442, 444
 Ireland, C.T., 536

 Jennings, D.E., 590
 Jensen, D.R., 526
 Johnson, N.L., 279
 Jonckheere, A.R., 453
 Jowett, G.H., 94
 Juran, J.M., 473

 Kahn, H.A., 142, 492
 Kalbfleisch, J.D., 613
 Kalton, G., 141
 Kaplan, E.L., 611
 Kastenbaum, M.A., 427, 428, 524
 Katz, D., 492
 Keiding, N., 615, 618
 Kelsey, Jennifer L., 491
 Kendall, M.G., 67, 482, 531
 Kerrich, J.E., 95
 Kim, Ch., 483
 Kim, P.J., 406
 Kimball, A.W., 513
 Kitagawa, T., 141
 Klar, Janelle, 493
 Kleppmann, W., 477
 Kohnen, R., 481, 483
 Koller, S., 91
 Kolmogoroff, A.N., 108, 243, 405
 Kramer, C.Y., 526

- Krummenauer, F., 537
 Kruskal, W.H., 442, 444
 Ku, H.H., 536
 Kuehl, R.O., 477
 Kullback, S., 536
 Kymn, K.O., 545
- Laan, P. van der, 400
 Lababidi, Z., 557
 Landis, J.R., 538, 541
 Laplace, P.S. de, 108, 112, 251, 307
 Lawal, H.B., 523
 Lee, Elisa T., 619
 Leibniz, Gottfried Wilhelm, 24, 47
 Lemeshow, S., 493, 617, 618
 Leone, F.C., 520
 Lepage, Y., 376
 Levene, H., 369, 423
 Levy, P., 204
 Lewontin, R.C., 508
 Ligges, U., 633
 Lilliefors, H.W., 339, 342
 Lindeberg, J.W., 204
 Lorenz, M.O., 83
 Lotz, R., 483
- Méré, A.G. de, 108, 125
 Madansky, A., 93
 Maindonald, J., 568
 Manly, B.F.J., 411
 Mann, H.B., 391, 395
 Mantel, N., 491, 617
 Markoff, A.A., 162
 Martin Andrés, A., 497, 498
 Martin, S.A., 545
 Martini, P., 508
 Mason, A.L., 339
 Massey, F.J., 288
 McCornack, R.L., 402
 McHugh, R.B., 495
 McKean, J.W., 536
 McNemar, Q., 90, 499
 Mee, R.W., 452
 Meier, P., 611
 Meis, T., 32
 Mendenhall, W., 271
 Meyer-Bahlburg, H.F.L., 372
 Michaelis, J., 456, 457
 Miller, L.H., 338
 Moivre, A. de, 108, 173, 176
 Molenaar, I.W., 385
 Montgomery, D.C., 474, 477
- Moore, G.H., 364
 Mosteller, F., 410
 Mould, R.F., 486
 Murphy, K.R., 405
 Myors, B., 405
- Natrella, Mary G., 104, 105
 Neill, J.J., 545
 Nelson, W.B., 616
 Nemenyi, P., 449
 Newton, Isaac, 47
 Neyman, J., 108, 139, 252, 318, 322
 Nikulin, M.S., 335
- Odeh, R.E., 299, 300, 456, 457, 464, 544
 Ord, J.K., 159
 Overall, J.E., 496
 Owen, D.B., 406
- Page, E.B., 462, 464
 Pascal, B., 52, 108, 125
 Pawitan, Y., 246
 Pearson, E.S., 108, 252, 257, 260, 318, 322, 328, 342, 347, 419, 450, 461, 491
 Pearson, K., 72, 108, 159, 215, 245, 508, 531
 Pillai, K.C.S., 371
 Poisson, S.D., 173
 Potthoff, R.F., 556
 Prentice, R.L., 613
- Ramsay, P.H., 436
 Rasch, D., 667
 Reibnegger, G., 259
 Reid, N., 477
 Rinne, H., 667
 Ripley, B.D., 596
 Roberts, G.C., 334
 Rogosa, D., 557
 Rosenbaum, S., 410, 413
 Royen, Th., 528, 529, 531
 Rump, S.M., 32
- Sachs, L., 4, 342, 408, 410, 428, 440, 460, 477, 481, 501, 514, 515, 526, 528
 Samiuddin, M., 545
 Sampson, A.R., 547
 Sarkar, S., 431
 Scheffé, H., 382
 Seier, Edith, 272
 Sempos, Ch.T., 142, 492
 Severini, T.A., 246
 Shapiro, M.M., 524

Shaw, G.B., 205
 Shiffler, R.E., 271
 Siddharta, R.D., 587
 Siegel, S., 372
 Simpson, E.H., 57
 Smirnoff, N.W., 405
 Smith, Judith A., 495
 Snedecor, G.W., 507
 Steiger, J.H., 545
 Stephens, M.A., 328, 342
 Stoline, M.R., 439
 Storer, B.E., 483
 Stuart, A., 159, 482
 Student, 652
 Städtler, P., 249
 Swed, Frida S., 363

Tate, M.W., 536
 Therneau, T.M., 591, 630
 Thompson, W.A., 280
 Tibshirani, R., 277
 Tiede, M., 91
 Tietjen, G.L., 330
 Timm, Neil H., 560
 Trapletti, A., 362
 Tschebyscheff, P.L., 154
 Tukey, J.W., 67, 93, 307, 372, 413

Upton, G.J.G., 523

Van der Waerden, B.L., 273
 Venables, W.N., 39, 596
 Vollandt, R., 440

Wald, A., 8, 108, 362
 Wallis, J., 147
 Wallis, W.A., 300, 364, 420, 442, 444
 Walter, E., 147, 404
 Watrous, J.B., Jr., 507
 Weir, J.B. de V., 383
 Welch, B.L., 279
 Wellek, S., 464
 Werner, J., 639
 Westergaard, H., v
 Westfall, P., 581
 Westlake, W.J., 413
 Wetzell, W., 303
 Whitney, D.R., 391, 395
 Wilcox, R.R., 557
 Wilcox, Roberta A., 449, 456, 461, 462
 Wilcoxon, F., 449, 456, 461, 462
 Wilkinson, L., 339

Williams, C.B., 205
 Williams, E.J., 545
 Wilrich, P.T., 284
 Wilson, G.A., 545
 Winsor, C.P., 69
 Wolfowitz, J., 362
 Wright, T., 258
 Wu, C.F.J., 477
 Wyatt, J.K., 405
 Yates, F., 381, 483, 548
 Yule, G.U., 183
 Zahn, D.A., 334
 Zar, J.H., 160, 379, 558
 Zimmermann, H., 502

Sachverzeichnis

- Abhängigkeit
 - gerichtete oder ungerichtete 121
 - kausale 491
- Ablehnungsbereich für H_0 (K_α) 311–312
- Abnahmeprüfung 319
- Abszisse (x-Koordinate) 43
- Abweichungen oder Residuen 6
- Abweichungen, zufällige gegenüber systematischen 14
- Abweichungsquadrate 92
- Achsenabschnitt (intercept) 43, 91
- Adaptive Verfahren 343
- Additionssatz 114
- adjustiertes Chancenverhältnis 595
- Änderungen, relative 77
- Änderungsrate, durchschnittliche 103
- Äquivalenzbereich 324
 - kritische Grenzen 355
- Äquivalenzgrenzen 415
- Äquivalenzintervall 416
- Äquivalenztest 324
- AIC-Kriterium
 - Allgemein 575
 - Cox-Regression 628
 - Modellbildung in R 596
 - Variablen-Auswahl 596
- Akaike Information Criterion 575
- Alpha-Fehler
 - welchen Wert sollte er nicht überschreiten? 308
- Alternativ- oder Nullhypothese 321
- Alternativmerkmale 15
- Analyse eines Problems 5
- Analyse von Vierfeldertafeln 479
- Anderson-Darling Test 342
- Annahmebereich ($\bar{K}_\alpha$) 311
- Annahmekennlinie 319
- Annahmezahl 319
- Anordnungswerte 63
- ANOVA, Analysis of Variance (im linearen Modell) 580
- Anpassung an eine Poisson-Verteilung 341
- Anpassungstests 333
 - Anpassung an eine Poisson-Verteilung 340
 - Vergleich einer empirischen mit einer möglichen theoretischen Verteilung 327
- Ansari-Bradley-Test 372, 375
- Anteil an einer Grundgesamtheit 299–301
- AOQ, Average Outgoing Quality (mittlerer Durchschlupf) 321
- Aposteriori Wahrscheinlichkeit 129
- Apriori Wahrscheinlichkeit 129
- Arbeits- und Wartezeiten 76
- Arbeitshypothese (H_A) 307
- Area Under Curve (AUC) 46, 416
- Arithmetischer Mittelwert 68
- Arithmetisches Mittel
 - $\bar{x}$ 68, 69
 - gewichtetes 75
 - gewogenes $\bar{x}_{gew}$ 74
- Arzneimittelprüfung
 - Unbedenklichkeit und Wirksamkeit 308
- Asbestfasern, Exposition von 92
- Asymptotische Effizienz 325
- AUC, Area Under Curve 46, 416
- Ausgleichsgerade 85
- Ausreißer
 - problem 343
 - Modellbildung (influential points) 597
 - robuste Regression 564
- Ausreißer (Extremwerte) 67
 - ja oder nein? 344
- Ausreißertest nach
 - Dixon 346
 - Grubbs 345
- Aussage
 - Schärfe oder Sicherheit? 264
- Ausschuss
 - Kontrolle 366
 - Quote 320
- Auswahl der Variablen (Regressionsmodell) 573, 595
- Auswahlsatz 239, 266
- Axiomatischer Wahrscheinlichkeitsbegriff 113
- Axiome nach Kolmogoroff 113
- $B(n; p)$ -Verteilung 165
- Backward Elimination oder Forward Selection? 596
- Badewannenkurve 611
- Bakterien 77
- Bartlett-Test 421
 - Beispiel (auch in R) 422

