ERWIN KREYSZIG

Statistische Methoden und ihre Anwendungen

7. AUFLAGE

VANDENHOECK & RUPRECHT

Herrn Professor
Dr. FERDINAND WEINHANDL †
zum Gedächtnis

CIP-Kurztitelaufnahme der Deutschen Bibliothek

Kreyszig, Erwin:
Statistische Methoden und ihre Anwendungen / von Erwin Kreyszig. — Unveränd. Nachdr. d. 7. Aufl. — Göttingen: Vandenhoeck und Ruprecht, 1982.

ISBN 3-525-40717-3

1. Auflage 1965
2. Auflage 1967
3. Auflage 1968
(1. Nachdruck 1969)
(2. Nachdruck 1969)
(3. Nachdruck 1970)
(4. Nachdruck 1972)
4. Auflage 1973
(Nachdruck 1974)
5. Auflage 1975
6. Auflage 1977
7. Auflage 1979
(Nachdruck 1982)

Druck: Hubert & Co., Göttingen

Vorwort

zur 1. bis 7. Auflage

Die Bedeutung der modernen Statistik in der Wissenschaft, Technik, Wirtschaft und Politik wächst ständig. Das hat zwei Hauptgründe: 1. Die Statistik leistet heute mehr als früher. Früher mußte sie sich auf eine bloße empirische Beschreibung beschränken. Heute kann sie der Beurteilung von Situationen dienen, weil ihre mathematischen Grundlagen entsprechend entwickelt worden sind. 2. Es besteht dringender Bedarf für eine solche mathematisch fundierte Statistik, denn die Lebens- und Produktionsgemeinschaften werden immer größer und komplexer, die auftretenden Probleme immer komplizierter und die „guten alten“ empirischen Verfahren demgemäß immer unzulänglicher.

Bekanntlich wird die mathematische Statistik in den deutschsprachigen Ländern noch immer nicht genügend beachtet und angewendet. Um den Vorsprung des Auslandes wettzumachen, ist es notwendig, das Interesse an der mathematischen Statistik und das Verständnis für deren Wesen *in breiten Kreisen* zu wecken. Hierzu soll die vorliegende Einführung einen Beitrag leisten. Diese ist aus der praktischen statistischen Arbeit und aus Vorlesungen entstanden, die ich in Aachen, Columbus, Graz, Karlsruhe, Ottawa und Windsor für Hörer verschiedener Interessengebiete (Technik, Mathematik, Physik, Medizin, Volkswirtschaft usw.) gehalten habe.

Der Stoff ist möglichst einfach dargestellt. Die für das Verständnis notwendigen Vorkenntnisse gehen nicht über die elementare Differential- und Integralrechnung hinaus. Auf die Hilfsmittel der Maß- und Integrationstheorie wird also verzichtet. Allgemein verständliche praktische Probleme stehen im Vordergrund. Alle Zahlenwerte entstammen tatsächlichen Beobachtungen, so daß der Leser auf Schritt und Tritt sozusagen „den Zufall bei der Arbeit sehen kann“. Nur so erwirbt man sich ein tieferes Verständnis. Um zu eigenen Versuchen anzuregen, ist ein Teil der Beispiele besonders einfach gehalten. Eine enge Verbindung zwischen der Theorie und den Anwendungen wird angestrebt. Es kommt mir nicht darauf an, möglichst „viel“ zu bringen, sondern den Leser mit der typischen Denkweise des behandelten Gebietes vertraut zu machen.

Das Buch hat drei Teile:

I. Beschreibende Statistik,
II. Wahrscheinlichkeitstheorie,
III. Beurteilende Statistik.

Teil III. Beurteilende Statistik

Im ersten Teil befassen wir uns mit der graphischen und tabellarischen Darstellung und Beschreibung gegebener statistischer Daten, im zweiten Teil mit der wahrscheinlichkeitstheoretischen Grundlage der Methoden zur mathematischen Auswertung statistischer Daten und im dritten Teil mit dieser Auswertung selbst.

Wir behandeln also zunächst die einfachere Aufgabe der übersichtlichen Darstellung statistischen Materials und entwickeln erst später mathematische Modellvorstellungen, auf die sich die Beurteilung statistischer Daten gründet.

Dieser Weg soll es dem Anfänger ermöglichen, sich zuerst einmal in die Gedankenwelt der Statistik etwas einzufühlen, bevor er zur Anwendung mathematischer Hilfsmittel übergeht.

Praktische Übung ist in der Statistik wichtiger als in vielen anderen Gebieten der Mathematik. Aus diesem Grunde bilden die Aufgaben am Schluß der Abschnitte einen wesentlichen Teil des Buches. Sie sollen dem Leser auch eine Ahnung davon verschaffen, wie weit und vielseitig die Anwendungsmöglichkeiten der mathematischen Statistik sind. Einen Teil der Lösungen findet man in Anhang 2.

Englische Fachausdrücke sind in Anhang 4 angegeben, um den Zugang zu der umfangreichen englischsprachigen statistischen Literatur zu erleichtern.

Herrn Prof. Dr. F. REUTTER (Aachen) danke ich herzlich für sein Interesse am Entstehen des Buches und für die durch sein Institut geleistete Hilfe bei der Herstellung des Manuskriptes, den Herren Prof. Dr. K. STANGE (Aachen), Dr. H. HILDEN (Aachen), Dipl.-Math. W. SCHULZ (Frankfurt), Dr. W. VOLK (Darmstadt), Prof. Dr. H. FLORIAN (Graz), Dr. J. GÖLLES (Graz) und Akad. Oberrat Dr. H. KÖLLING (Göttingen) für wertvolle Hinweise und Diskussionen, dem Verlag und der Druckerei für die angenehme Zusammenarbeit und die gefällige Ausstattung des Buches.

Die vorliegende 7. Auflage wurde erfreulicherweise schon jetzt notwendig. Wie ihre Vorgängerin enthält sie eine Einführung in die **Entscheidungstheorie** (Kap. 21), ein modernes Gebiet der Statistik, das für die Anwendungen zunehmendes Interesse besitzt. Ferner wurde Anhang 1 erweitert, insbesondere durch Bemerkungen zur Mengenlehre im Zusammenhang mit der Wahrscheinlichkeitstheorie.

Für Leserzuschriften danke ich sehr. Für weitere Hinweise werde ich ebenfalls jederzeit sehr dankbar sein.

Erwin Kreyszig

Inhalt

Wichtige Bezeichnungen

(Formelzusammenstellung siehe Anhang 6)

		Abschnitt
b	Regressionskoeffizient der Stichprobe	94
β	Regressionskoeffizient der Grundgesamtheit	97
E	Erwartung	28, 30, 58
$\tilde{f}$	Häufigkeitsfunktion	4
f	Wahrscheinlichkeitsfunktion bzw. Dichte	22, 25, 26
$\tilde{F}$	Summenhäufigkeitsfunktion	7
F	Verteilungsfunktion	24
h	Relative Häufigkeit	4
μ	Mittelwert der Grundgesamtheit	28
n	Stichprobenumfang	4
P	Wahrscheinlichkeit	14, 15
p	Wahrscheinlichkeit bei der Binomialverteilung	39
r	Korrelationskoeffizient der Stichprobe	106
ϱ	Korrelationskoeffizient der Grundgesamtheit	107
s	Standardabweichung der Stichprobe	8
σ	Standardabweichung der Grundgesamtheit	29
s^2	Varianz der Stichprobe	8
σ^2	Varianz der Grundgesamtheit	29
X, Y	Zufallsvariable	21, 52
$x_1, x_2, \ldots$	Stichprobenwerte	8
$\bar{x}$	Mittelwert der Stichprobe	8

Kapitel 1

Einleitung

Statistiken über Geburten, Verkehrsunfälle, Börsenkurse, Wahlresultate, Niederschläge, Geschäftsumsätze und viele andere Dinge sind uns aus dem täglichen Leben wohlbekannt. Wir wissen, daß es sich dabei jeweils um Zahlentabellen oder entsprechende graphische Darstellungen handelt. Daneben wird das Wort „Statistik" noch in einem anderen Sinne gebraucht: Mit diesem Wort oder genauer mit „**mathematischer Statistik**" bezeichnet man die Lehre von den mathematischen Methoden zur Gewinnung und Auswertung von Statistiken.

Im vorliegenden Kapitel wollen wir uns eine erste grobe Vorstellung vom Wesen und von der Aufgabe der mathematischen Statistik bilden und auf einige wichtige Anwendungsgebiete hinweisen.

1 Vorbemerkungen über das Wesen der mathematischen Statistik

In der mathematischen Statistik betrachtet man **Massenerscheinungen.** Das sind Erscheinungen, die entweder eine große Anzahl von Individuen betreffen oder eine große Menge von Einzelerscheinungen zusammenfassen. Dabei zeigt sich oft, daß die Massenerscheinungen gewisse Gesetzmäßigkeiten aufweisen, die sich nicht für die Einzelerscheinung formulieren lassen. Zwei einfache Beispiele mögen dies erläutern.

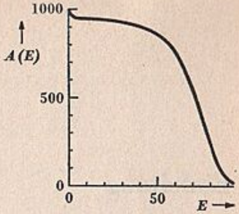
Über das Lebensalter, das eine bestimmte Person erreichen wird, kann man keine genauen Vorhersagen machen. Wohl aber lassen sich für eine Bevölkerungsgruppe, etwa für alle Einwohner eines Landes, auf Grund von Beobachtungen empirische Gesetzmäßigkeiten feststellen. Man kann z. B. ermitteln, wie viele von je 1000 Neugeborenen nach einer gewissen Anzahl von Jahren jeweils noch am Leben sind. Auf diese Weise erhält man eine sogenannte *Sterbetafel* (vgl. Abb. 1.1), die als Grundlage für die Berechnung der Lebens- und Altersversiche-

E	$A(E)$
5	955
10	952
15	950
20	944
25	935
30	927

E	$A(E)$
35	918
40	907
45	892
50	868
55	829
60	764

E	$A(E)$
65	673
70	554
75	408
80	249
85	111
90	31

Tabellarische Form



Graphische Form

Abbildung 1.1. Sterbetafel für männliche Personen in Deutschland
E = Erreichtes Alter (vollendetes Altersjahr); $A(E)$ = Anzahl der Personen, die von je 1000 männlichen Lebendgeborenen das Lebensalter E erreichen. (Nach Angaben für 1958/59 im Stat. Jahrbuch f. d. Bundesrep. Deutschland 1961, S. 66)

rungsprämien benutzt wird. Die durch eine solche Tafel ausgedrückte empirische Gesetzmäßigkeit bezieht sich also auf die Gesamtheit der betrachteten Personen, nicht aber auf eine bestimmte Einzelperson.

Ähnlich ist die Lage beim Würfeln: Welche Zahl der nächste Wurf bringen wird, kann man nicht vorhersagen. Wohl aber beobachtet man, daß in einer großen Anzahl von Würfelergebnissen die Zahlen 1, 2, 3, 4, 5, 6 jeweils *etwa* gleich oft vorkommen, sofern der Würfel aus homogenem Material besteht, eine geometrisch einwandfreie Form besitzt und ordnungsgemäß geworfen wird.

Bei unseren beiden Beispielen steht der „**zufälligen Unregelmäßigkeit**" der Einzelerscheinung eine „**statistische Regelmäßigkeit**" oder „**statistische Gesetzmäßigkeit**" der Massenerscheinung gegenüber.

Die Aufdeckung solcher Gesetzmäßigkeiten ist ein grundlegendes Problem. In diesem Zusammenhang fragt man sich: Auf welche Weise und in welchem Umfange sammelt man in einem konkreten Falle **statistische Daten,** d. h. Zahlenmaterial, aus dem man einzelne Besonderheiten und Eigenschaften von Massenerscheinungen ermitteln will? Wie stellt man diese Daten tabellarisch oder graphisch in übersichtlicher Weise zusammen? Wie wertet man die Daten aus? Welche Schlüsse lassen sich ziehen? Welchen Zuverlässigkeitsgrad besitzen die so gewonnenen Urteile über eine Massenerscheinung?

Diese und ähnliche Fragen gilt es zu beantworten. Nur in einfachen Fällen ist das ohne wesentliche mathematische Hilfsmittel möglich. Bei den weitaus meisten in der Praxis anfallenden Problemen muß man Verfahren der modernen mathematischen Statistik anwenden, um zu sinnvollen Urteilen zu gelangen und vernünftige Entscheidungen im Falle von Ungewißheiten zu treffen. Solche Methoden sind in den letzten Jahrzehnten entwickelt worden. Sie ermöglichen es, statistisches Zahlenmaterial auszuwerten, also von einer bloßen *Beschreibung* zu einer *Beurteilung* gegebener Situationen überzugehen. Der Charakter statistischer Überlegungen hat sich dadurch grundlegend gewandelt: An die Stelle der empirischen beschreibenden Statistik ist die wissenschaftlich fundierte mathematische Statistik getreten, deren praktische Bedeutung ständig wächst.

Schon aus der Antike sind uns Statistiken über den Staatshaushalt, über Pachteinkünfte und über Bevölkerungszahlen bekannt. Und man hat schon frühzeitig versucht, statistische Daten auszuwerten. Solange aber solche empirische Versuche keine mathematische Grundlage besaßen, war die Gefahr von Fehlschlüssen sehr groß. So entstand das Sprichwort, daß man „mit Hilfe von Statistiken alles beweisen kann“. Es kennzeichnet aber lediglich die Mißerfolge kritikloser Beurteilung statistischer Daten. (Interessante Beispiele findet man in [7], [14] und [15], siehe Anhang 3 am Schluß.)

Allerdings ist die mathematische Statistik kein Automat, in den man nur seinen Groschen hineinzustecken braucht, um sinnvolle Resultate zu erhalten. Vielmehr muß man sich in die Denkweise dieses Gebietes einleben, damit man die Anwendungsmöglichkeiten sehen lernt und in einem konkreten Fall das richtige Verfahren auswählen kann.

2 Einige Anwendungsgebiete der mathematischen Statistik

Die Beispiele und Aufgaben des Buches entstammen der Industrie und Technik, der Physik, Meteorologie, Geodäsie, Medizin, Botanik und Zoologie, Landwirtschaft, dem Verkehrswesen usw. Das sind einige Anwendungsgebiete der mathematischen Statistik. Viele andere ließen sich leicht aufzählen. Man denke etwa an den Arbeits- und Wohnungsmarkt, den Außenhandel und die Zollpolitik, die Jugenderziehung, das Versicherungs- und Fürsorgewesen, die angewandte Psychologie und die Sozialwissenschaften.

In jedem dieser Gebiete ist die Anwendungsmöglichkeit der mathematischen Statistik meist sehr vielgestaltig. Das zeigt schon ein Blick auf die genannten Aufgaben. Beispielsweise bedient man sich in der Industrie und Wirtschaft statistischer Methoden bei der Materialprüfung, Fertigungssteuerung und -regelung, End- und Abnahmekontrolle, Entwicklung, Personalauswahl, Markt- und Produktforschung, Wirtschaftsprüfung, Rationalisierung von Arbeitsverfahren, Planung von Versuchen, Kapitalanlagepolitik usw. Zahlreiche große Firmen haben in letzter Zeit für derartige Zwecke besondere statistische Abteilungen eingerichtet oder stehen wenigstens in enger Verbindung mit statistischen Büros, denen die auftretenden Probleme zur Bearbeitung übertragen werden.

Manche Probleme lassen sich in kurzer Zeit und mit einfachen Hilfsmitteln erledigen. Andere erfordern erhebliche Zeit und Kosten. Große Untersuchungen waren die englische Erforschung der Wirksamkeit von Streptomycin bei der Tuberkulosebehandlung im Jahre 1946 und eine amerikanische Prüfung, bei der 126 Fahrzeuge 2 Jahre lang 19 Stunden je Tag und 6 Tage je Woche über Versuchsstraßen fuhren. Dieser „A. A. S. H. O. Road Test" kostete 27 Millionen Dollar. Solche Versuche machen sich durch die Erfahrung, die man dabei gewinnt, reichlich bezahlt.

Bei größeren statistischen Vorhaben leisten elektronische Rechenmaschinen wertvolle Dienste, insbesondere, wenn der Zeitfaktor eine Rolle spielt.

3 Zur Arbeitsweise in der mathematischen Statistik

Wir haben gesehen, daß die mathematische Statistik in ganz verschiedenartigen Gebieten angewendet wird. Und in jedem Gebiet können ganz verschiedenartige Fragestellungen auftreten. Zum Beispiel kann man sich, wie die Aufgaben des Teiles III zeigen, in der Landwirtschaft für das günstigste Pflanzdatum von Baumwolle, den Stärkegehalt von Kartoffeln, die Brauchbarkeit von Sämaschinen, die Anzahl von Erdbakterien, den Heuertrag in Abhängigkeit von der Düngung, den Zusammenhang zwischen dem Humusgehalt und dem Stickstoffgehalt des Bodens interessieren. Oder in der Medizin kann man fragen nach der Wirksamkeit von Medikamenten und Heilverfahren, nach dem Einfluß des Rauchens auf den Lungenkrebs, des Autofahrens auf Wirbelsäulenschäden usw.

Trotz der großen Heterogenität der Probleme besteht deren statistische Bearbeitung im allgemeinen aus denselben Schritten. Wir illustrieren diese Schritte durch ein einfaches Beispiel. Dabei erwähnen wir einige wichtige Grundbegriffe, kümmern uns aber einstweilen noch nicht um deren präzise Fassung.

Ein Ingenieur einer Firma sei beauftragt, die Dichte des Kokses zu untersuchen, den die Firma verwendet.

1. Schritt. Formulierung des Problems. Dieser Schritt ist im vorliegenden Falle einfach: Der Ingenieur entschließt sich, die durchschnittliche Dichte des Kokses zu bestimmen. In anderen Fällen erfordert der Schritt meist mehr Überlegung, als man naiverweise denkt, weil man oftmals erst klare Begriffe schaffen und die Fragestellung präzisieren und so einengen muß, daß die Arbeit zu bewältigen ist. (Beispiel: Untersuchung über die Verbreitung der Geisteskrankheiten in einem Land. Wie weit oder wie eng faßt man da den Begriff der Geisteskrankheit? Soll man nur klinisch behandelte Fälle heranziehen oder auch andere, bei denen man genaue Unterlagen wohl wesentlich schwerer erhält? Soll man nach Geschlechtern trennen? Nach Altersgruppen? Nach Stadt- und Landbevölkerung? Wenn ja, wo liegt die Grenze zwischen diesen beiden Begriffen? Auf welchen Zeitraum soll man sich beschränken? Und so weiter.)

2. Schritt. Planung des Experimentes. Hätten alle Koksbrocken dieselbe Dichte, so brauchte der Ingenieur nur einen einzigen herzunehmen und dessen Dichte zu bestimmen. Dann käme die Statistik überhaupt nicht ins Spiel. Nun zeigt aber die Erfahrung, daß diese Dichte von Brocken zu Brocken variiert. Der Ingenieur sollte demgemäß nun eigentlich alle Koksstücke hernehmen und messen. Das würde aber mehr kosten, als der ganze Koks wert ist, und außerdem sehr viel Zeit beanspruchen. Ein Brocken ist zuwenig, alle sind zuviel. Der Ingenieur entschließt sich sozusagen für den goldenen Mittelweg: Er faßt den Plan, einige Brocken ganz zufällig herauszugreifen und deren Dichte zu bestimmen.

3. Schritt. Ausführung des Experimentes. Er führt seinen Plan durch. Der genannte Vorgang des zufälligen Auswählens und Messens wird in der Statistik als ein **Zufallsexperiment** oder eine **Zufallsbeobachtung** (kurz: **Experiment** oder **Beobachtung**) bezeichnet. Der Einfachheit halber nehmen wir an, der Ingenieur habe nur 5 Brocken ausgewählt und gemessen. Er habe dabei die Werte

$$1{,}40 \quad 1{,}45 \quad 1{,}39 \quad 1{,}44 \quad 1{,}38$$

[g/cm³] (die einem tatsächlich ausgeführten Experiment entstammen, vgl. Aufgabe 72.2) erhalten. Diese Werte werden als eine **Stichprobe** aus der „**Grundgesamtheit**" aller möglichen Meßwerte bezeichnet, die bei der Ausführung des Experimentes denkbar sind. Ihre Anzahl (5) heißt der *Umfang* der Stichprobe.

4. Schritt. Tabellierung und Beschreibung des experimentellen Ergebnisses. Berechnung von Maßzahlen. Bei umfangreichen Untersuchungen, bei denen viele Zahlenwerte anfallen, muß man nun dafür sorgen, daß man diese in eine übersichtliche tabellarische Form bringt und vielleicht sogar noch graphisch veranschaulicht. Wie man das macht, lernen wir in Teil I. Im Falle unserer einfachen kleinen Stichprobe ist nichts dergleichen zu tun. Der Ingenieur berechnet lediglich zwei Maßzahlen, nämlich die durchschnittliche Dichte

$$\bar{x} = \frac{1}{5}(1{,}40 + 1{,}45 + 1{,}39 + 1{,}44 + 1{,}38) = 1{,}412$$

und die sogenannte Varianz, die ein Maß dafür ist, wie sehr die Stichprobenwerte streuen. (Die Formel für die Varianz folgt später.)

5. Schritt. Schluß von der Stichprobe auf die Grundgesamtheit. Aus der Stichprobe schließt der Ingenieur, daß die Grundgesamtheit *etwa* die durchschnittliche Dichte 1,412 g/cm³ hat. Er kann weiter die Genauigkeit dieses Näherungswertes abschätzen, nach einem Verfahren, das wir später (in Abschn. 72) kennenlernen werden.

Warum und wie man derartige Schlüsse von Stichproben auf die zugehörige Grundgesamtheit ziehen kann, lernen wir in Teil III des Buches. Wie sich der Begriff der Wahrscheinlichkeit, der doch zunächst nur ein Wort unserer Umgangssprache ist, so fassen läßt, daß man mathematisch damit arbeiten kann, überlegen wir uns in Teil II.

Kann der Ingenieur sein Ergebnis verbessern und einen zuverlässigeren Näherungswert für die genannte Dichte gewinnen?

Ja, das kann er tun, indem er zu größeren Stichproben (20 Meßwerte, 100 Meßwerte oder dgl.) übergeht. Wie wir in Teil III sehen werden, steigt dann die Genauigkeit der Schlüsse. Es ist klar, daß diese Verbesserung durch erhöhte Kosten und größeren Zeitbedarf erkauft wird. Beim 2. Schritt handelt es sich infolgedessen darum, den Versuchsplan nach Grundsätzen der mathematischen Statistik so anzulegen, daß man bei gegebenem Aufwand an Zeit und Kosten ein Maximum an Information über die Grundgesamtheit erhält.

Man mache sich klar, daß es vollkommen sichere Schlüsse von einer Stichprobe auf die Grundgesamtheit nicht gibt. An dieser prinzipiellen Tatsache kann die beste mathematische Theorie nichts ändern. Wenn man 10000 billige Massenartikel (Schräubchen, Hosenknöpfe, Rundfunkwiderstände usw.) vor sich hat und ganz genau wissen wollte, wie viele davon unbrauchbar (Ausschuß) sind, so bliebe nicht anderes übrig, als alle 10000 Artikel Stück für Stück zu prüfen und dabei die Ausschußstücke zu zählen.

Erfreulicherweise sind derart genaue Angaben in der Praxis fast nie wertvoller als die Aufschlüsse, die eine Stichprobe über die Grundgesamtheit liefert. So versteht man, daß die Stichprobenentnahme zu einem allgemein angewendeten Verfahren der modernen Statistik geworden ist.

Außer der Kosten- und Zeitfrage gibt es übrigens noch andere Gründe für die praktische Bedeutung der Stichprobenentnahme:

Die Prüfung kann zur Zerstörung führen, z. B. bei Experimenten über die Qualität von Sprengladungen, über die Stromstärke, bei der Sicherungen durchbrennen, über den Heizwert von Öl usw. Die Prüfung kann die Dinge verändern und damit ihrer Zweckbestimmung entziehen (Beispiel: Keimversuche an Saatgut). Die Grundgesamtheit kann zu groß sein (Beispiel: Bestimmung der Durchschnittsgröße der Bewohner eines Landes) oder nicht vollständig zugängig sein (Beispiel: Untersuchung über die Verbreitung von Seuchen durch Ratten, Blutuntersuchung, bei der Blut entnommen werden muß).

Schlüsse von einer Stichprobe auf die Grundgesamtheit, wie sie die mathematische Statistik ermöglicht, sind nur gültig, wenn wirklich eine Zufallsauswahl vorliegt, d. h. jedes Ding der Menge von Dingen, für die man sich interessiert, muß dieselbe Chance (oder wenigstens: eine genau angebbare Chance) haben, bei der Stichprobenentnahme gezogen zu werden. So muß man den Versuch anlegen. Das ist oft leichter gesagt als getan. Auf dieses grundlegende Problem gehen wir später (zu Anfang des Teiles III) näher ein.

TEIL I

BESCHREIBENDE STATISTIK

Im ersten Teil des Buches wollen wir uns überlegen, wie man statistische Daten in praktisch brauchbarer Form tabellarisch oder graphisch darstellt. Wir gehen dabei von Stichproben aus, wie sie bei Zufallsexperimenten anfallen, und kümmern uns einstweilen weder darum, wie diese Daten gewonnen wurden, noch, wie sie weiterverarbeitet werden sollen, d. h. wie man aus ihnen Aufschluß über die zugehörige Grundgesamtheit gewinnt.

Weiterhin werden wir sehen, daß man jeder Stichprobe gewisse **Maßzahlen** zuordnen kann, die die Stichprobe als Ganzes kennzeichnen. Die beiden praktisch wichtigsten dieser Zahlen sind der sogenannte Mittelwert und die sogenannte Varianz einer Stichprobe.

Um den vorliegenden ersten Teil des Buches möglichst einfach zu gestalten, betrachten wir nur Stichproben, bei denen jeder Stichprobenwert aus einer einzelnen Zahl besteht. Was wir über Stichproben von Zahlen*paaren*, Zahlen*tripeln* usw. wissen müssen, erläutern wir erst später.

Kapitel 2

Häufigkeitsverteilungen

Wir befassen uns im vorliegenden Kapitel mit der übersichtlichen Darstellung statistischen Materials. Die Begriffe der absoluten und relativen Häufigkeit, der Häufigkeitsfunktion $\tilde{f}(x)$ und der Summenhäufigkeitsfunktion oder Verteilungsfunktion $\tilde{F}(x)$ einer Stichprobe, die dabei vorkommen, sind auch für alle späteren Betrachtungen wichtig.

4 Tabellarische Darstellung. Häufigkeit

Bei statistischen Erhebungen werden die Beobachtungsergebnisse üblicherweise in der Reihenfolge nacheinander in Formulare oder Hefte eingetragen, in der sie anfallen. Die dabei entstehende Liste heißt das **Protokoll** oder die **Urliste.**

Tab. 4.1 ist eine solche Urliste. Sie enthält 100 Werte, nämlich die Größe von 100 zufällig aus der Gesundheitskartei ausgewählten Grazer Mittelschülerinnen. Das sind also n (= 100) Zahlen als Ergebnisse von 100 Beobachtungen (Längenmessungen). Diese bilden zusammen eine **Stichprobe** vom **Umfange** n, aus der man später Schlüsse auf die zugehörige **Grundgesamtheit** ziehen will. Die Zahlen heißen die **Stichprobenwerte.**

Übrigens bezeichnet man oft auch die Menge der ausgewählten Personen bzw. Dinge selbst als die Stichprobe.

Hätten wir gleichzeitig zwei Merkmale gemessen, etwa bei jeder der 100 Schülerinnen die Größe und das Gewicht, so hätten wir eine Stichprobe erhalten, die aus 100 Zahlen*paaren* besteht. Entsprechend erhält man Zahlen*tripel*, wenn man gleichzeitig drei Merkmale beobachtet, usw.

Wie stellt man nun eine Stichprobe übersichtlich dar?

Bei kleinen Stichproben hilft es oft schon, daß man die Werte der Größe nach ordnet. Bei unserer Stichprobe in Tab. 4.1 ist damit

Tabelle 4.1. Urliste. Größe von 100 achtzehnjährigen Mittelschülerinnen des Schuljahrs 1961–62. (Statistisches Amt des Magistrates Graz)

Lfd. Nr.										
1– 10	161	162	166	161	171	159	160	174	165	163
11– 20	161	178	157	156	160	172	167	162	164	156
21– 30	177	162	167	168	157	164	176	166	171	169
31– 40	171	155	170	158	171	167	161	172	169	161
41– 50	160	164	162	170	168	165	173	159	173	166
51– 60	170	154	165	162	174	158	156	165	160	165
61– 70	172	167	173	166	164	168	175	158	163	169
71– 80	171	166	159	162	159	171	163	158	167	168
81– 90	163	153	172	170	158	164	162	175	165	169
91–100	170	155	169	159	163	159	166	157	166	175

Tabelle 4.2. Strichliste zu Tab. 4.1

151		161	𝍸	171	𝍸 I
152		162	𝍸 II	172	IIII
153	I	163	𝍸	173	III
154	I	164	𝍸	174	II
155	II	165	𝍸 I	175	III
156	III	166	𝍸 II	176	I
157	III	167	𝍸	177	I
158	𝍸	168	IIII	178	I
159	𝍸 I	169	𝍸	179	
160	IIII	170	𝍸	180	

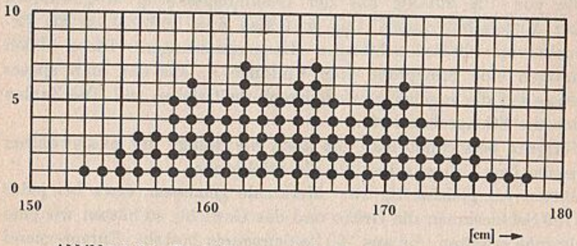


Abbildung 4.1. Punktdiagramm der Stichprobe in Tab. 4.1

wenig gewonnen. Wir können aber außerdem zahlenmäßig gleiche Werte zusammenfassen: Wir durchlaufen Tab. 4.1 zeilenweise. Jeden Wert tragen wir dabei als Strich in die entsprechende Zeile der Tab. 4.2

Tabelle 4.3. Häufigkeitsverteilung der Stichprobe in Tab. 4.1

Größe x [cm]	Absolute Häufigkeit	Relative Häufigkeit $h(x)$
153	1	0,01
154	1	0,01
155	2	0,02
156	3	0,03
157	3	0,03
158	5	0,05
159	6	0,06
160	4	0,04
161	5	0,05
162	7	0,07
163	5	0,05
164	5	0,05
165	6	0,06
166	7	0,07
167	5	0,05
168	4	0,04
169	5	0,05
170	5	0,05
171	6	0,06
172	4	0,04
173	3	0,03
174	2	0,02
175	3	0,03
176	1	0,01
177	1	0,01
178	1	0,01
Summe	100	1

ein, die als **Strichliste** bezeichnet wird. Daraus gewinnen wir dann schließlich die Tab. 4.3.

Statt der Strichliste können wir auch auf kariertem Papier das in Abb. 4.1 gezeigte **Punktdiagramm** anlegen. Die Anzahl der Punkte über einem Meßwert gibt an, wie oft dieser Zahlenwert in der Stichprobe vorkommt. Diese Anzahl heißt die **absolute Häufigkeit** des betreffenden Wertes in der Stichprobe. Zum Beispiel hat der Wert 171 die absolute Häufigkeit 6. Dividieren wir die absolute Häufigkeit durch den Stichprobenumfang n ($= 100$ in unserem Beispiel), so erhalten wir die **relative Häufigkeit** des betreffenden Wertes in der Stichprobe.

Die relative Häufigkeit eines Wertes a bezeichnen wir mit $h(a)$. Kommt also ein Wert a genau k-mal in der Stichprobe vor, so hat er die relative Häufigkeit

$$h(a) = \frac{k}{n} = \frac{\text{Anzahl der Versuche mit } a \text{ als Ergebnis}}{\text{Stichprobenumfang}}.$$

Zum Beispiel hat der Wert 171 in der obigen Stichprobe die relative Häufigkeit $h(171) = 0{,}06 = 6\%$.

Kommt eine Zahl a in einer Stichprobe überhaupt nicht vor, so ist $h(a) = 0$. Sind alle n Beobachtungen einer Stichprobe gleich a, so ist $h(a) = n/n = 1$. Es gilt also stets die Ungleichung

$$\boxed{0 \leqq h(a) \leqq 1.} \tag{4.1}$$

In Worten: *Die relative Häufigkeit ist eine nicht negative Zahl, die höchstens gleich 1 sein kann.*

Vorgelegt sei nun eine Stichprobe von n Werten, unter denen sich insgesamt m zahlenmäßig verschiedene Werte befinden, die wir mit

$$x_1, x_2, \ldots, x_m$$

bezeichnen. Für die zugehörigen relativen Häufigkeiten $h(x_1), h(x_2), \ldots$ schreiben wir einfacher

$$h_1, h_2, \ldots, h_m.$$

Wir führen nun eine Funktion $\tilde{f}(x)$ ein, die für x_1 den Wert h_1, für x_2 den Wert h_2 usw. hat und für jede Zahl x, die nicht in der Stichprobe vorkommt, gleich null ist. In Formeln:

$$\tilde{f}(x) = \begin{cases} h_j & \text{für } x = x_j \qquad (j = 1, 2, \ldots, m) \\ 0 & \text{für alle übrigen } x. \end{cases} \tag{4.2}$$

Diese Funktion heißt die **Häufigkeitsfunktion der Stichprobe.** Sie gibt an, wie die Werte in der Stichprobe verteilt sind, oder, wie wir auch sagen, sie bestimmt die **Häufigkeitsverteilung** der Stichprobe. Ein Beispiel gibt Tab. 4.3.

Die Summe aller absoluten Häufigkeiten in einer Stichprobe muß gleich dem Stichprobenumfang n sein (warum?). Die Summe aller relativen Häufigkeiten in der Stichprobe ist demnach gleich 1. Also ist

$$\tilde{f}(x_1) + \tilde{f}(x_2) + \cdots + \tilde{f}(x_m) = 1.$$

Wir können das Summenzeichen verwenden und dies kürzer schreiben:

$$\sum_{j=1}^{m} \tilde{f}(x_j) = 1. \tag{4.3}$$

Um dem mathematisch weniger geschulten Leser zu helfen, sei gesagt, daß diese abgekürzte Schreibweise von Summen sehr nützlich ist, weil sie die Übersichtlichkeit erhöht. Das große griechische Sigma (Σ) heißt **Summationszeichen** und bedeutet: Man bilde die Summe aller Ausdrücke, die man aus dem rechts neben dem Σ stehenden Ausdruck $\tilde{f}(x_j)$ erhält, wenn man für den „*Summationsbuchstaben*" j lauter ganze Zahlen einsetzt, und zwar mit dem unterhalb des Summationszeichens stehenden Wert $j = 1$ beginnend und der Reihe nach fortfahrend, bis man den oberhalb des Summationszeichens stehenden Wert m erreicht. So erhält man der Reihe nach

$$\tilde{f}(x_1), \quad \tilde{f}(x_2), \quad \tilde{f}(x_3), \quad \ldots, \quad \tilde{f}(x_m).$$

Das sind die Ausdrücke, deren Summe zu bilden ist. Diese Summe heißt „die Summe der $\tilde{f}(x_j)$ für j von 1 bis m". Weitere typische Beispiele sind

$$\sum_{j=1}^{4} 3^j = 3^1 + 3^2 + 3^3 + 3^4 = 3 + 9 + 27 + 81 = 120$$

$$\sum_{j=2}^{6} a_j = a_2 + a_3 + a_4 + a_5 + a_6.$$

Statt j kann man auch einen anderen Summationsbuchstaben wählen, der sonst noch nicht in anderer Bedeutung vorkommt. Beispiel:

$$\sum_{\alpha=1}^{4} 3^\alpha = 3^1 + 3^2 + 3^3 + 3^4 = 120.$$

Die Zahl unterhalb des Σ heißt die *untere Summationsgrenze*. Die Zahl oberhalb des Σ heißt die *obere Summationsgrenze*.

Aufgaben zu Abschnitt 4

4.1—4.8 Für die folgenden Stichproben stelle man ein Punktdiagramm und eine Tabelle von der Art der Tab. 4.3 her.

4.1 Gewichte [g] von 20 Ilvico-Dragees einer Packung (Hersteller E. Merck, Darmstadt)

0,65	0,65	0,64	0,65	0,65	0,62	0,63	0,61	0,65	0,64
0,65	0,68	0,66	0,67	0,62	0,61	0,62	0,68	0,68	0,62

4.2 Länge [cm] der Fiederblättchen eines Blattes der Robinie (Robinia pseudoacácia L.)

4,6	5,5	6,2	6,5	6,7	6,4	6,5	6,2	5,4	5,7
5,1	6,0	5,9	6,4	6,6	6,8	6,0	5,9	5,1	

4.3 Kohlenstoffgehalt [%] verschiedener Kohleproben (G. JURANEK, B. BUSSMANN u. G. SEIFERT, Brennstoff-Chemie 44, 1963, 19)

87	86	85	87	86	87	86	81	77	85
86	84	83	83	82	84	83	79	82	73

4.4 Schalldämmzahl [db] von 10 cm starken Gipsdielenwänden bei 400 Hz (U. PISKE, Lärmbekämpfung 7, 1963, 11)

25	24	27	24	24	23	28	26

4.5 Zugfestigkeit [kg/mm^2] von Blechen (A. SLATTENSCHEK u. G. SCHNEEWEISS, Maschinenbau u. Wärmewirtsch. 12, 1957, 277)

44	43	41	41	44	44	43	44	42	45	43	43	44	45	46
42	45	41	44	44	43	44	46	41	43	45	45	42	44	44

4.6 Heuerträge [dz/ha] für Rotschwingel (F. ZÜRN, Veröff. Bundesanst. alp. Landw. Admont, 1953, 120)

92	88	87	92	98	87	72	84
89	85	95	84	96	91	91	101

4.7 Würfe mit einer Münze (0 = „Wappen", 1 = „Kopf")

0 0 1 1 1 0 1 0 1 1

4.8 Würfe mit einem Würfel

2 5 3 2 4 1 6 2 4 5

4.9 Man würfle 50mal mit einem Würfel und erstelle aus der Urliste der Ergebnisse das Punktdiagramm und eine Tabelle der Häufigkeiten.

4.10 Man bestimme die Häufigkeiten der Buchstaben *a*, *b*, ... auf irgendeiner Seite irgendeines Buches.

5 Graphische Darstellungen

Indem wir die Punkte in Abb. 4.1 durch vertikale Striche verbinden, erhalten wir das in Abb. 5.1 gezeigte **Stabdiagramm.**

Verbreitern wir alle Stäbe dieses Diagramms gleichmäßig und symmetrisch nach rechts und links zu Rechtecken, die aneinanderstoßen, so ergibt sich aus Abb. 5.1 das in Abb. 5.2 dargestellte **Histogramm** oder **Staffelbild.** Offenbar sind die Rechteckflächen jeweils proportional zu den entsprechenden Häufigkeiten.

Verbinden wir die Endpunkte der Stäbe des Stabdiagramms durch Strecken und lassen die Stäbe weg, so entsteht das **Häufigkeitspolygon** in Abb. 5.3.

Zwei weitere Arten graphischer Darstellungen zeigt die Abb. 5.4.

Mit graphischen Darstellungen statistischer Daten kann man übrigens auch erheblichen Schwindel treiben, indem man zwar rich-

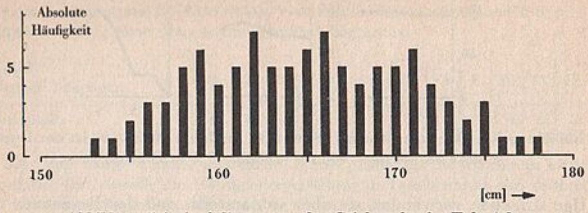


Abbildung 5.1. Stabdiagramm der Stichprobe in Tab. 4.3

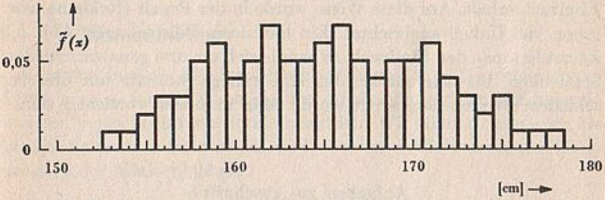


Abbildung 5.2. Histogramm der Stichprobe in Tab. 4.3

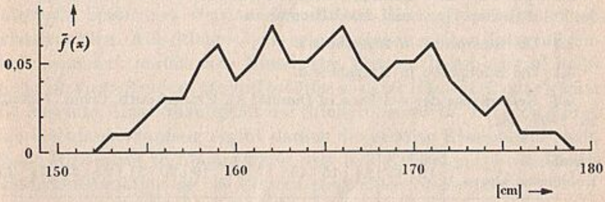
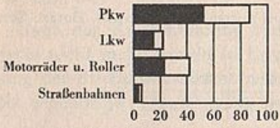
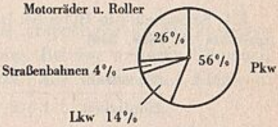


Abbildung 5.3. Häufigkeitspolygon der Stichprobe in Tab. 4.3



Gegliedertes Stabdiagramm
(schwarz = stadtwärts, weiß = nach auswärts)



Kreisdiagramm
(Fahrzeuge in beiden Richtungen zusammengenommen)

Abbildung 5.4. Verkehrszählung in der Münzgrabenstraße in Graz (18. 9. 1962, 14.46–15.16 Uhr, 100 m nordwestlich der Brockmanngasse)

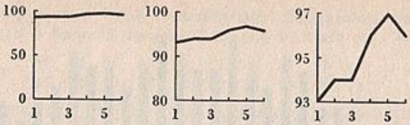


Abbildung 5.5. Weltmarktpreise (SCHULZE) im Jahre 1959. (100 ist der Durchschnitt des Jahres 1950.) 1 = Januar, 2 = Februar usw.

tige Angaben verwendet, sie aber so darstellt, daß der Beschauer bei mangelnder Schulung oder Aufmerksamkeit einen ganz verkehrten Eindruck erhält. Auf diese Weise wurde in der Praxis (Reklame usw.) schon viel Unheil angerichtet. Ein harmloses Beispiel zeigt Abb. 5.5. Zuerst hat man den Eindruck, es handle sich um drei ganz verschiedene Statistiken, bis man merkt, daß eine einzige Statistik auf drei verschiedene Weisen dargestellt wurde. Man muß also vorsichtig sein.

Aufgaben zu Abschnitt 5

5.1—5.4 Für die folgenden Stichproben zeichne man das Stabdiagramm, das Häufigkeitspolygon und das Histogramm.

5.1 Die Stichprobe in Aufgabe 4.3.

5.2 Die Stichprobe in Aufgabe 4.5.

5.3 Kettenlänge der n-Alkane in Dieselöl (G. HILDEBRAND, Chem. Technik **15**, 1963, 484)

Anzahl der Kohlenstoffatome	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Häufigkeit [%]	2	3	4	7	11	13	14	12	10	8	7	4	3	2

5.4 Samenzahl pro Hülse bei Indigofera australis (F. LUDWIG, Botan. Zentralblatt **73**, 1898, 348)

Samenzahl	3	4	5	6	7	8	9	10	11
Häufigkeit	1	2	8	13	22	45	63	23	1

5.5 Der Farbengeschmack von 36 Verkäuferinnen und Lehrtöchtern in einem Betrieb der Damenkonfektionsbranche wurde folgendermaßen getestet: Jeder Teilnehmerin wurden 12 Karten vorgelegt. Auf jeder Karte befanden sich 4 Farbpaare, davon 3 ästhetisch befriedigende und 1 unbefriedigendes. Letzteres war zu bestimmen. Das Ergebnis zeigt die Tabelle (A. MÜLLER, Industr. Organisation 28, 1958, 335). Vier Damen lösten also genau 6 der 12 Aufgaben richtig, usw. Man zeichne ein Stabdiagramm.

Anzahl richtiger Lösungen	0	1	2	3	4	5	6	7	8	9	10	11	12
Häufigkeit	0	0	1	1	1	1	4	3	10	3	7	2	3

5.6 Man würfle 100mal mit einem Würfel, halte die Ergebnisse in einer Strichliste fest, erstelle die Häufigkeitsverteilung in Tabellenform und zeichne das zugehörige Stabdiagramm und das Histogramm.

6 Klassenbildung

Kommen in einer Stichprobe sehr viele zahlenmäßig verschiedene Werte vor, so ist die Tabelle oder die Zeichnung der Häufigkeitsfunktion meist noch recht unübersichtlich. Wir können dann aber die Stichprobe weiter vereinfachen, und zwar durch die sogenannte *Gruppierung* oder **Klassenbildung.**

Dabei gehen wir von einem Intervall aus, in dem alle Stichprobenwerte liegen. Dieses unterteilen wir in Teilintervalle, die als **Klassenintervalle** bezeichnet werden. Die Mitten dieser Intervalle heißen **Klassenmitten.** Alle Stichprobenwerte in einem solchen Intervall bilden zusammen jeweils eine **Klasse** von Werten. Deren Anzahl heißt die *zu dem betreffenden Intervall gehörige absolute Häufigkeit* oder kürzer die **absolute Klassenhäufigkeit** der Stichprobenwerte. Division durch den Stichprobenumfang ergibt daraus die **relative Klassenhäufigkeit.** Diese Häufigkeit in Abhängigkeit von den Klassenmitten heißt die **Häufigkeitsfunktion** *der in Klassen eingeteilten Stichprobe.*

Zum Beispiel erhält man aus der Tab. 6.1 durch Klassenbildung die Tab. 6.2.

Nach der Klassenbildung treten die ursprünglichen Stichprobenwerte nicht mehr einzeln in Erscheinung. Bei der weiteren Verarbeitung der Stichprobe nimmt man an, daß alle Werte einer Klasse in der zugehörigen Klassenmitte liegen.

Je weniger Klassen man bildet, desto einfacher wird zwar die Stichprobe, desto mehr Information, die in den ursprünglichen Stichprobenwerten steckt, geht aber verloren. Man sollte so klassifizieren, daß nur unwesentliche Einzelheiten ausgeschieden werden. In der Praxis wählt man meist 10–20 Klassen und mehr als 20 höchstens

Tabelle 6.1. Druckfestigkeit [kg/cm²] von Betonwürfeln mit 20 cm Kantenlänge (W. Eberl und G. Schneeweiss, Öst. Ing. Arch. **11**, 1957, 173)

358	392	368	324	307	308	235	228	237	317
346	276	299	284	293	330	376	381	333	389
371	333	334	364	443	489	401	434	354	366
328	341	374	279	302	320	453	458	410	261
279	244	353	345	361	301	402	379	250	230
278	335	342	300	290	352	358	239	349	315
359	397	394	324	336	352	328	302	316	285
285	303	314	318	355	271	245	209	246	272
317	322	386	328	378	368	353	419	344	355

Tabelle 6.2. Stichprobe in Tabelle 6.1 nach der Klassenbildung

Klassenmitte [kg/cm²]	Klassenintervall [kg/cm²]	Absolute Häufigkeit Strichliste	Wert	Relative Häufigkeit
200	187,5 – 212,5	I	1	0,011
225	212,5 – 237,5	IIII	4	0,044
250	237,5 – 262,5	𝍸 I	6	0,067
275	262,5 – 287,5	𝍸 IIII	9	0,100
300	287,5 – 312,5	𝍸 𝍸	10	0,111
325	312,5 – 337,5	𝍸 𝍸 𝍸 IIII	19	0,211
350	337,5 – 362,5	𝍸 𝍸 𝍸 II	17	0,189
375	362,5 – 387,5	𝍸 𝍸 I	11	0,122
400	387,5 – 412,5	𝍸 II	7	0,078
425	412,5 – 437,5	II	2	0,022
450	437,5 – 462,5	III	3	0,033
475	462,5 – 487,5		0	0,000
500	487,5 – 512,5	I	1	0,011

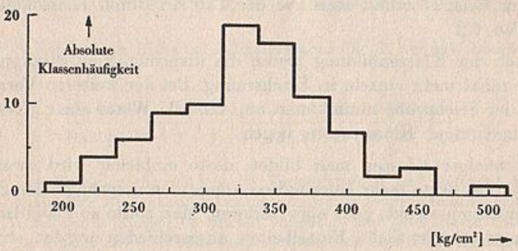


Abbildung 6.1. Histogramm der Stichprobe in Tab. 6.2

bei sehr umfangreichen Stichproben. Unnötige Komplikationen bei späteren Rechnungen lassen sich vermeiden, wenn man die folgenden **Regeln** beachtet:

(1) *Die Klassenintervalle wählt man gleich lang.*

(2) *Die Klassenmitten sollen möglichst einfachen Zahlen, d. h. Zahlen mit möglichst wenigen Ziffern, entsprechen.*

(3) *Ein Wert, der auf einen Intervallendpunkt fällt, wird je zur Hälfte in jedem der beiden angrenzenden Klassenintervalle mitgezählt. Oftmals kann man die genannten Endpunkte ohne Mühe so wählen, daß sie nicht mit Stichprobenwerten zusammenfallen. Dann sollte man dies tun.*

Aufgaben zu Abschnitt 6

6.1—6.6 Für die folgenden in Klassen eingeteilten Werte zeichne man das Histogramm.

6.1 Höhen von neunjährigen Kiefern (Vater, Mitt. Sächs. Versuchsanstalt Tharandt 2, 1925, 8).

6.2 Milchertrag steirischer Kühe (Tätigkeitsbericht des Steirischen Pinzgauer Zuchtverbandes 1948).

Zu Aufgabe 6.1

Höhe [cm]	Anzahl
55— 75	2
75— 95	0
95—115	4
115—135	8
135—155	17
155—175	27
175—195	30,5
195—215	19,5
215—235	10
235—255	5
255—275	2

Zu Aufgabe 6.2

Milchertrag [kg/Jahr]	Anzahl Kühe
0—1000	11
1000—2000	556
2000—3000	1169
3000—4000	326
4000—5000	32
5000—6000	4

6.3 Gewicht männlicher ausgetragener Neugeborener (Univ.-Frauenklinik Graz, Dir. Prof. Dr. E. Navratil, 1961).

6.4 Heiratsalter der Männer in Österreich im Jahre 1955 (Stat. Handbuch Österr., Wien 1956, S. 19).

Zu Aufgabe 6.3

Gewicht [g]	Anzahl
2400 – 2600	2
2600 – 2800	10
2800 – 3000	21
3000 – 3200	23
3200 – 3400	26
3400 – 3600	24
3600 – 3800	22
3800 – 4000	12
4000 – 4200	6
4200 – 4400	0
4400 – 4600	0
4600 – 4800	3
4800 – 5000	1

Zu Aufgabe 6.4

Alter [in vollendeten Lebensjahren]	Anzahl geschlossener Ehen (abgerundet)
15 – 19	1000
20 – 24	17000
25 – 29	18000
30 – 34	8000
35 – 39	3000
40 – 44	3000
45 – 49	2000
50 – 54	2000
55 – 59	1000
60 – 64	1000

6.5 Radar-Rückstreuquerschnitt von Schiffsobjekten im Hamburger Hafen (R. MANTHEY, Telefunken-Z. **35**, 1962, 204).

6.6 Korngrößen bei Drehofenklinker (E. MATOUSCHEK, Radex-Rundschau 1962, 15).

Zu Aufgabe 6.5

Querschnitt [m²]	Anzahl Schiffe
0 – 1	1
1 – 10	0
10 – 100	4
100 – 1000	21
1000 – 10000	21
10000 – 100000	16
100000 – 1000000	3

Zu Aufgabe 6.6

Korngrenzen [mm]	Häufigkeit [%]
0 – 2	1
2 – 4	24
4 – 6	9
6 – 8	16
8 – 10	18
10 – 12	12
12 – 14	15
14 – 16	5

6.7 In Aufgabe 6.1 fasse man die Klassen paarweise zusammen zu den Klassen 55 – 95, 95 – 135, ..., 255 – 295 und vergleiche das zugehörige Histogramm mit dem ursprünglichen.

6.8 Die Gestalt des Histogramms einer gruppierten Stichprobe hängt nicht nur von der Länge der Klassenintervalle, sondern auch von der Wahl der Klassenmitten ab. Man illustriere dies, indem man die Stichprobe in Tab. 6.1 wie in Tab. 6.2 gruppiert, dabei aber die Klassenmitten verschiebt (z. B. 210, 235, 260, ... wählt) und das Histogramm mit dem in Abb. 6.1 vergleicht.

6.9 Bei ganz kritischer Betrachtung kann man behaupten, daß im Falle einer stetigen Verteilung die ursprüngliche Stichprobe eigentlich auch schon in Klassen eingeteilt ist. Wieso?

7 Summenhäufigkeitsfunktion einer Stichprobe

Die Häufigkeitsfunktion $\tilde{f}(x)$ einer Stichprobe gibt die relativen Häufigkeiten an, mit der die einzelnen Zahlenwerte in der Stichprobe vorkommen. So sehen wir aus der Tab. 7.1 z. B. sofort, daß Kamillenblüten mit 17 Blütenblättern in der Stichprobe die relative Häufigkeit 13% haben. Fragen wir nun nach der relativen Häufigkeit der Blüten mit 17 *oder weniger* Blütenblättern, so können wir die Antwort zwar nicht unmittelbar aus der Tabelle ablesen, wir können sie aber dadurch gewinnen, daß wir die Werte $\tilde{f}(x)$ für $x = 12$ bis 17 summieren. Dabei erhalten wir

$$0{,}02 + 0{,}07 + 0{,}03 + 0{,}06 + 0{,}08 + 0{,}13 = 0{,}39 = 39\%.$$

Tun wir dies für jeden Wert von x in Tab. 7.1, so ergibt sich die Tab. 7.2 der „**relativen Summenhäufigkeiten**" unserer Stichprobe.

Wir führen nun die folgende Funktion ein:

$\tilde{F}(x) =$ *Summe der relativen Häufigkeiten aller Stichprobenwerte, die kleiner als x oder gleich x sind.*

Diese Funktion heißt die **Summenhäufigkeitsfunktion** oder **Verteilungsfunktion** der Stichprobe.

Tabelle 7.1. Blütenblätteranzahl x von 100 Blüten der Falschen Kamille (Matricaria inodora. Schörgelhof Graz, 20. 9. 1962)

x	Relative Häufigkeit
12	0,02
13	0,07
14	0,03
15	0,06
16	0,08
17	0,13
18	0,06
19	0,12
20	0,13
21	0,22
22	0,06
23	0,02

Tabelle 7.2. Relative Summenhäufigkeiten für die Stichprobe in Tab. 7.1

x	Relative Summenhäufigkeit
12	0,02
13	0,09
14	0,12
15	0,18
16	0,26
17	0,39
18	0,45
19	0,57
20	0,70
21	0,92
22	0,98
23	1,00

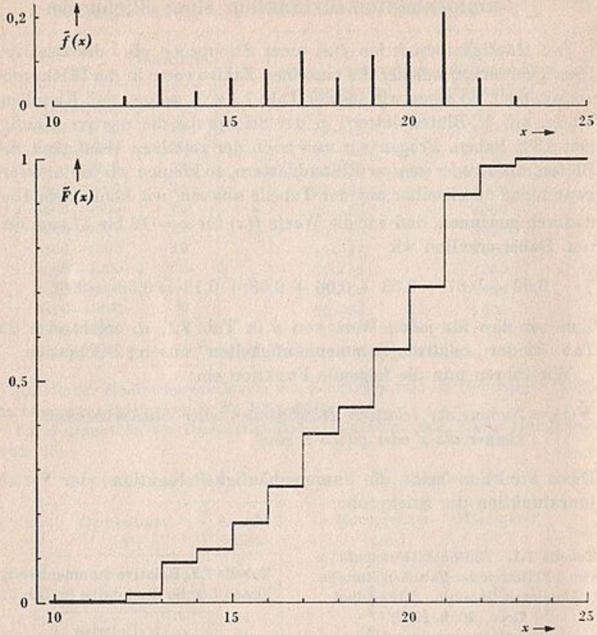


Abbildung 7.1. Häufigkeitsfunktion $\tilde{f}(x)$ und Summenhäufigkeitsfunktion $\tilde{F}(x)$ der Stichprobe in Tab. 7.1

Abb. 7.1 zeigt $\tilde{f}(x)$ und $\tilde{F}(x)$ für die Stichprobe in Tab. 7.1. Die Sprunghöhe von $\tilde{F}(x)$ für ein ganzzahliges x ist gleich der entsprechenden Stablänge im oberen Teil der Abb. 7.1. (Warum ?) Man kann demzufolge die graphische Darstellung von $\tilde{F}(x)$ in einfacher Weise aus dem Stabdiagramm von $\tilde{f}(x)$ gewinnen und umgekehrt.

Jede der beiden genannten Funktionen bestimmt unsere Stichprobe in allen Einzelheiten. Man muß zugeben, daß die Summenhäufigkeitsfunktion weniger anschaulich als die Häufigkeitsfunktion ist. Daß man sie trotzdem einführt und ihr bei vielen Überlegungen sogar den Vorzug vor der Häufigkeitsfunktion gibt, hat theoretische und praktische Gründe, die wir später (in Teil II und III) erörtern.

Unsere obige Definition von $\tilde{F}(x)$ können wir leicht in eine kurze Formel fassen:

$$\tilde{F}(x) = \sum_{t \leq x} \tilde{f}(t). \tag{7.1}$$

Das Summationszeichen $\sum$, das schon in Abschn. 4 vorkam, besagt, es ist die Summe der Werte von $\tilde{f}(t)$ zu bilden, und zwar derjenigen Werte, für die das zugehörige t höchstens gleich dem jeweiligen Wert x ist. Das drückt die Angabe $t \leq x$ aus. Wir haben t als Summationsbuchstaben genommen, weil x schon als Summationsgrenze verwendet wird. Da $\tilde{f}$ für nur endlich viele Werte von null verschieden ist, handelt es sich natürlich immer um eine Summe mit nur endlich vielen Gliedern. Und $\tilde{F}(x)$ ist stets eine **Treppenfunktion** (= stückweise konstante Funktion), die genau an denjenigen Stellen x nach oben springt, an denen $\tilde{f}(x)$ nicht null ist. Die Sprunghöhe ist gleich dem Wert von $\tilde{f}(x)$ an der betreffenden Stelle. Der erste Sprung liegt beim kleinsten Stichprobenwert und der letzte Sprung beim größten. Beim letzten Sprung erreicht $\tilde{F}(x)$ den Wert 1. (Warum?)

$\tilde{F}(x)$ heißt im Englischen *distribution function of the sample*, bisweilen auch *cumulative distribution function of the sample* (insbesondere, wenn Autoren die Häufigkeitsfunktion der Stichprobe mit *distribution function* bezeichnen, was manchmal getan wird).

Im Falle einer in Klassen eingeteilten Stichprobe benutzt man die Funktion

$\tilde{F}(x)$ = *Summe der relativen Klassenhäufigkeiten aller Klassen, deren Mitten kleiner als x oder gleich x sind.*

Diese heißt die *Summenhäufigkeits-* oder *Verteilungsfunktion* der betreffenden Stichprobe. Sie ist eine Treppenfunktion mit Sprüngen an den Stellen der Klassenmitten. Die Sprunghöhe ist jeweils gleich der zugehörigen relativen Klassenhäufigkeit.

Ersetzt man in dieser Definition von $\tilde{F}(x)$ das Wort „Mitten" durch „rechte Endpunkte", so erhält man eine Treppenfunktion mit Sprüngen an den Klassenendpunkten. Diese Funktion wird in der Literatur oftmals als Summenhäufigkeits- oder Verteilungsfunktion der in Klassen eingeteilten Stichprobe bezeichnet. Wie im vorigen Abschnitt gesagt wurde, tut man bei der Weiterbearbeitung der Stichprobe so, als ob alle Werte einer Klasse in der zugehörigen Klassenmitte lägen. Daraus erkennt man, daß die obige Definition von $\tilde{F}(x)$ logisch vorzuziehen ist. Wir werden dies später (in Abschn. 66) noch etwas näher erörtern.

Aufgaben zu Abschnitt 7

7.1—**7.5** Für die folgenden Stichproben bestimme man die Werte der Verteilungsfunktion und stelle diese Funktion graphisch dar.

7.1 Die Stichprobe in Aufgabe 4.7.

7.2 Würfe mit einem Würfel

6 2 4 1 2 4 3 3 2 1 6 5 6 3 4 1 6 2 5 3

7.3 Die Stichprobe in Aufgabe 4.5.

7.4 Die Stichprobe in Tab. 6.2.

7.5 Die Stichprobe in Aufgabe 6.1.

7.6 Man bestimme und zeichne die Verteilungsfunktion für die Bruttomonatsverdienste männlicher Industriearbeiter in der Bundesrepublik Deutschland im Oktober 1957 (Stat. Jahrbuch f. d. Bundesrep. Deutschland 1960, S. 512).

Verdienst [DM]	0—200	200—400	400—600
Häufigkeit (abgerundet)	2000	88000	358000

Verdienst [DM]	600—800	800—1000	1000—1200
Häufigkeit (abgerundet)	109000	13000	2000

7.7 Man bestimme die Verteilungsfunktion der Stichprobe in Tab. 6.2 auf graphischem Wege unter Benutzung der Abb. 6.1.

7.8 Bei Versuchen zur prophylaktischen Bekämpfung der Nonne (Lymantria monacha L.) wurden an 30 Probestämmen die folgenden Eizahlen beobachtet (C. PALLY, Arch. f. Forstwesen **12**, 1963, 827):

125	212	284	176	100	132	52	319	410	181
273	186	43	11	109	20	76	30	73	47
121	518	129	22	314	144	381	225	257	138

Man überlege sich eine geeignete Klasseneinteilung und stelle die zugehörige Verteilungsfunktion graphisch dar.

Kapitel 3

Mittelwert und Varianz einer Stichprobe

Jede Stichprobe wird durch ihre Häufigkeitsfunktion oder ebensogut durch ihre Summenhäufigkeitsfunktion in allen Einzelheiten beschrieben und bestimmt. Daneben kann man eine Stichprobe auch mehr „summarisch" kennzeichnen, und zwar durch gewisse Werte, die man aus der Häufigkeitsfunktion berechnet und die als **Maßzahlen der Stichprobe** bezeichnet werden. Von diesen Zahlen handelt das vorliegende Kapitel.

8 Mittelwert und Varianz einer Stichprobe

Die beiden praktisch wichtigsten Maßzahlen sind der Mittelwert, der die durchschnittliche Größe der Stichprobenwerte kennzeichnet, und die Varianz, die mißt, wie stark diese Werte streuen.

Der **Mittelwert** einer Stichprobe $x_1, \ldots, x_n$ ist definiert als das arithmetische Mittel der Stichprobenwerte und wird mit $\bar{x}$ bezeichnet. Es ist also

$$\boxed{\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}.} \tag{8.1}$$

Unter Verwendung des Summenzeichens schreiben wir kürzer

$$\bar{x} = \frac{1}{n} \sum_{j=1}^{n} x_j .$$

Beispiel 8.1. Fünf Proben von Permalloy-Legierungen mit rechteckiger Hystereseschleife hatten den prozentualen Nickelgehalt

79,4 79,0 78,9 79,2 78,9

(F. Pfeifer, Zeitschr. angew. Physik **13**, 1961, 177). Diese 5 Stichprobenwerte haben die Summe 395,4. Die Stichprobe hat also den Mittelwert

$$\bar{x} = \frac{395,4}{5} = 79,08\ [\%].$$

Das ist der durchschnittliche prozentuale Nickelgehalt.

Die Stichproben

$$1 \quad 2 \quad 4 \quad 5 \qquad \text{und} \qquad 2{,}7 \quad 3{,}0 \quad 3{,}1 \quad 3{,}2$$

haben beide den Mittelwert 3, unterscheiden sich aber trotzdem recht wesentlich voneinander, denn die Werte der ersten Stichprobe liegen viel weiter auseinander (und auch weiter vom Mittelwert weg) als die Werte der zweiten. Um diesen Unterschied zu erfassen, braucht man noch eine weitere Maßzahl. Geeignet ist hierzu offenbar eine Zahl, die die Abweichung der Stichprobenwerte $x_1, \ldots, x_n$ vom Mittelwert $\bar{x}$ mißt. Dabei fordern wir, daß, ähnlich wie beim Mittelwert, *jeder* Stichprobenwert in gewisser Weise mitberücksichtigt wird.

Die wohl am nächsten liegende Möglichkeit, die Summe der Einzelabweichungen $x_1 - \bar{x}, \ldots, x_n - \bar{x}$ zu benutzen, scheidet allerdings aus, denn es gilt immer die Beziehung

$$(x_1 - \bar{x}) + (x_2 - \bar{x}) + \cdots + (x_n - \bar{x}) = 0. \tag{8.2}$$

In der Tat steht ja links

$$x_1 + x_2 + \cdots + x_n - n\bar{x},$$

und die Summe der Stichprobenwerte ist gemäß (8.1) gleich $n\bar{x}$.

Unser Fehlschlag kommt daher, daß die Summe in (8.2) im allgemeinen positive und negative Glieder enthält. Er wird vermieden, indem wir entweder zu den Absolutbeträgen oder zu den Quadraten der Glieder $x_j - \bar{x}$ übergehen. Der zweite Weg erfordert zwar etwas mehr Rechenarbeit, erweist sich aber im Hinblick auf die mathematische Begründung der zu entwickelnden statistischen Verfahren als einfacher. Die Maßzahl, die wir auf diesem Wege erhalten, heißt die **Varianz** der Stichprobe $x_1, \ldots, x_n$. Sie wird mit s^2 bezeichnet und durch die Formel

$$\boxed{s^2 = \frac{1}{n-1} \sum_{j=1}^{n} (x_j - \bar{x})^2} \qquad (n > 1) \tag{8.3}$$

definiert.

Sind alle x_j zahlenmäßig gleich, so wird $\bar{x} = x_j$, und die Varianz ist gleich null. In jedem anderen Falle gilt

$$s^2 > 0.$$

Die nichtnegative Quadratwurzel der Varianz heißt die **Standardabweichung** und wird mit s bezeichnet.

Daß man gleich beide Größen s^2 *und* s mit einem Namen belegt, kommt daher, daß man beide in der Praxis verwendet. s^2 hat den Vorteil, daß man sich nicht mit Quadratwurzeln herumzuärgern braucht. s hat den Vorteil, daß es dieselbe Dimension (z. B. cm oder kg usw.) wie die x_j besitzt.

Beispiel 8.2. Die beiden obigen Stichproben haben den Umfang $n = 4$, den Mittelwert $\bar{x} = 3$ und die Varianz

$$s^2 = \frac{1}{3}\,[(1-3)^2 + (2-3)^2 + (4-3)^2 + (5-3)^2] \approx 3{,}3$$

bzw.

$$s^2 = \frac{1}{3}\,[(2{,}7-3{,}0)^2 + (3{,}0-3{,}0)^2 + (3{,}1-3{,}0)^2 + (3{,}2-3{,}0)^2] \approx 0{,}05.$$

Die Varianz der zweiten Stichprobe ist also wesentlich kleiner als die der ersten, und dasselbe gilt für die Standardabweichungen

$$s \approx \sqrt{3{,}3} \approx 1{,}8 \qquad \text{bzw.} \qquad s \approx \sqrt{0{,}05} \approx 0{,}22.$$

Im Englischen heißt s^2 einheitlich *variance*, und s heißt *standard deviation*. Im Deutschen nennt man s^2 manchmal auch *Streuung*. Leider bezeichnen gewisse Autoren die Standardabweichung s als Streuung. Wenn man solche Literatur liest, muß man also aufpassen.

Man wundert sich vielleicht, daß in (8.3) im Nenner $n - 1$ statt n steht, wie man erwarten sollte. Ein Grund dafür wird in Abschn. 64 angegeben. Man bezeichnet die Zahl $n - 1$ als die Anzahl der *Freiheitsgrade* der Summe der n Quadrate in (8.3). nicht gut genug

Aufgaben zu Abschnitt 8

8.1 Man stelle die Häufigkeitsfunktion der Stichproben in Beispiel 8.2 graphisch dar.

8.2—8.4 Man berechne den Mittelwert, die Varianz und die Standardabweichung der folgenden Stichproben.

8.2 1, 2, 3.

8.3 Wirkungsgrad [%] der Kessel einer Hochdruckdampfanlage (R. Wlodawer, Techn. Rundschau Sulzer **43**, 1961, 34)

90,3 91,6 90,9 90,4 90,3 91,0 87,9 89,4.

8.4 Die Stichprobe in Aufgabe 4.4.

8.5 Man konstruiere eine Stichprobe mit dem Umfang 2, dem Mittelwert 0 und der Varianz 1.

8.6 Man beweise: Sind die Werte einer Stichprobe nicht alle zahlenmäßig gleich, so liegt der Mittelwert zwischen dem kleinsten und dem größten Stichprobenwerte.

8.7 Die Differenz zwischen dem größten und dem kleinsten Stichprobenwert heißt die **Spannweite** der betreffenden Stichprobe. Welche Vor- und Nachteile hat es, diese Zahl für die Variabilität der Stichprobenwerte zu benutzen?

8.8 Man berechne die Spannweite der Stichprobe in Aufgabe 8.2.

8.9 Bei gegebenem $x_1, \ldots, x_n$ ist

$$H(t) = \sum_{j=1}^{n} (x_j - t)^2$$

eine Funktion von t. Für welches t wird $H(t)$ am kleinsten?

8.10 Die Werte einer Stichprobe seien in Zentimetern ausgedrückt. Es sei $\bar{x} = 10$, $s^2 = 4$, also $s = 2$. Welche Zahlenwerte ergäben sich für $\bar{x}$, s^2 und s, wenn man die Stichprobenwerte in Millimetern ausdrückte?

9 Vereinfachte Berechnung des Mittelwertes und der Varianz

Wir wollen nun einige Formeln angeben, die für die praktische Berechnung des Mittelwertes und der Varianz meist günstiger als die Definitionsformeln (8.1) und (8.3) sind. Auch der mathematisch weniger Geübte sollte die Mühe nicht scheuen und sich mit den neuen Formeln anfreunden. Es lohnt sich.

Multiplizieren wir jedes Quadrat in (8.3) aus, so besteht jedes Glied aus 3 Summanden. Dann können wir die Summe in (8.3) entsprechend in 3 Summen zerlegen. Ersetzen wir noch $\bar{x}$ gemäß (8.1) durch $\sum x_j/n$, so ergibt sich (vgl. auch Aufgabe 9.1)

$$\textbf{(9.1)} \qquad \boxed{s^2 = \frac{1}{n-1}\left[\sum_{j=1}^{n} x_j^2 - \frac{1}{n}\left(\sum_{j=1}^{n} x_j\right)^2\right].}$$

Der Vorteil gegenüber (8.3) ist, daß die Differenzen $x_j - \bar{x}$ nicht auftreten, die kleine Differenzen großer Zahlen sein, also Genauigkeitsverluste bedingen können (die man beim elektronischen Rechnen nicht einmal bemerken würde).

Ersetzen wir $\sum x_j$ in (9.1) gemäß (8.1) durch $n\bar{x}$, so folgt

$$(9.2) \qquad s^2 = \frac{1}{n-1}\left[\sum_{j=1}^{n} x_j^2 - n\bar{x}^2\right].$$

Dies ist nicht so gut wie (9.1), weil man beim Bilden von $\bar{x}^2$ durch n^2 dividiert und dann wieder mit n multipliziert.

Aber (9.1) hat für die Tischrechenmaschine oft noch den Nachteil, daß große Zahlen vorkommen. Dem läßt sich durch eine „**Nullpunktverschiebung**" abhelfen, mit der man zugleich auch $\bar{x}$ einfacher erhält: Wir setzen

$$\textbf{(9.3)} \qquad x_j = c + x_j^* \qquad (\text{also } x_j^* = x_j - c)$$

und wählen dabei die Konstante c so, daß die $x_j{}^*$ kleine handliche Werte haben. Aus der gegebenen Stichprobe $x_1, \ldots, x_n$ erhalten wir dann die „transformierte Stichprobe" $x_1{}^*, \ldots, x_n{}^*$. Wir berechnen nun zuerst den Mittelwert $\bar{x}^*$ der transformierten Stichprobe. Daraus ergibt sich der Mittelwert

$$\text{(9.4)} \qquad \bar{x} = c + \bar{x}^* \qquad \left(\text{mit } \bar{x}^* = \frac{1}{n}\sum_{j=1}^{n} x_j{}^*\right)$$

der ursprünglichen Stichprobe. Weiterhin hat diese Stichprobe dieselbe Varianz wie die transformierte. Die einfachen Beweise dieser Tatsachen überlassen wir dem Leser (s. Aufgabe 9.3 und 9.4).

Beispiel 9.1. Für die Stichprobe in Beispiel 8.1 erhalten wir aus (8.3) und Tab. 9.1 die Varianz

$$s^2 = \frac{0{,}1880}{4} = 0{,}047.$$

Aus (9.1) und Tab. 9.2 ergibt sich ebenfalls

$$s^2 = \frac{1}{4}\left[31268{,}42 - \frac{1}{5}(395{,}4)^2\right] = 0{,}047.$$

Viel einfacher wird die Rechnung, wenn wir (9.3) mit $c = 79$ anwenden. Dann ist

$$x_j{}^* = x_j - 79.$$

Gemäß Tab. 9.3 haben die transformierten Werte den Mittelwert

$$\bar{x}^* = \frac{0{,}4}{5} = 0{,}08.$$

Aus Tab. 9.3 und Formel (9.1) [mit $x_j{}^*$ statt x_j] ergibt sich die Varianz

$$s^{*2} = \frac{1}{4}\left[0{,}22 - \frac{0{,}4^2}{5}\right] = \frac{0{,}188}{4} = 0{,}047.$$

Wegen (9.4) hat also die gegebene Stichprobe den Mittelwert

$$\bar{x} = 79 + 0{,}08 = 79{,}08$$

und die Varianz $s^2 = 0{,}047$.

Tabelle 9.1—9.3. Zu Beispiel 9.1

Tabelle 9.1

x_j	$x_j - \bar{x}$	$(x_j - \bar{x})^2$
79,4	0,32	0,1024
79,0	−0,08	0,0064
78,9	−0,18	0,0324
79,2	0,12	0.0144
78,9	−0,18	0,0324
	0,00	0,1880

Tabelle 9.2

x_j	x_j^2
79,4	6304,36
79,0	6241,00
78,9	6225,21
79,2	6272,64
78,9	6225,21
395,4	31268,42

Tabelle 9.3

x_j	$x_j{}^*$	$x_j{}^{*2}$
79,4	0,4	0,16
79,0	0,0	0,00
78,9	−0,1	0,01
79,2	0,2	0,04
78,9	−0,1	0,01
	0,4	0,22

Manchmal erzielt man noch weitere Rechenvorteile, indem man eine Nullpunktsverschiebung und zugleich eine Änderung der Skaleneinheit, also insgesamt eine „**lineare Transformation**"

(9.5) $$x_j = c_1 x_j^* + c_2 \qquad \left[\text{also } x_j^* = \frac{1}{c_1}(x_j - c_2)\right]$$

ausführt. Die Konstanten c_1 und c_2 wählt man dabei so, daß die transformierten Werte $x_1^*, \ldots, x_n^*$ möglichst einfache Zahlen sind. Man berechnet dann zuerst den Mittelwert $\bar{x}^*$ und die Varianz s^{*2} der transformierten Stichprobe und daraus weiter den Mittelwert

(9.6) $$\bar{x} = c_1 \bar{x}^* + c_2$$

der gegebenen Stichprobe sowie deren Varianz

(9.7) $$s^2 = c_1^2 s^{*2}.$$

Den Beweis dieser beiden Formeln überlassen wir dem Leser (vgl. Aufgabe 9.9). Dieses Rechenverfahren heißt „**Verschlüsselung**".

Beispiel 9.2. In Tab. 9.4 bedeutet x_j die Punktezahl von 6 Studenten bei einer Mathematikklausur im Sommer 1962. Weniger als 60 Punkte war ungenügend. 100 Punkte konnte man höchstens erreichen. Gefragt ist nach dem Mittelwert und der Varianz der Stichprobe.

Da die Stichprobenwerte ganzzahlige Vielfache von 10 sind, wollen wir

$$x_j = 10 x_j^* + 20 \qquad (\text{also } x_j^* = 0{,}1 x_j - 2)$$

setzen. Aus Tab. 9.4 erhalten wir dann den Mittelwert

$$\bar{x}^* = \frac{21}{6} = 3{,}5$$

und die Varianz

$$s^{*2} = \frac{1}{5}\left(111 - \frac{21^2}{6}\right) = \frac{37{,}5}{5} = 7{,}5.$$

Also hat die gegebene Stichprobe gemäß (9.6) den Mittelwert

$$\bar{x} = 10 \cdot 3{,}5 + 20 = 55$$

und gemäß (9.7) die Varianz

$$s^2 = 10^2 \cdot 7{,}5 = 750.$$

Tabelle 9.4. Zu Beispiel 9.2

x_j	x_j^*	x_j^{*2}
20	0	0
30	1	1
50	3	9
60	4	16
80	6	36
90	7	49
	21	111

Aufgaben zu Abschnitt 9

9.1 Man gewinne (9.1) aus (8.3).

9.2 Man gewinne (9.2) aus (9.1).

9.3 Man beweise (9.4).

9.4 Man beweise, daß sich die Varianz einer Stichprobe bei einer Nullpunktsverschiebung nicht ändert.

9.5—**9.8** Man berechne den Mittelwert und die Varianz der folgenden Stichproben auf möglichst einfache Weise.

9.5 2 4 6 8.

9.6 Punktbewertung von 15 Personen bei einem Gedächtnistest (S. M. NEWHALL und M. H. HEIM, Journ. Applied Psychology **13**, 1929, 62. Jede Person betrachtete dabei verschiedene Reklameinserate für je 5 Sekunden Dauer und mußte dann später die Handelsnamen der inserierten Produkte nennen. Die größte mögliche Punktezahl betrug 81).

12 21 21 23 27 28 30 34 37 39 39 39 40 49 56

9.7 Die Stichprobe in Aufgabe 8.3.

9.8 Die Stichprobe in Aufgabe 4.2.

9.9 Man beweise (9.6) und (9.7).

9.10 Eine Zahl, die in einer Stichprobe am häufigsten vorkommt, heißt ein **häufigster Wert** der Stichprobe. Man bestimme die häufigsten Werte der Stichprobe in Tab. 4.3.

9.11 Ordnet man die Werte einer Stichprobe der Größe nach, so steht, wenn der Stichprobenumfang n eine ungerade Zahl ist, ein Stichprobenwert genau in der Mitte dieser Anordnung. Dieser Wert heißt der **Medianwert** (auch *Zentralwert*) der Stichprobe. Ist n gerade, so stehen in der genannten Anordnung zwei Stichprobenwerte am weitesten in der Mitte, und meistens wird dann deren arithmetisches Mittel als Medianwert bezeichnet. Welchen Medianwert haben die Stichproben in Beispiel 8.1 und 8.2?

10 Berechnung des Mittelwertes und der Varianz aus der Häufigkeitsfunktion

Beispiel 10.1. Zehn blaue Holznägel hatten die Länge [mm]

20 21 21 22 21 22 20 22 21 21

Gemäß (8.1) hat diese Stichprobe den Mittelwert

$$\bar{x} = \frac{1}{10}(20 \cdot 2 + 21 \cdot 5 + 22 \cdot 3) = 21{,}1 \text{ [mm]}.$$

Wie wir sehen, kommen in der Summe in den Klammern alle *zahlenmäßig* verschiedenen Werte unserer Stichprobe vor, und jeder solche Wert ist mit seiner absoluten Häufigkeit multipliziert. Wir können auch $1/n = 1/10$ hineinmultiplizieren und erhalten dann

$$\bar{x} = 20 \cdot 0{,}2 + 21 \cdot 0{,}5 + 22 \cdot 0{,}3 .$$

Nun kommt jeder *zahlenmäßig* verschiedene Wert unserer Stichprobe vor, und jeder solche Wert x_j ist mit seiner relativen Häufigkeit multipliziert, also mit dem zugehörigen Wert $\tilde{f}(x_j)$ der Häufigkeitsfunktion. [Die absolute Häufigkeit ist $n\,\tilde{f}(x_j) = 10\,\tilde{f}(x_j)$.]

Ganz entsprechend ergibt sich aus (8.3) die Formel

$$s^2 = \frac{1}{9}\left[(20-21{,}1)^2 \cdot 2 + (21-21{,}1)^2 \cdot 5 + (22-21{,}1)^2 \cdot 3\right] = 0{,}54 \text{ [mm}^2\text{]}.$$

Liegt nun allgemein eine Stichprobe von n Werten vor, unter denen *m zahlenmäßig verschiedene* sind, so können wir die Numerierung so wählen, daß

$$x_1, x_2, \ldots, x_m \qquad (m \leqq n)$$

gerade diese zahlenmäßig verschiedenen Werte sind. Dann gilt

$$(10.1) \qquad \bar{x} = \frac{1}{n} \sum_{j=1}^{m} x_j \, n \tilde{f}(x_j) = \sum_{j=1}^{m} x_j \tilde{f}(x_j),$$

$$(10.2) \qquad s^2 = \frac{1}{n-1} \sum_{j=1}^{m} (x_j - \bar{x})^2 \, n \tilde{f}(x_j) = \frac{n}{n-1} \sum_{j=1}^{m} (x_j - \bar{x})^2 \tilde{f}(x_j).$$

Man beachte: $n\tilde{f}(x_j)$ ist die absolute Häufigkeit von x_j. In (8.1) und (8.3) liefert *jeder* Stichprobenwert einen Summanden. Dagegen gibt es nun nur so viele Summanden, wie *zahlenmäßig verschiedene* Werte in der Stichprobe vorkommen.

Ganz entsprechend gewinnt (9.1) nun die Form

$$s^2 = \frac{1}{n-1} \left[\sum_{j=1}^{m} x_j^2 n \tilde{f}(x_j) - \frac{1}{n} \left(\sum_{j=1}^{m} x_j n \tilde{f}(x_j) \right)^2 \right].$$

Beim praktischen Rechnen sind die absoluten Häufigkeiten meist bequemer als die relativen. In (9.6) wird nun

$$\bar{x}^* = \frac{1}{n} \sum_{j=1}^{m} x_j^* \, n \tilde{f}(x_j^*) = \sum_{j=1}^{m} x_j^* \tilde{f}(x_j^*)$$

und so weiter. Dabei ist selbstredend $\tilde{f}(x_j^*) = \tilde{f}(x_j)$.

Beispiel 10.2. Tab. 10.1 zeigt, wie man den Mittelwert und die Varianz der Stichprobe in Tab. 4.3 berechnen kann. In der ersten Spalte stehen die zahlenmäßig verschiedenen Stichprobenwerte, in der zweiten die transformierten Werte

$$x_j^* = x_j - 160$$

und in der dritten die absoluten Häufigkeiten. Aus der vierten Spalte und (10.1) folgt

$$\bar{x}^* = \frac{505}{100} = 5{,}05 \text{ [cm]}.$$

Wegen (9.4) mit $c = 160$ hat die Stichprobe also den Mittelwert

$$\bar{x} = 160 + 5{,}05 = 165{,}05 \text{ [cm]}.$$

Die Varianz ist dieselbe wie die der transformierten Werte. Aus der vierten und der letzten Spalte ergibt sich damit

$$s^2 = \frac{1}{99} \left(5947 - \frac{505^2}{100} \right) = 34{,}31 \text{ [cm}^2\text{]}.$$

Tabelle 10.1. Rechnung zu Beispiel 10.2

x_j	x_j^*	$100\tilde{f}(x_j)$	$100x_j^*\tilde{f}(x_j)$	x_j^{*2}	$100x_j^{*2}\tilde{f}(x_j)$
153	−7	1	− 7	49	49
154	−6	1	− 6	36	36
155	−5	2	−10	25	50
156	−4	3	−12	16	48
157	−3	3	− 9	9	27
158	−2	5	−10	4	20
159	−1	6	− 6	1	6
160	0	4	0	0	0
161	1	5	5	1	5
162	2	7	14	4	28
163	3	5	15	9	45
164	4	5	20	16	80
165	5	6	30	25	150
166	6	7	42	36	252
167	7	5	35	49	245
168	8	4	32	64	256
169	9	5	45	81	405
170	10	5	50	100	500
171	11	6	66	121	726
172	12	4	48	144	576
173	13	3	39	169	507
174	14	2	28	196	392
175	15	3	45	225	675
176	16	1	16	256	256
177	17	1	17	289	289
178	18	1	18	324	324
	Summe	100	505		5947

Im Falle der **Klassenbildung** treten die ursprünglichen Stichprobenwerte nicht mehr einzeln in Erscheinung. Man tut dann so, als ob jeweils alle Werte, die man zu einer Klasse zusammengefaßt hat, in der zugehörigen Klassenmitte lägen (vgl. Abschn. 6). Wie aus (10.1) bzw. (10.2) verständlich wird, definiert man den Mittelwert einer in Klassen eingeteilten Stichprobe demgemäß durch

$$\bar{x} = \frac{1}{n}\sum_{j=1}^{K} x_j\, n\tilde{f}(x_j) = \sum_{j=1}^{K} x_j\,\tilde{f}(x_j) \tag{10.3}$$

und die Varianz durch

$$s^2 = \frac{1}{n-1}\sum_{j=1}^{K}(x_j-\bar{x})^2\, n\tilde{f}(x_j) = \frac{n}{n-1}\sum_{j=1}^{K}(x_j-\bar{x})^2\tilde{f}(x_j)\,. \tag{10.4}$$

Hierbei ist K die Anzahl der Klassen, x_j die Mitte des j-ten Klassenintervalls und $\tilde{f}(x_j)$ die zugehörige relative Klassenhäufigkeit. $\tilde{f}(x_j)$ ist also die Häufigkeitsfunktion (vgl. Abschn. 6).

Bei der praktischen Berechnung kann man die Klassenmitten verschlüsseln, wie das folgende Beispiel illustriert.

Beispiel 10.3. Die Stichprobe in Tab. 10.1 läßt sich in Klassen einteilen, wie Tab. 10.2 zeigt. Wir setzen

$$x_j = 170 + 5x_j^* \qquad \text{(also } x_j^* = 0{,}2x_j - 34\text{)}.$$

Dann erhalten wir $\bar{x}^* = -100/100 = -1$, also

$$\bar{x} = 170 + 5 \cdot (-1) = 165 \text{ [cm]}.$$

Dies weicht um 0,05 cm vom Mittelwert der ursprünglichen Stichprobe ab (vgl. Beispiel 10.2). Die Abweichung ist durch die Klassenbildung bedingt. Sie kann in anderen Fällen auch größer sein. Vgl. Aufgabe 10.8.

Die Varianz der verschlüsselten Werte ist

$$s^{*2} = \frac{1}{99}\left(240 - \frac{100^2}{100}\right) = \frac{140}{99} = 1{,}414.$$

Mit (9.7) folgt hieraus

$$s^2 = 5^2 \cdot 1{,}414 = 35{,}35 \text{ [cm}^2\text{]}.$$

Dies weicht infolge der Klassenbildung von dem Ergebnis im vorigen Beispiel ab.

Tabelle 10.2. Zu Beispiel 10.3

Klassen-intervall	Klassen-mitte x_j	Absolute Klassen-häufigkeit $100\tilde{f}(x_j)$	x_j^*	$100x_j^*\tilde{f}(x_j)$	x_j^{*2}	$100x_j^{*2}\tilde{f}(x_j)$
152,5 – 157,5	155	10	−3	−30	9	90
157,5 – 162,5	160	27	−2	−54	4	108
162,5 – 167,5	165	28	−1	−28	1	28
167,5 – 172,5	170	24	0	0	0	0
172,5 – 177,5	175	10	1	10	1	10
177,5 – 182,5	180	1	2	2	4	4
	Summe	100		−100		240

Aufgaben zu Abschnitt 10

10.1 Man beweise (10.1) und (10.2).

10.2 – 10.5 Man berechne den Mittelwert und die Varianz der folgenden Stichproben auf möglichst einfache Weise.

10.2 Die Stichprobe in Aufgabe 4.5.

10.3 Die Stichprobe in Aufgabe 5.3.

10.4 PS-Zahl eines Zweitaktmotors bei verschiedenen Kraftstoffen (H. Luther, H. Ihrig u. H. Lies, Bergbauwiss. **10**, 1963, 267)

3,20 3,05 3,15 3,10 3,20 3,05 3,00 2,95 3,15.

10.5 Lebendsauer von Glühlampen in Stunden (A. K. Gupta, Biometrika 39, 1952, 260)

Lebensdauer	Häufigkeit
950—1050	4
1050—1150	9
1150—1250	19
1250—1350	36
1350—1450	51
1450—1550	58

Lebensdauer	Häufigkeit
1550—1650	53
1650—1750	37
1750—1850	20
1850—1950	9
1950—2050	3
2050—2150	1

10.6 Man beweise: Der Bruchteil der Stichprobenwerte, die von $\bar{x}$ um mindestens M abweichen, ist höchstens gleich s^2/M^2.

10.7 Man teile die Stichprobe

3,6	2,9	3,2	3,4	3,5	3,1	3,0	3,0	3,4	3,1
2,9	2,8	3,1	3,2	3,8	3,5	3,1	3,5	3,4	3,2

(Geburtsgewicht [kg] ausgetragener Mädchen, Univ.-Frauenklinik Graz, Direktor Prof. Dr. E. Navratil, 1962) unter Verwendung der folgenden Klassenintervalle in Klassen ein:

(a) 2,65—2,95 2,95—3,25 3,25—3,55 3,55—3,85
(b) 2,75—3,05 3,05—3,35 3,35—3,65 3,65—3,95
(c) 2,75—3,25 3,25—3,75 3,75—4,25.

Man berechne jeweils den Mittelwert und vergleiche diesen mit dem Mittelwert der ursprünglichen Stichprobe.

10.8 Man beweise: Teilt man eine Stichprobe unter Benutzung gleichlanger Klassenintervalle der Länge l in Klassen ein, so kann sich der Mittelwert dabei um höchstens $l/2$ ändern.

11 Analogie zwischen Häufigkeits- und Massenverteilungen

Zwischen den Häufigkeitsverteilungen von Stichproben und den Verteilungen von Massenpunkten mit der Gesamtmasse 1 längs einer Geraden besteht eine vollkommene quantitative Analogie. Wir erwähnen diese, weil sie dem Leser, der mit den Grundbegriffen der elementaren Mechanik vertraut ist, zu einem besseren Verständnis unserer bisherigen Überlegungen verhilft.

Vorgelegt sei eine Stichprobe, in der insgesamt m zahlenmäßig verschiedene Werte $x_1, x_2, \ldots, x_m$ vorkommen mögen, und zwar mit den relativen Häufigkeiten

$$\tilde{f}(x_1) = h_1, \ldots, \tilde{f}(x_m) = h_m.$$

Das mechanische Analogon dieser Stichprobe ist eine Verteilung, bestehend aus m Massen, nämlich der Masse h_1 im Punkte x_1 der x-Achse, der Masse h_2 im Punkte x_2 usw. Der Summenhäufigkeitsfunktion (7.1) entspricht die Funktion

$$\tilde{F}(x) = \sum_{x_j \leq x} h_j.$$

Für jedes x ist dies die Summe der Massen, die in dem Punkte x oder links davon liegen.