- Bartlett-Verfahren 93
- Basisrisiko 622, 623
- Baumdiagramm und Pfadregeln 119, 120
- Bayessches Theorem 128
 - und Pfadregel 129
- Bedingte
 - Dichtefunktionen 229
 - Verteilung und Unabhängigkeit 228
 - Wahrscheinlichkeit 116
- Befunde mit praktischer Relevanz 4
- Behrens-Fisher-Problem 382
- Bereichsschätzung (Intervallschätzung) 240
- Berkson's Fallacy 238
- Bernoulli-Kette vom Umfang n 165
- Bernoulli-Versuch 164
- Bernoulli-Verteilung 165
- Beschreibende Statistik 1
- Bestandsmassen 58
- Bestimmtheitsmaß 565
 - $\hat{B} = r^2$ 92
 - nichtlineares 102
- Beta-Fehler
 - wovon hängt er ab? 315
- Beurteilende Statistik 2, 10, 11
- Bevölkerungsdichte, durchschnittliche 79
- Bewegungsmassen 58
- Bewertende „Gewichte“ 75
- Beziehungszahlen 57
- Bias 34, 244
- Bindungen 63
- Bindungen bei Rangkorrelation 89
- Binomialentwicklung 165
- Binomialkoeffizient 49
- Binomialtest 347
 - Approximation durch die Normalverteilung 349–350
 - Likelihood-Quotienten-Test 351
 - wie viele Beobachtungen werden benötigt? 350–351
- Binomialverteilung 165, 246
 - Approximation durch die Poisson-Verteilung 172
 - Approximation durch die Standardnormalverteilung 171
 - Beispiele 166–169
 - ML-Schätzer 247
 - negative 179
 - oder Poisson-Verteilung? 179
 - Parameter 170
 - Test auf Anpassung an 333
- Binomialwahrscheinlichkeiten
 - einige tabellierte Werte 168
- Bioäquivalenz 324
- Bioäquivalenz-Test 416
 - Beispiel in R 417
- Bland-Altman Verfahren 302, 537
- Blindversuche 473
- Blockbildung 388, 470, 472
- Blockinterne Vergleiche 387
- Blockvarianzanalyse 454–455
 - Beispiel in R 455
- Blutgruppen 56, 61, 62
- Body-Mass-Index 69, 81
- Bonferroni
 - χ^2 -Tabelle 526
 - Holm-Test 430
 - Ungleichung 114, 123
 - Verfahren 525
- Bootstrap
 - Perzentilmethode 276
 - Schätzungen 276
 - Standardfehler 277
 - Stichprobe 276
 - t-Methode 277
- Bowker-Test auf Symmetrie 535
- Bowley-Koeffizient 160
- Box-Plot 66
- Brandt-Snedecor-Test 507, 508, 512
 - Beispiel (auch in R) 508
- Breslow-Nomogramm („ n für Psi“), Hinweis auf 494
- Briggssche Logarithmen 30
- Brown-Forsythe-Version des Levene-Tests 369, 423
- Challenger Katastrophe 587
 - Datenauswertung in R 589
- Chancen-Verhältnis (Odds Ratio) 118, 488
- Chancenverhältnis (odds ratio) 594
 - adjustiertes 595
- Charakteristische Gleichung 41
- Chemotherapie
 - Vergleiche (Beispiel mit R) 612
 - Vergleiche mit R 613
- Chevalier de Méré 169
- Chiquadrat
 - $k \cdot 2$ -Felder-Homogenitätstest 510
 - additive Eigenschaft 484
 - Anpassungstest 333
 - Komponenten 512–514
 - Test für eine $r \cdot c$ -Tafel 520
 - Variationsbereich für $r \cdot c$ -Tabellen 523
- Chiquadrat-Verteilung (χ^2) 215, 216

- ein- und zweiseitige Schranken für einen Freiheitsgrad 483
- exakte Wahrscheinlichkeiten für einen Freiheitsgrad 482
- Parameter 216
- Schranken 217
- Cholesterinwert als Funktion des Alters (Beispiel mit R) 562
- CM-Test 408
- Cochran
 - Kombination von Vierfeldertafeln 505, 506
 - Q-Test 536
 - Vergleich mehrerer Varianzen 420
- Cochran-Armitage Test auf Trend 516–519
 - Beispiele (auch in R) 517
- Cohen's Kappa-Koeffizient 537
- Confounding 488
- Cox-Regressionsmodell 621
 - Auswahl von Einflussgrößen 627
 - Cox-Snell-Residuen 629
 - Interaktionen zwischen Einflussgrößen 626
 - Modellrechnungen in R 627
 - Residuenanalyse 629
 - Skalierung der Einflussgrößen 625–627
- Cramér-von Mises Test 408
- Cross-Over Design 503
- Data Editing 9
- Data Splitting 322
- Daten 1, 161
 - Definition, Art, Gewinnung, Struktur 18
 - medizinische 8
 - multivariate; Reduktion ihrer Dimensionalität 9
 - sind sie „sauber“? 9
- Daten und Modelle 3
- Datenanalyse 4
 - konfirmative 10
- Datenbeschreibung 8
 - formalisierte? 318
- Datenfolge
 - zufällig verteilt? 360, 361
- de Morgan-Gesetze 110
- Design-Matrix im linearen Modell 577
- Deskriptive Statistik 8, 55
 - Maßzahlen und Skalenarten 55
- Determinanten 39
- Deterministische Komponente 561
- Devianz 590
 - Residuen 597
 - log-lineares Modell 603
 - Statistik 600
 - Differenzen (G-Statistik) 590, 593
- Dezile 64
- Diagnostischer Test 132, 133
- Diagramm-Varianten 59, 60
- Dichtefunktion 147, 148
- Dichtefunktion, gemeinsame
 - Beispiel 225, 226, 229
- Dichtemittel 56
- Differenz $\bar{G}$ der Devianzen 590
- Differenzenstreuung, sukzessive 360
- Differenzenvorzeichen-Iterationstest 364
- Direkter Schluss 260
- Disjunkte Mengen 23
- Diskrete Gleichverteilung 163
- Diskrete Zufallsvariable 145, 148
- Dispersion nach Gini-Simpson 57
- Dispersionsindex 179
- Distanzmaß nach Akaike (AIC) 575
- Diversität 57
- Dot-Plot 66
- Drei-Sigma-Regel 197
- Dreidimensionale Kontingenztafeln (Kontingenzquader) 604
- Dummy-Codierung 576
- Durchleuchtung der Brust mit Röntgenstrahlen 132
- Durchschlupf, mittlerer 321
- Durchschnittliche
 - Bevölkerungsdichte 79
 - Stückzeit 79
- Durchschnittlicher Korrelationskoeffizient 552
- $E(Z)$, Erwartungswert von Z 202
- Ecksumme 60
- EDA, Explorative (erkundende) Datenanalyse 9
- Effekt 387, 491
- Effekt-Parametrisierung
 - Dummy-Codierung 578
 - Effekt-Codierung 579
- Effizienz 241
- Eigenwerte und Eigenvektoren 41
- Ein- bzw. zweiseitiger Test 310
- Einfaktorielle Varianzanalyse im linearen Modell 577
- Einflussfunktion 97
- Einflussgrößen 14, 418, 470, 560
- Einheitskreis 45
- Einseitiger Test 316
- Einstichprobentests
 - Gauß-Test
 - am Beispiel 312–313

- Likelihood-Quotient 313
- Median-Test 358
- t-Test 352–354
 - P-Wert 353
 - wie viele Beobachtungen werden benötigt? 356–357
- Äquivalenz-Test 355–356
 - Beispiel zur Mikrozirkulation 355
- Elementarereignisse 109
- Elementarhypothesen 429
- Empirische Kovarianz 85, 286
- Empirische Verteilung; knappe Beschreibung 161
- Empirische Verteilungsfunktion 82, 145, 150
- Empirischer Korrelationskoeffizient 87
- Endlichkeitskorrektur 257, 266
- Entscheidungen im Falle von Ungewissheit 8, 307
- Entscheidungsprinzipien 306
- Entscheidungsprozesse 560
- Entsprechungszahlen 57
- Enzymkinetik 105
- Epidemiologie 139
- Ereignis 111
 - disjunktion 128
 - massen 58
 - raum 109
 - sicheres 109, 110
 - unmögliches 110
- Ereignisse
 - korrelierte 124
 - praktisch sichere 122
 - unvereinbare 123
- Ereigniszeiten
 - rechts zensiert 612
- Erfahrungen sollten wiederholbar sein 7
- Erfolgswahrscheinlichkeit (π) 348, 586
- Erhebung 10
- Erhebung typischer Einzelfälle 13
- Erkenntnisgewinnung: datengesteuert oder hypothesengesteuert? 10
- Erkrankungswahrscheinlichkeit 141
- Erwartungshäufigkeiten 480
- Erwartungshäufigkeiten gleich Eins 334
- Erwartungstreue 241
- Erwartungswert 112, 151
 - Beispiele und Rechenregeln 151, 152
 - einer Zielgröße im linearen Modell 584
 - Parametrisierung 578
- Euler-Symbol 49
- Eulersche Gammafunktion 212
- Eulersche Konstante 27
- Exakter Fisher-Test 496
- Experiment 10
- Experten beurteilen Weine 537
- Explorative Studien 322
- Exponentialfunktion 46
- Exponentialfunktion, nichtlineare Regression 103
- Exponentialpapier 105
- Exponentialverteilung 207–209, 618
 - Beispiele 208
 - Parameter 208
- Exponentielles Wachstum 77
- Extremabweichungen, standardisierte 346, 347
- Extremwert(e)
 - einer Stichprobe 301
 - noch brauchbar? 344
 - Studentisierte 436
- Exzess (kurtosis) 155, 160
- F-Test
 - Alternative 371, 375
 - wie viele Beobachtungen werden benötigt? 370
- F-Verteilung 218
 - 0,95-Quantile (obere 5%-Schranken) 219
 - 0,975-Quantile (obere 2,5%-Schranken) 220
 - nichtzentrale 415
 - Parameter 218
- Fagan-Nomogramm 138
- Faktoren 470
- Faktorielle Experimente 473, 474
- Fakultät: Begriff 28
- Fall-Kontroll-Studie 487
 - wie viele Beobachtungen werden benötigt? 493
- Fallzahl (sample size) 311
- Fallzahlabschätzung 316
 - Binomialtest 350–351
 - Vierfeldertest 484
 - Zweistichproben-t-Test, Beispiel (auch in R) 385–387
- Falsch positives Ergebnis 428
- Fehlentscheidung im statistischen Test 308
- Fehler 1. und 2. Art 307–308
- Fehlerbalkendiagramm 73
- Fernsehzeiten und Hausaufgaben (Beispiel) 225, 226, 229, 231
- Fisher Scoring 588
- Fisher-Pitman-Randomisierungstest 411
- Fisher-Test 481
- Fisher-Verteilung (F) 216

- Fisher-Verteilung, nichtzentrale 355
- Fleming-Harrington Schätzer 616
- Fläche unter der Kurve der Wahrscheinlichkeitsdichte 148
- Fläche unter der ROC-Kurve 136
- Flächen unter einer Funktion - Integrale 46
- Flügelspannweite und Alter von Sperlingen 295–296
- Formale Korrelation 91
- Forward Selection oder Backward Elimination? 596
- Fragestellung 12
 - was ist zu bedenken? 5
- Fraktil 64
- Freiheitsgrad (FG) 212
- Friedman-Rangsummen
 - paarweise multiple Vergleiche und Vergleiche mit einer Kontrolle 459, 460
- Friedman-Test 418, 443, 447, 456–459
- Funktion 43
- Funktion, logistische 586
- Funktionalparameter 150
- Funktionspapier 104
- Fünf-Zahlen-Maße (Tukey) 161
- Gambler's Fallacy 171
- Games-Howell-Verfahren; Varianzheterogenität zugelassen 439
- Gamma-Funktion 212, 215
 - wichtige Eigenschaften 213
- Gauss-Transformation 41
- Geburtsstagsproblem 126, 177
- Geburtsstagsverfahren 239
- Gegenhypothese 307
- Gehaltserhöhungen 76
- Geometrische Verteilung 185
 - Parameter und Beispiel 186
- Geometrischer Mittelwert 76
- Geometrisches Mittel $\bar{x}_G$ 76
 - gewogenes 76
- Gepaarte Beobachtungen 387
- Gesamtmittel 74
- Geschichtlicher Überblick 108
- Geschwindigkeitsdurchschnitt 79
- Gesetze der großen Zahlen 243
 - schwaches bzw. starkes 243
- Gesetzmäßigkeiten 8
 - der Schluss auf allgemeine 10
- Gesichtsfalten-Reduktion mit Botulin (Beispiel in R) 540
- Gewichte, bewertende 75
- Gewichtete lineare Kontraste 432
- Gini-Index 72, 83
- Gini-Simpson-Index 57
- Gleichmäßig bester Test 314
- Gleichung von Wilks 301
- Gleichung zweiten Grades 99, 101
- Gleichverteilung 163
 - Prüfung auf 334
 - stetige 190
- Gliederungszahlen 57
- Globales Signifikanzniveau 429
- Globalhypothese 428
- Glockenkurven 192
- Grenzwertsatz von de Moivre und Laplace 204
- Größe eines Wurfes bei Mäusen (Beispiel in R) 568
- Große Zahlen anschaulich gemacht 26
- Grubbs-Test 345
- Grundgesamtheit 2, 12, 144, 161
 - anteile 299
 - Population wie groß? 188
- Grundrechenarten 24, 26
- Gruppenfehlschluss 91
- Gruppierung, zusammenfassende 523, 524
- Güte eines Tests, Power 311
- Gütefunktion (power function) 314
- H-Test von Kruskal und Wallis 369, 418, 442, 443, 447
 - Beispiel (auch in R) 443
 - mit Stichproben-Untergruppen 448
 - paarweise Vergleiche mittlerer Ränge 445
 - Variante: Vergleiche mit einer Kontrolle 451
 - wie viele Beobachtungen werden benötigt? 427
- H_0
 - Nullhypothese 307
 - plausibel? 324
 - und P-Wert 324
- H_A
 - Alternativhypothese (Arbeitshypothese) 307
- Häufigkeiten
 - absolute und relative 56
 - bedingte 62
 - relative, Vergleiche 477
- Harmonischer Mittelwert 78
- Harmonisches Mittel $\bar{x}_H$ 78
 - gewichtetes 78
 - gewogenes 78
- Hartley-Test 419

- Hat-Matrix (Hut-M.) und weitere Schätzungen 567
- Haupteffekte 476
- Hazardfunktion, Risikofunktion 610
- Hazensche Gerade 331
- Heavy Tailed Distributions 192
- Herkunft von Ausschussware 131
- Hierarchisch geordnete Hypothesen zur Unabhängigkeit 606
- Histogramm 80
- Historischer Überblick 108
- Höhenlinien (Linien gleicher Wahrscheinlichkeit) 231, 232
- Holm-Prozedur 430
- Homogenitätstest
 - einer $r \cdot c$ -Tafel 520
 - für mehrere verbundene Stichproben 536
 - nach Ryan 514
- Hsu-Ansatz 268
- Hypergeometrische Verteilung 186, 496
 - Beispiele 187–189
 - drei Approximationen 190
 - Parameter 187
 - verallgemeinerte 187
- Hypothesen
 - einfache bzw. zusammengesetzte 310
 - prüfen und gültige anreichern 11
 - sind vor der Datengewinnung zu formulieren 322
 - statistische (H_0 und H_A) 307–308
- Hypothesentest als Entscheidungshilfe 322
- Hypothesentest im logistischen Regressionsmodell 589
- Identifikationsgrößen 14
- Index 58
- Indexkorrelation 90
- Indexzahl 58
- Indikatorfunktion 394
- Indikatorvariablen (Designvariablen) 591
- Indirekter Schluss 259
- Influential Points 597, 598
- Inhomogenitäten 332
- Inhomogenitätskorrelation 90
- Inklusionsschluss 267
- Integral 46
- Interaction-Plot 584
- Interaktionseffekt 584
- Interaktionsterm 584
- Interquartilbereich (IQR) 65
- Intervall- und Verhältnis-Skala 15
- Intervall-Zensierung 612
- Intervallinklusion 324
- Intervallinklusionsprinzip 416
- Intervallschätzung (Bereichsschätzung) 240, 253
- Inverse Matrix 39
- Inverse Prädiktion aus einer linearen Regression 296
- Inversionen 67
- Inzidenz 140
 - und Prävalenz 139
- Inzidenzdichte-Verhältnis 142
- Inzidenzraten 488
- Irrtumswahrscheinlichkeit 305, 307
- Irrtumswahrscheinlichkeit, empirische 323
- Iterationstest 361
- Iterationszyklus 4
- Jonckheere-Trendtest 418, 452–454
- $K_{\alpha'}$, Ablehnungsbereich für H_0 311
- $k \cdot 2$ -Felder- χ^2 -Test nach Brandt und Snedecor 507
- $k \cdot 2$ -Feldertafel, Zerlegung der Freiheitsgrade 511, 512
- $k \cdot 2$ -Feldertafel, Trend 516
- $k\sigma$ -Bereiche für unterschiedliche Verteilungstypen 204
- k -Permutationen 49
- K-S-Test für Blutzucker- und Cholesterinwerte 338
- k -tes zentrales Moment 154
- $K_{\alpha'}$, Ablehnungsbereich für H_0 311
- Kaplan-Meier Schätzung 611
 - Anmerkungen 616
 - Beispiel in R 613
 - graphische Darstellung 615, 618
- Kappa
 - Beispiel (auch in R) 539
 - Details und Beurteilung 538, 539
 - für Mehrfachbeurteilungen 541
 - gewichtet (Beispiel, auch in R) 540
 - Konfidenzintervall 539
 - Übereinstimmungsmaß zweier Beobachter 537, 538
- Kategoriale oder qualitative Merkmale 477
- Kausale Abhängigkeit 121, 491
- Kausale Korrelationen 90
- Kausalität 90
- Kausalitätskriterien 491
- Kehrmatrix 39
- Kendall, Kontingenzkoeffizient nach 531
- Kerrich-Verfahren 95
- KI für einige Parameter (griech. Buchstaben)

- α und β (Regression) 291
- $\beta_1 - \beta_2$ 556
- γ (Variationskoeffizient) 279
- κ (Cohen's Kappa) 539
- λ (Poisson-Verteilung) 280
 - Crow-Gardner Tabelle 281
- μ 264–267
 - Bootstrap-Stichprobe in R 276
 - mit t -Verteilung, R und Beispiel 265
 - weitere Details und Beispiele 266
- $\mu_1 - \mu_2$ 267, 268, 379
 - Beispiel 268
- μ_1/μ_2 (nach Chakravarti) 269
- μ_d (Paardifferenzen) 269, 390
 - Beispiel mit R 269
- ω (Odds Ratio) 491
- π 260
- π (Approximation) 256
- $\pi_1 - \pi_2$ 261
- $\pi_1 \rightarrow \pi_2$ (Zunahme) 261
- $\pi_i - \pi_{i'}$ 510
- ψ (relatives Risiko) 491
- ρ (Korrelationskoeffizient) 297
 - Beispiele, auch in R 297–299
- σ^2 bzw. σ 278
- σ_1^2/σ_2^2 279, 368
- $\tilde{\mu}$ (Median) 273, 274
 - Beispiel in R 275, 277
- $\tilde{\mu}_1 - \tilde{\mu}_2$ 400
- $\tilde{\mu}_d$ (Paardifferenzen) 403
- ϱ (Korrelationskoeffizient) 552
- v (Verhältnis) 270
- KI, Konfidenzintervall 241, 252–255, 263
 - Achsenabschnitt (Regression) 291
 - Allgemeines 267, 268
 - Ansätze für eine Homogenitätstafel (r.c) 523
 - Anteilswert aus einer dichotomen Grundgesamtheit 254
 - Berechnung für π mit R 255
 - Details und t-Test 379
 - einseitiges 265
 - Erwartungswert μ 265
 - Erwartungswert von Y an der Stelle $X = x$ 292
 - Kappa 539
 - Korrelationskoeffizient ρ 297
 - Median 273
 - Mittlere absolute Abweichung vom Median 271–272
 - Nullergebnisse und Vollergebnisse 258
 - Odds Ratio und Relatives Risiko 491
 - Quantile, mit Beispiel 274, 275
 - Regressionsgerade 292, 294
 - Beispiel, auch mit R 294
 - Regressionskoeffizient 291
 - Restvarianz 291
 - Variationskoeffizient γ , mit Beispiel 279
 - Wahrscheinlichkeit 260
 - Weibull-Gerade, mit Beispiel in R 286
- Klassierte Messwerte 74
- Klumpen-Verfahren 238
- Klumpung oder regelmäßiger Wechsel? 362
- Kodierungen 19
- Koeffizientenmatrix 41
- Körpergröße 160
- Kohorten-Studie 488
 - wie viele Beobachtungen werden benötigt? 494
- Kollektive Korrelation 91
- Kolmogoroff-Smirnoff Test 406
- Kolmogoroff-Smirnoff-Anpassungstest 337
- Kombination von Vierfeldertafeln 505
- Kombinationen: vier Varianten 50, 51
- Kombinationsvergleiche 474
- Kombinatorik 47
- Komplementärmenge 23
- Konservativer Test 309, 314
- Konsistenter Test 314
- Konsistenz 242
- Konstanten 31
- Konsumentenrisiko 320
- Kontingenzkoeffizient
 - maximaler nach Kendall 531
 - nach H. Cramér 531
 - nach Pearson 530, 531
- Kontingenzquader
 - hierarchische Unabhängigkeitshypothesen 606
 - Nullhypothesen 606
 - Unabhängigkeitsvarianten 604
- Kontingenztabelle 477
 - log-lineares Modell 598
 - log-lineares Modell (Beispiel in R) 601, 603
 - stochastische Unabhängigkeit 520
- Kontingenzwürfel 599
- Kontinuitätskorrektur 257
 - für Scoring 511
 - zum Vierfeldertest 483
- Konzentration von Marktanteilen 84
- Konzentrationsmaß nach Gini 83
- Korrelation 123