Die genannte Massenverteilung hat den Schwerpunkt

$$\bar{x} = \sum_{j=1}^{m} x_j h_j.$$

Das Trägheitsmoment bezüglich des Schwerpunktes ist

$$T = \sum_{j=1}^{m} (x_j - \bar{x})^2 h_j.$$

Wegen (10.1) bzw. (10.2) gilt also:

Der Mittelwert einer Stichprobe entspricht in der Mechanik dem Schwerpunkt der analogen Massenverteilung. Die Varianz entspricht im wesentlichen [d. h. bis auf einen Faktor $n/(n-1)$] *dem Trägheitsmoment bezüglich des Schwerpunktes.*

Ferner ist bekanntlich das statische Moment bezüglich des Schwerpunktes gleich null. Dies ist das mechanische Analogon unserer Formel (8.2).

TEIL II

WAHRSCHEINLICHKEITSTHEORIE

Bisher haben wir uns überlegt, wie man statistische Daten in eine übersichtliche tabellarische oder graphische Form bringt und durch gewisse Maßzahlen kennzeichnet. Wollten wir uns damit begnügen, so wären wir schon am Ende unserer Überlegungen angelangt.

Tatsächlich wollen wir aber von dieser rein beschreibenden Statistik zur mathematischen Statistik fortschreiten. Das heißt, wir wollen von der bloßen Beschreibung statistischen Materials zu dessen Beurteilung übergehen, indem wir Schlüsse von Stichproben auf die zugehörigen Grundgesamtheiten ziehen. Hierzu müssen wir mathematische Modellvorstellungen schaffen und dann Verbindungen zu den empirischen Beobachtungen herstellen, also Zusammenhänge zwischen der Theorie und der Wirklichkeit ermitteln.

Den Ausgangspunkt für dieses Vorhaben bildet die sogenannte mathematische Wahrscheinlichkeitstheorie, die damit die Grundlage der mathematischen Statistik darstellt. So ist es notwendig, daß wir uns nun zunächst einmal mit der genannten Theorie vertraut machen, soweit wir sie für unsere Zwecke brauchen.

Im vorliegenden II. Teil betrachten wir zuerst die Grundbegriffe der Wahrscheinlichkeitstheorie, insbesondere den Begriff der Wahrscheinlichkeit. Dann kommen verschiedene wichtige Wahrscheinlichkeitsverteilungen zur Sprache. Diese Verteilungen dienen als theoretische Modelle von Grundgesamtheiten, im Gegensatz zu den bisher betrachteten Häufigkeitsverteilungen von Stichproben, die empirische Verteilungen sind.

Die mathematische Wahrscheinlichkeitstheorie wird uns übrigens auch die mathematische Begründung der statistischen Methoden in Teil III liefern helfen.

Kapitel 4

Grundbegriffe

Historisch entsprang die mathematische Wahrscheinlichkeitstheorie dem Wunsch, die Gewinnaussichten bei Glücksspielen zu berechnen (Jakob Bernoulli, Ars conjectandi, 1713. Abraham de Moivre, The Doctrine of Chances, 1718. Pierre Simon de Laplace, Théorie analytique des probabilités, 1812). Ihre Anwendbarkeit auf andere Gebiete (Versicherungswesen, Meßfehlertheorie) zeigte sich gar bald.

Die genannte Theorie soll ein mathematisches Modell beobachtbarer empirischer Sachverhalte der Wirklichkeit darstellen, bei denen der Zufall eine Rolle spielt. Dementsprechend müssen wir den Begriff der mathematischen Wahrscheinlichkeit definieren. Im Bereich der Spiele ist das einfach. Dort benützt man den sogenannten klassischen Wahrscheinlichkeitsbegriff von Laplace (Abschn. 14). Leider ist dieser Begriff für die meisten statistischen Zwecke zu speziell. Wie man einen für die Statistik geeigneten Wahrscheinlichkeitsbegriff einführt, der übrigens den klassischen mitumfaßt, überlegen wir uns in Abschn. 15.

12 Zufallsexperiment. Ereignis. Häufigkeit

Im vorliegenden Abschnitt erläutern wir einige grundlegende Begriffe, die zum Teil bereits früher erwähnt wurden.

Unter einem *Zufallsexperiment* oder einer *Zufallsbeobachtung* — kurz: **Experiment** oder **Beobachtung** — verstehen wir einen beliebig oft wiederholbaren Vorgang, der nach einer ganz bestimmten Vorschrift ausgeführt wird und dessen Ergebnis „vom Zufall abhängt", das soll heißen, nicht im voraus eindeutig bestimmt werden kann.

Beispiele sind das Werfen eines Würfels, das Ziehen einer Spielkarte, das in Abschn. 3 beschriebene Experiment und ähnliche technische Messungen, die zufällige Auswahl einer Person und die Feststellung ihrer Größe, ihres Blutdrucks, ihres täglichen Zigarettenverbrauchs oder ihrer Ansicht über den Wert zeitgenössischer Musik, die zufällige Auswahl einer Kuh und die Bestimmung ihres Milchertrages, das zufällige Herausgreifen und Prüfen einer Glühbirne aus einer Produktion usw.

Bei jedem solchen Experiment können wir gewisse **Ereignisse** unterscheiden, deren Eintreffen vom Zufall abhängt.

Zum Beispiel können wir beim Würfeln die 6 Ereignisse des Würfelns einer

1, 2, 3, 4, 5 oder 6

betrachten. Beim Münzenwurf können wir zwischen den Ereignissen

„Kopf" oder „Wappen"

unterscheiden, bei der Prüfung einer Glühbirne zwischen

„Brauchbare Birne" oder „Unbrauchbare Birne" (Ausschuß)

usw.

Erzielen wir bei einem Münzenwurf „Kopf", so sagen wir, bei der betreffenden Ausführung des Experimentes (d. h. bei dem betreffenden Wurf) sei das Ereignis „Kopf" **eingetroffen** oder **realisiert**. Ganz entsprechend kümmert man sich in anderen Fällen jeweils um das Eintreffen oder Nichteintreffen von Ereignissen.

Ereignisse bezeichnen wir im folgenden mit großen kursiven Buchstaben, etwa A, B, C usw.

Wir führen irgendein Zufallsexperiment n-mal nacheinander aus. Trifft dabei ein Ereignis A genau k-mal ein, so heißt k die **absolute Häufigkeit** und k/n die **relative Häufigkeit** des Ereignisses A bei dieser Versuchsreihe. Diese relative Häufigkeit bezeichnen wir mit $h(A)$. Es ist also

$$\text{(12.1)} \qquad \boxed{h(A) = \frac{k}{n} = \frac{\text{Anzahl der Versuche, bei denen } A \text{ eintrifft}}{\text{Gesamtzahl der Versuche in der Reihe}}.}$$

Da k in jedem Falle mindestens gleich 0 sein muß und höchstens gleich n sein kann, so gilt stets

$$\text{(12.2)} \qquad \boxed{0 \leq h(A) \leq 1\,.}$$

Die relative Häufigkeit ist also eine reelle nicht negative Zahl, die höchstens gleich 1 sein kann.

In dem Sonderfall, daß ein Ereignis A durch eine Zahl a gekennzeichnet ist (Beispiel: Würfelwurf), wird unsere gegenwärtige Betrachtung natürlich mit der in Abschn. 4 durchgeführten identisch.

Beispiel 12.1. An der Technischen Hochschule Graz werden die Prüfungsnoten

Vorzüglich Sehr Gut Gut Genügend Ungenügend

benutzt, die wir der Reihe nach mit A, B, C, D und E bezeichnen wollen. Aus der Kartei aller im Jahre 1962 am 1. Mathematischen Institut geprüften Studenten wurde eine Stichprobe entnommen, indem einige Karten ganz zufällig gezogen wurden. Die Noten auf diesen Karten lauteten:

Note	A	B	C	D	E
Absolute Häufigkeit	2	4	8	3	8

Wie wir sehen, ist in dieser Stichprobe

$$h(A) = \frac{2}{25}, \qquad h(B) = \frac{4}{25}, \qquad h(C) = \frac{8}{25}$$

usw.

13 Summe und Produkt von Ereignissen

A und B seien Ereignisse, die bei einem Zufallsexperiment eintreffen können. Dann bezeichnen wir mit

$$A + B$$

das Ereignis, das genau dann eintrifft, wenn A oder B eintrifft oder wenn beide Ereignisse gleichzeitig eintreffen. Das Ereignis $A + B$ heißt die **Summe** der Ereignisse A und B.

Weiterhin bezeichnen wir mit

$$AB$$

das Ereignis, das genau dann eintrifft, wenn A und B gleichzeitig eintreffen. AB heißt das **Produkt** der Ereignisse A und B.

Nun kann es aber sein, daß A und B bei dem betreffenden Experiment gar nicht gleichzeitig eintreffen können. Dann heißen A und B **einander ausschließende Ereignisse** oder **sich gegenseitig ausschließende Ereignisse.**

Beispiel 13.1. Die Geburt eines einzelnen Kindes können wir als ein Experiment auffassen, bei dem zwei Ereignisse möglich sind:

A: *Geburt eines Mädchens,*
B: *Geburt eines Jungen.*

Diese Ereignisse A und B schließen einander aus. AB trifft also nie ein, während $A + B$ (Geburt eines Jungen oder Mädchens) stets eintrifft.

Die Summe $A + B$ von Ereignissen A und B heißt auch **Vereinigung** der Ereignisse A und B, und das Produkt AB heißt auch **Durchschnitt** von A und B. Das hängt mit der Mengenlehre zusammen. Der mit der Mengenlehre etwas vertraute Leser findet Näheres dazu in Anhang 1.

Das Bilden von Summen und Produkten von Ereignissen kann man sich durch geeignete Skizzen erleichtern. Wir erläutern das Prinzip an einem einfachen Beispiel.

Beispiel 13.2. Wir würfeln mit einem Würfel und betrachten die Ereignisse

A: *Würfeln einer geraden Zahl*,
B: *Würfeln einer durch 3 teilbaren Zahl*,
C: *Würfeln einer Eins*.

Um Summen und Produkte dieser Ereignisse zu bilden, kann man zuerst jedes der 6 möglichen Würfelergebnisse 1, 2, 3, 4, 5, 6 durch einen Punkt in der Ebene darstellen. Auf die Anordnung dieser 6 Punkte kommt es nicht an. Dann wird A, B und C durch je einen Bereich dargestellt, der die entsprechenden Punkte enthält, wie Abb. 13.1 zeigt. Der Summe

$A + B$: *Würfeln einer geraden oder einer durch 3 teilbaren Zahl*

entspricht dann der gestrichelt umrandete Bereich. Dieser enthält 2, 3, 4, 6. Das Ereignis $A + B$ trifft also dann und nur dann ein, wenn eine 2, 3, 4 oder 6 gewürfelt wird. Weiterhin entspricht das Produkt

AB: *Würfeln einer geraden durch 3 teilbaren Zahl*

dem schraffierten Bereich, der gleichzeitig den beiden A und B entsprechenden Bereichen ganz angehört. Wie wir sehen, trifft AB dann und nur dann ein, wenn eine Sechs gewürfelt wird. Entsprechend liest man aus der Abb. 13.1 ab, daß sich A und C gegenseitig ausschließen, weil keiner der 6 genannten Punkte zugleich in dem zu A und dem zu C gehörigen Bereich liegt.

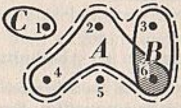


Abbildung 13.1. Graphische Darstellung von Summen und Produkten von Ereignissen. (Zu Beispiel 13.2)

A und B seien Ereignisse, die in einer Versuchsreihe mit der relativen Häufigkeit $h(A)$ bzw. $h(B)$ aufgetreten sind. Dann hat, wie wir zeigen werden, das Ereignis $A + B$ die relative Häufigkeit (vgl. Abb. 13.2)

$$(13.1) \qquad \boxed{h(A + B) = h(A) + h(B) - h(AB)\,.}$$

Abbildung 13.2. Zu Formel (13.1)

Schließen sich A und B gegenseitig aus, so ist $h(AB) = 0$ und

$$(13.2) \qquad \boxed{h(A + B) = h(A) + h(B)\,.}$$

In Worten:

Bei sich gegenseitig ausschließenden Ereignissen A und B ist die relative Häufigkeit des Ereignisses $A + B$ gleich der Summe der relativen Häufigkeiten von A und B.

Wir beweisen (13.1). Die Versuchsreihe bestehe aus n Versuchen. Hinsichtlich der n Ergebnisse gibt es 4 Fälle:

(F 1) A und B sind gleichzeitig eingetroffen.
(F 2) A ist eingetroffen, B nicht.
(F 3) B ist eingetroffen, A nicht.
(F 4) Weder A noch B ist eingetroffen.

(F 1) komme in der Reihe n_1-mal vor, (F 2) komme n_2-mal vor, usw. Die Fälle schließen sich gegenseitig aus. Deshalb ist insgesamt

A genau $n_1 + n_2$-mal, $\quad B$ genau $n_1 + n_3$-mal
AB genau n_1-mal, $\quad A + B$ genau $n_1 + n_2 + n_3$-mal

eingetroffen. So erhalten wir die relativen Häufigkeiten

$$h(A) = \frac{n_1 + n_2}{n}, \qquad h(B) = \frac{n_1 + n_3}{n}, \tag{13.3}$$
$$h(AB) = \frac{n_1}{n}, \qquad h(A + B) = \frac{n_1 + n_2 + n_3}{n}.$$

Hieraus folgt (13.1) unmittelbar.

Aufgaben zu Abschnitt 13

13.1 Um eine Sendung von 1000 Bakelitringen für Telefonhörer zu prüfen, wurden 50 Ringe ganz zufällig herausgegriffen. Einer davon hatte Kratzer und Sprünge, 3 hatten nur Kratzer und 2 hatten nur Sprünge. Wie groß ist die relative Häufigkeit der Ereignisse

A = *Ring mit Sprüngen*, $\quad B$ = *Zerkratzter Ring*,

AB und $A + B$?

13.2 Man zeige: Sind $A_1, A_2, \ldots, A_k$ alle möglichen Ereignisse, die bei einem bestimmten Experiment eintreffen können, und schließen sich diese Ereignisse gegenseitig aus, so gilt für die relativen Häufigkeiten bei mehrmaliger Ausführung des Experimentes

$$h(A_1) + h(A_2) + \cdots + h(A_k) = 1.$$

13.3 In (13.3) stehen relative Häufigkeiten, die sich auf *alle* n Ausführungen beziehen. Mit $h(B \mid A)$ bezeichnen wir die relative Häufigkeit von B unter den Ausführungen, bei denen A realisiert ist. Entsprechend bezeichnet $h(A \mid B)$ die relative Häufigkeit von A unter den Ausführungen, bei denen B realisiert ist. Man zeige, daß

$$h(B \mid A) = \frac{n_1}{n_1 + n_2}, \qquad h(A \mid B) = \frac{n_1}{n_1 + n_3} \tag{13.4}$$

ist, und gewinne hieraus vermöge (13.3)

$$h(A)\, h(B \mid A) = h(AB), \qquad h(B)\, h(A \mid B) = h(AB). \tag{13.5}$$

13.4 Wie groß ist $h(A \mid B)$ und $h(B \mid A)$ in Aufgabe 13.1?

13.5 K. PEARSON (Phil. Trans. (A) **195**, 1900, 138) beobachtete die folgenden Häufigkeiten von Augenfarben bei Vätern und Söhnen:

		Vater	
		hell	dunkel
Sohn	hell	471	148
	dunkel	151	230

Wie groß ist die relative Häufigkeit der Ereignisse

$$A = \textit{Vater helläugig}, \quad B = \textit{Sohn helläugig},$$

AB und $A + B$? Wie groß ist $h(B \mid A)$?

14 Der klassische mathematische Wahrscheinlichkeitsbegriff

Beim Werfen eines Würfels führt jeder Wurf zu einer der Zahlen

$$1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6.$$

Diese 6 Ereignisse schließen einander aus. Ist der Würfel **regelmäßig,** d. h. aus homogenem Material und von genau kubischer Gestalt, so ist bei ordnungsgemäßem Würfeln keines dieser 6 Ereignisse hinsichtlich des Eintreffens vor dem anderen ausgezeichnet. Wir sagen, es gibt bei diesem Zufallsexperiment 6 *gleichmögliche* oder *gleichwahrscheinliche Ereignisse* oder kürzer 6 *gleichmögliche Fälle.*

In ähnlicher Weise kann man auch bei anderen Spielen endlich viele **gleichmögliche Fälle** angeben, das heißt endlich viele einander ausschließende gleichwahrscheinliche Ereignisse, die alle Möglichkeiten erschöpfen.

Gibt es nun bei einem Spiel insgesamt m gleichmögliche Fälle, so kann ein Spieler S diese m Fälle in zwei Klassen einteilen, nämlich in die für ihn günstigen Fälle, bei deren Eintreffen er gewinnt, und in die für ihn ungünstigen Fälle, bei deren Eintreffen er verliert. Sind g Fälle günstig für S, so ist die Zahl g/m offenbar ein Maß für die Gewinnaussicht des Spielers S bei dem betrachteten Spiel. Dieser Gedankengang legt die folgende Definition nahe, die von LAPLACE stammt.

Klassische Definition der mathematischen Wahrscheinlichkeit. *Die Wahrscheinlichkeit $P(A)$ eines Ereignisses A bei einem Zufallsexperiment ist durch*

(14.1) $$\boxed{P(A) = \frac{g}{m}}$$

gegeben. Hierbei ist

g die Anzahl der Fälle, bei denen A eintrifft, und
m die Anzahl aller gleichmöglichen Fälle

bei dem betreffenden Experiment.

Beispiel 14.1. Die Wahrscheinlichkeit, bei einem Wurf mit einem regelmäßigen Würfel eine Sechs zu würfeln, beträgt $1/_6$, denn 6 Fälle sind möglich, und bei einem dieser Fälle trifft das Ereignis „Sechs" ein.

Beim **Ziehen aus einem Kartenspiel** nehmen wir stets an, daß die Wahrscheinlichkeit, eine bestimmte Karte zu ziehen, für alle Karten dieselbe ist, ohne daß dies jedesmal wieder ausdrücklich gesagt wird. Entsprechend nehmen wir beim **Ziehen aus einer Urne** an, daß die Wahrscheinlichkeit, einen bestimmten Gegenstand zu ziehen, für alle Gegenstände in der Urne dieselbe ist.

Beispiel 14.2. Wie groß ist die Wahrscheinlichkeit P, aus einer Urne bei einmaligem Ziehen eine rote Kugel zu erhalten, wenn sich insgesamt 3 rote und 7 schwarze Kugeln in der Urne befinden?

Unter den 10 gleichmöglichen Fällen sind 3 Fälle, bei denen das betrachtete Ereignis eintrifft. Also ist $P = 0{,}3 = 30\,\%$.

Beispiel 14.3. Wie groß ist die Wahrscheinlichkeit P, bei einem einzelnen Wurf mit 2 regelmäßigen Würfeln gleichzeitig zwei gerade Zahlen zu würfeln?

Es gibt insgesamt 36 gleichmögliche Fälle, nämlich $(1, 1)$, $(1, 2)$, $\ldots$, $(1, 6)$, $(2, 1)$, ... usw. Hierbei bezeichnet die erste Zahl jeweils das mit dem ersten Würfel erzielte Ergebnis und die zweite Zahl das mit dem zweiten Würfel erzielte Ergebnis. Bei 9 dieser 36 Fälle, nämlich bei den Fällen

$$(2, 2)\quad (2, 4)\quad (2, 6)\quad (4, 2)\quad (4, 4)\quad (4, 6)\quad (6, 2)\quad (6, 4)\quad (6, 6)$$

trifft das fragliche Ereignis ein. Also ist $P = 9/36 = 1/4$.

Kann man dieses Ergebnis auch einfacher erhalten, indem man nur 4 gleichmögliche Fälle betrachtet?

Wie wir sehen, besteht die klassische Definition der mathematischen Wahrscheinlichkeit darin, daß der Begriff der Wahrscheinlichkeit auf den Begriff der Gleichwahrscheinlichkeit (Gleichmöglichkeit) zurückgeführt wird. Letzterer gilt als grundlegend und unterliegt keiner weiteren Definition.

Damit ist klar, daß der klassische Wahrscheinlichkeitsbegriff nur bei Zufallsexperimenten angewendet werden kann, bei denen die Natur des Experimentes einleuchtende Anhaltspunkte für die Einteilung aller möglichen Ereignisse in endlich viele gleichwahrscheinliche Fälle liefert, denn die Definition dieses Begriffes gibt selbst keine derartigen Anhaltspunkte. Nun sind aber viele Zufallsexperimente nicht von der genannten Art, und der klassische Wahrscheinlichkeitsbegriff versagt dann.

Dies gilt schon für kompliziertere Spiele, z. B. für das Würfeln mit einem unsymmetrischen Würfel, und erst recht für die meisten praktisch wichtigen einschlägigen Probleme der Naturwissenschaft, Technik, Wirtschaft und anderer Gebiete. Man denke z. B. an die Frage nach der Wahrscheinlichkeit des Heilerfolges durch ein gewisses Medikament oder nach der Wahrscheinlichkeit, daß eine bestimmte Maschine Ausschuß produziert, usw. Wie soll man bei derartigen und ähnlichen Problemen zu einer Einteilung in gleichwahrscheinliche Fälle gelangen? Das ist unmöglich.

Wir brauchen also einen allgemeineren Wahrscheinlichkeitsbegriff. Einen solchen wollen wir im nächsten Abschnitt einführen. Wir folgen dabei A. N. Kolmogoroff [8], soweit dies im Rahmen einer elementaren Darstellung möglich ist. (Die Angabe [8] bezieht sich auf Anhang 3 am Schluß des vorliegenden Buches.)

Aufgaben zu Abschnitt 14

14.1—14.5 Man bestimme die Wahrscheinlichkeit der folgenden Ereignisse.

14.1 „Kopf" beim einzelnen Wurf einer Münze.

14.2 Wenigstens einen „Kopf" beim gleichzeitigen Wurf zweier Münzen.

14.3 „As" beim Ziehen einer Karte aus einem Skatspiel.

14.4 „Linksgängige Schraube" beim ganz zufälligen Herausgreifen einer Schraube aus einer Schachtel, die 20 linksgängige und 30 rechtsgängige Schrauben enthält.

14.5 Wenigstens 10 Augen beim Wurf zweier regelmäßigen Würfel.

14.6 G. E. Lessing schrieb am 15. 12. 1770 an Madame König, daß er bei der Hamburger Lotterie auf Los Nr. 19 gewonnen habe und wieder Lose genommen habe, „nur Nr. 19 nicht, wofür ich 7 gewählt habe: denn 19 wird doch nicht des Henkers sein und sich wieder herausziehen lassen". Stimmt dieser Schluß?

15 Der Wahrscheinlichkeitsbegriff in der Statistik

Mit einem Bleistift und einem Lineal können wir Dinge zeichnen, die man in der Umgangssprache Punkte und Geraden nennt. An diesen Dingen können wir dann gewisse Eigenschaften und Beziehungen empirisch feststellen. Um von dieser „empirischen Geometrie" zur Geometrie im mathematischen Sinne fortzuschreiten, sind zwei Schritte notwendig:

Zuerst postulieren wir die Existenz entsprechender mathematischer Begriffe (Punkt, Gerade, usw.).

Beim zweiten Schritt stellen wir diejenigen empirisch beobachteten Eigenschaften und Beziehungen, die uns als grundlegend erscheinen, in idealisierter Form zusammen, z. B.:

Durch zwei Punkte geht stets eine Gerade.

Durch zwei Punkte geht nicht mehr als eine Gerade. Und so weiter.

Diese Aussagen heißen *Axiome*. Die Zusammenstellung, die man auf diese Weise schließlich erhält, wird als ein *Axiomensystem* bezeichnet. Dieses bildet die Grundlage, auf der man die Geometrie aufbaut. Damit ist gemeint, daß man alle geometrischen Sätze aus diesem Axiomensystem herleitet. (Vgl. z. B. DAVID HILBERT, Grundlagen der Geometrie. 8. Auflage. Stuttgart: Teubner, 1956.) Die Axiome sind so zu wählen, daß die Sätze, die man daraus erhält, mit unserer geometrischen Anschauung im Einklang stehen. Anderenfalls ist die erhaltene Geometrie kein brauchbares mathematisches Modell empirisch beobachtbarer geometrischer Tatsachen.

Ganz entsprechend gehen wir bei der Grundlegung der Wahrscheinlichkeitstheorie vor. Den Ausgangspunkt bildet dabei die Erfahrung, daß das Eintreffen von Ereignissen bei den meisten Zufallsexperimenten auf lange Dauer gewissen Gesetzmäßigkeiten unterliegt. Insbesondere erweist sich die relative Häufigkeit eines Ereignisses in großen Versuchsreihen als nahezu konstant. Das heißt, man erhält bei mehreren solchen Reihen Werte dieser Häufigkeit, die nur wenig differieren.

Im Bereich der Glücksspiele haben dies namhafte Forscher wiederholt geprüft und bestätigt. Bei Münzenwürfen fand man z. B. die in Tab. 15.1 gezeigten Werte.

Tabelle 15.1. Münzenwürfe

	Anzahl der Würfe	Anzahl des Eintreffens des Ereignisses „Kopf“	Relative Häufigkeit
BUFFON	4040	2048	0,5069
K. PEARSON	12000	6019	0,5016
	24000	12012	0,5005

Abb. 15.1 zeigt ein Beispiel aus der Bevölkerungsstatistik. Man sieht, daß die relative Häufigkeit der Knabengeburten mit wachsender Anzahl der in zeitlicher Reihenfolge betrachteten Geburten offenbar

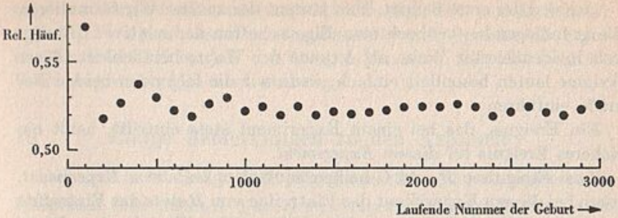


Abbildung 15.1. Relative Häufigkeit von Knabengeburten (Geburtenbuch des Standesamtes Graz, beginnend am 1. 1. 1962)

immer weniger schwankt. Schon seit langer Zeit hat man beobachtet, daß diese relative Häufigkeit bei größeren Bevölkerungsgruppen von Jahr zu Jahr nahezu konstant bleibt. Weitere Beispiele (Größenverteilung, Durchschnittsgewicht, Lebensalter von Tieren, klimatische Daten usw.) ließen sich leicht anführen.

Sind die relativen Häufigkeiten der Ereignisse bei einem Experiment in dem angegebenen Sinne nahezu konstant, so sagt man, das Experiment zeige eine **statistische Regelmäßigkeit** oder **Stabilität der relativen Häufigkeiten,** im Gegensatz zu der *zufälligen Unregelmäßigkeit* der Ergebnisse der einzelnen Versuche. Die meisten praktisch wichtigen Experimente besitzen diese Stabilitätseigenschaft.

So liegt die Vermutung nahe, daß die relative Häufigkeit eines bestimmten Ereignisses A bei oftmaliger Ausführung eines bestimmten derartigen Experimentes praktisch mit Gewißheit ungefähr gleich einer angebbaren Zahl P ist (wobei „ungefähr gleich" noch präzisiert werden muß, damit die Aussage einen Sinn erhält).

Deshalb postulieren wir nun die Existenz einer bestimmten Zahl P, die die **Wahrscheinlichkeit** des betreffenden Ereignisses A bei dem betreffenden Zufallsexperiment heißt.

Die Aussage „A hat bei dem genannten Experiment die Wahrscheinlichkeit P" bedeutet dann konkret: Bei oftmaliger Ausführung des Experimentes ist es praktisch gewiß, daß die relative Häufigkeit $h(A)$ ungefähr gleich P ist. (Dies muß noch präzisiert werden. Das können wir erst in Abschn. 51 tun.)

Die so eingeführte Wahrscheinlichkeit P ist das theoretische Gegenstück der empirischen relativen Häufigkeit.

Das war der erste Schritt. Nun kommt der zweite: Wir formulieren die grundlegenden beobachteten Eigenschaften der relativen Häufigkeit in idealisierter Weise als Axiome der Wahrscheinlichkeit. Diese Axiome lauten besonders einfach, wenn wir die folgenden beiden Begriffe einführen:

Ein Ereignis, das bei einem Experiment stets eintrifft, heißt ein **sicheres Ereignis** *bei diesem Experiment.*

Zwei Ereignisse B und C heißen **äquivalent** *bei einem Experiment,* wenn bei diesem Experiment das Eintreffen von B stets das Eintreffen von C zur Folge hat und umgekehrt. (Beispiel: „Würfeln einer Eins" und „Würfeln einer Zahl, die kleiner als 2 ist" sind äquivalente Ereignisse.)

Die Wahrscheinlichkeit von Ereignissen $A, B, \ldots$ bezeichnen wir fortan mit $P(A), P(B), \ldots$ Dabei müssen wir im Sinn behalten, daß sich diese Zahlen stets auf ein ganz bestimmtes Zufallsexperiment beziehen.

Die **Axiome der Wahrscheinlichkeit** lauten nun, wie folgt.

Axiom 1. *Die Wahrscheinlichkeit $P(A)$ eines Ereignisses A bei einem Experiment ist eine eindeutig bestimmte reelle nichtnegative Zahl, die höchstens gleich 1 sein kann,*

$$\boxed{0 \leqq P(A) \leqq 1.} \tag{15.1}$$

Axiom 2. *Für ein sicheres Ereignis S bei einem Experiment gilt*

$$P(S) = 1.$$

Für äquivalente Ereignisse B und C bei einem Experiment gilt

$$P(B) = P(C).$$

Axiom 3. *Schließen sich zwei Ereignisse A und B bei einem Experiment gegenseitig aus* (s. Abschn. 13), *so gilt bei diesem Experiment*

$$\boxed{P(A + B) = P(A) + P(B).} \tag{15.2}$$

Wie man sieht, entsprechen die Formeln (15.1) und (15.2) den Formeln (12.2) und (13.2) für die relativen Häufigkeiten.

Damit wir von Axiomen im strengen Sinne reden können, muß der Begriff des Ereignisses genau definiert werden. Dies erfordert Kenntnisse aus der sogenannten Mengenlehre und Maßtheorie und übersteigt das Niveau einer elementaren Einführung. Vgl. [2] und [8] in Anhang 3. Wer mit Mengen etwas vertraut ist, sieht, daß es sich bei den Ereignissen um Mengen handelt, daß die Wahrscheinlichkeit eine nichtnegative Mengenfunktion ist und daß wir, einer Idee von L. Schmetterer [12] folgend, den Äquivalenzbegriff von Ereignissen als Ersatz für den Gleichheitsbegriff von Mengen benutzen, der uns nicht zur Verfügung steht. Vgl. auch Anhang 1.

16 Einige Bemerkungen zu den Axiomen

Das zu einem Ereignis A **entgegengesetzte** oder **komplementäre Ereignis** $\bar{A}$ (lies „nicht A") bei einem Experiment ist definiert als das Ereignis, das bei dem Experiment genau dann eintrifft, wenn A nicht eintrifft.

A und $\bar{A}$ schließen sich gegenseitig aus. $A + \bar{A}$ ist ein sicheres Ereignis. Gemäß Axiom 2 und 3 gilt also stets

$$P(A + \bar{A}) = P(A) + P(\bar{A}) = 1.$$

Hieraus folgt die praktisch wichtige Formel

$$\boxed{P(A) = 1 - P(\bar{A}).} \tag{16.1}$$

Diese wendet man an, wenn $P(\bar{A})$ einfacher zu bestimmen ist als die gesuchte Wahrscheinlichkeit $P(A)$.

Wie wir nun sehen, genügt es übrigens, (15.1) durch

$$P(A) \geqq 0 \tag{16.2}$$

zu ersetzen. Denn $P(\bar{A})$ ist eine Wahrscheinlichkeit, also nichtnegativ, und hieraus folgt mit (16.1) sofort $P(A) \leqq 1$, die zweite Ungleichung in (15.1).

Sind $A_1, A_2, \ldots, A_m$ Ereignisse, dann bezeichnen wir mit

$$A_1 + A_2 + \cdots + A_m \tag{16.3}$$

das Ereignis, das genau dann eintrifft, wenn wenigstens eines der genannten Ereignisse eintrifft. Das Ereignis (16.3) heißt die **Summe** dieser Ereignisse.

Mit

$$A_1 A_2 \cdots A_m \tag{16.4}$$

bezeichnen wir das Ereignis, das genau dann eintrifft, wenn alle m Ereignisse $A_1, \ldots, A_m$ gleichzeitig eintreffen. Das Ereignis (16.4) heißt das **Produkt** der genannten Ereignisse.

Kommen Zweifel daran auf, ob man in einem konkreten Falle die Wahrscheinlichkeiten wirklich richtig gewählt hat, so kann man eine Nachprüfung vornehmen. Darauf gehen wir in Teil III des Buches näher ein.

18 Additionssatz für beliebige Ereignisse

Die Formel (15.2) bezieht sich auf einander ausschließende Ereignisse. Allgemeiner gilt der

Satz 18.1 (Additionssatz für beliebige Ereignisse). *Haben zwei beliebige Ereignisse A und B bei einem Experiment die Wahrscheinlichkeit $P(A)$ bzw. $P(B)$, so hat das Ereignis $A + B$ bei diesem Experiment die Wahrscheinlichkeit*

(18.1) $$\boxed{P(A + B) = P(A) + P(B) - P(AB).}$$

Beweis. Das Ereignis $A + B$ trifft genau dann ein, wenn eines der folgenden 3 Ereignisse eintrifft:

(1) AB (Gleichzeitiges Eintreffen von A und B),
(2) $A\bar{B}$ (A trifft ein, B nicht),
(3) $\bar{A}B$ (B trifft ein, A nicht).

Wir sagen, *das Ereignis $A + B$ ist in diesen 3 sich gegenseitig ausschließenden* **Formen** *realisierbar*. Also gilt gemäß (16.5) zunächst

(18.2) $$\begin{aligned} P(A + B) &= P(AB + A\bar{B} + \bar{A}B) \\ &= P(AB) + P(A\bar{B}) + P(\bar{A}B). \end{aligned}$$

Weiterhin ist A in den beiden sich gegenseitig ausschließenden Formen AB und $A\bar{B}$ realisierbar. Demnach gilt gemäß Axiom 3

$$P(A) = P(AB + A\bar{B}) = P(AB) + P(A\bar{B}).$$

Entsprechend ergibt sich

$$P(B) = P(AB + \bar{A}B) = P(AB) + P(\bar{A}B).$$

Durch Addition der letzten beiden Formeln erhalten wir

$$P(A) + P(B) = 2P(AB) + P(A\bar{B}) + P(\bar{A}B).$$

Der Vergleich der rechten Seite mit der rechten Seite in (18.2) liefert

$$P(A) + P(B) = P(A + B) + P(AB).$$

Hieraus folgt (18.1), und der Satz ist bewiesen.

Dieser Satz läßt sich auch auf mehr als zwei Ereignisse ausdehnen (vgl. Aufgabe 18.1).

Beispiel 18.1. Wie groß ist die Wahrscheinlichkeit, bei zweimaligem Würfeln mindestens eine Sechs zu erzielen?

Die Ereignisse

$$A = \textit{Sechs beim 1. Wurf}$$
$$B = \textit{Sechs beim 2. Wurf}$$

haben die Wahrscheinlichkeit $P(A) = 1/6$ und $P(B) = 1/6$. Das Ereignis

$$AB = \textit{Sechs bei beiden Würfen}$$

hat die Wahrscheinlichkeit $P(AB) = 1/36$, denn es gibt $6 \cdot 6 = 36$ gleichmögliche Fälle, und in einem davon trifft AB ein. So liefert der Satz 18.1 die Antwort

$$P(A+B) = \frac{1}{6} + \frac{1}{6} - \frac{1}{36} = \frac{11}{36} = 31\,\%.$$

Aufgaben zu Abschnitt 18

18.1 Man verallgemeinere den Satz 18.1 auf den Fall von drei Ereignissen.

18.2 Jemand zieht aus zwei Skatspielen je eine Karte. Wie groß ist die Wahrscheinlichkeit, dabei wenigstens ein Herz-As zu erhalten?

18.3 Man beantworte die Frage in Beispiel 18.1 mittels (16.1).

19 Bedingte Wahrscheinlichkeit. Multiplikationssatz

Wir betrachten ein beliebiges Experiment und nehmen an, daß wir bei diesem Experiment insgesamt n gleichmögliche Fälle unterscheiden können. Dabei treffe

in k Fällen ein gewisses Ereignis A ein,
in l Fällen ein gewisses Ereignis B ein und
in m Fällen das Ereignis AB ein.

Die klassische Wahrscheinlichkeitsdefinition läßt sich anwenden und ergibt

$$P(A) = \frac{k}{n}, \qquad P(B) = \frac{l}{n}, \qquad P(AB) = \frac{m}{n}.$$

Wir wollen nun die Wahrscheinlichkeit von B *unter der zusätzlichen Bedingung* ermitteln, daß insgesamt nur noch die Fälle betrachtet werden, in denen A eintrifft. Diese Wahrscheinlichkeit bezeichnen wir mit dem Symbol $P(B \mid A)$. Unter der genannten Bedingung bleiben insgesamt nur noch k Fälle übrig, nämlich alle die Fälle, in denen A eintrifft. Hierbei sei $k \neq 0$. Unter diesen sind m Fälle, in denen außer A obendrein noch B eintrifft. So ergibt sich sofort

$$P(B \mid A) = \frac{m}{k}.$$

Aus Axiom 3 folgt durch Induktion der

Satz 16.1 (Additionssatz der Wahrscheinlichkeit). *Sind $A_1, \ldots, A_m$ Ereignisse, die sich bei einem Zufallsexperiment gegenseitig ausschließen, so gilt bei diesem Experiment*

$$\boxed{P(A_1 + A_2 + \cdots + A_m) = P(A_1) + P(A_2) + \cdots + P(A_m).} \tag{16.5}$$

Bei Experimenten, bei denen man zwischen abzählbar unendlich vielen Ereignissen $A_1, A_2, \ldots, A_m, \ldots$ unterscheiden kann, bezeichnet

$$A_1 + A_2 + \cdots + A_m + \cdots \quad \text{oder kürzer} \quad A_1 + A_2 + \cdots$$

das Ereignis, das genau dann eintrifft, wenn mindestens eines der genannten unendlich vielen Ereignisse eintrifft.

Bei einem solchen Experiment ist Axiom 3 zu ersetzen durch

Axiom 3*. *Sind $A_1, A_2, \ldots$ abzählbar unendlich viele Ereignisse, die sich bei einem Experiment gegenseitig ausschließen, so gilt bei diesem Experiment*

$$P(A_1 + A_2 + \cdots) = P(A_1) + P(A_2) + \cdots. \tag{16.6}$$

Ein Ereignis, das bei einem Experiment niemals eintreffen kann, heißt ein **unmögliches Ereignis** bei diesem Experiment. Ist A ein solches Ereignis, so gilt

$$P(A) = 0. \tag{16.7}$$

Den Beweis überlassen wir dem Leser (s. Aufgabe 16.1).

Aus (16.7) *folgt aber nicht, daß A notwendigerweise ein unmögliches Ereignis ist.* Das werden wir uns in Abschn. 26 auf ganz einfache Weise klarmachen. Einstweilen bemerken wir dazu: Die Häufigkeitsinterpretation der Wahrscheinlichkeit besagt nur, daß die relative Häufigkeit $h(A)$ bei oftmaliger Ausführung des betreffenden Experimentes ungefähr gleich null ist. A kann also auf lange Dauer höchstens in einem winzigen Bruchteil aller Ausführungen auftreten. Führen wir das Experiment nur einmal aus, so können wir es als *praktisch* sicher ansehen, daß A nicht eintrifft.

Entsprechend folgt aus $P(A) = 1$ nicht notwendigerweise, daß A ein sicheres Ereignis ist. Wir können dann nur schließen, daß A auf lange Dauer höchstens in einem winzigen Bruchteil aller Ausführungen des betreffenden Experimentes nicht eintrifft. Führen wir das Experiment nur einmal aus, so können wir also *praktisch* sicher damit rechnen, daß A eintrifft.

Aufgaben zu Abschnitt 16

16.1 Man beweise: Ist A unmöglich, so ist $P(A) = 0$.

16.2 Man zeige: Hat das Eintreffen von A stets das Eintreffen von B zur Folge, so gilt $P(A) \leq P(B)$.

16.3 Man beweise den Satz 16.1.

16.4 Wie groß ist die Wahrscheinlichkeit, beim gleichzeitigen Wurf dreier regelmäßiger Würfel lauter verschiedene Zahlen zu erhalten? Man bearbeite die Aufgabe mit und ohne (16.1).

17 Zur praktischen Bestimmung von Wahrscheinlichkeiten

Wir erörtern nun die Frage, wie man die Wahrscheinlichkeiten von Ereignissen in einem konkreten Falle bestimmt.

Der klassische Wahrscheinlichkeitsbegriff ist ein Sonderfall des soeben definierten Wahrscheinlichkeitsbegriffs, läßt sich also auch weiterhin verwenden, wo dies bisher möglich war.

In der Tat erfüllt die klassische Definition der Wahrscheinlichkeit das Axiom 1. Weiterhin lassen sich die in Axiom 2 und 3 ausgedrückten Eigenschaften bei dieser Definition beweisen, brauchen also nicht postuliert zu werden.

Liegt ein Zufallsexperiment vor, bei dem sich endlich viele gleichmögliche Fälle unterscheiden lassen, so können wir also die Wahrscheinlichkeit jedes Ereignisses unter Benutzung der klassischen Definition sofort angeben, ohne uns auf beobachtete relative Häufigkeiten zu stützen.

Bei anderen Zufallsexperimenten müssen wir die unbekannten Werte der Wahrscheinlichkeiten unter Benutzung beobachteter relativer Häufigkeiten bei langen Versuchsreihen festlegen, und zwar derart, daß die Axiome nicht verletzt werden.

Daß man auf diese Weise immer nur Näherungswerte angeben kann, stört so wenig, wie es im klassischen Fall von Belang ist, daß man z. B. dem Ereignis „Kopf" beim Münzenwurf die Wahrscheinlichkeit 0,5 zuordnet, obwohl vielleicht 0,502 oder 0,499 genauere Werte wären, weil ja eine Münze stets eine gewisse winzige Unsymmetrie besitzt.

Die Lage ist ähnlich wie etwa in der klassischen Mechanik: Man postuliert, jeder Körper habe eine ganz bestimmte Masse, und es stört dann beim Aufbau der Theorie nicht im geringsten, daß man die Masse eines Körpers praktisch immer nur näherungsweise ermitteln kann.

Nun ist auf Grund der obigen Wahrscheinlichkeiten

$$\frac{m}{k} = \frac{m/n}{k/n} = \frac{P(A\,B)}{P(A)}.$$

Demnach erhalten wir das Ergebnis

$$\text{(19.1)} \qquad \boxed{P(B \mid A) = \frac{P(A\,B)}{P(A)}} \qquad [P(A) \neq 0].$$

Unter Benutzung dieser Formel führen wir nun den folgenden wichtigen Begriff ein.

Definition. *Sind A und B irgendwelche Ereignisse und ist bei einem Experiment $P(A) \neq 0$, so heißt die durch* (19.1) *gegebene Größe $P(B \mid A)$ die Wahrscheinlichkeit des Ereignisses B bei dem genannten Experiment unter der Hypothese, daß das Ereignis A eingetroffen ist, oder kürzer die* **bedingte Wahrscheinlichkeit** *des Ereignisses B unter der Hypothese A.*

Die inhaltliche Bedeutung dieses Begriffs bei Experimenten, bei denen sich die klassische Wahrscheinlichkeitsdefinition anwenden läßt, haben wir oben angegeben, und bei anderen Experimenten ist sie ganz entsprechend.

Aus der Definition folgt, daß die bedingte Wahrscheinlichkeit eines Ereignisses A unter der Hypothese B durch

$$\text{(19.2)} \qquad P(A \mid B) = \frac{P(A\,B)}{P(B)} \qquad [P(B) \neq 0]$$

gegeben ist.

Schließen sich A und B gegenseitig aus, so ist $P(A \mid B) = 0$ und $P(B \mid A) = 0$, wie man sieht.

Aus (19.1) und (19.2) ergibt sich unmittelbar der grundlegende

Satz 19.1 (Multiplikationssatz). *Haben zwei Ereignisse A und B bei einem Experiment die Wahrscheinlichkeit $P(A)$ bzw. $P(B)$, so beträgt die Wahrscheinlichkeit des gleichzeitigen Eintreffens von A und B bei diesem Experiment*

$$\text{(19.3)} \qquad \boxed{P(A\,B) = P(A)\,P(B \mid A) = P(B)\,P(A \mid B).}$$

Diese Beziehung entspricht der Formel (13.5) für die relativen Häufigkeiten. Man wendet sie an, wenn die Berechnung der bedingten Wahrscheinlichkeit einfacher ist als die unmittelbare Berechnung von $P(A\,B)$. Ein Beispiel möge dies erläutern.

Zum Verständnis dieses und ähnlicher Beispiele bemerken wir zuvor, daß es zwei Arten des Ziehens von Dingen (etwa Karten aus einem Spiel, Kugeln aus einer Urne usw.) gibt:

1. Beim **Ziehen mit Zurücklegen** wird das jeweils gezogene Ding wieder zu den übrigen Dingen zurückgelegt (und es wird erneut gut durchgemischt), ehe man den nächsten Zug tut. Man stellt also jeweils den ursprünglichen Zustand vollkommen wieder her. Das Ergebnis eines Zuges hängt demgemäß nicht von den Ergebnissen der vorhergehenden Züge ab.

2. Beim **Ziehen ohne Zurücklegen** wird das jeweils gezogene Ding beiseite gelegt, also nicht wieder zurück zu den übrigen Dingen.

Beispiel 19.1. In einer Schachtel liegen 10 Verschlußkappen, darunter 3 defekte. Wie groß ist die Wahrscheinlichkeit, daß man lauter brauchbare Kappen erhält, wenn man 2 Kappen nacheinander ganz zufällig und ohne Zurücklegen herausnimmt?

Das Ereignis

$$A = \textit{Brauchbare Kappe beim 1. Zug}$$

hat die Wahrscheinlichkeit $P(A) = 7/10$. Ist dieses Ereignis eingetroffen, so sind noch 9 Kappen, davon 6 brauchbare, in der Schachtel. Also hat das Ereignis

$$B = \textit{Brauchbare Kappe beim 2. Zug}$$

unter der Bedingung, daß A eingetroffen ist, die Wahrscheinlichkeit $P(B \mid A) = 6/9 = 2/3$. Gemäß Satz 19.1 erhalten wir damit die Antwort

$$P(AB) = \frac{7}{10} \cdot \frac{2}{3} \approx 0{,}47.$$

Die Multiplikationsformel (19.3) läßt sich auf mehr als zwei Ereignisse ausdehnen (s. Aufgabe 19.2). Für m Ereignisse $A_1, \ldots, A_m$ erhält man

$$\begin{aligned} &P(A_1 A_2 \cdots A_m) \\ &= P(A_1)\, P(A_2 \mid A_1)\, P(A_3 \mid A_1 A_2) \cdots P(A_m \mid A_1 A_2 \cdots A_{m-1}). \end{aligned} \tag{19.4}$$

Aufgaben zu Abschnitt 19

19.1 Wie groß ist in Beispiel 19.1 die Wahrscheinlichkeit, bei dreimaligem Ziehen ohne Zurücklegen lauter brauchbare Verschlußkappen zu erhalten?

19.2 Man beweise (19.4).

19.3 Wie groß ist in Beispiel 19.1 die Wahrscheinlichkeit, bei viermaligem Ziehen ohne Zurücklegen wenigstens 3 brauchbare Verschlußkappen zu erhalten?

19.4 Man zeige, daß $P(A \mid B)$ die Axiome der Wahrscheinlichkeit erfüllt.

19.5 Wir nehmen an, daß eine Firma bei einer Abnahmekontrolle folgendermaßen verfährt: Aus einer Packung von 50 Artikeln (z. B. Sicherungen) werden 5 ganz zufällig und ohne Zurücklegen herausgegriffen. Sind alle 5 einwandfrei, wird die Packung angenommen, andernfalls nicht. Wie groß ist die Wahrscheinlichkeit, daß eine Packung angenommen wird, obwohl sie 20 % Ausschuß (= unbrauchbare Artikel) enthält?

19.6 Wie groß ist die Wahrscheinlichkeit in Aufgabe 19.5, wenn die Packung aus N Artikeln, darunter M defekten, besteht und n Artikel zufällig und ohne Zurücklegen herausgegriffen werden?

20 Unabhängige Ereignisse

Gilt für zwei Ereignisse A und B bei einem Experiment die Beziehung

$$\boxed{P(AB) = P(A)\,P(B)\,,} \tag{20.1}$$

so gewinnt (19.3) die Form

$$P(A)\,P(B) = P(A)\,P(B \mid A) = P(B)\,P(A \mid B).$$

Ist $P(A) \neq 0$ und $P(B) \neq 0$, so folgt hieraus sofort

$$P(A \mid B) = P(A), \qquad P(B \mid A) = P(B).$$

Diese Formeln besagen offenbar, daß die Wahrscheinlichkeit von A gar nicht davon abhängt, ob B eingetroffen ist oder nicht, und umgekehrt.

Ereignisse A und B, für die (20.1) gilt, werden deshalb als *stochastisch unabhängige Ereignisse* oder kurz als **unabhängige Ereignisse** bei dem betreffenden Experiment bezeichnet. Dies ist ein sehr wichtiger Begriff.

„**Stochastisch**" bedeutet ganz allgemein „mit Zufallsexperimenten und Wahrscheinlichkeiten zusammenhängend".

Beispiel 20.1. Wie groß ist die Wahrscheinlichkeit, bei zweimaligem Ziehen einer Karte aus einem Skatspiel mit Zurücklegen zwei Asse zu erhalten?

Da 4 der 32 Karten Asse sind, beträgt die Wahrscheinlichkeit, beim ersten Zug ein As zu erhalten, $4/32 = 1/8$. Da die gezogene Karte wieder ins Spiel gesteckt und dieses gemischt wird, ist die Ausgangssituation beim zweiten Zug dieselbe wie beim ersten. Das Ergebnis des ersten Zuges hat also keinen Einfluß auf das Ergebnis des zweiten. Demnach handelt es sich um unabhängige Ereignisse. Die Wahrscheinlichkeit, beim zweiten Zug ein As zu erhalten, beträgt ebenfalls $1/8$, und die Antwort lautet

$$P = \frac{1}{8} \cdot \frac{1}{8} = \frac{1}{64}.$$

Bei Glücksspielen wird es im allgemeinen leicht möglich sein zu entscheiden, ob zwei Ereignisse unabhängig sind oder nicht. Bei anderen Experimenten ist Vorsicht geboten: Man sollte Ereignisse dann und nur dann als unabhängig ansehen, wenn man bei hinreichender

Kenntnis der Natur des Experimentes zu dem Schluß kommt, daß zwischen diesen Ereignissen kein Kausalzusammenhang bestehen kann.

Der Begriff der Unabhängigkeit läßt sich auch auf mehr als zwei Ereignisse ausdehnen:

m Ereignisse $A_1, \ldots, A_m$ heißen **unabhängig** bei einem Experiment, wenn bei diesem Experiment

$$(20.2) \qquad P(A_{j_1} A_{j_2} \cdots A_{j_k}) = P(A_{j_1}) P(A_{j_2}) \cdots P(A_{j_k})$$

für alle $1 \leqq j_1 < j_2 < \cdots < j_k \leqq m$ und $k = 2, 3, \ldots, m$ gilt.

Ähnlich wie oben bedeutet dies, daß die Wahrscheinlichkeit jedes dieser Ereignisse nicht davon abhängt, welche und wie viele der übrigen Ereignisse eingetroffen sind.

Aufgaben zu Abschnitt 20

20.1 Wie groß ist die Wahrscheinlichkeit, bei viermaligem Werfen einer Münze zuerst zweimal „Kopf" und dann zweimal „Wappen" zu erhalten?

20.2 Wie groß ist die Wahrscheinlichkeit, bei viermaligem Werfen einer Münze zweimal „Kopf" und zweimal „Wappen" zu erhalten?

20.3 Sind mehrere Ereignisse unabhängig, so sind auch je zwei dieser Ereignisse unabhängig. Sind unter mehreren Ereignissen jeweils alle Paare unabhängig, so folgt hieraus *nicht* allgemein, daß diese Ereignisse insgesamt unabhängig sind. Um dies einzusehen, stelle man sich vor, eine Urne enthalte 4 Lose mit den Nummern 112, 121, 211, 222. Ein Los wird gezogen. Man zeige, daß die Ereignisse

A: *Die 1. Ziffer auf dem gezogenen Los ist eine Eins*
B: *Die 2. Ziffer auf dem gezogenen Los ist eine Eins*
C: *Die 3. Ziffer auf dem gezogenen Los ist eine Eins*

zwar paarweise aber nicht insgesamt unabhängig sind.

20.4 Ein Gehäuse bestehe aus einem Oberteil, einem Unterteil und einer Dichtung. Die Wahrscheinlichkeit, daß Ober- und Unterteil Ausschuß sind, betrage je 5%, die Wahrscheinlichkeit einer fehlerhaften Dichtung sei 10%, und es bestehe Unabhängigkeit. Wie groß ist die Wahrscheinlichkeit, daß ein Gehäuse völlig einwandfrei ist?

Kapitel 5

Wahrscheinlichkeitsverteilungen

Der Begriff der Wahrscheinlichkeitsverteilung erwächst aus der Frage, wie sich bei einem Zufallsexperiment die Wahrscheinlichkeiten auf die verschiedenen Ereignisse verteilen, d. h. welche Wahrscheinlichkeit diese Ereignisse bei dem betreffenden Experiment jeweils besitzen.

Wir unterscheiden zwei Klassen von Verteilungen, die *diskreten* (Abschn. 22) und die *stetigen* (Abschn. 26). Bei Zufallsexperimenten, bei denen man *zählt* (Ausschußstücke, Fahrzeuge, Grippefälle usw.) treten diskrete Verteilungen auf. Bei Experimenten, bei denen man *mißt*, d. h. eine kontinuierlich veränderliche Größe beobachtet (Länge, Temperatur, Ernteertrag usw.), treten stetige Verteilungen auf.

Jede diskrete bzw. stetige Verteilung wird mathematisch durch ihre sogenannte Wahrscheinlichkeitsfunktion (Abschn. 22) bzw. Dichte (Abschn. 26) oder ebensogut durch ihre Verteilungsfunktion (Abschn. 24) beschrieben und bestimmt.

(Im vorliegenden Kapitel betrachten wir ausschließlich Verteilungen einer einzelnen Zufallsvariablen. Verteilungen mehrerer Variablen werden in Kap. 10 behandelt.)

Dem Leser wird es auffallen, daß die Begriffe im vorliegenden Kapitel so ähnliche Namen haben wie die in Teil I des Buches. Hierzu ist folgendes zu sagen:

In Teil I ging es uns darum, die empirisch beobachtbare Wirklichkeit zu beschreiben, und alle Begriffe bezogen sich auf Stichproben. Gegenwärtig sind wir dabei, theoretische Modelle von Grundgesamtheiten zu schaffen, und alle Begriffe in Teil II sind demzufolge theoretischer Natur. In Teil III werden Beziehungen zwischen Theorie und Wirklichkeit hergestellt. Dabei wird sich herausstellen, daß Begriffe in Teil I und II mit ähnlichen Namen einander jeweils entsprechen. Zum Beispiel wird sich die Wahrscheinlichkeitsverteilung als theoretisches Gegenstück der Häufigkeitsverteilung erweisen usw.

Beide Begriffsarten sind natürlich prinzipiell verschieden. Um es an unserem Beispiel zu erläutern: Eine Grundgesamtheit besitzt eine einzige wohlbestimmte Wahrscheinlichkeitsverteilung. Die Stichproben, die man aus dieser Grundgesamtheit entnehmen kann, unterscheiden sich im allgemeinen voneinander und besitzen demzufolge im allgemeinen auch verschiedene Häufigkeitsverteilungen.

21 Zufallsvariable

In den meisten Fällen läßt sich das Ergebnis einer einzelnen Ausführung eines Zufallsexperimentes durch eine oder mehrere Zahlen kennzeichnen. Im vorliegenden Kapitel befassen wir uns ausschließlich mit *Zufallsexperimenten, bei denen das Ergebnis einer einzelnen Ausführung jeweils durch eine einzelne Zahl ausgedrückt werden kann.*

Dies gilt z. B. für das Würfeln mit einem einzelnen Würfel. Dabei erhalten wir als Ergebnis eines Wurfs eine der Zahlen $1, 2, \ldots, 6$. Dieses Ergebnis wird also durch eine variable Größe, nämlich die Augenzahl, die wir mit X bezeichnen wollen, gekennzeichnet. X ist demnach eine Funktion, die bei jedem Wurf einen der Werte $1, 2, \ldots$ oder 6 annimmt. Welchen dieser Werte sie bei einem bestimmten Wurf annimmt, das hängt „vom Zufall" ab. Eine solche Funktion, die das Ergebnis eines Zufallsexperimentes ausdrückt, wird als *Zufallsvariable* oder *stochastische Variable* bezeichnet.

Ganz entsprechend ist die Lage bei anderen Experimenten, bei denen man eine Messung oder Zählung durchführt. Man denke z. B. an das Messen der Größe der Personen einer Bevölkerungsgruppe oder der Kapazität seriengefertigter Kondensatoren, an das Zählen der monatlichen Verkehrsunfälle in einer Stadt oder der täglichen Ferngespräche in einer Zentrale usw. Dabei ist die genannte Länge, Kapazität bzw. monatliche Unfallszahl jeweils eine Zufallsvariable.

Aber auch, wenn die bei einem Experiment denkbaren Ereignisse zunächst nicht durch Zahlenwerte gekennzeichnet sind, kann man jedes mögliche Ereignis durch eine Zahl bezeichnen und hat dann dieselbe Situation wie bei den vorstehenden Beispielen. Beim Werfen einer Münze können wir etwa die beiden Ereignisse „Kopf" und „Wappen" mit 0 bzw. 1 bezeichnen. Bei der Feststellung der Augenfarbe können wir die Ereignisse „blaue Augen", „braune Augen" und „andersfarbige Augen" mit 1, 2, bzw. 3 bezeichnen. Welche Zahlen wir wählen, ist gleichgültig. Wir müssen nur im Sinn behalten, welches Ereignis die jeweils gewählte Zahl symbolisiert.

Die Beispiele zeigen, daß Zufallsvariable in den verschiedensten Gebieten auftreten. Obwohl diese Beispiele (und andere, die sich der Leser leicht selbst bilden kann) inhaltlich ganz verschieden sind, haben sie, mathematisch gesehen, die folgende Gemeinsamkeit: In jedem Beispiel kommt eine Variable vor, die verschiedener Werte fähig ist, und man kann nie mit Sicherheit vorhersagen, welchen dieser Werte die betreffende Variable annehmen wird, weil dies vom Einfluß unkontrollierbarer zufälliger Umstände abhängt.

Grob gesprochen ist also eine Zufallsvariable X eine Funktion, deren Werte reelle Zahlen sind und „vom Zufall" abhängen.

Die Definition der Zufallsvariablen X fassen wir nun so, daß sich stets eine Wahrscheinlichkeit erklären läßt, mit der X einen gewissen Zahlenwert oder allgemeiner einen beliebigen Wert aus einem gewissen Intervall der Zahlengeraden annimmt. (Definition des Intervallbegriffs siehe in Anhang 1.)

Definition. *Eine Funktion X heißt eine* **Zufallsvariable** *oder* **stochastische Variable,** *wenn sie einem Zufallsexperiment zugeordnet ist und folgende Eigenschaften hat:*

1. *Die Werte von X sind reelle Zahlen.*
2. *Für jede Zahl a und für jedes Intervall I auf der Zahlengeraden ist die Wahrscheinlichkeit des Ereignisses „X hat den Wert a" bzw. „X liegt in dem Intervall I" im Einklang mit den Axiomen der Wahrscheinlichkeit* (Abschn. 15) *erklärt.*

Den Leser mag es zunächst bekümmern, daß diese Definition offenbar eine Fülle verschiedenartigster Funktionen einschließt. Zum Trost sei aber gesagt, daß die Anzahl der praktisch bedeutsamen Arten von Zufallsvariablen (und ihrer „Verteilungen") relativ klein und gut überschaubar ist. Hiervon werden wir in Kap. 8–11 eine Vorstellung erhalten.

Trifft bei der Ausführung eines bestimmten Experimentes das Ereignis, das einem Zahlenwert a entspricht, ein, so sagen wir, daß die zu diesem Experiment gehörige Zufallsvariable X den Wert a **angenommen** habe oder daß der Wert $X = a$ **beobachtet** worden sei. Wir sprechen auch kurz von *dem Ereignis* $X = a$ bei dem betreffenden Experimente.

Die Wahrscheinlichkeit, daß X den Wert a annimmt, bezeichnen wir mit

$$P(X = a).$$

Die Wahrscheinlichkeit des Ereignisses

„X nimmt irgendeinen Wert in dem Intervall $a < X < b$ an"

bezeichnen wir mit

$$P(a < X < b),$$

usw. Entsprechend bezeichnet

$$P(X \leqq c)$$

die Wahrscheinlichkeit des Ereignisses

$X \leqq c$ („*X nimmt irgendeinen Wert an, der höchstens gleich c ist*"), und

$$P(X > c)$$

bezeichnet die Wahrscheinlichkeit des Ereignisses

$X > c$ („*X nimmt irgendeinen Wert an, der größer als c ist*").

Die beiden zuletzt genannten Ereignisse schließen sich für jeden beliebigen Wert von c aus. Gemäß Axiom 3 in Abschn. 15 gilt also

$$P(X \leqq c) + P(X > c) = P(-\infty < X < \infty).$$

Auf Grund von Axiom 2 ist die Wahrscheinlichkeit auf der rechten Seite gleich 1, denn $-\infty < X < \infty$ ist ein sicheres Ereignis, da X ja stets irgendeinen Wert auf der Zahlengeraden annehmen muß. So erhalten wir die wichtige Beziehung

$$\text{(21.1)} \qquad \boxed{P(X > c) = 1 - P(X \leqq c)} \qquad (c \text{ beliebig, reell}),$$

die wir später noch oft anwenden werden.

Zum Schluß noch ein einfaches Beispiel zur Erläuterung unserer Bezeichnungen:

Beispiel 21.1. Es sei X die beim Wurf eines regelmäßigen Würfels erzielte Zahl. Dann ist

$$P(X = 1) = \frac{1}{6}, \qquad P(X = 2) = \frac{1}{6} \qquad \text{usw.}$$

$$P(1 < X < 2) = 0, \quad P(1 \leqq X < 2) = \frac{1}{6}, \quad P(1 \leqq X \leqq 2) = \frac{1}{3}$$

$$P(1 \leqq X < 6) = \frac{5}{6}, \quad P(1 \leqq X \leqq 6) = 1, \quad P(-\infty < X < \infty) = 1$$

$$P\left(\frac{3}{2} < X < \frac{5}{2}\right) = \frac{1}{6}, \quad P(0{,}3 \leqq X \leqq 3{,}2) = \frac{1}{2} \qquad \text{usw.}$$

Aufgaben zu Abschnitt 21

21.1 Es sei X die Anzahl der „Köpfe" beim zweimaligen Wurf einer Münze. Man bestimme die Wahrscheinlichkeiten $P(X = 0)$, $P(X = 1)$, $P(X = 2)$, $P(1 < X < 2)$, $P(X \leq 1)$, $P(X \geq 1)$, $P(X > 1)$, $P(0{,}5 < X < 10)$.

21.2 In einer Schachtel liegen 4 rechtsgängige und 6 linksgängige Muttern. 2 davon werden zufällig und ohne Zurücklegen gezogen. X sei die Anzahl der linksgängigen unter den gezogenen. Man bestimme die Wahrscheinlichkeiten $P(X = 0)$, $P(X = 1)$ usw. wie in Aufgabe 21.1.

21.3 Man zeige: Ist $b < c$, so ist $P(X \leq b) \leq P(X \leq c)$.

22 Diskrete Verteilung. Wahrscheinlichkeitsfunktion

Die meisten bei praktischen Problemen vorkommenden Zufallsvariablen lassen sich in zwei Klassen einteilen, nämlich in die sogenannten diskreten und die sogenannten stetigen Variablen. Wir betrachten zuerst die diskreten und dann später die stetigen.

Eine Zufallsvariable X und ihre Verteilung heißen **diskret**, wenn folgendes gilt:

1. Die Variable X kann nur endlich viele oder abzählbar unendlich viele (reelle) Werte mit positiver Wahrscheinlichkeit annehmen.

2. In jedem endlichen Intervall der reellen Zahlengeraden liegen nur endlich viele der genannten Werte. Für jedes Intervall $a < X \leqq b$, das keinen solchen Wert enthält, ist die zugehörige Wahrscheinlichkeit $P(a < X \leqq b)$ gleich null.

Die Werte, für die X eine positive Wahrscheinlichkeit besitzt, bezeichnen wir mit

$$x_1, \quad x_2, \quad x_3, \ldots$$

und die zugehörigen Wahrscheinlichkeiten mit

$$p_1, \quad p_2, \quad p_3, \ldots$$

Es ist also $P(X = x_1) = p_1$ usw. Wir führen nun die Funktion

$$f(x) = \begin{cases} p_j & \text{für } x = x_j, \\ 0 & \text{für alle übrigen } x \end{cases} \qquad (j = 1, 2, \ldots) \tag{22.1}$$

ein. Diese heißt die **Wahrscheinlichkeitsfunktion** der betreffenden Zufallsvariablen X. Sie läßt sich durch ein Stabdiagramm graphisch darstellen, wie die Beispiele im nächsten Abschnitt erläutern.

Da X stets irgendeinen Wert annimmt, muß die Summe aller $f(x_j)$ gleich 1 sein,

$$\Sigma f(x_j) = 1. \tag{22.2}$$

Kennen wir die Wahrscheinlichkeitsfunktion einer diskreten Zufallsvariablen X, so können wir die Wahrscheinlichkeit $P(a < X \leqq b)$

für jedes beliebige Intervall $a < X \leq b$ in einfacher Weise berechnen. Wir brauchen lediglich die Summe aller Wahrscheinlichkeiten p_j zu bilden, deren zugehörige x_j in dem Intervall liegen:

$$P(a < X \leq b) = \sum_{a < x_j \leq b} f(x_j) = \sum_{a < x_j \leq b} p_j. \tag{22.3}$$

Ganz entsprechend erhält man die Wahrscheinlichkeit im Falle eines offenen, abgeschlossenen oder unendlichen Intervalls durch Summation. Wir sagen hierfür kurz:

Durch die Wahrscheinlichkeitsfunktion $f(x)$ ist die „**Wahrscheinlichkeitsverteilung**" oder die „**Verteilung**" der betreffenden Zufallsvariablen X vollständig bestimmt.

23 Einige einfache Beispiele

Beispiel 23.1. Die Zufallsvariable

X = *Augenzahl beim Wurf eines Würfels*

hat, wenn der Würfel regelmäßig ist, die Wahrscheinlichkeitsfunktion (siehe Abb. 23.1)

$$f(x) = 1/6 \qquad \text{für } x = 1, 2, \ldots, 6$$

und $f(x) = 0$ für alle übrigen x.

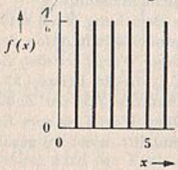


Abbildung 23.1. Wahrscheinlichkeitsfunktion in Beispiel 23.1

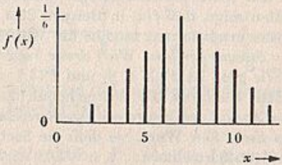


Abbildung 23.2. Wahrscheinlichkeitsfunktion in Beispiel 23.2

Beispiel 23.2. Die Wahrscheinlichkeitsfunktion $f(x)$ der Zufallsvariablen

X = *Augensumme beim Wurf zweier regelmäßiger Würfel*

hat die folgenden Werte (vgl. Abb. 23.2):

x	2	3	4	5	6	7	8	9	10	11	12
$f(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

In der Tat hat jeder der $6 \cdot 6 = 36$ gleichmöglichen Fälle $(1, 1), (1, 2), \ldots, (6, 6)$ die Wahrscheinlichkeit $1/36$. Hierbei bezeichnet die erste Zahl jeweils das Ergebnis auf dem ersten Würfel und die zweite Zahl das Ergebnis auf dem anderen. $(1, 1)$ ist der einzige Fall, in dem $X = 2$ eintrifft. Also ist $P(X = 2) = 1/36$. Der Wert $X = 3$ trifft in den Fällen $(1, 2)$ und $(2, 1)$ ein. Demnach ist $P(X = 3) = 2/36$, usw.

Beispiel 23.3. Wir werfen eine Münze und betrachten die Zufallsvariable

X = *Anzahl der Würfe bis zum Eintreffen des „ersten Kopfes"*

(der Wurf, bei dem „Kopf" eintrifft, mitgezählt). Dann bedeutet also

$X = 1$ „Kopf" beim 1. Wurf

$X = 2$ „Wappen" beim 1. Wurf, „Kopf" beim 2. Wurf

und so weiter. Da „Kopf" und „Wappen" gleichwahrscheinlich und die Einzelereignisse unabhängig sind, ist offenbar

$$P(X=1)=\frac{1}{2}, \qquad P(X=2)=\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{4}, \qquad P(X=3)=\frac{1}{8}$$

usw. So erhalten wir die Wahrscheinlichkeitsfunktion

$$f(x)=\frac{1}{2^x} \qquad \text{für } x=1,2,3,\ldots$$

und $f(x) = 0$ für alle übrigen x. Das ist also eine diskrete Verteilung, bei der die Zufallsvariable für abzählbar unendlich viele Werte eine positive Wahrscheinlichkeit hat.

Aufgaben zu Abschnitt 23

23.1 Man zeichne das Stabdiagramm der Wahrscheinlichkeitsfunktion in Beispiel 23.3.

23.2 Wie ändert sich $f(x)$ in Beispiel 23.3, wenn man X dadurch ändert, daß man den Wurf, bei dem „Kopf" eintrifft, nicht mitzählt?

23.3 Man zeige, daß $f(x)$ in Beispiel 23.3 die Beziehung (22.2) erfüllt.

23.4 Man ermittle und zeichne die Wahrscheinlichkeitsfunktion der Variablen X = *Summe der beim Wurf dreier regelmäßiger Würfel erzielten Zahlen.*

23.5 Wie groß ist $P(X \leq 8)$ und $P(3{,}1 < X \leq 6{,}4)$ in Beispiel 23.2?

23.6 Man bestimme die Wahrscheinlichkeitsfunktion $f(x)$ der Zufallsvariablen X = *Anzahl der Würfe mit einem regelmäßiger Würfel bis zum Eintreffen der ersten Sechs* (der Wurf, bei dem die Sechs eintrifft, nicht mitgezählt) und zeige, daß (22.2) erfüllt ist.

23.7 Man ermittle und zeichne die Wahrscheinlichkeitsfunktion der Variablen X = *Anzahl der „Köpfe" beim gleichzeitigen Wurf von 4 Münzen.*

24 Verteilungsfunktion einer Zufallsvariablen

Ist X eine Zufallsvariable, die zu irgendeinem bestimmten Zufallsexperiment gehört, so existiert zu jeder gegebenen reellen Zahl x definitionsgemäß die Wahrscheinlichkeit

$$P(X \leqq x),$$

mit der X irgendeinen Wert in dem unendlichen Intervall $X \leqq x$ annimmt. Diese Wahrscheinlichkeit hängt natürlich von x ab. Sie ist eine Funktion von x, die wir mit $F(x)$ bezeichnen. Dann ist also

$$\boxed{F(x) = P(X \leqq x).} \tag{24.1}$$

Diese Funktion $F(x)$, die für alle reellen x erklärt ist, heißt die **Verteilungsfunktion** der betreffenden Zufallsvariablen X.

Beispiel 24.1. Man bestimme die Verteilungsfunktion $F(x)$ der Zufallsvariablen

X = *Anzahl der „Köpfe" beim einmaligen Wurf einer Münze.*

Da X keine negativen Werte annehmen kann, ist $P(X < 0) = 0$, also auch $F(x) = 0$ für $x < 0$. Aus Symmetriegründen sind „Kopf" und „Wappen" gleichwahrscheinlich. Demnach ist $P(X = 0) = 0{,}5$, also

$$F(0) = P(X \leqq 0) = 0{,}5.$$

Diesen Wert behält $F(x)$ bei, bis der nächste Wert erreicht wird, den X annehmen kann. Dies ist der Wert $X = 1$, und es gilt $P(X = 1) = 0{,}5$. Also wird

$$F(1) = P(X \leqq 1) = P(X = 0) + P(X = 1) = 0{,}5 + 0{,}5 = 1.$$

Diesen Wert behält $F(x)$ für alle $x \geqq 1$ bei. $F(x)$ hat also die folgenden Werte (s. Abb. 24.1):

x	$x < 0$	$0 \leqq x < 1$	$x \geqq 1$
$F(x)$	0	0,5	1

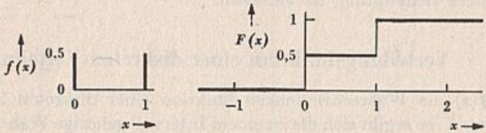


Abbildung 24.1. Wahrscheinlichkeitsfunktion $f(x)$ und Verteilungsfunktion $F(x)$ in Beispiel 24.1

Bisweilen wird die Verteilungsfunktion $F(x)$ in der Literatur durch $F(x) = P(X < x)$ definiert. Hierauf sollte man achten. Im Englischen heißt $F(x)$ *distribution function* bzw. *cumulative distribution function* je nachdem, ob $f(x)$ als *frequency function* oder als *distribution function* bezeichnet wird.

Wir wollen nun zeigen, daß sich aus der Verteilungsfunktion $F(x)$ die Wahrscheinlichkeit $P(a < X \leqq b)$ von

$a < X \leqq b$ [„*X nimmt irgendeinen Wert zwischen a* (ausschließlich) *und b* (einschließlich) *an.*"]

berechnen läßt. a und b sind hierbei irgendwelche reelle Zahlen, und es ist $b > a$. Wir sagen hierfür kurz: Durch $F(x)$ ist die **„Wahrscheinlichkeitsverteilung"** oder die **„Verteilung"** der betreffenden Zufallsvariablen eindeutig bestimmt.

Zum Beweis bemerken wir zuerst, daß

$$X \leqq a \quad \text{und} \quad a < X \leqq b$$

Ereignissen entsprechen, die sich gegenseitig ausschließen. Der Summe dieser Ereignisse entspricht

$$X \leqq b.$$

Gemäß Axiom 3 in Abschn. 15 erhalten wir demnach

$$P(X \leqq b) = P(X \leqq a) + P(a < X \leqq b)$$

oder

$$P(a < X \leqq b) = P(X \leqq b) - P(X \leqq a). \tag{24.2}$$

In dieser Formel ist

$$P(X \leqq b) = F(b) \quad \text{und} \quad P(X \leqq a) = F(a).$$

So ergibt sich die grundlegende Beziehung

$$\boxed{P(a < X \leqq b) = F(b) - F(a),} \tag{24.3}$$

und unsere Behauptung ist bewiesen.

25 Verteilungsfunktion einer diskreten Verteilung

Ist $f(x)$ die Wahrscheinlichkeitsfunktion einer diskreten Zufallsvariablen X, so ergibt sich die zu einem Intervall gehörige Wahrscheinlichkeit von X durch Summation. Dies haben wir uns in Abschn. 22 überlegt. Insbesondere beträgt die Wahrscheinlichkeit, daß X irgendeinen Wert annimmt, der nicht größer als eine gegebene Zahl x ist,

$$P(X \leqq x) = \sum_{x_j \leqq x} f(x_j) \qquad \text{[vgl. (22.1)].}$$

Hierbei wird also über alle x_j summiert, die höchstens gleich der gegebenen Zahl x sind. Nun ist aber definitionsgemäß $P(X \leqq x) = F(x)$. Dies bedeutet:

Zwischen der Verteilungsfunktion $F(x)$ und der Wahrscheinlichkeitsfunktion (22.1) *besteht der Zusammenhang*

$$F(x) = \sum_{x_j \leqq x} f(x_j). \tag{25.1}$$

Abb. 24.1 zeigt ein einfaches Beispiel zu (25.1). Ein weiteres ist das folgende

Beispiel 25.1. Abb. 25.1 zeigt die Wahrscheinlichkeitsfunktion $f(x)$ und die Verteilungsfunktion $F(x)$ der Zufallsvariablen

X = *Erzielte Zahl beim Wurf eines regelmäßigen Würfels.*

$F(x)$ hat die folgenden Werte:

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	1/3	1/2	2/3	5/6	1

$= \sum f(x$

In der Tat ist 1 der kleinste Wert, den X annehmen kann. Also wird $F(x) = 0$ für $x < 1$. Im Punkte $x = 1$ springt $F(x)$ auf den Wert 1/6 $[= P(X = 1)]$. Dann bleibt $F(x)$ konstant, bis der nächste Wert ($x = 2$) erreicht wird, zu dem eine positive Wahrscheinlichkeit gehört, usw.

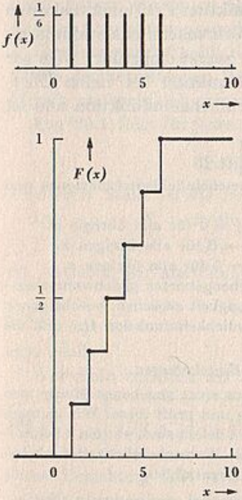


Abbildung 25.1. Wahrscheinlichkeitsfunktion $f(x)$ und Verteilungsfunktion $F(x)$ in Beispiel 25.1. (Die Punkte am linken Ende der Strecken kennzeichnen den Wert der Funktion $F(x)$ an den Sprungstellen.)

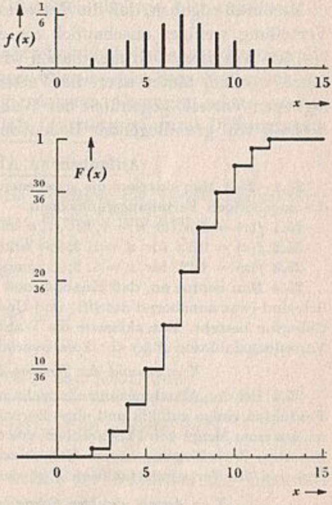


Abbildung 25.2. Wahrscheinlichkeitsfunktion $f(x)$ und Verteilungsfunktion $F(x)$ der Zufallsvariablen $X =$ *Summe der mit zwei regelmäßigen Würfeln erzielten Zahlen* (vgl. Beispiel 23.2)

Die Verteilungsfunktionen in Abb. 24.1, 25.1 und 25.2 besitzen Sprünge an den Stellen, die positiven Wahrscheinlichkeiten von X entsprechen. Dazwischen verlaufen sie konstant. Sie sind also sogenannte **Treppenfunktionen.** Dies ist typisch für diskrete Verteilungen. Genauer:

Die Werte $x_1, x_2, x_3, \ldots$, für die eine diskrete Zufallsvariable X eine positive Wahrscheinlichkeit besitzt, (und nur diese Werte) nennen wir die **möglichen Werte** von X. In jedem Intervall, in dem kein möglicher Wert liegt, verläuft $F(x)$ konstant. $F(x)$ ist also eine Treppenfunktion, die an der Stelle $x = x_j$ um den Betrag $p_j = P(X = x_j)$ in die Höhe springt und zwischen zwei benachbarten Sprungstellen konstant verläuft.

Man muß zugeben, daß die Verteilungsfunktion $F(x)$ einer diskreten Verteilung weniger anschaulich als die Wahrscheinlichkeitsfunktion $f(x)$ ist. Aus diesem Grunde haben wir $f(x)$ zuerst eingeführt. Wie wir sehen werden, bietet aber die Verteilungsfunktion bei vielen Überlegungen Vorteile gegenüber der Wahrscheinlichkeitsfunktion und ist deshalb von grundlegender Bedeutung.

Aufgaben zu Abschnitt 25

25.1—25.3 Man skizziere die gegebenen Wahrscheinlichkeitsfunktionen und die zugehörigen Verteilungsfunktionen.

25.1 $f(x) = 1/n$ für $x = 1, 2, \ldots, n$ und $f(x) = 0$ für alle übrigen x.

25.2 $f(x) = 0{,}1\,x$ für $x = 1, 2, 3, 4$ und $f(x) = 0$ für alle übrigen x.

25.3 $f(x) = 1/2^x$ für $x = 1, 2, \ldots$ und $f(x) = 0$ für alle übrigen x.

25.4 Man nehme an, daß Knaben- und Mädchengeburten gleichwahrscheinlich sind (was annähernd zutrifft) und Unabhängigkeit zwischen verschiedenen Geburten besteht. Man skizziere die Wahrscheinlichkeitsfunktion $f(x)$ und die Verteilungsfunktion $F(x)$ der Zufallsvariablen

X = *Anzahl der Knaben bei 3 Einzelgeburten*.

25.5 Bei der Abnahmekontrolle zieht man aus einer gegebenen Menge von Produkten einige zufällig und ohne Zurücklegen und prüft diese. Wir nehmen an, aus einer Menge von 10 Schrauben, von denen 4 defekt sind, werden 2 Schrauben ohne Zurücklegen gezogen. Man skizziere die Wahrscheinlichkeitsfunktion $f(x)$ und die Verteilungsfunktion $F(x)$ der Zufallsvariablen

X = *Anzahl defekter Schrauben unter den gezogenen*.

26 Stetige Verteilung

Eine Zufallsvariable X und deren Verteilung heißen **stetig***, wenn die zugehörige Verteilungsfunktion

$$F(x) = P(X \leq x)$$

in Integralform dargestellt werden kann,

$$\boxed{F(x) = \int_{-\infty}^{x} f(v)\,dv\,,} \tag{26.1}$$

* Auch *vom stetigen Typ*, um zu betonen, daß aus der Stetigkeit von $F(x)$ *nicht* die Existenz einer Darstellung (26.1) folgt.

wobei der Integrand eine nichtnegative und bis auf höchstens endlich viele Punkte stetige Funktion ist.

$F(x)$ ist natürlich überall stetig.

Wir verwenden ausschließlich den Riemannschen Integralbegriff der elementaren Integralrechnung. Die Integrationsvariable haben wir mit v bezeichnet, da x schon in der oberen Grenze vorkommt. Für den mit der Differential- und Integralrechnung wenig vertrauten Leser haben wir einige Grundtatsachen, die wir hier benutzen, in Anhang 1 zusammengestellt.

Der Integrand f heißt die **Wahrscheinlichkeitsdichte** oder kurz die **Dichte** der betreffenden Verteilung. Dies ist dem Sprachgebrauch der Physik nachgebildet, wie wir im nächsten Abschnitt sehen werden.

Aus (26.1) folgt für jedes x, in dem $f(x)$ stetig ist, durch Differentation

$$\boxed{F'(x) = f(x).}$$

In diesem Sinne ist die Dichte die Ableitung der Verteilungsfunktion.

Da $-\infty < X < \infty$ ein sicheres Ereignis und damit

$$P(-\infty < X < \infty) = 1$$

ist, so sehen wir aus (26.1), daß

$$\int_{-\infty}^{\infty} f(v)\, dv = 1 \tag{26.2}$$

sein muß.

Weiterhin erhalten wir aus (24.3) und (26.1) nun

$$\boxed{P(a < X \leqq b) = F(b) - F(a) = \int_a^b f(v)\, dv\,.} \tag{26.3}$$

Diese Beziehung läßt sich auf Grund der Definition des Integralbegriffs anschaulich fassen:

Die Wahrscheinlichkeit des Ereignisses $a < X \leqq b$ ist gleich der Fläche unter der Kurve der Wahrscheinlichkeitsdichte $f(x)$ zwischen $x = a$ und $x = b$. Vgl. Abb. 26.1.

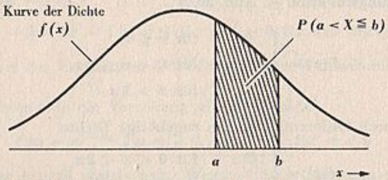


Abbildung 26.1. Beispiel zur Erläuterung der Formel (26.3)

Für sehr kurze Intervalle mit der Länge Δx und dem Mittelpunkt $x = c$ ist die genannte Wahrscheinlichkeit demnach näherungsweise gleich

$$f(c)\,\Delta x,$$

denn dies ist die Fläche eines Rechtecks mit der Breite Δx und der Höhe $f(c)$.

Da $F(x)$ stetig ist, so gehört übrigens zu einem Intervall $a < X \leqq b$ dieselbe Wahrscheinlichkeit wie zu dem Intervall $a < X < b$ oder $a \leqq X < b$ oder auch $a \leqq X \leqq b$. Dies ist also anders als bei den diskreten Verteilungen.

Aus demselben Grunde gilt auch für jede Zahl a

$$P(X = a) = 0.$$

Dies bedeutet natürlich *nicht*, daß $X = a$ ein unmögliches Ereignis ist. Sonst wären ja alle diese Ereignisse unmöglich. Hierauf haben wir in Abschn. 16 schon hingewiesen.

Da Wahrscheinlichkeiten nicht negativ sind und (26.3) für jedes Intervall gilt, so können wir für alle x

$$f(x) \geqq 0$$

voraussetzen.

Wie im diskreten Fall ist auch im stetigen die Wahrscheinlichkeitsverteilung jeweils durch die Verteilungsfunktion $F(x)$ [oder die Dichte $f(x)$] eindeutig bestimmt.

Beispiel 26.1 (Gleichförmige Verteilung). Beim Roulettespiel können wir die Endstellungen des sich drehenden Zeigers durch die Zufallsvariable

$$X = \textit{Winkel des Zeigers mit einer festen Richtung}, \quad 0 \leqq X < 2\pi$$

messen, wenn wir von der Feldeinteilung (auf die es beim tatsächlichen Spiel ankommt) absehen. Ist das Roulette völlig in Ordnung, so können wir annehmen, daß für jedes x zwischen 0 und 2π die Wahrscheinlichkeit $P(X \leq x)$ gleich dem Verhältnis dieses Winkels x zum gesamten Winkel 2π ist. So ergibt sich die Verteilungsfunktion (s. Abb. 26.2)

$$F(x) = \begin{cases} 0 & \text{für } x \leq 0 \\ x/2\pi & \text{für } 0 < x \leqq 2\pi \\ 1 & \text{für } x > 2\pi \end{cases}$$

und hieraus durch Differentiation die zugehörige Dichte

$$f(x) = \begin{cases} 1/2\pi & \text{für } 0 < x < 2\pi \\ 0 & \text{für alle übrigen } x. \end{cases}$$

Eine Verteilung mit der Wahrscheinlichkeitsdichte

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{für } a < x < b \\ 0 & \text{für alle übrigen } x \end{cases} \tag{26.4}$$

heißt eine **gleichförmige Verteilung** oder **Rechteckverteilung.** Unser Problem führt also auf eine spezielle solche Verteilung.

Weitere stetige Verteilungen werden später (in Kapitel 9–11) behandelt.

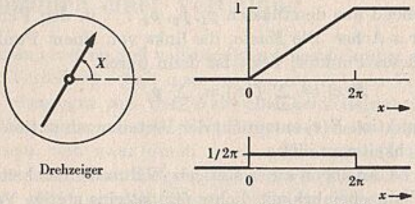


Abbildung 26.2. Zu Beispiel 26.1

Aufgaben zu Abschnitt 26

26.1 Man zeige, daß für jede Verteilungsfunktion $F(x)$ die folgenden Beziehungen gelten:

$$F(b) \geqq F(a) \quad (b > a), \qquad \lim_{x\to\infty} F(x) = 1.$$

26.2 Man zeige: Ist $P(X = a) = 0$, so ist die zugehörige Verteilungsfunktion $F(x)$ für $x = a$ stetig und umgekehrt.

26.3 Die Zufallsvariable X besitze eine Rechteckverteilung im Intervall $-1 \leq x \leq 1$. Wie lautet die Dichte und die Verteilungsfunktion? Wie groß sind die Wahrscheinlichkeiten $P(X \geq 0)$, $P(0 \leq X \leq 0{,}3)$, $P(X = 0{,}5)$?

26.4 Welchen Wert muß die Konstante k haben, damit die Funktion

$$f(x) = kx \text{ für } 0 \leqq x \leqq 2, \quad f(x) = 0 \text{ für alle übrigen } x$$

die Dichte einer stetigen Verteilung bildet? Wie lautet die Verteilungsfunktion?

26.5 Warum kann

$$f(x) = kx \text{ für } -1 \leqq x \leqq 1, \quad f(x) = 0 \text{ für alle übrigen } x$$

für keinen Wert der Konstanten k die Dichte einer Wahrscheinlichkeitsverteilung bilden?

26.6 Man betrachte die Verteilung mit der Dichte

$$f(x) = ce^{-\alpha x} \text{ für } x \geqq 0, \quad f(x) = 0 \text{ für } x < 0.$$

Der Parameter α muß positiv sein. Warum? Man bestimme c als Funktion von α. Man zeichne $f(x)$ und $F(x)$ für $\alpha = 0{,}5$ und 1.

27 Analogie zwischen Wahrscheinlichkeits- und Massenverteilungen

Zwischen den Wahrscheinlichkeitsverteilungen einer Zufallsvariablen X und den Verteilungen einer Masse vom Gesamtbetrage 1 längs einer Geraden besteht eine vollkommene Analogie:

Das mechanische Analogon einer diskreten Wahrscheinlichkeitsverteilung mit der Wahrscheinlichkeitsfunktion (22.1) ist eine diskrete Verteilung, bestehend aus den Massen $p_1, p_2, p_3, \ldots$ in den Punkten $x_1, x_2, x_3, \ldots$ der x-Achse Die Masse, die links von einem Punkte x (mit Einschluß dieses Punktes) liegt, ist dann durch

$$F(x) = \sum_{x_j \leq x} f(x_j) = \sum_{x_j \leq x} p_j$$

gegeben. Diese Funktion $F(x)$ entspricht der Verteilungsfunktion $F(x)$ der Wahrscheinlichkeitsverteilung.

Das mechanische Analogon einer stetigen Wahrscheinlichkeitsverteilung mit der Wahrscheinlichkeitsdichte $f(x)$ ist eine stetige Verteilung der Gesamtmasse 1 mit der Massendichte $f(x)$. Zwischen $-\infty$ und einem Wert x liegt dann die Masse

$$F(x) = \int_{-\infty}^{x} f(v)\, dv,$$

und in einem endlichen Intervall $a < x \leq b$ liegt die Masse

$$F(b) - F(a) = \int_{a}^{b} f(v)\, dv.$$

Diese Funktion $F(x)$ entspricht offenbar der Verteilungsfunktion $F(x)$ der betreffenden stetigen Wahrscheinlichkeitsverteilung.

Die betrachtete Analogie ist nicht nur qualitativ, sondern quantitativ. Sie läßt sich deshalb in vielen Fällen als Hilfsvorstellung bei Wahrscheinlichkeitsuntersuchungen verwenden. So kann man z. B. Beziehungen über Schwerpunkte, Trägheitsmomente usw. zur Verdeutlichung wahrscheinlichkeitstheoretischer Sachverhalte heranziehen. Hierauf wird später noch hingewiesen. Man spricht im Bereich der Wahrscheinlichkeitstheorie oft geradezu von einer „**Wahrscheinlichkeitsmasse**", die gemäß einer gegebenen Verteilungsfunktion $F(x)$ längs der x-Achse verteilt ist.

Aufgaben zu Abschnitt 27

27.1—27.3 Welche Massenverteilung entspricht der jeweils angegebenen Verteilung?