- Korrelation und Regression
 - einfaches Beispiel, auch mit R 289
- Korrelationsanalyse 543
- Korrelationskoeffizient 85, 543
 - dimensionsloses Zusammenhangsmaß 230
- Korrelationskoeffizient ρ_{XY} 230
- Korrelationskoeffizient nach Kendall 67
- Korrelationskoeffizient, empirischer 87, 287
- Korrelationsziffer 297
- Kovarianz
 - $Cov(X, Y)$ 230
 - empirische (s_{xy}) 85, 86, 286
 - Zerlegung 91
- Krankheitsdauer, mittlere 141
- Krankheitsursachen aufspüren 139
- Kreisprozess; Struktur und Details 3, 4
- KS-Zweistichprobentest 405–407
- Kubikzahlen 29
- Kumulierte Risikofunktion 622
- Kurtosis, Steilheit, Wölbung 330
- Kurtosis-Varianten 156, 160
- Kurvenformen I bis IV 100
- Kyphose nach Wirbelsäulenoperation
 - Beispiel in R 591
 - Devianz- und Pearson-Residuen 598
- LAD-Methode 97
- Lage-Test nach Rosenbaum 410
- Lageschätzer bei Kontamination 343
- Lambda; Tabelle oberer Vertrauensgrenzen für λ 283
- Lateinische Quadrate 473, 474
- Lawal-Upton Korrektur 523
- Least Absolute Deviation (LAD) 97
- Lebensdaueranalysen 205
- Leistungsvergleich von drei Schülern 75
- Lepage-Test 376
- Letalität 126
- Levene-Test 369, 421, 423
- Likelihood-Funktion
 - exponentielles Überlebenszeit-Modell 619
 - log-lineares Modell 600
 - Logistische Regression 588
- Likelihood-Quotient 137
- Likelihood-Quotienten-Test
 - Auswahl der Variablen bei der Modellbildung 595
 - Cox-Regression 628
 - Log-lineares Modell 600
 - Logistische Regression 590
 - Neyman-Pearson Lemma 313
- Lilliefors-Modifikation des Kolmogoroff-Smirnoff-Tests 339
- Lineare Funktionen 43
- Lineare Kontraste
 - gewichtete Kontraste 432
 - Maximalzahl 435
 - nach Scheffé 431–436
- Lineare Regression 91, 562
 - multiple 566
- Lineares Modell
 - Erwartungswert der Zielgröße 584
 - Hypothesentest und KI 571–573
 - Prädiktionsintervall 572
 - Varianzanalyse 577
- Lineares Zusammenhangsmaß 232
- Linearisierung von Punktwolken 106, 107
- Linearitätsprüfung einer Regression (Beispiel in R) 553
- Linearkombinationen 41
- Linkfunktion 561
 - Logit-Transformation 591
- Links-Zensierung 612
- Linkssteile Verteilungen 204
- Log-lineares Modell 598
 - Devianz-Statistik und AIC-Kriterium 605
 - Drogenbeispiel in R 605
 - Einschränkungen und Hinweise 608
 - Interpretation der Modellparameter 608
 - Modellauswahl 605
 - Unabhängigkeitshypothesen 606
 - zwei Faktoren 602
- Logarithmen 30
- Logarithmische Funktion 46
- Logarithmische Normalverteilung 205
- Logische Operatoren 20
- Logistische Funktion 586
- Logistische Regression 561, 585
 - Hypothesentest 589
 - Interpretation der Regressionskoeffizienten (odds) 594
 - Likelihood-Funktion 588
 - Maximum-Likelihood Schätzung 588
 - Residuenanalyse 597
- Logit-Transformation 586
- Loglineares Modell 561
- Lognormalverteilung 204–207
 - Beispiel 206
 - Bioäquivalenz 416
 - Parameter und Kennzahlen 206
- Logrank-Test 616
 - Beispiel in R 617

- Lokale Kontrolle 471
- Lokales Signifikanzniveau 429
- Lorenzkurve 83
- LSD-Test nach Hayter (mit Beispiel) 434, 436
- Lückentest für geordnete μ_i (mit Beispiel) 434, 436
- Lückentest nach Ryan 514
- Lungenfunktion 19
- M-Schätzung nach Huber 98
- Mantel-Haenszel-Test 503
 - Beispiel (auch in R) 504
 - Kontinuitätskorrektur 504
- Mantel-Haenszel-Teststatistik 504
- Markoffsche Ketten 162
- Martingal-Residuen
 - Cox-Regression 630
- Matched Pairs 487
- Matching 388
- Materialermüdung, Ausfälle durch 611
- Matrixaddition und -subtraktion 35
- Matrixalgebra 34
- Maximalabweichung, standardisierte 428
- Maximax-Kriterium 306
- Maximum-Likelihood Schätzung 246
 - Ansatz 244
 - log-lineares Modell 600
 - Logistische Regression 588
- Maximum-Test für Paardifferenzen 404
- Maßzahlen der zentralen Lage 156
- McNemar-Test 497, 537
 - Beispiel (auch in R) 499
 - Konfidenzintervall 500
 - Kontinuitätskorrektur 499
 - Power und benötigte Fallzahl 502
 - Überkreuzversuch 503
- Mean Survival 615
- Median
 - wert $\tilde{x}$ 63, 69
 - Deviation (MAD) 65
 - Quartile-Test 391, 414
 - Test 413
 - Test nach Wilcoxon 358–359
 - Vertrauensgrenzen 404
- Median Survival 615
- Mehrdimensionale Kontingenztafeln 602
- Mehrfachtests 428, 525
 - t-Test nach Bonferroni 430
 - unterscheide lokales, globales und multiples Signifikanzniveau 429
 - Wahrscheinlichkeit für P_0 und $P_{\geq 1}$ 428
- Mehrfelder-Chiquadratstest 519–521
- Mehrstichprobenverfahren 418
- Mehrstichprobenvergleiche, verteilungsunabhängige 418
- Mengen 21
- Mengenlehre: einige Verknüpfungen 109, 110
- Mengenoperationen 22
- Merkmal, Merkmalsausprägung und Merkmalsträger 12, 161
- Merkmal; intensives gegenüber extensives 79
- Merkmalskombinationen 523
- Messen 15
- Messreihen vergleichbar gemacht 72
- Messwerte
 - klassierte 74
 - Vergleich zweier Methoden 302
- Messzahlen 58
- Methode der kleinsten Fehlerquadrate (OLS) 251
- Methode der kleinsten Quadrate 92, 244
- Metrische Daten 68
- Michaelis-Menten Gleichung 105
- Mindestumfänge (Häufigkeiten) 262
- Minimales n zur Schätzung von
 - S , mit Beispiel 280
 - $\bar{X}$ 271
 - π (Beispiel) 262–263
- Minimax-Kriterium 306
- Minimum-Effekt- H_0 405
- Mischverteilungen 332
- Mittel, quadratisches 80
- Mittelwert der Zuwachsraten 76
- Mittelwert-Vergleich
 - Bemerkungen 384
 - Varianten 384
- Mittelwerte, die robust sind 69, 70
- Mittelwertgruppen, Bildung homogener 434, 436
- Mittlere absolute Abweichung vom Median 271
 - KI mit Beispiel in R 272
- Mittlere absolute Abweichung vom Medianwert (MAD) 65
- Mittlerer quadratischer Fehler (Mean Squared Error, MSE) 243–244
- Mitursachen 139
- ML-Schätzer (Beispiele)
 - Binomialverteilung 247
 - Münzwurf 246, 247
 - negative Binomialverteilung 248, 249
 - Normalverteilung 250
 - Poisson-Verteilung 250
- ML-Schätzung 246

- Beispiel in R 248
- Eigenschaften 246–247
- Modalwert 56
- Modell (in der Statistik) 4–7, 11
 - Abweichung 343
 - Bildung 560–561
 - Auswahl der Variablen 595
 - Verteilungsmodelle 560
 - Devianz, Zerlegung der 597
 - Matrixschreibweise 563
- Modell(e)
 - gesättigtes (saturated) 590
 - Bernoulli-Verteilung 165
 - Binomialverteilung 166
 - diskrete Zufallsvariablen 144
 - Exponentialverteilung 207–209
 - geometrische Verteilung 185
 - hypergeometrische Verteilung 186
 - Lognormalverteilung 204–207
 - negative Binomialverteilung 181–185
 - Normalverteilung 191–204
 - Poisson-Verteilung 173–181
 - Polyhypergeometrische Verteilung 188
 - Polynomialverteilung 233–234, 600
 - Weibull-Verteilung (2 Parameter) 209–210
 - zweifache Varianzanalyse 469
- Modellierung zufallsabhängiger Befunde 11
- Modellkomponenten
 - systematische 561
 - zufällige 561
- Modellvoraussetzungen, Test der 324
- MOM-Schätzer
 - Beispiele und Eigenschaften 245
- Momente
 - $g_1 = \sqrt{b_1}$ und $g_2 = b_2$ 329
 - empirische; Berechnung von g_1 und g_2 155
 - Schiefe und Exzess 154
 - zentrierte 156
- Momentenschätzer (Method of Moments, MOM) 245
- Monte-Carlo-Simulation 243
- Moore-Penrose (inverse Matrix) 39
- Morbidität 126
- Mortalität 126, 141
- Mortalitätsverhältnis, standardisiertes 142
- Mosaikplot 61
- Mosteller-Schnelltest 410
- MSE, mittlerer quadratischer Fehler 243–244
- Multi-Rater Kappa 541
 - Beispiel in R 542
- Multinomialkoeffizient 54
- Multinomialverteilung (Polynomialverteilung) 233–234, 600
- Multiple lineare Regression 566–571
 - Beispiel ausführlich in R 568–571
 - Kollinearität 571
 - Singularität 571
 - Variablenauswahl 573
- Multiple logistische Regression 591
- Multiple Vergleiche 428–431
 - nach Dunnett 440
 - nach Tukey 581
 - nach Tukey-Kramer 436
- Multiples Signifikanzniveau 429
- Multiples Testproblem 428–431
- Multiplikation zweier Matrizen 37
- Multiplikationssatz 117
- My, μ , Mittelwert der Grundgesamtheit
 - $\mu \pm 3\sigma$, Verteilungsanteile für unterschiedliche Verteilungstypen 204
 - Schluss von $\bar{X}$ auf μ 263, 267
- $N(0; 1)$, Standardnormalverteilung 194–196
 - $F(z)$ für $[-2, 99 \leq z \leq 0]$ 195
- $N(\mu; \sigma)$, Normalverteilung 191–204
 - Hazensche Gerade 331
- n-Fakultät 28, 48
- $n=30$, Aussagekraft einer Stichprobe dieses Umfangs 301
- $n_{\min}$, um S zu schätzen, mit Beispiel 280
- $n_{\min}$, um ein $\bar{X}$ zu schätzen 271
- $n_{\min}$, um ein π zu schätzen (Beispiel) 262–263
- Natürliche Logarithmen 31
- Nebenwirkungen 258, 259
- Negative Binomialverteilung 179, 181–185
 - Beispiele 182, 183
 - ML-Schätzer 248
 - Parameter 183
 - Spezialfall; Geometrische Verteilung 185
- Negativer Voraussagewert 132
- Nelson-Aalen Schätzer 616
 - Cox-Snell-Residuen 629
- Nemenyi-Vergleiche 449
- Neuerkrankungen 140
- Neuerkrankungsraten 488
- Neutrales Element 37
- Neyman-Pearson Lemma 313
- Nichtlineare Funktionen 44