27.1 Die Verteilung in Beispiel 23.1.

27.2 Die Verteilung in Beispiel 23.2.

27.3 Die Verteilung in Beispiel 26.1.

Kapitel 6

Maßzahlen einer Verteilung

Einen genauen und vollständigen Überblick über alle Eigenschaften einer Verteilung erhält man aus der zugehörigen Verteilungsfunktion oder ebensogut aus der Wahrscheinlichkeitsfunktion bzw. Dichte. Daneben kann man eine Verteilung auch mehr „summarisch" kennzeichnen, und zwar durch gewisse Konstanten („Maßzahlen"), die sich aus der Verteilungsfunktion gewinnen lassen. Am wichtigsten davon sind der Mittelwert μ (Abschn. 28) und die Varianz σ^2 (Abschn. 29). Als weitere Maßzahl betrachten wir in Abschn. 32 noch die Schiefe γ.

μ und σ^2 sind Sonderfälle der sogenannten Momente einer Verteilung (Abschn. 31). Bei der Berechnung der Momente kann man sich manchmal gewisser Hilfsfunktionen bedienen, die wir in Abschn. 33 kennenlernen.

28 Mittelwert einer Verteilung

Der *Mittelwert einer Verteilung* wird mit μ bezeichnet. Er ist im Falle einer diskreten Verteilung durch

(28.1 a) $$\boxed{\mu = \sum_j x_j f(x_j)}$$

definiert und im Falle einer stetigen Verteilung durch

(28.1 b) $$\boxed{\mu = \int_{-\infty}^{\infty} x f(x)\, dx.}$$

Dabei ist $f(x)$ die Wahrscheinlichkeitsfunktion bzw. die Dichte der betreffenden Zufallsvariablen X. In (28.1a) wird über alle möglichen Werte x_j (vgl. Abschn. 25) summiert.

Statt *Mittelwert einer Verteilung* sagt man auch *Mittelwert der betreffenden Zufallsvariablen X*.

Beispiel 28.1. Die Zufallsvariable

$$X = \textit{Augenzahl beim Wurf eines regelmäßigen Würfels}$$

hat die Wahrscheinlichkeitsfunktion

$$f(x) = 1/6 \qquad \text{für } x = 1, 2, \ldots, 6$$

und $f(x) = 0$ für alle übrigen x. So ergibt sich aus (28.1a) der Mittelwert (vgl. Abb. 28.1)

$$\mu = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3{,}5.$$

Dies ist der durchschnittlich zu erwartende Wert in dem Sinne, daß zum Beispiel bei 1000 Würfen die Summe aller gewürfelten Zahlen *etwa* $1000 \cdot 3{,}5 = 3500$ betragen wird.

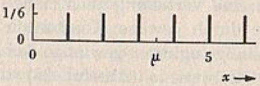


Abbildung 28.1. Wahrscheinlichkeitsfunktion $f(x)$ und Mittelwert μ in Beispiel 28.1

Beispiel 28.2. Die gleichförmige Verteilung mit der Dichte (26.4) hat den Mittelwert

$$\mu = \int_{-\infty}^{\infty} x f(x)\, dx = \int_a^b \frac{x}{b-a}\, dx = \frac{1}{b-a}\left(\frac{b^2}{2} - \frac{a^2}{2}\right) = \frac{a+b}{2}.$$

Weitere wichtige Beispiele folgen in Kapitel 8 und 9.

Statt *Mittelwert* sagt man auch *mathematischer Erwartungswert von* X oder kurz **Erwartung von** X und schreibt deshalb statt μ auch $E(X)$. Hierauf kommen wir im übernächsten Abschnitt zurück.

Bei der Definition des Mittelwertes wird vorausgesetzt, daß die Reihe in (28.1a) absolut konvergiert bzw. das Integral

$$\int_{-\infty}^{\infty} |x|\, f(x)\, dx$$

existiert. Trifft dies nicht zu, so sagt man, die betreffende Verteilung hat keinen Mittelwert. Derartige Verteilungen kommen praktisch nur selten vor (vgl. Aufgabe 28.6).

Der Mittelwert entspricht in der Mechanik dem Schwerpunkt der analogen Massenverteilung (vgl. Abschn. 27).

In der Tat entspricht (28.1) gerade der bekannten Formel für die Koordinate des Schwerpunktes der genannten Verteilung.

Eine Verteilung mit der Wahrscheinlichkeitsfunktion bzw. Dichte $f(x)$ heißt **symmetrisch** *bezüglich einer Zahl* $x = c$, wenn für jedes a

$$f(c + a) = f(c - a) \tag{28.2}$$

ist.

Satz 28.1. *Hat eine bezüglich $x = c$ symmetrische Verteilung einen Mittelwert μ, so ist*

$$\mu = c. \tag{28.3}$$

Dies ist auf Grund des mechanischen Analogons evident. Den einfachen formalen Beweis überlassen wir dem Leser. Aus diesem Satz folgen die Ergebnisse in Beispiel 28.1 und 28.2 nun ohne jede weitere Rechnung.

Aufgaben zu Abschnitt 28

28.1 Welchen Mittelwert hat die Verteilung in Beispiel 23.2?

28.2 Welche Summe der gewürfelten Zahlen kann man bei zwanzigmaligem Würfeln etwa erwarten? Man führe das Experiment aus.

28.3 Man bestimme den Mittelwert der Verteilung mit der Dichte $f(x) = xe^{-x}$ für $x > 0$ und $f(x) = 0$ für $x < 0$.

28.4 Man beweise den Satz 28.1 für eine stetige Verteilung.

28.5 Wie groß ist der Mittelwert der Verteilung in Beispiel 23.3?

28.6 Man zeige, daß die sogenannte **Cauchy-Verteilung**, deren Dichte $f(x) = 1/\pi(1 + x^2)$ ist, keinen Mittelwert besitzt.

28.7 Man zeige, daß $f(x)$ in Aufgabe 28.6 der Bedingung (26.2) genügt und daß $F(x) = \frac{1}{2} + \frac{1}{\pi} \arctan x$ die zugehörige Verteilungsfunktion ist.

28.8 Man zeige: Bei diskreten Zufallsvariablen mit nur *endlich* vielen möglichen Werten und stetigen Zufallsvariablen, die nur Werte in einem *endlichen* Intervall annehmen können, existiert der Mittelwert stets.

28.9 Wie groß ist der Mittelwert der Variablen X = *Anzahl der Würfe mit einem regelmäßigen Würfel bis zum Eintreffen der ersten Sechs* (der Wurf, bei dem sie eintrifft, mitgezählt)?

28.10 Man zeige: Kann eine Zufallsvariable nur Werte aus einem endlichen Intervall annehmen, so liegt auch ihr Mittelwert in diesem Intervall.

29 Varianz einer Verteilung

Die *Varianz einer Verteilung* wird mit σ^2 bezeichnet. Sie ist im diskreten Falle durch

$$\boxed{\sigma^2 = \sum_j (x_j - \mu)^2 f(x_j)} \tag{29.1a}$$

(Summation über alle möglichen Werte x_j; vgl. Abschn. 25) und im stetigen Falle durch

(29.1 b)
$$\boxed{\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx}$$

definiert, vorausgesetzt, daß die Reihe in (29.1a) konvergiert bzw. das Integral in (29.1b) existiert. Hierbei ist $f(x)$ die Wahrscheinlichkeitsfunktion bzw. Dichte der betreffenden Verteilung.

Statt *Varianz einer Verteilung* sagt man auch *Varianz der betreffenden Zufallsvariablen.*

Ist $f(x)$ im Falle einer diskreten Verteilung in einem einzigen Punkt gleich 1 und sonst 0, so ist die Varianz gleich null. Dieser Fall ist ohne praktisches Interesse. In allen übrigen Fällen gilt

(29.2)
$$\sigma^2 > 0.$$

Die positive Quadratwurzel der Varianz heißt die **Standardabweichung** und wird mit σ bezeichnet.

Im Englischen hat sich für σ^2 die Bezeichnung *variance* durchgesetzt. Im Deutschen ist die Terminologie leider noch nicht einheitlich: σ^2 wird bisweilen als **Streuung** bezeichnet. Andere Autoren verstehen unter „Streuung" unsere Standardabweichung σ.

Die Varianz ist, grob gesprochen, ein Maß für die Streuung der Werte, die die betreffende Zufallsvariable X annehmen kann.

In der Tat: Ist die Varianz im diskreten Falle klein, so muß jedes Glied in (29.1a) klein sein. Also müssen Werte x_j, die weit von μ entfernt liegen, eine geringe Wahrscheinlichkeit besitzen. Große Abweichungen der Variablen X vom Mittelwert μ sind demnach bei kleiner Varianz ziemlich unwahrscheinlich. Umgekehrt folgt bei großer Varianz, daß nicht alle x_j nahe bei μ liegen können. Ähnlich schließt man im stetigen Falle.

Aus (29.1) erkennt man weiterhin:

Das mechanische Analogon (s. Abschn. 27) *der Varianz ist das Trägheitsmoment der entsprechenden Verteilung einer Einheitsmasse bezüglich des Schwerpunktes.*

Beispiel 29.1. Die Zufallsvariable

X = Anzahl der „Köpfe" beim einmaligen Wurf einer Münze

hat die möglichen Werte $X = 0$ und $X = 1$ mit den Wahrscheinlichkeiten $P(X = 0) = 1/2$ und $P(X = 1) = 1/2$. Der Mittelwert ist

$$\mu = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2}.$$

Aus (29.1a) erhalten wir demnach die Varianz

$$\sigma^2 = \left(0 - \frac{1}{2}\right)^2 \cdot \frac{1}{2} + \left(1 - \frac{1}{2}\right)^2 \cdot \frac{1}{2} = \frac{1}{4}.$$

Beispiel 29.2. Die Rechteckverteilung (26.4) hat den Mittelwert $\mu = (a+b)/2$ und gemäß (29.1b) die Varianz

$$\sigma^2 = \int_a^b \left(x - \frac{a+b}{2}\right)^2 \frac{1}{b-a}\,dx = \frac{(b-a)^2}{12}.$$

Die Sonderfälle mit den Dichten

$$f_1(x) = \begin{cases} 1 & \text{für } 0 \leqq x \leqq 1 \\ 0 & \text{für alle übrigen } x \end{cases} \qquad f_2(x) = \begin{cases} 1/3 & \text{für } -1 \leqq x \leqq 2 \\ 0 & \text{für alle übrigen } x \end{cases}$$

haben denselben Mittelwert (1/2), aber die Varianz der ersten Verteilung (1/12) ist kleiner als die der zweiten (3/4), wie man schon aus Abb. 29.1 schließen kann.

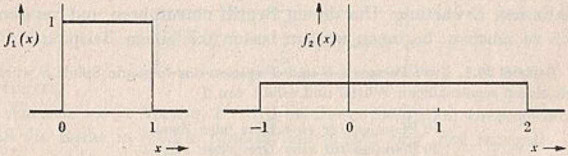


Abbildung 29.1. Beispiel gleichförmiger Verteilungen mit demselben Mittelwert und verschiedener Varianz

Aufgaben zu Abschnitt 29

29.1 Man berechne den Mittelwert und die Varianz der Verteilung mit der Dichte

$$f(x) = x/2 \text{ für } 0 \leqq x \leqq 2, \quad f(x) = 0 \text{ für alle übrigen } x.$$

29.2 Man gebe je ein möglichst einfaches Beispiel einer diskreten und einer stetigen Verteilung mit einer sehr großen Varianz (etwa $\sigma^2 > 1000$) und einer sehr kleinen Varianz (etwa $\sigma^2 < 1/1000$) an.

29.3 Eine Maschine stellt Bolzen für Bauzwecke her. Keine Produktion ist so vollkommen, daß alle Stücke absolut gleich ausfallen. In diesem Sinne ist der Durchmesser X [cm] der Bolzen eine Zufallsvariable. Wir nehmen an, ihre Verteilung habe die Dichte

$$f(x) = k(x - 0{,}9)(1{,}1 - x) \qquad \text{für } 0{,}9 < x < 1{,}1$$

und $f(x) = 0$ für alle übrigen x. Man bestimme die Konstante k aus (26.2), stelle $f(x)$ graphisch dar und bestimme den Mittelwert und die Varianz.

29.4 Wieviel Prozent Ausschuß hat man in Aufgabe 29.3 zu erwarten, wenn man Bolzen zum Ausschuß rechnet, deren Durchmesser um mehr als 0,6 mm von 1 cm abweichen?

29.5 Wie muß man in Aufgabe 29.3 die Grenzen der höchstens zulässigen Abweichung wählen, damit man nicht mehr als 10% Ausschuß zu erwarten hat?

29.6 Eine Tankstelle auf einer Farm wird jeweils am Wochenende mit Benzin beliefert. Die pro Woche verbrauchte Benzinmenge X [cbm] ist eine Zufallsvariable. Wir nehmen an, X besitze die Dichte

$$f(x) = 6x(1-x) \qquad \text{für } 0 \leqq x \leqq 1$$

und $f(x) = 0$ für alle übrigen x. Man berechne den Mittelwert und die Varianz.

29.7 Wie groß muß in Aufgabe 29.6 der Tank sein, damit die Wahrscheinlichkeit, daß der Vorrat in einer bestimmten Woche erschöpft wird, nur 10% beträgt?

30 Mathematische Erwartung

In den folgenden Abschnitten benötigen wir den Begriff der mathematischen Erwartung. Um diesen Begriff einzuführen und verständlich zu machen, beginnen wir am besten mit einem Beispiel.

Beispiel 30.1. Zwei Personen S und T spielen das folgende Spiel: S würfelt mit einem regelmäßigen Würfel und erhält von T

10 Pfennige für eine Eins oder Zwei,
20 Pfennige für eine Drei oder Vier,
40 Pfennige für eine Fünf und
80 Pfennige für eine Sechs.

S sollte an T vor jedem Spiel einen Betrag zahlen, der gleich seiner durchschnittlichen Gewinnerwartung pro Spiel ist. Diese beträgt

$$(30.1) \quad 10\cdot\frac{1}{6} + 10\cdot\frac{1}{6} + 20\cdot\frac{1}{6} + 20\cdot\frac{1}{6} + 40\cdot\frac{1}{6} + 80\cdot\frac{1}{6} = 30 \text{ [Pfennige]},$$

denn die Wahrscheinlichkeit, eine Eins zu würfeln, beträgt 1/6, und er erhält dann 10 Pfennige, usw.

Wir haben hier jedem möglichen Wert der Zufallsvariablen

X = *Erzielte Zahl beim Wurf eines regelmäßigen Würfels*

eine Zahl, nämlich den jeweiligen Gewinn von S, zugeordnet und auf diese Weise eine Funktion von X erhalten. Diese bezeichnen wir mit $g(X)$. Ihre Funktionswerte sind:

X	1	2	3	4	5	6
$g(X)$	10	10	20	20	40	80

Da die Werte von X vom Zufall abhängen, gilt dies natürlich auch für den Wert, den $g(X)$ bei einem Spiel jeweils annimmt. $g(X)$ ist also selbst eine Zufallsvariable.

Der Ausdruck (30.1), der die (durchschnittliche Gewinn-) Erwartung des Spielers S pro Spiel angibt, hängt natürlich von $g(X)$ ab, und wir bezeichnen ihn mit $E(g(X))$. Er hat offenbar die Form

$$\boxed{E(g(X)) = \sum_j g(x_j)\, f(x_j).} \tag{30.2a}$$

Hierbei ist $f(x)$ die zu X gehörige Wahrscheinlichkeitsfunktion. Summiert wird über alle möglichen Werte x_j (vgl. Abschn. 25).

Ganz allgemein, d. h. für eine beliebige diskrete Zufallsvariable X und eine für alle möglichen Werte von X definierte reellwertige Funktion $g(X)$ heißt der durch (30.2a) gegebene Ausdruck $E(g(X))$ der *mathematische Erwartungswert der Funktion* $g(X)$ oder kurz die **Erwartung** *von* $g(X)$.

Entsprechend ist im Falle einer stetigen Verteilung mit der Dichte $f(x)$ die Erwartung einer für alle X erklärten Funktion $g(X)$ durch

$$\boxed{E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x)\, dx} \tag{30.2b}$$

definiert.

Ähnlich wie in Abschn. 28 wird bei dieser Definition vorausgesetzt, daß die Reihe in (30.2a) absolut konvergiert bzw. das Integral

$$\int_{-\infty}^{\infty} |g(x)|\, f(x)\, dx$$

existiert.

Oftmals hat man die Erwartung einer Funktion der Form

$$a g(X) + b h(X) \qquad (a, b \text{ konstant})$$

aus den Erwartungen von $g(X)$ und $h(X)$ zu berechnen. Dies geschieht mittels der wichtigen Formel

$$E(a g(X) + b h(X)) = a E(g(X)) + b E(h(X)). \tag{30.3}$$

Den einfachen Beweis dieser Formel überlassen wir dem Leser.

Aufgaben zu Abschnitt 30

30.1 Man beweise (30.3).

30.2 Ein Spieler S würfelt mit zwei regelmäßigen Würfeln. Sein Mitspieler T zahlt an S so viele Pfennige, wie das Produkt der beiden jeweils erzielten Zahlen beträgt. Wieviel sollte S an T pro Spiel zahlen, damit das Spiel fair ist?

30.3 Bei einer Lotterie werden 20000 Lose zum Preise von je 1 Mark verkauft. Der Gewinn ist ein Wagen im Werte von 10000 Mark. Wie groß ist die Gewinnerwartung eines Teilnehmers, der 3 Lose kauft?

31 Momente einer Verteilung

Wir wollen uns überlegen, daß der Mittelwert und die Varianz Sonderfälle allgemeinerer Größen sind, die man als *Momente* bezeichnet.

Wählen wir in (30.2) speziell $g(X) = X^k$ $(k = 1, 2, \ldots)$, so ergibt sich im diskreten bzw. stetigen Falle

$$(31.1) \qquad E(X^k) = \sum_j x_j^k f(x_j) \quad \text{bzw.} \quad E(X^k) = \int_{-\infty}^{\infty} x^k f(x)\, dx.$$

Diese Größe heißt das k-**te Moment** der betreffenden Verteilung oder der betreffenden Zufallsvariablen X.

Für $k = 1$ wird (31.1) mit (28.1) identisch. Dies bedeutet:

Das 1. Moment ist der Mittelwert μ der betreffenden Verteilung,

$$(31.2) \qquad \boxed{\mu = E(X).}$$

Wählen wir in (30.2) speziell $g(X) = (X - \mu)^k$, so erhalten wir im diskreten bzw. stetigen Falle

$$(31.3) \qquad E([X - \mu]^k) = \sum_j (x_j - \mu)^k f(x_j) \quad \text{bzw.} \quad \int_{-\infty}^{\infty} (x - \mu)^k f(x)\, dx.$$

Diese Größe heißt das k-**te zentrale Moment** der betreffenden Verteilung oder der betreffenden Zufallsvariablen X.

Existiert das 1. zentrale Moment, so hat es den Wert null.

Mechanisches Analogon: *Das auf den Schwerpunkt bezogene statische Moment hat den Wert null.*

Den sehr einfachen Beweis überlassen wir dem Leser.

Für $k = 2$ wird (31.3) mit (29.1) identisch. Es gilt also:

Das 2. zentrale Moment ist die Varianz der betreffenden Verteilung,

$$(31.4) \qquad \boxed{\sigma^2 = E([X - \mu]^2).}$$

Die zentralen Momente lassen sich durch die Momente $E(X)$, $E(X^2)$, ... ausdrücken.

Zum Beispiel erhalten wir unter Benutzung von (30.3) und

$$E(1) = \sum_j f(x_j) = 1 \qquad \text{bzw.} \qquad E(1) = \int_{-\infty}^{\infty} f(x)\, dx = 1$$

für das 2. zentrale Moment sofort

$$\sigma^2 = E([X - \mu]^2) = E(X^2 - 2\mu X + \mu^2) = E(X^2) - 2\mu E(X) + \mu^2.$$

Auf der rechten Seite ist $E(X) = \mu$, und so ergibt sich einfach

$$\boxed{\sigma^2 = E(X^2) - \mu^2\,.} \tag{31.5}$$

Entsprechend beweist man

$$E([X - \mu]^3) = E(X^3) - 3\mu E(X^2) + 2\mu^3 \tag{31.6}$$

usw.

Beispiel 31.1. Im Falle der Zufallsvariablen

$$X = \textit{Erzielte Zahl beim Wurf eines regelmäßigen Würfels}$$

ist

$$\mu = E(X) = \frac{1}{6}(1 + 2 + \cdots + 6) = \frac{7}{2},$$

$$E(X^2) = \frac{1}{6}(1^2 + 2^2 + \cdots + 6^2) = \frac{91}{6}.$$

Aus (31.5) ergibt sich damit

$$\sigma^2 = \frac{91}{6} - \frac{49}{4} = \frac{35}{12}.$$

Aufgaben zu Abschnitt 31

31.1 Man beweise: Existiert das 1. zentrale Moment einer Verteilung, so hat es den Wert null.

31.2 Man leite (31.6) her.

31.3 Der Ausdruck

$$E([X - c]^2) \qquad (c \text{ irgendeine reelle Zahl})$$

heißt das **2. Moment von X bezüglich** c. Man zeige: Existiert dieses Moment, so hat es genau dann seinen kleinsten Wert, wenn $c = \mu$ ist.

31.4 Was ist das mechanische Analogon der Aussage in Aufgabe 31.3?

31.5 Man beweise

$$\sigma^2 = E(X[X - 1]) + \mu - \mu^2. \tag{31.7}$$

31.6 Der **Steinersche Satz** der Mechanik lautet im Falle einer eindimensionalen Massenverteilung:

$$I_P = I_S + a^2 M.$$

Hierbei ist M die Gesamtmasse, a der Abstand eines Punktes P vom Schwerpunkt S und I_P bzw. I_S das Trägheitsmoment bezüglich P bzw. S. Man überlege sich, daß (31.5) das Analogon dieses Satzes ist.

31.7 Man berechne σ^2 in Beispiel 31.1, ohne (31.5) anzuwenden, und überzeuge sich davon, daß (31.5) Vorteile bietet.

31.8 Man berechne das 3. zentrale Moment in Beispiel 31.1. Kann man auch ohne Rechnung einsehen, daß sich null ergeben muß?

32 Schiefe einer Verteilung

Wie der Leser beweisen möge (s. Aufgabe 32.1), gilt der

Satz 32.1. *Ist eine Verteilung symmetrisch bezüglich $x = \mu$ (vgl. Abschn. 28) und existiert das 3. zentrale Moment der Verteilung, so ist dieses null.*

Ist das genannte Moment null, so braucht die betreffende Verteilung aber nicht symmetrisch zu sein. Trotzdem sieht man dieses Moment und ebenso die sogenannte **Schiefe**

$$\boxed{\gamma = \frac{1}{\sigma^3} E([X - \mu]^3)} \tag{32.1}$$

bei den in der Praxis vorkommenden Verteilungen gern als Maße für die Asymmetrie an. Man unterscheidet dabei zwischen symmetrischen Verteilungen und Verteilungen mit positiver und negativer Schiefe.

Beispiel 32.1. Die Verteilung in Beispiel 31.1 ist symmetrisch bezüglich $x = 7/2$. Aus Satz 32.1 folgt demnach, daß das 3. zentrale Moment der Verteilung null sein muß.

Beispiel 32.2. Für die Verteilung mit der Dichte (s. Abb. 32.1)

$$f(x) = \begin{cases} 0 & \text{für negative } x \\ x e^{-x} & \text{für positive } x \end{cases}$$

erhält man unter wiederholter Teilintegration den Mittelwert

$$\mu = E(X) = \int_0^\infty x^2 e^{-x}\, dx = 2$$

und weiterhin die Momente

$$E(X^2) = \int_0^\infty x^3 e^{-x}\, dx = 6, \qquad E(X^3) = \int_0^\infty x^4 e^{-x}\, dx = 24.$$

Gemäß (31.5) ergibt sich die Varianz

$$\sigma^2 = 6 - 4 = 2$$

und gemäß (31.6) das 3. zentrale Moment

$$E([X - \mu]^3) = 24 - 3 \cdot 2 \cdot 6 + 2 \cdot 8 = 4.$$

Es handelt sich also um eine Verteilung mit positiver Schiefe $\left(\gamma = 4/2\sqrt{2} = \sqrt{2}\right)$. Dieses Ergebnis ist verständlich, denn die Kurve von $f(x)$ hat einen langen „Schwanz" nach der positiven Seite hin. Die Potenz $(x - \mu)^3$ hat für große x einen sehr großen Wert, so daß der $x > \mu = 2$ entsprechende positive Beitrag zum 3. zentralen Moment den $0 \le x < 2$ entsprechenden negativen Beitrag überwiegt.

Aufgaben zu Abschnitt 32

32.1. Man beweise den Satz 32.1.

32.2 Man führe die Integralberechnungen in Beispiel 32.2 durch.

32.3 Man berechne die Schiefe der Verteilung mit der Dichte

$f(x) = 2(1 - x)$ für $0 < x < 1$, $f(x) = 0$ für alle übrigen x.

32.4 Man berechne das k-te Moment der Verteilung in Aufgabe 32.3.

32.5 Man zeige: Die Zufallsvariable X mit $P(X = 1) = 1/2$, $P(X = -4) = 1/3$, $P(X = 5) = 1/6$ hat $\gamma = 0$, ist aber nicht symmetrisch verteilt.

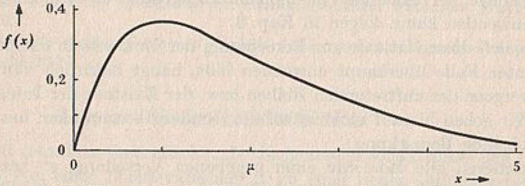


Abbildung 32.1. Wahrscheinlichkeitsdichte in Beispiel 32.2

33 Momenterzeugende und charakteristische Funktion

Wie wir später sehen werden, berechnet man bei manchen speziellen Verteilungen die Momente am einfachsten direkt aus den Definitionsgleichungen, während man in anderen Fällen durch die Benutzung einer geeigneten Hilfsfunktion, nämlich der Erwartung von e^{tX}, leichter zum Ziele kommt. Diese Erwartung $E(e^{tX})$ heißt die **momenterzeugende Funktion** der betreffenden Verteilung und wird mit $G(t)$ bezeichnet. Auf Grund der Definition der Erwartung einer Funktion ist also im diskreten bzw. im stetigen Falle

$$(33.1)\qquad G(t) = E(e^{tX}) = \sum_j e^{tx_j} f(x_j) \quad \text{bzw.} \quad \int_{-\infty}^{\infty} e^{tx} f(x)\,dx.$$

Hierbei ist $e = 2{,}718\ldots$ die Basis der natürlichen Logarithmen.

Nehmen wir an, daß wir unter dem Summen- bzw. Integralzeichen nach t differenzieren dürfen, so ergibt sich

$$G'(t) = \sum_j x_j e^{tx_j} f(x_j) \quad \text{bzw.} \quad \int_{-\infty}^{\infty} x e^{tx} f(x)\,dx.$$

Entsprechend liefert k-malige Differentiation nach t

$$\frac{d^k G}{dt^k} = \sum_j x_j^k e^{tx_j} f(x_j) \quad \text{bzw.} \quad \int_{-\infty}^{\infty} x^k e^{tx} f(x)\,dx.$$

Setzen wir nun $t = 0$, so hat die Exponentialfunktion den Wert 1, und auf der rechten Seite steht dann gerade das k-te Moment. Demnach gilt

$$\text{(33.2)} \qquad \boxed{E(X^k) = G^{(k)}(0)} \qquad \left(G^{(k)}(0) = \frac{d^k G}{dt^k}\bigg|_{t=0}\right).$$

Insbesondere wird für $k = 1$

$$\text{(33.3)} \qquad \mu = E(X) = G'(0).$$

Beispiele, bei denen man die momenterzeugende Funktion mit Vorteil anwenden kann, folgen in Kap. 8.

Ob sich diese Methode zur Berechnung der Momente in einem bestimmten Falle überhaupt anwenden läßt, hängt natürlich von der Konvergenz der auftretenden Reihen bzw. der Existenz der Integrale ab. Wir gehen hierauf nicht näher ein, sondern beschränken uns auf die folgende Bemerkung:

Existieren alle Momente einer gegebenen Verteilung, so hat die Taylorentwicklung von $G(t)$ nach Potenzen von t wegen (33.2) die Form

$$G(t) = \sum_{k=0}^{\infty} \frac{G^{(k)}(0)}{k!} t^k = \sum_{k=0}^{\infty} \frac{E(X^k)}{k!} t^k.$$

Man kann zeigen: Hat diese Reihe einen positiven Konvergenzradius, so ist die betreffende Verteilung durch $G(t)$ [d. h. durch die Momente $E(X^k)$, $k = 1, 2, \ldots$] eindeutig bestimmt.

Aufgaben zu Abschnitt 33

33.1 Ein Ziel wird beschossen. p sei die Trefferwahrscheinlichkeit je Schuß. p sei konstant. Die Zufallsvariable X bezeichne die Anzahl der Schüsse, die dem ersten Treffer vorausgehen. Man überlege sich, daß X die Wahrscheinlichkeitsfunktion

$$f(x) = q^x p \qquad \text{mit} \qquad q = 1 - p \qquad (x = 0, 1, \ldots)$$

hat, und bestimme die momenterzeugende Funktion. (Die zugehörige Verteilung heißt die **geometrische Verteilung.**)

33.2 Man berechne den Mittelwert und die Varianz der geometrischen Verteilung in Aufgabe 33.1.

33.3 Man kontrolliere die Ergebnisse in Beispiel 31.1 unter Benutzung der momenterzeugenden Funktion.

33.4 Wie lautet die momenterzeugende Funktion der Verteilung in Beispiel 23.3?

33.5 Konvergenzschwierigkeiten, wie sie in (33.1) auftreten können, bestehen nicht, wenn man die komplexe Funktion

$$\text{(33.4)} \qquad \Psi(t) = E(e^{itX}), \qquad \left(i = \sqrt{-1}\right)$$

die sogenannte **charakteristische Funktion der Verteilung**, benutzt. Man zeige, daß diese bei jeder diskreten oder stetigen Verteilung für jedes reelle t existiert.

33.6 Man zeige, daß, falls $E(X^k)$ existiert,

$$E(X^k) = \Psi^{(k)}(0)/i^k$$

mit $\Psi^{(k)} = d^k \Psi/dt^k$ und $i = \sqrt{-1}$ ist.

33.7 Man berechne den Mittelwert und die Varianz der Verteilung mit der Dichte

$$f(x) = ae^{-ax} \quad \text{für} \quad x \geqq 0, \qquad f(x) = 0 \quad \text{für} \quad x < 0,$$

und zwar (a) direkt aus den Definitionsformeln, (b) mittels der momenterzeugenden Funktion, (c) mittels der charakteristischen Funktion. Hierbei sei $a > 0$.

33.8 Für welche Werte von t existiert die momenterzeugende Funktion in Aufgabe 33.7?

34 Lineare Skalentransformation

Oft erfordern theoretische oder rechentechnische Gründe den Übergang von einer Zufallsvariablen X zu einer neuen Zufallsvariablen

$$\boxed{X^* = c_1 X + c_2} \qquad (c_1 \neq 0). \tag{34.1}$$

Im Zusammenhang mit einer solchen „*linearen Skalentransformation*" muß man dann aus den zu X gehörenden Größen (Mittelwert μ, Varianz σ^2 usw.) die zu X^* gehörigen berechnen. Letztere kennzeichnen wir durch einen Stern.

Für den Mittelwert ergibt sich wegen (30.3) sofort

$$\mu^* = E(X^*) = E(c_1 X + c_2) = c_1 E(X) + c_2 E(1).$$

Wegen $E(X) = \mu$ und $E(1) = 1$ ist dies gleichbedeutend mit

$$\boxed{\mu^* = c_1 \mu + c_2.} \tag{34.2}$$

Damit wird nun

$$X^* - \mu^* = (c_1 X + c_2) - (c_1 \mu + c_2) = c_1 (X - \mu).$$

Für das k-te zentrale Moment gilt also wegen (30.3) einfach

$$E([X^* - \mu^*]^k) = E(c_1{}^k [X - \mu]^k) = c_1{}^k E([X - \mu]^k). \tag{34.3}$$

Insbesondere haben wir demnach für die Varianz ($k = 2$)

$$\boxed{\sigma^{*2} = c_1{}^2 \, \sigma^2 .} \tag{34.4}$$

Da c_2 in (34.4) nicht auftritt, gilt der einleuchtende

Satz 34.1. *Die Varianz ist invariant gegen Nullpunktsverschiebungen, d. h. Transformationen der Form $X^* = X + c_2$.*

Wählen wir in (34.1) speziell $c_1 = 1/\sigma$ und $c_2 = -\mu/\sigma$, so hat X^* die Form $(X - \mu)/\sigma$. Diese Variable bezeichnen wir mit Z. Aus (34.2) und (34.4) ergibt sich dann, daß Z den Mittelwert 0 und die Varianz 1 hat. Es gilt also der

Satz 34.2. *Hat die Zufallsvariable X den Mittelwert μ und die Varianz σ^2, so hat die Zufallsvariable*

$$Z = \frac{X - \mu}{\sigma}$$

den Mittelwert 0 und die Varianz 1.

Z bezeichnet man bisweilen als die zu X gehörige **Zufallsvariable in Standardform.**

Aufgaben zu Abschnitt 34

34.1 Man beweise: Hat X die momenterzeugende Funktion $G(t)$, so hat $X^* = c_1 X + c_2$ die momenterzeugende Funktion

$$G^*(t) = e^{c_2 t}\, G(c_1 t). \tag{34.5}$$

34.2 Welchen Mittelwert und welche Varianz hat $Y = 2X - 2$, wenn X die Dichte

$$f(x) = 0{,}75(1 - x^2) \qquad \text{für } -1 \leq x \leq 1$$

und $f(x) = 0$ für alle übrigen x hat?

34.3 Man zeige, daß

$$\sigma^2 = G^{*\prime\prime}(0) \qquad \text{mit} \qquad G^*(t) = E(e^{(X-\mu)t}) \tag{34.6}$$

ist.

34.4 Man zeige, daß in Aufgabe 34.3

$$G^*(t) = e^{-\mu t} G(t)$$

ist, und gewinne daraus (31.5), indem man zweimal differenziert und (33.2) benutzt.

Kapitel 7

Permutationen und Kombinationen

Im vorliegenden Kapitel betrachten wir die grundlegenden Sätze über Permutationen und Kombinationen und die wichtigsten Eigenschaften der Fakultäten und Binomialkoeffizienten. Diese Sätze benötigt man in der Wahrscheinlichkeitstheorie und Statistik häufig. Zahlentafeln findet man in Anhang 5 (Tafel 1a—1c).

35 Permutationen. Fakultäten

Gegeben seien n verschiedene Dinge („**Elemente**"). Jede Zusammenstellung, die dadurch entsteht, daß man sämtliche n Elemente in irgendeiner Reihenfolge nebeneinander setzt, heißt eine **Permutation** der gegebenen Elemente.

Zum Beispiel gibt es 6 Permutationen der 3 Buchstaben a, b, c, nämlich

$$\begin{matrix} abc & bac & cab \\ acb & bca & cba. \end{matrix}$$

Allgemein gilt der

Satz 35.1. *Die Anzahl der Permutationen von n verschiedenen Elementen ist gleich*

(35.1) $$n! = 1 \cdot 2 \cdot 3 \cdots n$$ (lies „n Fakultät").

Beweis. Dies ergibt sich durch vollständige Induktion oder auch so: Man hat n Möglichkeiten, den ersten Platz in der Zusammenstellung zu besetzen. Dann sind noch $n - 1$ Elemente übrig. Also hat man nun $n - 1$ Möglichkeiten für die Besetzung des 2. Platzes, usw.

Beispiel 35.1. In einer Urne befinden sich 6 Lose mit den Nummern von 1 bis 6. Wie groß ist die Wahrscheinlichkeit, diese Lose nacheinander ohne Zurücklegen in der Reihenfolge 1, 2, ..., 6 zu ziehen?

Es gibt $6! = 720$ gleichmögliche Fälle, d. h. 720 verschiedene Reihenfolgen, die gleichwahrscheinlich sind und sich gegenseitig ausschließen. Also beträgt die gesuchte Wahrscheinlichkeit nur 1/720 oder 0,14%.

Für Permutationen von Elementen, die nicht alle verschieden sind, gilt der

Satz 35.2. *Gegeben seien n Elemente, die nicht alle voneinander verschieden sind, sondern in c Klassen untereinander gleicher Elemente zerfallen. Die 1. Klasse umfasse n_1,* die *2. Klasse $n_2, \ldots$ und die c-te Klasse n_c Elemente. Elemente aus verschiedenen Klassen seien verschieden. Dann ist die Anzahl der Permutationen dieser Elemente gleich*

$$\frac{n!}{n_1!\,n_2!\cdots n_c!} \qquad (n_1+n_2+\cdots+n_c=n).$$

Diesen Satz möge der Leser beweisen (s. Aufgabe 35.1). Ein einfaches Anwendungsbeispiel findet man in Aufgabe 35.3.

Nun folgen noch einige wichtige Bemerkungen über die **Fakultät.** Man definiert

$$0! = 1. \tag{35.2}$$

Grundlegend ist die Funktionalbeziehung

$$\boxed{(n+1)! = (n+1)\,n!} \tag{35.3}$$

Tafel 1a in Anhang 5 läßt erkennen, daß die Werte der Fakultät außerordentlich rasch anwachsen. Deshalb begnügt man sich bei großem n oft mit Näherungswerten von $n!$, die sich einfach berechnen lassen. Solche Werte liefert die **Stirling-Formel**

$$\boxed{n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n} \qquad (e = 2{,}718\ldots) \tag{35.4}$$

(Beweis siehe z. B. in A. OSTROWSKI, Vorlesungen über Differential- und Integralrechnung. Basel: Birkhäuser, 1945–1954).

Man kann zeigen: *Der relative Fehler der Näherung* (35.4) *strebt mit wachsendem n gegen null, während der absolute Fehler dann gegen unendlich strebt.* Dies sollte man bei der Anwendung der STIRLING-Formel im Bewußtsein behalten.

Tabelle 35.1. Zur Genauigkeit der STIRLING-Formel (35.4)

n	Näherung (35.4)	$n!$ exakt	Fehler (abgerundet) absolut	Fehler (abgerundet) relativ
4	23,5	24	0,5	2 %
6	710	720	10	1,4 %
8	39902	40320	400	1 %
10	3598696	3682800	30000	0,8 %

Aufgaben zu Abschnitt 35

35.1 Man beweise den Satz 35.2.

35.2 Man bestätige das Ergebnis in Beispiel 35.1, indem man (19.4) auf die Ereignisse

$$A_k\text{: Los-Nr. } k \text{ beim } k\text{-ten Zug} \qquad (k = 1, \ldots, 6)$$

anwendet.

35.3 Aus einer Urne, die 2 schwarze, 3 weiße und 4 rote Kugeln enthält, wird eine Kugel nach der anderen ohne Zurücklegen gezogen. Wie groß ist die Wahrscheinlichkeit, daß man dabei zuerst die beiden schwarzen, dann die 3 weißen und bei den letzten 4 Zügen die 4 roten Kugeln zieht?

35.4 Wie kann man Aufgabe 35.3 mittels (19.4) behandeln?

35.5 Man schreibe alle Permutationen der Buchstaben $a\,b\,c\,d$ auf.

35.6 Ein zerstreuter Professor hat 3 Briefe geschrieben und die zugehörigen Umschläge adressiert. In jeden Umschlag legt er einen Brief, und zwar ganz zufällig. Wie groß ist die Wahrscheinlichkeit, daß dabei wenigstens ein Brief in den richtigen Umschlag kommt?

36 Kombinationen ohne Wiederholung

Sind n Elemente gegeben, greift man k ($\leqq n$) Elemente heraus und stellt man diese in irgendeiner Anordnung nebeneinander, so nennt man die erhaltene Zusammenstellung eine **Kombination k-ter Ordnung** der gegebenen Elemente.

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.**

Zum Beispiel können wir aus den 3 Buchstaben a, b, c, die folgenden Kombinationen 2. Ordnung bilden:

$$\begin{array}{ccc} ab & ac & bc \\ ba & ca & cb. \end{array}$$

Berücksichtigen wir die Anordnung, so sind dies 6 verschiedene Kombinationen. Berücksichtigen wir die Anordnung nicht, so sind die beiden jeweils untereinanderstehenden Kombinationen als gleich anzusehen, und es bleiben dann also nur 3 verschiedene Kombinationen übrig.

Sind die gegebenen Elemente alle verschieden, so sind natürlich auch die Elemente der daraus gebildeten Kombinationen alle verschieden, und man spricht dann von **Kombinationen ohne Wiederholung,** da keines der Elemente wiederholt auftreten kann. Wie der Leser zeigen möge, gilt der

Satz 36.1 (Kombinationen ohne Wiederholung mit Berücksichtigung der Anordnung). *Für n gegebene voneinander verschiedene Elemente ist die Anzahl der Kombinationen k-ter Ordnung mit Berücksichtigung der Anordnung gleich*

$$n(n-1)(n-2)\cdots(n-k+1) = \frac{n!}{(n-k)!}.$$

Die in Satz 36.1 genannten Kombinationen heißen im Englischen *permutations of n different things taken k at a time, without repetitions.* Hier weicht also der Sprachgebrauch vom Deutschen ab.

Nimmt man auf die Anordnung der Elemente in den Kombinationen keine Rücksicht, so fallen von den Kombinationen in Satz 36.1 jedesmal diejenigen zu einer einzigen zusammen, die die gleichen Elemente in verschiedener Anordnung enthalten. Gemäß Satz 35.1 sind dies bei Kombinationen k-ter Ordnung jedesmal $k!$ solcher Kombinationen. Aus dem Satz 36.1 folgt damit der

Satz 36.2 (Kombinationen ohne Wiederholung und ohne Berücksichtigung der Anordnung). *Die Anzahl der Kombinationen k-ter Ordnung von n verschiedenen Elementen ohne Berücksichtigung der Anordnung beträgt*

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)\cdots(n-k+1)}{1\cdot 2\cdots k}.$$

Die in diesem Satz genannten Kombinationen heißen im Englischen *combinations of n different things taken k at a time, without repetitions.*

Einige wichtige Formeln für Binomialkoeffizienten stellen wir im übernächsten Abschnitt zusammen.

Aufgaben zu Abschnitt 36

36.1 In einer Urne liegen 5 Zettel mit den Buchstaben a, b, c, d bzw. e. Ein Spieler zieht nacheinander 3 Zettel ohne Zurücklegen. Zieht er zuerst a, dann b und beim letzten Zug c, so hat er gewonnen und erhält 5 Mark. Anderenfalls hat er verloren und zahlt 10 Pfennige. Ist es wahrscheinlich, daß er bei oftmaliger Wiederholung des Spiels einen Überschuß erzielt?

36.2 Man beweise den Satz 36.1.

36.3 Wie ändert sich die Lage in Aufgabe 36.1, wenn man vereinbart, daß der Spieler schon gewinnt, wenn er a, b und c in irgendeiner Reihenfolge zieht?

36.4 Man löse Aufgabe 36.1 unter Anwendung von (19.4) auf die drei Ereignisse „*a beim 1. Zug*", „*b beim 2. Zug*", „*c beim 3. Zug*".

37 Kombinationen mit Wiederholung

Bei den Kombinationen ohne Wiederholung tritt jedes Element höchstens einmal auf. Wir wenden uns nun den Kombinationen mit Wiederholung zu, bei denen jedes Element mehrmals auftreten kann.

Zum Beispiel können wir aus den 3 Buchstaben a, b, c die folgenden Kombinationen 2. Ordnung mit Wiederholung bilden:

$$\begin{array}{ccc} ab & ac & bc \\ ba & ca & cb \\ aa & bb & cc. \end{array}$$

In den ersten beiden Zeilen stehen die schon zu Beginn des vorigen Abschnittes betrachteten Kombinationen. Neu hinzugekommen sind die in der letzten Zeile, weil wir nun auch Wiederholungen zulassen. Berücksichtigen wir die Anordnung, so haben wir $3^2 = 9$ verschiedene Kombinationen. Berücksichtigen wir die Anordnung nicht, so haben wir nur 6 verschiedene Kombinationen.

Bilden wir Kombinationen k-ter Ordnung von n verschiedenen unbeschränkt oft wiederholbaren Elementen, so haben wir n verschiedene Möglichkeiten, den 1. Platz in der Kombination zu besetzen, sodann unabhängig davon n Möglichkeiten bei der Besetzung des 2. Platzes, denn wir dürfen ja wiederholen, usw. So ergibt sich der

Satz 37.1 (Kombinationen mit Wiederholung und mit Berücksichtigung der Anordnung). *Die Anzahl der Kombinationen k-ter Ordnung von n verschiedenen unbeschränkt oft wiederholbaren Elementen mit Berücksichtigung der Anordnung beträgt*

$$n^k.$$

Durch Induktion nach k beweist man den

Satz 37.2 (Kombinationen mit Wiederholung ohne Berücksichtigung der Anordnung). *Die Anzahl der Kombinationen k-ter Ordnung von n verschiedenen unbeschränkt oft wiederholbaren Elementen ohne Berücksichtigung der Anordnung beträgt*

$$\binom{n+k-1}{k} = \frac{(n+k-1)(n+k-2)\cdots(n+1)\,n}{1\cdot 2\cdots k}.$$

Die in Satz 37.1 (bzw. Satz 37.2) genannten Kombinationen heißen im Englischen *permutations* (bzw. *combinations*) *of n different things taken k at a time, with repetitions.*

Aufgaben zu Abschnitt 37

37.1 Man beweise den Satz 37.2.

37.2 In einer Urne befinden sich Lose. Auf jedem Los stehen 3 (gleiche oder verschiedene) der Zahlen 1, 2, ..., 6 derart, daß jedes ungeordnete Tripel von Zahlen, das man aus den genannten Zahlen bilden kann, auf genau einem dieser Lose vorkommt. Unter Benutzung des Satzes 37.2 zeige man, daß die Wahrscheinlichkeit, bei einem einzelnen Zug das Los (6, 6, 6) zu erhalten, $1/56 \approx 2\%$ beträgt.

37.3 Bei der Chiffrierung von Telegrammen wird der Text mit Hilfe eines gegebenen Schlüssels verschlüsselt und beim Übersenden zu Gruppen von je 5 Buchstaben zusammengefaßt. Wieviel verschiedene solche Fünfergruppen der 26 Buchstaben des Alphabetes kann man bilden?

37.4 Wieviel verschiedene Kraftfahrzeugkennzeichen kann man herstellen, wenn diese aus 5 Symbolen bestehen, von denen die ersten beiden Symbole Buchstaben und die restlichen Symbole Ziffern sind?

37.5 Bei einem Zufallsexperiment sei die Wahrscheinlichkeit, einen positiven oder negativen Wert zu erhalten, gleich groß, also jeweils gleich 0,5. Man schreibe die 16 bei viermaliger Ausführung des Experimentes vorhandenen Möglichkeiten auf, nämlich $++++$, $-+++$, $+-++$ usw. Man unterklammere gleichartige nebeneinanderstehende Symbole und die dann noch übrigen Einzelsymbole ebenfalls, also

$$\underbrace{+\,+\,+\,+} \qquad \underbrace{-}\,\underbrace{+\,+\,+} \qquad \underbrace{+}\,\underbrace{-}\,\underbrace{+\,+}$$

usw. Man bestimme die Wahrscheinlichkeitsfunktion der Zufallsvariablen $V =$ *Anzahl der Klammern*, die wir später (in Abschn. 117) brauchen werden. (Diese Anzahl heißt die *Anzahl der Iterationen*.)

37.6 Man sehe die Plus- und Minuszeichen in Aufgabe 37.5 als Vorzeichen der Zahlen 1 2 3 4 an und bestimme die Verteilung der Zufallsvariablen $U =$ *Summe der negativen Zahlen*, die wir später brauchen werden.

38 Binomialkoeffizienten

Die Binomialkoeffizienten sind durch

$$\binom{\alpha}{k} = \frac{\alpha(\alpha-1)(\alpha-2)\cdots(\alpha-k+1)}{k!} \tag{38.1}$$

definiert. Hierbei ist k stets eine nichtnegative ganze Zahl, während α keiner Einschränkung unterliegt. Weiterhin definiert man

$$\binom{\alpha}{0} = 1, \quad \text{insbesondere } \binom{0}{0} = 1. \tag{38.2}$$

Diese Größen treten bekanntlich als Koeffizienten in der **binomischen Reihenentwicklung**

$$(a+b)^\alpha = \sum_k \binom{\alpha}{k} a^{\alpha-k} b^k \qquad (|b| < |a|) \tag{38.3}$$

auf. Hierbei ist über alle nichtnegativen ganzzahligen k zu summieren, für die $\binom{\alpha}{k}$ nicht null ist.

Aus (38.1) folgt, wie man sofort sieht bzw. durch Nachrechnung bestätigt,

$$\binom{n}{k} = \binom{n}{n-k} \qquad (n \geqq 0,\ 0 \leqq k \leqq n,\ \text{beide ganzzahlig}), \tag{38.4}$$

$$\binom{\alpha}{k} + \binom{\alpha}{k+1} = \binom{\alpha+1}{k+1} \qquad (k \geqq 0,\ \text{ganzzahlig}), \tag{38.5}$$

$$\binom{r}{k} = 0 \qquad (k > r \geqq 0,\ \text{beide ganzzahlig}). \tag{38.6}$$

Mittels (38.5) kann man die Binomialkoeffizienten rekursiv berechnen. Für nichtnegative ganzzahlige α ordnet man die Koeffizienten dabei zweckmäßigerweise in der Gestalt des bekannten **Pascalschen Dreiecks** an:

$$\begin{array}{ccccccccc}
 & & & & 1 & & & & \\
 & & & 1 & & 1 & & & \\
 & & 1 & & 2 & & 1 & & \\
 & 1 & & 3 & & 3 & & 1 & \\
1 & & 4 & & 6 & & 4 & & 1 \\
 & & & & \cdots\cdots & & & &
\end{array}$$

Zahlenwerte der Binomialkoeffizienten findet man in Anhang 5, Tafel 1c.

Für beliebiges positives m folgt aus (38.1)

$$\binom{-m}{k} = \frac{-m(-m-1)(-m-2)\cdots(-m-k+1)}{k!}$$

$$= (-1)^k \frac{m(m+1)\cdots(m+k-1)}{k!}.$$

Dies ist gleichbedeutend mit

$$\binom{-m}{k} = (-1)^k \binom{m+k-1}{k} \qquad (m > 0). \tag{38.7}$$

Zwischen den Binomialkoeffizienten gibt es eine überraschend große Anzahl von Beziehungen (vgl. z. B. E. Netto, Lehrbuch der

Kombinatorik, Berlin: Teubner, 1927). Wir erwähnen (vgl. Aufgabe 38.1, 38.3 und 38.4)

$$\sum_{k=0}^{r} \binom{p}{k} \binom{q}{r-k} = \binom{p+q}{r}, \tag{38.8}$$

$$\sum_{k=0}^{p} \binom{p}{k}^2 = \binom{p}{0}^2 + \binom{p}{1}^2 + \cdots + \binom{p}{p}^2 = \binom{2p}{p}, \tag{38.9}$$

$$\sum_{s=0}^{n-1} \binom{k+s}{k} = \binom{n+k}{k+1} \quad (k \geqq 0,\ n \geqq 1, \text{ beide ganzzahlig}). \tag{38.10}$$

Aufgaben zu Abschnitt 38

38.1 Man gewinne (38.8) durch Anwendung von (38.3) auf

$$(1+x)^p (1+x)^q = (1+x)^{p+q}.$$

38.2 Man berechne $\binom{10}{3}, \binom{-2}{3}, \binom{0{,}5}{4}, \binom{3}{4}$.

38.3 Man beweise (38.10).

38.4 Man gewinne (38.9) aus (38.8).

Kapitel 8

Spezielle diskrete Verteilungen

Wir betrachten nun nacheinander drei diskrete Verteilungen, die für die Statistik besonders wichtig sind:

Die Binomialverteilung (Abschn. 39—41) erhält man, wenn man sich für die Anzahl des Eintreffens eines bestimmten Ereignisses A bei der wiederholten Ausführung eines Zufallsexperimentes interessiert, bei dem A bei jeder Ausführung dieselbe Wahrscheinlichkeit p besitzt und sich die Ergebnisse der verschiedenen Ausführungen gegenseitig nicht beeinflussen.

Wird das genannte Experiment sehr oft ausgeführt und ist p klein, so läßt sich die Binomialverteilung durch die in vieler Hinsicht bequemere Poisson-Verteilung (Abschn. 42—44) annähern.

Die Binomialverteilung ist auch beim Ziehen mit Zurücklegen von Bedeutung. Die entsprechende Rolle spielt beim Ziehen ohne Zurücklegen die etwas kompliziertere hypergeometrische Verteilung (Abschnitt 45, 46). Das Ziehen ohne Zurücklegen ist übrigens für die Praxis (z. B. bei der Endkontrolle und Abnahmeprüfung) wichtiger als das Ziehen mit Zurücklegen. In manchen Fällen besteht jedoch zwischen den entsprechenden Wahrscheinlichkeiten kein großer Unterschied, wie wir noch im einzelnen sehen werden.

39 Bernoulli- oder Binomialverteilung

Hat ein Ereignis A bei einem Zufallsexperiment die Wahrscheinlichkeit

$$p,$$

so ist die Wahrscheinlichkeit, daß A nicht eintrifft, gleich

$$q = 1 - p.$$

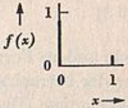


Abbildung 39.1. Wahrscheinlichkeitsfunktion (39.2) mit $p = 1/6$

Wir betrachten nun die Zufallsvariable

$$X = \textit{Anzahl des Eintreffens von } A.$$

Wird das Experiment nur einmal ausgeführt, so kann X natürlich nur die Werte 0 oder 1 annehmen je nachdem, ob A nicht eintrifft oder eintrifft. Die zugehörigen Wahrscheinlichkeiten sind

$$P(X = 0) = q \quad \text{und} \quad P(X = 1) = p.$$

Also hat die zu X gehörige Wahrscheinlichkeitsfunktion $f(x)$ die Werte

$$f(0) = q, \qquad f(1) = p \tag{39.1}$$

und $f(x) = 0$ für alle übrigen x. Wie man leicht nachprüft, lassen sich die beiden Formeln (39.1) zu einer einzigen Formel zusammenfassen:

$$f(x) = p^x q^{1-x} \quad 1^1 \cdot 1^{1-1} = 1 \qquad (x = 0, 1). \tag{39.2}$$

Ist A z. B. das Ereignis „Würfeln einer Sechs mit einem regelmäßigen Würfel", so ist $p = 1/6$, und man erhält die Funktion in Abb. 39.1.

Wird das obige Zufallsexperiment mehrmals ausgeführt, etwa n-mal hintereinander, so kann X die Werte $0, 1, \ldots, n$ annehmen. Wir fragen nach den zugehörigen Wahrscheinlichkeiten unter der *Voraussetzung, daß die Ergebnisse der einzelnen Ausführungen voneinander unabhängig sind, d. h. sich gegenseitig nicht beeinflussen.*

Das Ereignis

$X = x$ („*A trifft bei den n Ausführungen genau x-mal ein*")

ist z. B. realisiert, wenn wir (in der Reihenfolge des Auftretens) die Einzelereignisse

$$\underbrace{AA\cdots A}_{x\text{ mal}}\ \underbrace{BB\cdots B}_{n-x\text{ mal}} \qquad (B = \bar{A}, \text{d. h. } A \textit{ trifft nicht ein})$$

erhalten. Diese Einzelereignisse sind unabhängig. Also hat unsere Form der Realisierung des Ereignisses $X = x$ die Wahrscheinlichkeit

$$\underbrace{pp\cdots p}_{x\text{ mal}}\ \underbrace{qq\cdots q}_{n-x\text{ mal}} = p^x q^{n-x}. \tag{39.3}$$

Jede andere Form der Realisierung des Ereignisses $X = x$ ergibt sich aus der betrachteten Form durch Änderung der Reihenfolge der Einzelereignisse. Die zugehörige Wahrscheinlichkeit ist offenbar ebenfalls durch (39.3) gegeben. Da sich diese Formen gegenseitig ausschließen, folgt aus Satz 16.1, daß die Wahrscheinlichkeit des Ereignisses $X = x$ gleich der Summe der Wahrscheinlichkeiten aller möglichen Formen ist, also gleich $p^x q^{n-x}$ mal der Anzahl dieser Formen. Diese Anzahl ergibt sich so: Wir numerieren die n Ausführungen von 1 bis n durch.

x dieser n Zahlen greifen wir heraus. Das sind die Nummern der Ausführungen, bei denen A eintrifft. Auf die Reihenfolge der herausgegriffenen Zahlen kommt es nicht an. Also haben wir es mit Kombinationen x-ter Ordnung von n verschiedenen Elementen (den n Nummern der Ausführungen) ohne Berücksichtigung der Anordnung zu tun. Gemäß Satz 36.2 ist deren Anzahl gleich

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

Damit wird die Wahrscheinlichkeit $P(X = x)$ des Ereignisses $X = x$ gleich

(39.4) $$\boxed{f(x) = \binom{n}{x} p^x q^{n-x}} \qquad (x = 0, 1, \ldots, n).$$

Dies ist also die Wahrscheinlichkeit, daß ein Ereignis A bei n unabhängigen Ausführungen eines Experimentes genau x-mal eintrifft, wenn A bei der Einzelausführung die Wahrscheinlichkeit p besitzt und $q = 1 - p$ ist.

Die durch die Wahrscheinlichkeitsfunktion (39.4) (mit $f = 0$ für alle übrigen x) bestimmte Verteilung heißt die **Bernoulli-** oder **Binomialverteilung.** Sie ist wohl die wichtigste diskrete Verteilung. Das Eintreffen des genannten Ereignisses A wird gewöhnlich als **Erfolg** bezeichnet und das Nichteintreffen als **Mißerfolg.** Die Zahl p heißt die *Erfolgswahrscheinlichkeit beim Einzelversuch.* Ein solches Experiment, bei dem nur zwei verschiedene sich gegenseitig ausschließende Ereignisse eintreffen können, heißt ein **Bernoulli-Experiment,** nach Jakob Bernoulli, der sich zuerst mit solchen Experimenten befaßt hat.

Aus der Definition der Binomialkoeffizienten folgt unmittelbar

(39.5) $$\binom{n}{x+1} = \frac{n-x}{x+1}\binom{n}{x}.$$

Damit erhalten wir die nützliche Rekursionsformel

(39.6) $$f(x+1) = \left(\frac{n-x}{x+1}\right)\frac{p}{q} f(x) \qquad (x = 0, 1, \ldots, n-1).$$

Aus (39.4) ergibt sich die Verteilungsfunktion $F(x) = 0$ für $x < 0$ und

(39.7) $$F(x) = \sum_{k \leq x} \binom{n}{k} p^k q^{n-k} \qquad \text{für } x \geqq 0,$$

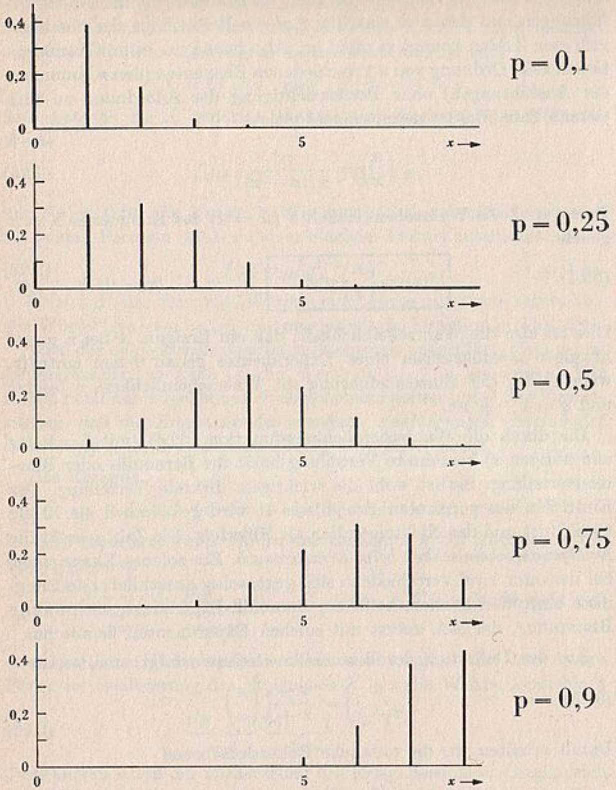


Abbildung 39.2. Wahrscheinlichkeitsfunktion (39.4) der Binomialverteilung für $n = 8$ und verschiedene Werte von p

wobei also über alle nichtnegativen ganzzahligen k zu summieren ist, die höchstens gleich x sind.

Zahlentafeln findet man in Anhang 5 (Tafel 1d) und in [19] (siehe Anhang 3).

Aufgaben zu Abschnitt 39

39.1 Man gewinne (39.5) aus (38.1).

39.2 Man zeichne $f(x)$ für $n = 8$ und $p = 0{,}2,\ 0{,}3,\ 0{,}4,\ 0{,}6,\ 0{,}7,\ 0{,}8$ (vgl. Tafel 1d in Anhang 5).

39.3 Vermöge (39.6) verschaffe man sich einen qualitativen Überblick über das Verhalten der Funktionswerte von $f(x)$ bei festem n und p. Wo hat $f(x)$ ein Maximum?

39.4 Man berechne und zeichne $f(x)$ und $F(x)$ für $n = 4$ und $p = 0{,}15$.

39.5 Man zeige die Richtigkeit der folgenden Formeln für die Wahrscheinlichkeiten der jeweils genannten Anzahl der Erfolge bei n Ausführungen eines Bernoulli-Experimentes mit der Erfolgswahrscheinlichkeit p beim Einzelversuch und $q = 1 - p$.

Höchstens m Erfolge:

$$P(X \leqq m) = \sum_{k=0}^{m} \binom{n}{k} p^k q^{n-k} = F(m).$$

Mindestens l Erfolge:

(39.8) $$P(X \geqq l) = \sum_{k=l}^{n} \binom{n}{k} p^k q^{n-k} = 1 - F(l-1).$$

Mindestens 1 Erfolg:

(39.9) $$P(X \geqq 1) = 1 - F(0) = 1 - q^n.$$

Mindestens l und höchstens m Erfolge:

(39.10) $$P(l \leqq X \leqq m) = \sum_{k=l}^{m} \binom{n}{k} p^k q^{n-k} = F(m) - F(l-1).$$

40 Mittelwert und Varianz der Binomialverteilung

Formeln für den Mittelwert und die Varianz der Binomialverteilung erhalten wir am einfachsten unter Benutzung der momenterzeugenden Funktion $G(t)$. Gemäß (33.1) und (39.4) ist

(40.1) $$G(t) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x q^{n-x}.$$

Dies können wir auch in der folgenden Form schreiben:

$$G(t) = \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x q^{n-x}.$$

Auf Grund der binomischen Formel (38.3) gilt daher

(40.2) $$G(t) = (pe^t + q)^n.$$

Differentiation ergibt

$$G'(t) = n(pe^t + q)^{n-1}\, pe^t, \quad \text{also} \quad G'(0) = np,$$

denn es ist $p + q = 1$. Gemäß (33.3) hat die Binomialverteilung demnach den Mittelwert

(40.3) $$\mu = np.$$

Ähnlich erhält man für die Varianz den Ausdruck

$$\sigma^2 = npq. \tag{40.4}$$

Denn durch Differentiation von $G'(t)$ ergibt sich zunächst

$$G''(t) = n(n-1)(pe^t + q)^{n-2}(pe^t)^2 + n(pe^t + q)^{n-1}pe^t.$$

Unter Benutzung von (33.2) wird also

$$E(X^2) = G''(0) = n(n-1)p^2 + np.$$

Setzen wir dies in (31.5) ein, so folgt

$$\sigma^2 = E(X^2) - \mu^2 = n(n-1)p^2 + np - n^2p^2 = npq.$$

41 Einige Anwendungen der Binomialverteilung

Die Binomialverteilung spielt beim **Ziehen mit Zurücklegen** (vgl. Abschn. 19) eine Rolle. Gegeben seien N Dinge, z. B. Schrauben, darunter M unbrauchbare („*Ausschuß*"). Die Wahrscheinlichkeit, beim zufälligen Herausgreifen einer Schraube eine unbrauchbare zu erhalten, ist dann gleich

$$p = \frac{M}{N}.$$

Also ist die Wahrscheinlichkeit, bei n Zügen mit Zurücklegen genau x unbrauchbare Schrauben zu erhalten, gemäß (39.4) gleich

$$f(x) = \binom{n}{x}\left(\frac{M}{N}\right)^x\left(1 - \frac{M}{N}\right)^{n-x}. \tag{41.1}$$

Praktisch wichtiger ist das Ziehen ohne Zurücklegen. Das führt aber auf eine kompliziertere Verteilung, und in Fällen, in denen sich diese Verteilung nicht sehr von der Binomialverteilung unterscheidet, benutzt man daher lieber die letztere. Einzelheiten folgen in Abschn. 45 und 46.

Beispiel 41.1. Aus einem Skatspiel werden nacheinander 6 Karten mit Zurücklegen gezogen. Wie groß ist die Wahrscheinlichkeit P, dabei mindestens 3 Asse zu ziehen?

Unter den $N = 32$ Karten sind $M = 4$ Asse. Also ist $p = 1/8$, $q = 7/8$ und

$$P = \sum_{k=3}^{6}\binom{6}{k}\left(\frac{1}{8}\right)^k\left(\frac{7}{8}\right)^{6-k} = \frac{7638}{262144} \approx 3\%.$$

Viele Spiele, bei denen mehr als 2 Ereignisse eintreffen können, lassen sich als BERNOULLI-Experimente behandeln. Hierzu faßt man alle Ereignisse, die für einen bestimmten Spieler zum Gewinn des betreffenden Spiels führen, zu einem Ereignis A („*Erfolg*") und die übrigen, die zum Verlust führen, zu einem Ereignis B („*Mißerfolg*") zusammen.

Beispiel 41.2. Mit welcher Wahrscheinlichkeit P erzielt man bei 3 Würfen eines regelmäßigen Würfels mindestens eine Fünf oder Sechs?

Beim einzelnen Wurf hat das Ereignis

A: *Fünf oder Sechs*

die Wahrscheinlichkeit $p = 2/6 = 1/3$. Also ist $q = 2/3$, und aus (39.9) folgt

$$P = 1 - \left(\frac{2}{3}\right)^3 = \frac{19}{27} \approx 70\%.$$

Beispiel 41.3. Wie groß ist die Wahrscheinlichkeit P, mit 10 Schüssen aus einem eingespannten Gewehr ein gegebenes Ziel wenigstens einmal zu treffen, wenn die Treffwahrscheinlichkeit je Schuß 1/10 beträgt?

Mit $n = 10$, $p = 0{,}1$, also $q = 0{,}9$, erhalten wir aus (39.9) die Antwort

$$P = 1 - (0{,}9)^{10} \approx 0{,}65.$$

Einige weitere Anwendungen illustrieren die folgenden Aufgaben.

Aufgaben zu Abschnitt 41

41.1 Wie groß ist die Wahrscheinlichkeit, daß in einer Familie mit 4 Kindern (*a*) 2 Jungen und 2 Mädchen, (*b*) 3 Jungen und 1 Mädchen, (*c*) lauter Jungen sind, wenn man annimmt, daß Jungen- und Mädchengeburten gleichwahrscheinlich sind und Unabhängigkeit besteht?

41.2 Man zeige, daß (39.4) für $p = 0{,}5$ die folgende einfache Form gewinnt:

$$f(x) = \frac{1}{2^n}\binom{n}{x}.$$

41.3 Wie groß ist die Wahrscheinlichkeit, daß wenigstens 14 von 15 Meßergebnissen eine geradzahlige letzte Ziffer besitzen (wie ich dies unlängst in einer Fachzeitschrift sah), wenn gerade und ungerade letzte Ziffern gleichwahrscheinlich sind?

41.4 Sechs unter den acht am 1. 10. 1961 in der Universitäts-Frauenklinik Graz geborenen Kindern waren Jungen. Wie groß ist die Wahrscheinlichkeit eines solchen Ereignisses, wenn Jungen- und Mädchengeburten gleichwahrscheinlich sind?

41.5 Bei einem Experiment sei die Wahrscheinlichkeit, einen positiven bzw. einen negativen Meßwert zu erhalten, jeweils gleich 1/2. Wie groß ist dann die Wahrscheinlichkeit, bei 9 Ausführungen des Experimentes höchstens einen einzigen negativen Wert zu erhalten?

41.6 Welche Anzahl von „Köpfen" hat bei 2, 4, 10 bzw. 100 nacheinander ausgeführten Münzenwürfen die größte Wahrscheinlichkeit?

41.7 Man ziehe nacheinander 4 Kugeln mit Zurücklegen aus einer Schachtel, die Kugeln enthält, von denen der dritte Teil rot sind, und notiere die Anzahl gezogener roter Kugeln. Man führe diesen Versuch 81mal aus und vergleiche die erhaltene Verteilung mit der theoretischen.

41.8 Man benutze in Aufgabe 41.7 statt der Kugeln die Ziffern der Tafel 5 in Anhang 5, indem man zum Beispiel 3, 6, 9 an die Stelle der roten Kugeln treten läßt, 1, 2, 4, 5, 7, 8 an die Stelle der anderen und 0 wegläßt.

41.9 Durch Versuche sei festgestellt worden, daß 5% der Zwiebeln einer großen Menge einer bestimmten Blumenzwiebelsorte nicht keimen. Diese Zwiebelsorte wird in Zehnerpackungen auf den Markt gebracht, und es wird eine Keimgarantie von 90% gegeben. Wie groß ist die Wahrscheinlichkeit, daß eine bestimmte Packung dieses Garantieversprechen nicht erfüllt?

41.10 Wie ändert sich die Lage in Aufgabe 41.9, wenn eine Keimgarantie von nur 80% gegeben wird?

41.11 Fünf Arbeiter, die unabhängig arbeiten, benötigen elektrischen Strom, und zwar jeder mit Unterbrechungen durchschnittlich etwa 10 Minuten je Stunde. Genügt es, die Stromversorgung so einzurichten, daß 3 Arbeiter gleichzeitig Strom entnehmen können, oder entstehen dann erhebliche Wartezeiten, indem 4 oder 5 Arbeiter gleichzeitig Strom entnehmen wollen?

41.12 Wie viele Schüsse sind notwendig, um mit einer Wahrscheinlichkeit von 0,9 wenigstens einen Treffer zu erzielen, wenn die Treffwahrscheinlichkeit je Schuß gleich 0,1 ist?

41.13 Bei einem Produktionsprozeß trete auf lange Dauer 5% Ausschuß auf, und es bestehe Unabhängigkeit zwischen der Herstellung der einzelnen Stücke. Man bestimme und zeichne die Wahrscheinlichkeit, unter 10 Stücken 0, 1, 2, ..., 10 Ausschußstücke zu erhalten.

41.14 Bei einem Fabrikationsprozeß elektrischer Sicherungen will man erreichen, daß höchstens 1% der Produktion Ausschuß ist. Hierzu kontrolliert man den Prozeß stündlich, indem man aus der Produktion der letzten Stunde 10 Sicherungen zufällig auswählt und prüft. Sind nicht alle diese 10 Sicherungen brauchbar, so stoppt und überprüft man den Prozeß. Man berechne die Wahrscheinlichkeit, daß der Prozeß nicht gestoppt wird, wenn tatsächlich einmal 2% Ausschuß produziert wird. Auf Grund des Ergebnisses fälle man ein Urteil über die Zweckmäßigkeit des Verfahrens.

41.15 Die sogenannte **Pascal-Verteilung** hat die Wahrscheinlichkeitsfunktion

$$f(x) = \binom{k+x-1}{x} p^k q^x, \qquad q = 1-p, \quad x = 0, 1, \ldots \tag{41.2}$$

Man überlege sich, daß dies die Wahrscheinlichkeit des Eintreffens des k-ten Erfolges bei der $(k+x)$ten Ausführung eines BERNOULLI-Experimentes ist. k ist fest vorgegeben. p ist die Erfolgswahrscheinlichkeit beim Einzelversuch. (Dies ist ein einfaches „Wartezeitproblem".)

41.16 Man überlege sich, daß die geometrische Verteilung ein Sonderfall der PASCAL-Verteilung ist.

42 Poisson-Verteilung

Bei vielen Anwendungen, die mit Bernoulli-Experimenten zusammenhängen, ist die Erfolgswahrscheinlichkeit p beim einzelnen Experiment klein, der Erfolg also ein seltenes Ereignis, während die Anzahl n der Ausführungen sehr groß ist. In einem solchen Falle bringt es Vorteile, die Binomialverteilung mit ihren für großes n recht unbequemen Binomialkoeffizienten durch die Verteilung zu approximieren, die sich ergibt, wenn p gegen null und n gegen unendlich strebt, und zwar derart, daß der Mittelwert

$$\mu = np \tag{42.1}$$

gegen einen endlichen Wert strebt.

Um die genannte Verteilung zu gewinnen, gehen wir von der Wahrscheinlichkeitsfunktion (39.4) der Binomialverteilung aus. Wegen (42.1) ist

$$p = \frac{\mu}{n}, \qquad \text{also} \qquad p^x = \frac{\mu^x}{n^x},$$

und weiterhin

$$q^{n-x} = (1-p)^{n-x} = \left(1 - \frac{\mu}{n}\right)^{n-x} = \left[1 - \frac{\mu}{n}\right]^n \left\{1 - \frac{\mu}{n}\right\}^{-x}.$$

Setzen wir dies in (39.4) ein, so gewinnt (39.4) sofort die Form

$$\frac{\mu^x}{x!} \frac{n(n-1)\cdots(n-x+1)}{n^x} \left[1 - \frac{\mu}{n}\right]^n \left\{1 - \frac{\mu}{n}\right\}^{-x}$$

Für $n \to \infty$ strebt

$$\frac{n(n-1)\cdots(n-x+1)}{n^x} = \left(\frac{n}{n}\right)\left(\frac{n-1}{n}\right)\cdots\left(\frac{n-x+1}{n}\right) \to 1,$$

der Ausdruck in der eckigen Klammer gegen $e^{-\mu}$ und der Ausdruck in der geschweiften Klammer gegen 1. Also strebt (39.4) dann gegen die Funktion

$$\boxed{f(x) = \frac{\mu^x}{x!} e^{-\mu}} \qquad (x = 0, 1, \ldots). \tag{42.2}$$

Die durch diese Wahrscheinlichkeitsfunktion gegebene diskrete Verteilung heißt die **Poisson-Verteilung.** Sie wurde von S. D. Poisson im Jahre 1837 eingeführt. Zahlenwerte enthält die Tafel 2 in Anhang 5.

Die Verteilung hat den **Mittelwert** μ, wie aus der Herleitung folgt und im übernächsten Abschnitt auch noch direkt bestätigt wird.

Tab. 42.1 illustriert, daß die Binomialverteilung durch die POISSON-Verteilung auch bei relativ kleinem n schon recht gut angenähert wird.

Zum ersten Male bei unseren Überlegungen begegnen wir hier dem wichtigen Sachverhalt, daß sich gewisse Verteilungen in speziellen Fällen (Grenzfällen) durch andere, unter Umständen einfachere Verteilungen approximieren lassen.

Tabelle 42.1. Zur Approximation der Binomialverteilung durch die POISSON-Verteilung

x	Binomialverteilung $n=4,\ p=\frac{1}{4}$	Binomialverteilung $n=8,\ p=\frac{1}{8}$	Binomialverteilung $n=100,\ p=\frac{1}{100}$	POISSON-Verteilung $\mu=1$
0	0,316	0,344	0,366	0,368
1	0,422	0,393	0,370	0,368
2	0,211	0,196	0,185	0,184
3	0,047	0,056	0,061	0,061
4	0,004	0,010	0,015	0,015
5	—	0,001	0,003	0,003

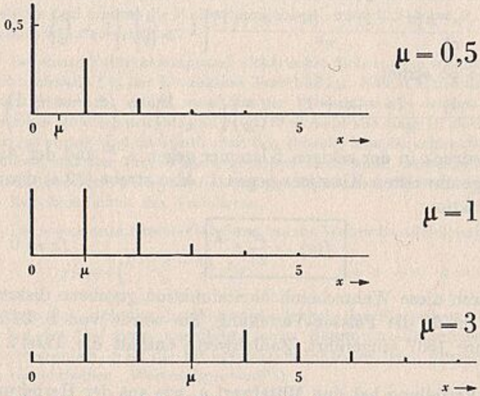


Abbildung 42.1. Wahrscheinlichkeitsfunktion (42.2) der POISSON-Verteilung für verschiedene Mittelwerte

Aus (42.2) erhalten wir die Verteilungsfunktion

$$F(x) = e^{-\mu} \sum_{s \leq x} \frac{\mu^s}{s!} \qquad \text{für } x \geqq 0 \tag{42.3}$$

und $F(x) = 0$ und für $x < 0$.

43 Anwendungen der Poisson-Verteilung

Die Poisson-Verteilung ist ein im Zusammenhang mit verschiedenartigen praktischen Problemen vielseitig verwendbares Modell. Drei typische Anwendungen zeigen die Tab. 43.1–43.3. In Tab. 43.1 ist

$$\bar{x} = \frac{1}{200}(1 \cdot 65 + 2 \cdot 22 + 3 \cdot 3 + 4 \cdot 1) = 0{,}61.$$

Nimmt man dies als μ in (42.2) und multipliziert mit der Gesamtzahl von Jahren (200), so erhält man die in der letzten Spalte angegebenen Werte. Entsprechend ergibt sich die letzte Spalte in Tab. 43.2 und 43.3.

Tabelle 43.1. Tote durch Hufschlag in 10 preußischen Kavallerieregimentern während 20 Jahren (L. v. Bortkiewicz, Das Gesetz der kleinen Zahlen. Leipzig, 1898)

x	Anzahl von Jahren mit x Toten pro Regiment pro Jahr	
	Beobachtet	Theoretisch (abgerundet)
0	109	109
1	65	66
2	22	20
3	3	4
4	1	1
$\geqq 5$	0	0

Ein weiteres Beispiel ist die Verteilung von Bomben in einem Zielgebiet, das man quadratisch unterteilt hat, wobei die Quadrate die entsprechende Rolle spielen wie die Zeitintervalle in den Tabellen. Die Verteilung von Druckfehlern pro Seite in Büchern, Unkrautsamen unter dem Weizen, Rosinen in einem Kuchen und Fadenbrüche pro Zeitspanne in einer Spinnerei ließen sich als weitere Beispiele anführen.

Es ist also $G'(0) = \mu$ und wegen (33.2) weiterhin

$$E(X^2) = G''(0) = \mu + \mu^2.$$

Demnach hat die Poisson-Verteilung wegen (31.5) die Varianz

$$\sigma^2 = \mu. \tag{44.3}$$

Aus (44.2c) folgt mit (33.2) und (31.6) weiterhin

$$E([X - \mu]^3) = \mu, \tag{44.4}$$

wie man leicht nachrechnet. Die Poisson-Verteilung hat also gemäß (32.1) die positive Schiefe

$$\gamma = \frac{1}{\sqrt{\mu}}. \tag{44.5}$$

Mit wachsendem μ strebt γ gegen null. Es gilt:

Für große Werte von μ ist die Poisson-*Verteilung nahezu symmetrisch.* (Vgl. Abb. 42.1.)

Aufgaben zu Abschnitt 44

44.1 Man leite (44.1) her.

44.2 Aus (44.2a) sieht man, daß der Parameter μ tatsächlich der Mittelwert der Poisson-Verteilung ist. Warum?

44.3 Aus (44.1) gewinne man mittels (34.5) die momenterzeugende Funktion $G^*(t)$ der Variablen $X - \mu$ und hieraus (44.3) vermöge (34.6).

44.4 Man bestätige (44.4) durch Nachrechnung.

45 Hypergeometrische Verteilung

Eine Schachtel enthalte N Dinge, z. B. Schrauben, darunter M unbrauchbare („Ausschuß"). n Schrauben werden nacheinander zufällig gezogen. Wie groß ist die Wahrscheinlichkeit, dabei genau x unbrauchbare Schrauben zu erhalten?

Ziehen wir mit Zurücklegen, so ist die Antwort durch (41.1) gegeben. Ziehen wir ohne Zurücklegen, so lautet die Antwort

$$\boxed{f(x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}} \qquad (x = 0, 1, \ldots, n). \tag{45.1}$$

Dies folgt aus Satz 36.2 (vgl. Aufgabe 45.1). Die Verteilung mit der Wahrscheinlichkeitsfunktion (45.1) heißt die **hypergeometrische Verteilung.** Der Name rührt daher, daß sich die zugehörige momenterzeugende Funktion durch die hypergeometrische Funktion ausdrükken läßt. Wir gehen hierauf nicht näher ein. Diese Verteilung spielt also beim Ziehen ohne Zurücklegen dieselbe Rolle wie die Binomialverteilung beim Ziehen mit Zurücklegen. Übrigens kann $f(x)$ für einige der angegebenen x auch null sein, wie man aus (38.6) sieht.

Die hypergeometrische Verteilung hat den Mittelwert

$$\mu = n\frac{M}{N} \tag{45.2}$$

(vgl. Aufgabe 45.3) und die Varianz (vgl. Aufgabe 45.4)

$$\sigma^2 = \frac{nM(N-M)(N-n)}{N^2(N-1)}. \tag{45.3}$$

Aufgaben zu Abschnitt 45

45.1 Man beweise die Behauptung, die (45.1) enthält.

45.2 Man berechne und zeichne $f(x)$ für $N = 12$, $n = 8$ und verschiedene M, etwa $M = 3, 4, 6$ (vgl. Tafel 1c in Anhang 5).

45.3 Man gewinne (45.2) aus (28.1a) unter Benutzung von (38.8).

45.4 Man beweise (45.3). Anleitung: Vermöge (38.8) zeige man, daß

$$E(X[X-1]) = n(n-1)\,M(M-1)/N(N-1)$$

ist. Dann benütze man (31.7) in Aufgabe 31.5.

45.5 Aus einem Skatspiel werden nacheinander 6 Karten ohne Zurücklegen gezogen. Wie groß ist die Wahrscheinlichkeit P, daß mindestens 3 der gezogenen Karten Asse sind?

45.6 Aus einer Urne, die 9 blaue und 3 schwarze Kugeln enthält, werden 8 Kugeln ohne Zurücklegen gezogen. Wie groß ist die Wahrscheinlichkeit, daß genau x der gezogenen Kugeln blau sind?

45.7 Eine Firma liefert Dichtungen in Packungen zu 100 Stück. Eine Packung darf laut Liefervertrag höchstens 10% Ausschuß enthalten. Jede Packung wird geprüft, indem man 10 Stück zufällig und ohne Zurücklegen entnimmt. Sind diese 10 Stück alle einwandfrei, wird die Packung angenommen. Anderenfalls wird sie zurückgewiesen. Wie groß ist bei diesem Prüfverfahren die Wahrscheinlichkeit ungerechtfertigter Reklamationen, indem eine Packung zurückgewiesen wird, obwohl sie den Lieferbedingungen entspricht?

45.8 Wie groß ist in Aufgabe 45.7 die Wahrscheinlichkeit, daß eine Packung angenommen wird, obwohl sie 20 defekte Dichtungen enthält?

45.9 Die Prüfung in Aufgabe 45.7 (**Endkontrolle** des Herstellers oder **Abnahmekontrolle** des Kunden) verläuft offenbar nach dem folgenden Schema: Aus einer Partie von N Stück wird (durch Ziehen ohne Zurücklegen) eine Stichprobe von n Stück entnommen. Enthält diese nicht mehr als eine zugelassene Zahl c fehlerhafter Stücke, so wird die Partie angenommen. Die Wahrscheinlichkeit P, daß eine Partie angenommen wird, heißt die *Annahmewahrscheinlichkeit*. Diese ist natürlich eine Funktion des relativen Anteils M/N der fehlerhaften Stücke in der gesamten Partie (M = Anzahl fehlerhafter Stücke in der

Tabelle 43.2. Zählung von Alphateilchen (E. RUTHERFORD u. H. GEIGER, Phil. Mag. (6) 20, 1910, 698)

x	Anzahl von Zeitintervallen (Länge 7,5 sec) mit x Teilchen pro Zeitintervall	
	Beobachtet	Theoretisch (abgerundet)
0	57	54
1	203	210
2	383	407
3	525	526
4	532	509
5	408	394
6	273	254
7	139	140
8	45	68
9	27	29
10	10	11
11	4	4
12	2	1
$\geqq 13$	0	1

Tabelle 43.3. Verkehrszählung in Graz, Münzgrabenstraße zwischen Brockmanngasse und Stremayrgasse, am 28. 9. 1963, 10.20—11.10 Uhr

x	Anzahl von Zeitintervallen (Länge 30 sec) mit x Personenkraftwagen pro Intervall	
	Beobachtet	Theoretisch (abgerundet)
0	6	6
1	18	17
2	21	24
3	26	22
4	16	15
5	8	9
6	2	4
7	1	2
8	2	1
$\geqq 9$	0	0

Aufgaben zu Abschnitt 43

43.1 Eine Ladung Saatgut wird in Pfundpäckchen verkauft. Jedes Päckchen enthält rund 1000 Samenkörner. Von früheren Prüfungen her sei bekannt, daß etwa 1% der Körner nicht der Sorte des Saatgutes angehören. Wie groß ist die Wahrscheinlichkeit, daß in einem bestimmten Päckchen mehr als 10 solche „fremde" Körner sind?

43.2 Seriengefertigte Rundfunkwiderstände von 50 Ohm Nennwert werden in Packungen zu je 100 Stück geliefert. Dabei wird die Garantie gegeben, daß alle Widerstände zwischen 45 und 55 Ohm liegen. Wie groß ist die Wahrscheinlichkeit, daß eine bestimmte Packung diese Zusage erfüllt, wenn die Wahrscheinlichkeit, einen Widerstand zu produzieren, der nicht zwischen 45 und 55 Ohm liegt, erfahrungsgemäß nur 2‰ beträgt?

43.3 Wie groß ist die Wahrscheinlichkeit, daß in einem Dorfe mit 500 Einwohnern wenigstens einer am 24. Dezember Geburtstag hat?

43.4 Man zeige, daß (42.3) die Beziehung $F(\infty) = 1$ erfüllt.

43.5 In einer Stadt, die 10000 erwachsene (d. h. mindestens 18 Jahre alte) Einwohner zählt, soll das Netz der Omnibuslinien geändert werden. Um sich über die öffentliche Meinung zu diesem Vorhaben zu informieren, werden 100 zufällig herausgegriffene erwachsene Einwohner der Stadt befragt. Wie groß ist die Wahrscheinlichkeit, daß sich unter den Befragten weniger als 5 Personen ablehnend verhalten, wenn 10% aller Erwachsenen der Stadt das Vorhaben ablehnen?

43.6 Es sei bekannt, daß 0,005% einer Bevölkerungsgruppe jährlich durch einen gewissen Unfall getötet wird. Bei einer Versicherung sind 10000 Personen aus der genannten Gruppe gegen diesen Unfall versichert. Wie groß ist die Wahrscheinlichkeit, daß in einem gegebenen Jahr mehr als drei dieser Versicherten durch den genannten Unfall umkommen?

43.7 Eine Fernsprechvermittlung erhalte während der Hauptbetriebszeit durchschnittlich 300 Anrufe stündlich. Sie kann maximal 10 Verbindungen je Minute herstellen. Man benütze die Poisson-Verteilung zur Abschätzung der Wahrscheinlichkeit, daß die Vermittlung während einer beliebigen gegebenen Minute in der Hauptbetriebszeit überlastet ist, d. h. mehr Anrufe erhält, als sie Verbindungen herstellen kann.

43.8 Für welches x hat (42.2) bei gegebenem μ den größten Wert?

44 Varianz und Schiefe der Poisson-Verteilung

Gemäß (33.1) und (42.2) hat die Poisson-Verteilung die momenterzeugende Funktion

$$G(t) = e^{-\mu}\, e^{\mu e^t} \tag{44.1}$$

(vgl. Aufgabe 44.1). Es ist $G(0) = 1$. Wiederholte Differentiation ergibt nacheinander

$$\begin{aligned} &\text{(a)}\quad G'(t) = e^{-\mu}\, e^{\mu e^t}\, \mu e^t = \mu e^t G(t),\\ &\text{(b)}\quad G''(t) = \mu e^t [G(t) + G'(t)],\\ &\text{(c)}\quad G'''(t) = \mu e^t [G(t) + 2G'(t) + G''(t)]. \end{aligned} \tag{44.2}$$

Partie). Diese Funktion kann man graphisch darstellen, indem man die Funktionswerte in der üblichen Weise als Punkte in ein Koordinatensystem einträgt. Die Kurve durch diese Punkte heißt die **Annahmekennlinie**. Man berechne und zeichne diese für den Fall $N = 100$, $n = 3$, $c = 0$.

45.10 Man bestimme die Annahmekennlinie für den Fall $N = 100$, $n = 4$, $c = 1$.

45.11 (Französische Lotterie). Bei der bis zum Ende des 18. Jahrhunderts in Frankreich üblichen staatlichen Lotterie wurden nacheinander 5 Lose aus einer Urne mit 90 von 1 bis 90 numerierten Losen ohne Zurücklegen gezogen. Der Teilnehmer konnte vor der Ziehung 1, 2, 3, 4 oder 5 Karten mit von ihm genannten Nummern (von 1 bis 90) kaufen. Stimmten die 1, 2, 3, 4 bzw. 5 Nummern der gekauften Karten mit den Nummern gezogener Lose überein, so erhielt er das 15-, 270- 5500-, 75000- bzw. 1000000fache des Preises der Karten. Wie groß ist die Wahrscheinlichkeit zu gewinnen in jedem der 5 genannten Fälle?

46 Vergleich der hypergeometrischen Verteilung und der Binomialverteilung

Ist im vorigen Abschnitt n klein gegenüber N, M und $N-M$, so kann man erwarten, daß zwischen den Wahrscheinlichkeiten beim Ziehen mit Zurücklegen und ohne Zurücklegen kein großer Unterschied besteht. In der Tat gilt für die Wahrscheinlichkeitsfunktion (45.1) der hypergeometrischen Verteilung

$$(46.1)\quad \binom{n}{x}\left(p-\frac{x}{N}\right)^x\left(q-\frac{n-x}{N}\right)^{n-x} < f(x) < \binom{n}{x}p^x q^{n-x}\left(1-\frac{n}{N}\right)^{-n}$$

(vgl. Aufgabe 46.3 und 46.4). Hierbei ist

$$(46.2)\qquad p = \frac{M}{N}$$

und $q = 1 - p$. Halten wir n und x fest und lassen N und M derart gegen unendlich streben, daß (46.2) einem Wert zwischen 0 und 1 zustrebt, so sehen wir, daß der erste und der letzte Ausdruck in (46.1) gegen die Wahrscheinlichkeitsfunktion (41.1) der Binomialverteilung streben. Dies bedeutet praktisch:

Für großes N, M und $N-M$ und im Vergleich dazu kleines n kann die hypergeometrische Verteilung durch die Binomialverteilung mit dem durch (46.2) *gegebenen p brauchbar angenähert werden.*

Tab. 46.1 gibt eine Vorstellung von der Güte der Näherung bei großem N. Für kleines N ist die Näherung verständlicherweise nicht befriedigend (vgl. Tab. 46.2).