- Nichtlineare Regression 99
- Nichtparametrische Methoden 325
- Nichtzentrale F-Verteilung 415
 - Quantile 355
- Nichtzufälligkeitsprüfung 362
- Nominalskala 15
 - Einflussgrößen, nominal-skaliert 576
 - Merkmale, nominal-skaliert 477
- Norm eines Vektors 38
- Normalgleichungen 99
- Normalverteilte Schätzfunktion 242
- Normalverteilung 191–204
 - Anpassung an 333
 - logarithmische 205
 - ML-Schätzer 250
 - Prüfung auf 335
 - Standardnormalverteilung 194
 - Wahrscheinlichkeitsdichte 193
 - zweidimensionale 231, 543
- Normierter Vektor 38
- Null-Eins-Verteilung 164
- Nullhypothese (H_0) 305, 307, 309
 - mögliche Fehlentscheidungen 308
- Nullklasse 183, 185
- Nullmatrix 36
- Nullmodell 591, 595
- OC-Kurve 319
 - für einen Stichprobenplan 320
- Odds Ratio 118, 489, 490, 494, 496
 - adjustiertes (Beispiel auch in R) 505
 - Beispiel in R 493
- Ökonometrie 184
- Oktile 160
- OLS-Methode 251
 - Schätzer 251
 - Schätzer zur Regression, lineare und nicht-lineare 252, 253
 - Schätzung im linearen Modell 563
- Operationscharakteristik (OC) 314, 319
- Operatoren 20
- Ordinalskala
 - Datenbeschreibung 62
 - Merkmale, ordinal-skaliert 477
- Ordinary Least Squares (OLS) 92, 251
- Ordinate (y-Koordinate) 43
- Orthogonale
 - Regressionsgeraden 93
 - kleinste Quadrate 96
 - Matrix 42
 - Projektion 566
 - Vektoren 38
- Orthonormale Vektoren 38
- $P(-1, 96 \leq Z \leq 1, 96) = 0, 95$ 197
- P-Wert 307
 - adjustierter 527
 - mittlerer 148
 - multiple Testproblem 428
 - und H_0 324
 - und Sternsymbolik 323
- P-Werte, nach Holm/Hochberg geordnet (Beispiel in R) 431
- Paarhypothesen 429
- Paarige Stichproben 387
- Page-Test 418, 462
- PAR, Population Attributable Risk 494, 495
 - weiterführende Tabellen 495
- Parabel 44
- Parameter 4, 6
 - Hypothesen 307
 - Raum 310
 - Test 310, 352
 - einer Verteilung 162
 - Schätzung für einen faktoriellen 2^3 -Plan (Beispiel in R) 475
- Parameterzahl optimieren nach dem AIC-Kriterium 575
- Partial-Likelihood Estimation 623
- Pascalsches Dreieck (mit 5 Identitäten) 53
- Pearson-Residuen 597
 - log-lineares Modell 603, 607
- Periodische Funktionen 45
- Permutationen 47
- Permutationstest 410
 - Beispiel mit R 412
- Perzentile 64
- Pfadregeln 119
- Pferdehufschlagtote 179
- Phasenhäufigkeitstest von Wallis und Moore 364
- Π , π , relative Häufigkeit in der Grundgesamtheit
 - π ist mit kleinstem n zu schätzen (Beispiel) 262–263
 - 95%-Konfidenzintervalle, ausgewählte 254
- Pillai-Buenaventura-Test (Streuungsvergleich) 370
- Pilotstudien 322
- Planen 4
- Poisson-Verteilung 173–181, 183
 - Approximation durch die Standardnormalverteilung 181
 - Beispiele 174, 176–178, 184

- Details zu λ 175
- einige tabellierte Wahrscheinlichkeiten 176
- Einstichproben-Lambda-Test 366–367
- Form 175
- Konfidenzintervall 280
- ML-Schätzer 250
- Parameter λ 174
- Prüfung auf 337
- Test auf Anpassung an 333
- verallgemeinerte 184
- wie stark ist die Nullklasse besetzt? 179
- zusammengesetzte 179
- Polyhypergeometrische Verteilung
 - Beispiele 188
 - Parameter 188
- Polynomfunktionen 44
- Polynomialverteilung (Multinomialverteilung) 600
 - Entstehung, Parameter und Beispiele 233
- Population Attributable Risk 494
- Positiver Voraussagewert 133
- Posttest-Chance 138
- Posttest-Wahrscheinlichkeit 138
- Potenzen und Wurzeln 29
- Potenzmenge 22
- Potenzmomente 156
- Power 309, 315
- Power eines χ^2 -Tests 524
- Power eines Tests
 - wovon hängt sie ab? 315
- Power und Fallzahl für den McNemar-Test 502
- Power zum Vierfeldertest 484
- Powerfunktion 314
- Prädiktion, inverse aus einer linearen Regression 296–297
- Prädiktionsintervall (lineares Modell) 572
- Prätest-Chance 138
- Prätest-Wahrscheinlichkeit 138
- Prävalenz 133, 138
 - Stufen 136
 - eines Risikofaktors 494
 - und Inzidenz 139
- Praktische Relevanz 4
- Prediction Interval (Voraussagebereich) 293
- Preisanstieg für Fische und Meeresfrüchte 97
- Prinzipien der Versuchsplanung 470
- Probability P 112
- Probandenpaare 388
- Probit-Transformation 381
- Problem: Überlegungen und Lösungsstrategien 5
- Produktdefinition der Unabhängigkeit 120
- Produktzeichen 28
- Produzentenrisiko 320
- Profildigramm (interaction plot) 584
- Programm R 24
- Projektion, orthogonale 566
- Proportional-Hazards Modell 622
 - Schätzung der Parameter 623
- Proportionale Risikofunktionen 622
- Proversionen 67
- Prozentpunkte 59
- Prozentsatzdifferenzen, minimale 478
- Prozentuale Zunahme? 261, 262
- Prozentwerte, Prozentzahlen 59
 - Umgang mit 254
- Prüfgröße (Teststatistik) 306, 307
- Prüfgrößen (Testverteilungen) 210–218
- Prüfplan 319
- Prüfung der Gleichheit zweier Varianzen paariger Stichproben 390
- Prüfung der Linearität einer Regression 552
- Prüfung der Nullhypothese: $\lambda = \lambda_x$ 282
- Prüfung des Rang-Korrelationskoeffizienten ϱ_S 557
- Prüfung einer Zeitreihe auf Trendänderung 364
- Prüfung einiger Nullhypothesen:
 - $H_0 : \alpha_1 = \alpha_2$ 557
 - $H_0 : \alpha_{0;yx} = \alpha_{yx}$ 555
 - $H_0 : \beta_1 = \beta_2$ 555
 - $H_0 : \beta_{0;yx} = \beta_{yx}$ 554
 - $H_0 : \beta_{yx} = 0$ 553
 - $H_0 : \lambda = \lambda_0$ 366–367
 - $H_0 : \rho = 0$ 544, 546
 - $H_0 : \sigma_1^2 = \dots = \sigma_k^2$ nach Levene 369
 - $H_0 : \varrho_1 = \varrho$ 549
 - $H_0 : \varrho_1 = \varrho_2 = \dots = \varrho$ 551
 - $H_0 : \varrho_1 = \varrho_2$ 549
 - $H_0 : \varrho_S = 0$ 557
- Prüfung von m Vierfeldertafeln 617
- Prüfverteilungen 211
- Pseudozufallszahlen 191
- Punktnotation 61
- Punktschätzung 240
- Punktwolke 85, 86, 88, 92
- Punktwolken, Linearisierung von 106
- Q-Symbolik 377
 - Q_x, Q_y, Q_{xy} 286
- Q-Test nach Cochran 536

- Q-Test nach Dixon 345, 346
- QQ-Plot 330
- Quadratische Formen 42
- Quadratisches Mittel 80
- Quadratzahlen 29
- Qualitative und quantitative Merkmale 12
- Qualitätskontrolle 319
- Qualitätsüberwachung 188, 346
- Quantile 64
 - einseitige KI 299
 - KI, mit Beispiel 274, 275
- Quantile-Quantile Plot, QQ-Plot 331
- Quantilmaße zu Schiefe und Exzess 160
- Quartile 64, 160
- r-c-Tafel 519, 520
 - Ansätze nach Royen 528
 - Beispiel (auch in R) 521
 - Lokalisationsansatz nach Hommel 526
 - schlecht besetzte, daher zu vereinfachen 521
 - schwach besetzte (Lawal-Upton Korrektur) 523
 - Trend? 532–535
 - Ursachen einer möglichen Signifikanz 524
 - Zerlegung in unabhängige Komponenten 524, 525
- r-r-Tafel
 - Symmetrie 535
- r, Stichprobenkorrelationskoeffizient 544
 - einige Prüfungen 545–548
 - Schätzung - wie viele Beobachtungen werden benötigt? 547
 - Umrechnung in $\hat{z}$ 548, 549
- R/s - Quotient
 - $N(\mu, \sigma)$ 327, 328
- Randomisierte Blöcke 472, 473
- Randomisierung 471, 472
- Randomisierungstest 410
- Randsommen in Tabellen 60
- Randverteilungen und Unabhängigkeit 226
- Rang 391
- Rang einer Matrix 40
- Rang- oder Ordinalskala 15
- Rang-Block-Varianzanalyse nach Friedman 456
- Rang-Korrelationskoeffizient r_S 88, 557
 - kritische Schranken 558
- Rangdaten 325
- Rangdispersionstest von Siegel und Tukey 371
- Rangliste 63
- Rangsummentest 325, 391
- Rangzahlen 164
- Realisierung von Zufallsvariablen 144
- Receiver Operating Characteristic 136
- Rechenschema, altväterliches 32
- Rechteckdiagramm 61
- Rechts-Zensierung 612
- Regression
 - lineare, Schätzung einiger Standardabweichungen 286, 288–289
 - mehrere Einflussgrößen 566
 - nichtlineare 99
 - robuste lineare 97
 - Sperlingsbeispiel mit R 295–296
 - von Y auf X 91
- Regressionsgerade 85
 - spezielle Schätzungen 93
- Regressionskoeffizient 91, 287
 - Standardfehler, KI und Teststatistik 565
- Regressionsmodell 562
 - nach Cox 621
 - Varianzkomponenten 565
- Regressionsparameter
 - Prüfung verschiedener Nullhypothesen 553
- Reguläre Matrix 40
- Reihenuntersuchung 136
- Relationen, mathematische 20
- Relative Häufigkeit und Wahrscheinlichkeit 112
- Relative Häufigkeiten, Vergleich mit einem vorgegebenen Verhältnis 479
- Relativer Variationskoeffizient V_r 73
- Relatives Risiko 118, 488, 489, 494
- Relatives Risiko und Exposition 491
- Repräsentationsschluss 267
- Repräsentativität einer Stichprobe 235
- Resampling-Verfahren 275
- Residualvarianz 564
- Residuen 92
- Residuen, nichtlineare Regression 102
- Residuenanalyse 564
 - Cox-Regression 629
 - logistische Regression 597
- Resistente Schätzverfahren 10
- Restmenge 23
- Resultatvalidität eines diagnostischen Tests 134
- Rho (ρ), Korrelationskoeffizient 543
- Risiko
 - Maße 118
 - mit Beispielen (auch in R) 490

- Zeiten 140
- für Lungenkrebs 495
- konstantes 618
- kumuliertes 629
- relatives 118
- zuschreibbares 118
- Risikofaktor 139, 490
- Risikofunktion (Hazardfunktion) 610
 - Graphik nach Weibull-Verteilung 621
 - kumulierte 622
- Robuste lineare Regression 97
- Robuste Mittelwerte 69, 70
- Robuste Verfahren 343
- Robustheit 242
- ROC - Analyse 136
- Rosenbaumsche Schnelltests 410
- RSS, Residual Sum of Squares 567
- Rückschluss und direkter Schluss 267
- Rückwärts-Elimination 574
- Rundungsfehler 33, 34
- Rundungsregeln 32
- Ryan-Lückentest 514
- SAR-Verteilung 439
- Satz von Glivenko und Cantelli 150
- Schadefekt 142
- Schätzfunktion 239–241
 - aus normalverteilten Grundgesamtheiten 202
 - Beispiele 242
 - Eigenschaften 241
- Schätzwert (estimator) 240, 241
 - Eigenschaften 242
 - einer Verteilung 162
 - für σ^2 , mit Beispiel 278
 - Parameter der Weibull-Verteilung 284
- Scharparameter 150
- Scheffé, lineare Kontraste; Beispiele (auch in R) 431–436
- Schichten 238
- Schichtenbildung 471
- Schiefe (skewness) 154, 160, 328
- Schließende (beurteilende) Statistik 8
- Schlussfolgerungen 11, 307
- Schlussziffernauswahl 239
- Schmerzintensität: Skalierung 63
- Schnelltests 326
- Schnittmenge 22
- Schoenfeld-Residuen
 - Cox-Regression 631
- Schranken der Studentisierten Maximum Modulus Verteilung 293
- Schrankenwert 1, 96 198
- Schwankungsintervalle, zentrale 196, 197
- Schwerpunkt der Punktwolke $(\bar{x}, \bar{y})$ 93
- Scoring I (Homogenitätstest) 511
- Scoring II (Homogenitätstest) 522
- Selektionseffekte 12
- Selektionskorrelation 238
- Sensitivität 132
- Sequentiell und simultan verwerfende Bonferroni-Prozedur 430
- Shapiro-Wilk Test 341
- Sheppard-Korrektur 74, 158
- Siegel-Tukey-Test 368, 371
 - Beispiel 373
 - Funktion in R 374
 - kritische Werte für R_1 373
- Sigma-Bereiche einer $N(\mu; \sigma)$ 196, 197
- Signifikante Ziffern 32
- Signifikanz
 - Begriff 309
 - Niveau; Varianten im Fall von Mehrfachtests 429
 - Test, Hypothesentest 322
- Simes-Hochberg-Prozedur 430, 527
- Simultane Konfidenzintervalle 430
 - nach Tukey 582
- Simultane multiple Vergleiche 429
- Simultane Paarvergleiche
 - mit einer Kontrolle 529
 - nach Royen 528
- Singuläre Matrix 40
- Skalare 36
- Skalarprodukt 38
- Skalenarten 15–18
- Skalentransformationen 17
- Skalierung von Variablen 15
- SMM-Verteilung 293
- Spaltenvektor 35
- Spaltungsziffern 334
- Spannweite (Range R) 65
- Spearman'sche Rangkorrelation bei Bindungen 89
- Spezifität 132
- Sprache der Statistik 25
- SR-Verteilung 436, 437
- Stamm-Blatt Darstellung 83
- Stammbäume 48
- Stammfunktion 47
- Standardabweichung
 - einer Zufallsvariablen (σ) 152
 - empirische (s) 70

- Standardisierte Extremabweichungen 346
- Standardisierte Messreihen 72
- Standardisierungen; Beispiele 141
- Standardized Mortality Ratio (SMR) 142
- Standardnormalverteilung $N(0; 1)$ 194
 - Zusammenhang mit anderen Verteilungen 221
 - zweidimensionale 232
- Standardschätzfehler 288
- Standardverfahren der Beurteilenden Statistik 6
- Statistik: Aufgaben, Definition und Umfeld 1, 108
- Statistisch prüfbare Hypothesen 2
- Statistische Maßzahlen 162
- Statistische Methoden 6
- Steigung (slope) 43, 91
- Steilheit, Wölbung, Kurtosis 330
- Stepwise Regression Modelling 573
- Sterbefunktion 610
- Sterbetafel 118
- Sterbeziffern, standardisierte 141
- Sternsymbolik 323
- Stetige Gleichverteilung 190
- Stetige Zufallsvariable 145, 148
- Stichprobe 8
 - repräsentative 235
- Stichproben
 - Funktionen, Verteilung von 222
 - Korrelationskoeffizient (r) 544
 - Verfahren 12, 238
 - Verteilung; knappe Beschreibung 161
 - Werte zufällig? 362
 - Ziehen ohne Zurücklegen 257
- Extremwerte 301
 - paarige 387
 - Umfang und Test 308
 - Umfänge zum Zweistichproben- t -Test 386
- Stirlingsche Formel 176
- Stochastische
 - Abhängigkeit 121
 - Unabhängigkeit 120, 520, 543
 - für n Ereignisse 122
- Stochastische (zufallsbedingte) Experimente 2
- Strata 238
- Streuung 7
- Streuungsvergleich anhand zweier Stichproben 370
- Student t -Verteilung 211
- Studien, explorative 322
- Stutzen 70
- Störfall-Kontrolle 366
- Störgrößen 14
- Stückzeit, durchschnittliche 79
- Suffizienz 242
- Sukzessive Differenzenstreuung 360
- Summe der Abweichungsquadrate 567
- Summen, spezielle 27
- Summenhäufigkeitsprozente 330
- Summenprozentlinie 331
- Supermarkt-Kunden-Beispiel 225, 227, 229
- Survival Analysis 609
- Symbolik für Prüfgrößen-Schranken; geändert im Vergleich zu älteren Auflagen 216
- Systematische Fehler 2, 13, 471
- Systematische Stichprobe 238
- t -Test 378
 - Paardifferenzen 387, 389–390
 - unabhängige Stichproben 377
 - ungleiche Varianzen ($\sigma_1^2 \neq \sigma_2^2$) 382
 - Untergruppen 381
- t -Verteilung 211–213
 - Parameter 214
 - Schranken für die 2- und die 1-seitige Fragestellung 214
 - Wahrscheinlichkeitsdichte 212
- t -Werte 213
- Tabellen 60
 - r Zeilen und c Spalten 60
 - Matrix-Struktur in R 60
- Tee-Test-Experiment 305
- Teilmengen 52
- Terrorismus im Flugverkehr 131
- Test
 - multipler 428
 - auf Bioäquivalenz 416
 - auf Normalverteilung 327
 - ein- bzw. zweiseitig 310
 - ein- oder zweiseitige Fragestellung 316, 317
 - konservativer 309
 - kritische Einschätzung 308–309
 - multipler 431
 - nach Mantel-Haenszel 503
 - statistischer (Prüfgröße) 306, 310
 - und Stichprobenumfang 308
 - verteilungsunabhängiger 317
 - Voraussetzungen erfüllt? 318, 322
- Testentscheidung 311
- Testkriterien 314
- Teststatistik (Prüfgröße) 306, 307
- Teststärke (Power) 311, 315

- Teststärkekurven (Gütefunktionen) 317
- Testverteilungen (Prüfgrößen) 210–218
- Testwiederholung 388
- Theorie wiederholbarer Ereignisse 11
- Therapie-Effekt 480
- Therapievergleich anhand des Proportional-Hazard-Modells (Beispiel in R) 623
- Toleranzfaktoren 300
- Toleranzgrenzen 299
- Toleranzgrenzen, verteilungsunabhängige 301
- Totale Wahrscheinlichkeit 128
- Transformation
 - linearisierende 104, 105
 - logistische 586
 - standardisierende 72
- Transponierte einer Matrix 35
- Trefferwahrscheinlichkeiten 125
- Trend 360, 365
- Trendtest
 - nach Jonckheere 452
 - nach Page 462
- Trennscharfer Test 316
- Trennschärfe 315
- Trennwert 136
- Treppenfunktion 145
 - Kaplan-Meier Schätzung 615, 618
- Trigonometrische Funktionen 45
- Trugschlüsse 238
- Trugschlüsse beim Vierfeldertest 487
- Tschebyscheff, P.L.: Ungleichung 154
- Tukey's five numbers 161
- Tukey-Kramer-Vergleiche 436
 - Beispiel (auch in R) 438
- Tumoren der Lunge 92
- U-Test 391
 - bei Rangaufteilung 398
 - Beispiele (auch mit R) 396–398
 - Bemerkungen 400
 - kritische Werte 393, 396
 - Voraussetzungen und Prinzip 391, 392
- Überdeckungswahrscheinlichkeit 253
- Übereinstimmung
 - noch zufällig? 537, 538
 - von Messwerten 302
- Überkreuzversuch (Cross-Over Design) 503
- Überlebende im Alter x 119
- Überlebensfunktion 610, 622
 - exponentielles Modell 620
 - Graphik nach Weibull-Verteilung 621
 - nach Kaplan-Meier geschätzt 611
- Überlebenszeit
 - durch Regressionsmodelle angenähert 621
- Logrank-Test 616
- mediane 615
- Medianwert im exponentiellen Modell 619
- mittlere 610, 615
- nach Chemotherapie (Beispiel in R) 613
- parametrische Modelle 618
- Weibull-Verteilung (Beispiel in R) 620
- Überlebenszeitanalyse 609
- Überschreitungswahrscheinlichkeit 322, 323
- Unabhängigkeit und Mosaikplot 62
- Unabhängigkeit von Ereignissen 120
- Unabhängigkeitstest für eine Kontingenztafel 600
- Ungleichung nach/von
 - Barrow und Bernoulli 165
 - Bienaymé und Tschebyscheff 203
 - Bonferroni 114, 123, 429
 - Cauchy für Mittelwerte 80
 - Tschebyscheff 153
- Untergruppen-Effekt 333
- Untergruppen-t-Test 381
- Untergruppenbildung (Stratifizierung) 332
- Unvereinbarkeit und stochastische Unabhängigkeit 123
- Unverfälschter Test 314
- Unvollständige faktorielle Experimente 473
- Urnenmodell 161, 181, 186
- Ursache 491
- $\text{Var}(Z)$, Varianz von Z 202
- Variabilität 7
- Variabilität der zentralen Tendenz 365
- Variabilitätskoeffizient 205
- Variabilitätstest nach Rosenbaum 410
- Variablen 14
- Variablen-Auswahl
 - Regressionsmodell 573, 595
 - Verfahren zur Modellbildung 573
- Varianz (von Zufallsvariablen) 152, 153
- Varianz, empirische (s^2) 71
 - Vergleich mit ihrem Parameter 359–360
- Varianz, gewogene s_{gew}^2 74
- Varianzanalyse
 - Beispiele (auch in R) 425–427
 - Einführung 423–425
 - für Messwiederholungen 454
 - im linearen Modell 577
 - wie viele Beobachtungen werden benötigt? 427
 - zweifach 585
 - zweifache 465–470