Tabelle 46.1. Werte der Wahrscheinlichkeitsfunktionen der hypergeometrischen Verteilung mit $M = 20$, $N = 1000$, $n = 100$ [vgl. (45.1)] und der Binomialverteilung mit $p = 0{,}02$ und $n = 100$

x	Hypergeometrische Verteilung	Binomialverteilung
0	0,119	0,133
1	0,270	0,271
2	0,288	0,273
3	0,192	0,182
4	0,089	0,090
5	0,031	0,035
6	0,008	0,011
7	0,002	0,003
8	0,000	0,001

Tabelle 46.2. Werte der Wahrscheinlichkeitsfunktionen der hypergeometrischen Verteilung mit $M = 8$, $N = 16$, $n = 8$ [vgl. (45.1)] und der Binomialverteilung mit $p = 0{,}5$ und $n = 8$

x	Hypergeometrische Verteilung	Binomialverteilung
0	0,000	0,004
1	0,005	0,031
2	0,061	0,109
3	0,244	0,219
4	0,381	0,273
5	0,244	0,219
6	0,061	0,109
7	0,005	0,031
8	0,000	0,004

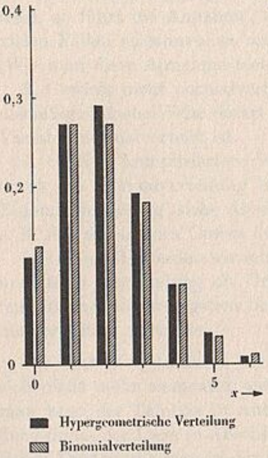


Abbildung 46.1. Werte in Tab. 46.1

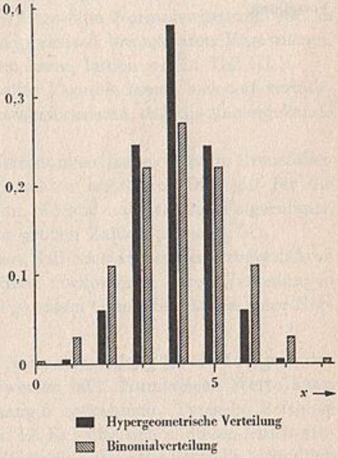


Abbildung 46.2. Werte in Tab. 46.2

Aus dieser Näherung und dem Zusammenhang zwischen der Binomial- und der Poisson-Verteilung ergibt sich weiterhin:

Für kleines p, großes n und im Vergleich mit n sehr großes N läßt sich die hypergeometrische Verteilung durch die Poisson-*Verteilung mit $\mu = np$ brauchbar annähern.*

Aufgaben zu Abschnitt 46

46.1 Man zeige, daß die Varianz der hypergeometrischen Verteilung für $n > 1$ kleiner als die der Binomialverteilung mit p gemäß (46.2) ist.

46.2 Man zeige, daß die Verteilungen in Aufgabe 46.1 denselben Mittelwert besitzen.

46.3 Man leite den zweiten Teil der Doppelungleichung (46.1) her.

46.4 Man gewinne den ersten Teil der Doppelungleichung (46.1).

46.5 Eine Urne enthält 4 blaue und 6 schwarze Kugeln. 3 Kugeln werden nacheinander gezogen. Man betrachte die Zufallsvariable

$$X = \textit{Anzahl der blauen Kugeln unter den gezogenen}$$

und berechne und vergleiche die zugehörigen Wahrscheinlichkeitsfunktionen, den Mittelwert und die Varianz beim Ziehen mit und ohne Zurücklegen.

46.6 Man berechne die Annahmekennlinie in Aufgabe 45.9 näherungsweise unter Benutzung der Binomialverteilung und vergleiche die Ergebnisse.

46.7 Man benutze in Aufgabe 46.6 statt der Binomialverteilung die Poisson-Verteilung.

Kapitel 9

Normalverteilung

Die GAUSS-Verteilung oder Normalverteilung wurde von GAUSS im Zusammenhang mit der Theorie der Meßfehler eingeführt, auf die wir später (in Kap. 19) eingehen werden. Sie ist die wichtigste stetige Verteilung. Hierfür gibt es verschiedene Gründe:

1. Viele Zufallsvariable, die bei Experimenten und Beobachtungen in der Praxis auftreten, sind normalverteilt.

2. Andere Zufallsvariable sind annähernd normalverteilt. Vermutet man, etwa auf Grund einer Stichprobe, daß eine bestimmte Grundgesamtheit eine eingipflige Verteilung besitzt, ohne daß man Näheres weiß, so führt die Annahme, es liege eine Normalverteilung vor, in vielen Fällen zu sinnvollen und praktisch brauchbaren Ergebnissen. (Wie man diese Annahme testen kann, lernen wir in Teil III.)

3. Gewisse nicht normalverteilte Variable lassen sich auf verhältnismäßig einfache Weise derart transformieren, daß die sich ergebende Variable normalverteilt ist.

4. Gewisse kompliziertere Verteilungen lassen sich in Grenzfällen durch die Normalverteilung brauchbar annähern. Dies gilt für die Binomialverteilung (siehe Abschn. 50) und hat wichtige Folgerungen, z. B. das sogenannte Gesetz der großen Zahlen (Abschn. 51).

5. In Teil III werden wir sehen, daß bei statistischen Prüfverfahren und deren Begründung oft Größen vorkommen, deren Verteilungen normal sind oder wenigstens bei gewissen Grenzübergängen einer Normalverteilung zustreben.

Die Verteilungsfunktion der Normalverteilung ist ein Integral, das sich nicht mehr elementar auswerten läßt. Numerische Werte kann man aber der Tafel 3a in Anhang 5 entnehmen. Typische Beispiele dazu findet der Leser in Abschn. 49. Er sollte sich in diesen handwerklichen Dingen, die man in der Statistik notwendig braucht, gründlich üben.

47 Gauß- oder Normalverteilung

Die stetige Verteilung mit der Wahrscheinlichkeitsdichte

$$\boxed{f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}} \qquad (-\infty < x < \infty,\ \sigma > 0) \tag{47.1}$$

heißt **Gauß-Verteilung** oder **Normalverteilung.** Eine Zufallsvariable mit dieser Verteilung heißt kurz eine **normalverteilte Zufallsvariable.**

Die Kurve von $f(x)$, die sogenannte **Glockenkurve,** ist symmetrisch bezüglich μ. Das Integral $\int_{-\infty}^{\infty} |x|\, f(x)\, dx$ existiert. Also ist der Parameter μ in (47.1) wegen Satz 28.1 tatsächlich der Mittelwert. Abb. 47.1 zeigt $f(x)$ für $\mu = 0$. Für $\mu > 0$ bzw. $\mu < 0$ erhält man Kurven derselben Form, die lediglich um den Betrag $|\mu|$ nach rechts bzw. links verschoben sind.

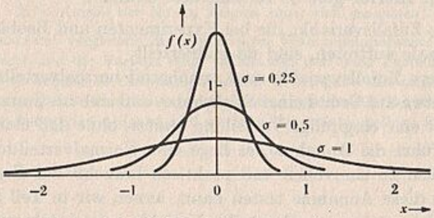


Abbildung 47.1. Wahrscheinlichkeitsdichte (47.1) mit $\mu = 0$ für verschiedene Werte σ

Daß der Parameter σ in (47.1) tatsächlich die Standardabweichung ist, zeigen wir später (in Aufgabe 48.3). Je kleiner σ^2 ist, desto größer ist das Maximum von $f(x)$ und desto rascher fällt die Kurve nach beiden Seiten hin ab, wie Abb. 47.1 veranschaulicht. Dies ist auf Grund der Bedeutung der Varianz verständlich.

48 Verteilungsfunktion der Normalverteilung

Wie wir aus (47.1) ersehen, hat die Normalverteilung die Verteilungsfunktion

$$\boxed{F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{v-\mu}{\sigma}\right)^2} dv.} \tag{48.1}$$

Besitzt eine Zufallsvariable X diese Verteilung, so ist die Wahrscheinlichkeit, daß X irgendeinen Wert aus einem Intervall $a < X \leqq b$

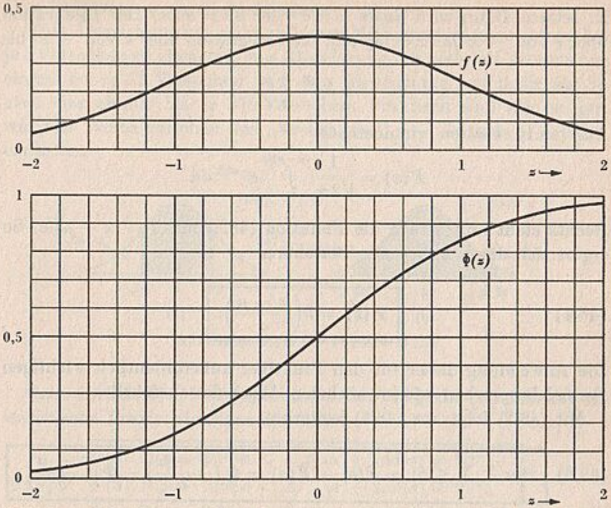


Abbildung 48.1. Dichte $f(z)$ und Verteilungsfunktion $\Phi(z)$ [vgl. (48.3)] der Normalverteilung mit dem Mittelwert 0 und der Varianz 1

annimmt, also gemäß (26.3) durch

$$(48.2) \qquad P(a < X \leqq b) = F(b) - F(a) = \frac{1}{\sigma\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}\left(\frac{v-\mu}{\sigma}\right)^2} dv$$

gegeben.

Das Integral in (48.1) läßt sich zwar nicht mehr elementar auswerten, kann aber durch das Integral (siehe Abb. 48.1)

$$(48.3) \qquad \boxed{\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-u^2/2} du,}$$

also durch die Verteilungsfunktion der Normalverteilung mit dem Mittelwert 0 und der Varianz 1, ausgedrückt werden, deren Werte vertafelt sind (s. Tafel 3a in Anhang 5). Hierzu brauchen wir in (48.1) nur

$$\frac{v-\mu}{\sigma} = u$$

zu setzen. Dann wird $du/dv = 1/\sigma$ oder $dv = \sigma\, du$. Der Integration über v von $-\infty$ bis x entspricht die Integration über u von $-\infty$ bis

$$z = \frac{x-\mu}{\sigma}.$$

Aus (48.1) erhalten wir demnach

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(x-\mu)/\sigma} e^{-u^2/2}\, du\,.$$

Rechts steht nun gerade die Funktion (48.3) mit $z = (x-\mu)/\sigma$. So ergibt sich die grundlegende Beziehung

(48.4) $$\boxed{F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right).}$$

Die Anwendung dieser für den Praktiker außerordentlich wichtigen Formel besprechen wir im nächsten Abschnitt ausführlich.

Mit (48.2) folgt aus (48.4) weiterhin

(48.5) $$\boxed{P(a < X \leqq b) = F(b) - F(a) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).}$$

Damit kann man Wahrscheinlichkeiten normalverteilter Zufallsvariablen unter Benutzung der Tafel 3a in Anhang 5 berechnen. Beispiele folgen noch.

Insbesondere hat die rechte Seite von (48.5) für $a = \mu - \sigma$ und $b = \mu + \sigma$ die Form $\Phi(1) - \Phi(-1)$, für $a = \mu - 2\sigma$ und $b = \mu + 2\sigma$ die Form $\Phi(2) - \Phi(-2)$, usw. Die Zahlenwerte stehen in Tafel 3a, und so erhalten wir insgesamt aus (48.5) die Formeln

(48.6) $$\boxed{\begin{aligned} &\text{(a)} \quad P(\mu - \sigma < X \leqq \mu + \sigma) \approx 68\% \\ &\text{(b)} \quad P(\mu - 2\sigma < X \leqq \mu + 2\sigma) \approx 95{,}5\% \\ &\text{(c)} \quad P(\mu - 3\sigma < X \leqq \mu + 3\sigma) \approx 99{,}7\%\,. \end{aligned}}$$

Man kann also erwarten, daß sich die beobachteten Werte einer normalverteilten Zufallsvariablen X bei einer großen Anzahl von Versuchen ungefähr folgendermaßen verteilen:

(a) *Etwa 2/3 aller Werte liegen zwischen $\mu - \sigma$ und $\mu + \sigma$.*

(b) *Etwa 95% aller Werte liegen zwischen $\mu - 2\sigma$ und $\mu + 2\sigma$.*

(c) *Etwa $99^3/_4$% aller Werte liegen zwischen $\mu - 3\sigma$ und $\mu + 3\sigma$.*

Dies kann man auch so ausdrücken:

Eine Abweichung um mehr als σ vom Mittelwert ist etwa einmal bei je 3 Versuchen zu erwarten, eine Abweichung um mehr als 2σ etwa nur einmal bei je 22 Versuchen und eine Abweichung um mehr als 3σ etwa nur einmal bei je 370 Versuchen. Praktisch sind also so gut wie alle Werte zwischen den „**3σ-Grenzen**" $\mu - 3\sigma$ und $\mu + 3\sigma$ zu erwarten.

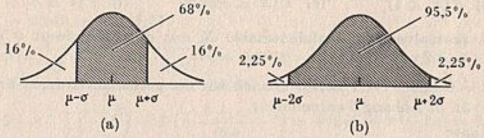


Abbildung 48.2. Zu Formel (48.6)

Ganz ähnlich wie (48.6) ergeben sich aus (48.5) die ebenfalls für die statistische Praxis wichtigen Formeln

$$(48.7)\qquad \boxed{\begin{aligned} &\text{(a) } P(\mu - 1{,}96\sigma < X \leqq \mu + 1{,}96\sigma) = 95\% \\ &\text{(b) } P(\mu - 2{,}58\sigma < X \leqq \mu + 2{,}58\sigma) = 99\% \\ &\text{(c) } P(\mu - 3{,}29\sigma < X \leqq \mu + 3{,}29\sigma) = 99{,}9\%. \end{aligned}}$$

Aufgaben zu Abschnitt 48

48.1 Man beweise

$$(48.8)\qquad \Phi(\infty) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-u^2/2}\, du = 1.$$

Anleitung: Man bilde $\Phi^2(\infty)$ und führe in dem Doppelintegral Polarkoordinaten ein.

48.2 Man zeige, daß die Kurve der Funktion (47.1) zwei Wendepunkte hat, die bei $x = \mu \pm \sigma$ liegen.

48.3 Unter Benutzung von Teilintegration und (48.8) beweise man, daß der Parameter σ in (47.1) tatsächlich die Standardabweichung ist.

48.4 Man beweise

$$(48.9)\qquad \Phi(-z) = 1 - \Phi(z).$$

48.5 Es sei X normalverteilt mit dem Mittelwert 0 und der Varianz 1. Man zeige, daß dann $P(-c \leqq X \leqq c) = \gamma$ gleichbedeutend ist mit $\Phi(c) = (1 + \gamma)/2$.

48.6 Es sei X normalverteilt mit dem Mittelwert μ und der Varianz σ^2. Für welches c wird $P(c \leqq X \leqq c + a)$ bei fest gegebenem positiven a am größten?

49 Zum Gebrauch der Tafeln 3a und 3b in Anhang 5

Die folgenden Beispiele sind typisch. Sie sollen dem Leser Gelegenheit geben, sich im Gebrauch der Tafeln 3a und 3b in Anhang 5 zu üben. Das ist für die Praxis sehr wichtig.

Beispiel 49.1. Wie groß ist die Wahrscheinlichkeit

(a) $P(X \leqq 2{,}44)$ (b) $P(X \leqq -1{,}16)$ (c) $P(X \leqq 1{,}923)$
(d) $P(X \geqq 1)$ (e) $P(X \geqq -2{,}9)$ (f) $P(2 \leqq X \leqq 10)$

für eine normalverteilte Zufallsvariable X mit dem Mittelwert 0 und der Varianz 1?

Da $\mu = 0$ und $\sigma^2 = 1$ ist, so können wir die gesuchten Zahlen direkt aus der Tafel 3a in Anhang 5 entnehmen:

(a) 0,9927
(b) 0,1230
(c) 0,9728 (durch lineare Interpolation)
(d) $1 - P(X \leqq 1) = 1 - 0{,}8413 = 0{,}1587$ [vgl. (21.1)]
(e) $1 - P(X \leqq -2{,}9) = 1 - \Phi(-2{,}9) = 0{,}9981$
(f) $\Phi(10) = 1{,}0000$ (wieso?), $\Phi(2) = 0{,}9772$, $\Phi(10) - \Phi(2) = 0{,}0228$.

Beispiel 49.2. Man bestimme die sechs im vorigen Beispiel genannten Wahrscheinlichkeiten für eine normalverteilte Zufallsvariable X mit dem Mittelwert $\mu = 0{,}8$ und der Varianz $\sigma^2 = 4$.

Aus (48.5) und Tafel 3a in Anhang 5 ergibt sich nun:

(a) $F(2{,}44) = \Phi\left(\frac{2{,}44 - 0{,}80}{2}\right) = \Phi(0{,}82) = 0{,}7939$
(b) $F(-1{,}16) = \Phi(-0{,}98) = 0{,}1635$
(c) $F(1{,}923) = \Phi(0{,}5615) = 0{,}7128$
(d) $1 - P(X \leqq 1) = 1 - F(1) = 1 - \Phi(0{,}1) = 0{,}4602$
(e) $1 - P(X \leqq -2{,}9) = 1 - \Phi\left(\frac{-2{,}9 - 0{,}8}{2}\right) = 1 - \Phi(-1{,}85) = 0{,}9678$
(f) $F(10) - F(2) = \Phi(4{,}6) - \Phi(0{,}6) = 1 - 0{,}7257 = 0{,}2743$.

Beispiel 49.3. Wie groß muß die Konstante c sein, damit

(a) $P(X \geqq c) = 10\%$ (b) $P(X \leqq c) = 5\%$
(c) $P(0 \leqq X \leqq c) = 45\%$ (d) $P(-c \leqq X \leqq c) = 99\%$

gilt? X sei hierbei normalverteilt mit dem Mittelwert $\mu = 0$ und der Varianz $\sigma^2 = 1$.

Aus Tafel 3b in Anhang 5 folgt:

(a) $1 - P(X \leqq c) = 1 - \Phi(c) = 0{,}1$, $\Phi(c) = 0{,}9$, $c = 1{,}282$
(b) $c = -1{,}645$
(c) $\Phi(c) - \Phi(0) = \Phi(c) - 0{,}5 = 0{,}45$, $\Phi(c) = 0{,}95$, $c = 1{,}645$
(d) $c = 2{,}576$.

Beispiel 49.4. X sei normalverteilt mit dem Mittelwert $\mu = -2$ und der Varianz $\sigma^2 = 0{,}25$. Man bestimme c derart, daß folgendes gilt:

(a) $P(X \geqq c) = 0{,}2$
(b) $P(-c \leqq X \leqq -1) = 0{,}5$
(c) $P(-2 - c \leqq X \leqq -2 + c) = 0{,}9$
(d) $P(-2 - c \leqq X \leqq -2 + c) = 99{,}6\%$.

Wir benutzen Tafel 3b in Anhang 5 und erhalten:

(a) $1 - P(X \leqq c) = 1 - \Phi\left(\frac{c+2}{0{,}5}\right) = 0{,}2, \; \Phi(2c + 4) = 0{,}8,$
$2c + 4 = 0{,}842, \; c = -1{,}579$

(b) $\Phi\left(\frac{-1+2}{0{,}5}\right) - \Phi\left(\frac{-c+2}{0{,}5}\right) = 0{,}9772 - \Phi(4 - 2c) = 0{,}5,$
$\Phi(4 - 2c) = 0{,}4772, \; 4 - 2c = -0{,}057, \; c = 2{,}03$

(c) $\Phi\left(\frac{-2+c+2}{0{,}5}\right) - \Phi\left(\frac{-2-c+2}{0{,}5}\right) = \Phi(2c) - \Phi(-2c) = 0{,}9,$
$2c = 1{,}645, \; c = 0{,}823$

(d) $\Phi(2c) - \Phi(-2c) = 99{,}6\%, \; 2c = 2{,}878, \; c = 1{,}439.$

Beispiel 49.5. Eine Metallhobelmaschine stellt Platten her. Kein Produktionsvorgang ist so vollkommen, daß alle Stücke ganz gleich ausfallen. Material und Maschine bedingen eine gewisse Variabilität, die wir als zufällig ansehen müssen, weil sie durch winzige Störursachen bedingt ist, die wir gar nicht alle kennen, geschweige denn in ihrer Wirkung vorherbestimmen können. So läßt sich die Plattendicke X [mm] als Zufallsvariable auffassen, die von Platte zu Platte etwas andere Werte annimmt. X sei normalverteilt und habe bei einer bestimmten Maschineneinstellung den Mittelwert $\mu = 10$ mm und die Standardabweichung $\sigma = 0{,}02$ mm.

Wieviel Prozent Ausschuß sind dann zu erwarten, wenn die Platten (a) mindestens 9,97 mm stark sein sollen, (b) höchstens 10,05 mm stark sein dürfen, (c) um maximal $\pm$ 0,03 mm vom Sollwert 10 mm abweichen dürfen? (d) Wie muß man die Toleranzgrenzen $10 - c$ und $10 + c$ wählen, damit man nicht mehr als 5% Ausschuß erhält? (e) Wie ändert sich der Ausschußprozentsatz für die in Frage (d) bestimmten Toleranzgrenzen, wenn sich μ (zum Beispiel infolge Abnutzung des Hobelstahls) nach 10,01 mm verschiebt?

Aus (48.5) und Tafel 3a in Anhang 5 folgt:

(a) $P(X \leqq 9{,}97) = \Phi\left(\frac{9{,}97 - 10{,}00}{0{,}02}\right) = \Phi(-1{,}5) = 0{,}0668 = 6{,}7\%$

(b) $P(X \geqq 10{,}05) = 1 - P(X \leqq 10{,}05) = 1 - \Phi\left(\frac{10{,}05 - 10{,}00}{0{,}02}\right)$

$$= 1 - \Phi(2{,}5) = 1 - 0{,}9938 = 0{,}6\%$$

(c) $P(9{,}97 \leqq X \leqq 10{,}03) = \Phi\left(\frac{10{,}03 - 10{,}00}{0{,}02}\right) - \Phi\left(\frac{9{,}97 - 10{,}00}{0{,}02}\right)$

$$= \Phi(1{,}5) - \Phi(-1{,}5) = D(1{,}5) = 0{,}8664.$$

Die Antwort lautet also $1 - 0{,}8664 = 13\%$.

(d) 95% der Platten sollen Stärken zwischen $10 - c$ und $10 + c$ aufweisen.

Aus (48.7a) folgt demnach $c = 1{,}96\,\sigma = 0{,}039$. Die Grenzen sind also 9,961 mm und 10,039 mm.

(e) $$P(9{,}961 \leq X \leq 10{,}039) = \Phi\left(\frac{10{,}039 - 10{,}010}{0{,}02}\right) - \Phi\left(\frac{9{,}961 - 10{,}010}{0{,}02}\right)$$
$$= \Phi(1{,}45) - \Phi(-2{,}45) = 0{,}9265 - 0{,}0071 = 92\,\%.$$

Der zu erwartende Ausschußanteil erhöht sich auf 8 %, also beträchtlich.

Aufgaben zu Abschnitt 49

49.1 Man beantworte die Frage (b) in Beispiel 49.1 und die Frage (a) in Beispiel 49.3 näherungsweise, indem man Abb. 48.1 benützt.

49.2 Man veranschauliche das Ergebnis in Beispiel 49.5, Frage (e), indem man die Kurven der Dichte in Frage (d) und (e) in ein gemeinsames Koordinatensystem zeichnet.

49.3 Man gewinne die Antwort auf Frage (d) in Beispiel 49.5 aus Tafel 3b.

49.4 Wie stark verringert sich der Ausschußanteil in Beispiel 49.5, Frage (c), wenn man eine bessere Hobelmaschine einsetzt, bei der $\sigma = 0{,}01$ mm ist?

49.5 Wie groß müßte σ in Beispiel 49.5, Frage (c), sein, damit man nur 1 % Ausschuß erhält?

49.6 Die Zufallsvariable X sei normalverteilt mit dem Mittelwert 10. Wie groß darf σ höchstens sein, damit die Wahrscheinlichkeit, daß X irgendeinen Wert $x \geq 15$ annimmt, nicht größer als 5 % ist?

49.7 Wie ändert sich das Ergebnis in Aufgabe 49.6, wenn man $x \geq 15$ durch $x \geq 11$ ersetzt?

49.8 Im Kriege mußten die Brötchen durchschnittlich 50 g wiegen. Ein Bäcker, der bewußt etwas leichtere Brötchen herstellte, las die ganz leichten aus, wenn er merkte, daß die Kontrolle kam. Wie könnte der Kontrolleur diesen Schwindel vermuten, wenn er weiß, daß das Brötchengewicht etwa normalverteilt sein muß?

49.9 Eine Firma stellt Luftpostumschläge her, deren Gewicht erfahrungsgemäß normalverteilt ist mit dem Mittelwert $\mu = 1{,}95$ g und der Standardabweichung $\sigma = 0{,}05$ g. Wie viele Umschläge, die 2 g oder mehr wiegen, muß man dann in einem Päckchen von 100 Umschlägen etwa mit in Kauf nehmen?

49.10 Die Variablen X_0 und X_1 seien normalverteilt mit dem Mittelwert $\mu_0 = 10$ bzw. $\mu_1 = 20$ und derselben Varianz $\sigma^2 = 36$. Man bestimme die Zahl c derart, daß die Wahrscheinlichkeit α, mit der X_0 irgendeinen Wert größer als c annimmt, nur 5 % beträgt. Wie groß ist dann die Wahrscheinlichkeit β, daß X_1 irgendeinen Wert größer als c annimmt? Man skizziere die beiden Dichten und c in einem gemeinsamen Koordinatensystem und deute α und β als Flächeninhalte anschaulich.

49.11 Wie ändert sich β in Aufgabe 49.10, wenn man σ^2 verkleinert, indem man z. B. $\sigma^2 = 4$ wählt?

49.12 Die Zufallsvariable Y sei normalverteilt mit dem Mittelwert $\varkappa$ und der Varianz λ^2. Man setze $Y = \ln X$ und zeige, daß X dann die Dichte

$$f(x) = \frac{1}{x\lambda\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{\ln x - \varkappa}{\lambda}\right)^2} \qquad \text{für } x > 0 \tag{49.1}$$

und $f(x) = 0$ für $x < 0$ hat. Diese Verteilung heißt die **logarithmische Normalverteilung.**

49.13 Man zeichne die Dichte der logarithmischen Normalverteilung mit $\varkappa = 0$ und $\lambda = 1$.

49.14 Man zeige, daß die logarithmische Normalverteilung den Mittelwert

$$\mu = e^{\varkappa + \frac{1}{2}\lambda^2}$$

hat.

50 Annäherung der Binomialverteilung durch die Normalverteilung

Wie wir aus Abschn. 39 und 40 wissen, hat die Binomialverteilung die Wahrscheinlichkeitsfunktion

(50.1) $$f(x) = \binom{n}{x} p^x q^{n-x} \qquad (x = 0, 1, \ldots, n),$$

den Mittelwert

(50.2) $$\mu = np$$

und die Varianz

(50.3) $$\sigma^2 = npq.$$

Für große Werte von n läßt sich die Binomialverteilung durch die Normalverteilung mit dem Mittelwert (50.2) *und der Varianz* (50.3), *also mit der Dichte*

(50.4) $$f^*(x) = \frac{1}{\sqrt{2\pi}\sqrt{npq}} e^{-z^2/2}, \qquad z = \frac{x - np}{\sqrt{npq}}$$

annähern. Für $x = 0, 1, \ldots, n$ ist dann $f(x) \approx f^*(x)$. Weiterhin gilt

(50.5) $$\begin{gathered} P(a \leqq X \leqq b) = \sum_{x=a}^{b} \binom{n}{x} p^x q^{n-x} \approx \Phi(\beta) - \Phi(\alpha), \\ \alpha = \frac{a - np - 0{,}5}{\sqrt{npq}}, \quad \beta = \frac{b - np + 0{,}5}{\sqrt{npq}}. \end{gathered}$$

Der praktische Wert einer solchen Annäherung leuchtet unmittelbar ein, wenn man bedenkt, wie unhandlich (50.1) für große n, etwa $n = 100$ oder gar 1000, wird. Tab. 50.1 zeigt übrigens, daß die Näherung sogar für relativ kleine n schon recht gut ist, sofern p nahe bei 1/2 liegt.

Auf die theoretische Untersuchung des Fehlers der Näherung können wir in der vorliegenden elementaren Einführung nicht eingehen. Wohl aber wollen wir die Sätze formulieren, aus denen sich die obigen Näherungen ergeben.

Satz 50.1 (Lokaler Grenzwertsatz von de Moivre und Laplace). *Für* $0 < p < 1$ *und jedes* $x = 0, 1, \ldots, n$ *genügt die Wahrscheinlichkeitsfunktion* (50.1) *der Binomialverteilung der Beziehung*

$$f(x) \sim f^*(x) \qquad (n \to \infty) \tag{50.6}$$

mit $f^*(x)$ *gemäß* (50.4) [*und zwar gleichmäßig für alle* x, *für die* z *in einem endlichen Intervall liegt*].

Das Zeichen

$\sim$ (lies „**asymptotisch gleich**")

in (50.6) bedeutet, daß das Verhältnis beider Seiten für $n \to \infty$ gegen 1 strebt. Für große n wird also $f(x)$ durch $f^*(x)$ brauchbar angenähert.

Der Satz wird als *lokaler Satz* bezeichnet, weil sich (50.6) auf das Verhalten von $f(x)$ an festen Stellen x bezieht. Den Beweis des Satzes findet man in Anhang 1.

Satz 50.2 (Grenzwertsatz von de Moivre und Laplace). *Für die Wahrscheinlichkeit, daß eine Zufallsvariable* X *mit der Wahrscheinlichkeitsfunktion* (50.1) *und* $0 < p < 1$ *irgendeinen Wert in* $a \leqq X \leqq b$ $(a, b = 0, 1, \ldots, n, a \leqq b)$ *annimmt, gilt*

$$P(a \leqq X \leqq b) \sim \frac{1}{\sqrt{2\pi}} \int_\alpha^\beta e^{-u^2/2}\, du = \Phi(\beta) - \Phi(\alpha) \tag{50.7}$$

mit α *und* β *wie in* (50.5) *und* Φ *gemäß* (48.3).

Tabelle 50.1. Wahrscheinlichkeitsfunktion (50.1) und Näherung (50.4) für verschiedene n und p

x	$n = 8,\ p = 0{,}2$		$n = 8,\ p = 0{,}5$		$n = 25,\ p = 0{,}2$	
	Näherung	Exakt	Näherung	Exakt	Näherung	Exakt
0	0,130	0,168	0,005	0,004	0,009	0,004
1	0,306	0,336	0,030	0,031	0,027	0,024
2	0,331	0,294	0,104	0,109	0,065	0,071
3	0,164	0,147	0,220	0,219	0,121	0,136
4	0,037	0,046	0,282	0,273	0,176	0,187
5	0,004	0,009	0,220	0,219	0,199	0,196
6	0,000	0,001	0,104	0,109	0,176	0,163
7	0,000	0,000	0,030	0,031	0,121	0,111
8	0,000	0,000	0,005	0,004	0,065	0,062
9					0,027	0,029
10					0,009	0,012
11					0,002	0,004

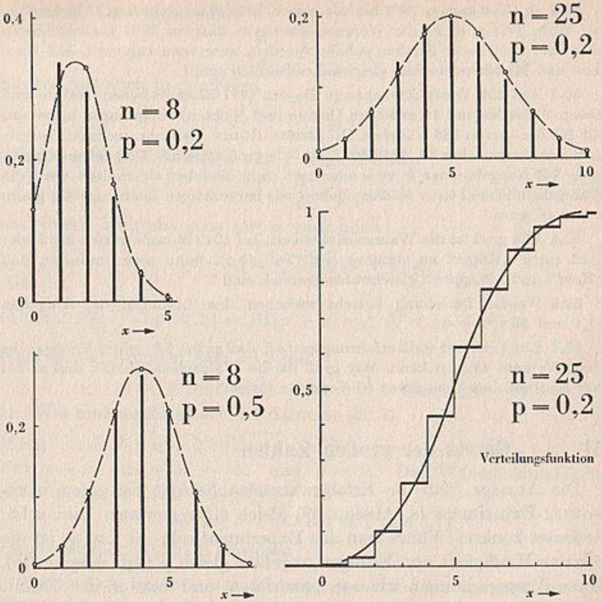


Abbildung 50.1. Zu Tab. 50.1

Die Verteilungsfunktion der betrachteten Binomialverteilung ist also asymptotisch gleich der Verteilungsfunktion der Normalverteilung mit dem Mittelwert (50.2) und der Varianz (50.3). Den Beweis dieses Satzes findet man ebenfalls in Anhang 1.

Aufgaben zu Abschnitt 50

50.1 Eine Firma liefert Glimmlämpchen in Kartons zu je 1000 Stück. Wie groß ist die Wahrscheinlichkeit, daß ein solcher Karton nicht mehr als 1% Ausschuß (= defekte Lämpchen) enthält, wenn man den Produktionsvorgang als Bernoulli-Experiment mit $p = 1\%$ (= Wahrscheinlichkeit, ein defektes Lämpchen zu produzieren) ansehen kann?

50.2 Man bestimme die Wahrscheinlichkeit, bei 10 Münzenwürfen mindestens 4 „Köpfe" und höchstens 6 „Köpfe" zu erhalten, aus (50.5) und vergleiche mit dem exakten Wert.

50.3 In Graz kamen 1962 bei den ersten 3000 Einzelgeburten 1578 Knaben zur Welt. Wie groß ist die Wahrscheinlichkeit, daß bei 3000 Einzelgeburten 1578 oder noch mehr Knaben geboren werden, wenn man annimmt, daß Knaben- und Mädchengeburten gleichwahrscheinlich sind?

50.4 Aus 299 Verwandtenehen in Bayern (287 Ehen zwischen Vettern und Basen 1. Grades und 12 zwischen Onkeln und Nichten) entsprangen insgesamt 709 Kinder, davon 386 Mädchen (E. ZERBIN-RÜDIN, Zeitschr. menschl. Vererb. u. Konstitutionslehre **35**, 1959/60, 274). Wie groß wäre die Wahrscheinlichkeit, unter 709 Neugeborenen so viele oder noch mehr Mädchen zu erhalten, wenn die Wahrscheinlichkeit einer Mädchengeburt wie bei sonstigen Ehen ungefähr gleich $p = 0{,}48$ wäre?

50.5 Wie groß ist die Wahrscheinlichkeit, bei 4040 Münzenwürfen 2048 oder noch mehr „Köpfe" zu erhalten (vgl. Tab. 15.1), wenn man annimmt, daß „Kopf" und „Wappen" gleichwahrscheinlich sind?

50.6 Welche Beziehung besteht zwischen den Lösungen der Aufgaben 43.1 und 50.1?

50.7 Ein Vertreter weiß erfahrungsgemäß, daß er bei 5% seiner Erstbesuche einen Verkauf tätigen kann. Wie groß ist die Wahrscheinlichkeit, daß er bei 200 Erstbesuchen wenigstens 10 Verkäufe tätigt?

51 Gesetz der großen Zahlen

Die Aussage, daß die Erfolgswahrscheinlichkeit bei einem BERNOULLI-Experiment (s. Abschn. 39) gleich einer gewissen Zahl p ist, bedeutet konkret: Führt man das Experiment sehr oft aus, so ist die relative Häufigkeit von Erfolgen ungefähr gleich p (vgl. Abschn. 15). Diese Aussage können wir nun präzisieren, und zwar in der Gestalt des sogenannten klassischen Gesetzes der großen Zahlen. Dieses Gesetz und seine Verallgemeinerungen gehören zu den wichtigsten Sätzen der Wahrscheinlichkeitstheorie.

Satz 51.1 (Gesetz der großen Zahlen von Bernoulli). *Es sei A ein Ereignis, das bei einem Zufallsexperiment die Wahrscheinlichkeit p $(0 < p < 1)$ besitzt. X bezeichne die Anzahl des Eintreffens von A bei n unabhängigen Ausführungen dieses Experimentes. Weiterhin sei eine beliebige positive Zahl ε vorgegeben. (ε darf beliebig klein, aber nicht null sein.) Dann gilt*

$$P\left(\left|\frac{X}{n} - p\right| < \varepsilon\right) \to 1 \qquad \textit{für } n \to \infty .$$

Mit wachsendem n strebt also die Wahrscheinlichkeit, daß die durchschnittliche Anzahl des Auftretens von A um mehr als eine beliebig vorgegebene positive Zahl ε von p abweicht, gegen null.

Beweis. Die Ungleichung

$$\left|\frac{X}{n} - p\right| < \varepsilon$$

können wir auch in der Form

$$-\varepsilon < \frac{X}{n} - p < \varepsilon$$

schreiben. Durch Addition von p folgt hieraus

$$p - \varepsilon < \frac{X}{n} < p + \varepsilon$$

und durch Multiplikation mit n schließlich

$$[p - \varepsilon]\, n < X < [p + \varepsilon]\, n.$$

Also ist

$$P\left(\left|\frac{X}{n} - p\right| < \varepsilon\right) = P([p - \varepsilon]\, n < X < [p + \varepsilon]\, n). \tag{51.1}$$

Nun wenden wir den Satz 50.2 mit

$$a = [p - \varepsilon]\, n \qquad \text{und} \qquad b = [p + \varepsilon]\, n$$

an. Wie man nachrechnet, ist dann in (50.7)

$$\alpha = -(\varepsilon n + 0{,}5)/\sqrt{n p q} \quad \text{und} \quad \beta = (\varepsilon n + 0{,}5)/\sqrt{n p q}. \tag{51.2}$$

Für $n \to \infty$ strebt $\alpha \to -\infty$ und $\beta \to \infty$. Die Wahrscheinlichkeit in (51.1) strebt dann also gegen

$$\Phi(\infty) - \Phi(-\infty) = 1 - 0 = 1.$$

Damit ist der Satz 51.1 bewiesen.

Beispiel 51.1. Im Falle eines Münzenwurfs ist $p = 0{,}5$, und (51.2) hat die Form

$$\alpha = -2\varepsilon\sqrt{n} - \frac{1}{\sqrt{n}} \qquad \text{und} \qquad \beta = 2\varepsilon\sqrt{n} + \frac{1}{\sqrt{n}}.$$

Die Wahrscheinlichkeit in (51.1) ist dann also gemäß Satz 50.2 näherungsweise gleich

$$\Phi\left(2\varepsilon\sqrt{n} + \frac{1}{\sqrt{n}}\right) - \Phi\left(-2\varepsilon\sqrt{n} - \frac{1}{\sqrt{n}}\right). \tag{51.3}$$

BUFFON erhielt bei 4040 Würfen die relative Häufigkeit 0,5069, die um $\varepsilon = 0{,}0069$ von $p = 0{,}5$ abweicht (vgl. Abschn. 15). Mit $n = 4040$ und $\sqrt{n} = 63{,}561$ ergibt sich dann aus (51.3) und Tafel 3a in Anhang 5 die Wahrscheinlichkeit

$$P\left(\left|\frac{X}{4040} - 0{,}5\right| < 0{,}0069\right) = \Phi(0{,}893) - \Phi(-0{,}893) = 63\%.$$

Die beobachtete Abweichung ist also unter der Annahme $p = 0{,}5$ recht wahrscheinlich. Wir können auch sagen: Die Annahme $p = 0{,}5$ verträgt sich recht gut mit der beobachteten Wirklichkeit. Ein sehr kleiner Wert (z. B. $P = 1\%$) gäbe uns Grund, die Richtigkeit der genannten Annahme zu bezweifeln. Wir gehen auf diese Schlußweise in Teil III ausführlich ein.

Kapitel 10

Wahrscheinlichkeitsverteilungen mehrerer Zufallsvariablen

Die in der Kapitelüberschrift genannten Verteilungen sind aus zwei Gründen wichtig:

Erstens betrachtet man bei Zufallsexperimenten oftmals mehrere Größen gleichzeitig, und dann hat man es mit einer entsprechenden Anzahl von Zufallsvariablen gleichzeitig zu tun. Beispiele sind die gleichzeitige Feststellung mehrerer Rassemerkmale von Tieren oder Personen (Gewicht, Größe, Augenfarbe, Schädelform), die gleichzeitige Bestimmung mehrerer Eigenschaften eines Werkstoffes (bei Stahl etwa Kohlenstoffgehalt, Zugfestigkeit, Brinellhärte) und so weiter.

Zweitens spielen Verteilungen mehrerer Variablen bei der theoretischen Begründung statistischer Prüfverfahren (Tests) eine Rolle. Wie wir später sehen werden, kann man diese Tests praktisch anwenden, auch ohne daß man über die zugrunde liegende Theorie genau Bescheid weiß. Für jemanden, der sich damit begnügt, ist das vorliegende und auch das nächste Kapitel in diesem Zusammenhange entbehrlich. Wer sich aber für die theoretischen Grundlagen der Tests interessiert, sollte sich in die Ideen, die wir im folgenden erörtern, hineindenken.

Wir beginnen unsere Betrachtung mit dem einfachsten Fall, nämlich mit zwei Zufallsvariablen. Das hat den Vorteil, daß wir viele Zusammenhänge und Begriffe anschaulich erfassen können.

Weitere Tatsachen über Verteilungen mehrerer Zufallsvariablen behandeln wir später (in Kap. 17 und 18), wenn wir schon einen Einblick in das Wesen statistischer Tests besitzen.

52 Zweidimensionale Verteilungen

Zu einem Zufallsexperiment, bei dem man eine einzelne Größe beobachtet, gehört eine einzelne Zufallsvariable, die wir mit X bezeichnen wollen. Das Ergebnis einer Ausführung des Experimentes ist durch eine einzelne Zahl x, nämlich den Wert, den X jeweils annimmt, gekennzeichnet. Dieser Wert entspricht einem Punkt auf der X-Achse. Kennen wir für jedes Intervall $a < X \leqq b$ die zugehörige Wahrscheinlichkeit

$$P(a < X \leqq b),$$

mit der X irgendeinen Wert aus diesem Intervall annimmt, so ist die Wahrscheinlichkeitsverteilung der Zufallsvariablen X bekannt. Wir wissen, daß sich diese Verteilung durch die zugehörige Verteilungsfunktion

$$F(x) = P(X \leqq x)$$

eindeutig kennzeichnen läßt. In der Tat gilt für beliebige Zahlen a und b $(b > a)$ die Beziehung

$$P(a < X \leqq b) = F(b) - F(a).$$

Vgl. Abschn. 24.

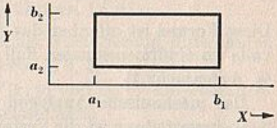


Abbildung 52.1. Zum Begriff der zweidimensionalen Verteilung

Zu einem Zufallsexperiment, bei dem man *gleichzeitig zwei Größen* beobachtet, gehören *zwei Zufallsvariable*, die wir mit X und Y bezeichnen wollen. Das Ergebnis einer Ausführung des Experimentes ist dann jeweils durch ein geordnetes Zahlenpaar $X = x$, $Y = y$ oder kürzer (x, y) gekennzeichnet. Hierbei ist x bzw. y der Wert, den X bzw. Y bei der genannten Ausführung angenommen hat. Diesem Zahlenpaar (x, y) entspricht ein Punkt in der XY-Ebene mit der Abszisse $X = x$ und der Ordinate $Y = y$.

Kennen wir für jedes Rechteck (siehe Abb. 52.1)

$$a_1 < X \leqq b_1, \qquad a_2 < Y \leqq b_2$$

die zugehörige Wahrscheinlichkeit

$$P(a_1 < X \leqq b_1, a_2 < Y \leqq b_2),$$

mit der (X, Y) irgendein Paar von Werten annimmt, das einem Punkt in diesem Rechteck entspricht, so sagen wir, daß die **zweidimensionale Wahrscheinlichkeitsverteilung** der beiden Zufallsvariablen X und Y oder der **zweidimensionalen Zufallsvariablen** (X, Y) bekannt sei.

Die Funktion

$$\text{(52.1)} \qquad \boxed{F(x, y) = P(X \leqq x, Y \leqq y)}$$

heißt die **Verteilungsfunktion** der betreffenden zweidimensionalen Verteilung. Wie man sieht, gibt diese Funktion die Wahrscheinlichkeit an, mit der X einen beliebigen Wert annimmt, der kleiner als x oder gleich x ist, und mit der Y gleichzeitig einen Wert annimmt, der kleiner als y oder gleich y ist.

Die Verteilungsfunktion $F(x, y)$ bestimmt die betreffende zweidimensionale Verteilung eindeutig. In der Tat gilt nämlich für jedes Rechteck $a_1 < X \leqq b_1$, $a_2 < Y \leqq b_2$ die Beziehung

$$\begin{aligned} &P(a_1 < X \leqq b_1,\ a_2 < Y \leqq b_2) \\ &= F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2). \end{aligned} \tag{52.2}$$

Diese Formel ist offenbar das Analogon der obigen Formel für Intervalle im eindimensionalen Fall. Der Beweis sei dem Leser überlassen (s. Aufgabe 52.1).

Das **mechanische Analogon** einer zweidimensionalen Wahrscheinlichkeitsverteilung ist die Verteilung einer Masse vom Gesamtbetrage 1 in der XY-Ebene. Dabei entspricht die Masse in einem Rechteck R der Wahrscheinlichkeit, mit der (X, Y) irgendein Wertepaar in R annimmt.

Wie im Falle eindimensionaler Verteilungen unterscheiden wir die beiden praktisch besonders wichtigen Typen der diskreten und der stetigen zweidimensionalen Verteilungen, die wir nun nacheinander betrachten.

Aufgaben zu Abschnitt 52

52.1 Man beweise (52.2). Anleitung: Man überlege sich zuerst, welche Wahrscheinlichkeit jedes der vier Glieder auf der rechten Seite von (52.2) darstellt.

52.2 In sinngemäßer Verallgemeinerung von (52.1) definiere man die Verteilungsfunktion $F(x, y, z)$ einer dreidimensionalen Zufallsvariablen (X, Y, Z) und überlege sich dann das Analogon der Formel (52.2).

53 Diskrete zweidimensionale Verteilung

Eine zweidimensionale Zufallsvariable (X, Y) und deren Wahrscheinlichkeitsverteilung heißen **diskret**, wenn folgendes gilt: Die Variable (X, Y) kann nur endlich viele oder abzählbar unendlich viele Wertepaare (x, y) mit positiver Wahrscheinlichkeit annehmen. Dabei enthält jedes beschränkte Gebiet in der XY-Ebene nur endlich viele dieser Wertepaare. Zu jedem Gebiet, das keines dieser Paare enthält, gehört die Wahrscheinlichkeit null. (Beschränktheit eines Gebietes bedeutet, daß jeder Punkt des Gebietes einen endlichen Abstand vom Nullpunkt hat.)

Wir bezeichnen die zu einem solchen Paar (x_i, y_j) gehörige Wahrscheinlichkeit mit p_{ij}, so daß also

$$P(X = x_i, Y = y_j) = p_{ij}$$

ist. (Hierbei kann p_{ij} für gewisse Indexkombinationen natürlich gleich null sein.) Dann können wir unsere Verteilung, ähnlich wie im eindimensionalen Falle, durch eine **Wahrscheinlichkeitsfunktion**

$$f(x, y) = \begin{cases} p_{ij} & \text{für } x = x_i,\ y = y_j \\ 0 & \text{für alle übrigen } (x, y) \end{cases} \tag{53.1}$$

kennzeichnen. Die Indizes i und j dienen der Numerierung, durchlaufen also (unabhängig voneinander) die Werte 1, 2, 3, ...

Zwischen der Verteilungsfunktion $F(x, y)$ und der Wahrscheinlichkeitsfunktion $f(x, y)$ besteht der Zusammenhang

$$F(x, y) = \sum_{x_i \leq x} \sum_{y_j \leq y} f(x_i, y_j). \tag{53.2}$$

Diese Formel ist das Analogon der Formel (25.1) im eindimensionalen Fall. Statt (22.2) gilt nun

$$\sum_i \sum_j f(x_i, y_j) = 1.$$

Hierbei wird über alle Wertepaare summiert, für die $f(x, y)$ nicht null ist.

Beispiel 53.1. Beim einmaligen Wurf zweier unterscheidbarer Münzen sind 4 Ereignisse

$$WW \qquad WK \qquad KW \qquad KK$$

möglich. K bedeutet „Kopf", und W bedeutet „Wappen". Der erste Buchstabe bezeichnet das bei der 1. Münze eingetroffene Ereignis und der zweite das bei der 2. Münze eingetroffene. Wir betrachten die Zufallsvariablen

X = *Anzahl der „Köpfe" bei der 1. Münze*,

Y = *Anzahl der „Köpfe" bei der 2. Münze*.

Dann bedeuten natürlich

$X = 0$ das Ereignis W bei der 1. Münze,

$X = 1$ das Ereignis K bei dieser Münze,

usw. Und die obigen 4 Ereignisse entsprechen der Reihe nach den folgenden Paaren $(X = x,\ Y = y)$:

$$(0, 0) \qquad (0, 1) \qquad (1, 0) \qquad (1, 1).$$

Da diese Ereignisse gleichwahrscheinlich sind, hat jedes die Wahrscheinlichkeit 1/4. So ergeben sich die in Tab. 53.1 gezeigten Werte der Wahrscheinlichkeitsfunktion $f(x, y)$ und daraus die in Tab. 53.2 angegebenen Werte der Verteilungsfunktion $F(x, y)$. Abb. 53.1 zeigt das zu $f(x, y)$ gehörige Stabdiagramm und Abb. 53.2 einen Teil der durch $F(x, y)$ dargestellten Treppenfläche über der xy-Ebene.

Tabelle 53.1. Werte der Wahrscheinlichkeitsfunktion $f(x, y)$ in Beispiel 53.1

	$y = 0$	$y = 1$
$x = 0$	1/4	1/4
$x = 1$	1/4	1/4

Tabelle 53.2. Werte der Verteilungsfunktion $F(x, y)$ in Beispiel 53.1

	$y < 0$	$0 \leq y < 1$	$y \geq 1$
$x < 0$	0	0	0
$0 \leq x < 1$	0	1/4	1/2
$x \geq 1$	0	1/2	1

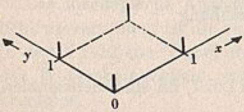


Abbildung 53.1. Stabdiagramm der Wahrscheinlichkeitsfunktion $f(x, y)$ in Beispiel 53.1 (vgl. Tab. 53.1)

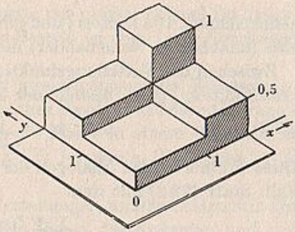


Abbildung 53.2. Verteilungsfunktion $F(x, y)$ in Beispiel 53.1 (vgl. Tab. 53.2)

Aufgaben zu Abschnitt 53

53.1 Aus einem Skatspiel werden nacheinander 3 Karten mit Zurücklegen gezogen. Man zeige, daß die Wahrscheinlichkeitsfunktion $f(x, y)$ der Zufallsvariablen (X, Y),

$X =$ *Anzahl der Könige unter den gezogenen Karten*,

$Y =$ *Anzahl der Asse und Zehnen unter den gezogenen Karten*,

die Form

$$f(x, y) = \frac{3!}{x!\,y!\,(3-x-y)!}\left(\frac{1}{8}\right)^x \left(\frac{1}{4}\right)^y \left(\frac{5}{8}\right)^{3-x-y} \qquad (x + y \leqq 3),$$

$f(x, y) = 0$ für alle übrigen Paare (x, y), und die in Tab. 53.3 angegebenen Werte hat.

Tabelle 53.3. Zu Aufgabe 53.1

$x \backslash y$	0	1	2	3
0	125/512	150/512	60/512	8/512
1	75/512	60/512	12/512	0
2	15/512	6/512	0	0
3	1/512	0	0	0

53.2 Man stelle die Wahrscheinlichkeitsfunktion in Aufgabe 53.1 und auch die zugehörige Verteilungsfunktion graphisch dar.

53.3 Die durch die Wahrscheinlichkeitsfunktion

$$f(x_1, x_2, \ldots, x_r) = \frac{n!}{x_1!\, x_2! \cdots x_r!} p_1^{x_1} p_2^{x_2} \cdots p_r^{x_r}$$

gegebene Verteilung heißt die **Multinomialverteilung.** Man zeige: $f(x_1, \ldots, x_r)$ ist die Wahrscheinlichkeit, daß bei n unabhängigen Ausführungen eines Experimentes ein Ereignis A_1 genau x_1-mal, ..., ein Ereignis A_r genau x_r-mal eintrifft, wenn bei der Einzelausführung eines und nur eines der Ereignisse $A_1, \ldots, A_r$ eintrifft, und zwar mit der Wahrscheinlichkeit $p_1, \ldots$ bzw. p_r.

53.4 Man zeige, daß die Binomialverteilung und die Verteilung in Aufgabe 53.1 Sonderfälle der Multinomialverteilung sind.

53.5 Bei einem Produktionsprozeß seien die obere und untere Toleranzgrenze derart festgelegt, daß die Wahrscheinlichkeit der Zurückweisung eines Artikels gleich $p_1 = 3\%$ bzw. $p_2 = 5\%$ ist. Wie groß ist dann die Wahrscheinlichkeit, daß in einer Stichprobe von 20 Stück genau x_1 bzw. x_2 Artikel zurückgewiesen werden, weil sie oberhalb der oberen bzw. unterhalb der unteren Toleranzgrenze liegen?

54 Stetige zweidimensionale Verteilung

Eine zweidimensionale Zufallsvariable (X, Y) und deren Wahrscheinlichkeitsverteilung heißen **stetig,** wenn sich die zugehörige Verteilungsfunktion $F(x, y)$ durch ein Doppelintegral in der Form

$$F(x, y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(x^*, y^*)\, dx^*\, dy^* \tag{54.1}$$

darstellen läßt, wobei $f(x, y)$ in der ganzen Ebene definiert, nichtnegativ und beschränkt ist und mit Ausnahme von höchstens endlich vielen (stetig differenzierbaren) Kurven überall stetig ist. Die Funktion $f(x, y)$ heißt die **Wahrscheinlichkeitsdichte** der betreffenden Verteilung.

Die zu einem Rechteck R: $a_1 < X \leqq b_1$, $a_2 < Y \leqq b_2$ gehörige Wahrscheinlichkeit ist dann durch

$$P(a_1 < X \leqq b_1, a_2 < Y \leqq b_2) = \int_{a_2}^{b_2} \int_{a_1}^{b_1} f(x, y)\, dx\, dy \tag{54.2}$$

gegeben. Diese Wahrscheinlichkeit läßt sich natürlich vermöge (52.2) durch die Verteilungsfunktion ausdrücken.

Weiterhin gilt

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1. \tag{54.3}$$

Beispiel 54.1. (Gleichförmige Verteilung). Der eindimensionalen gleichförmigen Verteilung in einem Intervall (vgl. Beispiel 26.1) entspricht im zweidimensionalen Falle die gleichförmige Verteilung in einem Rechteck R: $\alpha_1 < X \leq \beta_1$, $\alpha_2 < Y \leq \beta_2$ mit der Wahrscheinlichkeitsdichte

$$f(x, y) = \begin{cases} 1/k & \text{für } (x, y) \text{ in } R \\ 0 & \text{für } (x, y) \text{ außerhalb } R. \end{cases}$$

Die Konstante $k = (\beta_1 - \alpha_1)(\beta_2 - \alpha_2)$ ist hierbei der Flächeninhalt dieses Rechtecks. Vgl. Abb. 54.1.

Der Wert der Verteilungsfunktion $F(x, y)$ in einem Punkt (x, y) ist gleich dem Flächeninhalt des Teiles von R, der in dem Bereich $X \leq x$, $Y \leq y$ der XY-Ebene liegt (vgl. Abb. 54.2).

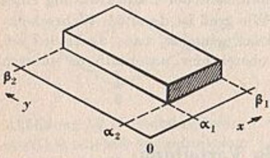


Abbildung 54.1. Wahrscheinlichkeitsfunktion in Beispiel 54.1

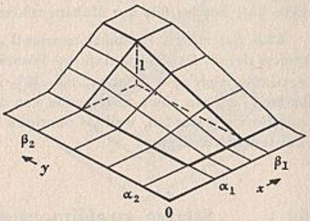


Abbildung 54.2. Verteilungsfunktion der gleichförmigen Verteilung in Beispiel 54.1

Aufgaben zu Abschnitt 54

54.1 Wie kann man die Verteilungsfunktion in Beispiel 54.1 formelmäßig darstellen?

54.2 Man skizziere die Höhenlinien $F(x, y) = konst$ der Verteilungsfunktion in Beispiel 54.1.

54.3 Es sei $f(x, y) = 2$ für $x > 0$, $y > 0$, $x + y < 1$ und $f(x, y) = 0$ außerhalb dieses Gebietes. Man stelle $f(x, y)$ und die zugehörige Verteilungsfunktion $F(x, y)$ graphisch dar.

55 Randverteilungen

Jeder zweidimensionalen Verteilung lassen sich zwei eindimensionale Verteilungen, die sogenannten *Randverteilungen* zuordnen. Um den Begriff der Randverteilung einzuführen, gehen wir am einfachsten von einem Beispiel aus.

Beispiel 55.1. In Aufgabe 53.1 haben wir die Verteilung der beiden Zufallsvariablen

$$X = \textit{Anzahl der Könige}$$
$$Y = \textit{Anzahl der Asse oder Zehnen}$$

beim Ziehen dreier Karten aus einem Skatspiel mit Zurücklegen und die zugehörige Wahrscheinlichkeitsfunktion $f(x, y)$ betrachtet.

Wir fragen nun nach der Wahrscheinlichkeit des Ereignisses

$$X = x, \quad Y \text{ beliebig}$$

d. h. wir fragen, mit welcher Wahrscheinlichkeit erhalten wir genau x Könige und irgendeine Anzahl von Assen und Zehnen, wobei es auf die zuletztgenannte Anzahl nun überhaupt nicht ankommen soll. Da sich die Ereignisse, zu denen in Tab. 53.3 eine positive Wahrscheinlichkeit gehört, gegenseitig ausschließen, so folgt aus Axiom 3 in Abschn. 15, daß wir die gesuchte Wahrscheinlichkeit, die natürlich eine Funktion von x ist, etwa

$$f_1(x) = P(X = x, \ Y \text{ beliebig}),$$

für $x = 0, 1, 2, 3$ jeweils durch Zeilensummenbildung aus der Tab. 53.3 erhalten. Es gilt also

$$f_1(x) = \sum_y f(x, y),$$

wobei alle Werte summiert werden, für die $f(x, y)$ bei dem betreffenden x nicht null ist. Zum Beispiel ist

$$\begin{aligned} f_1(0) &= f(0, 0) + f(0, 1) + f(0, 2) + f(0, 3) \\ &= \frac{125}{512} + \frac{150}{512} + \frac{60}{512} + \frac{8}{512} = \frac{343}{512} \end{aligned}$$

und so weiter.

Andererseits können wir uns aber auch nur um Y kümmern und X beiseite lassen, d. h. nach der Wahrscheinlichkeit

$$P(X \text{ beliebig}, \ Y = y)$$

fragen, mit der man genau y Asse und Zehnen erhält, wobei es auf die Zahl der Könige nicht ankommen soll. Offenbar ist diese, natürlich von y abhängige, Wahrscheinlichkeit gleich dem zugehörigen Wert der Funktion

$$f_2(y) = \sum_x f(x, y).$$

Hierbei werden alle Werte summiert, für die $f(x, y)$ bei dem betreffenden y nicht null ist. Die Werte dieser Funktion ergeben sich durch Spaltenaddition in Tab. 53.3.

Die Zeilen- und Spaltensummen, also die Werte der Funktionen f_1 und f_2, können wir am Rande der Tab. 53.3 eintragen. Dann erhalten wir die Tab. 55.1.

Tabelle 55.1. Werte der Wahrscheinlichkeitsfunktionen f_1 und f_2 der Randverteilungen in Beispiel 55.1

	$y = 0$	$y = 1$	$y = 2$	$y = 3$	$f_1(x)$
$x = 0$	125/512	150/512	60/512	8/512	343/512
$x = 1$	75/512	60/512	12/512	0	147/512
$x = 2$	15/512	6/512	0	0	21/512
$x = 3$	1/512	0	0	0	1/512
$f_2(y)$	216/512	216/512	72/512	8/512	

Unser Beispiel lehrt: Bei einer gegebenen diskreten Verteilung einer zweidimensionalen Zufallsvariablen (X, Y) mit der Wahrscheinlichkeitsfunktion $f(x, y)$ kann man nach der Wahrscheinlichkeit

$$P(X = x,\ Y \text{ beliebig})$$

fragen, mit der X einen bestimmten Wert x annimmt, wobei es ganz gleichgültig ist, welchen Wert Y annimmt. Es ist dann

$$\boxed{f_1(x) = P(X = x,\ Y \text{ beliebig}) = \sum_y f(x, y)\,.} \tag{55.1}$$

Bei der Summation sind alle Werte von $f(x, y)$ zu berücksichtigen, die für das betreffende x nicht null sind. $f_1(x)$ bestimmt eine eindimensionale Wahrscheinlichkeitsverteilung. Diese heißt die **Randverteilung** *der Variablen X bezüglich der gegebenen zweidimensionalen Verteilung*, und $f_1(x)$ heißt die zugehörige *Wahrscheinlichkeitsfunktion*. Durch Summation erhalten wir die zugehörige Verteilungsfunktion

$$F_1(x) = P(X \leqq x,\ Y \text{ beliebig}) = \sum_{x^* \leqq x} f_1(x^*). \tag{55.2}$$

Entsprechend bestimmt die Wahrscheinlichkeitsfunktion

$$\boxed{f_2(y) = P(X \text{ beliebig},\ Y = y) = \sum_x f(x, y)} \tag{55.3}$$

die sogenannte **Randverteilung** *der Variablen Y bezüglich der gegebenen zweidimensionalen Verteilung*. Unter dem Summenzeichen stehen nun jeweils alle Werte von $f(x, y)$, die für das betreffende y nicht null sind. Die *Verteilungsfunktion* dieser Randverteilung ist

$$F_2(y) = P(X \text{ beliebig},\ Y \leqq y) = \sum_{y^* \leqq y} f_2(y^*). \tag{55.4}$$

Wie man sieht, sind die Randverteilungen einer diskreten Verteilung ebenfalls diskret.

Im Falle einer *stetigen* zweidimensionalen Verteilung verlaufen die Überlegungen ganz entsprechend. An die Stelle von Summationen treten Integrationen. Es sei $f(x, y)$ die Wahrscheinlichkeitsdichte einer solchen Verteilung. Wir betrachten das Ereignis $X \leqq x$ ohne Rücksicht auf den Wert, den Y annimmt, also das Ereignis

$$(X \leqq x,\ Y \text{ beliebig}) \quad \text{oder} \quad (X \leqq x,\ -\infty < Y < \infty).$$

Dieses hat offenbar die Wahrscheinlichkeit

$$F_1(x) = P(X \leqq x, -\infty < Y < \infty) = \int_{-\infty}^{x} \left[\int_{-\infty}^{\infty} f(x^*, y)\, dy \right] dx^*.$$

Setzen wir hierbei

$$(55.5)\qquad \boxed{f_1(x) = \int_{-\infty}^{\infty} f(x, y)\, dy,}$$

so können wir einfacher

$$(55.6)\qquad F_1(x) = \int_{-\infty}^{x} f_1(x^*)\, dx^*$$

schreiben. $f_1(x)$ heißt die *Wahrscheinlichkeitsdichte* und $F_1(x)$ die *Verteilungsfunktion* der **Randverteilung** *von X bezüglich der gegebenen stetigen zweidimensionalen Verteilung.*

Entsprechend ist

$$(55.7)\qquad \boxed{f_2(y) = \int_{-\infty}^{\infty} f(x, y)\, dx}$$

die *Wahrscheinlichkeitsdichte* und

$$(55.8)\qquad F_2(y) = \int_{-\infty}^{y} f_2(y^*)\, dy^* = \int_{-\infty}^{y} \int_{-\infty}^{\infty} f(x, y^*)\, dx\, dy^*$$

die *Verteilungsfunktion* der **Randverteilung** *von Y bezüglich der gegebenen stetigen zweidimensionalen Verteilung.*

Wie wir sehen, sind die beiden Randverteilungen einer stetigen Verteilung ebenfalls stetige Verteilungen.

Den Begriff der Randverteilung können wir uns auch mechanisch veranschaulichen, indem wir das am Ende von Abschn. 52 erwähnte Analogon heranziehen. Wir stellen uns die Verteilung der Einheitsmasse in der XY-Ebene vor, die einer gegebenen zweidimensionalen Wahrscheinlichkeitsverteilung entspricht. Nun verschieben wir jedes Massenteilchen auf kürzestem Wege, also parallel zur Y-Achse, nach der X-Achse. So erhalten wir eine eindimensionale Massenverteilung auf der X-Achse. Diese entspricht offenbar der Randverteilung von X. Den Verschiebungsvorgang bezeichnet man auch als *Orthogonalprojektion* der Masse auf die X-Achse. Entsprechend ergibt sich das Analogon der Randverteilung von Y durch Orthogonalprojektion der ursprünglichen Massenverteilung auf die Y-Achse.

Bilden wir aus den Verteilungen zweier Variablen X und Y mit den Wahrscheinlichkeitsfunktionen bzw. Dichten $f_1(x)$ und $f_2(y)$ die zweidimensionale (X, Y)-Verteilung mit der Wahrscheinlichkeitsfunktion bzw. Dichte

$$f(x, y) = f_1(x) f_2(y),$$

so sind deren Randverteilungen natürlich gerade die ursprünglichen Verteilungen, von denen wir ausgegangen sind.

Indem wir in (53.2) bzw. (54.1) den Grenzübergang $x \to \infty$ bzw. $y \to \infty$ ausführen und (55.1)—(55.4) bzw. (55.5)—(55.8) berücksichtigen, folgt

$$F_1(x) = F(x, \infty) \quad \text{und} \quad F_2(y) = F(\infty, y).$$

Entsprechend gibt es im Falle einer n-dimensionalen Zufallsvariablen $(X_1, \ldots, X_n)$ mit der Verteilungsfunktion $F(x_1, \ldots, x_n)$ die n Randverteilungen der einzelnen Zufallsvariablen $X_1, \ldots,$ bzw. X_n mit den Verteilungsfunktionen

$$F_1(x_1) = F(x_1, \infty, \ldots, \infty), \ldots \quad \text{bzw.} \quad F_n(x_n) = F(\infty, \ldots, \infty, x_n)$$

(und auch Randverteilungen von Paaren X_j, X_k usw.).

Aufgaben zu Abschnitt 55

55.1 Man stelle die Randverteilungen in Beispiel 55.1 graphisch dar.

55.2—55.4 Man bestimme die Wahrscheinlichkeitsfunktionen bzw. Dichten der Randverteilungen der folgenden Verteilungen.

55.2 Die Verteilung in Beispiel 53.1.

55.3 Die Verteilung in Beispiel 54.1.

55.4 Die Verteilung in Aufgabe 54.3.

56 Unabhängige Zufallsvariable

Die beiden Zufallsvariablen X und Y einer zweidimensionalen (X, Y)-Verteilung mit der Verteilungsfunktion $F(x, y)$ heißen **unabhängig**, wenn für alle (x, y) die Beziehung

$$F(x, y) = F_1(x)\, F_2(y) \tag{56.1}$$

gilt. Ist dies nicht der Fall, so heißen die Variablen **abhängig.** Hierbei sind F_1 und F_2 die durch (55.2), (55.4), (55.6) bzw. (55.8) gegebenen Verteilungsfunktionen der Randverteilungen von X bzw. Y.

Die Variablen seien entweder beide diskret oder beide stetig. Notwendig und hinreichend für Unabhängigkeit ist dann, daß die zugehörigen Wahrscheinlichkeitsfunktionen bzw. Dichten $f_1(x)$ und $f_2(y)$ für alle (x, y) die folgende Beziehung erfüllen:

$$f(x, y) = f_1(x)\, f_2(y). \tag{56.2}$$

In der Tat erhält man im stetigen Falle (56.2) aus (56.1) durch Differentiation und umgekehrt (56.1) aus (56.2) durch Integration. Im diskreten Falle ergibt sich unsere Aussage aus dem folgenden

Satz 56.1. *Zwei Zufallsvariable X und Y sind dann und nur dann unabhängig, wenn für jedes Paar von Ereignissen der Form*

$$a_1 < X \leqq b_1 \quad \text{und} \quad a_2 < Y \leqq b_2$$

die Beziehung

$$P(a_1 < X \leqq b_1, a_2 < Y \leqq b_2) = P(a_1 < X \leqq b_1)\, P(a_2 < Y \leqq b_2) \tag{56.3}$$

gilt, d. h. wenn diese Ereignisse unabhängig sind [*vgl.* (20.1)].

Den Beweis dieses Satzes stellen wir als Übungsaufgabe (s. Aufgabe 56.1).

Die Variablen X und Y in Beispiel 53.1 sind unabhängig, wie sich der Leser überlegen möge.

Beispiel 56.1. In einer Schachtel liegen 10 Bohrer, darunter 4 stumpfe. 2 Bohrer werden nacheinander zufällig und ohne Zurücklegen entnommen. Wir betrachten die Zufallsvariablen

$$X = \textit{Anzahl der beim 1. Zug erhaltenen stumpfen Bohrer},$$

$$Y = \textit{Anzahl der beim 2. Zug erhaltenen stumpfen Bohrer}.$$

Diese können die Werte 0 oder 1 annehmen. Da 4 unter den 10 Bohrern stumpf und 6 scharf sind, so hat die Wahrscheinlichkeitsfunktion $f(x, y)$ der zweidimensionalen Variablen (X, Y) die Werte

$$f(0, 0) = \frac{6}{10}\cdot\frac{5}{9} = \frac{1}{3}, \qquad f(0, 1) = \frac{6}{10}\cdot\frac{4}{9} = \frac{4}{15},$$

$$f(1, 0) = \frac{4}{10}\cdot\frac{6}{9} = \frac{4}{15}, \qquad f(1, 1) = \frac{4}{10}\cdot\frac{3}{9} = \frac{2}{15}.$$

Und die Wahrscheinlichkeitsfunktionen der einzelnen Variablen (also der Randverteilungen) haben die Werte

$$f_1(0) = f(0, 0) + f(0, 1) = \frac{1}{3} + \frac{4}{15} = 0{,}6$$

$$f_2(0) = f(0, 0) + f(1, 0) = \frac{1}{3} + \frac{4}{15} = 0{,}6$$

usw. Die Variablen sind abhängig, denn es ist $f(0,0) = 1/3$, aber $f_1(0)\, f_2(0) = 0{,}36$.

Die Zufallsvariablen $X_1, \ldots, X_n$ einer n-dimensionalen $(X_1, \ldots, X_n)$-Verteilung mit der Verteilungsfunktion

$$F(x_1, \ldots, x_n) = P(X_1 \leqq x_1, \ldots, X_n \leqq x_n)$$

heißen **unabhängig**, wenn für alle $(x_1, \ldots, x_n)$ die Beziehung

$$F(x_1, \ldots, x_n) = F_1(x_1)\, F_2(x_2) \cdots F_n(x_n) \tag{56.4}$$

gilt. Ist dies nicht der Fall, so heißen die Variablen **abhängig.** Hierbei ist F_j die Verteilungsfunktion der Randverteilung von X_j.

Die genannten Variablen seien entweder alle diskret oder alle stetig. Notwendig und hinreichend für Unabhängigkeit ist dann die identische Gültigkeit der Formel

$$f(x_1, \ldots, x_n) = f_1(x_1)\, f_2(x_2) \cdots f_n(x_n) \tag{56.5}$$

für die Wahrscheinlichkeitsfunktionen bzw. Dichten, die wir mit entsprechenden Kleinbuchstaben bezeichnet haben. Ebenso gilt der Satz 56.1 in sinngemäß erweiterter Form.

Aufgaben zu Abschnitt 56

56.1 Man beweise den Satz 56.1.

56.2 Man zeige, daß die Variablen X und Y in Beispiel 53.1 unabhängig sind.

56.3 In einer Schachtel liegen N Dinge, darunter M defekte. Es wird ohne Zurücklegen gezogen. Man zeige, daß die Zufallsvariablen

$$X = \textit{Anzahl defekter Dinge beim 1. Zug}$$

$$Y = \textit{Anzahl defekter Dinge beim 2. Zug}$$

abhängig sind.

57 Funktionen mehrerer Zufallsvariablen

In Abschn. 30 haben wir Funktionen $g(X)$ einer einzelnen Zufallsvariablen X betrachtet. Entsprechend kann man bei mehreren Variablen verfahren. Der Einfachheit halber behandeln wir zuerst den Fall einer zweidimensionalen Verteilung.

(X, Y) sei eine zweidimensionale Zufallsvariable. $f(x, y)$ sei deren Wahrscheinlichkeitsfunktion bzw. Dichte und $F(x, y)$ deren Verteilungsfunktion. Weiterhin sei $g(x, y)$ eine in der ganzen Ebene definierte und stetige Funktion zweier Veränderlichen. Dann ist

$$Z = g(X, Y)$$

ebenfalls eine Zufallsvariable, und wir können nach der zugehörigen Wahrscheinlichkeitsfunktion bzw. Dichte oder nach der Verteilungsfunktion fragen.

Entsprechend ist natürlich auch eine Funktion

$$Z = g(X_1, \ldots, X_n),$$

die von n Zufallsvariablen $X_1, \ldots, X_n$ abhängt, selbst eine Zufallsvariable.

Beispiel 57.1. Wir würfeln gleichzeitig mit zwei regelmäßigen unterscheidbaren Würfeln. Es sei

$$X = \textit{Mit dem 1. Würfel gewürfelte Zahl},$$

$$Y = \textit{Mit dem 2. Würfel gewürfelte Zahl}.$$

Die Verteilung dieser beiden Zufallsvariablen hat die Wahrscheinlichkeitsfunktion

$$f(x, y) = \frac{1}{36} \qquad \text{für} \quad x = 1, \ldots, 6 \text{ und } y = 1, \ldots, 6.$$

Wir betrachten nun die Zufallsvariable

$$Z = g(X, Y) = X + Y,$$

also die Summe der beiden gewürfelten Zahlen. Die zugehörige Wahrscheinlichkeitsfunktion $f(z)$ erhalten wir offenbar dadurch, daß wir für jeden möglichen Wert von z ($= 2, 3, \ldots, 12$) jeweils alle diejenigen Werte von $f(x, y)$ addieren, für die $x + y = z$ ist. In Formeln:

$$f(z) = \sum_{x+y=z}\sum f(x, y).$$

Der Leser überzeuge sich, daß diese Summation tatsächlich auf die in Beispiel 23.2 angegebenen Funktionswerte führt.

Ganz allgemein erhalten wir im diskreten Falle aus der gegebenen Wahrscheinlichkeitsfunktion $f(x, y)$ einer zweidimensionalen Zufallsvariablen (X, Y) die Wahrscheinlichkeitsfunktion $f(z)$ einer Zufallsvariablen $Z = g(X, Y)$ durch geeignete Summation:

$$f(z) = P(Z = z) = \sum_{g(x,y)=z}\sum f(x, y). \tag{57.1}$$

Bei der Summation sind also jeweils alle diejenigen Werte $f(x, y)$ zu berücksichtigen, deren zugehörige (x, y) der Bedingung

$$g(x, y) = z$$

genügen.

Aus (57.1) ergibt sich die Verteilungsfunktion $F(z)$ der Variablen $Z = g(X, Y)$ in der Form

$$F(z) = P(Z \leqq z) = \sum_{g(x,y)\leqq z}\sum f(x, y). \tag{57.2}$$

Die Summe umfaßt nun jeweils alle Werte $f(x, y)$, deren zugehörige (x, y) der Bedingung

$$g(x, y) \leqq z$$

genügen.

Ähnlich ist die Lage bei *stetigen* zweidimensionalen Verteilungen. An die Stelle der Summationen treten Integrationen. Statt (57.2) gilt dann also

$$F(z) = P(Z \leqq z) = \iint_{g(x,y)\leqq z} f(x, y)\, dx\, dy. \tag{57.3}$$

Hierbei ist für jedes z über den durch $g(x, y) \leqq z$ bestimmten Bereich in der xy-Ebene zu integrieren. Und die Wahrscheinlichkeitsdichte der Zufallsvariablen Z ist die Ableitung

$$f(z) = F'(z).$$

Aufgaben zu Abschnitt 57

57.1 Die Zufallsvariable (X, Y) sei gleichförmig verteilt im Quadrat Q: $0 < X \leq 1$, $0 < Y \leq 1$. Man bestimme die Verteilungsfunktion $F(z)$ und die Dichte $f(z)$ der Variablen $Z = X + Y$.

57.2 Man berechne die Mittelwerte und die Varianzen der beiden Randverteilungen der (X, Y)-Verteilung in Aufgabe 57.1.

57.3 Man zeige: Zu $Z = g(X, Y) = X + Y$ gehört, wenn X und Y unabhängige stetige Zufallsvariable sind, die Verteilungsfunktion

$$F(z) = \int_{-\infty}^{\infty} f_2(y) \left[\int_{-\infty}^{z-y} f_1(x)\, dx \right] dy \tag{57.4}$$

und, wenn $f_1(x)$ stetig ist, die Dichte

$$f(z) = \int_{-\infty}^{\infty} f_1(z-y) f_2(y)\, dy. \tag{57.5a}$$

57.4 Man zeige: Außer (57.5a) gilt auch

$$f(z) = \int_{-\infty}^{\infty} f_2(z-x) f_1(x)\, dx. \tag{57.5b}$$

58 Mathematische Erwartung. Mittelwert

Im Falle einer zweidimensionalen Zufallsvariablen (X, Y) mit der Wahrscheinlichkeitsfunktion bzw. Dichte $f(x, y)$ wird der *mathematische Erwartungswert* – kurz: die **Erwartung** – *einer gegebenen Funktion* $g(X, Y)$ mit $E(g(X, Y))$ bezeichnet und durch

$$E(g(X, Y)) = \begin{cases} \sum\limits_x \sum\limits_y g(x, y) f(x, y) & (\textit{diskreter Fall}) \\ \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} g(x, y) f(x, y)\, dx\, dy & (\textit{stetiger Fall}) \end{cases} \tag{58.1}$$

definiert, vorausgesetzt, daß die Doppelreihe absolut konvergiert bzw. das Integral

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |g|\, f\, dx\, dy$$

existiert.

Ähnlich wie (30.3) beweist man

$$E(ag(X, Y) + bh(X, Y)) = aE(g(X, Y)) + bE(h(X, Y)). \tag{58.2}$$

Hängt g nur von einer einzigen Variablen, etwa von X, ab, so erhalten wir aus (58.1) im diskreten Fall

$$E(g(X)) = \sum_x g(x) \sum_y f(x, y) = \sum_x g(x) f_1(x).$$

Dabei ist $f_1(x)$ die Wahrscheinlichkeitsfunktion der Randverteilung von X bezüglich der Variablen (X, Y). Entsprechend schließt man im stetigen Falle. Dies bedeutet, daß unsere obige Definition der Erwartung mit derjenigen im eindimensionalen Falle im Einklang steht.

Demnach ergibt sich aus (58.2) insbesondere

$$E(X + Y) = E(X) + E(Y). \tag{58.3}$$

Indem wir Y durch eine Summe von zwei Zufallsvariablen ersetzen, erhalten wir hieraus die entsprechende Formel für drei Zufallsvariable und durch wiederholte Anwendung dieses Prozesses schließlich den wichtigen

Satz 58.1 (Additionssatz für Mittelwerte). *Der Mittelwert einer Summe von Zufallsvariablen, deren Mittelwerte existieren, ist gleich der Summe dieser Mittelwerte,*

$$\boxed{E(X_1 + X_2 + \cdots + X_n) = E(X_1) + E(X_2) + \cdots + E(X_n).} \tag{58.4}$$

Beispiel 58.1. Es sei p die Erfolgswahrscheinlichkeit bei einem Bernoulli-Experiment. Das Experiment werde n mal ausgeführt. Wir betrachten die Zufallsvariable

$$X_j = \textit{Anzahl der Erfolge bei der } j\textit{-ten Ausführung}.$$

Die zugehörige Wahrscheinlichkeitsfunktion hat die Werte [vgl. (39.1)]

$$f_j(0) = 1 - p \quad \text{und} \quad f_j(1) = p.$$

Hieraus ergibt sich der Mittelwert

$$E(X_j) = 0 \cdot (1 - p) + 1 \cdot p = p. \tag{58.5}$$

Die Gesamtzahl der Erfolge bei n Ausführungen ist

$$Z = X_1 + X_2 + \cdots + X_n.$$

Gemäß (58.4) und (58.5) hat diese Zufallsvariable den Mittelwert

$$E(Z) = E(X_1) + E(X_2) + \cdots + E(X_n) = np,$$

in Übereinstimmung mit (40.3).

Nicht ganz so einfach ist die Lage für den Mittelwert eines *Produktes*. So ist $E(X^2)$ im allgemeinen nicht gleich $[E(X)]^2$. Zum Beispiel hat die Zufallsvariable

$$X = \textit{Mit einem regelmäßigen Würfel erzielte Zahl}$$

den Mittelwert $E(X) = 7/2$, aber es ist

$$E(X^2) = 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + \cdots + 6^2 \cdot \frac{1}{6} = \frac{91}{6} \neq \left(\frac{7}{2}\right)^2.$$

Jedoch gilt für *unabhängige* Variable X und Y, deren Mittelwerte existieren,

(58.6) $$E(XY) = E(X)\,E(Y).$$

In der Tat erhalten wir wegen (56.2) im diskreten Fall

$$E(XY) = \sum_x \sum_y xyf(x, y) = \sum_x xf_1(x) \sum_y yf_2(y),$$

und der letzte Ausdruck ist gleich $E(X)\,E(Y)$. Die hierbei vorgenommene Umordnung der Doppelreihe ist wegen der absoluten Konvergenz zulässig. Entsprechend schließt man bei einer stetigen Verteilung.

Die Formel (58.6) läßt sich auf n unabhängige Zufallsvariable ausdehnen. So erhält man den

Satz 58.2 (Multiplikationssatz für Mittelwerte). *Für n* ***unabhängige*** *Zufallsvariable $X_1, \ldots, X_n$, deren Mittelwerte existieren, gilt*

(58.7) $$\boxed{E(X_1 X_2 \cdots X_n) = E(X_1)\,E(X_2) \cdots E(X_n).}$$

Aufgaben zu Abschnitt 58

58.1 Man beweise: Die momenterzeugende Funktion einer Summe unabhängiger Zufallsvariablen ist gleich dem Produkt der momenterzeugenden Funktionen dieser Variablen.

58.2 Man beweise (58.2).

58.3 In einer Urne liegen r rote und s schwarze Kugeln. n Kugeln werden mit Zurücklegen gezogen. Man bestimme den Mittelwert der Zufallsvariablen Z = *Anzahl der roten Kugeln unter den gezogenen.*

58.4 Ändert sich der Mittelwert in Aufgabe 58.3, wenn man ohne Zurücklegen zieht?

59 Varianz

Wir wollen uns nun überlegen, welche Beziehungen zwischen den Varianzen σ_1^2 und σ_2^2 zweier Zufallsvariablen X und Y und der Varianz σ^2 der Summe

(59.1) $$Z = X + Y$$

bestehen. Gemäß (31.5) haben wir zunächst

(59.2) $$\sigma^2 = E(Z^2) - [E(Z)]^2.$$

Wegen (59.1) und (58.2) ist hierbei

(59.3) $E(Z^2) = E(X^2 + 2XY + Y^2) = E(X^2) + 2E(XY) + E(Y^2)$.

Wegen (58.3) ist in (59.2) weiterhin

$$[E(Z)]^2 = [E(X + Y)]^2 = [E(X) + E(Y)]^2.$$

Multiplizieren wir die rechte Seite aus, so folgt

$$[E(Z)]^2 = [E(X)]^2 + 2E(X)\,E(Y) + [E(Y)]^2.$$

Setzen wir dies und (59.3) in (59.2) ein, so erhalten wir

$$\begin{aligned}\sigma^2 = E(X^2) - [E(X)]^2 + E(Y^2) - [E(Y)]^2 \\ + 2[E(XY) - E(X)\,E(Y)].\end{aligned}$$

In der ersten Zeile steht rechts die Summe der Varianzen von X und Y, wie wir aus (31.5) ersehen. Setzen wir

$$\boxed{\sigma_{XY} = E(XY) - E(X)\,E(Y),} \tag{59.4}$$

so lautet unsere Formel demnach einfach

$$\boxed{\sigma^2 = \sigma_1{}^2 + \sigma_2{}^2 + 2\sigma_{XY}.} \tag{59.5}$$

Die Größe σ_{XY} heißt die **Kovarianz** der Zufallsvariablen X und Y. Auf ihre Bedeutung gehen wir erst später (in Kap. 17) ein.

Sind X und Y unabhängig, so ist die Kovarianz wegen (58.6) gleich null, und (59.5) reduziert sich auf die Form

$$\sigma^2 = \sigma_1{}^2 + \sigma_2{}^2. \tag{59.6}$$

Dieses Resultat läßt sich auf mehr als zwei Zufallsvariable ausdehnen und ergibt dann den

Satz 59.1 (Additionssatz für Varianzen). *Die Varianz einer Summe* ***unabhängiger*** *Zufallsvariablen, deren Varianzen existieren, ist gleich der Summe dieser Varianzen.*

Beispiel 59.1 In Beispiel 57.1 hat jede der beiden Zufallsvariablen den Mittelwert 7/2 und die Varianz 35/12 (vgl. Beispiel 31.1). Also hat die Summe dieser Variablen gemäß Satz 58.1 den Mittelwert 7 (vgl. Abb. 23.2) und, da die Variablen unabhängig sind, gemäß Satz 59.1 die Varianz 35/6.

Beispiel 59.2 (Bernoulli-Experiment mit variabler Erfolgswahrscheinlichkeit). Wir führen ein BERNOULLI-Experiment (s. Abschn. 39) n mal aus und betrachten die Zufallsvariable

$$X_j = \textit{Anzahl der Erfolge bei der j-ten Ausführung}.$$

Die Erfolgswahrscheinlichkeit bei dieser Ausführung bezeichnen wir mit p_j. Und wir lassen zu, daß diese Wahrscheinlichkeit von Ausführung zu Ausführung wechselt.

X_j hat den Mittelwert $E(X_j) = p_j$. Weiterhin wird

$$E(X_j^2) = p_j,$$

wie man leicht nachrechnet. Demnach hat X_j gemäß (31.5) die Varianz

$$\sigma_j^2 = E(X_j^2) - [E(X_j)]^2 = p_j - p_j^2.$$

Wegen der vorausgesetzten Unabhängigkeit der Variablen $X_1, \ldots, X_n$ hat deren Summe

$$Z = X_1 + \cdots + X_n = \textit{Anzahl der Erfolge bei } n \textit{ Ausführungen}$$

also gemäß Satz 59.1 die Varianz

$$\sigma^2 = \sum_{j=1}^{n} (p_j - p_j^2) = \sum_{j=1}^{n} p_j - \sum_{j=1}^{n} p_j^2. \tag{59.7}$$

Führen wir die „mittlere Erfolgswahrscheinlichkeit"

$$p = \frac{1}{n} \sum_{j=1}^{n} p_j = \frac{1}{n}(p_1 + p_2 + \cdots + p_n) \tag{59.8}$$

ein, so gewinnt (59.7) die Form

$$\sigma^2 = n\,p - \sum_{j=1}^{n} p_j^2.$$

Wir denken uns nun n und p festgehalten. Dann hängt σ^2 noch von $p_1, \ldots, p_n$ ab. Wie wir sehen, ist σ^2 am größten, wenn $\sum p_j^2$ am kleinsten ist. Unter der Nebenbedingung (59.8) trifft dies genau dann zu, wenn $p_1 = p_2 = \cdots = p_n (= p)$ gewählt wird. Anschaulich heißt das: Bei konstanter Erfolgswahrscheinlichkeit p sind größere Zufallsabweichungen zu erwarten als bei variabler Erfolgswahrscheinlichkeit mit dem Durchschnittswert p. Dieses Ergebnis ist überraschend.

Kapitel 11

Testverteilungen

Die in der Statistik vorkommenden Verteilungen kann man ihrem Verwendungszweck nach in zwei Klassen einteilen:

1. Verteilungen, die im Zusammenhang mit mathematischen Modellen von Zufallsexperimenten auftreten. Hierzu gehören alle bisher betrachteten Verteilungen.

2. Testverteilungen (auch Prüfverteilungen genannt). Das sind Verteilungen, die die Grundlage statistischer Tests (= Prüfverfahren) bilden. Zwei besonders wichtige derartige Verteilungen betrachten wir im vorliegenden Kapitel, die zugehörigen Tests in Teil III des Buches. (Die sogenannte F-Verteilung wird erst in Abschn. 83 behandelt.)

Übrigens gibt es Verteilungen, die gleichzeitig beiden Klassen angehören, allen voran die Normalverteilung. Das werden wir in Teil III sehen.

60 Chi-Quadrat-Verteilung. Gammafunktion

$X_1, X_2, \ldots, X_n$ seien unabhängige Zufallsvariable, deren jede eine Normalverteilung mit dem Mittelwert 0 und der Varianz 1 besitzt. Die Summe der Quadrate dieser Variablen wird allgemein mit χ^2 bezeichnet. Man setzt also

(60.1) $$\chi^2 = X_1^2 + X_2^2 + \cdots + X_n^2.$$

Diese Bezeichnung ist nicht sehr schön, aber wohl nicht mehr auszurotten. Die zugehörige Verteilung heißt die **Chi-Quadrat-Verteilung** (χ^2-Verteilung). Sie hat die Wahrscheinlichkeitsdichte

(60.2) $$\boxed{f(x) = K_n x^{(n-2)/2}\, e^{-x/2}} \qquad \text{für } x > 0$$

und $f(x) = 0$ für negative x (Herleitung in Anhang 1). Hierbei ist n eine positive ganze Zahl. n heißt die *Anzahl der Freiheitsgrade* der Verteilung. K_n ist eine Konstante (s. unten). Abb. 60.1 zeigt $f(x)$ für verschiedene n. Für $n = 1$ und 2 fallen die Kurven monoton. Für $n > 2$

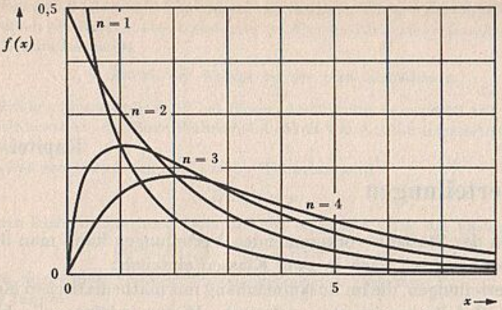


Abbildung 60.1. Dichte der Chi-Quadrat-Verteilung

haben sie ein Maximum bei $x = n - 2$. Dies ergibt sich, indem man $f'(x)$ gleich null setzt.

Aus (60.2) erhalten wir die Verteilungsfunktion

$$F(x) = K_n \int_0^x u^{(n-2)/2} e^{-u/2} du \qquad \text{für } x \geqq 0 \tag{60.3}$$

und $F(x) = 0$ für negative x.

Die Chi-Quadrat-Verteilung wurde von F. R. HELMERT eingeführt und bildet die Grundlage eines wichtigen Tests, den wir später (in Abschn. 84) betrachten.