- Varianzanalytische Methoden 418
 Varianzkomponenten im linearen Modell 580
 mit zwei Faktoren 583
 Variation zweier Zufallsvariablen 230
 Variationskoeffizient 72
 γ , KI mit Beispiel 279
 für die Regression 288
 relativer V_r 73
 VB, s.u. KI 252
 Venn-Diagramm 110
 Verallgemeinerung 235, 324
 Vereinigung von Mengen 22
 Vergleich
 einer empirischen Varianz mit ihrem Parameter 359
 geordneter P-Werte nach Holm und nach Hochberg (Beispiel in R) 431
 mehrerer Mittelwerte 418
 mehrerer Varianzen 418–423
 Übersicht 421
 mit einer Kontrolle nach Dunnett 440
 unabhängiger Stichproben nach Nemenyi 449
 zweier χ^2 -Werte aus Tafeln mit gleichem FG 532
 zweier relativer Häufigkeiten 477
 zweier Varianzen (F-Test) 367
 Vergleich dreier Antibiotika
 ANOVA-Modell 580
 Beispiel in R 577
 multiple Vergleiche nach Tukey 581
 Parametrisierung 578, 579
 zweifaktoriell (Beispiel in R) 582
 Verhältnisskala 16
 Verhältniszahlen 57, 58
 Verhältniszahlen, Schätzung von 270
 Verknüpfungen zwischen Ereignissen 111
 Verschiebungssatz von Steiner 242
 Verschlüsselung 19
 Versuchsanordnungen 473
 Versuchsplanung, Grundprinzipien 470
 Versuchspläne, fünf Ansätze 473–474
 Verteilung
 der Differenz von Stichproben-Mittelwerten 223
 der Stichprobenvarianz 222
 der Studentisierten Extremwerte (SR) 436
 des „Studentized Augmented Range“ (SAR) 439
 des Quotienten von Stichproben-Varianzen 224
 des Stichprobenmittelwertes 222
 linkssteil oder rechtssteil? 155
 unterdispers oder überdispers? 179
 zweidimensionaler Zufallsvariablen 224
 Verteilungen
 wie sie zusammenhängen 221
 Verteilungsanteile ($\mu \pm 3\sigma$) für unterschiedliche Verteilungstypen 204
 Verteilungsenden, stark besetzt 332
 Verteilungsfreier Test 317
 Verteilungsfunktion 145, 146, 149
 empirische 150
 Rechenregeln 149
 Verteilungsfunktion, empirische 82, 150
 Verteilungshypothesen 307
 Verteilungsunabhängige Toleranzgrenzen 301, 303
 Verteilungsunabhängige Verfahren 325, 326
 Verteilungsunabhängiger Test 317
 Vertrauensbereich, Konfidenzintervall 241, 252
 Vertrauensgrenze (confidence limit) 265
 bei Sensitivitäten und Spezifitäten kleiner als 100% 259
 für π 258, 259
 für den Median 404
 für Null- und Vollergebnisse 259
 obere für λ 283
 Vertrauenswahrscheinlichkeit 264
 Verursachungszahlen 57
 Verzerrung (Bias) 244
 Vierfelder-Chiquadrat-Test 481
 Beispiel in R 483
 kritische Schranken 482
 minimaler Stichprobenumfang 485
 Vierfeldertafel 477, 479
 kollabierte 527
 Kombination mehrerer Tafeln 505, 506
 und bedingte Wahrscheinlichkeiten 489
 Vierfeldertest
 H_0 (zwei Varianten) 481
 H_0 und H_A 480
 Vollerhebungen 13
 Vollständige Randomisierung 473
 Voraussage, inverse aus einer linearen Regression 296–297
 Voraussagebereich (Regression) für eine zukünftige Beobachtung Y an der Stelle $X = x$ 293
 Voraussagewert eines diagnostischen Tests 132, 134, 135
 Voraussetzungen eines Tests erfüllt? 318
 Vorhersage (Prädiktion) 594

- Vorhersagen 560
- Vortests 367
- Vorwärts-Einschluss 574
- Vorzeichen-Rang-Test von Wilcoxon 400
- Vorzeichen-Trendtest von Cox und Stuart 364
- Vorzeichentest 404
 - Schnellschätzung 405
 - Schranken 273
- Vorzeichentest von Dixon und Mood 404

- Wachstum, exponentielles 77
- Wachstumserscheinungen 76
- Wahrscheinlichkeit 112
 - Axiome 113
 - bedingte 116
 - Definition nach Laplace 111
 - und Odds 113
- Wahrscheinlichkeits-Plot (probability plot) 331
- Wahrscheinlichkeitsansteckung 179
- Wahrscheinlichkeitsaussagen 4
- Wahrscheinlichkeitsdichte 147
 - Normalverteilung 192
- Wahrscheinlichkeitselement 147
- Wahrscheinlichkeitsfunktion 146, 147
- Wahrscheinlichkeitsnetz 330
- Wahrscheinlichkeitsrechnung 108
- Wald-Statistik 589
 - Cox-Regression 627
- Wechselwirkungen (Interaktionen) 474, 584
- Wechselwirkungseffekt 468
- Weibull Accelerated Life Model 622
- Weibull-Diagramm 285
- Weibull-Gerade 285
- Weibull-Verteilung 209–210, 616
 - Beispiel 209
 - Beispiel in R 284
 - Parameter 209
 - Schätzung beider Parameter 284
 - Überlebenszeit 619
- Welch-Test 382
- Wettchancen (odds) 113
- Wiederholbare Erfahrungen 2
- Wiederholbarkeit der Zufallsstichprobe 162
- Wiederholung 324, 472
- Wilcoxon
 - Einstichproben-Mediantest 358–359
 - Paardifferenzentest 400–403
 - Kritische Werte 402
 - Rangsummentest 369, 391–398
 - Verteilung 393–395
- Wildlife Tracking 188
- Wilson-Hilferty-Approximation 215

- Winkeltransformation, Normalisierung durch 380
- Winsorisieren 70
- Wissenschaft 7, 8
- Wissenschaftliche Arbeitstechnik 3
- Wölbung (kurtosis) 155, 328, 330
- Würfel-Modell 144–146
- Würfelmodell: Erwartungswert 151
- Wurzelrechnung 29

- x-Koordinate (Abszisse) 43

- y-Koordinate (Ordinate) 43
- Yates-Korrektur 483

- Z, Zufallsvariable, die standardnormalverteilt ist 195
- Z-Intervalle 197
- z-Punkt ($\hat{z}$)-Transformation
 - nach R.A. Fisher 548
 - weitere Anwendungen 549
- Zahlenlotto 2
- Zeilenvektor 35
- Zeitreihe 364
- Zeitstudien 205
- Zensierungsarten (zensiert=unbeobachtet) 612
- Zentrale Schwankungsintervalle 196, 197
- Zentraler Grenzwertsatz 192, 204
- Zerlegung der χ^2 -Statistik 513, 516
- Zerlegung der FG einer χ^2 -Statistik 512
- Zerlegung einer Menge 52
- Zielfunktion 97
- Zielgröße 14, 470, 560
 - dichotome 586
- Zielgrößenoptimierung 476
- Ziffern, signifikante 32
- Zufallsergebnisse 1, 14
- Zufallsexperiment 109
- Zufallsfehler 2
- Zufallskomponente (ϵ_i) 561
- Zufallskomponenten-Modell der zweifachen Varianzanalyse 468
- Zufallsstichprobe 2, 8, 13
 - aus definierter Grundgesamtheit 235
 - Kontrolle einer Datenfolge 362
- Zufallsvariable 144, 149, 240
 - 5 Eigenschaften 145
 - Realisierung 6
 - standardnormalverteilt 194
 - zweidimensionale 225
- Zufallszahlen 164, 235, 471
 - Eigenschaften und Anwendung 236

- Gewinnung mit R 237
- Tabelle 236, 237
- Zufallszuteilung 471
- Zufällige Fehler 471
- Zufälligkeit der Stichprobe 471
- Zunahme, prozentuale (Beispiel) 261, 262
- Zusammenfassen geeigneter Merkmalskombinationen 523
- Zusammenhang
 - funktionaler 543
 - kurvilinearer 106
 - linearer 87
- Zusammenhangsanalyse 88, 90
- Zuschreibbares Risiko 118, 489
- Zusätzlicher Wert 72
- Zwei-Würfel-Modell 224
- Zweidimensionale Normalverteilung 231, 232
- Zweidimensionale Zufallsvariablen
 - bedingte Dichten 228
 - bedingte Verteilungen und Unabhängigkeit 228
 - Randverteilungen und Unabhängigkeit 226
 - Satz von Bayes 229
- Zweifache Varianzanalyse 465
 - 4 *SAQ*-Anteile 465, 466
 - Modell I mit festen Effekten 466, 467
 - Modell II mit zufälligen Effekten 468, 469
 - Modelle I, II und III 469, 470
- Zweifaktorielle Varianzanalyse 581
- Zweistichproben
 - „Schnelltest“ nach Tukey 412–413
 - Fisher-Test 486
 - Permutationstest 412
 - t-Test 377
 - Stichprobenumfänge 386
 - weitere Details 378
 - wie viele Beobachtungen werden benötigt? 385–387
 - Test auf Äquivalenz 414
 - Beispiel in R 415
 - Test bei starken Verteilungsformunterschieden 414

Hinweise zum Programm R: Eine kompakte Einführung in die Verwendung des Programmes **R** gibt das Kapitel 9. Wichtige Funktionen sind hier näher beschrieben. Das **Le-sezeichen** auf der rechten Seite kann herausgetrennt werden und dient somit als knappe Referenz zu häufig verwendeten Befehlen und Funktionen. Ausführliche Hilfestellung bietet das Programm **R** selbst bzw. es muss im Internet nachgelesen werden. Insbesondere wird auch an dieser Stelle darauf hingewiesen, dass die **R**-Befehle zu allen in diesem Buch aufgeführten **R**-Beispielen im Internet auf der Produktseite des Buches bei Springer abrufbar sind!