Die Konstante K_n in (60.2) und (60.3) muß so gewählt werden, daß $F(\infty) = 1$ wird. Aus dieser Forderung ergibt sich (vgl. Aufgabe 60.4)

$$K_n = \frac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)}. \tag{60.4}$$

Hierbei ist $\Gamma(\alpha)$ die sogenannte **Gammafunktion.** Diese ist wohl die wichtigste nichtelementare Funktion. Wir müssen von ihr bei der vorliegenden Betrachtung folgendes wissen:

$\Gamma(\alpha)$ ist durch das Integral

$$\boxed{\Gamma(\alpha) = \int_0^\infty e^{-t} t^{\alpha-1} dt} \qquad (\alpha > 0) \tag{60.5}$$

definiert. Schreiben wir $\alpha + 1$ statt α und integrieren wir partiell, so ergibt sich die wichtige Beziehung

$$\Gamma(\alpha + 1) = \alpha\Gamma(\alpha). \tag{60.6}$$

Für $\alpha = 1$ können wir (60.5) integrieren:

$$\Gamma(1) = \int_0^\infty e^{-t}\, dt = 1.$$

Wegen (60.6) erhalten wir damit nacheinander

$$\Gamma(2) = 1 \cdot \Gamma(1) = 1!, \qquad \Gamma(3) = 2 \cdot \Gamma(2) = 2!$$

und allgemein

$$\boxed{\Gamma(n+1) = n!} \qquad (n = 0, 1, \ldots). \tag{60.7}$$

Die Gammafunktion ist also eine Verallgemeinerung der elementaren Fakultät.

Ist n geradzahlig, so ist demnach in (60.4)

$$\Gamma\left(\frac{n}{2}\right) = \left(\frac{n}{2} - 1\right)! \quad (n = 2, 4, \ldots).$$

Übrig bleibt der Fall ungeradzahliger n. Für $n = 1$ ist (s. Aufgabe 60.3)

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}. \tag{60.8}$$

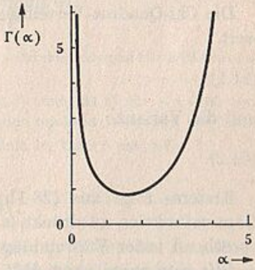


Abbildung 60.2. Gammafunktion $\Gamma(\alpha)$ für positive α

Indem wir (60.6) anwenden, erhalten wir damit der Reihe nach

$$\Gamma\left(\frac{3}{2}\right) = \frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \frac{1}{2}\sqrt{\pi}, \qquad \Gamma\left(\frac{5}{2}\right) = \frac{3}{2}\Gamma\left(\frac{3}{2}\right) = \frac{3}{4}\sqrt{\pi}$$

und so weiter.

Eine Zahlentafel der Gammafunktion findet man in Anhang 5 (Tafel 1f).

= 1,18 · Γ(1,18) = 1,089

0,9237

Aufgaben zu Abschnitt 60

60.1 Man gewinne (60.6) aus (60.5).

60.2 Vermöge (60.6) und Tafel 1f berechne man $\Gamma(2{,}18)$, $\Gamma(3{,}99)$ und $\Gamma(4{,}61)$.

= 3,89 · 2,99 · Γ(1,99) =

0,9958

60.3 Man gewinne (60.8) aus (48.8) in Aufgabe 48.1.

60.4 Die Chi-Quadrat-Verteilung ist ein Sonderfall der sogenannten **Gamma-Verteilung**, die die Dichte

$$f(x) = e^{-x} x^{\alpha-1} / \Gamma(\alpha) \qquad \text{für } x \geqq 0 \tag{60.9}$$

und $f(x) = 0$ für $x < 0$ hat. Hierbei ist $\alpha > 0$. Man zeige, daß die zugehörige Verteilungsfunktion für $x \to \infty$ gegen 1 strebt.

60.5 Die Verteilung mit der Dichte

$$f(x) = x^{\alpha-1}(1-x)^{\beta-1}/B(\alpha, \beta) \qquad (\alpha > 0,\ \beta > 0) \tag{60.10}$$

für $0 \leq x \leq 1$ und $f(x) = 0$ für die übrigen x heißt die **Beta-Verteilung**. Hierbei ist $B(\alpha, \beta) = \Gamma(\alpha)\,\Gamma(\beta)/\Gamma(\alpha+\beta)$ die sogenannte Betafunktion. Man zeichne $f(x)$ für $\alpha = 2$, $\beta = 2$.

61 Weitere Eigenschaften der Chi-Quadrat-Verteilung

Die Chi-Quadrat-Verteilung mit der Dichte (60.2) hat den Mittelwert

$$\mu = n \tag{61.1}$$

und die Varianz

$$\sigma^2 = 2n. \tag{61.2}$$

Ersteres folgt aus (28.1b), indem man das Integral durch die Gammafunktion ausdrückt (s. Aufgabe 61.1), und letzteres ganz entsprechend unter Verwendung von (31.5) und (60.6).

Wie man zeigen kann, läßt sich die χ^2-Verteilung für großes n durch die Normalverteilung brauchbar annähern. Genauer gilt:

(I) Die Zufallsvariable χ^2 ist asymptotisch normalverteilt mit dem Mittelwert (61.1) und der Varianz (61.2). Bei großem n gilt also für (60.3) [vgl. (48.4)]

$$F(x) \approx \Phi\left(\frac{x-n}{\sqrt{2n}}\right). \tag{61.3}$$

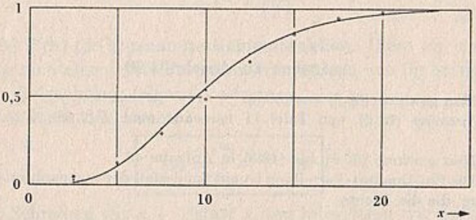


Abbildung 61.1. Kurve der Verteilungsfunktion der Chi-Quadrat-Verteilung mit $n = 10$ Freiheitsgraden und Näherungen (61.3) (kleine Kreise) und (61.4) (Punkte)

(II) Die Zufallsvariable $\sqrt{2\chi^2}$ ist asymptotisch normalverteilt mit dem Mittelwert $\sqrt{2n-1}$ und der Varianz 1. Bei großem n gilt also für (60.3)

$$F(x) \approx \Phi\left(\sqrt{2x} - \sqrt{2n-1}\right). \tag{61.4}$$

Die Näherung (61.4) ist besser als (61.3) und schon bei relativ kleinem n recht befriedigend, wie Abb. 61.1 zeigt. Aus (II) folgen auch die Näherungsausdrücke in der letzten Spalte der Tafel 6 in Anhang 5.

Aufgaben zu Abschnitt 61

61.1 Man leite (61.1) her.

61.2 Man gewinne (61.2) aus (31.5), indem man das Integral für $E(X^2)$ durch die Gammafunktion ausdrückt.

61.3 Man vergleiche die Näherungen (61.3) und (61.4) für $n = 100$ und einige in Tafel 6 vorkommende x-Werte mit den exakten Werten in der Tafel.

61.4 Wie ergeben sich die Näherungsformeln in Tafel 6 aus (61.4)?

62 t-Verteilung von Student

Die sogenannte Studentsche t-Verteilung bildet die Grundlage wichtiger Tests, die wir in Abschn. 72, 79 und 81 besprechen werden. Sie wurde von W. S. Gosset eingeführt, der unter dem Pseudonym „Student" veröffentlichte.

Die t-Verteilung ist die Verteilung der Zufallsvariablen

$$T = \frac{X}{\sqrt{Y/n}}, \tag{62.1}$$

wobei folgendes vorausgesetzt wird: n ist eine positive ganze Zahl und heißt die *Anzahl der Freiheitsgrade* der t-Verteilung. X und Y sind unabhängige Zufallsvariable. X ist normalverteilt mit dem Mittelwert 0 und der Varianz 1, und Y besitzt eine Chi-Quadrat-Verteilung mit n Freiheitsgraden.

Die Variable T wird auch oft mit t bezeichnet. Dies tun wir nicht, weil wir für Zufallsvariable einheitlich Großbuchstaben verwenden.

Die t-Verteilung hat die Wahrscheinlichkeitsdichte (s. Abb. 62.1)

$$\boxed{f(z) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \cdot \frac{1}{\left(1 + \frac{z^2}{n}\right)^{(n+1)/2}}} \tag{62.2}$$

(Herleitung in Anhang 1) und demnach die Verteilungsfunktion

(62.3) $$F(z) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \int_{-\infty}^{z} \frac{du}{\left(1 + \frac{u^2}{n}\right)^{(n+1)/2}}.$$

Für $n = 1$ erhalten wir aus (62.2) wegen (60.8) die in Aufgabe 28.6 angegebene Dichte der **Cauchy-Verteilung**, die keinen Mittelwert besitzt. Für $n = 2, 3, \ldots$ hat die t-Verteilung wegen der Symmetrie von $f(z)$ den Mittelwert

(62.4) $$\mu = 0.$$

Für $n = 1$ und $n = 2$ hat die t-Verteilung keine Varianz. Für $n = 3, 4, \ldots$ ergibt sich (s. Aufgabe 62.1)

(62.5) $$\sigma^2 = \frac{n}{n-2}.$$

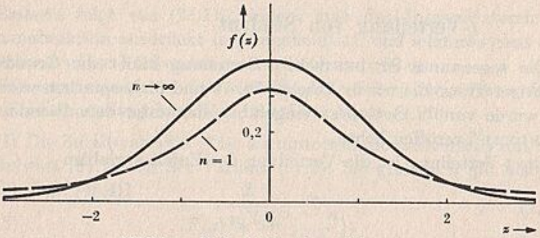


Abbildung 62.1. Dichte der t-Verteilung

Mit wachsendem n strebt die Verteilungsfunktion der t-Verteilung gegen die Verteilungsfunktion der Normalverteilung mit dem Mittelwert 0 und der Varianz 1.

Aufgaben zu Abschnitt 62

62.1 Man beweise (62.5). Anleitung: In (29.1b) verwandle man das Integral in ein Integral von 0 bis ∞, setze dann

$$1 + \frac{z^2}{n} = \frac{1}{t}$$

und benutze (60.11) in Anhang 1.

62.2 Man zeige, daß (62.3) für $z \to \infty$ gegen 1 strebt.

TEIL III

BEURTEILENDE STATISTIK

Im ersten Teil des Buches handelte es sich darum, statistische Daten in eine übersichtliche Form zu bringen. Im zweiten Teil haben wir, auf dem Wahrscheinlichkeitsbegriff aufbauend, Modellvorstellungen statistischer Gesetzmäßigkeiten entwickelt.

Im dritten Teil wollen wir nun die Verbindung zwischen der Theorie und der Wirklichkeit herstellen. Dabei interessiert uns die Frage: Welche Schlüsse kann man von einer Stichprobe auf die zugehörige Grundgesamtheit ziehen und welche Zuverlässigkeit besitzen derartige Schlüsse? Diesen Fragenkomplex kann man aufgliedern. Die Liste auf S. 165–166 gibt einen Überblick über Situationen und Fragestellungen, wie sie in der Praxis auftreten. Sie soll zugleich als Leitfaden durch die in Teil III hauptsächlich behandelten Problemkreise und Methoden dienen.

Bevor wir beginnen, müssen wir den Begriff der **Stichprobe aus einer Grundgesamtheit** präzisieren. Bisher genügte es zu wissen, daß es sich dabei um eine Auswahl von n Dingen aus einer gegebenen Gesamtheit G handelt oder, da das beobachtete Merkmal und weniger das Ding als solches interessiert, um n Zahlen (bzw. Zahlenpaare bei gleichzeitiger Beobachtung zweier Merkmale usw.).

Die Stichprobe soll Aufschluß über die genannte Gesamtheit G geben, die man aus finanziellen, zeitlichen oder prinzipiellen Gründen nicht als Ganzes untersuchen kann (Beispiele siehe Abschn. 3). Damit sich dieser Aufschluß gewinnen läßt, muß die Stichprobe eine **Zufallsauswahl** darstellen. Dies bedeutet, sie muß so entnommen werden, daß jedes Ding in G eine ganz bestimmte angebbare Wahrscheinlichkeit hat, gezogen, d. h. der Stichprobe einverleibt zu werden. Nur wenn diese Forderung (wenigstens annähernd) erfüllt ist, liefern die Verfahren, die wir nun in Teil III betrachten, vernünftige Ergebnisse.

Weiterhin machen wir die Annahme, daß die n Ausführungen des Experimentes, durch die man die Stichprobe erhält, voneinander un-

abhängig sind, d. h. daß das Ergebnis der jeweiligen Ausführung nicht davon abhängt, welche Ergebnisse bei den vorangegangenen Ausführungen erzielt wurden.

Die unter dieser Annahme entwickelte Theorie der Stichprobenentnahme heißt die *Stichprobenentnahme aus einer* **unendlichen Grundgesamtheit.** Diese Bezeichnung hat folgenden Grund: Wir können die Stichprobenwerte $x_1, \ldots, x_n$ als Ergebnisse des Ziehens aus einer Grundgesamtheit ansehen, wobei die Grundgesamtheit in ihrer Zusammensetzung ungeändert bleiben muß, damit die geforderte Unabhängigkeit gewährleistet ist.

Beim Ziehen mit Zurücklegen trifft dies zu. Denn durch das jeweilige Zurücklegen wird ja gerade der ursprüngliche Zustand wieder hergestellt, ehe man den nächsten Zug tut. So besteht z. B. bei n-maligem Würfeln die Stichprobe aus den n gewürfelten Zahlen, und die Grundgesamtheit besteht aus den beliebig vielen Zahlen (1, 2, 3, 4, 5 oder 6), die bei beliebig oftmaligem Würfeln als Ergebnisse denkbar sind. Eine solche Grundgesamtheit ist natürlich nur hypothetisch.

Beim Ziehen ohne Zurücklegen wäre die geforderte Unabhängigkeit nur dann gewährleistet, wenn die Grundgesamtheit beliebig groß („unendlich") wäre. Praktisch sieht man die Grundgesamtheit als unendlich an, wenn sie gegenüber der Stichprobe „sehr groß" ist, d. h. dann nimmt man auch beim Ziehen ohne Zurücklegen an, daß die genannte Unabhängigkeit besteht. Ein Beispiel ist die Stichprobe der (auf ganze Zentimeter abgerundeten) Größen von 1000 Personen, die aus der Einwohnerschaft einer Großstadt zufällig ausgewählt wurden.

Handelt es sich bei dem letzten Beispiel jedoch um ein Dorf, aus dessen Einwohnerschaft 1000 Personen ausgewählt werden, so kann die Grundgesamtheit nicht mehr als unendlich angesehen werden. Es liegt dann eine Stichprobenentnahme aus einer *endlichen* Grundgesamtheit vor.

Was hier für Stichproben gesagt wurde, deren Werte einzelne Zahlen sind, gilt sinngemäß auch für Stichproben von Zahlenpaaren, -tripeln usw., wie sie auftreten, wenn man gleichzeitig 2, 3 oder noch mehr Zufallsgrößen beobachtet.

Die Forderung, daß jede Stichprobe eine Zufallsauswahl darstellen soll, ist oftmals gar nicht so einfach zu erfüllen, wie man naiverweise denkt. Es gibt aber Verfahren mit dem Zweck, jede absichtliche oder unabsichtliche persönliche Beeinflussung auszuschalten und damit der genannten Forderung zu genügen. Ein solches Verfahren wollen wir jetzt erwähnen.

Will man z. B. 10 Gegenstände aus 800 gegebenen Gegenständen auswählen, so numeriere man die letzteren. Nun könnte man 800 Zettel von 1 bis 800 durchnumerieren und in eine Urne werfen. Aus dieser könnte man dann 10 Zettel ziehen und die zugehörigen Gegenstände als die gewünschte Auswahl ansehen. Statt der Urne benützen wir aber einfacher die **Hilfstafel für die Zufallsauswahl** (Tafel 5 in Anhang 5). Wir schreiben uns zuerst zwei beliebige ganze Zahlen hin, etwa 1728 und 219. Die erste teilen wir durch 100 und die zweite durch 10. Der Rest 28 bzw. 9 ergibt die Nummer der Zeile bzw. Spalte, mit der wir in Tafel 5 anfangen. Dort steht 83814. Wir schreiten von dort aus die Spalte nach unten zu fort. Dann erhalten wir der Reihe nach

83814   87802   67056   61340   usw.

Die letzten beiden Ziffern lassen wir jeweils weg, weil wir nur Zahlen bis 800 brauchen können, denn es sind ja 800 Gegenstände gegeben. Dann bleibt

838   878   670   613   usw.

Zahlen, die dann noch größer als 800 sind oder die wir zum zweiten Male erhalten, lassen wir auch noch weg. So ergeben sich schließlich die 10 Zahlen

670   613   783   528   90   658   765   242   324   600.

Die 10 Gegenstände mit diesen Nummern stellen die gewünschte Auswahl dar.

Einige statistische Grundaufgaben

Situation, Fragestellung	Wo behandelt
Die Verteilung der Grundgesamtheit enthalte unbekannte Konstanten („Parameter").	
Wie erhält man Näherungswerte für diese Konstanten?	Kap. 12
Wie erhält man Aussagen über die Güte der Näherung?	Kap. 13, insbesondere Normalverteilung Abschn. 69, 72, 74; Binomialverteilung Abschn. 75; Beliebige Verteilungen Abschn. 76
Wie kann man prüfen, ob eine solche Konstante einen gewissen Wert (z. B. den geforderten Sollwert) hat?	Kap. 14

Situation, Fragestellung	Wo behandelt
Es mögen Stichproben aus mehreren normalverteilten Grundgesamtheiten vorliegen, die sich möglicherweise durch den Mittelwert oder die Varianz unterscheiden.	
Wie prüft man, ob die Mittelwerte übereinstimmen?	Abschn. 81, Kap. 16
Wie prüft man, ob die Varianzen übereinstimmen?	Abschn. 83
Wie prüft man die Gleichheit der Mittelwerte bei beliebigen Verteilungen?	Abschn. 115
Man vermutet (z. B. aus Erfahrungs- oder theoretischen Gründen), daß eine Zufallsvariable eine gewisse Verteilungsfunktion hat. Wie prüft man eine solche Vermutung?	Kap. 15
Wie prüft man die Zufälligkeit der Reihenfolge der Stichprobenwerte?	Abschn. 117
Wie stellt man Zufallsauswahlen her?	S. 165
Wie beurteilt man die Genauigkeit physikalischer Messungen?	Kap. 19
Eine Stichprobe aus einer zweidimensionalen (X, Y)-Grundgesamtheit liege vor.	
Wie prüft man die Annahme, Y hänge linear von X ab?	Abschn. 103
Wie bestimmt man die zugehörige „Regressionsgerade“?	Abschn. 94
Wie erhält man Genauigkeitsaussagen über deren Steigungsmaß?	Abschn. 98
Wie prüft man, ob das Steigungsmaß einen gewissen Wert hat?	Abschn. 100, 102
Wie bestimmt man eine „Regressionsparabel“?	Abschn. 104
Wie prüft man, ob eine gewisse Art von Regressionskurve vorliegt?	Abschn. 105
Wie prüft man, ob die Y-Werte mit wachsendem X eine steigende (oder fallende) Tendenz („Trend“) haben?	Abschn. 100, 102, 116
Wie berechnet man den „Korrelationskoeffizienten“?	Abschn. 106
Wie erhält man Aussagen über die Güte dieser Näherung?	Abschn. 109
Wie prüft man, ob die Grundgesamtheit den Korrelationskoeffizienten null hat?	Abschn. 109

Kapitel **12**

Näherungswerte für unbekannte Konstanten

Bei praktischen Problemen steht man oft vor der Aufgabe, unter Benutzung einer Stichprobe Näherungswerte für unbekannte Konstanten zu gewinnen, die in Verteilungen auftreten (z. B. p in der Binomialverteilung, μ und σ in der Normalverteilung usw.). Die Konstanten werden als **Parameter** der Verteilung bezeichnet und die gestellte Aufgabe als die **Schätzung der Parameter.** Hiervon handelt das vorliegende Kapitel.

63 Mittelwert. Varianz. Momentenmethode

Es liegt nahe, den Mittelwert $\bar{x}$ einer Stichprobe $x_1, \ldots, x_n$ als Näherung für den Mittelwert μ der Wahrscheinlichkeitsverteilung der zugehörigen Grundgesamtheit anzusehen. So erhält man die Schätzung

$$\mu \approx \bar{x} = \frac{1}{n}(x_1 + \cdots + x_n). \tag{63.1}$$

Entsprechend kann man die Varianz s^2 der Stichprobe als Näherung für die Varianz σ^2 der genannten Verteilung auffassen. Dies ergibt

$$\sigma^2 \approx s^2 = \frac{1}{n-1} \sum_{j=1}^{n} (x_j - \bar{x})^2. \tag{63.2}$$

Bei einigen Verteilungen, z. B. bei der Normalverteilung und der Poisson-Verteilung, tritt μ explizit als Parameter auf, und (63.1) ist dann eine Parameterabschätzung. Bei der Binomialverteilung ist $p = \mu/n$ [vgl. (40.3)], und so liefert (63.1) den Näherungswert

$$p \approx \frac{\bar{x}}{n}.$$

Wir erwähnen, daß die Schätzung (63.1) ein ganz spezieller Fall der sogenannten **Momentenmethode** ist. Diese besteht darin, daß man den oder die abzuschätzenden Parameter durch die Momente der Verteilung (s. Abschn. 31) ausdrückt und in der resultierenden Formel die

Momente durch die entsprechenden Momente der Stichprobe ersetzt. Die so abgeänderte Formel ergibt dann die gesuchte Näherung. Hierbei ist das k-te Moment einer Stichprobe $x_1, \ldots, x_n$ durch

$$m_k = \frac{1}{n} \sum_{j=1}^{n} x_j^k$$

definiert.

Teilt man eine Stichprobe aus einer stetigen Grundgesamtheit in Klassen ein, so kann man der Auswirkung der Klasseneinteilung auf die Momente dadurch etwas entgegenwirken, daß man die Momente mit einer Korrektur, der sogenannten **Sheppard-Korrektur,** versieht. Für m_1 ist diese null. m_2 ist durch

$$m_2 - \frac{l^2}{12} \qquad (l = \text{Länge der Klassenintervalle})$$

zu ersetzen. Einzelheiten s. [2] (vgl. Anhang 3, am Schluß).

Bei der Poisson-Verteilung ist $\mu = \sigma^2$, und wir können demnach (63.1) *oder* (63.2) als Näherungsformel für den Parameter dieser Verteilung benutzen. Ganz allgemein hat man stets die Wahl zwischen mehreren Schätzungen. Dies ist keine Besonderheit unseres Problems, denn auch in anderen Gebieten der Mathematik gibt es für irgendeinen Zweck meist mehrere Näherungsformeln. Wir sollten uns aber überlegen, welche Gesichtspunkte im vorliegenden Falle bei der Wahl einer „guten" Näherungsformel zu beachten sind. Dies wollen wir im folgenden Abschnitt tun.

64 Schätzfunktion. Erwartungstreue. Wirksamkeit

Wir betrachten den Fall eines einzelnen unbekannten Parameters u in einer Verteilung und stellen uns vor, wir kennen eine Formel, mit deren Hilfe wir aus n Stichprobenwerten $x_1, \ldots, x_n$ einen Näherungswert $\tilde{u}$ von u berechnen können. $\tilde{u}$ hängt dann natürlich von diesen Werten ab, ist also eine Funktion dieser Werte, sagen wir $g(x_1, \ldots, x_n)$. Und unsere Formel hat demnach ganz allgemein die Gestalt

$$\tilde{u} = g(x_1, \ldots, x_n).$$

g heißt hierbei eine **Schätzfunktion** für u.

Beispiele sind (63.1) und (63.2).

Zu gegebenen Stichprobenwerten $x_1, \ldots, x_n$ gehört ein ganz bestimmter numerischer Wert von g. Diese Zahl heißt ein **Schätzwert** des Parameters u.

Nun kommt ein einfacher, aber sehr wichtiger Gedanke. *Bisher haben wir die Stichprobenwerte $x_1, \ldots, x_n$ als n beobachtete Werte einer einzelnen Zufallsvariablen X angesehen. Wir können diese n Werte aber ebensogut als einzelne beobachtete Werte von n Zufallsvariablen $X_1, \ldots, X_n$ ansehen, die alle dieselbe Verteilungsfunktion (nämlich diejenige von X) haben und unabhängig sind, weil es sich voraussetzungsgemäß um n unabhängige Ausführungen des betreffenden Experimentes handelt.*

Dann ist $\tilde{u} = g(x_1, \ldots, x_n)$ aufzufassen als ein einzelner beobachteter Wert der Zufallsvariablen

$$\tilde{U} = g(X_1, \ldots, X_n). \tag{64.1}$$

Aus Abschn. 57 wissen wir, daß eine nichtkonstante Funktion von Zufallsvariablen selbst eine Zufallsvariable ist.

Beispiel 64.1. Zwei nacheinander gewürfelte Zahlen, etwa $x_1 = 5$ und $x_2 = 3$, können wir als beobachteten Wert der Zufallsvariablen

$$X_1 = \textit{Beim 1. Wurf erzielte Zahl}$$

bzw.

$$X_2 = \textit{Beim 2. Wurf erzielte Zahl}$$

auffassen, die Summe

$$z = g(x_1, x_2) = x_1 + x_2 = 8$$

als beobachteten Wert der Zufallsvariablen

$$Z = X_1 + X_2 = \textit{Summe der Zahlen bei zweimaligem Würfeln}$$

und den Mittelwert

$$\bar{x} = \frac{1}{2}(x_1 + x_2) = 4$$

als beobachteten Wert der Zufallsvariablen

$$\bar{X} = \frac{1}{2}(X_1 + X_2).$$

Die Variable Z hat für einen regelmäßigen Würfel die in Abb. 23.2 gezeigte Wahrscheinlichkeitsfunktion. Warum?

Der in (63.1) stehende Ausdruck

$$\bar{x} = g(x_1, \ldots, x_n) = \frac{1}{n}(x_1 + \cdots + x_n)$$

läßt sich nun als ein beobachteter Wert der Zufallsvariablen

$$\bar{X} = g(X_1, \ldots, X_n) = \frac{1}{n}(X_1 + \cdots + X_n) \tag{64.2}$$

ansehen. Da X_j den Mittelwert $E(X_j) = \mu$ hat, so hat die Summe $X_1 + \cdots + X_n$ gemäß Satz 58.1 den Mittelwert $n\mu$, und $\bar{X}$ hat demnach wegen (34.2) den Mittelwert

$$E(\bar{X}) = \frac{1}{n} n\mu = \mu. \tag{64.3}$$

Eine Schätzfunktion $g(x_1, \ldots, x_n)$ für einen Parameter u heißt **erwartungstreu,** wenn

$$E(g(X_1, \ldots, X_n)) = u$$

ist.

Beispiel 64.2. Aus (64.3) ersehen wir, daß (63.1) eine erwartungstreue Schätzfunktion für den Mittelwert μ einer Verteilung ist.

Beispiel 64.3. Die Schätzfunktion (63.2) für die Varianz σ^2 ist erwartungstreu, denn die entsprechende Zufallsvariable

$$S^2 = \frac{1}{n-1} \sum_{j=1}^{n} (X_j - \bar{X})^2 \tag{64.4}$$

hat die Erwartung

$$E(S^2) = \sigma^2, \tag{64.5}$$

wie der Leser zeigen möge (s. Aufgabe 64.3). Würde man in (63.2) den Nenner $n - 1$ durch n ersetzen, so würde die Erwartungstreue zerstört. Deshalb haben wir seinerzeit in (8.3) im Nenner $n - 1$ statt n geschrieben.

Die Erwartungstreue allein ist natürlich noch kein Maß für die Güte einer Schätzfunktion $g(x_1, \ldots, x_n)$ für einen Parameter u. Vielmehr wird man weiterhin fordern, daß die Variable (64.1) mit großer Wahrscheinlichkeit Werte $\tilde{u}$ in der Nähe von u annimmt. Ein naheliegendes Maß für diese Eigenschaft ist

$$E([\tilde{U} - u]^2). \tag{64.6}$$

Je kleiner diese Zahl ist, desto besser ist die betreffende Schätzfunktion.

Ist die Schätzfunktion erwartungstreu, so hat $\tilde{U}$ den Mittelwert u, und (64.6) ist dann gerade die Varianz von $\tilde{U}$.

Eine erwartungstreue Schätzfunktion g für u heißt **wirksam,** wenn $\tilde{U} = g(X_1, \ldots, X_n)$ eine endliche Varianz hat und wenn es für u keine andere Schätzfunktion g^* derart gibt, daß $g^*(X_1, \ldots, X_n)$ eine kleinere Varianz als $\tilde{U}$ besitzt.

Aufgaben zu Abschnitt 64

64.1 Man bestimme die Werte der Wahrscheinlichkeitsfunktion der Zufallsvariablen $\bar{X}$ in Beispiel 64.1.

64.2 Warum hat Z in Beispiel 64.1 die in Abb. 23.2 gezeigte Wahrscheinlichkeitsfunktion? Wie ergibt sich aus dieser die Wahrscheinlichkeitsfunktion von $\overline{X}$ in Beispiel 64.1?

64.3 Man zeige, daß $\overline{X}$ die Varianz

$$E([\overline{X} - \mu]^2) = \frac{\sigma^2}{n} \tag{64.7}$$

hat, und beweise (64.5).

65 Konsistente Schätzfunktion

Aus (64.7) in Aufgabe 64.3 sehen wir, daß die Varianz der Zufallsvariablen (64.2) mit wachsendem n kleiner wird. Wir können also erwarten, daß die Schätzfunktion (63.1) desto bessere Näherungen liefert, je umfangreicher die Stichprobe ist.

Diesen Gedanken wollen wir noch etwas weiterverfolgen. Hierzu benötigen wir den

Satz 65.1. *Es sei c eine beliebige reelle Zahl und X eine beliebige Zufallsvariable, für die $E([X - c]^2)$ endlich ist. Dann gilt für jedes $\varepsilon > 0$ die* **Ungleichung von Tschebyscheff**

$$P(|X - c| \geqq \varepsilon) \leqq \frac{1}{\varepsilon^2} E([X - c]^2). \tag{65.1}$$

Der Beweis folgt am Ende des Abschnittes.

Ist übrigens c der Mittelwert μ von X, so steht rechts in (65.1) gerade die Varianz σ^2, und (65.1) gewinnt die Form

$$P(|X - \mu| \geqq \varepsilon) \leqq \frac{\sigma^2}{\varepsilon^2}. \tag{65.2}$$

Wir betrachten nun eine Schätzfunktion $g(x_1, \ldots, x_n)$ für einen Parameter u, die auch für beliebig große n noch definiert ist. Dann lautet (65.1) mit $c = u$ für die zugehörige Variable $\tilde{U} = g(X_1, \ldots, X_n)$ folgendermaßen:

$$P(|\tilde{U} - u| \geqq \varepsilon) \leqq \frac{1}{\varepsilon^2} E([\tilde{U} - u]^2). \tag{65.3}$$

Hat nun unsere Schätzfunktion die Eigenschaft

$$E([\tilde{U} - u]^2) \to 0 \quad \text{für} \quad n \to \infty, \tag{65.4}$$

so strebt die Wahrscheinlichkeit in (65.3) für jedes feste $\varepsilon > 0$ bei beliebig wachsendem n gegen null. Für jedes feste $\varepsilon > 0$ gilt dann also

$$P(|\tilde{U} - u| < \varepsilon) = 1 - P(|\tilde{U} - u| \geqq \varepsilon) \to 1 \quad \text{für} \quad n \to \infty.$$

Dies bedeutet in Worten: Für jede (noch so kleine) feste positive Zahl ε strebt die Wahrscheinlichkeit, daß sich die Näherung für u um weniger als ε von dem wahren Parameterwert u unterscheidet, mit wachsendem Stichprobenumfang gegen 1. Eine Schätzfunktion mit dieser Eigenschaft heißt eine **konsistente Schätzfunktion.**

Beispiel 65.1. Die Schätzfunktion (63.1) ist für jede Verteilung, deren Varianz endlich ist, konsistent. In der Tat hat (65.4) für die Zufallsvariable

$$\tilde{U} = \overline{X} = \frac{1}{n}(X_1 + \cdots + X_n)$$

wegen (64.7) in Aufgabe 64.3 die Form

$$E([\tilde{U} - u]^2) = E([\overline{X} - \mu]^2) = \frac{\sigma^2}{n} \to 0 \quad \text{für} \quad n \to \infty.$$

Wir beweisen nun den Satz 65.1.

Für eine stetige Zufallsvariable X mit der Dichte $f(x)$ ist

$$P(|X - c| \geqq \varepsilon) = \int\limits_{|x-c| \geqq \varepsilon} f(x)\, dx.$$

Integriert wird von $-\infty$ bis $c - \varepsilon$ und dann von $c + \varepsilon$ bis ∞. Im Integrationsbereich ist

$$|x - c| \geqq \varepsilon, \quad \text{also} \quad \frac{(x-c)^2}{\varepsilon^2} \geqq 1.$$

Demnach kann der Wert unseres Integrals nicht größer sein als der Wert des Integrals

$$\int\limits_{|x-c| \geqq \varepsilon} \frac{(x-c)^2}{\varepsilon^2} f(x)\, dx.$$

Da der Integrand nicht negativ ist und wir in positiver Richtung integrieren, so ist der Wert des Integrals nicht größer als der des Integrals

$$\frac{1}{\varepsilon^2} \int\limits_{-\infty}^{\infty} (x-c)^2 f(x)\, dx = \frac{1}{\varepsilon^2} E([X - c]^2),$$

das sich aus dem vorhergehenden durch Vergrößerung des Integrationsintervalles ergibt. Der Satz 65.1 ist damit für eine stetige Zufallsvariable bewiesen. Im diskreten Fall verläuft der Beweis entsprechend.

66 Wahrscheinlichkeitspapier

Im Falle der Normalverteilung kann man Näherungswerte für den Mittelwert μ und die Varianz σ^2 auch auf graphischem Wege mit Hilfe des sogenannten **Wahrscheinlichkeitspapiers** oder **Wahrscheinlichkeitsnetzes** erhalten (Hersteller Schleicher und Schüll, Einbeck, Hannover). Dies ist ein funktionales Papier, bei dem die Ordinatenskala derart verzerrt ist, daß sich die s-förmig geschwungene Kurve der Verteilungsfunktion einer Normalverteilung auf diesem Papier zu einer Geraden streckt. Der Grundgedanke ist also derselbe wie z. B. bei dem wohlbekannten einfachlogarithmischen Papier oder anderen funktionalen Papieren.

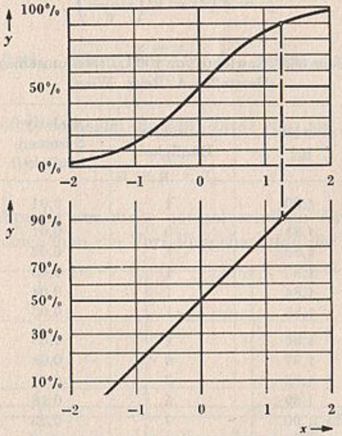


Abbildung 66.1. Zur Entstehung des Wahrscheinlichkeitsnetzes. Kurve der Verteilungsfunktion $y = F(x)$ der Normalverteilung mit $\mu = 0$, $\sigma = 1$ und zugehörige Gerade im Wahrscheinlichkeitsnetz

Abb. 66.1 veranschaulicht, wie das Wahrscheinlichkeitsnetz aus dem Millimeternetz entsteht. Wie man sieht, nehmen die Ordinatenabstände von der 50%-Linie aus nach oben und unten hin immer mehr zu, und allzu große oder allzu kleine Prozentwerte haben auf dem Papier keinen Platz mehr.

Als Ordinatenskala wird die zur Verteilungsfunktion

$$y = \Phi(w)$$

[vgl. (48.3)] inverse Funktion

$$w = \Psi(y) \tag{66.1}$$

benutzt. Da die Normalverteilung mit dem Mittelwert μ und der Varianz σ^2 gemäß (48.4) die Verteilungsfunktion

$$y = F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

Tabelle 66.1. Gewichte von 100 Luftpostumschlägen (Hersteller A. Reiss, Wien)

Gewicht [g]	Absolute Häufigkeit	Relative Summenhäufigkeit
1,80	1	0,01
1,81	0	0,01
1,82	1	0,02
1,83	1	0,03
1,84	1	0,04
1,85	1	0,05
1,86	1	0,06
1,87	2	0,08
1,88	3	0,11
1,89	5	0,16
1,90	7	0,23
1,91	6	0,29
1,92	8	0,37
1,93	8	0,45
1,94	9	0,54
1,95	4	0,58
1,96	11	0,69
1,97	3	0,72
1,98	4	0,76
1,99	3	0,79
2,00	7	0,86
2,01	2	0,88
2,02	4	0,92
2,03	5	0,97
2,04	1	0,98
2,05	2	1,00

besitzt, so ergibt sich hieraus mit (66.1)

$$w = \Psi(y) = \frac{x-\mu}{\sigma}. \tag{66.2}$$

Dies ist die Darstellung einer Geraden, wie behauptet wurde.

In (66.2) ist $w = 0$ für $x = \mu$. Wegen $\Phi(0) = 1/2$ gilt also:

Der Mittelwert μ ist die Abszisse des Schnittpunktes der Geraden (66.2) *mit der 50%-Linie.*

Weiterhin ist $\Phi(1) \approx 0{,}84$, wie wir aus Tafel 3a in Anhang 5 sehen. Also ist

$$\Phi\left(\frac{x-\mu}{\sigma}\right) \approx 0{,}84 \qquad \text{für} \qquad \frac{x-\mu}{\sigma} = 1 \qquad \text{oder} \qquad x = \sigma + \mu\,.$$

Demnach gilt:

Die Gerade (66.2) *schneidet die 84%-Linie etwa im Punkte mit der Abszisse*

$$x = \mu + \sigma.$$

Ist eine Stichprobe aus einer normalverteilten Grundgesamtheit gegeben, so zeichnet man die Verteilungsfunktion der Stichprobe als

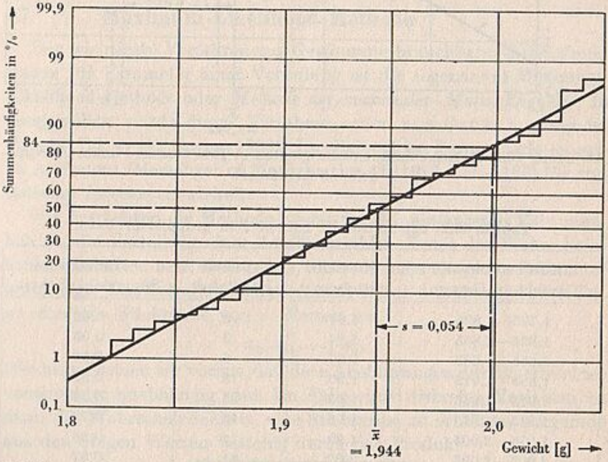


Abbildung 66.2. Verteilungsfunktion der Stichprobe in Tab. 66.1 im Wahrscheinlichkeitsnetz

Treppenkurve ins Wahrscheinlichkeitsnetz und zieht dann nach Augenmaß eine Gerade, die sich der Treppenkurve möglichst gut anpaßt. Wie aus der vorstehenden Überlegung folgt, schneidet diese Gerade die 50%-Linie im Punkt $x \approx \bar{x}$, und dies ist eine Näherung für μ. Entsprechend ist s etwa gleich der Differenz der x-Werte der Schnitt-

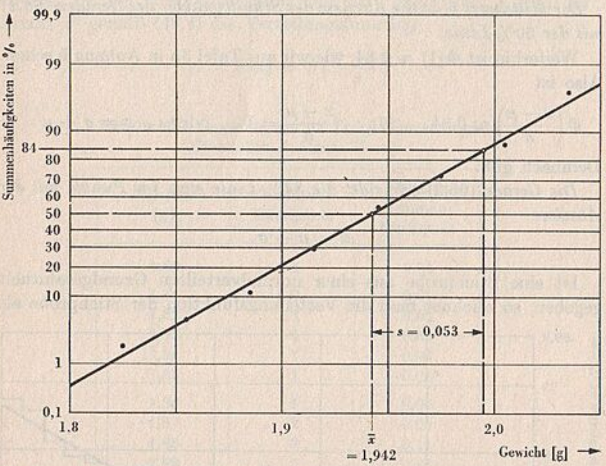


Abbildung 66.3. Relative Summenhäufigkeiten in Tab. 66.2 als Punkte im Wahrscheinlichkeitsnetz

Tabelle 66.2. Stichprobe in Tab. 66.1 in gruppierter Form

Klassenintervall [Gewicht in Gramm]	Klassenmittelpunkt	Absolute Häufigkeit	Relative Summenhäufigkeit
1,795—1,825	1,81	2	0,02
1,825—1,855	1,84	3	0,05
1,855—1,885	1,87	6	0,11
1,885—1,915	1,90	18	0,29
1,915—1,945	1,93	25	0,54
1,945—1,975	1,96	18	0,72
1,975—2,005	1,99	14	0,86
2,005—2,035	2,02	11	0,97
2,035—2,065	2,05	3	1,00

punkte der Geraden mit der 84%-Linie und der 50%-Linie, und dies ist eine Näherung für σ. Ein Beispiel zeigt die Abb. 66.2.

Ist die Stichprobe in Klassen eingeteilt, so ist es praktisch etwas einfacher, statt der Treppenkurve der Verteilungsfunktion nur die relativen Summenhäufigkeiten als Punkte vertikal über dem rechten Endpunkt des betreffenden Klassenintervalles (nicht über der Intervallmitte) ins Wahrscheinlichkeitsnetz einzutragen, nach Augenmaß eine Ausgleichsgerade durch diese Punkte zu legen und weiterzuverfahren wie zuvor. Ein Beispiel zeigt Abb. 66.3.

Aufgaben zu Abschnitt 66

66.1—66.3 Man bestimme den Mittelwert und die Varianz der folgenden Stichproben auf graphischem Wege aus dem Wahrscheinlichkeitsnetz.

66.1 Länge männlicher ausgetragener Neugeborener (Univ.-Frauenklinik Graz, Direktor Prof. Dr. E. Navratil, 1961)

Länge [cm]	46	47	48	49	50	51	52	53	54	55	56
Anzahl	1	2	11	8	15	6	9	6	1	0	1

66.2 Die Stichprobe in Aufgabe 6.1.

66.3 Die Stichprobe in Tab. 4.3.

67 Maximum-Likelihood-Methode

Das wichtigste Verfahren zur Gewinnung brauchbarer Schätzfunktionen für Parameter einer Verteilung ist die sogenannte **Maximum-Likelihood-Methode** oder *Methode der maximalen Mutmaßlichkeit*. In Sonderfällen wurde dieses Verfahren schon von Gauss angewendet (Werke Bd. 4, Göttingen 1880). In allgemeiner Form wurde es von R. A. Fisher (Messenger of Mathematics **41**, 1912, 155—160) für statistische Zwecke entwickelt.

Wir betrachten die Methode zuerst für den einfachsten Fall, nämlich für die Verteilung einer Zufallsvariablen X mit der Wahrscheinlichkeitsfunktion bzw. Dichte $f(x)$, die von einem einzelnen Parameter u abhängt. Das betreffende Experiment werde n-mal ausgeführt. Die so erhaltene Stichprobe von n Werten sei

$$x_1, x_2, \ldots, x_n.$$

Wie bisher setzen wir voraus, daß die n Ausführungen des Experimentes voneinander unabhängig sind. Im Falle einer diskreten Variablen ist dann die Wahrscheinlichkeit, eine Stichprobe zu erhalten, die gerade aus den obigen Werten besteht, durch das Produkt

$$\boxed{l = f(x_1)\, f(x_2) \cdots f(x_n)} \tag{67.1}$$

gegeben, denn $f(x_1)$ ist die Wahrscheinlichkeit, mit der X den Wert x_1 annimmt, usw. Ist X stetig verteilt, so ist ganz entsprechend die Wahrscheinlichkeit, eine Stichprobe zu erhalten, die gerade aus n Werten besteht, die in den kleinen Intervallen

$$x_1 \leqq x \leqq x_1 + \Delta x,\ x_2 \leqq x \leqq x_2 + \Delta x, \ldots, \text{ bzw. } x_n \leqq x \leqq x_n + \Delta x$$

liegen, durch den Ausdruck

(67.2) $$f(x_1)\,\Delta x\, f(x_2)\,\Delta x \cdots f(x_n)\,\Delta x = l\,(\Delta x)^n$$

gegeben.

Die Werte $f(x_1), \ldots, f(x_n)$ hängen von u ab. l hängt demnach von $x_1, \ldots, x_n$ und auch von u ab. Denken wir uns $x_1, \ldots, x_n$ fest vorgegeben, so ist l also eine Funktion des Parameters u und wird, so aufgefaßt, als **Likelihood-Funktion** oder *Mutmaßlichkeitsfunktion* bezeichnet.

Die Maximum-Likelihood-Methode besteht nun darin, daß man als Näherung für den unbekannten Parameter u einen Wert nimmt, für den l einen möglichst großen Wert besitzt. Eine notwendige Bedingung dafür, daß l, als differenzierbare Funktion von u, ein Maximum besitzt (Maxima am Rande ausgenommen), ist das Verschwinden der Ableitung nach u,

(67.3) $$\frac{\partial l}{\partial u} = 0.$$

Wir schreiben hier die partielle Ableitung, weil l auch von den Größen $x_1, \ldots, x_n$ abhängt, die wir bei der vorliegenden Betrachtung konstant halten. Eine von den $x_1, \ldots, x_n$ abhängige Lösung dieser Gleichung heißt eine **Maximum-Likelihood-Schätzfunktion** für den betreffenden Parameter u.

Ganz entsprechend erhält man im Falle einer Verteilung, die mehrere, sagen wir r Parameter $u_1, \ldots, u_r$ enthält, die r Gleichungen

(67.4) $$\frac{\partial l}{\partial u_1} = 0, \ \ldots, \ \frac{\partial l}{\partial u_r} = 0$$

zur Bestimmung von Schätzfunktionen für diese Parameter.

Da $f(x)$ nicht negativ ist, so ist l an der Stelle eines Maximums im allgemeinen positiv. Da der natürliche Logarithmus $\ln l$ eine monoton wachsende Funktion von l ist, so hat er genau dort ein Maximum, wo l ein Maximum hat. Dies legt es nahe, statt l die Funktion $\ln l$ zu verwenden, also (67.3) durch

(67.5) $$\boxed{\frac{\partial \ln l}{\partial u} = 0}$$

zu ersetzen und (67.4) entsprechend durch

$$\frac{\partial \ln l}{\partial u_1} = 0, \ \ldots, \ \frac{\partial \ln l}{\partial u_r} = 0. \tag{67.6}$$

Dann hat man statt der lästigen Differentiation von Produkten nur Summen zu differenzieren. Trotzdem kann die Lösung der erhaltenen Gleichung (bzw. Gleichungen) noch schwierig sein, so daß man sich in manchen Fällen lieber mit weniger guten aber oft einfacher zu erhaltenden Schätzfunktionen begnügen wird, wie sie etwa die Momentenmethode (s. Abschn. 63) liefert.

Gibt es für einen Parameter eine wirksame Schätzfunktion (vgl. Abschn. 64), so erhält man diese aus (67.5).

Für den Beweis dieser und anderer wichtiger Tatsachen über Maximum-Likelihood-Schätzfunktionen vgl. [12] oder [2].

68 Beispiele zur Maximum-Likelihood-Methode

Beispiel 68.1 (Poisson-Verteilung). Unter Benutzung einer Stichprobe $x_1, \ldots, x_n$ gewinne man eine Maximum-Likelihood-Schätzfunktion für den Parameter μ der durch (42.2) bestimmten Poisson-Verteilung.

Aus (42.2) und (67.1) erhalten wir die Likelihood-Funktion

$$l = \frac{\mu^{x_1}}{x_1!} e^{-\mu} \frac{\mu^{x_2}}{x_2!} e^{-\mu} \cdots \frac{\mu^{x_n}}{x_n!} e^{-\mu}.$$

Indem wir die Exponentialfaktoren und ebenso die Potenzen zusammenfassen, folgt wegen (8.1) sofort

$$l = \frac{1}{x_1! \cdots x_n!} \mu^{x_1 + \cdots + x_n} e^{-n\mu} = \frac{1}{x_1! \cdots x_n!} \mu^{n\bar{x}} e^{-n\mu}.$$

Der natürliche Logarithmus dieser Funktion ist

$$\ln l = -\ln(x_1! \cdots x_n!) + n\bar{x} \ln \mu - n\mu.$$

Also hat (67.5) im vorliegenden Falle die Form

$$\frac{\partial \ln l}{\partial \mu} = \frac{n\bar{x}}{\mu} - n = 0.$$

Durch Auflösung nach μ ergibt sich die gesuchte Schätzfunktion $\tilde{\mu}$ in der Form

$$\tilde{\mu} = \bar{x} = \frac{1}{n}(x_1 + \cdots + x_n).$$

Beispiel 68.2 (Binomialverteilung). Ein Ereignis A habe bei einem Zufallsexperiment die unbekannte Wahrscheinlichkeit p. Bei 100 Ausführungen des Experimentes sei A genau 63mal eingetroffen. Das Experiment sei derart, daß sich die Ergebnisse der einzelnen Ausführungen gegenseitig nicht beeinflussen. Man schätze p ab.

Die Zufallsvariable

$$X = \textit{Anzahl des Eintreffens von } A \textit{ beim Einzelversuch}$$

kann die Werte

$$X = 0 \quad (A \textit{ trifft nicht ein}) \qquad \text{und} \qquad X = 1 \quad (A \textit{ trifft ein})$$

annehmen. Die zugehörige Wahrscheinlichkeitsfunktion $f(x)$ hat also die Werte

$$f(0) = P(X = 0) = 1 - p \qquad \text{und} \qquad f(1) = P(X = 1) = p.$$

Da A in unserer Stichprobe von $n = 100$ Werten genau $k = 63$mal vorkommt und $n - k = 37$mal nicht, so ist die Likelihood-Funktion

$$l = p^k (1 - p)^{n-k}.$$

Bilden wir den Logarithmus, so erhalten wir

$$\ln l = k \ln p + (n - k) \ln (1 - p).$$

Die Gleichung (67.5) lautet also

$$\frac{\partial \ln l}{\partial p} = \frac{k}{p} - \frac{n-k}{1-p} = 0.$$

Die Auflösung nach p ergibt die gesuchte Schätzfunktion

$$\tilde{p} = \frac{k}{n}.$$

Dies ist gerade die relative Häufigkeit von A in unserer Versuchsreihe. Für unser Zahlenbeispiel wird also $p \approx \tilde{p} = 0{,}63$.

Beispiel 68.3 (Normalverteilung). Man wende die Maximum-Likelihood-Methode auf den Fall einer normalverteilten Variablen X an, deren Mittelwert μ und Varianz σ^2 beide unbekannt sind.

X hat die Dichte (47.1). Gemäß (67.1) hat also die Likelihood-Funktion die Form

$$l = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x_1-\mu)^2}{2\sigma^2}} \cdots \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x_n-\mu)^2}{2\sigma^2}}$$

Durch Zusammenfassen ergibt sich

$$l = \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{1}{\sigma}\right)^n e^{-h}.$$

Hierbei haben wir zur Abkürzung

$$h = \frac{1}{2\sigma^2} \sum_{k=1}^{n} (x_k - \mu)^2 \tag{68.1}$$

gesetzt. Nehmen wir den Logarithmus, so folgt

$$\ln l = -n \ln \sqrt{2\pi} - n \ln \sigma - h. \tag{68.2}$$

Im vorliegenden Fall sind zwei Parameter abzuschätzen. Die beiden zugehörigen Gleichungen (67.6) gewinnen wir durch Differentiation aus (68.2). Indem wir (68.1) berücksichtigen, ergibt sich

$$\frac{\partial \ln l}{\partial \mu} = -\frac{\partial h}{\partial \mu} = \frac{1}{\sigma^2} \sum_{k=1}^{n} (x_k - \mu) = 0 \tag{68.3}$$

und

$$\frac{\partial \ln l}{\partial \sigma} = -\frac{n}{\sigma} - \frac{\partial h}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{k=1}^{n} (x_k - \mu)^2 = 0. \tag{68.4}$$

Aus (68.3) folgt die Gleichung

$$\sum_{k=1}^{n} (x_k - \mu) = \sum_{k=1}^{n} x_k - n\mu = 0.$$

Die Lösung lautet

$$\tilde{\mu} = \frac{1}{n} \sum_{k=1}^{n} x_k = \bar{x}. \tag{68.5}$$

Dies ist die gesuchte Schätzfunktion für den Parameter μ. Setzen wir diese in (68.4) ein, so ergibt sich sofort

$$-\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{k=1}^{n} (x_k - \bar{x})^2 = 0.$$

Durch Auflösen nach σ^2 erhalten wir als Schätzfunktion für σ^2 den Ausdruck

$$\tilde{\sigma}^2 = \frac{1}{n} \sum_{k=1}^{n} (x_k - \bar{x})^2. \tag{68.6}$$

Diese Funktion ist nicht erwartungstreu. Man erkennt das am einfachsten daraus, daß sie sich um einen konstanten Faktor von der erwartungstreuen Schätzung (63.2) unterscheidet (vgl. Beispiel 64.3).

Unser Beispiel lehrt, daß eine Maximum-Likelihood-Schätzfunktion nicht erwartungstreu zu sein braucht. Daß unser Ergebnis trotzdem praktisch wertvoll ist, insbesondere bei großem n, versteht sich beinahe von selbst.

Aufgaben zu Abschnitt 68

68.1 Man überzeuge sich, daß man im Falle der Parameter a und b der gleichförmigen Verteilung (26.4) das Maximum der Likelihood-Funktion nicht durch Nullsetzen der Ableitung erhalten kann. Welcher Weg führt zum Ziel?

68.2 Man bestimme die Maximum-Likelihood-Schätzung für den Parameter μ einer Normalverteilung mit bekannter Varianz $\sigma^2 = \sigma_0^2$.

68.3 Man beweise: Ist $\tilde{u}$ die Maximum-Likelihood-Schätzung für einen Parameter u und ist $\varphi(u)$ eine einwertige monoton wachsende Funktion von u, so ist $\varphi(\tilde{u})$ die Maximum-Likelihood-Schätzung für $\varphi(u)$.

68.4 Das 4. zentrale Moment der Normalverteilung ist gleich $3\sigma^4$. Welche zugehörige Maximum-Likelihood-Schätzung ergibt sich damit wegen Aufgabe 68.3 für das genannte Moment?

Kapitel **13**

Konfidenzintervalle

Im vorigen Kapitel haben wir Näherungsformeln für Parameter in Wahrscheinlichkeitsverteilungen kennengelernt. Zum Beispiel hat die Stichprobe in Tab. 66.1 den Mittelwert $\bar{x} = 1{,}944$, wie man durch Nachrechnung bestätigt, und dies ist eine Näherung für den Mittelwert μ der zugehörigen Grundgesamtheit.

Nun erhebt sich die Frage, wie genau eine derartige Näherung ist. Denn wo immer man in der Mathematik Näherungswerte benutzt, sollte man festzustellen versuchen, wie weit ein solcher Wert ungünstigstenfalls von dem unbekannten wahren Wert abweichen kann. Zum Beispiel gibt es bei der numerischen Integration „Fehlerformeln" zur Berechnung der maximal möglichen Abweichung. Nehmen wir an, in einem konkreten Fall habe sich 2,46 als Näherung für ein Integral und $\pm 0{,}02$ als zugehörige maximal mögliche Abweichung ergeben. Dann wissen wir, daß die Größen

$$2{,}46 - 0{,}02 = 2{,}44 \qquad \text{und} \qquad 2{,}46 + 0{,}02 = 2{,}48$$

mit Sicherheit den unbekannten genauen Wert u unseres Integrals „*einschließen*", d. h. mit Sicherheit ist 2,44 kleiner als u oder gleich u und 2,48 größer als u oder gleich u.

Das entsprechende Problem bei unserem obigen statistischen Beispiel wäre die Bestimmung zweier von den Stichprobenwerten abhängiger Größen, etwa 1,940 und 1,955 oder dergleichen, die den unbekannten Mittelwert mit Sicherheit einschließen. Sichere Schlüsse von einer Stichprobe auf die Grundgesamtheit gibt es aber nicht. Deshalb müssen wir bescheiden sein und unser Problem, das wir nun gleich für irgendeinen Parameter u formulieren, modifizieren:

Man bestimme zwei Größen $\tilde{U}_1$ und $\tilde{U}_2$, die mit einer nach Belieben gewählten großen Wahrscheinlichkeit (z. B. 95% oder 99%) Werte annehmen, die den genauen unbekannten Parameterwert u einschließen.

Diese Werte seien dabei jeweils aus gegebenen Stichprobenwerten $x_1, \ldots, x_n$ zu errechnen. Letztere können wir wie in Abschn. 64 als einzelne beobachtete Werte von n Zufallsvariablen $X_1, \ldots, X_n$ ansehen. $\tilde{U}_1$ und $\tilde{U}_2$ sind dann Funktionen dieser Variablen, also selbst Zufallsvariable. Bezeichnen wir die genannte Wahrscheinlichkeit mit γ, so sollen also $\tilde{U}_1$ und $\tilde{U}_2$ derart beschaffen sein, daß

$$P(\tilde{U}_1 \leqq u \leqq \tilde{U}_2) = \gamma$$

ist. Kennen wir die Formeln solcher $\tilde{U}_1$ und $\tilde{U}_2$ als Funktionen von $X_1, \ldots, X_n$, so können wir nach Vorgabe einer Stichprobe den zugehörigen Wert von $\tilde{U}_1$ und $\tilde{U}_2$ ausrechnen. Das Intervall mit diesen beiden Werten als Endpunkten heißt ein **Konfidenzintervall** (auch **Vertrauensintervall** oder **Vertrauensbereich**) für den unbekannten Parameterwert u. Die Endpunkte heißen *Konfidenzgrenzen*. Die Zahl γ heißt die zugehörige **Konfidenzzahl.** Praktisch wählt man γ gleich 95% oder 99%, manchmal auch 99,9%.

γ ist also die Wahrscheinlichkeit, ein Konfidenzintervall zu erhalten, das den wahren unbekannten Parameterwert enthält, wenn man vorhat, eine Stichprobe zu entnehmen und dann ein Konfidenzintervall zu bestimmen.

Wird z. B. $\gamma = 95\%$ gewählt, so kann man erwarten, daß *etwa* bei 95% aller Stichproben, die man (wirklich oder nur in Gedanken) entnehmen will, die zugehörigen Konfidenzintervalle den Wert u enthalten und *etwa* 5% nicht. Hat man solche Intervalle bestimmt, so ist also die Aussage, daß ein derartiges Intervall den Wert u enthält, in etwa 19 von 20 Fällen richtig und in etwa einem von 20 Fällen falsch. Bei $\gamma = 0{,}99$ ist diese Aussage sogar in etwa 99 von 100 Fällen richtig, aber die Konfidenzintervalle sind dann jeweils länger, wie wir noch sehen werden.

Die Frage, welchen Wert γ man im konkreten Falle wählen soll, ist also keine mathematische Frage, sondern muß von der Art der Anwendung her beantwortet werden. Man muß sich nämlich überlegen, welches Risiko einer falschen Aussage der obengenannten Art man ohne Schaden in Kauf nehmen kann.

Den Begriff des Konfidenzintervalles hat J. Neyman (Annals of Math. Statistics 6, 1935, 111–116) geschaffen.

69 Konfidenzintervalle für den Mittelwert einer Normalverteilung mit bekannter Varianz

Vorgelegt sei eine Stichprobe $x_1, \ldots, x_n$ aus einer normalverteilten Grundgesamtheit mit bekannter Varianz σ^2. Der Mittelwert μ sei unbekannt, und wir stellen uns die Aufgabe, ein Konfidenzintervall für μ zu bestimmen.

Diese Situation und Fragestellung hat praktische Bedeutung. Handelt es sich z. B. um eine Abmessung irgendeines Massenartikels (Länge eines Bolzens, Stärke einer Beilagscheibe usw.), so kann es nämlich sein, daß man μ nicht kennt, weil μ von der jeweiligen Einstellung des Automaten abhängt, auf dem der Artikel hergestellt wird, daß man aber σ aus früherer Erfahrung kennt, weil σ in erster Linie von der Güte der genannten Maschine und fast nicht von deren spezieller Einstellung abhängt. Wir kommen auf diesen Punkt in Abschn. 72 zurück.

Wie man das Konfidenzintervall bestimmt, ersieht man aus Tab. 69.1. Warum man so vorgeht, überlegen wir uns im nächsten Abschnitt.

Tabelle 69.1. Bestimmung eines Konfidenzintervalles für den Mittelwert μ einer Normalverteilung mit bekannter Varianz σ^2

1. Schritt. Man wähle eine Konfidenzzahl γ (95%, 99% oder dgl.).

2. Schritt. Man bestimme das zugehörige c:

γ	0,90	0,95	0,99	0,999
c	1,645	1,960	2,576	3,291

3. Schritt. Man berechne den Mittelwert $\bar{x}$ der Stichprobe $x_1, x_2, \ldots, x_n$.

4. Schritt. Man berechne

(69.1) $$a = c\sigma/\sqrt{n}\,.$$

Das Konfidenzintervall für den Mittelwert μ der Grundgesamtheit ist

(69.2) $$\mathrm{KONF}\{\bar{x} - a \leqq \mu \leqq \bar{x} + a\}\,.$$

Beispiel 69.1. Man bestimme ein 95%-Konfidenzintervall für den Mittelwert einer Normalverteilung mit der Varianz $\sigma^2 = 9$ unter Benutzung einer Stichprobe mit dem Mittelwert $\bar{x} = 5$ und dem Umfang $n = 100$.

1. Schritt. $\gamma = 0{,}95$ ist gefordert.

2. Schritt. Dazu gehört $c = 1{,}960$.

3. Schritt. $\bar{x} = 5$ ist bereits vorgegeben.

4. Schritt. Es ist

$$a = 1{,}960 \cdot 3/\sqrt{100} = 0{,}588.$$

So wird $\bar{x} - a = 4{,}412$, $\bar{x} + a = 5{,}588$, und wir erhalten das Konfidenzintervall

$$\mathrm{KONF}\{4{,}412 \leqq \mu \leqq 5{,}588\}.$$

Manchmal schreibt man dies auch in der Form

$$\mu = 5 \pm 0{,}588.$$

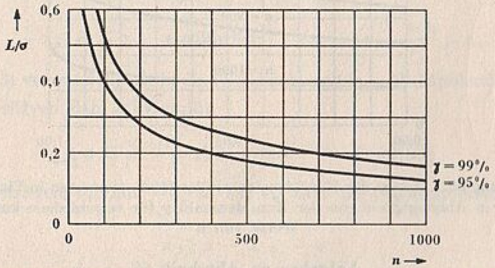


Abbildung 69.1. Länge des Konfidenzintervalles (69.2) (gemessen in Vielfachen von σ) in Abhängigkeit vom Stichprobenumfang n

Beispiel 69.2 (Bestimmung des Stichprobenumfanges zu gewünschter Länge des Konfidenzintervalles). Wie groß muß im vorigen Beispiel der Stichprobenumfang n genommen werden, damit man ein 95%-Konfidenzintervall mit der Länge $L = 0{,}4$ erhält?

Das Konfidenzintervall in Tab. 69.1 hat die Länge

$$L = 2a = 2c\sigma/\sqrt{n}\,. \tag{69.3}$$

Durch Quadrieren und Auflösen nach n erhalten wir

$$n = (2c\sigma/L)^2.$$

Die Antwort lautet demnach im vorliegenden Fall

$$n = (2 \cdot 1{,}960 \cdot 3/0{,}4)^2 \approx 870.$$

Allgemein nimmt L mit wachsendem n ab (s. Abb. 69.1). Je kürzere Intervalle man haben will, desto größer muß man die Stichprobe wählen. Um L zu halbieren, muß man n vervierfachen.

Abb. 69.2 zeigt, wie L bei konstantem n zunimmt, wenn man die Konfidenzzahl γ erhöht. Auch dies ist verständlich. Für $\gamma \to 1$ muß ja $L \to \infty$ gehen.

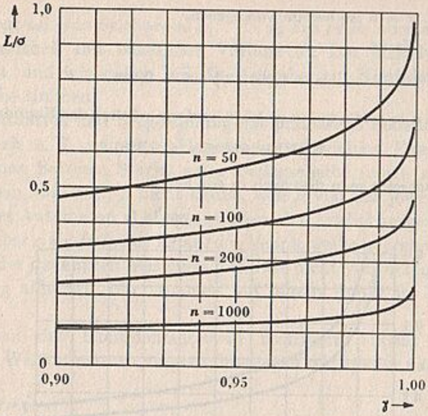


Abbildung 69.2. Länge des Konfidenzintervalles (69.2) (gemessen in Vielfachen von σ) in Abhängigkeit von der Konfidenzzahl γ für verschiedene konstante Werte von n

Aufgaben zu Abschnitt 69

69.1 Welches 95%-Konfidenzintervall erhielte man in Beispiel 69.1, wenn (a) $n = 10$, (b) $n = 1000$ wäre?

69.2—69.3 Man bestimme ein 99%-Konfidenzintervall für den Mittelwert einer Normalverteilung mit der Varianz σ^2 unter Benutzung einer Stichprobe mit dem Mittelwert $\bar{x}$ und dem Umfange n. Hierbei sei:

69.2 $\bar{x} = 18{,}4$, $\sigma^2 = 1{,}69$, $n = 36$.

69.3 $\bar{x} = 7{,}26$, $\sigma^2 = 5{,}76$, $n = 225$.

69.4 Wie groß muß in Beispiel 69.1 der Stichprobenumfang sein, damit man ein 99%-Konfidenzintervall mit der Länge $L = 0{,}4$ erhält?

69.5 Frage wie in Aufgabe 69.4, mit $L = 0{,}2$ statt 0,4.

70 Theorie zu Abschnitt 69

Satz 70.1. *$X_1, \ldots, X_n$ seien unabhängige normalverteilte Zufallsvariable mit demselben Mittelwert μ und derselben Varianz σ^2. Dann ist die Zufallsvariable*

$$Z = \sqrt{n}\,\frac{\overline{X} - \mu}{\sigma} \tag{70.1}$$

normalverteilt mit dem Mittelwert 0 und der Varianz 1. Hierbei ist

$$\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n). \tag{70.2}$$

Diesen Satz beweisen wir im nächsten Abschnitt.

Wir wählen eine Zahl γ zwischen 0 und 1. Dazu bestimmen wir für die normalverteilte Variable (70.1) aus Tafel 3b in Anhang 5 die Zahl c derart, daß

$$P(-c \leqq Z \leqq c) = \gamma \tag{70.3}$$

ist. (Für $\gamma = 0{,}90$ usw. ergeben sich die in Tab. 69.1 befindlichen Zahlen c.) Die Ungleichung

$$-c \leqq Z \leqq c, \quad \text{ausführlich} \quad -c \leqq \sqrt{n}\,\frac{\bar{X} - \mu}{\sigma} \leqq c,$$

in (70.3) verwandeln wir in eine Ungleichung für μ. Multiplikation mit der positiven Zahl $\sigma/\sqrt{n}$ ergibt

$$-a \leqq \bar{X} - \mu \leqq a \quad \text{mit} \quad a = c\sigma/\sqrt{n}\,. \tag{70.4}$$

Multiplizieren wir mit -1, so kehren sich die Ungleichheitszeichen um:

$$a \geqq \mu - \bar{X} \geqq -a.$$

Addieren wir $\bar{X}$, so folgt schließlich

$$\bar{X} + a \geqq \mu \geqq \bar{X} - a. \tag{70.5}$$

Also können wir (70.3) in der folgenden Form schreiben:

$$P(\bar{X} - a \leqq \mu \leqq \bar{X} + a) = \gamma.$$

In Worten: Mit der gewählten Wahrscheinlichkeit γ nehmen die Zufallsvariablen $\bar{X} - a$ und $\bar{X} + a$ [mit a gemäß (70.4) und dem aus (70.3) bestimmten c] Werte an, die den Mittelwert μ einschließen.

Liegt nun eine Stichprobe $x_1, \ldots, x_n$ aus einer normalverteilten Grundgesamtheit vor, so können wir die n Stichprobenwerte als einzelne beobachtete Werte der Variablen $X_1, \ldots, X_n$ in Satz 70.1 auffassen. Dann ist der Mittelwert $\bar{x}$ der Stichprobe ein beobachteter Wert der Variablen $\bar{X}$. Setzen wir $\bar{x}$ in (70.5) ein, so ergibt sich gerade das Konfidenzintervall (69.2).

Aufgaben zu Abschnitt 70

70.1 Das Konfidenzintervall (69.2) kann man als Sonderfall eines allgemeineren Konfidenzintervalles ansehen, das man erhält, wenn man (70.3) durch

$$P(-c_1 \leqq Z \leqq c_2) = \gamma \tag{70.6}$$

ersetzt. Man zeige, daß dieses allgemeinere Konfidenzintervall die Form

$$\mathrm{KONF}\left\{\bar{x} - c_2 \frac{\sigma}{\sqrt{n}} \leqq \mu \leqq \bar{x} + c_1 \frac{\sigma}{\sqrt{n}}\right\} \tag{70.7}$$

hat. (Hierbei wird c_1 gewählt und dazu c_2 vermöge Tafel 3b in Anhang 5 bestimmt.)

70.2 Warum erhält man in (70.6) kein c_2, wenn man c_1 zu klein wählt? Welches ist das kleinstmögliche c_1 im Falle $\gamma = 0{,}99$?

70.3 Unter Benutzung der Stichprobe in Beispiel 69.1 bestimme man ein Konfidenzintervall (70.7) mit $\gamma = 0{,}95$ und $c_1 = 2{,}5$.

70.4 Unter Benutzung der Stichprobe in Beispiel 69.1 bestimme man ein 99%-Konfidenzintervall (70.7) mit dem rechten Endpunkt 7,7.

70.5 Man beweise: Unter allen Konfidenzintervallen (70.7) mit demselben γ, σ und n ist (69.2) am kürzesten.

70.6 Man zeige, daß (70.3) gleichbedeutend mit $\Phi(c) = (1 + \gamma)/2$ ist.

71 Summe normaler Zufallsvariablen. Lineare Transformation

Wir müssen nun noch den Satz 70.1 beweisen. Hierzu beweisen wir zuerst den wichtigen

Satz 71.1 (Summe unabhängiger normalverteilter Variablen). *$X_1, X_2, \ldots, X_n$ seien unabhängige normalverteilte Zufallsvariable mit den Mittelwerten $\mu_1, \mu_2, \ldots$ bzw. μ_n und den Varianzen $\sigma_1^2, \sigma_2^2, \ldots$ bzw. σ_n^2. Dann ist auch die Zufallsvariable*

$$X = X_1 + X_2 + \cdots + X_n$$

normalverteilt und hat den Mittelwert

$$\mu = \mu_1 + \mu_2 + \cdots + \mu_n$$

und die Varianz

$$\sigma^2 = \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2.$$

Beweis. Die Formeln für μ und σ^2 folgen unmittelbar aus den Sätzen 58.1 und 59.1. Zu zeigen bleibt noch, daß X normalverteilt ist. Dies beweisen wir induktiv, wie folgt.

Wir betrachten $n = 2$, also $X = X_1 + X_2$. Nach Voraussetzung hat X_1 die Dichte

$$f_1(x) = \frac{1}{\sqrt{2\pi}\,\sigma_1} e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2}.$$

Entsprechend hat X_2 die Dichte

$$f_2(x) = \frac{1}{\sqrt{2\pi}\,\sigma_2} e^{-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2}.$$

Also hat $X = X_1 + X_2$ gemäß (57.5a) die Dichte

$$f(x) = \int_{-\infty}^{\infty} f_1(x-y)\,f_2(y)\,dy$$

$$= \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{(x-y-\mu_1)^2}{\sigma_1^2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right)} dy.$$

Den Ausdruck in der großen runden Klammer bezeichnen wir mit V. Wie man durch Nachrechnen bestätigt, läßt sich V in der Form $V = V_1^2 + V_2^2$ schreiben, wobei

$$(71.1) \qquad V_1 = \frac{\sigma}{\sigma_1\sigma_2}\left(y - \frac{\sigma_1^2\mu_2 + \sigma_2^2(x-\mu_1)}{\sigma^2}\right) \quad \text{und} \quad V_2 = \frac{x-\mu}{\sigma}$$

ist; hierbei bedeutet natürlich

$$\mu = \mu_1 + \mu_2 \quad \text{und} \quad \sigma^2 = \sigma_1^2 + \sigma_2^2.$$

V_2^2 hängt nicht von y ab, kann also vor das Integral gezogen werden. So erhalten wir

$$f(x) = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{1}{2}V_2^2} \int_{-\infty}^{\infty} e^{-\frac{1}{2}V_1^2}\,dy.$$

Nun liegt es nahe, V_1 als neue Integrationsvariable einzuführen. Schreiben wir $\tau = V_1$, so folgt aus (71.1) unmittelbar $d\tau/dy = \sigma/\sigma_1\sigma_2$, also $dy = (\sigma_1\sigma_2/\sigma)\,d\tau$. Demnach ergibt sich

$$f(x) = \frac{1}{2\pi\sigma} e^{-\frac{1}{2}V_2^2} \int_{-\infty}^{\infty} e^{-\tau^2/2}\,d\tau.$$

Gemäß (48.3) und (48.8) hat das verbleibende Integral den Wert $\sqrt{2\pi}$. Indem wir V_2 wieder ausführlich schreiben, lautet unser Ergebnis also:

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

Für $n = 2$ ist X demnach normalverteilt.

Wir machen nun die Induktionsvoraussetzung, X sei für eine beliebige ganze Zahl $n = k\ (\geqq 2)$ normalverteilt, d. h. $Y_1 = X_1 + \cdots + X_k$ sei normalverteilt. Dann folgt aus dem bisher Bewiesenen, daß auch $X = Y_1 + Y_2$ mit $Y_2 = X_{k+1}$ normalverteilt ist. Damit ist der Satz 71.1 bewiesen.

Satz 71.2 (Lineare Skalentransformation). *Ist X normalverteilt mit dem Mittelwert μ und der Varianz σ^2, so ist*

$$X^* = c_1 X + c_2 \qquad (c_1, c_2 \textit{ konstant}, \; c_1 \neq 0)$$

normalverteilt mit dem Mittelwert

$$\mu^* = c_1 \mu + c_2 \tag{71.2}$$

und der Varianz

$$\sigma^{*2} = c_1^2 \sigma^2. \tag{71.3}$$

Beweis. X^* hat die Verteilungsfunktion

$$F(x) = P(X^* \leqq x) = P(c_1 X + c_2 \leqq x) = P\left(X \leqq \frac{x - c_2}{c_1}\right)$$
$$= \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{(x-c_2)/c_1} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt.$$

Wir setzen $t = (\tau - c_2)/c_1$. Wie man nachrechnet, wird dann

$$\frac{dt}{\sigma} = \frac{d\tau}{c_1 \sigma} = \frac{d\tau}{\sigma^*}, \qquad \frac{t-\mu}{\sigma} = \frac{\tau - \mu^*}{\sigma^*}$$

mit μ^* und σ^* gemäß (71.2) und (71.3). Also ergibt sich

$$F(x) = \frac{1}{\sqrt{2\pi}\,\sigma^*} \int_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{\tau-\mu^*}{\sigma^*}\right)^2} d\tau,$$

und der Satz 71.2 ist bewiesen. [(71.2) und (71.3) folgen natürlich auch aus (34.2) und (34.4).]

Satz 71.3. *Sind $X_1, \ldots, X_n$ unabhängige normalverteilte Zufallsvariable, deren jede den Mittelwert μ und die Varianz σ^2 besitzt, so ist*

$$\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n)$$

normalverteilt mit dem Mittelwert μ und der Varianz σ^2/n.

Dieser Satz folgt unmittelbar aus Satz 71.1 und 71.2.

Weiterhin folgt aus Satz 71.2 und 71.3 unmittelbar der Satz 70.1, den wir beweisen wollten.

Aufgaben zu Abschnitt 71

71.1 Welche Verteilung haben $2X$, $-X$ und $3X - 12$, wenn X normalverteilt ist mit dem Mittelwert 4 und der Varianz 25?

71.2 Welche Verteilung haben $X_1 + X_2$ und $X_1 - X_2$, wenn X_1 und X_2 normalverteilt sind mit dem Mittelwert 6 bzw. -2 und der Varianz 1 bzw. 16?

71.3 Eine Maschine packt X Gramm pulverförmiges Desinfektionsmittel in Y Gramm schwere Schachteln. X und Y seien dabei normalverteilt mit dem Mittelwert 100 g bzw. 5 g und der Standardabweichung 1 g bzw. 0,5 g. Mit wieviel Prozent fertiger Packungen zwischen 104 g und 106 g ist dann zu rechnen?

71.4 Welches Gewicht hat ein Karton von 100 der Packungen in Aufgabe 71.3, wenn das Leergewicht des Kartons normalverteilt ist mit dem Mittelwert 500 g und der Standardabweichung 10 g?

71.5 Zum Schutz gegen Überspannungen befinden sich in einem Stromkreis zwei Relais R_1 und R_2 mit der Schaltzeit X_1 (= Zeit vom Beginn des Auftretens der Überspannung bis zur Ausführung des Schaltvorganges) bzw. X_2. Die Schaltzeiten seien unabhängig und normalverteilt mit dem Mittelwert $\mu_1 = 1$ sec bzw. μ_2 („mittlere Schaltzeiten") und der Varianz $\sigma_1^2 = \sigma_2^2 = 0{,}1\ \mathrm{sec}^2$. Wie groß muß μ_2 gewählt werden, damit μ_2 zwar möglichst klein ist, aber R_2 unter 1000 Fällen höchstens einmal eher als R_1 und sonst immer später schaltet?

71.6 Welche Dicke hat ein Transformatorkern aus 50 Blechen und 49 Papierzwischenlagen, wenn die Dicke X bzw. Y des einzelnen Bleches bzw. Papiers normalverteilt ist mit dem Mittelwert 0,5 mm bzw. 0,05 mm und der Standardabweichung 0,05 mm bzw. 0,02 mm?

72 Konfidenzintervalle für den Mittelwert einer Normalverteilung mit unbekannter Varianz

Die in der Überschrift genannten Konfidenzintervalle bestimmt man aus einer gegebenen Stichprobe gemäß Tab. 72.1. Der Unterschied gegenüber Abschn. 69 ist der, daß die Varianz σ^2 der Grundgesamtheit nicht mehr als bekannt vorausgesetzt wird. Die theoretische Begründung des Verfahrens folgt im nächsten Abschnitt.

Beispiel 72.1. Man bestimme ein 99%-Konfidenzintervall für den Mittelwert μ der Grundgesamtheit, der die Stichprobe in Tab. 66.1 entstammt. Dabei setze man voraus, die Grundgesamtheit sei normalverteilt. (Diese Voraussetzung rechtfertigen wir später [in Aufgabe 87.1] durch einen Test.)

1. Schritt. $\gamma = 0{,}99$ ist gefordert.

2. Schritt. Wegen $n = 100$ gehört dazu $c/\sqrt{n} = 0{,}263$.

3. Schritt. Die Berechnung ergibt

$$\bar{x} = 1{,}944 \quad s^2 = 0{,}00282 \quad s = 0{,}053.$$

(Die Werte in Abb. 66.2 sind also sehr befriedigend.)

4. Schritt. Es ist

$$a = 0{,}053 \cdot 0{,}263 = 0{,}014.$$

Wegen $1{,}944 - 0{,}014 = 1{,}930$ und $1{,}944 + 0{,}014 = 1{,}958$ erhalten wir demnach das Konfidenzintervall

$$\mathrm{KONF}\,\{1{,}930 \leq \mu \leq 1{,}958\}.$$

Tabelle 72.1. Bestimmung eines Konfidenzintervalles für den Mittelwert μ einer Normalverteilung mit unbekannter Varianz

1. Schritt. Man wähle eine Konfidenzzahl γ (95%, 99% oder dgl.).

2. Schritt. Man bestimme die Lösung c der Gleichung

$$F(c) = \frac{1}{2}(1+\gamma) \tag{72.1}$$

aus der Tafel 8 in Anhang 5 für die t-Verteilung mit $n-1$ Freiheitsgraden (n = Stichprobenumfang). Man berechne $c/\sqrt{n}$. Für $\gamma = 0{,}95$ und 0,99 und einige gängige n ergibt sich:

n	$c/\sqrt{n}$ für $\gamma=0{,}95$	$c/\sqrt{n}$ für $\gamma=0{,}99$	n	$c/\sqrt{n}$ für $\gamma=0{,}95$	$c/\sqrt{n}$ für $\gamma=0{,}99$	n	$c/\sqrt{n}$ für $\gamma=0{,}95$	$c/\sqrt{n}$ für $\gamma=0{,}99$
2	8,99	45,0	8	0,836	1,24	30	0,373	0,503
3	2,48	5,73	9	0,769	1,12	40	0,320	0,428
4	1,59	2,92	10	0,715	1,03	50	0,284	0,379
5	1,24	2,06	15	0,554	0,769	100	0,198	0,263
6	1,05	1,65	20	0,468	0,640	200	0,139	0,184
7	0,925	1,40	25	0,413	0,559	500	0,088	0,116

3. Schritt. Man berechne den Mittelwert $\bar{x}$ und die Varianz s^2 der Stichprobe $x_1, \ldots, x_n$.

4. Schritt. Man berechne

$$a = sc/\sqrt{n}\,. \tag{72.2}$$

Das Konfidenzintervall ist

$$\mathrm{KONF}\{\bar{x} - a \leqq \mu \leqq \bar{x} + a\}. \tag{72.3}$$

In Tab. 72.1 gehören zu $n = 100$ die Werte $c/\sqrt{n} = 0{,}198$ und 0,263, also $c = 1{,}98$ und 2,63. Die letzteren unterscheiden sich von den zu $n \to \infty$ gehörenden c-Werten 1,96 bzw. 2,58 nur noch sehr wenig. Die Zahlen 1,96 und 2,58 kommen auch in Tab. 69.1 vor. Das ist so, weil die Verteilungsfunktion der t-Verteilung für $n \to \infty$ gegen die Verteilungsfunktion der Normalverteilung mit dem Mittelwert 0 und der Varianz 1 strebt. Praktisch bedeutet das:

Bei einer hinreichend großen Stichprobe unterscheidet sich s so wenig von σ, daß man fast dasselbe Konfidenzintervall für μ erhält, wenn man so tut, als sei σ bekannt und gleich s, und dann nach Tab. 69.1 vorgeht. Ist die Stichprobe aber klein, so geht das nicht (vgl. Abb. 72.2).

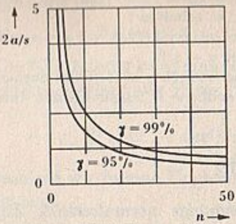


Abbildung 72.1. Länge des Konfidenzintervalls (72.3), gemessen in Vielfachen von s

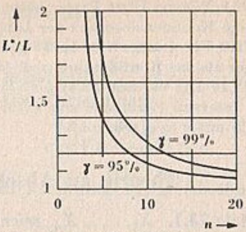


Abbildung 72.2. Verhältnis der Länge L^* des Konfidenzintervalls (72.3) zur Länge L des Konfidenzintervalls (69.2) in Abhängigkeit vom Stichprobenumfang n für gleiches σ und s

Aufgaben zu Abschnitt 72

72.1—72.7 Unter der Annahme, daß die folgenden Stichproben einer normalverteilten Grundgesamtheit entstammen, bestimme man jeweils ein 95%-Konfidenzintervall für den Mittelwert μ.

72.1 Prozentualer Stickstoffgehalt von Stahlproben (D. Blazejak-Ditges, Radex-Rundschau 1963, 348)

0,74 0,75 0,73 0,75 0,74 0,72.

72.2 Dichte [g/cm³] von Koks (H. Greif, Chem. Technik **15**, 1963, 491)

1,40 1,45 1,39 1,44 1,38.

72.3 Gewicht [g] von Abziehpapier

4,3 4,5 4,2 4,3 4,3 4,7 4,4 4,2 4,3 4,5.

72.4 Länge [mm] von Streichhölzern Marke Solo Edelweiß

44,4 44,2 44,3 44,2 44,6 44,4 44,7 44,3 44,5 43,9.

72.5 Die Stichprobe in Tab. 4.3.

72.6 Die Stichprobe in Aufgabe 6.1.

72.7 Gehirngewicht x [g] von 98 weiblichen Albinomäusen (I. Class, Z. Anatomie u. Entw.Gesch. **122**, 1960/61, 256)

x	Häufigkeit	x	Häufigkeit	x	Häufigkeit	x	Häufigkeit
0,36	1	0,41	6	0,46	9	0,51	2
0,37	1	0,42	12	0,47	7	0,52	1
0,38	1	0,43	12	0,48	6	0,53	0
0,39	3	0,44	14	0,49	5	0,54	2
0,40	4	0,45	9	0,50	3	0,55	0

72.8 Wieviel Blatt Papier hätte man in Aufgabe 72.3 etwa wiegen müssen, um ein Konfidenzintervall der Länge 0,1 g zu erhalten?

72.9 Für welches n ist das 95%-Konfidenzintervall (72.3) mit $s = s_0$ um 20% länger als das Konfidenzintervall (69.2) mit $\sigma = s_0$?

72.10 Die Varianz (62.5) der t-Verteilung ist größer als 1. Folgt daraus, daß das Intervall (72.3) mit $s = s_0$ für jedes endliche n länger als das Intervall (69.2) mit $\sigma = s_0$ sein muß?

73 Theorie zu Abschnitt 72

Satz 73.1. *$X_1, \ldots, X_n$ seien unabhängige normalverteilte Zufallsvariable mit demselben Mittelwert μ und derselben Varianz σ^2. Dann gilt: Die Variablen*

$$Z = \sqrt{n}\,\frac{\overline{X} - \mu}{\sigma} \quad \textit{mit} \quad \overline{X} = \frac{1}{n}(X_1 + \cdots + X_n) \tag{73.1}$$

und

$$Y = \frac{1}{\sigma^2}\sum_{j=1}^{n}(X_j - \overline{X})^2 \tag{73.2}$$

sind unabhängig. Y besitzt eine Chi-Quadrat-Verteilung mit $n-1$ Freiheitsgraden.

Den Beweis dieses Satzes findet man in Anhang 1.

Mit den Variablen in Satz 73.1 bilden wir die Variable

$$T = \frac{Z}{\sqrt{Y/(n-1)}}\,. \tag{73.3}$$

Diese hat die Form (62.1) mit Z statt X und $n-1$ statt n. Aus Satz 70.1 und Satz 73.1 sehen wir, daß Z und Y gerade die im Zusammenhang mit (62.1) genannten Voraussetzungen erfüllen. Also gilt der

Satz 73.2. *Unter den in Satz 73.1 genannten Voraussetzungen hat die Zufallsvariable* (73.3) *eine t-Verteilung mit $n-1$ Freiheitsgraden.*

Setzen wir (73.1) und (73.2) in (73.3) ein, so kürzt sich σ heraus, und wir erhalten unmittelbar

$$T = \sqrt{n}\,\frac{\overline{X} - \mu}{S}\,; \tag{73.4}$$

hierbei ist

$$S^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \overline{X})^2. \tag{73.5}$$

Wir wählen nun eine Zahl γ zwischen 0 und 1. Zu dieser Zahl γ können wir wegen Satz 73.2 aus Tafel 8 in Anhang 5 (mit $n-1$ Freiheitsgraden) eine Zahl c derart bestimmen, daß

$$(73.6) \qquad P(-c \leqq T \leqq c) = F(c) - F(-c) = \gamma$$

ist. Die t-Verteilung ist symmetrisch. Deshalb gilt $F(-c) = 1 - F(c)$, und (73.6) gewinnt die Form (72.1). Die Ungleichung

$$-c \leqq T \leqq c, \qquad \text{ausführlich} \qquad -c \leqq \sqrt{n}\,\frac{\bar{X} - \mu}{S} \leqq c,$$

verwandeln wir in eine Ungleichung für μ. Ähnlich wie in Abschn. 70 ergibt sich

$$(73.7) \qquad \bar{X} - A \leqq \mu \leqq \bar{X} + A \qquad \text{mit} \qquad A = cS/\sqrt{n}\,,$$

und gemäß (73.6) ist

$$P(\bar{X} - A \leqq \mu \leqq \bar{X} + A) = \gamma.$$

Liegt nun eine Stichprobe $x_1, \ldots, x_n$ aus einer normalverteilten Grundgesamtheit vor, so können wir die n Stichprobenwerte als einzelne beobachtete Werte der Variablen $X_1, \ldots, X_n$ in Satz 73.1 auffassen. Der Mittelwert $\bar{x}$ und die Varianz s^2 sind dann beobachtete Werte der Zufallsvariablen $\bar{X}$ [vgl. (73.1)] und (73.5). Setzen wir diese Werte in (73.7) ein, so ergibt sich gerade das Konfidenzintervall (72.3).

Wir vergleichen diese Herleitung mit der von (69.2). Damals spielte Z die entscheidende Rolle [vgl. (70.1)]. Z enthält aber σ und kann uns deshalb nicht mehr dienen, weil σ nicht mehr als bekannt vorausgesetzt wird. Ersetzen wir σ in Z durch S, so erhalten wir übrigens gerade T [vgl. (73.4)]. Damit wir mit T etwas anfangen konnten, mußten wir die zugehörige Verteilung kennen (siehe Satz 73.2). Das ist prinzipiell genau dieselbe Lage wie zuvor. *Überhaupt kommen jetzt und auch in den nächsten Abschnitten gar keine neuen Ideen ins Spiel sondern lediglich neue „Prüfgrößen" (statt Z nun T und so weiter) und deren Verteilungen.* Hat man also begriffen, was in Abschn. 69 und 70 vorging, so sollte man keine Schwierigkeiten haben, die gegenwärtigen und auch die folgenden Erörterungen zu verstehen.

74 Konfidenzintervalle für die Varianz der Normalverteilung

Die in der Überschrift genannten Konfidenzintervalle bestimmt man aus einer gegebenen Stichprobe gemäß Tab. 74.1. Der Mittelwert μ braucht dabei nicht bekannt zu sein.

Tabelle 74.1. Bestimmung eines Konfidenzintervalles für die Varianz σ^2 einer Normalverteilung

1. Schritt. Man wähle eine Konfidenzzahl γ (95%, 99% oder dgl.).

2. Schritt. Man bestimme Lösungen c_1 und c_2 der Gleichungen

$$F(c_1) = \frac{1}{2}(1-\gamma), \qquad F(c_2) = \frac{1}{2}(1+\gamma) \tag{74.1}$$

aus Tafel 6 in Anhang 5 für die Chi Quadrat-Verteilung mit $n-1$ Freiheitsgraden (n = Stichprobenumfang). Für $\gamma = 0{,}95$ und $0{,}99$ und einige Werte von n lauten die Lösungen folgendermaßen:

n	$\gamma = 0{,}95$		$\gamma = 0{,}99$	
	c_1	c_2	c_1	c_2
2	0,00	5,02	0,00	7,88
3	0,05	7,38	0,01	10,6
4	0,22	9,35	0,07	12,8
5	0,48	11,1	0,21	14,9
6	0,83	12,8	0,41	16,7
7	1,24	14,4	0,68	18,5
8	1,69	16,0	0,99	20,3
9	2,18	17,5	1,34	22,0
10	2,70	19,0	1,73	23,6
15	5,63	26,1	4,07	31,3
20	8,91	32,9	6,84	38,6
25	12,4	39,4	9,89	45,6
30	16,0	45,7	13,1	52,3
40	23,7	58,1	20,0	65,5
50	31,6	70,2	27,2	78,2
100	73,4	128	66,5	139

3. Schritt. Man berechne $(n-1)\,s^2$, wobei s^2 die Varianz der Stichprobe $x_1, x_2, \ldots, x_n$ ist.

4. Schritt. Man berechne

$$a_1 = (n-1)\,s^2/c_1, \qquad a_2 = (n-1)\,s^2/c_2. \tag{74.2}$$

Das Konfidenzintervall ist

$$\mathrm{KONF}\{a_2 \leq \sigma^2 \leq a_1\}. \tag{74.3}$$

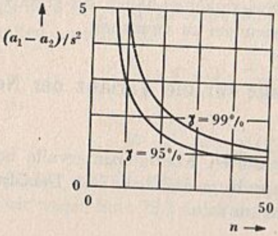


Abbildung 74.1. Länge des Konfidenzintervalls (74.3), gemessen in Vielfachen von s^2

Beispiel 74.1. Man bestimme ein 95%-Konfidenzintervall für die Varianz der Grundgesamtheit, aus der die Stichprobe in Tab. 66.1 entnommen wurde.

1. Schritt. $\gamma = 0{,}95$ ist gefordert.

2. Schritt. Wegen $n = 100$ gehört dazu $c_1 = 73{,}4$ und $c_2 = 128$.

3. Schritt. Die Berechnung ergibt $(n-1)s^2 = 0{,}2794$.

4. Schritt. So erhalten wir das Konfidenzintervall

$$\mathrm{KONF}\{0{,}0021 \leqq \sigma^2 \leqq 0{,}0038\}.$$

Das Konfidenzintervall (74.3) ergibt sich auf die folgende Weise: Die Variable (73.2) läßt sich wegen (73.5) in der Form

$$Y = (n-1)\frac{S^2}{\sigma^2}$$

schreiben und besitzt unter den in Satz 73.1 genannten Voraussetzungen eine Chi-Quadrat-Verteilung mit $n-1$ Freiheitsgraden. Also kann man eine Zahl γ zwischen 0 und 1 wählen und dazu aus Tafel 6 in Anhang 5 zwei Zahlen c_1 und c_2 derart bestimmen, daß

$$\begin{aligned} P(Y \leqq c_1) &= F(c_1) = \frac{1}{2}(1-\gamma), \\ P(Y \leqq c_2) &= F(c_2) = \frac{1}{2}(1+\gamma) \end{aligned} \tag{74.4}$$

[vgl. (74.1)] ist. Durch Subtraktion sieht man, daß dann

$$P(c_1 \leqq Y \leqq c_2) = P(Y \leqq c_2) - P(Y \leqq c_1) = \gamma \tag{74.5}$$

gilt. Die Ungleichung

$$c_1 \leqq Y \leqq c_2, \qquad \text{ausführlich} \qquad c_1 \leqq (n-1)\frac{S^2}{\sigma^2} \leqq c_2,$$

können wir in eine Ungleichung für σ^2 überführen:

$$\frac{n-1}{c_2} S^2 \leqq \sigma^2 \leqq \frac{n-1}{c_1} S^2. \tag{74.6}$$

Liegt nun eine Stichprobe $x_1, \ldots, x_n$ aus einer normalverteilten Grundgesamtheit vor, so können wir die n Stichprobenwerte als einzelne beobachtete Werte der Variablen $X_1, \ldots, X_n$ in Satz 73.1 auffassen. Dann ist die Stichprobenvarianz s^2 ein beobachteter Wert der Zufallsvariablen S^2, und indem wir s^2 in (74.6) einsetzen, erhalten wir gerade (74.3).

Aufgaben zu Abschnitt 74

74.1—74.7 Unter der Annahme, daß die folgenden Stichproben einer normalverteilten Grundgesamtheit entstammen, bestimme man jeweils ein 99%-Konfidenzintervall für die Varianz σ^2.

74.1 Die Stichprobe in Aufgabe 72.1.

74.2 Rockwell-Härte von Drehmeißeln (G. Richau, Die Technik 18, 1963, 475)

64,9 64,1 63,8 64,0.

74.3 Eine Stichprobe von 128 Werten (Zugfestigkeit von Blechen in kg/mm²) mit der Varianz $s^2 = 1{,}921$ (A. Slattenschek u. G. Schneeweiss, Maschinenbau u. Wärmewirtsch. 12, 1957, 4).

74.4 Kopfumfang [cm, gemessen über Kinn und Hinterhaupt] neugeborener männlicher Kinder (Univ.-Frauenklinik Graz, Direktor Prof. Dr. E. Navratil)

37 39 40 41 38 39 39 40 38 37.

74.5 Gewicht [g] von Theragran-Tabletten

1,19 1,23 1,18 1,21 1,27 1,17 1,15 1,14 1,19 1,20.

74.6 Die Stichprobe in Aufgabe 72.7

74.7 Die Stichprobe in Tab. 4.3.

74.8 Man wende die in Tafel 6 (Anhang 5) angegebene Näherungsformel mit $m = 100$ an und vergleiche die Werte mit den in Tafel 6 angegebenen.

75 Konfidenzintervalle für den Parameter p der Binomialverteilung

Wurden bei n Ausführungen eines Bernoulli-Experimentes k Erfolge beobachtet, so kann man unter Benutzung dieser Tatsache für den Parameter p in (39.4) ein Konfidenzintervall der Form

$$\text{KONF}\{p_1 \leqq p \leqq p_2\} \tag{75.1}$$

bestimmen. p_1 und p_2 erhält man für kleine n aus Tafel 1e in Anhang 5 und für große n als Wurzeln der quadratischen Gleichung

$$(n + c^2)\,p^2 - (2k + c^2)\,p + \frac{k^2}{n} = 0. \tag{75.2}$$

c ist dabei nach Wahl einer Konfidenzzahl γ der Tab. 69.1 zu entnehmen.

Sind außer n auch k und $n - k$ groß, so folgt aus der Formel für die Lösungen von (75.2), daß mit guter Näherung

$$\text{KONF}\left\{\frac{k}{n} - a \leqq p \leqq \frac{k}{n} + a\right\} \quad \text{mit} \quad a = \frac{c}{n}\sqrt{\frac{k(n-k)}{n}} \tag{75.3}$$

gilt.

Beispiel 75.1. In Graz waren im Jahre 1962 unter den ersten 3000 Lebendgeburten (Mehrlingsgeburten nicht mitgezählt) 1578 Knaben. Man bestimme ein 99%-Konfidenzintervall für die Wahrscheinlichkeit p des Ereignisses „Geburt eines Knaben".

In (75.3) ist

$$\frac{k}{n} = \frac{1578}{3000} = 0{,}526 \qquad \text{und} \qquad a = \frac{2{,}576}{3000}\sqrt{\frac{1578 \cdot 1422}{3000}} = 0{,}0235.$$

So erhalten wir das Konfidenzintervall

$$\text{KONF}\{0{,}502 \leqq p \leqq 0{,}550\}. \tag{75.4}$$

Beispiel 75.2. Bei einer Chlorophylluntersuchung an Maispflanzen fand E. W. Lindstrom (Cornell Univ. Agr. Exp. Station Memoir 13, 1918) unter

$n = 55$ Sämlingen $k = 42$ grüne (dominantes Merkmal) und 13 gelbe (rezessives Merkmal). Man bestimme ein 95%-Konfidenzintervall für p.

Es ist $k/n = 42/55 = 76\%$. Tafel 1e enthält zwar keine Kurven für $n = 55$. Aus den Kurven für $n = 50$ und 100 können wir aber schließen, daß das Konfidenzintervall etwa die folgende Gestalt haben muß:

$$\text{KONF}\{0{,}63 \leqq p \leqq 0{,}86\}.$$

Die Werte p_1 und p_2 in (75.1), die man der Tafel 1e entnimmt, ergeben sich aus

$$\sum_{x=k}^{n} \binom{n}{x} p_1^x (1-p_1)^{n-x} = \frac{1}{2}(1-\gamma) \tag{75.5}$$

und

$$\sum_{x=0}^{k} \binom{n}{x} p_2^x (1-p_2)^{n-x} = \frac{1}{2}(1-\gamma). \tag{75.6}$$

Die Summe

$$F = \sum_{x=0}^{k} f(x)$$

mit $f(x)$ gemäß (39.4) nimmt für jedes feste $k < np$ mit wachsendem p monoton ab, da die Ableitung von F nach p dann negativ ist. Grob gesprochen muß also p groß sein, wenn F klein sein soll, und p muß klein sein, wenn F groß oder $1 - F$ klein sein soll. So kann man sich klarmachen, warum (75.5) das linke und (75.6) das rechte Ende des Intervalles (75.1) liefert und nicht etwa umgekehrt. Die rechten Seiten von (75.5) und (75.6) sind ja klein, wenn man γ nahe bei 1 wählt.

Bei großem n ist die Zufallsvariable

$X =$ *Anzahl der Erfolge bei n Ausführungen des Bernoulli-Experimentes*

annähernd normalverteilt mit dem Mittelwert (50.2) und der Varianz (50.3). Also ist

$$Z = \frac{X - np}{\sqrt{npq}} \qquad (q = 1 - p) \tag{75.7}$$

dann annähernd normalverteilt mit dem Mittelwert 0 und der Varianz 1. Die Zahl c in (75.2) ist die Lösung der Gleichung

$$P(-c \leqq Z \leqq c) = \gamma,$$

wie man sie aus Tafel 3b oder dem daraus entnommenen Täfelchen in Tab. 69.1 erhält. In die Ungleichung $-c \leqq Z \leqq c$ setzen wir Z gemäß (75.7) ein und verwandeln sie in eine Ungleichung für p von der Form $p_1 \leqq p \leqq p_2$. Indem wir den beobachteten Wert $X = k$ einsetzen, ergeben sich dabei die Wurzeln der Gleichung (75.2) als Werte p_1 und p_2.

Aufgaben zu Abschnitt 75

75.1 WELDON würfelte bei einem seiner klassischen Experimente 49152mal und hatte dabei 25145mal „Erfolg" (= Würfeln einer 4, 5 oder 6). Man bestimme ein 95%-Konfidenzintervall für die Erfolgswahrscheinlichkeit p beim einzelnen Wurf.

75.2 Man würfle 30mal, bezeichne das Würfeln einer Fünf oder Sechs als „Erfolg" und bestimme unter Benutzung der erhaltenen Stichprobe ein Konfidenzintervall für die Erfolgswahrscheinlichkeit p beim einzelnen Wurf.

75.3 Unter Benutzung der Buffonschen und Pearsonschen Ergebnisse in Tab. 15.1 bestimme man jeweils ein 95%-Konfidenzintervall für p (= Wahrscheinlichkeit des Ereignisses „Kopf").

75.4 Aus (75.5) und (75.6) erhält man für $n = 50$ und $\gamma = 95\%$ die folgenden Werte:

k	10	20	30	40
p_1	0,10	0,27	0,45	0,66
p_2	0,34	0,55	0,73	0,90

Mit diesen vergleiche man das Konfidenzintervall (75.3).

75.5 Für $\gamma = 0{,}99$ und $n = 50$ stelle man die Länge des Konfidenzintervalles in Tafel 1e als Funktion von k graphisch dar. Warum ergibt sich eine Symmetrie bezüglich $k = 25$?

75.6 Bei einem Kreuzungsversuch glatthaariger und kraushaariger Mäuse waren in 32 Würfen zu 8 Mäusen die in der Tabelle angegebenen Anzahlen von Würfen mit x glatthaarigen Tieren (R. A. Fisher u. K. Mather, Annals Eugen. 7, 1936, 265—280). Man bestimme ein Konfidenzintervall für p (= Wahrscheinlichkeit des Ereignisses „Geburt einer glatthaarigen Maus").

x	0	1	2	3	4	5	6	7	8
Anzahl der Würfe mit genau x glatthaarigen Tieren	0	1	2	4	12	6	5	2	0

76 Konfidenzintervalle bei beliebigen Verteilungen

Bisher haben wir bei der Gewinnung der Konfidenzintervalle stets angenommen, die Art der Verteilung sei bekannt. Wir fragen uns nun, was sich tun läßt, wenn die Art der Verteilung nicht bekannt ist. Wir wollen zeigen, daß man sich dann auf eine **Näherungsmethode** stützen kann, die **bei großem Stichprobenumfang** brauchbare Ergebnisse liefert.

Diese Methode wollen wir erläutern, indem wir Konfidenzintervalle für den Mittelwert μ einer unbekannten Verteilung unter Benutzung einer gegebenen Stichprobe bestimmen.

Zunächst erinnern wir uns daran, daß für die Summe

$$X = X_1 + \cdots + X_n$$

unabhängiger Zufallsvariablen, die alle den Mittelwert μ und die Varianz σ^2 besitzen, folgendes gilt:

(A) X hat den Mittelwert $n\mu$ und die Varianz $n\sigma^2$ (vgl. Satz 58.1 und 59.1).

(B) Sind die Variablen $X_1, \ldots, X_n$ normalverteilt, so ist auch X normalverteilt (vgl. Satz 71.1).

Sind die genannten Variablen nicht normalverteilt, so gilt die Aussage (B) zwar nicht mehr, aber *für großes n* ist X wenigstens *annähernd* normalverteilt. Es gilt nämlich der folgende grundlegende

Satz 76.1 (Zentraler Grenzwertsatz). *$X_1, X_2, \ldots, X_n, \ldots$ seien unabhängige Zufallsvariable, die alle genau dieselbe Verteilungsfunktion, also auch denselben Mittelwert μ und dieselbe Varianz σ^2 besitzen. Weiterhin sei $Y_n = X_1 + \cdots + X_n$. Dann ist die Zufallsvariable*

$$Z_n = \frac{Y_n - n\mu}{\sigma\sqrt{n}} \tag{76.1}$$

„asymptotisch normalverteilt" *mit dem Mittelwert* 0 *und der Varianz* 1, *d. h. dann gilt für die zu Z_n gehörige Verteilungsfunktion $F_n(x)$ die Beziehung*

$$\lim_{n\to\infty} F_n(x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2}\, du.$$

Einen Beweis und eine etwas erweiterte Form dieses wichtigen Satzes findet man in [12] (s. Anhang 3).

Hat Z_n die angegebene Eigenschaft, so ist Y_n asymptotisch normalverteilt mit dem Mittelwert $n\mu$ und der Varianz $n\sigma^2$.

Praktisch gesehen bedeutet dieser Satz für unser gegenwärtiges Problem, daß die Zufallsvariable

$$\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n),$$

für die der Stichprobenmittelwert

$$\bar{x} = \frac{1}{n}(x_1 + \cdots + x_n)$$

einen beobachteten Wert darstellt, bei großem Stichprobenumfang n *annähernd* normalverteilt ist (und natürlich den Mittelwert μ und die Varianz σ^2/n besitzt).

Aus dieser Tatsache folgt:

Die früher behandelten Methoden zur Bestimmung von Konfidenzintervallen für den Mittelwert der Normalverteilung sind gleichzeitig brauchbare Näherungsmethoden bei beliebigen Verteilungen und umfangreichen Stichproben. Die Güte der Näherung wächst mit dem Stichprobenumfang n.

Also ist (69.2) bei großem n näherungsweise ein Konfidenzintervall für den Mittelwert μ einer beliebigen Verteilung mit bekannter Varianz, und (72.3) ist bei großem n näherungsweise ein Konfidenzintervall für

Eine Verkehrsgesellschaft glaubt, daß eine gewisse Fahrplanänderung für etwa 60% der Interessenten vorteilhaft sei. Das ist die Hypothese, die unter Benutzung einer Stichprobe (= Befragungsergebnisse zufällig ausgewählter Fahrgäste) getestet wird. Man entscheidet dann, ob man die Änderung durchführt oder nicht.

Oftmals handelt es sich um einen Vergleich, zum Beispiel der Wirksamkeit zweier Medikamente, des Ertrages bei verschiedener Düngung, der menschlichen Arbeitsleistung bei verschiedenen Arbeitsbedingungen, der Genauigkeit zweier Meßverfahren, der Qualität zweier Maschinen usw. Dann testet man die Hypothese, daß kein Unterschied besteht. Führt der Test zur Annahme der Hypothese, so entscheidet man sich bei den Medikamenten für das billigere (oder bequemer zu verabreichende) usw. Führt er zur Ablehnung, entscheidet man sich für das bessere Medikament usw. Oder man fragt sich weiter, ob der Unterschied den zusätzlichen Zeit- und Kostenaufwand rechtfertigt.

Woher stammen Hypothesen? Typische Fälle sind:

1. Die Hypothese entspringt einer Güteforderung, die man erfüllen soll (Sollwert). Diese Forderung kann dabei am „grünen Tisch" entstanden sein oder aus Voruntersuchungen stammen, bei denen man eine große Menge des betreffenden Produktes unter sorgfältig kontrollierten Bedingungen hergestellt hat. Die daraus errechneten statistischen Daten sieht man dann als Daten der Grundgesamtheit an, mit denen man die Produktionsdaten laufend vergleicht.

2. Man kennt die jeweils interessierenden Werte der Grundgesamtheit aus langer Erfahrung.

3. Die Hypothese ergibt sich aus einer Theorie, die man verifizieren möchte.

4. Die Hypothese ist eine reine Vermutung, angeregt durch Wünsche oder gelegentliche Wahrnehmungen.

Beim Testen von Hypothesen schließt man von Stichproben auf Grundgesamtheiten. Wir wissen, daß es dabei keine vollkommen sicheren Schlüsse gibt. Also muß es auch beim Testen von Hypothesen Fehlerrisiken geben. Auf diese Frage gehen wir später ein.

77 Ein Beispiel zur Einführung

Knaben- und Mädchengeburten sind ungefähr gleich häufig. Wir wollen untersuchen, ob die Wahrscheinlichkeit einer Knabengeburt genau 50% beträgt oder ob sie, wie in der Literatur behauptet wird, größer als 50% ist. (Mehrlingsgeburten betrachten wir dabei nicht mit.)

Die Wahrscheinlichkeit einer Knabengeburt bezeichnen wir mit p. Wir wollen dann also die Hypothese

$$p = 50\% = 0{,}5$$

testen. Hierzu benutzen wir eine Stichprobe von 3000 Grazer Geburten aus dem Jahre 1962, unter denen sich 1578 Knabengeburten befanden.

Trifft die Hypothese zu, sind ungefähr die Hälfte, also etwa 1500, Knabengeburten zu erwarten. Weicht die beobachtete Zahl „zu sehr" von 1500 ab, indem sie größer als ein kritischer Wert c ist, so fassen wir das als Anzeichen dafür auf, daß die Hypothese wohl nicht stimmen kann, und verwerfen diese.

Was heißt „zu sehr" ? Welches c nehmen wir, d. h. wo ziehen wir die Grenze zwischen kleinen zufallsbedingten und größeren signifikanten Abweichungen ? Gefühlsmäßig kann man verschiedener Meinung sein und sich auch darüber streiten. Gewonnen wird damit nichts. Vielmehr müssen wir, um hier die Grenzen der beschreibenden Statistik zu überschreiten, mathematische Hilfsmittel anwenden, übrigens recht einfache:

Wir bestimmen c derart, daß, falls die Hypothese richtig ist, die Wahrscheinlichkeit, mehr als c Knabengeburten in einer Stichprobe von 3000 Geburten zu beobachten, eine sehr kleine Zahl ist, die wir mit α bezeichnen. Üblicherweise wählt man $\alpha = 5\%$ oder 1%. Dabei riskiert man in etwa einem von 20 bzw. von 100 Fällen, daß man die Hypothese verwirft, obwohl sie richtig ist. Hierauf gehen wir im nächsten Abschnitt näher ein.

Wir können nun so vorgehen: Wir wählen $\alpha = 1\%$ und bestimmen dazu für die Variable

$$X = \textit{Anzahl der Knabengeburten unter 3000 Geburten}$$

einen kritischen Wert c aus der Gleichung

$$P(X > c)_{p=0{,}5} = \alpha = 0{,}01 \tag{77.1}$$

unter der Annahme, die Hypothese sei richtig (das deuten wir durch den Index $p = 0{,}5$ an). Ist der beobachtete Wert $x = 1578$ größer als c, so verwerfen wir die Hypothese. Ist hingegen $1578 \leq c$, so nehmen wir die Hypothese an.

Trifft die Hypothese zu, so hat X eine Binomialverteilung mit $p = 0{,}5$ und $n = 3000$, und diese können wir durch die Normalverteiden Mittelwert μ einer beliebigen Verteilung mit unbekannter Varianz. Da die Verteilungsfunktion der t-Verteilung für $n \to \infty$ gegen die Verteilungsfunktion der Normalverteilung strebt (vgl. Abschn. 62), so unterscheiden sich die beiden genannten Konfidenzintervalle bei großem n nur wenig voneinander.

Der zentrale Grenzwertsatz ist natürlich einer der Hauptgründe für die große Bedeutung der Normalverteilung in der Statistik.

Wie groß eine Stichprobe mindestens sein muß, damit man von der vorstehenden Näherung brauchbare Resultate erwarten kann, das hängt hauptsächlich von der Art der Parameter ab, für die man Konfidenzintervalle bestimmen will. Läßt die Stichprobe vermuten, daß die zugehörige Verteilung nicht allzu unsymmetrisch ist, so erhält man für den Mittelwert im allgemeinen befriedigende Ergebnisse, sobald n wenigstens etwa gleich 30 ist, und für die Varianz, wenn n wenigstens etwa gleich 100 ist.

Aufgaben zu Abschnitt 76

76.1—76.5 Unter Verwendung der gegebenen Stichprobe bestimme man jeweils ein 95%-Konfidenzintervall für den Mittelwert und die Varianz.

76.1 Die Stichprobe in Aufgabe 10.5.

76.2 Die Stichprobe in Aufgabe 6.1.

76.3 Gewicht [engl. Pfund] männlicher Erwachsener (British Assn., Anthropometric Comm., Final Report, 1883).

76.4 Die Stichprobe in Aufgabe 5.3. (Als Umfang nehme man 100 an.)

76.5 Gewicht [g] von österreichischen Schillingstücken, geprägt 1959—1961.

76.6 Wie umfangreich müßte die Stichprobe in Aufgabe 76.3 etwa sein, damit man für μ ein Konfidenzintervall der Länge 1 [Pfund] erhielte?

Zu Aufgabe 76.5

Gewicht	Anzahl
4,07	1
4,10	1
4,12	5
4,13	3
4,14	4
4,15	5
4,16	7
4,17	6
4,18	6
4,19	9
4,20	10

Gewicht	Anzahl
4,21	10
4,22	11
4,23	7
4,24	3
4,25	3
4,26	3
4,28	1
4,29	1
4,30	1
4,32	2
4,38	1

Zu Aufgabe 76.3

Gewicht (Klassenmitte) x_k	Häufigkeit
105	5
115	1
125	7
135	42
145	57
155	51
165	36
175	25
185	13
195	8
205	1
215	1

Kapitel 14

Testen von Hypothesen, Entscheidungen

Unter einer **Hypothese** versteht man in der Statistik eine Annahme über die Verteilung einer Zufallsvariablen, z. B. die Annahme, daß die betreffende Verteilung den Mittelwert 20,3 hat, usw. Und ein **Test** einer Hypothese ist ein Prüfverfahren, das man anwendet, wenn man wissen will, ob man die Hypothese „**annehmen**" kann oder „**verwerfen**" soll, d. h. ob man der Hypothese Vertrauen schenken und sie für richtig halten kann oder nicht.

Warum will man das wissen? Nun, weil man oftmals zwischen zwei Handlungen entscheiden muß und diese Entscheidung davon abhängt, ob die Hypothese annehmbar ist oder nicht. Es handelt sich also im Grunde genommen um eine „Theorie der Entscheidungen". Wir erläutern dies durch einige typische Beispiele.

Eine Firma kontrolliert die Qualität ihrer Produktion, indem sie stündlich eine Stichprobe entnimmt. Dabei wird das Merkmal gemessen, auf das es jeweils ankommt, z. B. die Festigkeit des gesponnenen Garns oder der Durchmesser der hergestellten Bolzen usw. Eine solche Größe ist eine Zufallsvariable. Sie unterliegt nämlich kleinen zufälligen Schwankungen, hervorgerufen durch Ungleichmäßigkeiten des Rohmaterials, des Verhaltens der Maschinen, auf denen man produziert, usw. Nehmen wir z. B. an, der genannte Durchmesser soll $\mu = 5$ [mm] betragen. Dann ist die Hypothese $\mu = 5$ zu testen. Weicht der Mittelwert $\bar{x}$ nicht „zu sehr" von der Hypothese ab, so nimmt man diese an und läßt die Produktion weiterlaufen. Ist die Abweichung „zu groß", so verwirft man die Hypothese, stoppt die Produktion und beginnt mit der Fehlersuche. Aber was heißt „zu sehr" und „zu groß"? Wo sollen wir die Grenze zwischen kleinen rein zufallsbedingten unvermeidlichen Abweichungen und größeren Abweichungen ziehen, die nicht mehr durch den Zufall allein erklärbar sind? Letztere heißen **signifikante Abweichungen,** im Gegensatz zu den rein zufallsbedingten. Die Antwort auf unsere Frage folgt aus der Theorie, die wir entwickeln wollen.

lung mit dem Mittelwert $\mu = np = 1500$ und der Varianz $\sigma^2 = npq = 750$ brauchbar annähern. (Den Summanden 0,5 in (50.5) lassen wir der Einfachheit halber unberücksichtigt.) Also folgt aus (77.1)

$$P(X > c) = 1 - P(X \leqq c) \approx 1 - \Phi\left(\frac{c - 1500}{\sqrt{750}}\right) = 0{,}01.$$

Aus Tafel 3b in Anhang 5 ergibt sich

$$\frac{c - 1500}{\sqrt{750}} = 2{,}326, \quad c = 1564.$$

Da $1578 > c$ ist, verwerfen wir die Hypothese und nehmen an, daß

$$p > 0{,}5$$

ist. Eine solche Gegenannahme heißt eine *Alternativhypothese* oder kurz eine **Alternative.** Damit ist der Test beendet.

Bei einer Stichprobe von nur 300 Werten hätte sich für die Variable

$$X = \textit{Anzahl der Knabengeburten unter 300 Geburten}$$

aus (77.1) der kritische Wert $c = 170$ ergeben. Hätte die kleinere Stichprobe denselben prozentualen Anteil (158) Knabengeburten aufgewiesen, so wäre die Hypothese $p = 0{,}5$ wegen $158 < c$ angenommen worden. Dies zeigt, daß die Brauchbarkeit des Tests mit dem Stichprobenumfang n wächst. Man wird n so groß wie nötig wählen, damit der Test über die Frage, die einen interessiert, Aufschluß gibt, andererseits aus Kosten- und Zeitgründen aber so klein wie möglich, denn es hat keinen Sinn, Einzelheiten, die praktisch belanglos sind, noch als statistisch signifikant nachweisen zu wollen. Anhaltspunkte für die Wahl von n kann man meist durch kurze Vorversuche gewinnen.

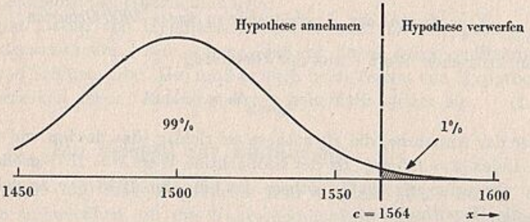


Abbildung 77.1. Lage des kritischen Wertes $c = 1564$ bezüglich der (angenäherten) Verteilung der Variablen X im Falle der Richtigkeit der Hypothese

Aufgaben zu Abschnitt 77

77.1 Unter Benutzung des Ergebnisses von BUFFON in Tab. 15.1 teste man die Hypothese, die Münze sei regelmäßig, gegen die Alternative, daß „Kopf" bevorzugt auftritt.

77.2 Welche Alternative bezüglich q (= Wahrscheinlichkeit des Ereignisses „Wappen") ist in Aufgabe 77.1 sinnvoll? Man führe den Test durch.

77.3 Man führe den Test in Aufgabe 77.1 unter Benutzung der PEARSONschen Ergebnisse (s. Tab. 15.1) durch.

77.4 Man bezeichne das Würfeln einer 5 oder 6 als „Erfolg". Man würfle 100mal, notiere die Anzahl der Erfolge und teste die Hypothese, der Würfel sei regelmäßig.

78 Typen von Alternativen. Fehler beim Testen

Im vorigen Abschnitt handelte es sich um eine Zufallsvariable X, deren Verteilung einen unbekannten Parameter enthält, den wir nun allgemein mit u (statt p) bezeichnen. Die Hypothese war $u = u_0 = 0{,}5$. Die Alternative

$$u > u_0 \tag{78.1}$$

wurde durch die Art des Problems nahegelegt. In anderen Fällen sind auch andersartige Alternativen sinnvoll, nämlich

$$u < u_0 \tag{78.2}$$

oder auch

$$u \neq u_0 . \tag{78.3}$$

Der Typ (78.2) tritt z. B. bei Festigkeitsuntersuchungen auf. u_0 ist dann der Sollwert (Mindestfestigkeit). Die Alternative bedeutet den unerwünschten Fall, daß die Festigkeit zu klein ist. Größer als u_0 kann sie sein, das macht nichts, darum braucht man sich nicht extra zu kümmern. Die „zweiseitige Alternative" (78.3) tritt z. B. bei Abmessungen auf, die weder zu klein noch zu groß sein dürfen, wie etwa der Durchmesser einer Welle. Abweichungen vom Sollwert u_0 nach oben oder nach unten sind gleich wichtig und beide unerwünscht. Man spricht dann von einem **„zweiseitigen Test"**, im Gegensatz zu dem **„einseitigen Test"** im Fall der Alternative (78.1) oder (78.2). Beim zweiseitigen Test hat man zwei kritische Werte c_1 und c_2, und der **„Verwerfungsbereich"** besteht aus zwei Teilen, wie Abb. 78.1 schematisch zeigt. Rechenbeispiele dazu folgen im nächsten Abschnitt.

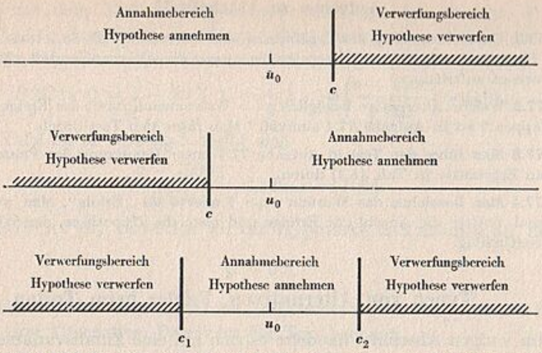


Abbildung 78.1. Test bei Alternative (78.1) (oberes Bild), (78.2) (mittleres Bild) und (78.3). Schematische Darstellung

Wir wollen uns nun überlegen, wie es mit Fehlentscheidungen beim Testen steht. $u = u_0$ sei die Hypothese. Als Alternative wählen wir der Einfachheit halber zuerst einen festen Wert

$$u = u_1.$$

u_1 sei größer als u_0. Dann liegt der kritische Wert c und auch der Verwerfungsbereich rechts von u_0 (vgl. Abb. 78.1, oberes Bild). Aus einer Stichprobe $x_1, \ldots, x_n$ wird ein Schätzwert

(78.4) $$\tilde{u} = g(x_1, \ldots, x_n)$$

für den unbekannten Parameterwert u bestimmt. Ist $\tilde{u} > c$, so verwirft man die Hypothese (wie z. B. im vorigen Abschnitt). Ist $\tilde{u} \leqq c$, so nimmt man die Hypothese an. $\tilde{u}$ ist als beobachteter Wert der Zufallsvariablen

$$\tilde{U} = g(X_1, \ldots, X_n)$$

aufzufassen.

Bei diesem Test gibt es 2 Fehlermöglichkeiten (vgl. auch Tab. 78.1):

Fehler 1. Art. Die Hypothese wird verworfen, obwohl sie richtig ist. Die Wahrscheinlichkeit, einen solchen Fehler zu begehen, wird mit α bezeichnet und heißt die **Signifikanzzahl** des Tests. Einen solchen Fehler begeht man, wenn die Hypothese richtig ist und $\tilde{U}$ einen Wert $\tilde{u} > c$ annimmt. Also gilt

(78.5) $$P(\tilde{U} > c)_{u=u_0} = \alpha.$$

Fehler 2. Art. Die Hypothese wird angenommen, obwohl sie falsch ist. Die zugehörige Wahrscheinlichkeit wird mit $1-\beta$ bezeichnet. Einen solchen Fehler begeht man, wenn die Hypothese falsch ist und $\tilde{U}$ einen Wert $\tilde{u} \leqq c$ annimmt. Also ist

$$P(\tilde{U} \leqq c)_{u=u_1} = 1 - \beta. \tag{78.6}$$

β ist die Wahrscheinlichkeit, einen Fehler 2. Art zu vermeiden, und heißt die **Macht** des Tests.

Zum Beispiel entnimmt man bei der *Annahmekontrolle* einer Sendung eine Stichprobe und testet die Hypothese, daß die Sendung den Lieferbedingungen entspricht. Weist man dabei eine Sendung zurück, obwohl sie den Lieferbedingungen entspricht, so begeht man einen Fehler 1. Art. Nimmt man eine Sendung an, obwohl sie den Lieferbedingungen nicht entspricht, so begeht man einen Fehler 2. Art. Die zugehörigen Wahrscheinlichkeiten α und $1-\beta$ bezeichnet man auch als **Risiko** 1. bzw. 2. Art oder auch als *Produzenten-* bzw. *Konsumentenrisiko.* i

Tabelle 78.1. Fehler 1. und 2. Art beim Testen einer Hypothese $u = u_0$ gegen eine Alternative $u = u_1$

Sendung entspricht n… den Bedingungen

Alternativhyp. ist aber richtig so wird nicht verworfen

d.h. Sendung wird angenommen.

Konsumentenri…

		Unbekannte Wirklichkeit	
		$u = u_0$	$u = u_1$
Angenommen	$u = u_0$	Richtige Entscheidung $P = 1 - \alpha$	Fehler 2. Art $P = 1 - \beta$
	$u = u_1$	Fehler 1. Art $P = \alpha$	Richtige Entscheidung $P = \beta$

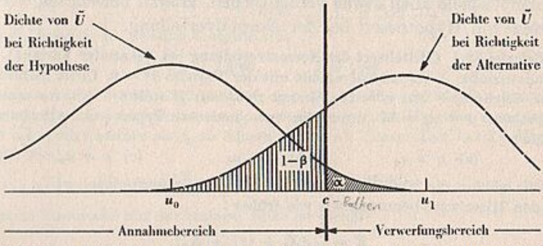


Abbildung 78.2. Zum Begriff des Fehlers 1. und 2. Art (schematisch)

Aus (78.5) und (78.6) sieht man, daß α und β von c abhängen. Also sollte man c so wählen, daß die Fehlerwahrscheinlichkeiten α und $1-\beta$ beide möglichst klein werden, daß also α möglichst klein und die Macht β möglichst groß wird. Aber diese beiden Forderungen widerstreiten einander. Dies illustriert die schematische Abb. 78.2: Um α zu verkleinern, muß man c nach rechts verschieben. Dabei vergrößert man leider gleichzeitig $1-\beta$. In der Praxis wählt man zuerst α (5%, 1% oder dgl.), bestimmt dazu c und berechnet schließlich β.

Aus der Annahme einer Hypothese auf Grund eines Tests folgt natürlich *nicht*, daß diese Hypothese die einzig mögliche ist. Außerdem muß man bedenken: Das Risiko 1. Art wählt man klein, aber das Risiko 2. Art $(1-\beta)$ kann groß ausfallen. Solange man über $1-\beta$ nichts weiß, muß man vorsichtig sein und sollte dies vielleicht sogar sprachlich betonen, indem man „*nicht verwerfen*" statt „annehmen" sagt.

Stellt es sich heraus, daß $1-\beta$ groß ist, weil β klein ist, so kann man sich helfen, indem man den Test mit einer größeren Stichprobe wiederholt. Dies läßt sich schon aus den Betrachtungen am Schluß des vorigen Abschnittes vermuten und wird im nächsten Abschnitt noch näher erläutert.

Ist die Alternative nicht auf einen speziellen Zahlenwert beschränkt, sondern hat sie die Form (78.1)—(78.3), so wird natürlich die Macht β eine Funktion von u. Diese bezeichnet man bisweilen als die **Gütefunktion** des Tests. Auch hierauf gehen wir im folgenden Abschnitt näher ein.

79 Anwendung auf die Normalverteilung

Die Überlegungen im vorigen Abschnitt sollten wir uns durch einige Zahlenbeispiele noch etwas verdeutlichen. Hierzu betrachten wir das Testen von Hypothesen bei der Normalverteilung.

Beispiel 79.1 (Mittelwert der Normalverteilung bei bekannter Varianz). Die Zufallsvariable X sei normalverteilt mit der Varianz $\sigma^2 = 9$. Unter Benutzung einer Stichprobe von $n = 10$ Werten mit dem Mittelwert $\bar{x}$ teste man die Hypothese $\mu = \mu_0 = 24$ gegen die verschiedenen Typen von Alternativen, nämlich

(a) $\mu > \mu_0$ (b) $\mu < \mu_0$ (c) $\mu \neq \mu_0$.

Wir wählen die Signifikanzzahl $\alpha = 0{,}05$. Als Schätzfunktion $g(X_1, \ldots, X_n)$ für den Mittelwert benutzen wir wie früher

$$\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n).$$

Trifft die Hypothese zu, so ist $\overline{X}$ gemäß Satz 71.3 normalverteilt mit dem Mittelwert $\mu = 24$ und der Varianz $\sigma^2/n = 0{,}9$. Bei der Bestimmung des kritischen Wertes c können wir also jeweils die Tafel 3b in Anhang 5 benutzen.

Fall (a). Wir bestimmen c aus

$$P(\overline{X} > c)_{\mu=24} = \alpha,$$

also

$$P(\overline{X} \leqq c)_{\mu=24} = \Phi\left(\frac{c-24}{\sqrt{0{,}9}}\right) = 1 - \alpha = 0{,}95.$$

Der genannten Tafel entnehmen wir

$$\frac{c-24}{\sqrt{0{,}9}} = 1{,}645 \qquad \text{oder} \qquad c = 25{,}56.$$

c liegt rechts von μ_0, wie in Abb. 78.1 oben. Ist $\bar{x} \leq 25{,}56$, so nehmen wir die Hypothese an. Ist $\bar{x} > 25{,}56$, so verwerfen wir sie. Dieser Test hat die Macht (s. Abb. 79.1)

$$\begin{aligned}\beta(\mu) &= P(\overline{X} > 25{,}56)_\mu = 1 - P(\overline{X} \leqq 25{,}56)_\mu \\ &= 1 - \Phi\left(\frac{25{,}56-\mu}{\sqrt{0{,}9}}\right) = 1 - \Phi(26{,}94 - 1{,}05\mu).\end{aligned} \tag{79.1}$$

Fall (b). Den kritischen Wert c bestimmen wir nun aus

$$P(\overline{X} \leqq c)_{\mu=24} = \Phi\left(\frac{c-24}{\sqrt{0{,}9}}\right) = \alpha = 0{,}05.$$

Vermöge Tafel 3b erhalten wir $c = 24 - 1{,}56 = 22{,}44$. Ist $\bar{x} \geq 22{,}44$, so nehmen wir die Hypothese an. Ist $\bar{x} < 22{,}44$, so lehnen wir sie ab. Dieser Test hat die Macht

$$\beta(\mu) = P(\overline{X} \leqq 22{,}44)_\mu = \Phi\left(\frac{22{,}44-\mu}{\sqrt{0{,}9}}\right) = \Phi(23{,}65 - 1{,}05\mu). \tag{79.2}$$

Fall (c). Da die Normalverteilung symmetrisch ist, wählen wir die kritischen Werte c_1 und c_2 symmetrisch zu $\mu = 24$ gelegen, also $c_1 = 24 - k$, $c_2 = 24 + k$, und bestimmen sie aus

$$P(24-k \leqq \overline{X} \leqq 24+k)_{\mu=24} = \Phi\left(\frac{k}{\sqrt{0{,}9}}\right) - \Phi\left(-\frac{k}{\sqrt{0{,}9}}\right) = 1 - \alpha = 0{,}95.$$

Vermöge Tafel 3b in Anhang 5 ergibt sich

$$k/\sqrt{0{,}9} = 1{,}960, \qquad k = 1{,}86.$$

So wird $c_1 = 24 - 1{,}86 = 22{,}14$ und $c_2 = 24 + 1{,}86 = 25{,}86$. Ist $\bar{x}$ nicht größer als c_2 und nicht kleiner als c_1, so nehmen wir die Hypothese an. Ist $\bar{x}$ größer als c_2 oder kleiner als c_1, so lehnen wir sie ab. Dieser Test hat die Macht (s. Abb. 79.1)

$$\beta(\mu) = P(\overline{X} < 22{,}14)_\mu + P(\overline{X} > 25{,}86)_\mu.$$

Der letzte Summand auf der rechten Seite ist gleich

$$1 - P(\overline{X} \leqq 25{,}86)_\mu.$$

So erhalten wir zunächst

$$\beta(\mu) = 1 + \Phi\left(\frac{22{,}14 - \mu}{\sqrt{0{,}9}}\right) - \Phi\left(\frac{25{,}86 - \mu}{\sqrt{0{,}9}}\right).$$

Durch Ausführung der Division folgt

$$\beta(\mu) = 1 + \Phi(23{,}34 - 1{,}05\mu) - \Phi(27{,}26 - 1{,}05\mu). \tag{79.3}$$

Die Funktion $1 - \beta(\mu)$ heißt die **Operations-Charakteristik** des betreffenden Tests. Wie wir wissen, ist dies die Wahrscheinlichkeit, einen Fehler 2. Art zu begehen (vgl. Tab. 78.1). Im Fall (c) unseres Beispiels ist (ausgezogene Kurve in Abb. 79.2)

$$1 - \beta(\mu) = \Phi(27{,}26 - 1{,}05\mu) - \Phi(23{,}34 - 1{,}05\mu).$$

Hätten wir den Test mit einer größeren Stichprobe, etwa mit $n = 100$ Werten (statt $n = 10$) durchgeführt, so wäre $\sigma^2/n = 0{,}09$ (statt 0,9). Dann würde $c_1 = 23{,}41$ und $c_2 = 24{,}59$, wie man durch Nachrechnen bestätigt. Dieser Test hätte also die Operations-Charakteristik

$$1 - \beta(\mu) = \Phi\left(\frac{24{,}59 - \mu}{\sqrt{0{,}09}}\right) - \Phi\left(\frac{23{,}41 - \mu}{\sqrt{0{,}09}}\right).$$

Vereinfachung ergibt

$$1 - \beta(\mu) = \Phi(81{,}97 - 3{,}33\mu) - \Phi(78{,}03 - 3{,}33\mu).$$

Abb. 79.2 zeigt, daß die zugehörige Kurve steilere Flanken hat als die zu $n = 10$ gehörige. Dies bedeutet, der Test hat durch die Vergrößerung des Stichprobenumfanges an Trennschärfe gewonnen, im Einklang mit der Schlußbemerkung im vorigen Abschnitt. n wählt man gerade so groß, daß der Test alle praktisch interessierenden Abweichungen

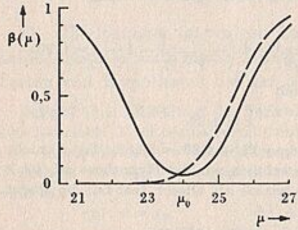


Abbildung 79.1. Macht $\beta(\mu)$ in Beispiel 79.1, Fall (a) (gestrichelt) und (c)

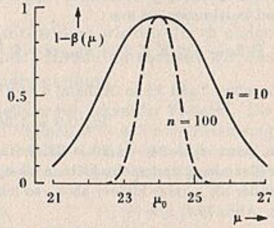


Abbildung 79.2. Kurven der Operations-Charakteristik in Beispiel 79.1, Fall (c), bei zwei verschiedenen Stichprobenumfängen n

zwischen μ und μ_0 aufdeckt, aber aus Kosten- und Zeitgründen nicht größer. Interessiert z. B. eine Abweichung von ± 2 Einheiten, so sieht man aus Abb. 79.2 auf den ersten Blick, daß eine Stichprobe von 10 Werten viel zu klein ist, denn für $\mu = 24 - 2 = 22$ bzw. $\mu = 26$ ist $1 - \beta$ noch fast 50%, das Risiko eines Fehlers 2. Art also viel zu groß. $n = 100$ reicht für den verlangten Zweck offenbar aus, wie man aus der gestrichelten Kurve ersieht, und man könnte sich fragen, ob vielleicht auch ein kleineres n genügt.

Beispiel 79.2 (Mittelwert der Normalverteilung bei unbekannter Varianz). Bei einer Stichprobe von $n = 16$ Messungen der Reißlast von 3 Zoll (= 76 mm) starken Manilahanftauen ergab sich der Mittelwert $\bar{x} = 4482$ kg und die Standardabweichung $s = 115$ kg (N. C. Wiley, 41. Jahrestagung der Amer. Soc. for Test. Materials). Man teste die Hypothese $\mu_0 = 4500$ kg gegen die Alternative $\mu_1 = 4400$ kg. Hierbei setze man voraus, daß die Reißlast eine normalverteilte Zufallsvariable ist. μ_0 könnte z. B. ein vom Hersteller angegebener Sollwert sein und μ_1 ein aus früheren Erfahrungen gewonnener Wert.

Als Signifikanzzahl wählen wir $\alpha = 5\%$. Ist die Hypothese richtig, so folgt aus Satz 73.2, daß die Zufallsvariable

$$T = \sqrt{n}\,\frac{\overline{X} - \mu_0}{S} = 4\,\frac{\overline{X} - 4500}{S}$$

eine t-Verteilung mit $n - 1 = 15$ Freiheitsgraden hat. Der kritische Wert ergibt sich aus

$$P(T \leqq c)_{\mu_0} = \alpha = 0{,}05\,.$$

Vermöge Tafel 8 in Anhang 5 erhalten wir $c = -1{,}75$. Die Stichprobe liefert als beobachteten Wert von T die Zahl

$$t = 4\,\frac{\bar{x} - 4500}{s} = 4\,\frac{4482 - 4500}{115} = -\,0{,}626\,.$$

Es ist $t > c$. Die Hypothese wird also angenommen.

Beispiel 79.3 (Varianz der Normalverteilung). Unter Benutzung einer Stichprobe mit dem Umfang $n = 15$ und der Varianz $s^2 = 13$ teste man die Hypothese, daß die Grundgesamtheit, die als normalverteilt vorausgesetzt wird, die Varianz $\sigma^2 = \sigma_0^2 = 10$ hat, gegen die Alternative $\sigma^2 = \sigma_1^2 = 20$.

Wir wählen die Signifikanzzahl $\alpha = 5\%$. Trifft die Hypothese zu, so hat die Variable

$$Y = (n - 1)\,\frac{S^2}{\sigma^2} = 14\,\frac{S^2}{10} = 1{,}4 S^2$$

(vgl. Abschn. 74) eine Chi-Quadrat-Verteilung mit $n - 1 = 14$ Freiheitsgraden. Aus der Gleichung

$$P(Y > c) = \alpha = 0{,}05 \qquad \text{oder} \qquad P(Y \leqq c) = 0{,}95$$

und Tafel 6 mit 14 Freiheitsgraden ergibt sich $c = 23{,}68$ als kritischer Wert für Y. Für $S^2 = Y/1{,}4$ erhalten wir also den kritischen Wert $c^* = 23{,}68/1{,}4 = 16{,}91$. Da $s^2 < c^*$ ist, nehmen wir die Hypothese an.

Träfe die Alternative zu, so hätte die Variable

$$Y_1 = 14\frac{S^2}{\sigma_1^2} = 0{,}7S^2$$

eine Chi-Quadrat-Verteilung mit 14 Freiheitsgraden. Also hat unser Test die Macht

$$\beta = P(S^2 > c^*)_{\sigma^2=20} = P(Y_1 > 0{,}7\,c^*)_{\sigma^2=20} = 1 - P(Y_1 \leqq 11{,}84)_{\sigma^2=20} \approx 62\,\%.$$

Obwohl sich die Alternative sehr von der Hypothese unterscheidet, ist also das Risiko 2. Art sehr groß, nämlich $1 - \beta = 38\%$. Abhilfe schafft man, indem man zu umfangreicheren Stichproben übergeht.

Aufgaben zu Abschnitt 79

Bei den folgenden Aufgaben nehme man an, die Grundgesamtheit, aus denen die Stichproben stammen, seien normalverteilt.

79.1 Man teste die Hypothese, daß die Stichprobe in Beispiel 10.1 einer Grundgesamtheit mit dem Mittelwert 20 mm (das war der angegebene Sollwert) entstammt.

79.2 Man teste die Hypothesen, daß die Grundgesamtheit, aus der die Stichprobe von Meßwerten an Permalloy-Legierungen in Beispiel 8.1 entstammt, den Mittelwert 80 [%] und die Standardabweichung 2 [‰] hat.

79.3 Unter Benutzung der Stichprobe in Aufgabe 74.5 teste man die Hypothese $\mu = 1{,}2$ [g] gegen die Alternative $\mu < 1{,}2$ [g].

79.4 In Aufgabe 79.3 wähle man als Hypothese

$$\text{(a) } \mu = 1{,}25 \text{ [g]}, \quad \text{(b) } \mu = 1{,}15 \text{ [g]},$$

überlege sich jeweils eine sinnvolle Alternative und führe den Test durch.

79.5 Bei der 143. Erdumkreisung des Satelliten Telstar am 25./26. 7. 1962 zeigte das Azimut die folgenden Abweichungen [in Neuminuten] von den vorherberechneten Werten (A. J. CLAUS, R. B. BLACKMAN, E. G. HALLINE, W. C. RIDGWAY, Bell Labor. Record **41**, 2963, 1380):

$$1 \quad -1 \quad 1 \quad 3 \quad -8 \quad 6 \quad 0$$

Unter Benutzung dieser Stichprobe teste man die Hypothese $\mu = 0$.

Tabelle 79.1. Relative Abweichung [%] der Kapazität vom Nennwert für 48 Tantal-Elektrolytkondensatoren (H. STRAUCH, Telefunken-Zeitung. **36**, 1963, 8)

Abweichung	Klassenmitte	Anzahl Kondensatoren
−13,5 ⋯ −10,5	−12	1
−10,5 ⋯ − 7,5	− 9	1
− 7,5 ⋯ − 4,5	− 6	4
− 4,5 ⋯ − 1,5	− 3	15
− 1,5 ⋯ 1,5	0	16
1,5 ⋯ 4,5	3	5
4,5 ⋯ 7,5	6	4
7,5 ⋯ 10,5	9	2

79.6 Unter Benutzung der Meßwerte in Tab. 79.1 teste man die Hypothese $\mu = 0$. Was bedeutet diese Hypothese praktisch?

79.7 Um zwei Methoden zur Stärkegehaltsbestimmung miteinander zu vergleichen, wurden 16 Kartoffeln halbiert und jeweils beide Methoden angewendet (C. v. Scheele, G. Svensson u. J. Rasmusson, Nordisk Jordbrugsforskning 1935, 22). Es ergaben sich die folgenden Unterschiede des Stärkegehalts [‰] zwischen den beiden Hälften:

$$2 \quad 0 \quad 0 \quad 1 \quad 2 \quad 2 \quad 3 \quad -3 \quad 1 \quad 2 \quad 3 \quad 0 \quad -1 \quad 1 \quad -2 \quad 1$$

Man teste die Hypothese, daß kein signifikanter Unterschied besteht.

79.8 Man zeichne die Operations-Charakteristik für die Fälle (a) und (b) in Beispiel 79.1.

80 Kontrollkarten

Kein Produktionsprozeß ist so vollkommen, daß alle hergestellten Stücke absolut gleich ausfallen. Material, Maschine und Mensch bedingen eine gewisse Variabilität. Deshalb muß man darauf bedacht sein, daß die vorgeschriebenen Sollwerte der Produkte (z. B. Abmessungen, Festigkeitseigenschaften oder was sonst jeweils wichtig ist) möglichst gut eingehalten werden. Unter Verwendung einer geeigneten Stichprobe testet man hierzu die Hypothese, daß ein bestimmter wichtiger Sollwert eingehalten ist, gegen eine sinnvolle Alternative. Tut man das hinterher, nachdem ein Auftrag (z. B. 10000 Kolbenringe) produktionsmäßig erledigt wurde, so kann man bei dieser „*Endkontrolle*" feststellen, ob der Produktionsvorgang zufriedenstellend abgelaufen ist oder nicht. Ändern kann man aber am Resultat hinterher natürlich nichts mehr. Deshalb ist es viel besser, dem Ablauf des Produktionsvorganges auf den Fersen zu bleiben, indem man in gewissen Zeitabständen (stündlich, halbstündlich) kleine Stichproben (3 bis 10 Produkte) entnimmt und sofort testet, ob der vorgeschriebene Sollwert μ_0 eingehalten wird. Nimmt man bei einem solchen Test die Hypothese $\mu = \mu_0$ an, so läßt man den Produktionsvorgang weiterlaufen. Führt ein Test zur Ablehnung, stoppt man den Prozeß, sucht und behebt die Ursache der Abweichung vom Sollwert. Dieses in regelmäßigen Zeitabständen wiederholte Testen einer Hypothese und das Eingreifen im Falle der Verwerfung der Hypothese bezeichnet man als **statistische Qualitätskontrolle** oder besser als **Qualitätsregelung.**

Ein Fehler 1. Art bedeutet hierbei, daß man den Produktionsprozeß stoppt, obwohl er ordentlich läuft, und nach einer Störungsursache fahndet, die gar nicht existiert. Einen Fehler 2. Art begeht man, wenn man den Prozeß laufen läßt, obwohl er gestört ist.

Daß die Qualitätsregelung so beliebt und wirkungsvoll ist, hängt auch damit zusammen, daß man die Ergebnisse des Tests und damit das Verhalten des Produktionsprozesses jeweils sofort graphisch veranschaulicht. Man trägt nämlich die Mittelwerte der Stichproben laufend in eine „**Kontrollkarte**" ein. Ein Beispiel zeigt Abb. 80.1. Die **Kontrollgrenzen** c_1 und c_2 entsprechen den kritischen Werten c_1 und c_2 in Abb. 78.1 und in Beispiel 79.1, Fall (c). Solange die Mittelwerte zwischen den Kontrollgrenzen (oder auf diesen Grenzen) liegen, nimmt man die Hypothese an und läßt den Fabrikationsprozeß weiterlaufen. Fällt ein Mittelwert außerhalb, verwirft man die Hypothese, stoppt die Fabrikation und beginnt mit der Fehlersuche.

Zieht man die Kontrollgrenzen zu eng, so riskiert man, daß der Fertigungsprozeß unberechtigt oft unterbrochen wird. Zieht man die Grenzen zu weit, so wird das Risiko 2. Art zu groß. Üblicherweise wählt man die Signifikanzzahl $\alpha = 1\%$. Unter der Voraussetzung, daß eine Normalverteilung vorliegt, erhält man gemäß Satz 71.3 und Tafel 3b die Werte

$$c_1 = \mu_0 - 2{,}58\frac{\sigma}{\sqrt{n}} \qquad \text{und} \qquad c_2 = \mu_0 + 2{,}58\frac{\sigma}{\sqrt{n}}.$$

σ wird dabei als bekannt vorausgesetzt. Liegt keine Erfahrung über σ vor, so berechne man die Standardabweichungen der ersten 20 oder 30 Stichproben und benutze das arithmetische Mittel dieser 20 oder 30 Zahlen als Näherung von σ.

Bisweilen zeichnet man auf Kontrollkarten zusätzlich noch die „*Warngrenzen*"

$$w_1 = \mu_0 - 1{,}96\frac{\sigma}{\sqrt{n}} \qquad \text{und} \qquad w_2 = \mu_0 + 1{,}96\frac{\sigma}{\sqrt{n}}$$

mit ein, die der Signifikanzzahl $\alpha = 5\%$ entsprechen. Fällt ein Mittelwert außerhalb der Warngrenzen (aber noch in den Bereich zwischen den Kontrollgrenzen), so entnimmt man gleich noch eine Stichprobe und wiederholt den Test.

Übrigens dienen die Verbindungsstrecken in Abb. 80.1 nur dazu, dem Auge das Durchlaufen der Werte in zeitlicher Reihenfolge zu erleichtern, und zu sonst nichts.

In der Praxis erlebt man sehr oft folgendes: Jedesmal, nachdem auf Grund der Kontrolle ein Fehler gefunden und beseitigt wurde, verkleinert sich die Variabilität der Mittelwerte. Man kann dann die Kontrollgrenzen enger ziehen und erreicht gar bald einen Zustand, bei dem die Kontrollgrenzen so eng sind, wie man dies nur wünschen

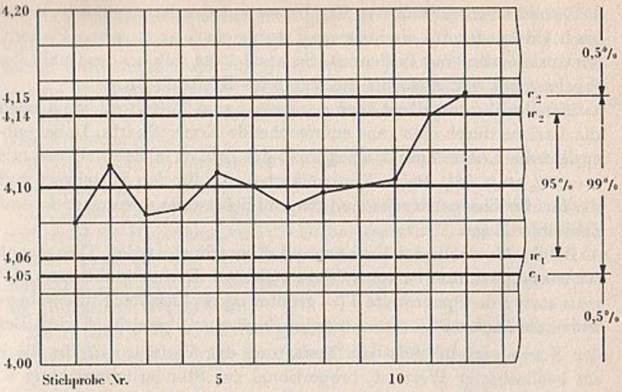


Abbildung 80.1. Kontrollkarte für den Mittelwert (Innendurchmesser eines Werkstücks). Sollwert 4,10 mm, Kontrollgrenzen $c_1 = 4{,}05$ mm, $c_2 = 4{,}15$ mm. Warngrenzen $w_1 = 4{,}06$ mm, $w_2 = 4{,}14$ mm

Tabelle 80.1. Meßwerte zu Abb. 80.1
(G. H. FISCHER, Telefunken-Zeitung **36**, 1963, 43)

Stichprobe Nr.	Stichprobenwerte					$\bar{x}$	s	$\tilde{s}$
1	4,06	4,08	4,08	4,08	4,10	4,080	0,014	0,04
2	4,10	4,10	4,12	4,12	4,12	4,112	0,011	0,02
3	4,06	4,06	4,08	4,10	4,12	4,084	0,026	0,06
4	4,06	4,08	4,08	4,10	4,12	4,088	0,023	0,06
5	4,08	4,10	4,12	4,12	4,12	4,108	0,018	0,04
6	4,08	4,10	4,10	4,10	4,12	4,100	0,014	0,04
7	4,06	4,08	4,08	4,10	4,12	4,088	0,023	0,06
8	4,08	4,08	4,10	4,10	4,12	4,096	0,017	0,04
9	4,06	4,08	4,10	4,12	4,14	4,100	0,032	0,08
10	4,06	4,08	4,10	4,12	4,16	4,104	0,038	0,10
11	4,12	4,14	4,14	4,14	4,16	4,140	0,014	0,04
12	4,14	4,14	4,16	4,16	4,16	4,152	0,011	0,02

kann, ohne daß man den Fabrikationsprozeß von Grund aus umwandelt. So führt die Qualitätsregelung meist zur Qualitätssteigerung. Dies ist verständlich, denn durch die Kontrolle gewinnt man einen vertieften Einblick in den Fabrikationsablauf und seine Schwächen.

Die anschauliche Form der Ergebnisse auf der Kontrollkarte liefert auch Unterlagen für eine bessere Zusammenarbeit zwischen Planern, Produzierenden und Prüfenden. So handelt es sich bei der Qualitätsregelung um weit mehr als um eine reine Kontrolltätigkeit.

Manchmal kontrolliert man zusätzlich zum Mittelwert auch noch die Varianz durch eine ganz entsprechende Kontrollkarte. Dabei entspricht die Kontrollgrenze sinngemäß der Zahl c^* in Beispiel 79.3, mit $\alpha = 1\%$ statt 5%. Beide Kontrollkarten, die für den Mittelwert und die für die Varianz, druckt man dann üblicherweise untereinander auf denselben Bogen.

Bei der Kontrolle der Varianz muß s^2 berechnet werden. Dies macht für wenig geschultes Personal Schwierigkeiten. Deshalb wählt man oftmals statt s die Spannweite $\tilde{s}$ (= größter minus kleinster Stichprobenwert), die auch mit R (von *range* im Englischen) bezeichnet wird. Bei der Normalverteilung ist die Erwartung der Variablen $\tilde{S}$, für die $\tilde{s}$ ein beobachteter Wert ist, proportional zur Standardabweichung σ, wie man zeigen kann. Es ist also $E(\tilde{S}) = \lambda_n \sigma$. Dabei hat λ_n die folgenden Werte:

n	2	3	4	5	6	7	8	9	10
λ_n	1,1	1,7	2,1	2,3	2,5	2,7	2,9	3,0	3,1

Es leuchtet ein, daß man die Ersparnis an Rechenarbeit mit Verlust an Information bezahlen muß, denn an s sind alle Stichprobenwerte beteiligt, während $\tilde{s}$ nur vom kleinsten und vom größten Wert abhängt. s ist eine bessere Schätzung für σ als $\tilde{s}$ und sollte verwendet werden, wenn die einzelnen Messungen relativ teuer sind.

81 Vergleich der Mittelwerte zweier Normalverteilungen

Getestet werden soll, ob die Mittelwerte μ_1 und μ_2 zweier normalverteilter Grundgesamtheiten gleich oder verschieden sind. Die Varianz braucht nicht bekannt zu sein, wird aber als gleich vorausgesetzt. (Wie wir diese Voraussetzung notfalls testen können, lernen wir im übernächsten Abschnitt. Was man machen kann, wenn die Voraussetzung nicht erfüllt ist, überlegen wir uns am Schluß des vorliegenden Abschnittes.)

Für den Test brauchen wir zwei Stichproben. Die folgenden beiden Fälle sind praktisch besonders wichtig:

1. Die beiden Stichproben sind gleich groß. Je ein Wert der einen und je ein Wert der anderen gehören zusammen, weil sie von demselben Individuum stammen. (Beispiel: An jedem Werkstück werden je zwei Meßwerte mit zwei verschiedenen Meßgeräten gewonnen. Je ein Meßwert stammt von jedem der beiden Augen eines Patienten oder Versuchstiers.)

2. Die beiden Stichproben sind unabhängig und nicht notwendigerweise gleich groß.

Im ersten Falle bildet man die Differenzen entsprechender Werte und testet die Hypothese, daß die Grundgesamtheit, aus der die Differenzen stammen, den Mittelwert 0 hat. Hierzu benutzt man die Methode in Abschn. 79. (Beispiel siehe Aufgabe 79.7; dort sind gleich die Differenzen angegeben.)

Im zweiten Falle testet man gemäß Tab. 81.1. Die zugehörige theoretische Begründung folgt im nächsten Abschnitt.

Tabelle 81.1. Test der Hypothese $\mu_1 = \mu_2$ gegen die Alternative $\mu_1 > \mu_2$ für Normalverteilungen mit gleicher Varianz unter Benutzung unabhängiger Stichproben

1. Schritt. Man wähle eine Signifikanzzahl α (5%, 1% oder dgl.).

2. Schritt. Man bestimme die Zahl c aus

$$P(T \leqq c) = 1 - \alpha \tag{81.1}$$

und Tafel 8 in Anhang 5 mit $n_1 + n_2 - 2$ Freiheitsgraden.

3. Schritt. Man berechne die Mittelwerte $\bar{x}$ und $\bar{y}$ und die Varianzen s_1^2 und s_2^2 der Stichproben

$$x_1, \ldots, x_{n_1} \qquad \text{und} \qquad y_1, \ldots, y_{n_2}.$$

4. Schritt. Man berechne

$$t_0 = \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}} \, \frac{\bar{x} - \bar{y}}{\sqrt{(n_1 - 1)\, s_1^2 + (n_2 - 1)\, s_2^2}} \,. \tag{81.2}$$

Ist $t_0 \leqq c$, so wird die Hypothese angenommen. Ist $t_0 > c$, so wird sie verworfen.

Hat man bei der Planung eines Experimentes die Wahl, sollte man den ersten Weg einschlagen, weil dabei die Variabilität zwischen den Versuchsobjekten eliminiert wird, man also sinnvollere Ergebnisse erhält.

Bei der 1. Art des Lesens wird die Bohne mit dem Daumen und Zeigefinger erfaßt und unter Drehung des Handgelenks in die hohle Hand befördert, die entleert wird, wenn sie voll ist. Bei der 2. Art wird die Bohne erfaßt wie zuvor und durch eine Bewegung des Handgelenks zur Seite geschleudert. Ist die eine Art besser als die andere? Es werde dabei vorausgesetzt, daß die Stichproben aus normalverteilten Grundgesamtheiten stammen. In der Erdnußindustrie wird übrigens ausschließlich die 1. Art benutzt.

Teilnehmer Nr.	1	2	3	4	5	6	7	8	9
1. Art	214	253	276	215	238	221	210	229	269
2. Art	281	279	260	230	267	253	205	265	299

81.6 Erntet man größere Kartoffeln, wenn man die Saatkartoffeln halbiert? Die Tabellen zeigen die Anzahl geernteter großer Kartoffeln (über 7 cm) in Prozent (G. Jähnl, Veröff. Bundesanst. f. alpine Landwirtsch. Admont 1952, 87)

Sorte „Böhms Allerfrüheste"

Geschnitten	6,3	10,0	6,0	7,5
Ungeschnitten	1,5	3,5	1,5	3,5

Sorte „Frühbote"

Geschnitten	10,0	3,5	8,0	3,0
Ungeschnitten	18,0	12,0	12,0	15,8

81.7 Das Umbrechen, Ansäen und Düngen führt bei einer Borstgrasfläche im Gebirge zu einer wesentlichen Steigerung des Heuertrages. Die Frage war, ob sich der Ertrag [Doppelzentner/ha] auch ohne Düngung allein durch Umbrechen und Aussäen steigern läßt oder dann vielleicht sogar zurückgeht. (F. Zürn, Veröff. Bundesanst. f. alpine Landwirtsch. Admont 1953, 36.) A = Nicht umgebrochen, ungedüngt. B = Umgebrochen, ungedüngt

A	20,3	13,7	13,8	12,4	16,0	12,5	19,4	30,1	35,6	24,6
B	19,9	13,2	8,8	11,7	14,6	14,1	21,8	25,5	35,1	25,5

81.8 Man konstruiere zwei möglichst einfache Stichproben, bei deren Verwendung die Hypothese, daß die zugehörigen Grundgesamtheiten denselben Mittelwert haben, angenommen wird, trotzdem sich die Mittelwerte der Stichproben erheblich unterscheiden.

81.9 Ist der Unterschied der Gewichtszunahme [kg in 4 Monaten] zweier Gruppen von je 7 Schweinen signifikant? Gruppe I erhielt Fischvollmehlzufütterung, Gruppe II Fischmehlzufütterung. Man wollte wissen, wie man Abfälle bei der Fischverarbeitung am besten verwerten kann. (E. Kraack, Ernährungsforschung 7, 1962, 145.)

Gruppe I	33	66	26	43	46	55	54
Gruppe II	53	53	37	73	58	61	38

82 Theorie zu Abschnitt 81

Aus der Definition der χ^2-Verteilung ergibt sich unmittelbar der

Satz 82.1. *Haben zwei unabhängige Zufallsvariable Z_1 und Z_2 eine χ^2-Verteilung mit m bzw. n Freiheitsgraden, so hat ihre Summe $Z_1 + Z_2$ eine χ^2-Verteilung mit $m + n$ Freiheitsgraden.*

Die Grundlage des Tests bildet der

Satz 82.2. *$X_1, \ldots, X_{n_1}$, $Y_1, \ldots, Y_{n_2}$ seien unabhängige normalverteilte Zufallsvariable mit der gleichen Varianz σ^2. Die X_j mögen den Mittelwert μ_1 besitzen und die Y_j den Mittelwert μ_2. Es sei*

$$\text{(82.1)} \qquad \overline{X} = \frac{1}{n_1}(X_1 + \cdots + X_{n_1}), \quad \overline{Y} = \frac{1}{n_2}(Y_1 + \cdots + Y_{n_2})$$

und weiterhin

$$\text{(82.2)} \qquad S_1^2 = \frac{1}{n_1 - 1}\sum_{j=1}^{n_1}(X_j - \overline{X})^2, \quad S_2^2 = \frac{1}{n_2 - 1}\sum_{j=1}^{n_2}(Y_j - Y)^2.$$

Dann hat die Zufallsvariable

$$\text{(82.3)} \qquad T = \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}} \cdot \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{(n_1 - 1)\, S_1^2 + (n_2 - 1)\, S_2^2}}$$

eine t-Verteilung mit $n_1 + n_2 - 2$ Freiheitsgraden.

Beweis. Gemäß Satz 73.1 hat die Zufallsvariable

$$V_1 = \frac{1}{\sigma^2}(n_1 - 1)\, S_1^2 \qquad \text{bzw.} \qquad V_2 = \frac{1}{\sigma^2}(n_2 - 1)\, S_2^2$$

eine χ^2-Verteilung mit $n_1 - 1$ bzw. $n_2 - 1$ Freiheitsgraden. Diese Variablen sind unabhängig, weil die X_j und Y_j unabhängig sind. Also hat ihre Summe $V = V_1 + V_2$ wegen Satz 82.1 eine χ^2-Verteilung mit $n_1 + n_2 - 2$ Freiheitsgraden. Wegen Satz 71.2 und 71.3 sind $\overline{X} - \mu_1$ und $-(\overline{Y} - \mu_2)$ normalverteilt mit dem Mittelwert 0 und der Varianz σ^2/n_1 bzw. σ^2/n_2. Diese Variablen sind unabhängig. Also ist ihre Summe

$$Z^* = \overline{X} - \mu_1 - (\overline{Y} - \mu_2) = \overline{X} - \overline{Y} - (\mu_1 - \mu_2)$$

wegen Satz 71.1 normalverteilt mit dem Mittelwert 0 und der Varianz

$$\sigma^{*2} = \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} = \frac{n_1 + n_2}{n_1 n_2}\sigma^2,$$

und die Variable $Z = Z^*/\sigma^*$ ist normalverteilt mit dem Mittelwert 0 und der Varianz 1. Ähnlich wie im Beweis des Satzes 73.1 ergibt sich, daß Z und V unabhängig sind, also völlig den Variablen X bzw. Y in

Ersetzt man die Alternative in Tab. 81.1 durch die Alternative $\mu_1 < \mu_2$ bzw. $\mu_1 \neq \mu_2$, so ist (81.1) durch

$$P(T \leqq c) = \alpha$$

bzw.

$$P(T \leqq c_1) = \frac{\alpha}{2} \qquad \text{und} \qquad P(T \leqq c_2) = 1 - \frac{\alpha}{2}$$

zu ersetzen.

Beispiel 81.1. An der Gebärklinik in Bologna wurde bei $n_1 = 288$ neugeborenen ausgetragenen Knaben das Durchschnittsgewicht $\bar{x} = 3300$ Gramm und die Standardabweichung $s_1 = 470$ Gramm festgestellt. Entsprechend erhielt man bei einer Stichprobe von $n_2 = 269$ Mädchen das Durchschnittsgewicht $\bar{y} = 3050$ Gramm und die Standardabweichung $s_2 = 460$ Gramm (C. GINI, Sulla variabilità dei due sessi, Cagliari 1910). Man teste die Hypothese, daß Knaben und Mädchen bei der Geburt gleich schwer sind, gegen die Alternative, daß Knaben schwerer als Mädchen sind.

1. Schritt. Wir wählen die Signifikanzzahl $\alpha = 1\%$.

2. Schritt. Tafel 8 in Anhang 5 enthält keine Spalte für

$$n_1 + n_2 - 2 = 555 \text{ Freiheitsgrade}.$$

Aus den Zahlenwerten der Tafel ist aber klar, daß die Gleichung

$$P(T \leqq c) = 1 - \alpha = 0{,}99$$

die Lösung $c = 2{,}33$ hat.

3. und 4. Schritt. Aus den beiden Stichproben folgt

$$t_0 = \sqrt{\frac{288 \cdot 269 \cdot 555}{557}} \, \frac{3300 - 3050}{\sqrt{287 \cdot 470^2 + 268 \cdot 460^2}} = 6{,}34.$$

Wegen $t_0 > c$ verwerfen wir die Hypothese und nehmen an, daß männliche ausgetragene Neugeborene schwerer als weibliche sind.

Für gleichen Stichprobenumfang $n_1 = n_2 = n$ gewinnt (81.2) die einfache Form

$$\text{(81.3)} \qquad \boxed{t_0 = \sqrt{n} \, \frac{\bar{x} - \bar{y}}{\sqrt{s_1{}^2 + s_2{}^2}}.}$$

Ist die Voraussetzung, daß die betrachteten Grundgesamtheiten dieselbe Varianz haben, nicht erfüllt, so wähle man für den Test gleichgroße Stichproben ($n_1 = n_2 = n$) und dabei n hinreichend groß (praktisch $n > 30$). Dann ist (81.3) ein beobachteter Wert einer bei Richtigkeit der Hypothese näherungsweise normalverteilten Zufallsvariablen mit dem Mittelwert 0 und der Varianz 1. Damit kann man den Test durchführen.

Im Gegensatz zu dem im übernächsten Abschnitt folgenden Test ist der vorliegende Test relativ unempfindlich gegen Abweichungen von der Normalität der Verteilungen, wie man zeigen kann.

Aufgaben zu Abschnitt 81

Voraussetzung: Die Grundgesamtheiten, aus denen die Stichproben stammen, sind normalverteilt und haben jeweils dieselbe Varianz.

81.1 Drei Tage alter Mörtel bzw. Beton mit demselben Wasserzementwert (0,4) hatte die Druckfestigkeit [kg/cm^2]

357 359 413 bzw. 346 302 358

(A. Meyer, Beton-Zeitung **29**, 1963, 392). Man teste die Hypothese, daß die zugehörigen Grundgesamtheiten denselben Mittelwert besitzen.

81.2 Lohnt es sich vom wirtschaftlichen Standpunkt aus, Fabrikräume zu klimatisieren? Würde die Arbeitsleistung steigen? Die Tabelle gibt Beobachtungen aus einem nichtklimatisierten englischen Zinnwalzwerk (J. J. B. Worth, Journ. Industr. Eng. **9**, 1958, 252). Man teste die Hypothese, daß die Arbeitsleistung nicht von der Jahreszeit abhängt.

Jahreszeit	Relative Arbeitsleistung			
Sommer	92,2	84,8	97,2	102,8
Winter	107,7	85,7	102,5	102,5

81.3 Zur Überprüfung der Reproduzierbarkeit einer bei Abkühlversuchen benutzten Meßmethode wurden zwei Versuchsreihen $V\,1$ und $V\,2$ angestellt. Ist der Unterschied der Mittelwerte dieser thermoelektrisch gemessenen Temperaturen [°C] signifikant? (M. Röhle, Untersuchung und Weiterentwicklung einer Probe zur Ermittlung des Durchmessereinflusses auf die Festigkeitseigenschaften wärmebehandelter Stähle. Dissertation T. U. Berlin 1962, S. 32.)

$V\,1$	106,9	106,3	107,0	106,0	104,9
$V\,2$	106,5	106,7	106,8	106,1	105,6

81.4 Verbrauchen Forellen in schnell fließendem Wasser mehr Sauerstoff [mm^3 pro Stunde und pro Gramm Lebendgewicht] als in langsam fließendem? (Washbourn, Journ. Exp. Biol. **13**, 1936, 145—147.)

Fluß	Sauerstoffverbrauch									
Schnell	108	122	144	129	107	115	114	97	96	126
Langsam	85	152	83	69	95	87	71	94	83	94

81.5 Die Aschenbrödel-Methode des Auslesens fehlerhafter Stücke mit der Hand spielt insbesondere in der Nahrungsmittelindustrie (Bohnen, Nüsse, Kaffee usw.) eine erhebliche Rolle. J. J. Moder und J. H. Oswalt (Journ. Industr. Eng. **10**, 1959, 217) verglichen den Zeitbedarf bei verschiedenen Arten des Auslesens. Es ergaben sich bei 9 Teilnehmern an dem Test die angegebenen Anzahlen ausgelesener fehlerhafter Stücke (gesprenkelter Bohnen aus weißen).

Abschn. 62 entsprechen. Deshalb hat

$$T = \frac{Z}{\sqrt{V/(n_1 + n_2 - 2)}} \tag{82.4}$$

eine t-Verteilung mit $n_1 + n_2 - 2$ Freiheitsgraden. Setzen wir die Ausdrücke für V und Z ein, so sehen wir, daß (82.4) mit (82.3) identisch ist. Der Satz 82.2 ist damit bewiesen.

Trifft die Hypothese $\mu_1 = \mu_2$ zu, so vereinfacht sich (82.3) etwas. Setzen wir dann für $\overline{X}$, $\overline{Y}$, S_1 und S_2 die beobachteten Werte $\bar{x}$, $\bar{y}$, s_1 bzw. s_2 ein, so erhalten wir (81.2). Dies bedeutet, daß t_0 ein beobachteter Wert von T ist. Die Wahrscheinlichkeit in (81.1) ist groß, weil α klein ist. Also ist die Wahrscheinlichkeit, daß T irgendeinen Wert $t_0 \leqq c$ annimmt, groß und die Wahrscheinlichkeit für irgendeinen Wert $t_0 > c$ klein. Wird trotzdem ein Wert $t_0 > c$ beobachtet, sind wir berechtigt, die Hypothese zu verwerfen. So ergibt sich die Vorschrift in Tab. 81.1.

Verwirft man die Hypothese, so sagt man auch, der Unterschied zwischen den Mittelwerten sei **statistisch gesichert.**

83 Vergleich der Varianzen zweier Normalverteilungen

Ob zwei Normalverteilungen, deren Mittelwerte nicht bekannt zu sein brauchen, gleiche Varianzen haben, kann man testen, wie Tab. 83.1 zeigt. Auch dies ist ein Problem von erheblicher praktischer Bedeutung. Denn die Varianz ist ja oftmals ein Maß für die Gleichmäßigkeit einer Produktion. Kleine Varianz kann z. B. bedeuten, daß sich die

Tabelle 83.1. Test der Hypothese $\sigma_1^2 = \sigma_2^2$ gegen die Alternative $\sigma_1^2 > \sigma_2^2$ für Normalverteilungen unter Benutzung unabhängiger Stichproben

1. Schritt. Man wähle eine Signifikanzzahl α (5% oder 1%).

2. Schritt. Man bestimme die Zahl c aus

$$P(V \leqq c) = 1 - \alpha \tag{83.1}$$

und Tafel 9a oder 9b (s. Anhang 5) der F-Verteilung mit $(n_1 - 1, n_2 - 1)$ Freiheitsgraden.

3. Schritt. Man berechne die Varianzen s_1^2 und s_2^2 der Stichproben

$$x_1, \ldots, x_{n_1} \qquad \text{und} \qquad y_1, \ldots, y_{n_2} \; .$$

4. Schritt. Man berechne $v_0 = s_1^2/s_2^2$. Ist $v_0 \leqq c$, so wird die Hypothese angenommen. Ist $v_0 > c$, so wird sie verworfen.

betreffenden Erzeugnisse bezüglich einer Größe, die von Belang ist (Abmessungen, Gewicht, Lebensdauer oder was es immer sein mag), nur wenig voneinander unterscheiden. Selbst bei Naturprodukten (Tomaten, Apfelsinen usw.) ist die Gleichmäßigkeit der Größe aus Gründen der Verpackung oftmals wesentlicher als ein großes Gewicht.

Beispiel 83.1. Bei je 16 Schrauben mit gewalztem bzw. gefrästem Gewinde wurde der Flankendurchmesser bestimmt. Es ergab sich der Mittelwert $\bar{x} = 23{,}189$ mm bzw. $\bar{y} = 23{,}277$ mm und die Varianz $s_1^2 = 0{,}001382$ mm² bzw. $s_2^2 = 0{,}000433$ mm² (J. WILHELM, Versuche zur Bestimmung des Berührungsfehlers von Gewinden. Dissertation, Dresden 1932). Unter der Annahme, daß Normalverteilungen vorliegen, teste man die Hypothese $\sigma_1^2 = \sigma_2^2$ gegen die Alternative $\sigma_1^2 > \sigma_2^2$.

1. Schritt. Wir wählen die Signifikanzzahl $\alpha = 5\%$.

2. Schritt. Es ist $n_1 - 1 = 15$ und $n_2 - 1 = 15$. Für (15, 15) Freiheitsgrade entnehmen wir der Tafel 9a als Lösung der Gleichung $P(V \leqq c) = 0{,}95$ die Zahl $c = 2{,}40$.

3. und 4. Schritt. Es ist $v_0 = 0{,}001382/0{,}000433 = 3{,}19$, also $v_0 > c$. Wir verwerfen deshalb die Hypothese zugunsten der Alternative. Der Unterschied zwischen den Varianzen ist also signifikant. Demgemäß zeigt der Test, daß die gefrästen Schrauben als qualitativ höherwertig anzusehen sind.

Wir wollen nun den Test theoretisch begründen. Wie sich zeigen läßt, gilt der folgende

Satz 83.1. *V_1 und V_2 seien unabhängige Zufallsvariable, die Chi-Quadrat-Verteilungen mit m bzw. n Freiheitsgraden besitzen. Dann hat die Zufallsvariable*

$$V = \frac{V_1/m}{V_2/n}$$

die Verteilungsfunktion $F(x) = 0$ für negative x und

$$F(x) = P(V \leqq x) = \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} m^{\frac{m}{2}} n^{\frac{n}{2}} \int_0^x \frac{t^{(m-2)/2}}{(mt+n)^{(m+n)/2}}\, dt \tag{83.2}$$

für positive x. Hierbei ist $\Gamma(\alpha)$ die Gammafunktion (vgl. Abschn. 60).

Die durch (83.2) gegebene Verteilung heißt die ***F*-Verteilung** mit (m, n) Freiheitsgraden. Sie wurde von R. A. FISHER (Proc. Int. Math. Congress, Toronto 1924) eingeführt.

Aus Satz 73.1 und 83.1 folgt der

Satz 83.2. *$X_1, \ldots, X_{n_1}, Y_1, \ldots, Y_{n_2}$ seien unabhängige normalverteilte Zufallsvariable. Die X_i mögen die Varianz σ_1^2 und die Y_j die Varianz*

σ_2^2 besitzen. Dann hat die Zufallsvariable

$$V = \left(\frac{S_1^2}{\sigma_1^2}\right) \Big/ \left(\frac{S_2^2}{\sigma_2^2}\right) \tag{83.3}$$

mit S_1^2 und S_2^2 gemäß (82.2) *eine F-Verteilung mit* $(n_1 - 1, n_2 - 1)$ *Freiheitsgraden.*

Beweis. Die Zufallsvariable

$$\frac{(n_1 - 1)\, S_1^2}{\sigma_1^2} = \frac{1}{\sigma_1^2} \sum_{j=1}^{n_1} (X_j - \bar{X})^2 \quad \text{bzw.} \quad \frac{(n_2 - 1)\, S_2^2}{\sigma_2^2} = \frac{1}{\sigma_2^2} \sum_{j=1}^{n_2} (Y_j - \bar{Y})^2$$

hat gemäß Satz 73.1 eine Chi-Quadrat-Verteilung mit $n_1 - 1$ bzw. $n_2 - 1$ Freiheitsgraden. Damit folgt Satz 83.2 aus Satz 83.1.

Ist die Hypothese $\sigma_1^2 = \sigma_2^2$ richtig, so reduziert sich (83.3) auf

$$V_0 = S_1^2 / S_2^2 . \tag{83.4}$$

v_0 in Tab. 83.1 ist ein beobachteter Wert dieser Variablen. Die Wahrscheinlichkeit (83.1) ist groß, weil α klein ist. Ist die Hypothese richtig, so ist es also sehr wahrscheinlich, daß V_0 irgendeinen Wert annimmt, der kleiner als c oder höchstens gleich c ist. Trifft dies zu, d. h. ist $v_0 \leqq c$, so nehmen wir die Hypothese an. Hingegen ist die Wahrscheinlichkeit α, mit der V_0 bei Richtigkeit der Hypothese irgendeinen Wert v_0 größer als c annimmt, recht gering. Trifft dieses Ereignis trotzdem ein, so dürfen wir annehmen, daß unsere Hypothese nicht richtig ist, und verwerfen sie.

Man kann zeigen, daß der betrachtete Test, im Gegensatz zu dem in Abschn. 81, empfindlich gegen Abweichungen von der Normalität ist. Will man bei Verteilungen, die stark von der Normalverteilung abweichen, testen, ob die Varianzen gleich sind, so bilde man aus der ersten Stichprobe die Werte $x_i^* = |x_i - \bar{x}|$, aus der zweiten $y_j^* = |y_j - \bar{y}|$. Die Mittelwerte der zugehörigen Zufallsvariablen sind proportional zu σ_1 bzw. σ_2, wie sich zeigen läßt. Also kann man die Hypothese $\sigma_1 = \sigma_2$ testen, indem man testet, ob die genannten Variablen denselben Mittelwert besitzen. Letzteres läßt sich nach dem in Abschn. 81 besprochenen Verfahren tun, weil dieses gegen Abweichungen von der Normalität relativ unempfindlich ist.

Aufgaben zu Abschnitt 83

Bei den folgenden Aufgaben setze man voraus, daß es sich um Stichproben aus Normalverteilungen handelt.

83.1 Im Rahmen einer umfangreichen Staubuntersuchung in Abgasen durch die British Coal Utilisation Research Association kam die Vermutung auf, daß das Staubgewicht [mg] in einem gewissen Teil A des untersuchten Rohrsystems stärker variiere als in einem anderen Teil B. Entsprechende Messungen ergaben (S. BROADBENT, Appl. Statistics **3**, 1954, 42):

A	75	20	70	70	85	90	100	40	35	65	90	35
B	20	35	55	50	65	40						

Ist der Unterschied zwischen den Varianzen signifikant?

83.2 Die folgenden Messungen wurden in der Schweiz gemacht, um Unterlagen bei der richtigen Gestaltung von Arbeitsplätzen zu erhalten (E. GRANDJEAN u. U. BURANDT, Industr. Organisation **31**, 1962, 240). Sind die Unterschiede zwischen den Mittelwerten und zwischen den Standardabweichungen signifikant?

	Mittelwerte für		Standardabweichungen	
	500 Männer	508 Frauen	Männer	Frauen
Körperlänge	169,0	158,8	9,5	10,5
Armlänge	70,4	63,6	4,7	4,5
Schulterbreite	43,5	41,2	2,5	2,7

83.3 Damit man den Test in Aufgabe 81.4 durchführen konnte, mußte man die Annahme machen, der Unterschied der beiden Varianzen sei nicht signifikant. Man teste diese Annahme.

83.4 Hängt die Varianz der Biegezugfestigkeit [kp/cm²] von Gießharzbeton in signifikanter Weise von der Körnigkeit des Zuschlages ab? (Werte aus einer Untersuchung der mechanischen Eigenschaften des Gießharzbetons von G. FRANZ und R. BOSSLER, Beton-Zeitschr. **29**, 1963, 411. Der grobe Zuschlag ist Rheinkiessand der Körnung 0/30. Der feine Zuschlag ist Rheinsand der Körnung 0/7.)

Grober Zuschlag	203	229	215	220	223	233	208	228	209
Feiner Zuschlag	181	207	181	173	165	190	185	184	182

83.5 Besteht in Aufgabe 83.4 ein signifikanter Unterschied zwischen den Mittelwerten?

83.6 Die nachstehende Tabelle zeigt Differenzen [°C] zwischen Abstich- und Gießtemperaturen bei Grauguß (K. SÄUBERLICH, J. GÜNTHER u. G. BRÜCKNER, Gießereitechnik **8**, 1962, 111). Besteht zwischen den beiden Pfannen, die mit verschiedenem Eisen gefüllt sind, ein signifikanter Unterschied hinsichtlich der Varianz?

Pfanne Nr.	Temperaturdifferenz [°C]						
1	50	90	100	90	110	80	
2	110	110	110	110	120	130	120

83.7 Die nachstehende Tabelle zeigt Anzahlen von Fadenbrüchen [pro 100 engl. Pfund Material] bei zwei verschiedenen Drehungen/Zoll. (R. E. PEAKE, Appl. Statistics **2**, 1953, 190.) Ist der Unterschied der Varianzen signifikant?

Drehungen/Zoll	Fadenbrüche/100 engl. Pfund							
1,69	6,0	9,7	7,4	11,5	17,9	11,9	10,2	7,8
1,78	6,4	8,3	7,9	8,8	10,1	11,5	8,7	9,7

Kapitel 15

Tests für Verteilungsfunktionen

Bisher haben wir uns um Näherungswerte für Parameter und um die Güte solcher Näherungen gekümmert. Die Art der Verteilung wurde dabei als bekannt vorausgesetzt oder sie spielte in gewissen Fällen überhaupt keine Rolle.

Im vorliegenden Kapitel werden wir sehen, wie man von der Stichprobenverteilung auf die Verteilung der Grundgesamtheit schließen kann. In den meisten praktischen Fällen handelt es sich darum, eine Vermutung über die Art der Verteilung der Grundgesamtheit unter Zuhilfenahme einer Stichprobe entweder zu bestätigen oder zu widerlegen. Eine solche Vermutung entspringt z. B. aus der Natur des betreffenden Experimentes, aus früheren Erfahrungen oder aus der graphischen Darstellung einer genügend großen Stichprobe.

Wir wollen also die Hypothese testen, daß eine Zufallsvariable X, die uns gerade interessiert, eine bestimmte Verteilungsfunktion $F(x)$ besitzt.

Das empirische Analogon dieser Funktion $F(x)$ ist offenbar die Verteilungsfunktion $\tilde{F}(x)$ einer Stichprobe aus der betreffenden Grundgesamtheit. $\tilde{F}(x)$ ist also eine Näherungsfunktion für $F(x)$, wobei die „Güte" der Näherung natürlich vom Stichprobenumfang abhängt. Grob gesprochen wird man die Hypothese, $F(x)$ sei die richtige Verteilungsfunktion, annehmen, falls $\tilde{F}(x)$ die Funktion $F(x)$ „befriedigend genau" annähert. Hingegen wird man die Hypothese verwerfen, falls die Übereinstimmung zwischen $\tilde{F}(x)$ und $F(x)$ „schlecht" ist.

Es ist klar, daß man, um eine solche Entscheidung fällen zu können, etwas darüber wissen muß, wie sehr sich $\tilde{F}(x)$ von $F(x)$ unterscheiden kann, falls die Hypothese richtig ist, und bei welcher Größe der Abweichung Zweifel an der Richtigkeit der genannten Hypothese berechtigt sind.

Was wir also zuerst einmal brauchen, ist ein Maß für die Abweichung zwischen $\tilde{F}(x)$ und $F(x)$. Weiterhin müssen wir die Wahrscheinlichkeitsverteilung des Abweichungsmaßes unter der Annahme, die Hypothese sei richtig, kennen. Beobachten wir dann in einem konkreten Fall einen Wert des genannten Maßes, der größer als eine Zahl c ist, und ist die Wahrscheinlichkeit, bei Richtigkeit der Hypothese irgendeinen Wert größer als c zu erhalten, gering, so entsteht ein berechtigter Zweifel an der Richtigkeit der Hypothese, und wir verwerfen sie. Entsprechend haben wir bei einer beobachteten Abweichung, die kleiner als c ist, keinen ausreichenden Grund, die Hypothese zu verwerfen.