Hinweise zur Lektüre: Folgende Hinweise sollen Ihnen helfen, den Inhalt des Buches einfacher zu erschließen. Namenverzeichnis und insbesondere das Sachverzeichnis werden Sie hierbei unterstützen.

Was lesen?

- 1. Verschaffen Sie sich einen ersten Einblick, indem Sie sich das Inhaltsverzeichnis ansehen und zunächst die Einführung in die Statistik (Kapitel 1) lesen. Besonders wichtig ist dabei auch ein Blick in das Vorwort.
- 2. Als Anfänger, zur Wiederholung oder auch zum Schließen von Wissenslücken in den Grundlagen sollten Sie dann mit der Lektüre des 2. Kapitels beginnen und die entsprechenden Abschnitte je nach Bedarf überfliegen oder durcharbeiten.

Wie lesen?

- 3. Bevor Sie einzelne Abschnitte durcharbeiten: (a) Vergewissern Sie sich anhand des Inhaltsverzeichnisses über Ihren Standort. (b) Überfliegen Sie die Textstellen der interessierenden Seiten; betrachten Sie aufmerksam die Überschriften, einige Formeln, die Abbildungen und Tabellen, und lesen Sie sorgfältig die Legenden zu den Abbildungen und Tabellen sowie die Einführungen und Zusammenfassungen. (c) Bemühen Sie sich, möglichst viele Fragen zu stellen, etwa nach der Bedeutung der Überschriften, der fett- oder kursivgedruckten Satzteile, der Abbildungen, der Tabellen, der Formeln, den hiermit zusammenhängenden Fragen und dem eigenen bereits vorhandenen Wissen darüber. Hierdurch wird Ihre Aufmerksamkeit geweckt (auch wenn ein Teil des Stoffes langweilig ist), Ihr Unterbewusstsein an bereits Bekanntes erinnert, und Sie erhalten eine Vorstellung von dem Umfang, dem Zeitbedarf und dem Zweck der Lektüre.
- 4. Die jetzt folgende gründliche Lektüre dient zur Beantwortung dieser Fragen. Neben den Abbildungen und Tabellen ist den Formeln besondere Aufmerksamkeit zu widmen; das Kleingedruckte und die Hinweise können bei der Erstlektüre überschlagen werden. Entscheidend wichtig sind Leseпаusen, in denen Sie das Gelesene mit eigenen Worten wiedergeben.
Im Text zitierte Arbeiten sind durch Autorennamen und in Klammern folgendem Literaturhinweis charakterisiert. Ein Blick auf das Literaturverzeichnis zeigt Ihnen, wo der Zeitschriftenaufsatz bzw. das Buch zitiert ist. Einige der zitierten Literaturstellen sollten Sie gelegentlich selbst einsehen. Bibliothekskataloge und das Internet weisen Ihnen den Weg.
- 5. Bei der zusammenfassenden Wiederholung: (a) wenige bedeutsame Begriffe markieren, (b) Randbemerkungen machen, (c) wichtige Aussagen wie z. B. Definitionen und ausgewählte Formeln notieren, (d) sich und an das Buch Fragen stellen, (e) wesentliche Teile des Gelesenen mit eigenen Worten wiedergeben.

Verschiedenes	
<code>q()</code> <code><-</code>	Beenden des Programms Zuweisung von Werten, Variablen; auch das Gleichheitszeichen ist möglich
<code>mat[,2]</code>	Auswahl von Daten über den Index; hier die 2. Spalte einer Matrix <i>mat</i>
<code>frame \$ var</code>	Bezeichnung für eine Variable <i>var</i> in einem Datenrahmen <i>frame</i>
NA	festе Bezeichnung für fehlende Angaben
<code>is.na (obj)</code> <code>is.numeric (obj)</code> <code>is.matrix (obj)</code> <code>library(MASS)</code>	logische Abfrage auf fehlende Angaben logische Abfrage auf den Datentyp logische Abfrage auf Objekttyp einbinden zusätzlicher Programmbibliotheken (Pakete)
Hilfestellung in R	
<code>help.start ()</code> <code>?cmd</code> <code>apropos("name")</code> <code>help.search("string")</code> <code>library(help=MASS)</code> <code>example(cmd)</code>	Start des Hilfesystems über ein entsprechendes Browser-Programm Hilfe zu einzelnen Befehlen oder Funktionen in R Suche in dem Hilfesystem von R nach Namen oder Zeichenketten Hilfe zur Programmbibliothek <i>MASS</i> Beispiele zu Funktionen und Befehlen
Eingabe und Ausgabe	
<code>source(file)</code> <code>read.table(file)</code> <code>data.entry (frame)</code> <code>edit (frame)</code> <code>vect <- scan()</code> <code>sink(file)</code> <code>sink ()</code> <code>write(obj, "file")</code> <code>write.table(obj, "file")</code>	Ausführen der Befehle aus einer Datei Einlesen von Daten aus einer Datei Dateneingabe und -korrektur in einem Tabellenschema Einlesen von Daten in einen Vektor Ausgabe in eine Datei beenden der Ausgabe in eine Datei Ausgabe eines Objektes in eine Datei Ausgabe einer Tabelle
Variablen und Objekte	
<code>attach(obj)</code> <code>detach(obj)</code> <code>ls()</code> <code>rm(obj)</code> <code>dim(mat)</code> <code>dimnames(mat)</code> <code>length(vect)</code> <code>1 : n</code> <code>seq (from, to, by=)</code> <code>rep(x, n)</code> <code>c(1, 2, 3)</code> <code>cbind(u, v, w)</code> <code>rbind(u, v, w)</code> <code>matrix(data, nrow=n, byrow=TRUE)</code> <code>data.frame(vector list)</code> <code>as.factor()</code> <code>as.matrix()</code> <code>as.data.frame()</code> <code>t(mat)</code> <code>which(x==a)</code>	Aufnehmen eines Objektes in den aktuellen Suchpfad in R Löschen eines Eintrags aus dem Suchpfad Liste aller zur Zeit aktiven Objekte Entfernen eines Objektes aus dem Arbeitsspeicher Dimensionen einer Matrix Namen zu Dimensionen einer Matrix Anzahl der Elemente in einem Vektor erzeugt den Vektor 1, 2, ..., <i>n</i> erzeugt eine Zahlenfolge von ... bis ... in einer festen Schrittweite wiederholt den Vektor <i>x</i> <i>n</i> -mal verbindet die Werte 1, 2, 3 zu einem Vektor verbindet die Vektoren <i>u</i> , <i>v</i> , <i>w</i> spaltenweise zu einer Matrix verbindet die Vektoren <i>u</i> , <i>v</i> , <i>w</i> zeilenweise zu einer Matrix erzeugt eine Matrix mit <i>n</i> Zeilen aus den Werten eines Vektors zeilenweise erstellt einen Datenrahmen aus einer Anzahl von Vektoren gleicher Länge Umwandlung in einen Faktor Umwandlung in eine Matrix Umwandlung in einen Datenrahmen transponieren einer Matrix; vertauschen von Zeilen und Spalten liefert den Index von <i>x</i> für die <i>x==a</i> zutrifft

Erklärungen zu den Parametern der Funktionen müssen über das Hilfesystem von **R** nachgelesen werden!

Ablaufsteuerung	
for (<i>i in vector</i>) { }	Wiederholung der folgenden Befehle
while (<i>cond</i>) { }	bedingte Befehlsverarbeitung
if (<i>cond</i>) { } else { }	bedingte Befehlsverarbeitung
function (<i>arg</i>) { }	Definition von Funktionen
break	Abbrechen der Befehlsverarbeitung
return (list())	Rückgabe von Argumenten; in der Regel als Liste (auch im letzten Befehl)
Arithmetik	
+, -, *, /, ^	Grundrechenarten: Addition, Subtraktion, Multiplikation, Division, Potenz
%*%	Matrixmultiplikation
ausgewählte Statistikfunktionen	
max (<i>v</i>), min (<i>v</i>)	Maximum, Minimum für einen Vektor
mean (<i>v</i>), median (<i>v</i>)	Mittelwert, Medianwert aus einem Vektor
sum (<i>v</i>), prod (<i>v</i>)	Summe, Produkt der Elemente eines Vektors
sd (<i>v</i>), var (<i>v</i>)	Standardabweichung, Varianz aus einem Vektor
rank (<i>v</i>) , sort (<i>v</i>)	Rangzahlen, Sortierung
summary (<i>frame</i>)	beschreibende Statistik zu den Variablen des Datenrahmens
apply (<i>x</i> , <i>n</i> , <i>function</i>)	anwenden einer Funktion auf mehrere Objekte
tabulate (<i>bin</i>)	bestimmt die Häufigkeiten (Verteilung) in einem Vektor (integer)
table (<i>A</i> , <i>B</i>)	erzeugen einer Häufigkeitstabelle
xtabs (<i>A</i> , <i>B</i>)	aus den Faktoren <i>A</i> und <i>B</i>
Verteilungsmodelle	
Mit dem ersten Buchstaben werden jeweils verschiedene Funktionstypen festgelegt: p - Verteilungsfunktion, d - Dichtefunktion, q - Quantilfunktion und r - Zufallszahlen.	
pbinom ()	Binomialverteilung
phyper ()	hypergeometrische Verteilung
ppois ()	Poisson-Verteilung
pnorm ()	Normalverteilung
pt ()	Student-Verteilung (t-Verteilung)
pf ()	Fisher-Verteilung
pchisq ()	Chiquadrat-Verteilung
ausgewählte statische Verfahren	
aov () , anova ()	Varianzanalyse
lm () , glm ()	lineare und verallg. lineare Modelle
t.test ()	t-Test (unabhängige bzw. verbundene Stichproben)
wilcox.test ()	Rangsummen-Test, Paardifferenzen-test
prop.test ()	Test für relative Häufigkeiten
binom.test ()	Binomial-Test
chisq.test ()	Chiquadrat-Test
fisher.test ()	exakter Test nach Fisher (Vierfelder)
cor ()	berechnet Korrelationskoeffizienten
cor.test ()	Testverf. zu Korrelationskoeffizienten
friedman.test ()	Friedman-Test
ausgewählte Graphiken (high level)	
par ()	Festlegung von allgemeinen Graphikparametern, vgl. ?par
plot ()	Basisbefehl für zahlreiche Darstellungen, insb. Punktwolken und Kurven
matplot ()	Plot zu Spalten aus Matrizen
pairs ()	Matrix mit Punktwolken
pie ()	Tortendiagramm
barplot ()	Balkendiagramm
boxplot ()	Box-Plot
stripchart ()	Punktplot bei kleiner Fallzahl
mosaicplot ()	Mosaikplot
hist ()	Histogramm
qqplot ()	Quantil-Plot