Letzteres schließt natürlich nicht aus, daß man andere Hypothesen finden könnte, die nach demselben Verfahren ebenfalls anzunehmen wären. Die ganze Sachlage ähnelt also in vieler Hinsicht derjenigen im vorhergehenden Kapitel. Lediglich die Fragestellung ist anders, indem wir nun eineVerteilung in ihrem Gesamtverlauf beurteilen wollen.

Wir besprechen nacheinander zwei wichtige Testverfahren, den sogenannten Chi-Quadrat-Test von K. PEARSON, der sich für diskrete und auch für stetige Verteilungen eignet, und anschließend daran den KOLMOGOROFF-SMIRNOW-Test für stetige Verteilungen.

84 Chi-Quadrat-Test

Der Grundgedanke des Chi-Quadrat-Tests ist sehr einfach: Man unterteilt die x-Achse in Teilintervalle, berechnet sich aus der hypothetischen Verteilungsfunktion $F(x)$ die zu diesen Intervallen gehörenden Wahrscheinlichkeiten der betreffenden Zufallsvariablen X und vergleicht diese mit den relativen Klassenhäufigkeiten (vgl. Abschn. 6) einer gegebenen Stichprobe. Ist die Diskrepanz zu groß, so wird die Hypothese, $F(x)$ sei die Verteilungsfunktion von X, verworfen. Liegt der Unterschied unterhalb eines gewissen (natürlich vom Stichprobenumfang abhängigen) Wertes, so wird die Hypothese angenommen. Einzelheiten zeigt Tab. 84.1. Dazu ist zu bemerken:

Zum 1. Schritt. I_1 und I_K sind natürlich unendliche Intervalle. Im Falle einer diskreten Verteilung dürfen die Intervallenden nicht mit einer Sprungstelle von $F(x)$ zusammenfallen.

Zum 2. Schritt. Ist jedes $e_j \geqq 5$, so fahre man mit dem Test fort. Andernfalls gehe man erst zu größeren Intervallen über, praktisch am einfachsten dadurch, daß man jedes Intervall mit $e_j < 5$ mit einem Nachbarintervall vereinigt. Ist n so klein, daß das nicht geht, so ist das Ergebnis des Tests mit Vorsicht zu verwenden. Daß jedes e_j

hinreichend groß sein soll und daß beim 5. Schritt die Chi-Quadrat-Verteilung ins Spiel kommt, ist begründet durch den

Satz 84.1. *Unter der Voraussetzung, daß die vorstehend genannte Hypothese zutrifft, hat die dem Wert χ_0^2 in Tab. 84.1 entsprechende Zufallsvariable eine Verteilungsfunktion, die für $n \to \infty$ gegen die Verteilungsfunktion der Chi-Quadrat-Verteilung mit $K - 1$ Freiheitsgraden strebt.*

Einen Beweis dieses Satzes findet man in [2].

Tabelle 84.1. Chi-Quadrat-Test für die Hypothese, daß $F(x)$ die Verteilungsfunktion der Grundgesamtheit ist, der die Stichprobe $x_1, \ldots, x_n$ entstammt

1. Schritt. Man unterteile die x-Achse in K Intervalle $I_1, I_2, \ldots, I_K$, und zwar derart, daß jedes Intervall mindestens 5 Werte der gegebenen Stichprobe $x_1, \ldots, x_n$ enthält. Für jedes Intervall I_j bestimme man die Anzahl b_j der Stichprobenwerte, die in I_j liegen. Ein Wert, der auf eine Intervallgrenze fällt, wird je zur Hälfte in den beiden angrenzenden Intervallen gezählt.

2. Schritt. Aus $F(x)$ berechne man für jedes Intervall I_j die Wahrscheinlichkeit p_j, mit der die betreffende Zufallsvariable X irgendeinen Wert aus I_j annimmt. Daraus berechne man die Anzahl

$$e_j = n p_j$$

der theoretisch in I_j zu erwartenden Stichprobenwerte.

3. Schritt. Man berechne die Abweichung

$$\chi_0^2 = \sum_{j=1}^{K} \frac{(b_j - e_j)^2}{e_j}. \tag{84.1}$$

4. Schritt. Man wähle eine Signifikanzzahl α (5%, 1% oder dgl.).

5. Schritt. Man bestimme die Lösung c der Gleichung

$$P(\chi^2 \leqq c) = 1 - \alpha$$

aus der Tafel 6 der Chi-Quadrat-Verteilung mit $K - 1$ Freiheitsgraden. Ist $\chi_0^2 \leqq c$, so wird die Hypothese angenommen. Ist $\chi_0^2 > c$, so wird sie verworfen.

Zum 5. Schritt. Enthält die zu testende Verteilungsfunktion r Parameter, deren genaue Werte nicht bekannt sind, so bestimme man zuerst Näherungen für diese Parameter mit Hilfe der Maximum-Likelihood-Methode (siehe Abschn. 67) und benutze dann die Chi-Quadrat-Verteilung mit $K - r - 1$ (statt $K - 1$) Freiheitsgraden.

Diese Regel folgt aus einem Satz von R. A. FISHER, dessen ziemlich komplizierten Beweis man in [2], S. 427–434, findet.

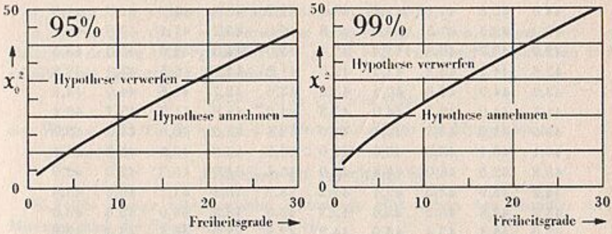


Abbildung 84.1. Graphische Darstellung der Tafel 6 der Chi-Quadrat-Verteilung

85 Zwei Beispiele zum Chi-Quadrat-Test

Beispiel 85.1 (Mendelsche Gesetze). G. MENDEL (Verh. Naturf. Verein. Brünn 4, 1865) erhielt bei einem seiner berühmten Kreuzungsversuche an 10 Erbsenpflanzen insgesamt 355 gelbe und 123 grüne Erbsen. Spricht dies für oder gegen die MENDELsche Theorie, nach der sich im vorliegenden Falle gelb : grün wie 3:1 verhalten müßte?

1. Schritt. Die beiden möglichen Ereignisse bezeichnen wir mit

$$X = 0 \text{ (gelbe Erbse) und } X = 1 \text{ (grüne Erbse)}$$

(oder sonst irgendwie). Wir wählen $K = 2$ Intervalle derart, daß jedes Intervall eines der beiden Ereignisse enthält. Dann ist $b_1 = 355$ und $b_2 = 123$.

2. Schritt. Wegen $n = 355 + 123 = 478$ erhalten wir

$$e_1 = 478 \cdot \frac{3}{4} = 358{,}5 \qquad \text{und} \qquad e_2 = 478 \cdot \frac{1}{4} = 119{,}5.$$

3. Schritt. So wird

$$\chi_0^2 = \frac{(355 - 358{,}5)^2}{358{,}5} + \frac{(123 - 119{,}5)^2}{119{,}5} = 0{,}137.$$

4. Schritt. Wir wählen die Signifikanzzahl $\alpha = 5\%$.

5. Schritt. Für $K - 1 = 1$ Freiheitsgrad hat die Gleichung

$$P(\chi^2 \leqq c) = 1 - \alpha = 0{,}95$$

die Lösung $c = 3{,}84$. Es ist also $\chi_0^2 < c$. Die Hypothese wird angenommen. Die Stichprobe spricht für die Richtigkeit der Theorie.

Beispiel 85.2 (Normalverteilung). Man teste die Hypothese, daß die Stichprobe in Tab. 85.1 einer normalverteilten Grundgesamtheit entstammt.

1. Schritt. Wir unterteilen die x-Achse in $K = 8$ Intervalle, die in Tab. 85.2 angegeben sind, und bestimmen die zugehörigen b_j. Dabei benutzen wir eine Strichliste.

Tabelle 85.1. Zugfestigkeit [in kg/mm²] von 0,3 mm dicken Stahlblechen parallel zur Walzrichtung (A. SLATTENSCHEK und G. SCHNEEWEISS, Maschinenbau u. Wärmewirtsch. **12**, 1957, 4)

44,0	42,8	40,8	41,4	44,4	43,9	42,8	44,0	42,2	44,8
43,3	42,5	43,5	44,7	45,8	42,0	45,2	41,1	43,8	43,8
42,9	43,7	45,8	41,4	42,6	45,0	44,5	41,6	44,3	43,5
43,8	44,4	43,2	42,3	42,0	41,2	44,1	45,5	43,0	39,8
43,2	44,9	42,6	40,1	43,2	43,0	42,7	43,5	44,0	44,8
44,5	44,0	42,7	44,0	42,3	44,2	44,8	41,7	43,7	42,4
43,5	44,3	43,7	45,4	44,6	42,4	45,5	40,8	44,3	43,7
44,1	46,1	45,3	43,6	43,0	46,8	44,8	42,9	45,3	44,1
43,8	42,5	46,0	44,4	42,0	45,4	44,0	45,3	42,0	42,9
44,2	43,9	43,5	42,1	44,2	44,2	43,8	41,7	46,5	43,5
44,5	44,8	45,2	43,6	42,3	43,5	43,5	46,0	43,6	43,0
42,0	41,1	43,4	45,0	44,2	42,8	42,0	46,1	43,4	42,6
46,7	44,5	44,0	44,8	43,3	42,7	43,1	43,8		

Tabelle 85.2. Berechnung zu Beispiel 85.2

x	$\dfrac{x-43{,}62}{1{,}368}$	$\Phi\left(\dfrac{x-43{,}62}{1{,}368}\right)$	p_j	e_j	b_j	A_j
$-\infty \cdots 42$	$-\infty \cdots -1{,}18$	$0{,}0000 \cdots 0{,}1190$	0,1190	15,23	15	0,003
$42 \cdots 42{,}5$	$-1{,}18 \cdots -0{,}82$	$0{,}1190 \cdots 0{,}2061$	0,0871	11,15	11	0,002
$42{,}5 \cdots 43$	$-0{,}82 \cdots -0{,}45$	$0{,}2061 \cdots 0{,}3264$	0,1203	15,40	15	0,010
$43 \cdots 43{,}5$	$-0{,}45 \cdots -0{,}09$	$0{,}3264 \cdots 0{,}4641$	0,1377	17,63	14	0,747
$43{,}5 \cdots 44$	$-0{,}09 \cdots 0{,}28$	$0{,}4641 \cdots 0{,}6103$	0,1462	18,71	22,5	0,768
$44 \cdots 44{,}5$	$0{,}28 \cdots 0{,}64$	$0{,}6103 \cdots 0{,}7389$	0,1286	16,46	19,5	0,561
$44{,}5 \cdots 45$	$0{,}64 \cdots 1{,}01$	$0{,}7389 \cdots 0{,}8438$	0,1049	13,43	12	0,152
$45 \cdots \infty$	$1{,}01 \cdots \infty$	$0{,}8438 \cdots 1{,}0000$	0,1562	19,99	19	0,049
		Summe	1,0000			2,292

2. Schritt. Die genauen Werte der Parameter μ und σ^2 kennen wir nicht. Aus der Stichprobe entnehmen wir die Maximum-Likelihood-Schätzungen [vgl. (68.5) und (68.6)]

$$\mu \approx \bar{x} = 43{,}62 \qquad \text{und} \qquad \sigma \approx \tilde{\sigma} = \sqrt{1{,}8705} = 1{,}368\,.$$

Wir haben demnach zu testen, ob die genannte Grundgesamtheit die Verteilungsfunktion [vgl. (48.3)]

$$F(x) = \Phi\left(\frac{x-43{,}62}{1{,}368}\right) \tag{85.1}$$

besitzt. Tab. 85.2 enthält die Werte dieser Funktion an den Intervallenden. Diese Werte ergeben sich aus Tafel 3a in Anhang 5. Die Differenz der beiden jeweils in einer Zeile stehenden Werte ist gleich p_j, und daraus erhalten wir

$$e_j = 128 p_j\,.$$

3. Schritt. In Tab. 85.2 ist $A_j = (b_j - e_j)^2/e_j$, und die Summe dieser Zahlen liefert

$$\chi_0^2 = 2{,}292 .$$

4. Schritt. Wir wählen die Signifikanzzahl $\alpha = 5\%$.

5. Schritt. Da wir $r = 2$ Parameter abgeschätzt haben, müssen wir in Tafel 6 in Anhang 5 die Spalte für $K - r - 1 = 5$ Freiheitsgrade benutzen und erhalten als Lösung der Gleichung

$$P(\chi^2 \leqq c) = 1 - \alpha = 0{,}95$$

den Wert $c = 11{,}07$. Es ist $\chi_0^2 < c$. Also nehmen wir die Hypothese an.

Aufgaben zu Abschnitt 85

85.1 Die Grazer telephonische Zeitansage gibt neben den Stunden und Minuten auch die Sekunden in der Form

0, 10, 20, 30, 40, 50 Sekunden

an. Ruft man die Zeitansage regellos in längeren Zeitabständen an, so sollte man erwarten, daß diese 6 Zahlen etwa gleich häufig vorkommen. Wird dies durch die folgenden Beobachtungen bestätigt?

Sekundenanzahl	0	10	20	30	40	50
Häufigkeit	16	19	18	17	17	13

85.2 Man teste die Hypothese, daß bei einem Münzenwurf „Kopf" und „Wappen" die gleiche Wahrscheinlichkeit besitzen, unter Benutzung einer Stichprobe von 50 Würfen, darunter 27 „Köpfen".

85.3 Wie oft muß „Kopf" bei 50 Münzenwürfen mindestens auftreten, damit die Hypothese in Aufgabe 85.2 bei Wahl von $\alpha = 5\%$ verworfen wird?

85.4 Kann man den von R. Wolf (Vierteljahresschr. Naturforsch. Ges. Zürich **27**, 1882, 242) benutzten Würfel als regelmäßig ansehen?

Gewürfelte Zahl x	1	2	3	4	5	6
Absolute Häufigkeit	3407	3631	3176	2916	3448	3422

85.5 Schwanken die nachstehenden Geburtenzahlen [in Einheiten von 1000] in den 12 Monaten des Jahres 1959 nur zufallsbedingt oder signifikant? (Stat. Jahrb. f. d. Bundesrep. Deutschland 1961, S. 59)

Monat	J	F	M	A	M	J	J	A	S	O	N	D
Anzahl	80	78	86	82	83	78	79	76	78	76	72	76

85.6 77 Grazer Hörer einer Statistikvorlesung wurden aufgefordert, 3 der 20 Zahlen 11, 12, 13, ..., 30 ganz beliebig auszuwählen und auf einen Zettel zu schreiben. Es ergab sich:

Zahl	11	12	13	14	15	16	17	18	19	20
Häufigkeit	11	10	20	8	13	9	21	9	16	8

Zahl	21	22	23	24	25	26	27	28	29	30
Häufigkeit	12	8	15	10	10	9	12	8	13	9

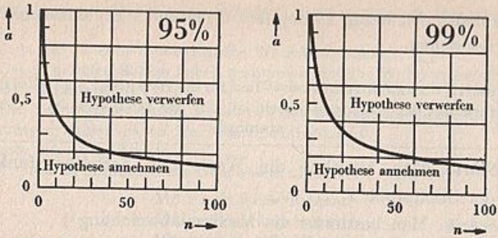


Abbildung 86.2. Graphische Darstellung der Tafel zum KOLMOGOROFF-SMIRNOW-Test (Tafel 7 in Anhang 5)

87 Beispiel zum Kolmogoroff-Smirnow-Test

Beispiel 87.1 (Normalverteilung). Entstammt die Stichprobe in Tab. 4.3 einer Normalverteilung mit dem Mittelwert $\mu = 165{,}05$ cm und der Varianz $\sigma^2 = 34{,}3106$ cm²? [Es ist $\bar{x} = 165{,}05$ und $s = \sqrt{34{,}3106} = 5{,}858$ (vgl. Beispiel 10.2).]

1. Schritt. Die Werte der Verteilungsfunktion $\tilde{F}(x)$ der Stichprobe in Tab. 87.1 erhalten wir durch Summenbildung aus der letzten Spalte der Tab. 4.3.

2. Schritt. Zu testen ist, ob die Grundgesamtheit die Verteilungsfunktion

$$F(x) = \Phi\left(\frac{x - 165{,}05}{5{,}858}\right)$$

hat. Deren Werte entnehmen wir der Tafel 3a in Anhang 5. Dann berechnen wir a_1 und a_2 in Tab. 87.1. In der ersten Zeile ist

$$a_1 = F(153) - 0 = 0{,}02 - 0 = 0{,}02$$
$$a_2 = F(153) - \tilde{F}(153) = 0{,}02 - 0{,}01 = 0{,}01\,.$$

In der zweiten Zeile ist

$$a_1 = F(154) - \tilde{F}(153) = 0{,}03 - 0{,}01 = 0{,}02$$
$$a_2 = F(154) - \tilde{F}(154) = 0{,}03 - 0{,}02 = 0{,}01$$

usw. Die größte aller dieser Zahlen ist $a = 0{,}07$.

3. Schritt. Wir wählen die Signifikanzzahl $\alpha = 5\%$.

4. Schritt. Die Stichprobe hat den Umfang $n = 100$. Aus Tafel 7 in Anhang 5 entnehmen wir als Lösung der Gleichung

$$P(A \leq c) = 1 - \alpha = 0{,}95$$

die Zahl $c = 0{,}134$. Es ist $a < c$. Wir nehmen also die Hypothese an.

Aufgaben zu Abschnitt 87

87.1—87.6 Unter Verwendung des KOLMOGOROFF-SMIRNOW-Testes prüfe man die Hypothese, daß die jeweilige Stichprobe aus einer normalverteilten Grundgesamtheit stammt.

Tabelle 87.1. Zu Beispiel 87.1

x	$\tilde{F}(x)$	$x - 165,05$	$\frac{x - 165,05}{5,858}$	$\Phi\left(\frac{x - 165,05}{5,858}\right)$	a_1	a_2
153	0,01	−12,05	−2,06	0,02	0,02	0,01
154	0,02	−11,05	−1,89	0,03	0,02	0,01
155	0,04	−10,05	−1,72	0,04	0,02	0,00
156	0,07	− 9,05	−1,55	0,06	0,02	0,01
157	0,10	− 8,05	−1,37	0,09	0,02	0,01
158	0,15	− 7,05	−1,20	0,12	0,02	0,03
159	0,21	− 6,05	−1,03	0,15	0,00	0,06
160	0,25	− 5,05	−0,86	0,19	0,02	0,06
161	0,30	− 4,05	−0,69	0,25	0,00	0,05
162	0,37	− 3,05	−0,52	0,30	0,00	**0,07**
163	0,42	− 2,05	−0,35	0,36	0,01	0,06
164	0,47	− 1,05	−0,18	0,43	0,01	0,04
165	0,53	− 0,05	−0,01	0,50	0,03	0,03
166	0,60	0,95	0,16	0,56	0,03	0,04
167	0,65	1,95	0,33	0,63	0,03	0,02
168	0,69	2,95	0,50	0,69	0,04	0,00
169	0,74	3,95	0,67	0,75	0,06	0,01
170	0,79	4,95	0,85	0,80	0,06	0,01
171	0,85	5,95	1,02	0,85	0,06	0,00
172	0,89	6,95	1,19	0,88	0,03	0,01
173	0,92	7,95	1,36	0,91	0,02	0,01
174	0,94	8,95	1,53	0,94	0,02	0,00
175	0,97	9,95	1,70	0,96	0,02	0,01
176	0,98	10,95	1,87	0,97	0,00	0,01
177	0,99	11,95	2,04	0,98	0,00	0,01
178	1,00	12,95	2,21	0,99	0,00	0,01

87.1 Die Stichprobe in Tab. 66.1.
87.2 Die Stichprobe in Tab. 66.2.
87.3 Die Stichprobe in Aufgabe 76.3.
87.4 Die Stichprobe in Beispiel 85.2.
87.5 Masse [g/m²] eines Gewirkes (105 Reihen/5 cm) für Unterbekleidung, hergestellt auf einer kurvenscheibengesteuerten Flachkettenwirkmaschine (H. DIETRICH, H. SCHMIDT, Deutsche Textiltechnik **13**, 1963, 418)

Klassenmitte	96	98	100	102	104	106	108	110	112
Häufigkeit	1	0	1	2	8	19	28	30	41

Klassenmitte	114	116	118	120	122	124	126	128	130
Häufigkeit	66	50	27	8	5	3	0	1	1

87.6 Die Stichprobe in Aufgabe 72.7.

Wäre die Auswahl wirklich ganz zufällig, so müßten folgende Hypothesen zutreffen:

1. Jede der 20 Zahlen hat dieselbe Wahrscheinlichkeit.
2. Die 10 geraden Zahlen haben zusammen dieselbe Wahrscheinlichkeit wie die 10 ungeraden.
3. Die 6 Primzahlen haben zusammen die Wahrscheinlichkeit 0,3, und die übrigen Zahlen haben die Wahrscheinlichkeit 0,7.

Man teste diese Hypothesen.

85.7 Gregor Mendel erhielt bei einem anderen Versuch

315 runde gelbe Erbsen
108 runde grüne Erbsen
101 kantige gelbe Erbsen
32 kantige grüne Erbsen.

Spricht dies für oder gegen die Theorie, nach der sich die vier Zahlen wie 9:3:3:1 verhalten müßten?

85.8 Ist die Grundgesamtheit, der die Stichprobe in Tab. 43.2 entstammt, nach Poisson verteilt?

85.9 Hat die Grundgesamtheit, aus der die Stichprobe in Tab. 43.1 entnommen wurde, eine Poisson-Verteilung?

85.10 Entstammt die Stichprobe in Tab. 43.3 einer nach Poisson verteilten Grundgesamtheit?

85.11 Ist die Anzahl der Jungen pro Familie binomialverteilt? (Beobachtungen an Familien mit 8 Kindern von A. Geissler, Zeitschr. Kgl. Sächs. Stat. Bureau 1889.)

Zu Aufgabe 85.11

Anzahl Jungen	Häufigkeit
0	215
1	1485
2	5331
3	10649
4	14959
5	11929
6	6678
7	2092
8	342

85.12 Entstammt die Stichprobe in Tab. 66.1 einer normalverteilten Grundgesamtheit?

86 Kolmogoroff-Smirnow-Test

Der Kolmogoroff-Smirnow-Test eignet sich nur für stetige Verteilungen. Wie zuvor handelt es sich darum, die Hypothese zu testen, daß eine Funktion $F(x)$ die Verteilungsfunktion einer Grundgesamtheit ist, aus der eine Stichprobe $x_1, \ldots, x_n$ entnommen wurde. Wie man vorgeht, zeigt Tab. 86.1.

Die Größe a ist dabei übrigens ziemlich einfach zu bestimmen, denn $\tilde{F}(x)$ ist eine Treppenfunktion (= stückweise konstante Funktion), und a muß deshalb an einer der Sprungstellen liegen. An jeder Sprungstelle berechne man die beiden Zahlen a_1 und a_2, die in der schematischen Abb. 86.1 eingetragen sind. Die größte aller so erhaltenen Zahlen ist a. Um triviale Irrtümer zu vermeiden, sollte man, solange man noch wenig Übung hat, $\tilde{F}(x)$ und $F(x)$ skizzenhaft graphisch darstellen.

Tabelle 86.1. KOLMOGOROFF-SMIRNOW-Test für die Hypothese, daß $F(x)$ die Verteilungsfunktion der Grundgesamtheit ist, der die Stichprobe $x_1, \ldots, x_n$ entstammt

1. Schritt. Man berechne die Werte der Verteilungsfunktion $\tilde{F}(x)$ der Stichprobe $x_1, \ldots, x_n$.

2. Schritt. Man bestimme die Maximalabweichung[1])

$$a = \max |\tilde{F}(x) - F(x)|$$

zwischen $\tilde{F}(x)$ und $F(x)$.

3. Schritt. Man wähle eine Signifikanzzahl α (5%, 1% oder dgl.).

4. Schritt. Man bestimme die Lösung c der Gleichung

$$P(A \leqq c) = 1 - \alpha \tag{86.1}$$

aus der dem Stichprobenumfang n entsprechenden Zeile der Tafel 7 in Anhang 5. Ist $a \leqq c$, so wird die Hypothese angenommen. Ist $a > c$, so wird sie verworfen.

Die Tafel 7 enthält, wie man sich wohl schon denken kann, die Verteilungsfunktion der Zufallsvariablen A, für die a in Tab. 86.1 ein beobachteter Wert ist. KOLMOGOROFF und SMIRNOW haben gezeigt, daß die Verteilung von A gar nicht von der speziellen Form von $F(x)$ abhängt, sondern für alle stetigen Funktionen $F(x)$ dieselbe ist.

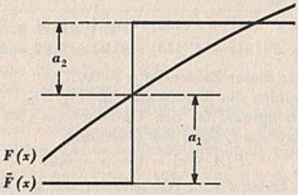


Abbildung 86.1. Zum KOLMOGOROFF-SMIRNOW-Test

[1] Genauer sollten wir schreiben

$$a = \sup_x |\tilde{F}(x) - F(x)|.$$

Kapitel 16

Varianzanalyse

Die monatliche Gewichtszunahme einer Anzahl von Tieren, z. B. Schweinen, variiert von Tier zu Tier, auch wenn die Futterart und -menge und die Lebensbedingungen für alle diese Tiere vollkommen gleich sind. Diesen wohlbekannten Sachverhalt müssen wir als Zufallsvariabilität ansehen, hervorgerufen durch Umstände, die sich unserer genauen Kenntnis und Kontrolle entziehen.

Füttern wir die Tiere unterschiedlich, so überlagert sich der Zufallsvariabilität möglicherweise eine Variabilität, die durch das verschiedene Futter hervorgerufen wird. Wollen wir wissen, ob die Futterart einen Einfluß auf die Gewichtszunahme besitzt, so müssen wir diesen Einfluß vom Zufallseinfluß zu trennen versuchen. Dies ist eine typische Aufgabe der **Varianzanalyse** oder *Varianzzerlegung* (auch **Streuungszerlegung**), genauer: der *einfachen* Varianzanalyse.

Wollen wir zwei Einflüsse gleichzeitig untersuchen (z. B. Futterart *und* Futtermenge), so müssen wir versuchen, die beiden Einflüsse voneinander und außerdem vom Zufallseinfluß zu trennen. Dies ist die Aufgabe der sogenannten *doppelten* Varianzanalyse.

Die Varianzanalyse beruht auf einer rein arithmetischen Zerlegung der „**Quadratsumme**" (= Summe der Quadrate der Abweichungen der Stichprobenwerte vom Mittelwert). Und zwar zerlegt man in eine Summe von Bestandteilen, deren jeder einer bestimmten Variationsursache entspricht. Bei der einfachen Varianzanalyse sind dies zwei Bestandteile. Der eine entspricht dem Zufallseinfluß, der andere dem Einfluß, den man untersuchen will (der Futterart im obigen Beispiel). Bei der doppelten Varianzanalyse zerlegt man in drei Bestandteile usw.

Die Varianzanalyse wurde von R. A. Fisher für biologische Zwecke entwickelt. Sie läßt sich aber auch auf anderen Gebieten anwenden, wie die Textbeispiele und die Aufgaben zeigen.

88 Vergleich der Mittelwerte mehrerer Normalverteilungen

n Versuchstiere werden gewogen und dann zufallsmäßig in r Gruppen eingeteilt. Diesen r Gruppen werden r verschiedene Futtermittel verabreicht. Nach einer gewissen Zeit werden die Tiere wieder gewogen, und die Gewichtszunahme wird festgestellt. Auf diese Weise erhält man eine Stichprobe von insgesamt n Werten (die Gewichtszunahme der einzelnen Tiere), die sich in r **Gruppen** gliedert, etwa

$$x_{11}, x_{12}, \ldots, x_{1n_1} \quad \text{(1. Tiergruppe)}$$

$$x_{21}, x_{22}, \ldots, x_{2n_2} \quad \text{(2. Tiergruppe)}$$

usw. Der erste Index bezeichnet die Gruppe, der zweite die Nummer des Tieres in der Gruppe. Die 1. Gruppe besteht aus n_1 Tieren, die 2. Gruppe aus n_2 Tieren usw. Natürlich ist hierbei

$$n_1 + n_2 + \cdots + n_r = n.$$

Geprüft werden soll, ob hinsichtlich der mittleren Gewichtszunahme bei den r Gruppen signifikante (d. h. durch die unterschiedliche Fütterung hervorgerufene) oder bloß zufallsbedingte Unterschiede bestehen. Im letzteren Falle wäre es, vom Standpunkt der Gewichtszunahme aus betrachtet, gleichgültig, mit welchem der genannten Futtermittel man mästet.

Wir behandeln dieses Problem unter der Annahme, daß die r Gruppen von Zahlen aus r normalverteilten Grundgesamtheiten entstammen, die alle dieselbe Varianz σ^2 besitzen. σ^2 braucht nicht bekannt zu sein. Geprüft werden soll, ob die Mittelwerte $\mu_1, \ldots, \mu_r$ der genannten Grundgesamtheiten ebenfalls übereinstimmen.

Wir testen demgemäß die Hypothese, daß alle diese r Mittelwerte gleich sind. (Für $r = 2$ wurde diese Aufgabe in Abschn. 81 behandelt.)

Statt der gewählten Einkleidung hätten wir ebensogut eine andere wählen können, bei der eine Stichprobe in natürlicher Weise in Gruppen gegliedert werden kann. Beispiele sind die Ernteerträge bei verschiedener Düngung, Eigenschaften von Rohstoffen bei verschiedener Vorbehandlung, Zeitbedarf verschiedener Arbeiter für eine bestimmte Arbeit und so weiter.

Das gestellte Problem lösen wir mit Hilfe der Varianzanalyse. Der Grundgedanke ist einfach: Wir zerlegen die „Quadratsumme" (= Summe von Quadraten)

$$q = \sum_{i=1}^{r} \sum_{k=1}^{n_i} (x_{ik} - \bar{x})^2 \tag{88.1}$$

in zwei Bestandteile q_1 und q_2,

$$\boxed{q = q_1 + q_2\,,}$$

von denen der erste von der Streuung zwischen den Gruppen und der zweite von der Streuung innerhalb jeder Gruppe herrührt. Diese beiden Bestandteile werden dann miteinander verglichen. Wie man vorgeht, zeigt Tab. 88.1. Dabei hat es sich eingebürgert, die für den Test wich-

Tabelle 88.1. Test der Hypothese, daß die normalverteilten Grundgesamtheiten gleicher Varianz, aus denen die im Text genannten r Gruppen stammen, alle denselben Mittelwert besitzen. (Vereinfachte Rechnung siehe im nächsten Abschnitt)

1. Schritt. Man berechne die r Mittelwerte $\bar{x}_1, \ldots, \bar{x}_r$ der Gruppen. Hierbei ist natürlich

$$\bar{x}_i = \frac{1}{n_i}(x_{i1} + x_{i2} + \cdots + x_{in_i}). \tag{88.2}$$

Weiterhin berechne man den Mittelwert der gesamten Stichprobe, also

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{r}\sum_{k=1}^{n_i} x_{ik} = \frac{1}{n}\sum_{i=1}^{r} n_i \bar{x}_i. \tag{88.3}$$

2. Schritt. Man berechne die „*Quadratsumme zwischen den Mittelwerten der Gruppen*“, gegeben durch

$$q_1 = \sum_{i=1}^{r} n_i(\bar{x}_i - \bar{x})^2, \tag{88.4}$$

und die „*Quadratsumme innerhalb der Gruppen*“, gegeben durch

$$q_2 = \sum_{i=1}^{r}\sum_{k=1}^{n_i}(x_{ik} - \bar{x}_i)^2. \tag{88.5}$$

Hieraus bilde man den Quotienten

$$v_0 = \frac{q_1/(r-1)}{q_2/(n-r)}. \tag{88.6}$$

3. Schritt. Man wähle eine Signifikanzzahl α (5% oder 1%).

4. Schritt. Man bestimme die Lösung c der Gleichung

$$P(V \leqq c) = 1 - \alpha \tag{88.7}$$

aus der Tafel 9a bzw. 9b der F-Verteilung mit $(r-1, n-r)$ Freiheitsgraden (siehe Anhang 5). Ist $v_0 \leqq c$, so wird die Hypothese $\mu_1 = \mu_2 = \cdots = \mu_r$ angenommen. Ist $v_0 > c$, so wird sie verworfen. Man nimmt dann an, daß die Mittelwerte nicht alle gleich sind.

tigsten Zahlenwerte gemäß Tab. 88.2 anzuordnen. Wesentliche rechnerische Vereinfachungen besprechen wir im nächsten Abschnitt. Die theoretische Begründung des Tests folgt im übernächsten Abschnitt.

Tabelle 88.2. Zur einfachen Varianzanalyse

Variation	Freiheitsgrade	Quadratsumme	Durchschnittsquadrat
Zwischen den Gruppen	$r-1$	q_1	$q_1/(r-1)$
Innerhalb der Gruppen	$n-r$	q_2	$q_2/(n-r)$
Insgesamt	$n-1$	q	

Beispiel 88.1 (Zugfestigkeit von Folien). Es soll festgestellt werden, ob die Zugfestigkeit von Folien aus einer Titanlegierung an allen Stellen dieselbe ist. 4 Folien werden untersucht, und es ergeben sich die in Tab. 88.3 gezeigten Werte (Youden, Industrial and Engin. Chem. 49, 1957, 71).

Tabelle 88.3. Stichprobe in Beispiel 88.1

Meßstelle	Meßwerte			
1. Gruppe (Ecke)	137	142	128	137
2. Gruppe (Mitte)	140	139	117	137
3. Gruppe (Kante)	142	140	133	141

1. Schritt. Die Gruppenmittelwerte sind

$$\bar{x}_1 = 136, \quad \bar{x}_2 = 133{,}25, \quad \bar{x}_3 = 139.$$

Der Mittelwert der gesamten Stichprobe ist

$$\bar{x} = \frac{1}{12}(4\bar{x}_1 + 4\bar{x}_2 + 4\bar{x}_3) = \frac{1}{3}(\bar{x}_1 + \bar{x}_2 + \bar{x}_3) = 136{,}083.$$

2. Schritt. Die Quadratsumme zwischen den Gruppen ist

$$\begin{aligned} q_1 &= 4[(\bar{x}_1 - \bar{x})^2 + (\bar{x}_2 - \bar{x})^2 + (\bar{x}_3 - \bar{x})^2] \\ &= 4[0{,}083^2 + 2{,}833^2 + 2{,}917^2] = 66{,}167. \end{aligned}$$

Die Quadratsumme innerhalb der Gruppen ist

$$q_2 = \sum_{i=1}^{3} \sum_{k=1}^{4} (x_{ik} - \bar{x}_i)^2$$

$$= (137-136)^2 + (142-136)^2 + \cdots + (141-139)^2 = 508{,}75.$$

Hieraus folgt

$$v_0 = \frac{q_1/2}{q_2/9} = \frac{33{,}083}{56{,}528} = 0{,}585.$$

3. Schritt. Wir wählen die Signifikanzzahl $\alpha = 0{,}05$.

4. Schritt. Es ist $r = 3$, $n = 12$, also $r - 1 = 2$, $n - r = 9$. Aus Tafel 9a in Anhang 5 ersehen wir, daß die Gleichung

$$P(V \leqq c) = 0{,}95$$

die Lösung $c = 4{,}26$ hat. Es ist $v_0 < c$. Deshalb wird die Hypothese

$$\mu_1 = \mu_2 = \mu_3$$

angenommen, d. h. auf Grund der vorgelegten Stichprobe nimmt man an, daß die Zugfestigkeit der Folien zwischen den verschiedenen Meßstellen nur zufallsbedingt schwankt, der Unterschied der Meßwerte also nicht signifikant ist.

Tabelle 88.4. Zu Beispiel 88.1

Variation	Freiheitsgrade	Quadratsumme	Durchschnittsquadrat
Zwischen den Gruppen	2	66,167	33,083
Innerhalb der Gruppen	9	508,750	56,528
Insgesamt	11	574,917	

89 Vereinfachte Rechnung

Einer der Gründe für die Beliebtheit der Varianzanalyse ist die Tatsache, daß man die Rechnungen wesentlich einfacher gestalten kann, als wir dies im vorigen Abschnitt getan haben. Um q_1 und q_2 zu erhalten, berechne man der Reihe nach die in Tab. 89.1 angegebenen Größen. Die Formeln in dieser Tabelle ergeben sich aus den ursprünglichen Formeln für q und q_1 so ähnlich wie die Formel (9.1) für die Varianz.

Tabelle 89.1. Rechengang bei der einfachen Varianzanalyse

$$G_i = \sum_{j=1}^{n_i} x_{ij} = \text{Summe der Werte der } i\text{-ten Gruppe}$$

$$G = G_1 + G_2 + \cdots + G_r = \text{Gesamtsumme}$$

$$K = G^2/n \quad \text{(Hilfsgröße)}$$

$$A = \sum_{i=1}^{r} \sum_{j=1}^{n_i} x_{ij}{}^2 = \text{Summe der Quadrate aller Werte}$$

$$q = A - K$$

$$q_1 = \left(\frac{G_1{}^2}{n_1} + \frac{G_2{}^2}{n_2} + \cdots + \frac{G_r{}^2}{n_r}\right) - K$$

$$q_2 = q - q_1$$

Die Rechnung in Beispiel 88.1 hat dann die in Tab. 89.2 gezeigte Form.

Tabelle 89.2. Beispiel 88.1, gerechnet nach dem in Tab. 89.1 gezeigten Schema

i	Stichprobe				G_i	G_i^2
1	137	142	128	137	544	295 936
2	140	139	117	137	533	284 089
3	142	140	133	141	556	309 136
			Summe		$G = 1633$	889 161

$$K = \frac{1633^2}{12} = 222\,224{,}083 \qquad A = 222\,799$$

$$q = 222\,799{,}000 - 222\,224{,}083 = 574{,}917$$

$$q_1 = \frac{889\,161{,}000}{4} - 222\,224{,}083 = 66{,}167$$

$$q_2 = 574{,}917 - 66{,}167 = 508{,}750$$

Natürlich können wir diese Rechnung noch weiter vereinfachen, indem wir

$$x_{ij} = x_{ij}^* + a \tag{89.1}$$

setzen und a dabei so wählen, daß die x_{ij}^* dem Betrage nach möglichst klein ausfallen. Man berechnet dann der Reihe nach die in Tab. 89.3 angegebenen Größen. Der Leser überzeuge sich, daß die Formeln in dieser Tabelle durch Einsetzen von (89.1) aus den vorigen hervorgehen.

Für $a = 130$ hat die Rechnung in Beispiel 88.1 die in Tab. 89.4 gezeigte Form.

Tabelle 89.3. Einfache Varianzanalyse, Rechengang unter Benutzung von (89.1)

$$G_i^* = \sum_{j=1}^{n_i} x_{ij}^*$$

$$G^* = G_1^* + \cdots + G_r^*$$

$$K^* = G^{*2}/n$$

$$A^* = \sum_{i=1}^{r} \sum_{j=1}^{n_i} x_{ij}^{*2}$$

$$q = A^* - K^*$$

$$q_1 = \sum_{i=1}^{r} \frac{1}{n_i} G_i^{*2} - K^*$$

$$q_2 = q - q_1$$

Tabelle 89.4. Beispiel 88.1, gerechnet nach dem in Tab. 89.3 angegebenen Schema mit $a = 130$ in (89.1)

i	Stichprobenwerte −130				G_i^*	G_i^{*2}
1	7	12	−2	7	24	576
2	10	9	−13	7	13	169
3	12	10	3	11	36	1296
			Summe		$G^* = 73$	2041

$$K^* = \frac{73^2}{12} = 444{,}083 \qquad A^* = 1019$$

$$q = 1019{,}000 - 444{,}083 = 574{,}917$$

$$q_1 = \frac{2041{,}000}{4} - 444{,}083 = 66{,}167$$

$$q_2 = 574{,}917 - 66{,}167 = 508{,}750$$

Aufgaben zu Abschnitt 89

Bei den folgenden Aufgaben nehme man jeweils an, daß es sich um Stichproben aus normalverteilten Grundgesamtheiten mit gleicher Varianz handelt.

89.1 Ist der Unterschied des Kupfergehaltes [%] der verschiedenen Bronzegüsse signifikant oder nur zufallsbedingt? (G. Wernimont, Industr. Quality Control **3**, 1947, 5)

Guß Nr.	Kupfergehalt [%]	
1	85,54	85,25
2	85,72	84,94
3	85,48	84,98

89.2 Man leite die Tab. 89.3 aus der Tab. 89.1 her.

89.3 In Aufgabe 81.5 haben wir gesehen, daß die 2. Art des Lesens (die in der Erdnußindustrie gerade nicht verwendet wird) besser als die 1. Art ist. Die Versuchspersonen standen dabei jeweils rechts vom laufenden Band. Ist es vielleicht günstiger, die Personen links vom Bande oder am Ende aufzustellen?

Teilnehmer Nr.	1	2	3	4	5	6	7	8	9
Rechts	281	279	260	230	267	253	205	265	299
Links	254	275	278	232	241	245	220	255	302
Am Ende	238	280	217	237	246	239	228	252	280

89.4 Die Tabelle (O. H. Latter, Biometrika **1**, 1901, 173) enthält die Länge [mm] von Kuckuckseiern in Nestern dreier verschiedener Vogelarten. Ist der Unterschied signifikant?

Braune Grasmücke	22,0	23,9	20,9	23,8	25,0	24,0	21,7	23,8
	22,8	23,1	23,1	23,5	23,0	23,0		
Rotkehlchen	21,8	23,0	23,3	22,4	22,4	23,0	23,0	23,0
	23,9	22,3	22,0	22,6	22,0	22,1	21,1	23,0
Zaunkönig	19,8	22,1	21,5	20,9	22,0	21,0	22,3	21,0
	20,3	20,9	22,0	20,0	20,8	21,2	21,0	

89.5 Beeinflußt das Pflanzdatum den Baumwollsamenertrag? Die 12 Werte aus einem Ort in Uganda wurden erhalten, indem man 4 Felder in je 3 Teile teilte und die 3 Teile den 3 Pflanzdaten zufällig zuordnete (R. L. Anderson u. H. L. Manning, Biometrics 4, 1948, 171—196).

Gepflanzt am	Baumwollsamenertrag [kg]			
1. Mai	3,35	1,49	2,44	2,44
15. Mai	3,86	2,71	2,18	1,95
29. Mai	1,99	2,89	1,68	2,13

89.6 Unterscheiden sich die Heuerträge [Doppelzentner/ha] der drei Rotschwingelsorten in signifikanter Weise? (F. Zürn, Veröff. Bundesanst. f. alpine Landw. Admont 1953, 120)

1. Sorte	92,4	87,7	86,6	92,2	98,2	87,2
2. Sorte	72,4	84,3	89,0	85,5	95,3	84,5
3. Sorte	96,1	91,5	91,0	100,8	95,2	107,4

89.7 Besteht hinsichtlich der Faserdicke [10^{-3} mm] einer gewissen Baumwollsorte an drei Stellen I, II, III des Fruchtstandes ein signifikanter Unterschied? (J. H. Moore, Journ. Agr. Research 62, 1941, 265)

I	17,8	16,0	16,0	17,3	18,3	17,5	15,7	16,2	18,3	17,0
II	16,8	14,1	14,9	16,5	17,0	15,7	16,5	16,0	14,7	17,5
III	17,5	14,4	14,7	17,3	17,3	15,5	15,2	13,4	15,7	15,7

89.8 Hängt der Ertrag [cwt/acre] einer gewissen Winterweizensorte (Hybrid 46) signifikant vom Furchenabstand [Zoll] ab und sollte man demgemäß eng säen? (J. H. Baldwin, Agriculture 70, 1963, 415)

Abstand [Zoll]	Ertrag [cwt/acre]				
4	44,0	43,8	47,7	39,8	43,8
8	41,6	42,4	44,9	40,1	42,2
12	39,5	40,7	42,3	38,9	40,4

89.9 Hängt der prozentuale Fettgehalt des Hähnchenfleisches in signifikanter Weise vom Alter ab? (Aus einer Untersuchung der Kreuzung von New Hampshire-, Rhodeländer- und Sussexhühnern mit Leghornhähnen. D. Baljosov, Archiv f. Gefl.-Zucht u. Kleintierkunde 12, 1963, 74)

Alter [Tage]	Fett [%]			
75	6,3	7,3	5,9	7,8
90	4,1	5,5	9,0	4,4
120	9,6	7,1	7,7	4,8

89.10 Die British Coal Utilisation Research Association hat umfangreiche Staubuntersuchungen in Abgasen durchgeführt. Die nachstehenden Werte dienten dazu festzustellen, ob die Gasgeschwindigkeit [Fuß/sec] zeitlich einigermaßen konstant ist oder ob signifikante Unterschiede bestehen. Zwischen den Meßwerten einer Gruppe liegt jeweils 5 Sekunden Zeitdifferenz und zwischen den Gruppen jeweils etwa 1 Minute. (S. BROADBENT, Appl. Statistics **3**, 1954, 43). Was läßt sich aus den Werten schließen?

Gruppe	Geschwindigkeit [Fuß/sec]									
A	20	21	20	20	23	21	26			
B	24	25	27	23	22	22	24	27	26	25
C	25	28	22	24	26	26				

90 Theorie zu Abschnitt 88

In (88.1) schreiben wir umständlicher

$$x_{ik} - \bar{x} = [x_{ik} - \bar{x}_i] + [\bar{x}_i - \bar{x}].$$

Hierbei ist $\bar{x}_i$ durch (88.2) in Tab. 88.1 gegeben. Durch Quadrieren und Ausmultiplizieren folgt

$$(x_{ik} - \bar{x})^2 = [x_{ik} - \bar{x}_i]^2 + 2[x_{ik} - \bar{x}_i][\bar{x}_i - \bar{x}] + [\bar{x}_i - \bar{x}]^2.$$

Dies summieren wir zunächst nur über k von 1 bis n_i. Dabei liefert das 2. Glied rechts keinen Beitrag, weil $\bar{x}_i$ der Mittelwert der i-ten Gruppe und demgemäß

$$\sum_{k=1}^{n_i} [x_{ik} - \bar{x}_i] = 0$$

ist. Das letzte Glied rechts enthält k nicht und liefert den Beitrag $n_i[\bar{x}_i - \bar{x}]^2$. So ergibt sich

$$\sum_{k=1}^{n_i} (x_{ik} - \bar{x})^2 = n_i[\bar{x}_i - \bar{x}]^2 + \sum_{k=1}^{n_i} [x_{ik} - \bar{x}_i]^2.$$

Summieren wir dies nun über i von 1 bis r, so erhalten wir

$$q = q_1 + q_2 \tag{90.1}$$

mit q, q_1 und q_2 gemäß (88.1), (88.4) und (88.5). Dies ist die im Zusammenhang mit (88.1) erwähnte Zerlegung.

Jeden Stichprobenwert x_{ik} können wir als Einzelbeobachtung einer normalverteilten Zufallsvariablen X_{ik} auffassen. Diese n Variablen sind unabhängig. q ist dann eine Einzelbeobachtung der Zufallsvariablen

$$Q = \sum_{i=1}^{r} \sum_{k=1}^{n_i} (X_{ik} - \bar{X})^2.$$

Hierbei ist $\bar{X}$ natürlich die Summe aller X_{ik}, geteilt durch n [vgl. (88.3)].

Wir nehmen nun einmal an, die Hypothese

$$\mu_1 = \mu_2 = \cdots = \mu_r$$

sei richtig. Dann haben alle n Variablen genau dieselbe Normalverteilung. $q/(n-1)$ ist dann die Varianz der gesamten Stichprobe. Also hat Q/σ^2 gemäß Satz 73.1 eine χ^2-Verteilung mit $n-1$ Freiheitsgraden. Die durch (88.4) und (88.5) gegebenen Zahlen q_1 und q_2 sind beobachtete Werte der Zufallsvariablen

$$Q_1 = \sum_{i=1}^{r} n_i\,(\bar{X}_i - \bar{X})^2$$

und

$$Q_2 = \sum_{i=1}^{r} \sum_{k=1}^{n_i} (X_{ik} - \bar{X}_i)^2.$$

Hierbei ist natürlich [vgl. (88.2)]

$$\bar{X}_i = \frac{1}{n_i} \sum_{k=1}^{n_i} X_{ik}.$$

Wie in Abschn. 73 ergibt sich, daß Q_1/σ^2 und Q_2/σ^2 eine χ^2-Verteilung mit $r-1$ bzw. $n-r$ Freiheitsgraden besitzen. Weiterhin sind Q_1 und Q_2 unabhängig. Aus (61.1) sehen wir, daß der Mittelwert einer χ^2-Verteilung gleich der Anzahl der Freiheitsgrade ist. Also hat Q_1/σ^2 den Mittelwert $r-1$, und Q_2/σ^2 hat den Mittelwert $n-r$. Wegen (34.2) ist also der Mittelwert der Variablen

$$S_1{}^2 = \frac{1}{r-1} Q_1 \qquad \text{und} \qquad S_2{}^2 = \frac{1}{n-r} Q_2$$

jeweils gleich σ^2. Aus Satz 83.1 folgt weiterhin, daß der Quotient

$$V = S_1{}^2/S_2{}^2$$

eine F-Verteilung mit $(r-1, n-r)$ Freiheitsgraden besitzt. Die Zahl v_0 [vgl. (88.6)] ist ein beobachteter Wert dieser Zufallsvariablen. Ist die Hypothese richtig, so ist die Wahrscheinlichkeit, daß V irgendeinen Wert $v_0 > c$ [vgl. (88.7)] annimmt, gleich α, also sehr gering. Liefert die Stichprobe trotzdem einen solchen großen Wert v_0, so haben wir demnach Grund zu der Annahme, daß unsere Hypothese nicht richtig ist, und verwerfen sie. Ist hingegen $v_0 \leq c$, so nehmen wir die Hypothese an.

Hierbei muß man sich noch überlegen, daß bei Nichtzutreffen der Hypothese jede Ungleichheit der Mittelwerte die Tendenz hat, v_0 zu vergrößern und nicht etwa zu verkleinern. Dies ergibt sich daraus, daß dann die Verteilung von Q_2 dieselbe bleibt wie zuvor, während die Verteilung von Q_1 und damit auch von V in dem genannten Sinne geändert wird.

91 Doppelte Varianzanalyse

Bisher haben wir den Fall einer Stichprobe betrachtet, die sich in Gruppen unterteilen ließ, wie dies einer *einzelnen* Variationsursache entspricht. Dann wendet man die *einfache* Varianzanalyse an. Lassen sich die Gruppen nun noch zusätzlich nach einem weiteren Gesichtspunkt untergliedern, so wendet man die *doppelte* Varianzanalyse an, der wir uns nun zuwenden. Mit dieser kann man also gleichzeitig den Einfluß zweier Variationsursachen untersuchen. Beispiele sind der Einfluß der Düngung *und* der Bewässerung auf den Ernteertrag, den Einfluß der Konservierungsart *und* der Lebensmittelart auf die Haltbarkeit, der Einfluß der Wohnungsverhältnisse *und* der Ernährung auf die Verbreitung von Kinderkrankheiten usw.

Die Teile, die wir bei der genannten Unterteilung der Gruppen einer Stichprobe erhalten, bezeichnen wir als **Klassen.**

Vorgegeben sei eine Stichprobe von insgesamt n Werten. Diese sei in r Gruppen geteilt. Jede Gruppe sei in p Klassen unterteilt. Wir behandeln den einfachsten Fall, daß jede Klasse nur einen einzigen Wert enthält. Dann ist natürlich $n = rp$. Die Stichprobenwerte bezeichnen wir wieder mit x_{ik}. Der erste Index ist wieder die Gruppennummer, und der zweite ist nun die Nummer der Klasse. Unsere Stichprobe läßt sich dann so anordnen:

$$\begin{array}{r}\\ r \text{ Zeilen} \\ \text{(Gruppen)}\end{array} \left\{ \overbrace{\begin{array}{cccc} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \cdot & \cdot & \cdots & \cdot \\ x_{r1} & x_{r2} & \cdots & x_{rp} \end{array}}^{p \text{ Spalten}} \right.$$

Wir machen die folgende **Voraussetzung**: Die n Beobachtungen entstammen aus n unabhängigen normalverteilten Grundgesamtheiten mit derselben Varianz σ^2 und den möglicherweise verschiedenen Mittelwerten $\mu_{11}, \ldots, \mu_{rp}$. Dabei braucht σ nicht bekannt zu sein. Zu testen sei die Hypothese, daß alle Mittelwerte gleich sind, also alle die n genannten Grundgesamtheiten vollkommen gleich verteilt sind.

Die waagerechten Reihen von Werten in unserem obigen Schema nennen wir **Zeilen** und die senkrechten Reihen **Spalten.**

Der Mittelwert der i-ten Zeile wird üblicherweise mit $\bar{x}_{i.}$ bezeichnet. Es ist also

$$\bar{x}_{i.} = \frac{1}{p} \sum_{k=1}^{p} x_{ik} = \frac{\text{Zeilensumme}}{\text{Anzahl der Werte pro Zeile}}. \tag{91.1}$$

Der Mittelwert der k-ten Spalte wird mit $\bar{x}_{.k}$ bezeichnet. Es ist also

$$(91.2)\qquad \bar{x}_{.k} = \frac{1}{r}\sum_{i=1}^{r} x_{ik} = \frac{\text{Spaltensumme}}{\text{Anzahl der Werte pro Spalte}}.$$

Die Punkte stehen zur Unterscheidung der beiden Arten von Mittelwerten, und zwar an der Stelle desjenigen Index, über den man summiert.

Wie bisher ist natürlich der Mittelwert der gesamten Stichprobe durch

$$(91.3)\qquad \bar{x} = \frac{1}{n}\sum_{i=1}^{r}\sum_{k=1}^{p} x_{ik}$$

gegeben.

Bei der einfachen Varianzanalyse hatten wir die Quadratsumme

$$(91.4)\qquad q = \sum_{i=1}^{r}\sum_{k=1}^{p} (x_{ik} - \bar{x})^2$$

in zwei Bestandteile zerlegt. Nun zerlegen wir q in eine Summe von drei Bestandteilen q_1, q_2, q_3. Diese rein algebraische Operation ist ganz ähnlich wie früher. Wir schreiben durch Zufügen und Abziehen von Gliedern umständlicher

$$x_{ik} - \bar{x} = (\bar{x}_{i.} - \bar{x}) + (\bar{x}_{.k} - \bar{x}) + (x_{ik} - \bar{x}_{i.} - \bar{x}_{.k} + \bar{x}).$$

Dann quadrieren wir auf beiden Seiten. Rechts multiplizieren wir so aus, daß wir 3 Quadrate $(\bar{x}_{i.} - \bar{x})^2$ usw. und 3 doppelte Produkte erhalten. Nun wird über i und k summiert. Ähnlich wie früher liefern dabei die doppelten Produkte keinen Beitrag. Übrig bleibt also

$$q = \sum_i\sum_k (\bar{x}_{i.} - \bar{x})^2 + \sum_i\sum_k (\bar{x}_{.k} - \bar{x})^2 + \sum_i\sum_k (x_{ik} - \bar{x}_{i.} - \bar{x}_{.k} + \bar{x})^2.$$

Die ersten beiden Doppelsummen lassen sich auf einfache Summen reduzieren, weil jeweils nur ein Index auftritt. So erhalten wir die gewünschte Zerlegung in der Form

$$(91.5)\qquad q = q_1 + q_2 + q_3.$$

Hierbei haben wir zur Abkürzung

$$(91.6)\qquad q_1 = p\sum_{i=1}^{r} (\bar{x}_{i.} - \bar{x})^2,$$

$$(91.7)\qquad q_2 = r\sum_{k=1}^{p} (\bar{x}_{.k} - \bar{x})^2,$$

$$(91.8)\qquad q_3 = \sum_{i=1}^{r}\sum_{k=1}^{p} (x_{ik} - \bar{x}_{i.} - \bar{x}_{.k} + \bar{x})^2$$

gesetzt.

q_1 heißt die **Quadratsumme zwischen den Mittelwerten der Zeilen.** q_2 heißt die **Quadratsumme zwischen den Mittelwerten der Spalten.** q_3 heißt die **quadratische Restsumme.**

Wie bei der einfachen Varianzanalyse fassen wir die n Stichprobenwerte x_{ik} als einzelne Beobachtungen von n Zufallsvariablen X_{ik} auf. Voraussetzungsgemäß sind die letzteren unabhängig und normalverteilt mit der gleichen Varianz σ^2 und den Mittelwerten μ_{ik}. Denken wir uns die x_{ik} in (91.1)–(91.8) durch die X_{ik} ersetzt, so erhalten wir Zufallsvariable, die wir mit dem jeweils entsprechenden Großbuchstaben $\bar{X}_{i.}$, $\bar{X}_{.k}$, $\bar{X}$, Q, Q_1, Q_2, Q_3 bezeichnen.

Ist die Hypothese richtig, so haben Q/σ^2, Q_1/σ^2, Q_2/σ^2, Q_3/σ^2 jeweils eine χ^2-Verteilung mit $n-1$, $r-1$, $p-1$ bzw. $(r-1)(p-1)$ Freiheitsgraden, und die letzten drei dieser Zufallsvariablen sind unabhängig. Weiterhin ist der Mittelwert der Variablen

$$S_1{}^2 = \frac{1}{r-1} Q_1, \qquad S_2{}^2 = \frac{1}{p-1} Q_2, \qquad S_3{}^2 = \frac{1}{(r-1)(p-1)} Q_3$$

gleich σ^2. Dies ergibt sich ähnlich wie im vorigen Abschnitt. Aus Satz 83.1 folgt, daß die Quotienten

$$V_1 = S_1{}^2/S_3{}^2 \qquad \text{und} \qquad V_2 = S_2{}^2/S_3{}^2$$

eine F-Verteilung mit

$$[r-1, (r-1)(p-1)] \text{ bzw. } [p-1, (r-1)(p-1)]$$

Freiheitsgraden besitzen.

92 Verlauf des Tests bei der doppelten Varianzanalyse

Wie man bei der doppelten Varianzanalyse vorzugehen hat, ersieht man aus Tab. 92.1. Dabei stellt man die Werte, die man beim 1. Schritt erhält, zusammen, wie Tab. 92.2 zeigt. Daß sich die Rechnung noch wesentlich vereinfachen läßt, überlegen wir uns im nächsten Abschnitt.

Ist $v_1 > c_1$, so wird die Hypothese verworfen. Die Aussage über die Spalten, die man beim 5. Schritt erhält, ist aber auch dann noch vernünftig, falls man annehmen kann, daß sich die beiden Variationsursachen, die man untersucht, additiv überlagern, so daß also jedes x_{ik} in der Form

$$x_{ik} = a_i + b_k + z_{ik}$$

darstellbar ist. Hierbei entsprechen die Konstanten a_i und b_k den beiden genannten Einflüssen. z_{ik} entspricht dem Zufallseinfluß, ist also ein beobachteter Wert einer normalverteilten Zufallsvariablen.

Tabelle 92.1. Doppelte Varianzanalyse unter Benutzung der zu Beginn des vorigen Abschnittes angegebenen Stichprobe, die in r Zeilen und p Spalten gegliedert ist. (Vereinfachte Rechnung siehe im nächsten Abschnitt)

1. Schritt. Man berechne q, q_1, q_2 und daraus $q_3 = q - q_1 - q_2$. Die Ergebnisse stelle man gemäß Tab. 92.2 zusammen.

2. Schritt. Man berechne die Quotienten

$$v_1 = s_1^2/s_3^2 \qquad \text{und} \qquad v_2 = s_2^2/s_3^2 .$$

3. Schritt. Man wähle eine Signifikanzzahl α (5% oder 1%).

4. Schritt. Man bestimme die Lösung c_1 der Gleichung

$$P(V \leqq c_1) = 1 - \alpha$$

aus der Tafel 9a bzw. 9b der F-Verteilung mit $[r-1, (r-1)(p-1)]$ Freiheitsgraden. Ist $v_1 \leqq c_1$, so wird angenommen, daß zwischen den *Zeilen* kein signifikanter Unterschied besteht. Ist $v_1 > c_1$, so wird angenommen, daß ein solcher Unterschied besteht.

5. Schritt. Man bestimme die Lösung c_2 der Gleichung

$$P(V \leqq c_2) = 1 - \alpha$$

aus der Tafel 9a bzw. 9b der F-Verteilung mit $[p-1, (r-1)(p-1)]$ Freiheitsgraden. Ist $v_2 \leqq c_2$, so wird angenommen, daß zwischen den *Spalten* kein signifikanter Unterschied besteht. Ist $v_2 > c_2$, so wird angenommen, daß ein solcher Unterschied besteht.

6. Schritt. Ist $v_1 \leqq c_1$ und $v_2 \leqq c_2$, so wird die Hypothese, daß alle n Mittelwerte μ_{ik} gleich sind, angenommen. In jedem anderen Fall wird sie verworfen.

Tabelle 92.2. Zum 1. Schritt der doppelten Varianzanalyse

Variation	Freiheits-grade	Quadrat-summe	Durchschnitts-quadrat
Zwischen den Zeilen	$r-1$	q_1	$s_1^2 = q_1/(r-1)$
Zwischen den Spalten	$p-1$	q_2	$s_2^2 = q_2/(p-1)$
Rest	$(r-1)(p-1)$	q_3	$s_3^2 = q_3/(r-1)(p-1)$
Insgesamt	$n-1$	q	

93 Beispiel. Vereinfachte Rechnung

Beispiel 93.1 (Schweinemast). Zwölf Schweine werden auf Grund ihres Gewichtes zu Beginn des Versuchs in $r = 4$ Gruppen zu je $p = 3$ Schweinen eingeteilt, denen dann jeweils verschiedene Futterarten (A, B, C) verabreicht werden. Nach einer gewissen Zeit stellt man die in Tab. 93.1 angegebenen Gewichtszunahmen [engl. Pfund] fest (G. DUNLOP, J. Agr. Sci. **25**, 1935, 445). Man teste die Hypothese, daß die Unterschiede zwischen den beobachteten Werten rein zufällig sind, daß also weder die verwendeten Futterarten noch die Anfangsgewichte einen Einfluß auf die Gewichtszunahme besitzen. Dabei nehme man an, daß eine normalverteilte Zufallsvariable vorliegt.

Tabelle 93.1. Stichprobe in Beispiel 93.1

	Futterart		
	A	B	C
Gruppe 1	7,0	14,0	8,5
„ 2	16,0	15,5	16,5
„ 3	10,5	15,0	9,5
„ 4	13,5	21,0	13,5

Tabelle 93.2. Erster Schritt in Beispiel 93.1

Variation	Freiheitsgrade	Quadrat-summe	Durchschnitts-quadrat
Zwischen den Zeilen (= zwischen den Gruppen)	$r - 1 = 3$	$q_1 = 87{,}73$	$s_1^2 = \frac{q_1}{3} = 29{,}24$
Zwischen den Spalten (= zwischen den Futterarten)	$p - 1 = 2$	$q_2 = 54{,}12$	$s_2^2 = \frac{q_2}{2} = 27{,}06$
Rest	$(r-1)(p-1) = 6$	$q_3 = 28{,}21$	$s_3^2 = \frac{q_3}{6} = 4{,}702$
Insgesamt	$n - 1 = 11$	$q = 170{,}06$	

1. Schritt. Die Berechnung ergibt die in Tab. 93.2 gezeigten Werte. (Einzelheiten der Berechnung siehe unten.)

2. Schritt. Es ist

$$v_1 = \frac{29{,}24}{4{,}702} = 6{,}219 \qquad \text{und} \qquad v_2 = \frac{27{,}06}{4{,}702} = 5{,}755.$$

3. Schritt. Wir wählen die Signifikanzzahl $\alpha = 0{,}05$.

4. Schritt. Für die F-Verteilung mit (3, 6) Freiheitsgraden hat die Gleichung $P(V \leqq c_1) = 1 - \alpha = 0{,}95$ die Lösung

$$c_1 = 4{,}76.$$

Es ist $v_1 = 6{,}219 > c_1 = 4{,}76$. Wir nehmen demgemäß an, daß zwischen den Gruppen ein signifikanter Unterschied besteht, d. h. daß das Anfangsgewicht die Gewichtszunahme beeinflußt. Damit kann die zu testende Hypothese bereits verworfen werden.

5. Schritt. Weiterhin können wir voraussetzen, daß bei dem vorliegenden Experiment die am Schluß des vorigen Abschnitts genannte Additivität besteht.

Für die F-Verteilung mit (2, 6) Freiheitsgraden hat die Gleichung $P(V \leq c_2) = 1 - \alpha = 0{,}95$ die Lösung

$$c_2 = 5{,}14.$$

Es ist $v_2 = 5{,}755 > c_2 = 5{,}14$. Unter der genannten Voraussetzung dürfen wir also annehmen, daß zwischen den Spalten ebenfalls ein signifikanter Unterschied besteht, d. h. daß auch die Futterart die Gewichtszunahme beeinflußt.

Die Rechnungen beim 1. Schritt lassen sich einfach gestalten. Um q_1, q_2 und q_3 zu erhalten, berechne man der Reihe nach die in Tab. 93.3 angegebenen Größen. Die Formeln dieser Tabelle erhält man aus (91.1)–(91.8). Sie ähneln denjenigen in Abschn. 89.

Tabelle 93.3. Rechenformeln zur doppelten Varianzanalyse, wenn jede Klasse aus nur einem Wert besteht

$$G_i = \sum_{k=1}^{p} x_{ik} = \text{Summe der Werte der } i\text{-ten Zeile}$$

$$H_k = \sum_{i=1}^{r} x_{ik} = \text{Summe der Werte der } k\text{-ten Spalte}$$

$$G = G_1 + G_2 + \cdots + G_r = \text{Gesamtsumme}$$

$$K = G^2/n$$

$$A = \sum_{i=1}^{r} \sum_{k=1}^{p} x_{ik}^2 = \text{Summe der Quadrate aller Werte}$$

$$q = A - K$$

$$q_1 = \frac{1}{p}(G_1^2 + G_2^2 + \cdots + G_r^2) - K$$

$$q_2 = \frac{1}{r}(H_1^2 + H_2^2 + \cdots + H_p^2) - K$$

$$q_3 = q - q_1 - q_2$$

Wiederum kann man

$$x_{ik} = x_{ik}^* + a$$

setzen und a dabei so wählen, daß die x_{ik}^* dem Betrage nach möglichst klein ausfallen. Den obigen Formeln entsprechen dann genau dieselben

Formeln in den x_{ik}^*. Bezeichnen wir die Größen in den neuen Formeln wie in Abschn. 89 mit einem Stern, so hat die Rechnung des 1. Schrittes in Beispiel 93.1, wenn wir $a = 15$ wählen, die in Tab. 93.4 gezeigte Form.

Tabelle 93.4. Rechnungen beim 1. Schritt in Beispiel 93.1

	Stichprobenwerte minus 15,0			G_i^*	G_i^{*2}
	−8,0	−1,0	−6,5	−15,5	240,25
	1,0	0,5	1,5	3,0	9,00
	−4,5	0	−5,5	−10,0	100,00
	−1,5	6,0	−1,5	3,0	9,00
H_k^*	−13,0	5,5	−12,0	$G^* = -19,5$	358,25
H_k^{*2}	169,00	30,25	144,00	343,25	

$$G^* = -19{,}5 \qquad K^* = \frac{19{,}5^2}{12} = 31{,}69 \qquad A^* = 201{,}75$$

$$q = 201{,}75 - 31{,}69 = 170{,}06$$

$$q_1 = \frac{358{,}25}{3} - 31{,}69 = 87{,}73$$

$$q_2 = \frac{343{,}25}{4} - 31{,}69 = 54{,}12$$

$$q_3 = 170{,}06 - 87{,}73 - 54{,}12 = 28{,}21$$

Aufgaben zu Abschnitt 93

Bei den folgenden Aufgaben nehme man jeweils an, daß die zu Beginn des Abschnittes 91 gemachte Voraussetzung erfüllt ist.

93.1 Man teste die Hypothese, daß die Unterschiede zwischen den Werten in Aufgabe 89.3 rein zufällig sind, daß also weder zwischen den Stellungen noch zwischen der Schnelligkeit der 9 Teilnehmer ein signifikanter Unterschied besteht.

93.2 Die folgenden Daten stammen aus einem Versuch über Befestigungsarten von Werkzeugen, um Unfälle und Zeitverluste bei Arbeiten an einem Raketensilo zu vermeiden. (H. J. Lavender u. J. A. Dinan, Journ. Industr. Eng. **13**, 1962, 477.) Angegeben ist die zum Entfernen eines durch 6 Schrauben gehaltenen Teils benötigte Zeit [sec]. Der Schraubenzieher war an einem Rie-

Halter	Arbeiter Nr. 1	2	3	4	5	6	7	8	9	10	11	12
Typ A	93	98	91	65	74	80	84	81	55	94	49	64
Typ B	97	62	100	70	76	68	73	73	61	85	61	61
Typ C	133	64	71	76	66	76	74	94	64	82	49	67
Ohne	108	62	62	62	68	78	74	67	53	71	47	63

men befestigt, dessen anderes Ende am Gürtel hing (Typ A) oder das Handgelenk fest (Typ B) bzw. lose (Typ C) umschloß. Weiterhin interessierten die Unterschiede des Arbeitens der nach Alter und Ausbildung differierenden Teilnehmer. Man teste die Hypothese, daß weder zwischen den 4 Befestigungsarten noch zwischen den 12 Teilnehmern ein signifikanter Unterschied besteht.

93.3 Im Zusammenhang mit der Untersuchung körperlicher Arbeit interessierte die Muskelkraftleistung des rechten Arms bei schnellen Bewegungen eines kleinen Gewichtes in Abhängigkeit von der Bewegungsrichtung und dem Gewicht. Für das Verhältnis der Maximalgeschwindigkeit zur Durchschnittsgeschwindigkeit ergaben sich die in der Tabelle gezeigten Werte. Dabei ist B_1 eine horizontale Bewegung von rechts nach links, B_2 eine horizontale Bewegung auf den Körper zu, B_3 eine vertikale Bewegung nach unten. Ausgeführt wurden jeweils 90 Bewegungen je Minute nach dem Takte eines Metronoms (D. Franz u. G. Nadler, Journ. Industr. Eng. **12**, 1961, 321). Man teste die Hypothese, daß weder die Bewegungsrichtungen noch die Gewichte einen signifikanten Einfluß haben.

Gewicht	Bewegungsart		
	B_1	B_2	B_3
0	1,59	1,55	1,85
400 g	1,53	1,53	1,60
800 g	1,50	1,47	1,80

93.4 Erdbakterien kann man zählen, indem man eine gewisse Menge Erde in Salzwasser aufschwämmt, diese Aufschwämmung weiterverdünnt, einen kleinen Teil davon auf eine Platte mit Nährboden bringt und die sich entwickelnden Kolonien zählt. Dieses Verfahren wird in landwirtschaftlichen Versuchsanstalten angewendet. R. A. Fisher, H. G. Thornton und W. A. Mackenzie (Annals Appl. Biol. **9**, 1922, 328) haben eine Erdprobe in 4 Teile A, B, C, D teilen lassen. Die Teile wurden unabhängig gelöst. Für jeden Teil wurden dann 5 Platten hergestellt. Deren Zählung ergab die in der Tabelle gezeigten Werte. Für die Brauchbarkeit des Verfahrens ist es entscheidend wichtig, daß weder zwischen den Teilen noch zwischen den Platten ein signifikanter Unterschied besteht. Man teste diese Hypothese.

Platte Nr.	Teil A	B	C	D
1	26	28	31	37
2	30	33	26	32
3	30	32	28	32
4	29	26	32	30
5	32	27	31	26

93.5 Besteht zwischen den angegebenen Gerstenerträgen ein signifikanter Unterschied hinsichtlich der Sorten oder der Orte (3 Orte in Minnesota)?

(F. R. IMMER, H. K. HAYES u. L. R. POWERS, Journ. Amer. Soc. Agron. 26, 1934, 403—419)

Ort	Ertrag [Bushels pro Acre]				
	Manchuria	Svansota	Velvet	Trebi	Peatland
1	27	31	33	33	30
2	41	43	44	57	42
3	31	30	32	45	37

93.6 Die nachstehende Tabelle zeigt Werte des Stroms [mA], der auf dem Schirm eines gewissen Typs von Fernsehröhren eine gegebene Helligkeit hervorruft. Ist der Einfluß der Glasart oder der Phosphorart signifikant? (C. R. HICKS, Industr. Quality Control 13, 1956, No. 3, S. 5)

Glasart	Phosphorart		
	A	*B*	*C*
I	285	302	282
II	235	245	225

93.7 Die Feuergeschwindigkeit eines Marinegeschützes sollte durch eine neue Lademethode erhöht werden. In Gruppe I, II bzw. III bestand die Mannschaft aus leichten, mittelschweren bzw. schweren Männern. Besteht zwischen den beiden Methoden und zwischen den 3 Gruppen ein signifikanter Unterschied? (C. R. HICKS, Industr. Quality Control 13, 1956, No. 4, S. 15)

Gruppe	Methode	
	alt	neu
I	15,9	24,4
II	15,3	23,4
III	14,0	23,0

93.8 Spannungsregler für Kraftwagen, die zwischen 15,8 Volt und 16,4 Volt schalten sollen, werden auf einer Einstellstation eingestellt und dann auf einer Prüfstation nochmals geprüft. Ist der Unterschied zwischen den 3 Einstellstationen oder zwischen den 4 Prüfstationen signifikant? (D. J. DESMOND, Appl. Statistics 3, 1954, 67. Die Tabelle zeigt Werte der geregelten Spannung eines Reglers, der zuerst auf Station *A* eingestellt wurde, dann auf den Prüfstationen 1—4 geprüft wurde, dann auf Station *B* erneut eingestellt wurde usw.)

Einstell-station	Prüfstation			
	1	2	3	4
A	16,5	16,5	16,6	16,6
B	16,0	16,1	16,0	16,1
C	16,0	16,0	15,9	16,3

Kapitel 17

Paare von Messungen. Regression

Bisher haben wir in Teil III durchweg Zufallsexperimente behandelt, bei denen nur eine einzelne Variable vorkam. Nun wenden wir uns Problemen zu, bei denen zwei Variable eine Rolle spielen. Diese bezeichnen wir mit X und Y. Man interessiert sich dabei vor allem dafür, ob irgendeine Beziehung zwischen den Variablen besteht und welcher Art diese ist.

Zum Beispiel kann man fragen nach der Größe Y der Söhne in Abhängigkeit von der Größe X der Väter, nach dem Eisengehalt Y von Erzen in Abhängigkeit von der Dichte X, nach dem Blutdruck Y in Abhängigkeit vom Alter X, nach der Gewichtszunahme Y von Tieren in Abhängigkeit von der Futtermenge X usw. Dabei sehen wir die eine Variable (X) als unabhängig und die andere als abhängig an. In der Analysis heißt Y dann eine *Funktion* von X. Hier in der Statistik, in der Y eine Zufallsvariable ist und X eine solche sein kann (wie bei den ersten beiden Beispielen) oder auch nicht (wie bei den letzten), sprechen wir statt dessen von der **Regression** *von Y bezüglich X*. Diese nichtssagende Bezeichnung (Regreß = Rückschritt) hat sich, von einem speziellen Beispiel (s. unten) herkommend, leider allgemein eingebürgert und erhalten.

Weiterhin kann man z. B. fragen nach dem Zusammenhang zwischen der Größe von Geschwisterpaaren, zwischen dem Heiratsalter der Männer und Frauen, zwischen dem Luftdruck an zwei verschiedenen Beobachtungsstellen, zwischen der Abnutzung der beiden Hinterreifen eines Fahrzeuges usw. In diesen Fällen, in denen man sich jeweils für Beziehungen zwischen zwei Zufallsvariablen X und Y interessiert, ohne die eine als abhängig und die andere als unabhängig anzusehen, spricht man von der **Korrelation** *zwischen X und Y*. Diese Betrachtungsweise, bei der X und Y in gewissem Sinne als gleichberechtigt aufgefaßt werden, folgt erst im nächsten Kapitel. Dabei wird sich zeigen, daß gewisse formelmäßige Beziehungen zu Begriffen der Regressionsrechnung bestehen.

94 Regressionsgerade. Prinzip der kleinsten Quadrate

Liegt eine Stichprobe von Beobachtungen

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$$

aus einer zweidimensionalen (X, Y)-Grundgesamtheit vor, so ist es meistens am besten, wenn wir diese n Zahlenpaare zuerst einmal als Punkte in die xy-Ebene in der üblichen Weise eintragen.

Stellt sich dabei heraus, daß diese Punkte nahezu auf einer Geraden liegen, so können wir eine solche „*Ausgleichsgerade*" oder *Regressionsgerade von Y bezüglich X*, die sich der Lage der Punkte möglichst gut anpaßt, ohne Mühe nach Augenmaß zeichnen. Aus dieser Geraden können wir ablesen, welcher Y-Wert zu einem vorgegebenen X-Wert etwa zu erwarten ist.

Abb. 94.1 zeigt als typisches Beispiel die Höhe von Sojabohnenpflanzen in Abhängigkeit vom Alter. Das Zeitintervall (5 Wochen) ist relativ kurz. Dementsprechend paßt sich das gezeichnete Geradenstück den Meßwerten recht gut an. Würden wir unser Intervall nach rechts hin immer mehr vergrößern, wäre es bald vorbei mit einer brauchbaren geradlinigen Annäherung, denn das Wachstum verlangsamt sich schließlich bis zum Stillstand. (Auf Regressions*kurven* gehen wir erst später ein.)

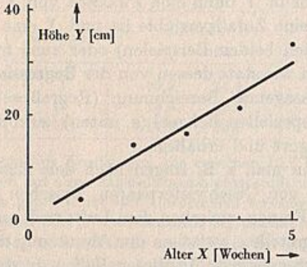


Abbildung 94.1. Regression der Höhe von Sojabohnenpflanzen bezüglich des Alters [J. B. WENTZ und R. T. STEWART, J. Amer. Soc. Agronomy **16**, 1924, 534]

Abb. 94.2 zeigt die Größe von Söhnen in Abhängigkeit von der Größe der Väter. Wie man sieht, besteht zwar die Tendenz, daß große Väter auch große Söhne haben, aber so, daß diese Söhne kleiner als die Väter sind. Es besteht also ein Rückschritt (= Regreß) zur Durchschnittsgröße der Menschen. Sinngemäß gilt dasselbe für kleine Söhne.

Das war zu erwarten. Man stelle sich vor, es wäre anders. Aus dieser Beobachtung, die zuerst von F. Galton gemacht wurde, leiten sich die Worte Regression, Regressionsgerade, Regressionsrechnung usw. her. Eine sehr spezielle Sache hat also hier dummerweise einem ganzen Gebiet den Namen aufgeprägt.

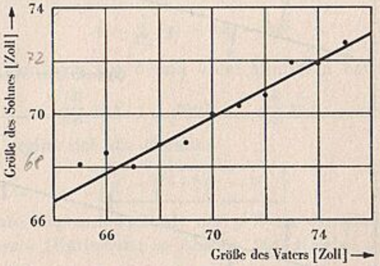


Abbildung 94.2. Regression der Größe des Sohnes bezüglich der Größe des Vaters (Mittelwerte von insgesamt über 1000 Beobachtungen nach K. Pearson und A. Lee, Biometrika 2, 1903, 357)

In Abb. 94.2 liegen die Punkte nicht mehr so schön auf einer Geraden, und wir werden bald noch krassere Beispiele kennenlernen, in denen die Punkte noch stärker streuen. Dann werden verschiedene Personen im allgemeinen verschiedener Meinung darüber sein, welche Gerade (bzw. Kurve) sich den Punkten „am besten anpaßt". Mit dem Verfahren nach Augenmaß sind wir dann am Ende, und um subjektive Einflüsse auszuschalten, müssen wir eine objektive Methode heranziehen. Eine solche ist das Gausssche **Prinzip der kleinsten Quadrate.** Für unser gegenwärtiges Problem besagt dieses Prinzip folgendes:

Die Gerade ist so zu legen, daß die Summe der Quadrate aller Abstände unserer Punkte von der Geraden möglichst klein wird.

Daß die genannten Abstände bei der Lösung unseres Problems ins Spiel kommen müssen, ist klar. Daß man deren Quadrate und nicht etwa deren Absolutbeträge benutzt, hat einen rein praktischen Grund: Man erhält dann einfachere Rechnungen und Ergebnisse. Deshalb gibt man auch sonst dem Prinzip der kleinsten Quadrate meist den Vorzug vor anderen denkbaren Möglichkeiten.

Unter dem Abstand eines Punktes von einer Geraden versteht man üblicherweise die Länge des Lotes von dem Punkt auf die Gerade, also den senkrechten Abstand (vgl. Abb. 94.3a). Diesen benutzen wir aber nicht. Sondern wir benutzen den in y-Richtung gemessenen Abstand, also den **vertikalen Abstand** (vgl. Abb. 94.3b). Erfahrungsge-

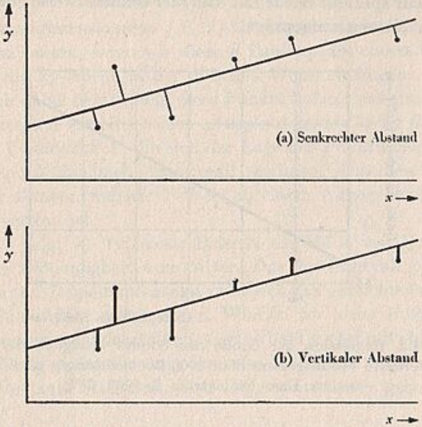


Abbildung 94.3. Abstand gegebener Punkte von einer Geraden

mäß ist das nämlich rechentechnisch und auch theoretisch viel einfacher. Hat man diese Erfahrung erst einmal gemacht, so kann man zugunsten des vertikalen Abstandes auch noch folgendes anführen: Da wir X als unabhängige Variable auffassen und die Gerade dazu verwenden wollen, den zu einem x gehörigen y-Wert abzuschätzen, ist der genannte Abstand besonders sachgemäß, denn er mißt die Differenz zwischen dem tatsächlich beobachteten y-Wert und dem genannten Schätzwert für das betreffende x.

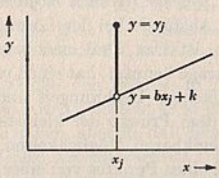


Abbildung 94.4. Berechnung des vertikalen Abstandes

Wir setzen voraus, daß die x-Werte unserer Stichprobe nicht alle gleich sind.

Ein Punkt (x_j, y_j) der Stichprobe hat von einer Geraden

$$y = bx + k \tag{94.1}$$

den vertikalen Abstand $|y_j - bx_j - k|$, vgl. Abb. 94.4. Die Summe der Abstandsquadrate ist also

$$a = \sum_{j=1}^{n} (y_j - bx_j - k)^2. \tag{94.2}$$

Damit diese Funktion von b und k ein Minimum hat, muß

$$\frac{\partial a}{\partial b} = 0 \qquad \text{und} \qquad \frac{\partial a}{\partial k} = 0$$

sein. Daraus ergibt sich die Gerade

$$\boxed{y - \bar{y} = b(x - \bar{x}),} \tag{94.3}$$

die sogenannte **Regressionsgerade** *der y-Werte bezüglich der x-Werte in der Stichprobe* (Herleitung in Abschn. 96). Hierbei ist

$$\bar{x} = \frac{1}{n}(x_1 + \cdots + x_n) \quad \text{und} \quad \bar{y} = \frac{1}{n}(y_1 + \cdots + y_n). \tag{94.4}$$

Das Steigungsmaß b heißt der zugehörige **Regressionskoeffizient** und hat die Form

$$\boxed{b = \frac{s_{xy}}{s_1^2}.} \tag{94.5}$$

Hierbei ist

$$s_1^2 = \frac{1}{n-1}\sum_{j=1}^{n} (x_j - \bar{x})^2 = \frac{1}{n-1}\left[\sum_{j=1}^{n} x_j^2 - \frac{1}{n}\left(\sum_{j=1}^{n} x_j\right)^2\right] \tag{94.6}$$

die Varianz der x-Werte in der Stichprobe und

$$\boxed{s_{xy} = \frac{1}{n-1}\sum_{j=1}^{n} (x_j - \bar{x})(y_j - \bar{y}).} \tag{94.7}$$

s_{xy} heißt die **Kovarianz** *der Stichprobe.* Der Leser mag zeigen, daß

$$\boxed{\begin{aligned} s_{xy} &= \frac{1}{n-1}\left(\sum_{j=1}^{n} x_j y_j - n\bar{x}\bar{y}\right) \\ &= \frac{1}{n-1}\left[\sum_{j=1}^{n} x_j y_j - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{j=1}^{n} y_j\right)\right] \end{aligned}} \tag{94.8}$$

gilt.

Wie man sieht, läuft die Regressionsgerade (94.3) durch den Punkt $(\bar{x}, \bar{y})$.

s_{xy} kann positiv, null oder negativ sein. Dasselbe gilt demnach auch für b. Ist b positiv bzw. negativ, so spricht man von *positiver* bzw. *negativer Regression.*

Ist X die Zeit, so nennt man die Regression auch **Trend.** Dieser Begriff wird insbesondere in der Wirtschaftsstatistik gebraucht, weil man sich in diesem Gebiet häufig für zeitliche Entwicklungen (im Außenhandel, in der Produktion und im Verbrauch, bei den Löhnen, Preisen, Börsenkursen usw.) interessiert.

Aufgaben zu Abschnitt 94

94.1 Man gewinne (94.8) aus (94.7).

94.2 Man konstruiere eine möglichst einfache Stichprobe, für die $b = 0$ ist.

95 Beispiele. Vereinfachte Rechnung

Beispiel 95.1 (Dichte und Eisengehalt von Erzen). Für die Stichprobe in Tab. 95.1 wollen wir die Regressionsgerade des Eisengehaltes y bezüglich der Dichte x bestimmen.

Aus den gegebenen Werten berechnen wir die Hilfswerte x_j^2 und $x_j y_j$ und dann die angegebenen Summen

$$\sum x_j = 28{,}1 \qquad \sum y_j = 267$$
$$\sum x_j^2 = 88{,}03 \qquad \sum x_j y_j = 837{,}2\,.$$

Wegen $n = 9$ sind die Mittelwerte

$$\bar{x} = \frac{28{,}1}{9} = 3{,}12 \qquad \bar{y} = \frac{267}{9} = 29{,}67\,.$$

Gemäß (94.6) und (94.8) erhalten wir weiterhin

$$s_1^2 = \frac{1}{8}\left(88{,}03 - \frac{28{,}1^2}{9}\right) = 0{,}03694$$
$$s_{xy} = \frac{1}{8}\left(837{,}2 - \frac{28{,}1 \cdot 267}{9}\right) = 0{,}4458\,.$$

Der Regressionskoeffizient hat den Wert

$$b = \frac{0{,}4458}{0{,}03694} = 12{,}07\,.$$

Die Regressionsgerade hat also die Darstellung

$$y - 29{,}67 = 12{,}07\,(x - 3{,}12)\,.$$

Dies können wir auch in der folgenden Form schreiben:

$$y = 12{,}07x - 7{,}99\,.$$

Die Berechnung der Regressionsgeraden läßt sich oft durch eine **lineare Transformation**

$$x_j = c_1 x_j^* + l_1 \qquad (95.1)$$
$$y_j = c_2 y_j^* + l_2$$

vereinfachen. Die vier Konstanten c_1, c_2, l_1 und l_2 wählt man so, daß die transformierten Werte x_j^* und y_j^* möglichst einfache Zahlen werden. Man berechnet dann die Mittelwerte $\bar{x}^*$ und $\bar{y}^*$, die Varianz s_1^{*2}

Tabelle 95.1. Regression des Eisengehaltes y [%] kieseliger Hämatiterze bezüglich der Dichte x [g/cm³] (H. Bottke, Bergbauwiss. **10**, 1963, 377)

Gemessene Werte		Hilfswerte	
x_j	y_j	x_j^2	$x_j y_j$
2,8	27	7,84	75,6
2,9	23	8,41	66,7
3,0	30	9,00	90,0
3,1	28	9,61	86,8
3,2	30	10,24	96,0
3,2	32	10,24	102,4
3,2	34	10,24	108,8
3,3	33	10,89	108,9
3,4	30	11,56	102,0
Summe 28,1	267	88,03	837,2

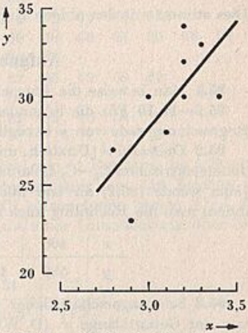


Abbildung 95.1. Zu Beispiel 95.1

und die Kovarianz s_{xy}^*, die zu den transformierten Werten gehört. Wegen (9.6) und (9.7) ergibt sich daraus

$$\begin{aligned} \bar{x} &= c_1\bar{x}^* + l_1, & \bar{y} &= c_2\bar{y}^* + l_2, \\ s_1^2 &= c_1^2 s_1^{*2}, & s_{xy} &= c_1 c_2 s_{xy}^*. \end{aligned} \tag{95.2}$$

Die Formel für s_{xy} möge der Leser beweisen (s. Aufgabe 95.1).

Beispiel 95.2. Setzen wir im vorigen Beispiel

$$x_j = 0{,}1 x_j^* + 3{,}0, \quad y_j = y_j^* + 30,$$

so ergibt sich (s. Tab. 95.2)

$$x_j^* = 10 x_j - 30, \quad y_j^* = y_j - 30.$$

Demnach wird

$$\bar{x}^* = \frac{11}{9} = 1{,}2, \qquad \bar{y}^* = \frac{-3}{9} = -0{,}33,$$

$$\sum x_j^{*2} = 43, \qquad \sum x_j^* y_j^* = 32.$$

Hieraus erhalten wir weiterhin

$$s_1^{*2} = \frac{1}{8}\left(43 - \frac{11^2}{9}\right) = 3{,}694,$$

$$s_{xy}^* = \frac{1}{8}\left(32 - \frac{11\cdot(-3)}{9}\right) = 4{,}458.$$

Tabelle 95.2. Zu Beispiel 95.2

x_j^*	y_j^*
−2	−3
−1	−7
0	0
1	−2
2	0
2	2
2	4
3	3
4	0
Summe 11	−3

Nun wenden wir schließlich (95.2) an. Wegen $c_1 = 0{,}1$ und $c_2 = 1$ folgt

$$\bar{x} = 0{,}1 \cdot 1{,}2 + 3{,}0 = 3{,}12 \qquad \bar{y} = -0{,}33 + 30 = 29{,}67$$

$$b = \frac{s_{xy}}{s_1^2} = \frac{c_2 s_{xy}^*}{c_1 s_1^{*2}} = \frac{4{,}458}{0{,}3694} = 12{,}07.$$

Dies stimmt mit den obigen Ergebnissen überein.

Aufgaben zu Abschnitt 95

95.1 Man beweise die Formel für s_{xy} in (95.2).

95.2—95.10 Für die folgenden Stichproben bestimme und zeichne man die Regressionsgerade von y bezüglich x.

95.2 Drehzahl x [Umdreh./min] und Leistung y [PS] des Dieselmotors einer dieselelektrischen C_0—C_0-Lokomotive (Techn. Rundschau Sulzer **41**, 1959, 27). [Man wende (95.2) an und überzeuge sich von der erzielten Vereinfachung, indem man die Rechnung auch noch ohne (95.2) ausführt.]

x	400	500	600	700	750
y	580	1030	1420	1880	2100

95.3 Schwangerschaftsdauer y (= Tragzeit post menstr.) in Abhängigkeit von der Geburtslänge x. (D. Wichmann, Arch. Gynäk. **177**, 1950, 261. Derartige Rückschlüsse von meßbaren Daten [z. B. Gewicht, Länge, Kopfumfang des neugeborenen Kindes] auf die Schwangerschaftsdauer benutzt man bei der gerichtlichen Vaterschaftsermittlung. Die Tabelle enthält $n = 5$ Wertepaare, und entsprechend rechne man. Um aber keine sachlich falschen Vorstellungen zu erwecken, sei gesagt, daß es sich in Wirklichkeit um 5 Mittelwerte aus zahlreichen [jeweils gleichvielen] Messungen handelt.)

x [cm]	48	49	50	51	52
y [Tage]	277,1	279,3	281,4	283,2	284,8

95.4 Sauerstoffgehalt y [mg/l] des Wörther Sees am 24. 9. 1962 in Abhängigkeit von der Tiefe x [m] (I. Findenegg, Österr. Wasserwirtsch. **15**, 1963, 67)

x	15	20	30	40	50	60	70
y	6,5	5,6	5,4	6,0	4,6	1,4	0,1

95.5 Sättigungs-Wassergehalt y [%] spezialbehandelter Wolle in Abhängigkeit vom Bisulfidgehalt x [% des Gehaltes unreduzierter Fasern] (A. R. Haly, Kolloid-Zeitschr. u. Zeitschr. f. Polymere **189**, 1963, 45)

x	10	15	30	40	50	55	80	100
y	50	46	43	43	36	39	37	33

95.6 Körpergröße x und „Sitzgröße" y (= Abstand von der Sitzfläche bis zum Auge) von 42 Arbeiterinnen in einer Stanzerei. (U. Burandt, Industr. Organisation **31**, 1962, 265. Die Daten wurden beim Neuentwurf von Tisch-Exzenterpressen nach modernen wissenschaftlichen Gesichtspunkten verwendet.)

x [cm]	y [cm]		
146	62		
150	60		
152	61		
154	60		
155	61	64	
158	61		
159	64	65	65
160	63	63	65
161	63		

x [cm]	y [cm]						
162	58	61	61				
163	63						
164	59	63	63	64	68		
165	60	60	65	65	65	66	67
166	62	66					
168	62	65	65	66	69		
170	68						
172	62	63	66	69			

95.7 Fehleranzahl y [pro Stunde und Funker] von 11 Funkern, die 9 Funksprüche zu je 1250 Zeichen mit einer Geschwindigkeit von 110 Zeichen je Minute bei verschiedenen Temperaturen x [°C] aufzunehmen hatten. (J. J. B. Worth, Journ. Industr. Eng. 9, 1958, 251, aus einer Untersuchung über die Klimatisierung von Arbeitsräumen.)

x	26	28	31	33
y	12,0	11,5	15,3	17,3

95.8 Halmfliegenbefall y [%] des Winterweizens in Abhängigkeit vom Schoßdatum x (= Tag im Juni, also $x = 2$ ist der 2. Juni usw. Nach R. Pfeiffer, Veröff. Bundesamt f. alpine Landw. Admont 1951)

x	2	3	3	5	6	7	8	8	9	10
y	4	10	12	12	8	38	32	52	32	28

x	10	11	11	12	13	13	13	14	16	16
y	55	40	60	72	51	63	65	69	76	78

95.9 Lufttemperaturen in Graz. x [°C] = mittlere Temperatur, y [°C] = Höchsttemperatur des betreffenden Monats. J = Januar usw. (Stat. Jahrb. d. Landeshauptstadt Graz 1960/61, S. 1)

	J	F	M	A	M	J	J	A	S	O	N	D
x	−3,1	3,3	7,8	12,7	12,9	18,4	17,6	18,3	16,8	11,0	4,0	−2,1
y	6,0	13,5	20,9	26,0	25,0	30,5	30,9	32,2	27,4	23,3	13,4	13,0

95.10 Diastolischer Blutdruck x [mm Hg] und Herzgewicht y [g] von 10 an Gehirnblutung gestorbenen Männern (W. M. Gibson u. G. H. Jowett, Appl. Statistics 6, 1957, 115)

x	121	120	95	123	140	112	92	100	102	91
y	521	465	352	455	490	388	301	395	375	418

96 Herleitung zu Abschnitt 94

Aus (94.2) folgt durch Differentiation

$$\frac{\partial a}{\partial b} = -2\sum x_j(y_j - bx_j - k),$$

$$\frac{\partial a}{\partial k} = -2\sum (y_j - bx_j - k).$$

(Summiert wird dabei jeweils über j von 1 bis n.) Diese Ausdrücke setzen wir gleich null. So ergibt sich

$$\sum x_j y_j - b\sum x_j^2 - k\sum x_j = 0,$$

$$\sum y_j - b\sum x_j - nk = 0.$$

Berücksichtigen wir (94.4), so erhalten wir

$$b\sum x_j^2 + nk\bar{x} = \sum x_j y_j,$$

$$b\bar{x} \quad + \quad k \quad = \bar{y}.$$

Dieses System zweier linearer Gleichungen für die Unbekannten b und k hat, da nicht alle x-Werte gleich sind, die Lösung

(96.1) $$b = \frac{\sum x_j y_j - n\bar{x}\bar{y}}{\sum x_j^2 - n\bar{x}^2}, \qquad k = \bar{y} - b\bar{x}.$$

Aus (94.6) und (94.8) sehen wir, daß der Zähler von b gleich $(n-1)s_{xy}$ ist und der Nenner gleich $(n-1)s_1^2$. Damit erhalten wir (94.5). Indem wir den Ausdruck für k in (94.1) einsetzen, wird

$$y = bx + \bar{y} - b\bar{x},$$

und dies ist gleichbedeutend mit (94.3). Damit ist die Darstellung der Regressionsgeraden aus dem Prinzip der kleinsten Quadrate hergeleitet.

Setzen wir den Ausdruck (96.1) für k in (94.2) ein, so ergibt sich

(96.2) $$a = \sum_{j=1}^{n} [(y_j - \bar{y}) - b(x_j - \bar{x})]^2.$$

Da alle Glieder der Summe Quadrate, also nichtnegativ sind, so ist a dann und nur dann gleich null, wenn jeder Summand null ist, wenn also für jedes $j = 1, 2, \ldots, n$ die Beziehung

$$y_j - \bar{y} = b(x_j - \bar{x})$$

gilt. Wegen (94.3) bedeutet dies:

Die Summe (94.2) *der Abstandsquadrate ist genau dann null, wenn alle Punkte der Stichprobe auf der Regressionsgeraden* (94.3) *liegen.*

Dieses Zwischenergebnis war zu erwarten.

Quadrieren wir in (96.2) aus, so folgt

$$a = \sum (y_j - \bar{y})^2 - 2b \sum (y_j - \bar{y})(x_j - \bar{x}) + b^2 \sum (x_j - \bar{x})^2.$$

Führen wir nun die Varianz

$$s_2^2 = \frac{1}{n-1} \sum_{j=1}^{n} (y_j - \bar{y})^2 \tag{96.3}$$

der y-Werte der Stichprobe ein, so ergibt sich wegen (94.6) und (94.7) sofort

$$a = (n-1)\,[s_2^2 - 2b s_{xy} + b^2 s_1^2].$$

Wegen (94.5) kann man die beiden letzten Glieder zusammenfassen und erhält einfach

$$\boxed{a = (n-1)\,(s_2^2 - b^2 s_1^2).} \tag{96.4}$$

Wegen (94.5) und unseres obigen Zwischenergebnisses folgt aus $a = 0$ demgemäß:

Alle Punkte der Stichprobe liegen genau dann auf der Regressionsgeraden (94.3), *wenn*

$$s_{xy}^2 = s_1^2 s_2^2 \tag{96.5}$$

ist.

97 Regressionsgerade der Grundgesamtheit

Welches Gegenstück haben die in Abschn. 94 auftretenden Größen jeweils bei der (X, Y)-Grundgesamtheit, der die betrachtete Stichprobe entstammt?

Den Mittelwerten $\bar{x}$ und $\bar{y}$ entsprechen natürlich die Mittelwerte der beiden Variablen X und Y, die wir mit μ_1 bzw. μ_2 bezeichnen,

$$\mu_1 = E(X), \qquad \mu_2 = E(Y). \tag{97.1}$$

Den Varianzen s_1^2 und s_2^2 entsprechen die Varianzen

$$\sigma_1^2 = E([X - \mu_1]^2), \qquad \sigma_2^2 = E([Y - \mu_2]^2)$$

der genannten Variablen.

Die Degenerationsfälle $\sigma_1 = 0$ *oder* $\sigma_2 = 0$, *die kein Interesse besitzen, schließen wir stets aus.*

Die Größe

$$\boxed{\sigma_{XY} = E([X - \mu_1]\,[Y - \mu_2])} \tag{97.2}$$

heißt die **Kovarianz** der Variablen X und Y. Indem wir die beiden eckigen Klammern ausmultiplizieren und (58.2) berücksichtigen, sehen wir, daß die rechte Seite gleich

$$E(XY) - \mu_1 E(Y) - \mu_2 E(X) + \mu_1\mu_2$$

ist. Wegen (97.1) wird also

(97.3) $$\sigma_{XY} = E(XY) - E(X)\,E(Y).$$

Dies stimmt mit (59.4) überein.

Ersetzen wir in (94.3) die Größen y, $\bar{y}$, b, x und $\bar{x}$ durch die entsprechenden Größen Y, μ_2,

(97.4) $$\boxed{\beta = \frac{\sigma_{XY}}{\sigma_1^{\,2}},}$$

X und μ_1, so erhalten wir die Gerade

(97.5) $$\boxed{Y - \mu_2 = \beta(X - \mu_1)\,.}$$

Diese heißt die **Regressionsgerade** *der Zufallsvariablen Y bezüglich der Zufallsvariablen X.* Das Steigungsmaß β heißt der zugehörige **Regressionskoeffizient.**

Die Regressionsgerade (94.3) der Stichprobe ergab sich aus der Forderung, daß (94.2) ein Minimum sein sollte. Entsprechend gilt für (97.5) der

Satz 97.1. *Unter allen Geraden $Y = \beta X + \varkappa$ ist* (97.5) *diejenige, für die die mittlere quadratische vertikale Abweichung (vgl. Abb. 97.1)*

(97.6) $$\alpha = E([Y - \beta X - \varkappa]^2)$$

ein Minimum hat. Dieses hat den Wert

(97.7) $$\alpha_{\min} = \sigma_2^{\,2} - \frac{\sigma_{XY}^{\,2}}{\sigma_1^{\,2}}\,.$$

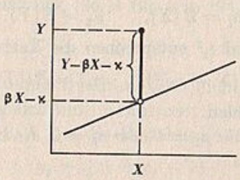


Abbildung 97.1. Zu Satz 97.1

Beweis. Wir schreiben α in der Form

$$\alpha = E\big([(Y - \mu_2) - \beta(X - \mu_1) + (\mu_2 - \beta\mu_1 - \varkappa)]^2\big).$$

Dann wenden wir (58.2) an. Wegen

$$E(Y - \mu_2) = 0, \qquad E(X - \mu_1) = 0$$

und (97.2) ergibt sich auf diese Weise

$$\alpha = \sigma_2^2 + \beta^2 \sigma_1^2 - 2\beta\sigma_{XY} + (\mu_2 - \beta\mu_1 - \varkappa)^2. \tag{97.8}$$

Die Werte von β und $\varkappa$, für die α ein Minimum hat, erhalten wir aus

$$\frac{\partial\alpha}{\partial\beta} = 2\beta\sigma_1^2 - 2\sigma_{XY} - 2\mu_1(\mu_2 - \beta\mu_1 - \varkappa) = 0,$$

$$\frac{\partial\alpha}{\partial\varkappa} = -2(\mu_2 - \beta\mu_1 - \varkappa) = 0.$$

Wegen der letzten Gleichung reduziert sich die erste auf

$$2\beta\sigma_1^2 - 2\sigma_{XY} = 0.$$

Die Lösung ist

$$\beta = \frac{\sigma_{XY}}{\sigma_1^2}.$$

Die andere Gleichung hat die Lösung

$$\varkappa = \mu_2 - \beta\mu_1.$$

Setzen wir diese Werte β und $\varkappa$ in $Y = \beta X + \varkappa$ ein, so ergibt sich (97.5). Setzen wir den Ausdruck für $\varkappa$ in (97.6) ein, so folgt

$$\alpha = E\big([Y - \mu_2 - \beta(X - \mu_1)]^2\big),$$

und dies ist gleich

$$\sigma_2^2 + \beta^2\sigma_1^2 - 2\beta\sigma_{XY}.$$

Durch Einsetzen des Ausdrucks für β ergibt sich (97.7). Der Satz ist damit bewiesen.

Aufgaben zu Abschnitt 97

97.1—97.3 Für die Verteilung mit der jeweils angegebenen Dichte $f(x, y)$ bestimme und zeichne man die Regressionsgerade (97.5).

97.1 $f(x, y) = x + y$ im Quadrat $0 \leq x \leq 1$, $0 \leq y \leq 1$ und $f(x, y) = 0$ außerhalb davon.

97.2 $f(x, y) = (5 - x - y)/3$ im Rechteck $1 \leq x \leq 3$, $1 \leq y \leq 2$ und $f(x, y) = 0$ außerhalb davon.

97.3 $f(x, y) = 4(1 - x)(1 - y)$ im Quadrat $0 \leq x \leq 1$, $0 \leq y \leq 1$ und $f(x, y) = 0$ außerhalb davon.

98 Konfidenzintervalle für den Regressionskoeffizienten

Bei den bisherigen Überlegungen zur Regression brauchten wir nichts über die Art der Verteilung der Variablen X und Y vorauszusetzen. Dies müssen wir aber tun, wenn wir nun dazu übergehen, Konfidenzintervalle anzugeben oder Hypothesen zu testen. Wir machen die folgende

Voraussetzung. *Für jedes feste $X = x$ ist Y normalverteilt mit dem Mittelwert*

$$\mu(x) = \beta x + \varkappa \tag{98.1}$$

und der Varianz σ^2. Letztere soll also nicht von x abhängen.

Tab. 98.1 zeigt, wie man ein Konfidenzintervall für den Regressionskoeffizienten β bestimmt. Die theoretische Begründung folgt in Abschn. 101.

Tabelle 98.1. Bestimmung eines Konfidenzintervalles für den Regressionskoeffizienten β unter der obigen Voraussetzung

1. Schritt. Man wähle eine Konfidenzzahl γ (95%, 99% oder dgl.).

2. Schritt. Man bestimme die Lösung c der Gleichung

$$F(c) = \frac{1}{2}(1 + \gamma) \tag{98.2}$$

aus der Tafel 8 der t-Verteilung mit $n - 2$ Freiheitsgraden (siehe Anhang 5).

3. Schritt. Aus der Stichprobe $(x_1, y_1), \ldots, (x_n, y_n)$ berechne man s_1^2, s_2^2, b und a gemäß (94.6), (96.3), (94.5) bzw. (96.4).

4. Schritt. Man berechne

$$l = c\sqrt{a}/s_1\sqrt{(n-1)(n-2)}. \tag{98.3}$$

Das Konfidenzintervall ist

$$\mathrm{KONF}\{b - l \leqq \beta \leqq b + l\}. \tag{98.4}$$

Beispiel 98.1 (Härte von Stahl bei Kaltverformung). In der Werkzeugherstellung hat sich das Kalteinsenken, ein Verfahren des Kaltverformens von Stahl, in den letzten Jahren insbesondere zur Herstellung schwieriger Innenformen immer mehr eingebürgert. Dabei ist es wesentlich, daß man weiß, wie sich die Eigenschaften des Stahls bei der Kaltverformung ändern. Tab. 98.2 zeigt einige Versuchsergebnisse dazu. Für diese bestimme man die Regressionsgerade von y bezüglich x und ein 95%-Konfidenzintervall für den Regressionskoeffizienten β.

Tabelle 98.2. Brinellfestigkeit von Kaltarbeitsstählen (556-5) in Abhängigkeit von der Einsenktiefe (K. Schimz, Industr. Organisation **26**, 1957, 107)

Einsenktiefe x_j [mm]	Brinellfestigkeit y_j [kg/mm²]
6	68
9	67
11	65
13	53
22	44
26	40
28	37
33	34
35	32

Es ist

$$\sum x_j = 183 \qquad \text{und} \qquad \sum y_j = 440.$$

Wegen $n = 9$ folgt hieraus

$$\bar{x} = \frac{183}{9} = 20{,}33 \qquad \text{und} \qquad \bar{y} = \frac{440}{9} = 48{,}89.$$

Weiterhin benötigen wir

$$\sum x_j^2 = 4665, \qquad \sum x_j y_j = 7\,701, \qquad \sum y_j^2 = 23\,232.$$

Gemäß (94.8) wird damit

$$s_{xy} = \frac{1}{8}\left(7\,701 - \frac{183 \cdot 440}{9}\right) = -155{,}71.$$

Entsprechend erhalten wir aus (94.6) den Wert

$$s_1^2 = \frac{1}{8}\left(4\,665 - \frac{183^2}{9}\right) = 118{,}00.$$

So ergibt sich der negative Regressionskoeffizient

$$b = \frac{-155{,}71}{118{,}00} = -1{,}32.$$

Die Regressionsgerade hat die Darstellung (vgl. Abb. 98.1)

(98.5) $$y - 48{,}89 = -1{,}32\,(x - 20{,}33).$$

Dies können wir auch folgendermaßen schreiben:

(98.6) $$y = 75{,}73 - 1{,}32x.$$

Wir bestimmen nun ein Konfidenzintervall für den Regressionskoeffizienten.

1. Schritt. $\gamma = 0{,}95$ ist gefordert.

2. Schritt. Für $(1 + \gamma)/2 = 0{,}975$ liefert Tafel 8 mit $n - 2 = 7$ Freiheitsgraden für die Gleichung (98.2) die Lösung $c = 2{,}37$.

3. Schritt. Mit den obigen Zahlenwerten wird

$$s_2^2 = \frac{1}{8}\left(23\,232 - \frac{440^2}{9}\right) = 215{,}11.$$

Damit ergibt sich gemäß (96.4) weiterhin

$$a = 8(215{,}11 - 1{,}32^2 \cdot 118{,}00) = 76$$

(genauer $a = 77{,}15$, wie man sieht, indem man höherstellig rechnet und erst am Schluß abrundet). Aus (98.3) erhalten wir demnach

$$l = \frac{2{,}37\sqrt{76}}{\sqrt{118}\sqrt{56}} = 0{,}254.$$

Wegen $b = -1{,}32$ (s. oben) ist also das Konfidenzintervall

$$\mathrm{KONF}\{-1{,}58 \leqq \beta \leqq -1{,}06\}. \tag{98.7}$$

Abb. 98.2 zeigt ein Geradenpaar mit den Steigungsmaßen $-1{,}58$ und $-1{,}06$.

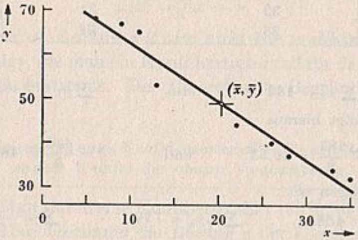


Abbildung 98.1. Stichprobenwerte und Regressionsgerade in Beispiel 98.1

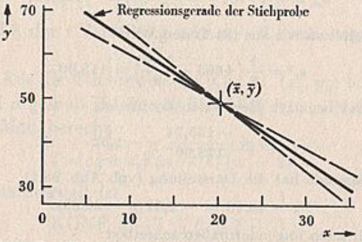


Abbildung 98.2. Zur Veranschaulichung des Konfidenzintervalls (98.7)

Aufgaben zu Abschnitt 98

98.1—98.4 Unter Verwendung der angegebenen Stichprobe bestimme man ein 95%-Konfidenzintervall für den Regressionskoeffizienten β. Dabei nehme man an, daß die Grundgesamtheit der zu Beginn des Abschnittes gemachten Voraussetzung genügt.

98.1 Die Stichprobe in Aufgabe 95.2.

98.2 Die Stichprobe in Beispiel 95.1.

98.3 Die Stichprobe in Aufgabe 95.8.

98.4 Ausdehnung y [%] einer Gelatineschicht in Abhängigkeit von der Luftfeuchtigkeit x [%] (K. Nakamura, Kolloid-Zeitschr. u. Zeitschr. f. Polymere 189, 1963, 131)

x	10	20	30	40
y	0,8	1,6	2,3	2,8

99 Konfidenzintervalle für den Mittelwert

Wir wollen nun ein Konfidenzintervall für den Mittelwert [vgl. (98.1)]

$$\mu = \beta x + \varkappa \tag{99.1}$$

bestimmen. Dabei machen wir wieder die am Anfang des vorigen Abschnittes genannte Voraussetzung. Wie man vorzugehen hat, zeigt Tab. 99.1. Die zugehörige Theorie folgt in Abschn. 101.

Tabelle 99.1. Bestimmung eines Konfidenzintervalles für den Mittelwert (99.1) unter der im vorigen Abschnitt gemachten Voraussetzung

1. Schritt. Man wähle eine Konfidenzzahl γ (95%, 99% oder dgl.).

2. Schritt. Man bestimme die Lösung c der Gleichung

$$F(c) = \frac{1}{2}(1 + \gamma) \tag{99.2}$$

aus der Tafel 8 der t-Verteilung mit $n - 2$ Freiheitsgraden (siehe Anhang 5).

3. Schritt. Aus der Stichprobe $(x_1, y_1), \ldots, (x_n, y_n)$ berechne man s_1^2, s_2^2, b und a gemäß (94.6), (96.3), (94.5) bzw. (96.4).

4. Schritt. Man berechne y gemäß (94.3) und

$$l = c\frac{h\sqrt{a}}{\sqrt{n-2}} \quad \text{mit} \quad h^2 = \frac{1}{n} + \frac{(x-\bar{x})^2}{(n-1)\,s_1^2}. \tag{99.3}$$

Das Konfidenzintervall ist

$$\text{KONF}\ \{y - l \leqq \mu \leqq y + l\}. \tag{99.4}$$

Das Konfidenzintervall (99.4) hat die Länge $2l$. Diese hängt nach Wahl einer Konfidenzzahl und Vorgabe einer Stichprobe noch von x ab, da h von x abhängt. Sie wird für $x = \bar{x}$ am kleinsten. Dies erkennt man aus (99.3) (vgl. auch Abb. 99.1). Je mehr man sich von $x = \bar{x}$ entfernt, desto größer wird die genannte Länge. Das wird verständlich, wenn man sich an das Konfidenzintervall für β (und die Abb. 98.2) erinnert.

Beispiel 99.1. Man bestimme ein 95%-Konfidenzintervall für den Mittelwert unter Benutzung der in Beispiel 98.1 betrachteten Stichprobe.

Die in Beispiel 98.1 berechneten Größen können wir zum Teil benutzen. Gemäß (98.6) ist

$$y = 75{,}73 - 1{,}32x.$$

Dies brauchen wir in (99.4). Gemäß (99.3) ist weiterhin

$$l = 2{,}37\,\frac{h\sqrt{76}}{\sqrt{7}} = 7{,}8h$$

mit

$$h = \sqrt{\frac{1}{9} + \frac{(x - 20{,}33)^2}{944}}.$$

h wird für $x = \bar{x} = 20{,}33$ am kleinsten. Dann ist $h = 1/3$ und $l = 7{,}8/3 = 2{,}6$. Das zugehörige Konfidenzintervall hat die Endpunkte

$$48{,}9 - 2{,}6 = 46{,}3 \qquad \text{und} \qquad 48{,}9 + 2{,}6 = 51{,}5.$$

Je mehr x von $\bar{x}$ abweicht, desto größer wird h und desto länger wird das Konfidenzintervall (99.4); vgl. Tab. 99.2.

Denken wir uns die beiden Endpunkte des Konfidenzintervalls für jedes x in die xy-Ebene eingetragen, so erhalten wir die Kurven C_1 und C_2 in Abb. 99.1. Aus diesen können wir das Intervall (99.4) für jedes x bequem (wenn auch etwas ungenau) entnehmen. Zum Beispiel ergibt sich für $x = 18$ aus der Zeichnung das Intervall

$$\text{KONF}\{49 \leq \mu \leq 55\}.$$

Den Bereich zwischen C_1 und C_2 bezeichnen wir als den **Konfidenzbereich.**

Tabelle 99.2. Konfidenzintervall (99.4) in Beispiel 99.1

x	h	$7{,}8\,h$	$75{,}7 - 1{,}32\,x$	Konfidenzintervall (99.4)
5	0,600	4,7	69,1	64,4 ··· 73,8
10	0,473	3,7	62,5	58,8 ··· 66,2
15	0,376	2,9	55,9	53,0 ··· 58,8
20	0,334	2,6	49,3	46,7 ··· 51,9
20,33	0,333	2,6	48,9	46,3 ··· 51,5
25	0,366	2,9	42,7	39,8 ··· 45,6
30	0,458	3,6	36,1	32,5 ··· 39,7
35	0,582	4,5	29,5	25,0 ··· 34,0

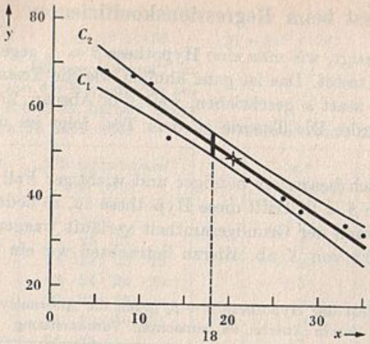


Abbildung 99.1. Konfidenzbereich für den Mittelwert in Beispiel 99.1

Aufgaben zu Abschnitt 99

99.1—99.4 Unter Benutzung der jeweils gegebenen Stichprobe bestimme man ein 95%-Konfidenzintervall für den Mittelwert (99.1) und stelle dieses ähnlich wie in Abb. 99.1 graphisch dar. Dabei nehme man an, daß die in Abschn. 98 getroffene Voraussetzung zutrifft.

99.1 Die Stichprobe in Beispiel 95.1.

99.2

x	1	2	4	6	8
y	2	0	2	4	3

99.3 Länge x [cm] und Kopfumfang y [cm] (fronto-occipital, also Hutmaß) ausgetragener neugeborener Knaben (Universitäts-Frauenklinik Graz, Direktor Prof. Dr. E. Navratil)

x	51	47	52	48	52	52	50	48	54	50
y	34	35	36	34	37	36	35	33	38	34

99.4 Zielachsenfehler y [cc] eines Theodoliten Theo 010 in Abhängigkeit von der Temperatur x [°C] (H. Krause, Vermessungstechnik **11**, 1963, 191)

x	8	9	9	12	13	14	14	16	16	16
y	−55	−55	−53	−50	−50	−35	−37	−35	−27	−24

100 Test beim Regressionskoeffizienten

Tab. 100.1 zeigt, wie man eine Hypothese $\beta = \beta_0$ gegen eine Alternative $\beta > \beta_0$ testet. Das ist ganz ähnlich wie die Tests in Kap. 14. Wir haben α^* statt α geschrieben, weil α in Abschn. 97 anderweitig verwendet wurde. Die Theorie zu dem Test folgt im nächsten Abschnitt.

Ein praktisch besonders häufiger und wichtiger Fall ist der Test der Hypothese $\beta = 0$. Trifft diese Hypothese zu, so bedeutet das, die Regressionsgerade der Grundgesamtheit verläuft waagerecht, und Y hängt gar nicht von X ab. Hierzu betrachten wir ein Beispiel.

Tabelle 100.1. Test der Hypothese $\beta = \beta_0$ gegen die Alternative $\beta > \beta_0$ unter der in Abschn. 98 gemachten Voraussetzung

1. Schritt. Man wähle eine Signifikanzzahl α^* (5%, 1% oder dgl.).

2. Schritt. Man bestimme die Zahl c aus

$$P(T \leqq c) = 1 - \alpha^* \tag{100.1}$$

und der Tafel 8 der t-Verteilung mit $n - 2$ Freiheitsgraden (siehe Anhang 5).

3. Schritt. Aus der Stichprobe $(x_1, y_1), \ldots, (x_n, y_n)$ berechne man $s_1{}^2$, $s_2{}^2$, b und a gemäß (94.6), (96.3), (94.5) bzw. (96.4).

4. Schritt. Man berechne

$$t_0 = s_1 \sqrt{(n-1)(n-2)}\, \frac{b - \beta_0}{\sqrt{a}}. \tag{100.2}$$

Ist $t_0 \leqq c$, so wird die Hypothese angenommen. Ist $t_0 > c$, so wird sie verworfen.

Beispiel 100.1 („Ellbogenabstand" in Abhängigkeit von der Körpergröße). Bei einer Untersuchung über die optimale Arbeitstischhöhe interessierte neben anderen Maßen auch der „Ellbogenabstand" (= Abstand zwischen der Sitzfläche des Stuhles und der Ellbogengelenkachse der auf dem Stuhl sitzenden Person bei vertikal gehaltenem Oberarm). Tab. 100.2 zeigt eine Stichprobe von 22 Wertepaaren in bequemer Schreibweise. y-Werte mit demselben x-Wert stehen jeweils in einer Zeile, und der x-Wert ist nur einmal hingeschrieben. Die zugehörige Abb. 100.1 zeigt, daß die Stichprobenwerte stark streuen. Wir wollen untersuchen, ob die Tendenz besteht, daß der Ellbogenabstand mit wachsender Körpergröße zunimmt. Hierzu bestimmen wir zuerst die Regressionsgerade der y-Werte bezüglich der x-Werte und testen dann die Hypothese $\beta = 0$ gegen die Alternative $\beta > 0$.

Wir setzen

$$x_j = x_j^* + 159, \qquad y_j = y_j^* + 23. \tag{100.3}$$

Tabelle 100.2. „Ellbogenabstand" in Abhängigkeit von der Körpergröße bei 22 Frauen (M. Enzmann, Industr. Organisation 27, 1958, 185)

Körpergröße x_j [cm]	„Ellbogenabstand" y_j [cm]
159	24 24
160	23 25 27
161	24 24 25 26
162	23 24 24 29
166	23 23 25
168	24 29 30 31
172	24 25

Tabelle 100.3. Zu Beispiel 100.1

$x_j^* = x_j - 159$	$y_j^* = y_j - 23$
0	1 1
1	0 2 4
2	1 1 2 3
3	0 1 1 6
7	0 0 2
9	1 6 7 8
13	1 2

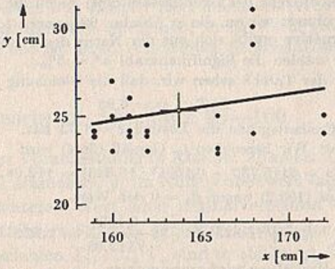


Abbildung 100.1. Werte in Tab. 100.2 und Regressionsgerade

Dann erhalten wir aus Tab. 100.2 die bequemeren Werte in Tab. 100.3 und daraus

$$\sum x_j^* = 0 \cdot 2 + 1 \cdot 3 + 2 \cdot 4 + \cdots + 13 \cdot 2 = 106$$

$$\sum x_j^{*2} = 0^2 \cdot 2 + 1^2 \cdot 3 + \cdots + 13^2 \cdot 2 = 864$$

$$\sum x_j^* y_j^* = 0 \cdot (1 + 1) + 1 \cdot (0 + 2 + 4) + \cdots + 13 \cdot (1 + 2) = 295$$

$$\sum y_j^* = 1 + 1 + 0 + \cdots + 8 + 1 + 2 = 50$$

$$\sum y_j^{*2} = 1^2 + 1^2 + \cdots + 8^2 + 1^2 + 2^2 = 234.$$

Wegen (95.2) ergeben sich demnach die Mittelwerte

$$\bar{x} = \bar{x}^* + 159 = \frac{106}{22} + 159 = 163{,}8$$

$$\bar{y} = \bar{y}^* + 23 = \frac{50}{22} + 23 = 25{,}27.$$

Weiterhin benötigen wir die Varianzen

$$s_1^2 = \frac{1}{21}\left[\sum x_j^{*2} - \frac{1}{22}(\sum x_j^*)^2\right] = \frac{864 - 510{,}727}{21} = 16{,}823$$

$$s_2^2 = \frac{1}{21}\left[\sum y_j^{*2} - \frac{1}{22}(\sum y_j^*)^2\right] = \frac{234 - 113{,}636}{21} = 5{,}732$$

(vgl. Abschn. 94 und 95) und die Kovarianz

$$s_{xy} = \frac{1}{21}\left[\sum x_j^* y_j^* - \frac{1}{22}\sum x_j^* \sum y_j^*\right] = \frac{295 - 240{,}909}{21} = 2{,}576\,.$$

Der Regressionskoeffizient (94.5) hat also den Wert

$$b = \frac{2{,}576}{16{,}823} = 0{,}1531\,.$$

Die Regressionsgerade der Stichprobe hat demnach die Darstellung [vgl. (94.3)]

(100.4) $$y - 25{,}27 = 0{,}1531\,(x - 163{,}8)\,.$$

Ob der gewonnene Wert von b signifikant oder nur zufallsbedingt von null abweicht, soll nun der Test der

$$\text{Hypothese } \beta = 0$$

(β = Regressionskoeffizent der Grundgesamtheit) gegen die Alternative $\beta > 0$ erweisen. Dabei nehmen wir an, die in Abschn. 98 gemachte Voraussetzung sei erfüllt. Die Alternative ergibt sich aus der Natur des Problems.

1. Schritt. Wir wählen die Signifikanzzahl $\alpha^* = 5\%$.

2. Schritt. Aus der Tafel 8 sehen wir, daß die Gleichung

$$P(T \leqq c) = 0{,}95$$

für $n - 2 = 20$ Freiheitsgrade die Lösung $c = 1{,}73$ hat.

3. und 4. Schritt. Wir berechnen t_0. Gemäß (96.4) wird

$$a = 21\,(5{,}732 - 0{,}1531^2 \cdot 16{,}823) = 112{,}08\,.$$

So erhalten wir aus (100.2) wegen $\beta_0 = 0$ den Wert

$$t_0 = \sqrt{16{,}823 \cdot 21 \cdot 20}\,\frac{0{,}1531}{\sqrt{112{,}08}} = 1{,}22\,.$$

Es ist $t_0 < c$. Die Hypothese wird angenommen.

Wir begegnen also hier dem Fall, daß die Hypothese $\beta = 0$ angenommen wird, obwohl b relativ groß (nämlich gleich 0,15) ist. Aber in t_0 kommt a vor, und dies ist groß, da unsere Stichprobenwerte stark streuen. So wird t_0 klein. Aus dem numerischen Wert von b allein können wir demnach noch keine Schlüsse ziehen. Vgl. auch Aufgabe 100.3 und 100.6.

Natürlich wäre es denkbar, daß eine andere Stichprobe von Körpergrößen und Ellbogenabständen vielleicht das entgegengesetzte Ergebnis liefert, denn jeder Test ist mit einem Risiko verbunden, wie wir uns in Kapitel 14 ausführlich überlegt haben. Kommen uns also Zweifel, so bleibt nichts anderes übrig, als weitere Stichproben zu erheben und mit deren Hilfe den Test zu wiederholen.

Aufgaben zu Abschnitt 100

100.1—100.4 Unter Benutzung der angegebenen Stichproben teste man die Hypothese $\beta = 0$ gegen die Alternative $\beta > 0$. Dabei nehme man an, daß die in Abschn. 98 gemachte Voraussetzung zutrifft.

100.1 Die Stichprobe in Aufgabe 99.4.

100.2 Erzeugung einer Quecksilberemulsion in Natriumzitratlösung durch Ultraschall; x [sec] = Beschallungsdauer, y [g/l] = dispergierte Phase (L. BERGMANN, Der Ultraschall, 3. Aufl., Berlin: VDI-Verlag, 1942, S. 323)

x	15	30	60	120	300	600
y	4,0	6,3	5,2	6,4	5,5	6,4

100.3

x	0	1	2
y	0	$1+p$	2

(p konstant)

100.4 Sterblichkeit y [%] von Termiten (Reticulitermes lucifugus Rossi) in Abhängigkeit von der Konzentration x [%] von Chlornaphthalin (G. SCHULTZE-DEWITZ, Beiträge zur Kenntnis der stimulierenden Wirkung von Holzschutzmitteln auf holzzerstörende Organismen. Diss. Hamburg 1961)

x	0,01	0,02	0,04	0,08	0,15	0,30	0,50	1,00	2,00
y	12	18	3	27	16	13	10	70	90

100.5 Man berechne a in Beispiel 100.1 direkt aus (94.2)

100.6 Man konstruiere eine Stichprobe, für die b sehr groß ist (z. B. $b = 5$ oder 10 oder 100) und bei deren Verwendung die Hypothese $\beta = 0$ angenommen wird.

101 Theorie zu Abschnitt 98—100

Auf Grund der Voraussetzung in Abschn. 98 spielt X gewissermaßen die Rolle eines Parameters. y_j im Stichprobenwert (x_j, y_j) können wir dann als beobachteten Wert einer normalverteilten Zufallsvariablen Y_j mit dem Mittelwert $\beta x_j + \varkappa$ [vgl. (98.1)] und der Varianz σ^2 auffassen. Diese Variablen $Y_1, \ldots, Y_n$ sind unabhängig. Der Regressionskoeffizient (94.5) ist dann ein beobachteter Wert der Zufallsvariablen

$$B = \frac{1}{(n-1)\,s_1^2} \sum_{j=1}^{n} (x_j - \bar{x})(Y_j - \bar{Y}). \tag{101.1}$$

Hierbei ist natürlich

$$\bar{Y} = \frac{1}{n}(Y_1 + \cdots + Y_n). \tag{101.2}$$

Die Konstante k in (96.1) ist ein beobachteter Wert der Zufallsvariablen

$$K = \bar{Y} - B\bar{x}. \tag{101.3}$$

Die Summe (94.2) ist ein beobachteter Wert der Zufallsvariablen

$$A = \sum_{j=1}^{n} (Y_j - Bx_j - K)^2. \tag{101.4}$$

Satz 101.1. *Unter der Voraussetzung in Abschn. 98 ist B normalverteilt mit dem Mittelwert*

$$\mu_B = \beta \tag{101.5}$$

und der Varianz

$$\sigma_B{}^2 = \frac{\sigma^2}{\sum_{j=1}^{n}(x_j - \bar{x})^2} = \frac{\sigma^2}{(n-1)\, s_1{}^2}. \tag{101.6}$$

Beweis. Da Y_j normalverteilt ist mit dem obengenannten Mittelwert und der Varianz σ^2, so ist

$$V_j = \lambda_j Y_j \quad \text{mit} \quad \lambda_j = \frac{x_j - \bar{x}}{(n-1)\, s_1{}^2} \tag{101.7}$$

wegen Satz 71.2 normalverteilt mit dem Mittelwert $\lambda_j(\beta x_j + \varkappa)$ und der Varianz $\lambda_j{}^2\sigma^2$. In (101.1) ist

$$\sum_{j=1}^{n}(x_j - \bar{x})(Y_j - \bar{Y}) = \sum_{j=1}^{n}(x_j - \bar{x})\, Y_j - \bar{Y}\sum_{j=1}^{n}(x_j - \bar{x}).$$

Die letzte Summe rechts ist null. So wird wegen (101.7)

$$B = \frac{1}{(n-1)\, s_1{}^2}\sum_{j=1}^{n}(x_j - \bar{x})\, Y_j = \sum_{j=1}^{n}\lambda_j\, Y_j = \sum_{j=1}^{n} V_j. \tag{101.8}$$

Die Summanden V_j sind unabhängig, weil die Y_j unabhängig sind. Also ist B gemäß Satz 71.1 normalverteilt mit dem Mittelwert

$$\mu_B = \sum_{j=1}^{n}\lambda_j(\beta x_j + \varkappa) = \beta\sum_{j=1}^{n}\lambda_j\, x_j + \varkappa\sum_{j=1}^{n}\lambda_j \tag{101.9}$$

und der Varianz

$$\sigma_B{}^2 = \sum_{j=1}^{n}\lambda_j{}^2\,\sigma^2. \tag{101.10}$$

Wir setzen λ_j [vgl. (101.7)] in (101.9) ein. Wegen $\sum(x_j - \bar{x}) = 0$ ist auch $\sum\lambda_j = 0$. Übrig bleibt

$$\mu_B = \frac{\beta}{(n-1)\, s_1{}^2}\sum_{j=1}^{n}(x_j - \bar{x})\, x_j.$$

In dieser Formel ist

$$\sum_{j=1}^{n}(x_j - \bar{x})\, x_j = \sum_{j=1}^{n} x_j{}^2 - n\bar{x}^2 = \sum_{j=1}^{n}(x_j - \bar{x})^2 = (n-1)\, s_1{}^2.$$

Damit ergibt sich (101.5). Die Formel (101.6) folgt sofort durch Einsetzen von λ_j in (101.10). Damit ist der Satz 101.1 bewiesen.

Satz 101.2. *Unter der Voraussetzung in Abschn. 98 hat*

$$T_1 = s_1 \sqrt{(n-1)(n-2)}\,\frac{B-\beta}{\sqrt{A}} \tag{101.11}$$

mit A und B gemäß (101.4) *bzw.* (101.1) *eine t-Verteilung mit $n-2$ Freiheitsgraden.*

Beweisskizze. Gemäß Satz 71.2 und 101.1 ist

$$W = \frac{B-\mu_B}{\sigma_B} = (B-\beta)\,\frac{s_1\sqrt{n-1}}{\sigma}$$

normalverteilt mit dem Mittelwert 0 und der Varianz 1. Die Variablen W und $U = A/\sigma^2$ sind unabhängig, und U hat eine Chi-Quadrat-Verteilung mit $n-2$ Freiheitsgraden. Diese Aussage, die wir nicht beweisen, ähnelt dem Satz 73.1. Sie wird plausibel, wenn man bedenkt, daß a, der beobachtete Wert von A, dem Wesen nach (von einem konstanten Faktor abgesehen) der Stichprobenvarianz im Falle einer einzelnen Zufallsvariablen entspricht. Daß man $n-2$ Freiheitsgrade (statt $n-1$) erhält, wird plausibel, wenn man beachtet, daß für $n=2$ stets $a=0$ gilt, weil ja eine Gerade durch 2 Punkte bestimmt ist. Also erfüllt $W/\sqrt{U/(n-2)}$ die in Abschn. 62 genannten Voraussetzungen (mit $n-2$ statt n), hat also eine t-Verteilung mit $n-2$ Freiheitsgraden. Diese Variable ist aber mit T_1 in (101.11) identisch, wie man leicht nachrechnet. Damit ergibt sich der Satz 101.2.

Auf diesen Satz gründet sich die Vorschrift in Tab. 98.1. Aus

$$P(-c \leqq T_1 \leqq c) = F(c) - F(-c) = \gamma$$

und der Symmetrie der t-Verteilung folgt (98.2). Setzt man (101.11) in $-c \leqq T_1 \leqq c$ ein, verwandelt man das Resultat in eine Ungleichung für β und ersetzt man schließlich A und B durch die beobachteten Werte a bzw. b, so ergibt sich (98.4).

Der Satz 101.2 bildet auch die Grundlage des Tests im vorigen Abschnitt. In der Tat ist (100.2) ein beobachteter Wert von (101.11) mit $\beta = \beta_0$. Die Wahrscheinlichkeit, daß diese Variable irgendeinen Wert größer als c annimmt, ist im Falle der Richtigkeit der Hypothese gering. Dies sieht man aus (100.1), denn $1-\alpha^*$ ist groß, weil α^* klein ist. So wird die Entscheidung in Tab. 100.1 verständlich.

Die in Tab. 99.1 angegebene Vorschrift folgt aus dem

Satz 101.3. *Unter der in Abschn. 98 genannten Voraussetzung hat die Zufallsvariable*

$$T_2 = \sqrt{n-2}\,\frac{(x-\bar{x})\,B + \bar{Y} - \mu}{h\sqrt{A}} \tag{101.12}$$

[mit A, B, $\bar{Y}$ und h gemäß (101.4), (101.1), (101.2) *bzw.* (99.3)] *eine t-Verteilung mit $n-2$ Freiheitsgraden.*

Beweisskizze. Gemäß Satz 71.1 ist $Y_1 + \cdots + Y_n$ normalverteilt mit dem Mittelwert

$$\sum_{j=1}^{n} (\beta x_j + \varkappa) = \beta n \bar{x} + n \varkappa$$

und der Varianz $n\sigma^2$. Also ist $\bar{Y}$ gemäß Satz 71.2 normalverteilt mit dem Mittelwert $\beta\bar{x} + \varkappa$ und der Varianz σ^2/n. Wegen Satz 71.2 und 101.1 ist $(x - \bar{x}) B$ normalverteilt mit dem Mittelwert $(x - \bar{x}) \beta$ und der Varianz

$$\sigma^2 \frac{(x - \bar{x})^2}{(n-1)\, s_1^{\,2}}.$$

$(x - \bar{x}) B$ und $\bar{Y}$ sind unabhängig, wie wir ohne Beweis angeben. Gemäß Satz 71.1 ist dann

$$\tilde{Y} = (x - \bar{x}) B + \bar{Y} \tag{101.13}$$

normalverteilt mit dem Mittelwert

$$(x - \bar{x}) \beta + \beta \bar{x} + \varkappa = x\beta + \varkappa = \mu$$

und der Varianz

$$\tilde{\sigma}^2 = \sigma^2 \frac{(x - \bar{x})^2}{(n-1)\, s_1^{\,2}} + \frac{\sigma^2}{n} = h^2 \sigma^2 \tag{101.14}$$

mit h^2 gemäß (99.3). Also ist

$$Z = \frac{\tilde{Y} - \mu}{\tilde{\sigma}} = \frac{(x - \bar{x}) B + \bar{Y} - \mu}{h\sigma}$$

normalverteilt mit dem Mittelwert 0 und der Varianz 1. Wie schon gesagt wurde, hat $U = A/\sigma^2$ eine Chi-Quadrat-Verteilung mit $n-2$ Freiheitsgraden. Z und U sind unabhängig, wie wir ohne Beweis angeben. Also hat der Quotient

$$\frac{Z}{\sqrt{U/(n-2)}} = \frac{(x - \bar{x}) B + \bar{Y} - \mu}{h\sigma \sqrt{A/\sigma^2 (n-2)}}$$

eine t-Verteilung mit $n-2$ Freiheitsgraden (vgl. Abschn. 62). Dieser Quotient ist aber gerade mit T_2 in (101.12) identisch. Damit erhalten wir den Satz 101.3.

In Tab. 99.1 ergibt sich (99.2) ganz entsprechend, wie dies für (98.2) oben erläutert wurde. (99.4) ergibt sich, indem man (101.12) in $-c \leqq T_2 \leqq c$ einsetzt, dies in eine Ungleichung für μ verwandelt und dann A, B und $\bar{Y}$ durch a, b, bzw. $\bar{y}$ ersetzt (vgl. Aufgabe 101.3).

Aufgaben zu Abschnitt 101

101.1 Aus $-c \leq T_1 \leq c$ gewinne man (98.4) auf die im Text angegebene Weise.

101.2 Man bestätige, daß (101.11) und $W/\sqrt{U/(n-2)}$ identisch sind.

101.3 Man gewinne (99.4) auf die im vorliegenden Abschnitt angegebene Weise aus $-c \leq T_2 \leq c$.

102 Regression und Varianzanalyse

Wir wollen uns überlegen, daß man die Methoden der Varianzanalyse auch bei Regressionsproblemen anwenden kann.

Hierzu schreiben wir jeden Wert y_i in einer vorgegebenen Stichprobe $(x_1, y_1), \ldots, (x_n, y_n)$ in der Form (vgl. Abb. 102.1)

(102.1) $$y_i = \bar{y} + y_i^* + d_i.$$

d_i ist hierbei der vertikale Abstand des Stichprobenwertes (x_i, y_i) von der Regressionsgeraden [vgl. (94.3)]

$$y - \bar{y} = b(x - \bar{x}).$$

Weiterhin ist

$$y_i^* = b(x_i - \bar{x}).$$

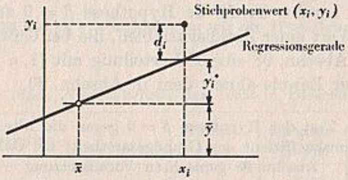


Abbildung 102.1. Zu Formel (102.1)

Die Summe der Quadrate der d_i bezeichnen wir mit q_2. Diese Größe q_2 ist mit a in (94.2), bezogen auf die Regressionsgerade, identisch. Aus (96.4) folgt demnach

$$q_2 = \sum_{i=1}^{n} d_i^2 = (n-1)\, s_2^2 - (n-1)\, b^2 s_1^2.$$

Den ersten Summanden auf der rechten Seite bezeichnen wir mit q und den letzten mit q_1. Dann gilt also

(102.2) $$q = q_1 + q_2,$$

und wegen (96.3) und (94.5) ist dabei

(102.3a) $$q = \sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2$$

und

$$(102.3\text{b}) \qquad q_1 = (n-1)\, b^2 s_1^2 = (n-1)\,\frac{s_{xy}^2}{s_1^2}\,.$$

q_1 hängt nur von b ab, hat also nur 1 Freiheitsgrad. q hat $n-1$ Freiheitsgrade, wie wir aus Abschn. 8 wissen, und $q_2 = q - q_1$ hat demnach $n-1-1 = n-2$ Freiheitsgrade. Der Zerlegung (102.2) entspricht die Tab. 102.1.

Tabelle 102.1. Schema der Varianzanalyse zur Zerlegung (102.2)

Variation	Freiheitsgrade	Quadratsumme	Durchschnittsquadrat
Auf der Regression	1	q_1	q_1
Abweichung von der Regression	$n-2$	q_2	$q_2/(n-2)$
Insgesamt	$n-1$	q	

Tab. 102.2 enthält die Vorschrift für einen Test, der auf der Zerlegung (102.2) beruht. Trifft die Hypothese $\beta = 0$ zu, so ist v_0 ein beobachteter Wert einer Zufallsvariablen, die bei Gültigkeit der Voraussetzung in Abschn. 98 eine F-Verteilung mit $(1, n-2)$ Freiheitsgraden hat. Der Beweis ähnelt dem in Abschn. 90.

Tabelle 102.2. Test der Hypothese $\beta = 0$ gegen die Alternative $\beta \neq 0$ (β = Regressionskoeffizient der Grundgesamtheit) bei Gültigkeit der in Abschn. 98 gemachten Voraussetzung

1. Schritt. Aus der gegebenen Stichprobe $(x_1, y_1), \ldots, (x_n, y_n)$ berechne man q und q_1 gemäß (102.3), sodann $q_2 = q - q_1$ und

$$v_0 = \frac{q_1}{q_2/(n-2)}\,.$$

2. Schritt. Man wähle eine Signifikanzzahl α (5% oder 1%).

3. Schritt. Man bestimme die Zahl c aus

$$P(V \leqq c) = 1 - \alpha$$

und der Tafel 9a bzw. 9b der F-Verteilung mit $(1, n-2)$ Freiheitsgraden (s. Anhang 5). Ist $v_0 \leqq c$, so wird die Hypothese $\beta = 0$ angenommen. Ist $v_0 > c$, so wird sie verworfen, und man nimmt an, daß $\beta \neq 0$ ist.

Beispiel 102.1. Man teste die Hypothese $\beta = 0$ unter Benutzung der Stichprobe in Tab. 100.2. Dabei nehme man an, daß die in Abschn. 98 angegebene Voraussetzung gilt.

1. Schritt. Aus Tab. 100.2 erhalten wir

$$q = \sum y_i^2 - \frac{1}{n}(\sum y_i)^2 = 120{,}36 .$$

Wir dürfen auch Tab. 100.3 verwenden (warum?) und erhalten einfacher

$$q = q^* = \sum y_i^{*2} - \frac{1}{n}(\sum y_i^*)^2 = 234 - \frac{1}{22}\cdot 50^2 = 120{,}36.$$

Weiterhin wird (vgl. Beispiel 100.1)

$$q_1 = 21 \cdot \frac{2{,}576^2}{16{,}823} = 8{,}28 \qquad \text{und} \qquad q_2 = q - q_1 = 112{,}08.$$

q_2 stimmt mit a in Beispiel 100.1 überein. Damit ergibt sich (vgl. Tab. 102.3)

$$v_0 = \frac{8{,}28}{5{,}60} = 1{,}48.$$

2. Schritt. Wir wählen $\alpha = 5\%$.

3. Schritt. Aus Tafel 9a entnehmen wir bei (1, 20) Freiheitsgraden als Lösung der Gleichung

$$P(V \leq c) = 0{,}95$$

den Wert $c = 4{,}35$. Da $v_0 < c$ ist, nehmen wir die Hypothese an, wie in Beispiel 100.1.

Tabelle 102.3. Zum 1. Schritt des Tests in Beispiel 102.1

Variation	Freiheitsgrade	Quadratsumme	Durchschnittsquadrat
Auf der Regression	1	8,28	8,28
Abweichung von der Regression	20	112,08	5,60
Insgesamt	21	120,36	

Aufgaben zu Abschnitt 102

102.1—102.5 Man wende den im Text erläuterten Test der Hypothese $\beta = 0$ unter Benutzung der folgenden Stichprobe an:

102.1

x	-1	0	0	1
y	-1	-1	1	1

102.2 (k konstant)

x	-1	0	1
y	-1	k	1

102.3 Parallelitätsfehler y (= Winkel zwischen den Gesenkmittelachsen) in einem Fallhammer mit Stangenführungen in Abhängigkeit vom Umformweg x (D. WATERMANN, Werkstattstechn. **53**, 1963, 419)

x [mm]	y [mm/m]
12	8 9 10 11
13	9
14	10 13
18	10 12 15

102.4 Die Stichprobe in Aufgabe 100.2.
102.5 Die Stichprobe in Aufgabe 99.4.

103 Test der Linearität der Regression

Bisher haben wir die Annahme gemacht, daß die jeweils betrachtete Regression linear ist. In vielen Fällen wird das zutreffen. In anderen Fällen wird man aber im Zweifel sein und möchte dann die genannte Annahme testen. Wir wollen zeigen, daß man dies mit Hilfe der Methode der Varianzanalyse tun kann.

Hierzu fassen wir in einer Stichprobe $(x_1, y_1), \ldots, (x_n, y_n)$ jeweils alle diejenigen y-Werte gruppenweise zusammen, für die das zugehörige x denselben Zahlenwert hat. Beispielsweise können wir also im Fall der Stichprobe

$$(1, 5) \quad (2, 6) \quad (1, 4) \quad (4, 11) \quad (1, 6) \quad (2, 10) \quad (4, 7)$$

drei Gruppen bilden und

$$\begin{array}{llllll} (1, 4) & (1, 5) & (1, 6) & & x = 1 & y = 4, 5, 6 \\ (2, 6) & (2, 10) & & \text{oder noch kürzer} & x = 2 & y = 6, 10 \\ (4, 7) & (4, 11) & & & x = 4 & y = 7, 11 \end{array}$$

schreiben. Für jede Gruppe berechnen wir den Mittelwert und schließen dann folgendermaßen:

Ist die Regression linear, so müssen die genannten Mittelwerte näherungsweise auf einer Geraden liegen, und zwar „verhältnismäßig genau". Das heißt, ihre Abweichung von der Regressionsgeraden darf nicht zu groß sein im Verhältnis zur Abweichung der Werte einer Gruppe von ihrem jeweiligen Mittelwert. Es kommt also darauf an, daß das Verhältnis

$$\frac{\text{Abweichung der Mittelwerte von der Regressionsgeraden}}{\text{Abweichung der } y\text{-Werte vom zugehörigen Mittelwert der Gruppe}}$$

nicht zu groß ist. Ist das Verhältnis nämlich groß, so muß der Zähler groß sein, verglichen mit dem Nenner. Das heißt, dann weichen die Gruppenmittelwerte stark von der Regressionsgeraden ab, obwohl die Werte innerhalb jeder Gruppe nur wenig vom Gruppenmittelwert abweichen, also wenig streuen, also nur geringe Zufallsabweichungen zeigen. In einem solchen Falle ist dann die Annahme oder Hypothese, daß lineare Regression vorliegt, zu verwerfen.

Diesen Gedankengang müssen wir nun in Formeln kleiden. Dazu brauchen wir zuerst einmal geeignete Bezeichnungen: Es sei (bzw. seien)

n der Stichprobenumfang,

r die Anzahl der Gruppen, also die Anzahl der zahlenmäßig verschiedenen x-Werte in der Stichprobe,

$x_1, x_2, \ldots, x_i, \ldots, x_r$ die genannten x-Werte,

n_i die Anzahl der y-Werte in der i-ten Gruppe, also die Anzahl der y-Werte in der Stichprobe, für die $x = x_i$ ist,

$y_{i1}, y_{i2}, \ldots, y_{ij}, \ldots$ die y-Werte in der i-ten Gruppe, wobei also der erste Index die Gruppe bezeichnet und der zweite die Nummer des y-Wertes in der Gruppe,

$\bar{y}_i$ der Mittelwert der y-Werte in der i-ten Gruppe,

$\bar{y}$ der Mittelwert aller y-Werte in der Stichprobe.

Natürlich ist

$$n_1 + n_2 + \cdots + n_r = n.$$

Weiterhin ist

(103.1) $$\bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij} = \frac{1}{n_i} (y_{i1} + y_{i2} + \cdots + y_{in_i})$$

und

(103.2) $$\bar{y} = \frac{1}{n} \sum_{i=1}^{r} \sum_{j=1}^{n_i} y_{ij} = \frac{1}{n} \sum_{i=1}^{r} n_i \, \bar{y}_i .$$

Die gewählten Bezeichnungen ähneln denen in Abschn. 88, aber statt x steht nun y.

Beispiel 103.1. Bei der obigen Stichprobe ist $n = 7$, $r = 3$, $n_1 = 3$, $n_2 = 2$, $n_3 = 2$, sowie

$$\bar{y}_1 = \frac{1}{3}(4 + 5 + 6) = 5, \quad \bar{y}_2 = \frac{1}{2}(6 + 10) = 8, \quad \bar{y}_3 = \frac{1}{2}(7 + 11) = 9$$

und

$$\bar{y} = \frac{1}{7}(4 + 5 + 6 + 6 + 10 + 7 + 11) = 7.$$

Die Stichprobe können wir auch in der Form der Tab. 103.1 anordnen und die Mittelwerte gleich mit eintragen.

Tabelle 103.1. Zu Beispiel 103.1

x_i	n_i	y_{ij}	Zeilen-summe	$\bar{y}_i$
1	3	4 5 6	15	5
2	2	6 10	16	8
4	2	7 11	18	9

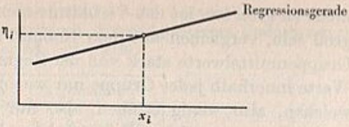


Abbildung 103.1. Zu Formel (103.3)

Nun kommt die Zerlegung der Quadratsumme

(103.3) $$q = \sum_{i=1}^{r} \sum_{j=1}^{n_i} (y_{ij} - \eta_i)^2.$$

Hierbei ist (siehe Abb. 103.1)

$$\eta_i = \bar{y} + b(x_i - \bar{x}).$$

Wir zerlegen q in zwei Bestandteile q_1 und q_2,

(103.4) $$q = q_1 + q_2.$$

Dabei rührt q_1 von der Streuung der Mittelwerte $\bar{y}_i$ um die Regressionsgerade her und hat die Form

(103.5) $$q_1 = \sum_{i=1}^{r} n_i (\bar{y}_i - \eta_i)^2.$$

Der zweite Bestandteil stammt von der Streuung innerhalb der Gruppen unserer Stichprobe und hat die Form

(103.6) $$q_2 = \sum_{i=1}^{r} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2.$$

Der Beweis dieser Zerlegung ähnelt dem in Abschn. 90 und sei dem Leser überlassen (s. Aufgabe 103.1). Der Zerlegung entspricht die Tab. 103.2. Unser jetziges q in (103.3) enthält die Abweichungen von η_i und nicht wie q in Abschn. 88 die Abweichungen von $\bar{y}$. Dadurch scheiden wir die Variation infolge der Regression aus. Wir verlieren dabei jeweils einen Freiheitsgrad. So ergeben sich die Freiheitsgrade in Tab. 103.2.

Tabelle 103.2. Schema der Varianzanalyse zur Zerlegung (103.4)

Variation	Freiheits-grade	Quadrat-summe	Durchschnitts-quadrat
Mittelwerte um die Regression	$r - 2$	q_1	$q_1/(r-2)$
Innerhalb der Gruppen	$n - r$	q_2	$q_2/(n-r)$
Insgesamt	$n - 2$	q	

Die Zerlegung (103.4) kann man dazu benutzen, um zu testen, ob die Regression als linear angesehen werden darf. Tab. 103.3 enthält die zugehörige Anweisung. Ist die Hypothese richtig, so ist (103.7) ein beobachteter Wert einer Zufallsvariablen V, die unter der Voraussetzung in Abschn. 98 eine F-Verteilung mit $(r-2, n-r)$ Freiheitsgraden hat. Dies ergibt sich ähnlich wie in Abschn. 90. Damit wird die Entscheidung in Tab. 103.3 verständlich.

Tabelle 103.3. Test der Hypothese, daß die Regression linear ist, unter der Voraussetzung, daß Y für jedes feste $X = x$ normalverteilt ist, wobei die Varianz nicht von X abhängt

1. Schritt. Aus der gegebenen Stichprobe berechne man q und q_1 gemäß (103.3) [oder (96.4)] bzw. (103.5), sodann $q_2 = q - q_1$ und

$$v_0 = \frac{q_1/(r-2)}{q_2/(n-r)}. \tag{103.7}$$

2. Schritt. Man wähle eine Signifikanzzahl α (5% oder 1%).

3. Schritt. Man bestimme die Lösung c der Gleichung

$$P(V \leqq c) = 1 - \alpha$$

aus der Tafel 9a bzw. 9b der F-Verteilung mit $(r-2, n-r)$ Freiheitsgraden. Ist $v_0 \leqq c$, so wird die Hypothese angenommen. Ist $v_0 > c$, so wird sie verworfen.

Daß man beim 1. Schritt auch (96.4) heranziehen kann, kommt daher, daß q mit a (bezogen auf die Regressionsgerade) in (94.2) identisch ist.

Wird die Hypothese angenommen, so bedeutet das natürlich keineswegs, daß eine Gerade die einzige mögliche oder die beste mit den Beobachtungen verträgliche Regressionslinie ist. Wird die Hypothese verworfen, so bedeutet dies, daß man die Regression beim besten Willen nicht mehr als linear ansehen kann (abgesehen von der Irrtumsmöglichkeit, die jedem Test innewohnt, wie wir wissen).

Beispiel 103.2. In Beispiel 100.1 und 102.1 hatten wir die Regression als linear angesehen. Diese Annahme können wir nun nachträglich mit Hilfe unseres Tests rechtfertigen:

1. Schritt. Aus Beispiel 100.1 folgt $q = a = 112{,}1$. Aus der Tab. 100.3 ergibt sich die Tab. 103.4. Hierbei ist [vgl. (100.3) und (100.4)]

$$\eta_i^* = y^*(x_i^*) = 0{,}1531\, x_i^* + 1{,}535.$$

Die y_i^* hängen mit den y_i durch die Nullpunktsverschiebung (100.3) zusammen. Deshalb ist in (103.5) einfach

$$\bar{y}_i - \eta_i = \bar{y}_i^* - \eta_i^*.$$

So ergibt sich aus Tab. 103.4 der Wert $q_1 = 47{,}2$. Gemäß (103.4) folgt hieraus

$$q_2 = q - q_1 = 112{,}1 - 47{,}2 = 64{,}9$$

und weiterhin aus Tab. 103.5 der Wert $v_0 = 9{,}4/4{,}3 = 2{,}2$.

2. Schritt. Wir wählen die Signifikanzzahl $\alpha = 5\%$.

3. Schritt. Für (5, 15) Freiheitsgrade liefert die Tafel 9a als Lösung der Gleichung

$$P(V \leqq c) = 1 - 0{,}05 = 0{,}95$$

den Wert $c = 2{,}90$. Es ist $v_0 < c$. Wir nehmen deshalb die Hypothese an.

Tabelle 103.4. Zum 1. Schritt in Beispiel 103.2

y_{ij}^*	n_i	$\bar{y}_i^*$	η_i^*	$(\bar{y}_i^* - \eta_i^*)^2$	$n_i(\bar{y}_i^* - \eta_i^*)^2$
1 1	2	1,000	1,535	0,286	0,572
0 2 4	3	2,000	1,688	0,097	0,291
1 1 2 3	4	1,750	1,841	0,008	0,032
0 1 1 6	4	2,000	1,994	0,000	0,000
0 0 2	3	0,667	2,607	3,764	11,292
1 6 7 8	4	5,500	2,913	6,693	26,772
1 2	2	1,500	3,525	4,101	8,202
				Summe	47,161

Tabelle 103.5. Schema der Varianzanalyse in Beispiel 103.2

Variation	Freiheitsgrade	Quadratsumme	Durchschnittsquadrat
Mittelwerte um die Regression	5	47,2	9,4
Innerhalb der Gruppen	15	64,9	4,3
Insgesamt	20	112,1	

Aufgaben zu Abschnitt 103

103.1 Man beweise die Zerlegung (103.4).

103.2 In Beispiel 103.2 berechne man q_2 zur Kontrolle direkt.

103.3—103.5 Unter Benutzung der gegebenen Stichproben teste man jeweils die Linearität der Regression. Dabei nehme man an, daß die in Tab. 103.3 genannte Voraussetzung erfüllt ist.

103.3

x	y	
0	−1	1
1	0	2
2	1	5
4	8	12
6	11	13

103.4

x	y	
1	0	2
2	5	7
3	9	13
4	10	14

103.5

x	y	
2	60	64
5	171	176
10	242	245
15	298	304
20	345	351
25	383	388

In Aufgabe 103.5 ist x die Stromstärke [in mA] und y die Leuchtdichte [in willkürlichen Einheiten] bei einem Kanalstrahlrohr. (R. Gebauer und E. Kreyszig, Zeitschr. f. Physik **135**, 1953, 349—360.)

104 Nichtlineare Regression. Prinzip der kleinsten Quadrate

In Abschn. 94 haben wir gesehen, wie man zu einer Stichprobe $(x_1, y_1), \ldots, (x_n, y_n)$ diejenige Gerade

$$y(x) = bx + k$$

bestimmt, für die die Summe

$$a = \sum_{i=1}^{n} [y_i - y(x_i)]^2 \tag{104.1}$$

der Quadrate der vertikalen Abstände ein Minimum ist. Nach dem entsprechenden Verfahren kann man statt einer Geraden ebensogut eine gekrümmte Kurve, etwa gegeben durch

$$y(x) = b_0 + b_1 x + \cdots + b_m x^m \tag{104.2}$$

bestimmen, wenn dies in einem praktischen Fall angezeigt erscheint. Hierzu berechnet man die n Werte

$$y(x_i) = b_0 + b_1 x_i + \cdots + b_m x_i^m$$

und setzt diese in (104.1) ein. a hat für diejenigen Werte $b_0, b_1, \ldots, b_m$ ein Minimum, für die

$$\frac{\partial a}{\partial b_0} = 0, \qquad \frac{\partial a}{\partial b_1} = 0, \qquad \ldots, \qquad \frac{\partial a}{\partial b_m} = 0 \tag{104.3}$$

ist. Diese $m + 1$ linearen Gleichungen in den $m + 1$ Unbekannten $b_0, \ldots, b_m$ heißen die **Normalgleichungen.** Ihre Lösung gibt die gesuchten Werte der Koeffizienten in (104.2). Theoretische Schwierigkeiten bestehen nicht. Praktisch kann das Lösen dieser Gleichungen natürlich oftmals noch erhebliche Rechenarbeit verursachen. Ein wichtiges Verfahren zur Lösung eines solchen Gleichungssystems ist der sogenannte **Gaußsche Algorithmus,** ein systematisches Eliminationsverfahren, das wir nachstehend an einem Beispiel (Beispiel 104.1) erläutern.

Im Falle einer Parabel

$$y(x) = b_0 + b_1 x + b_2 x^2 \tag{104.4}$$

ist

$$y(x_i) = b_0 + b_1 x_i + b_2 x_i^2.$$

Dann gewinnt (104.1) die Gestalt

$$a = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_i - b_2 x_i^2)^2.$$

Hieraus ergibt sich durch Differentiation

$$\frac{\partial a}{\partial b_0} = -2 \sum (y_i - b_0 - b_1 x_i - b_2 x_i^2) = 0$$

$$\frac{\partial a}{\partial b_1} = -2 \sum x_i (y_i - b_0 - b_1 x_i - b_2 x_i^2) = 0$$

$$\frac{\partial a}{\partial b_2} = -2 \sum x_i^2 (y_i - b_0 - b_1 x_i - b_2 x_i^2) = 0.$$

Das sind 3 lineare Gleichungen mit den 3 Unbekannten b_0, b_1 und b_2. Diese Normalgleichungen können wir in der folgenden Form schreiben:

$$\begin{aligned} b_0 n + b_1 \sum x_i + b_2 \sum x_i^2 &= \sum y_i \\ b_0 \sum x_i + b_1 \sum x_i^2 + b_2 \sum x_i^3 &= \sum x_i y_i \\ b_0 \sum x_i^2 + b_1 \sum x_i^3 + b_2 \sum x_i^4 &= \sum x_i^2 y_i \end{aligned} \tag{104.5}$$

Tabelle 104.1 Prozentuale Sterblichkeit von Neugeborenen (Schwangerschaftsdauer 280—289 Tage post menstr.) in Abhängigkeit von der Körperlänge bei der Geburt (H. Hosemann, Die Naturwiss. 37, 1950, 410)

Körperlänge		Anzahl Neugeborener in der Klasse	Davon gestorben	Sterblichkeit y [%]
Klassen-intervall	Klassen-mittelpunkt x_i [cm]			
46—49	47,5	124	14	11,29
49—52	50,5	1255	13	1,04
52—55	53,5	3149	28	0,89
55—58	56,5	1441	31	2,15
58—61	59,5	175	16	9,14

Beispiel 104.1 (Sterblichkeit Neugeborener in Abhängigkeit von der Körperlänge). Stellt man die Werte der gruppierten Stichprobe in Tab. 104.1 graphisch dar, so erhält man die Abb. 104.1. Daraus sieht man, daß es naheliegt, die

Regression der Sterblichkeit y bezüglich der Körperlänge x durch eine Parabel zu beschreiben. Die Koeffizienten dieser Parabel (104.4) erhalten wir durch Aufstellen und Lösen des Gleichungssystems (104.5).

Tabelle 104.2. Zu Beispiel 104.1

$x_i^* = x_i - 53{,}5$	x_i^{*2}	x_i^{*3}	x_i^{*4}	Anzahl	y_i	$x_i^* y_i$	$x_i^{*2} y_i$
−6	36	−216	1296	124	11,29	−67,74	406,44
−3	9	− 27	81	1255	1,04	− 3,12	9,36
0	0	0	0	3149	0,89	0	0
3	9	27	81	1441	2,15	6,45	19,35
6	36	216	1296	175	9,14	54,84	329,04

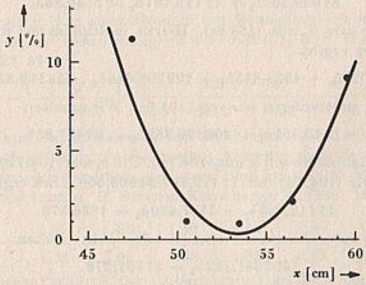


Abbildung 104.1. Die Stichprobe in Tab. 104.1 und die Parabel (104.9)

Die Rechnung wird bequemer, indem wir

$$x_i = x_i^* + 53{,}5 \qquad \text{(also } x_i^* = x_i - 53{,}5\text{)}$$

setzen. Wie wir wissen, denkt man sich bei einer in Klassen eingeteilten Stichprobe die Stichprobenwerte jeweils im Klassenmittelpunkte liegend. Aus Tab. 104.2 erhalten wir demnach

$$\sum x_i^* = 124 \cdot (-6) + 1255 \cdot (-3) + 1441 \cdot 3 + 175 \cdot 6 = 864$$
$$\sum x_i^{*2} = 124 \cdot 36 + 1255 \cdot 9 + 1441 \cdot 9 + 175 \cdot 36 = 35028$$
$$\sum x_i^{*3} = 124 \cdot (-216) + 1255 \cdot (-27) + 1441 \cdot 27 + 175 \cdot 216 = 16038$$
$$\sum x_i^{*4} = 124 \cdot 1296 + 1255 \cdot 81 + 1441 \cdot 81 + 175 \cdot 1296 = 605880$$

denn die Stichprobe besteht aus $n = 6144$ Wertepaaren, in denen man sich nach der Klassenbildung $x_1^* = -6$ genau 124mal, $x_2^* = -3$ genau 1255mal, $\ldots, x_5^* = 6$ genau 175mal vorkommend denkt. Entsprechend wird

$$\sum y_i = 124 \cdot 11{,}29 + 1255 \cdot 1{,}04 + \cdots = 10205{,}42$$
$$\sum x_i^* y_i = 124 \cdot (-6) \cdot 11{,}29 + 1255 \cdot (-3) \cdot 1{,}04 + \cdots = 6576{,}09$$
$$\sum x_i^{*2} y_i = 124 \cdot 36 \cdot 11{,}29 + 1255 \cdot 9 \cdot 1{,}04 + \cdots = 147610{,}71.$$

Die Normalgleichungen (104.5) lauten also im vorliegenden Falle

$$(104.6)\quad \begin{aligned} &(a)\quad 6144\, b_0 + 864\, b_1 + 35028\, b_2 = 10205{,}42\\ &(b)\quad 864\, b_0 + 35028\, b_1 + 16038\, b_2 = 6576{,}09\\ &(c)\quad 35028\, b_0 + 16038\, b_1 + 605880\, b_2 = 147610{,}71\,. \end{aligned}$$

Dieses Gleichungssystem lösen wir mit Hilfe des **Gaußschen Algorithmus,** der folgendermaßen verläuft:

1. Schritt. Elimination von b_0 aus (104.6b) *und* (104.6c). Um b_0 aus (104.6b) zu eliminieren, multiplizieren wir (104.6a) mit 864/6144. Dies ergibt

$$864 b_0 + 121{,}500 b_1 + 4925{,}813 b_2 = 1435{,}137\,.$$

Diese Gleichung subtrahieren wir von (104.6b). Wir erhalten

$$(104.7\text{a})\qquad 34906{,}500 b_1 + 11112{,}187 b_2 = 5140{,}953\,.$$

Nun eliminieren wir b_0 aus (104.6c). Hierzu multiplizieren wir (104.6a) mit 35028/6144. Dies ergibt

$$35028 b_0 + 4925{,}813 b_1 + 199700{,}648 b_2 = 58182{,}853\,.$$

Diese Gleichung subtrahieren wir von (104.6c). Wir erhalten

$$(104.7\text{b})\qquad 11112{,}187 b_1 + 406179{,}352 b_2 = 89427{,}857\,.$$

2. Schritt. Elimination von b_1 aus (104.7b). Um b_1 aus (104.7b) zu eliminieren, multiplizieren wir (104.7a) mit 11112,187/34906,500. Dies ergibt

$$11112{,}187 b_1 + 3537{,}470 b_2 = 1636{,}579\,.$$

Diese Gleichung subtrahieren wir von (104.7b). Wir erhalten

$$(104.8)\qquad 402641{,}882 b_2 = 87791{,}278\,.$$

Schlußschritt. Lösung des Gleichungssystems

$$\begin{aligned} &(104.6\text{a})\quad 6144{,}000\, b_0 + 864{,}000\, b_1 + 35028{,}000\, b_2 = 10205{,}420\\ &(104.7\text{a})\quad 34906{,}500\, b_1 + 11112{,}187\, b_2 = 5140{,}953\\ &(104.8)\quad 402641{,}882\, b_2 = 87791{,}278\,. \end{aligned}$$

Jede Lösung dieses Gleichungssystems ist eine Lösung des Systems (104.6) und umgekehrt, wie man sich auf Grund der Herleitung überlegt. Wir zeigen nun, daß unser neues System eine eindeutige Lösung hat. Aus (104.8) folgt

$$b_2 = \frac{87791{,}278}{402641{,}882} = 0{,}218038\,.$$

Setzen wir dies in (104.7a) ein, so erhalten wir

$$b_1 = \frac{1}{34906{,}5}\,(5140{,}953 - 11112{,}187\, b_2) = 0{,}077867\,.$$

Setzen wir dies und b_2 in (104.6a) ein, so ergibt sich

$$b_0 = \frac{1}{6144}\,(10205{,}42 - 864\, b_1 - 35028\, b_2) = 0{,}407016\,.$$

Die gesuchte Parabel hat also die Darstellung

$$y = 0{,}407 + 0{,}078x^* + 0{,}218x^{*2}\,.$$

Indem wir $x^* = x - 53{,}5$ einsetzen und vereinfachen, folgt

(104.9) $$y = 620{,}205 - 23{,}248x + 0{,}218x^2.$$

Abb. 104.1 zeigt diese Parabel.

Aufgaben zu Abschnitt 104

104.1—104.7 Unter Benutzung der folgenden Stichproben bestimme man jeweils die Parabel (104.4) nach dem Prinzip der kleinsten Quadrate und stelle diese Parabel und die Stichprobenwerte graphisch dar.

104.1

x	−1	0	0	1
y	2	0	1	2

104.2

x	0	1	2	4	6
y	3	1	0	1	4

104.3 Zugkraft y [kg] eines Graders mit Zehnganggetriebe im 1. Gang in Abhängigkeit von der Geschwindigkeit x [km/Std.] (E. F. Wahl, Straßen- u. Tiefbau **17**, 1963, 879)

x	1,4	1,8	2,3	3,0	4,0
y	7400	7500	7600	7500	7200

104.4 Zur Feuchtigkeitsmessung von Gießereiformsand mit einer Szintillationssonde (CZ-Szintillator, Po-Be-Quelle): Anzahl y gezählter Impulse/min in Abhängigkeit von der Sanddichte x [g/cm³] bei konstanter Feuchtigkeit (8% Wasser) (R. Schroller u. H. Stiede, Kernenergie **6**, 1963, 224)

x	1,09	1,28	1,36	1,44	1,60	1,65
y	1,35	1,58	1,68	1,85	2,23	2,38

104.5 Durchschnittskurve der Aktien einiger großer amerikanischer Konzerne im Jahre 1962 [Dow-Jones Average, Industrials, etwa vereinfachte Werte]. Monat 1 = Januar usw.

Monat	1	2	3	4	5	6	7	8	9	10	11	12
Kurs	730	720	720	720	680	630	580	590	600	600	580	630

104.6 Geburtsgewicht x [g] und Sterblichkeit y [%] von Neugeborenen (H. Hosemann, Die Naturwiss. **37**, 1950, 411)

Geburtsgewicht		Anzahl Neugeborener in der Klasse	Davon gestorben	Sterblichkeit y [%]
Klassen-intervall	Klassen-mittelpunkt x_i [g]			
2000—2500	2250	92	11	11,96
2500—3000	2750	769	13	1,69
3000—3500	3250	2589	31	1,20
3500—4000	3750	2074	25	1,21
4000—4500	4250	511	10	1,96
4500—5000	4750	90	4	4,44

104.7 Reaktionszeit y [sec] von Schichtarbeiterinnen bei taktgebundener einförmiger Arbeit in Abhängigkeit von der Zeitdauer x [Std.] der Arbeit (M. Haider, Sichere Arbeit 1963, 14. Die Schicht dauerte 7 Stunden. Beim Aufleuchten einer kleinen Signallampe hatte die Versuchsperson eine Fußtaste zu drücken. y ist die Zeit zwischen dem Aufleuchten und der Betätigung der Taste).

x	1	2	3	4	5	6
y	1,50	1,48	1,75	1,65	1,72	1,55

104.8 Untergliedert man Fließbandarbeit zu stark, so daß die Verweilzeit des Werkstücks je Arbeitsgang zu kurz wird, so besteht die Gefahr des Anwachsens der unproduktiven Arbeit, wie die Daten aus Chicagoer Radio- und Fernsehfabriken zeigen (M. D. Kilbridge, Journ. Industr. Eng. **12**, 1961, 156). Man bestimme die Parabel (104.4). Bei welchem x-Wert liegt deren Scheitel?

Verweilzeit x [min]	0,5	1	2	3	4	7
Nicht produktive Zeit y [%]	5,6	4,1	3,9	3,9	4,2	4,5

104.9 Man bestimme die Form der Normalgleichungen im Falle eines Polynoms

$$y = b_0 + b_1 x + b_2 x^2 + b_3 x^3.$$

104.10 Für die Stichprobe in Aufgabe 104.2 bestimme man das Polynom 3. Grades (vgl. Aufgabe 104.9) nach dem Prinzip der kleinsten Quadrate.

105 Test bei nichtlinearer Regression

Den in Abschn. 103 beschriebenen Test der Linearität können wir sinngemäß auch bei nichtlinearer Regression anwenden. Dabei ändert sich lediglich die Anzahl der Freiheitsgrade: In Tab. 103.3 hatte die F-Verteilung $(r - 2, n - r)$ Freiheitsgrade. Nun hat sie $(r - m - 1, n - r)$, wobei m der Grad des verwendeten Polynoms ist [vgl. (104.2)]. Es genügt, den Test an einem Beispiel zu erläutern.

Beispiel 105.1 (Vergrößerungszahl eines terrestrischen Fernrohrs). Blickt man mit dem einen Auge durch ein einäugiges Fernrohr nach einer vertikalen Meßlatte und mit dem anderen Auge direkt auf die Latte, so kann man vergleichen, welche Strecke mit bloßem Auge so groß erscheint wie eine Strecke der Länge 1 cm im Fernrohr. Das Längenverhältnis beider Strecken bezeichnen wir mit y und nennen es die *Vergrößerungszahl*. Tab. 105.1 zeigt, daß y von der Entfernung x (in Metern vom Auge des Beschauers ab gemessen) abhängt. Für 5 verschiedene Entfernungen wurden je 2 Messungen gemacht (Werte aus dem Physikalischen Institut der Technischen Hochschule Graz, Direktor Prof. Dr. R. Gebauer). Die beiden Messungen differieren jeweils etwas voneinander,

Tabelle 105.1. Zu Beispiel 105.1

x_i [m]	y_{ij}	
3	29,7	30,1
5	28,1	27,8
7	27,4	26,9
10	26,0	26,2
14	25,5	25,2

obwohl y in streng kausaler Weise von x abhängen muß. Sie sind, wie wir sagen, mit einem *Meßfehler* behaftet. Dieser wird durch kleine unkontrollierbare und deshalb als zufällig anzusehende Störeinflüsse auf die Messung hervorgerufen. Auf den Begriff des Meßfehlers gehen wir in Kap. 19 noch ausführlich ein.

Zeichnen wir unsere Meßwerte als Punkte in der xy-Ebene auf, so sehen wir, daß die Regression der y-Werte bezüglich der x-Werte offenbar nicht linear ist. Diese Vermutung bestätigt man, indem man die Regressionsgerade bestimmt und auf diese den in Abschn. 103 beschriebenen Test anwendet.

Bestimmen wir die Regressionsparabel (104.4), so ergibt sich (s. Abb. 105.1)

$$y = 32{,}450 - 1{,}021x + 0{,}037x^2. \tag{105.1}$$

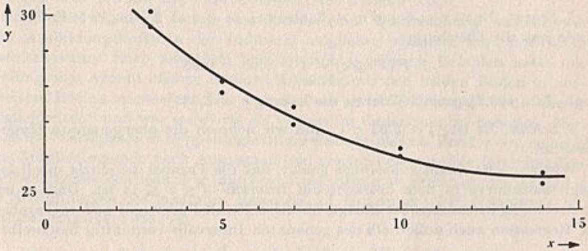


Abbildung 105.1. Meßwerte und Regressionsparabel in Beispiel 105.1

Tabelle 105.2. Rechnung zu Beispiel 105.1

x_i	$0{,}037x_i^2$	$1{,}021x_i$	$\eta_i^* = y^*(x_i)$	$y_{ij}^* = y_{ij} - 25$		$\bar{y}_i^*$	$(\bar{y}_i^* - \eta_i^*)^2$
3	0,333	3,063	4,72	4,7	5,1	4,90	0,032
5	0,925	5,105	3,27	3,1	2,8	2,95	0,102
7	1,813	7,147	2,12	2,4	1,9	2,15	0,001
10	3,700	10,210	0,94	1,0	1,2	1,10	0,026
14	7,252	14,294	0,41	0,5	0,2	0,35	0,004

Summe 0,165

Wir wollen die Hypothese testen, daß diese Parabel als Regressionskurve gewählt werden darf. Dabei setzen wir voraus, daß die y entsprechende Zufallsvariable Y für jedes feste $X = x$ normalverteilt ist, wobei die Varianz nicht von x abhängt.

1. Schritt. Für die bequemeren Werte

$$y_{ij}^* = y_{ij} - 25$$

hat die Parabel (105.1) die Darstellung

$$y^* = 7{,}450 - 1{,}021x + 0{,}037x^2. \tag{105.2}$$

Tab. 105.2 zeigt die notwendige Rechnung. Wegen $n_i = 2$ erhalten wir

$$q_1 = 2 \cdot 0{,}165 = 0{,}330.$$

Aus (103.6) folgt weiterhin

$$q_2 = \sum_{i=1}^{5} \sum_{j=1}^{2} (y_{ij} - \bar{y}_i)^2 = \sum_{i=1}^{5} \sum_{j=1}^{2} (y_{ij}^* - \bar{y}_i^*)^2 = 0{,}315.$$

Wir legen nun die Tab. 105.3 an. Aus dieser und (103.7) folgt

$$v_0 = \frac{0{,}165}{0{,}063} = 2{,}62.$$

2. Schritt. Wir wählen die Signifikanzzahl $\alpha = 5\%$.

3. Schritt. Es ist $r = 5$ und $m = 2$, also $r - m - 1 = 2$. Für (2, 5) Freiheitsgrade hat die Gleichung

$$P(V \leqq c) = 1 - 0{,}05 = 0{,}95$$

[vgl. Tab. 103.3] gemäß Tafel 9a die Lösung $c = 5{,}79$.

4. Schritt. Es ist $v_0 = 2{,}62 < c$, und wir nehmen die obengenannte Hypothese an.

Wiederum ist damit keineswegs gesagt, daß die Parabel die einzig mögliche Regressionskurve in dem betrachteten Intervall $3 \leqq x \leqq 14$ ist. Und es ist ebenso wenig etwas darüber gesagt, ob diese Parabel oder eine andere Parabel die Regression auch außerhalb des genannten Intervalls vernünftig beschreibt.

Tabelle 105.3. Schema der Varianzanalyse in Beispiel 105.1

Variation	Freiheitsgrade	Quadratsumme	Durchschnittsquadrat
Mittelwerte um die Regression	2	0,330	0,165
Innerhalb der Gruppen	5	0,315	0,063
Insgesamt	7	0,645	

Aufgaben zu Abschnitt 105

105.1 Für die gegebene Stichprobe (Temperatur y [°C] eines Thermometers in Abhängigkeit von der Zeit x [sec]) bestimme man die Parabel

$$y = b_0 + b_1 x + b_2 x^2$$

nach der Methode der kleinsten Quadrate. Man teste die Hypothese, daß diese Parabel als Regressionskurve gewählt werden darf. Dabei setze man voraus, daß die y entsprechende Zufallsvariable Y für jedes feste x normalverteilt ist, wobei die Varianz nicht von x abhängt.

x	10	20	30	60	90	120	180	240	300
y	27,9	25,9	24,3	19,6	15,9	12,9	8,5	5,6	3,6
	28,0	26,1	24,4	19,8	16,0	13,1	8,7	5,7	3,8

105.2 Bei Wachstums- und anderen Problemen muß man oftmals zu einer gegebenen Stichprobe eine Exponentialfunktion

$$y = b_0 e^{bx}$$

nach dem Prinzip der kleinsten Quadrate bestimmen. Man zeige, daß sich diese Aufgabe durch Logarithmieren auf die Aufgabe, eine Gerade zu bestimmen, zurückführen läßt.

105.3 Für die Stichprobe in Aufgabe 105.1 bestimme man die in Aufgabe 105.2 genannte Exponentialfunktion.

105.4 In Beispiel 105.1 bestimme man die Regressionsgerade und überzeuge sich davon, daß der Test in Abschn. 103 zur Verwerfung führt.

105.5 Man gewinne die Parabel (105.1) aus Tabelle 105.1.

105.6 Die folgenden Daten entstammen einer Untersuchung darüber, wie man Ausbildungskurse in der Industrie möglichst wirksam, kurz und billig gestalten kann. Nach akustisch bzw. optisch gegebenen Befehlen hatte der Prüfling eine Anzahl kleiner Radiowiderstände mit den beiden Enden in vorbereitete Löcher zu stecken. Die Aufgabe war „wirklichkeitsnah", bot „Lernmöglichkeiten" und war schwierig als Ganzes im Gedächtnis zu behalten. Man trage die Wertepaare in doppeltlogarithmischem Papier als Punkte ein, zeichne eine Ausgleichsgerade nach Augenmaß und gewinne daraus eine formelmäßige Darstellung des Zeitbedarfs y [min] pro Ausführung als Funktion der Anzahl x der Ausführungen des gesamten Versuchs. (J. Goldman u. H. Eisenberg, J. Industr. Eng. 14, 1963, 74)

x	1	2	3	5	10	15
	6,2	4,7	4,1	3,3	2,8	2,5
y	6,7	5,9	4,2	3,4	3,0	2,6
	7,0	6,1	4,4	3,5	3,2	2,7

105.7 Man löse Aufgabe 105.3 graphisch nach Augenmaß unter Verwendung von logarithmischem Papier.

Kapitel **18**

Korrelation

Einige allgemeine Bemerkungen über das Wesen und die Aufgabe der Korrelationsrechnung und die Unterschiede gegenüber der Regressionsrechnung haben wir bereits ganz zu Anfang des vorigen Kapitels gemacht. Darauf sei verwiesen.

106 Korrelationskoeffizient der Stichprobe

Gegeben sei eine Stichprobe aus einer zweidimensionalen XY-Grundgesamtheit. Die Stichprobe bestehe aus n Wertepaaren

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n).$$

Dann ist

$$\bar{x} = \frac{1}{n}(x_1 + \cdots + x_n)$$

der Mittelwert der x-Werte in der Stichprobe und

(106.1)
$$s_1^2 = \frac{1}{n-1}\sum_{j=1}^{n}(x_j - \bar{x})^2$$

die Varianz dieser Werte. Diese Größen haben wir schon in Abschn. 94 benutzt. Entsprechend ist

$$\bar{y} = \frac{1}{n}(y_1 + \cdots + y_n)$$

der Mittelwert der y-Werte in unserer Stichprobe und

(106.2)
$$s_2^2 = \frac{1}{n-1}\sum_{j=1}^{n}(y_j - \bar{y})^2$$

die Varianz dieser Werte. Weiterhin ist

(106.3)
$$s_{xy} = \frac{1}{n-1}\sum_{j=1}^{n}(x_j - \bar{x})(y_j - \bar{y})$$

die Kovarianz der Stichprobe. Vgl. (94.7).

Der Quotient

(106.4)
$$\boxed{r = \frac{s_{xy}}{s_1 s_2}} \qquad (s_1 > 0,\ s_2 > 0)$$

heißt der **Korrelationskoeffizient der Stichprobe.**

Wir zeigen zuerst, daß r mindestens gleich -1 sein muß und höchstens gleich $+1$ sein kann,

(106.5)
$$\boxed{-1 \leqq r \leqq 1.}$$

Gemäß (106.4) ist

$$s_{xy} = r s_1 s_2.$$

Setzen wir dies und

$$b = s_{xy}/s_1^2 = r s_2/s_1$$

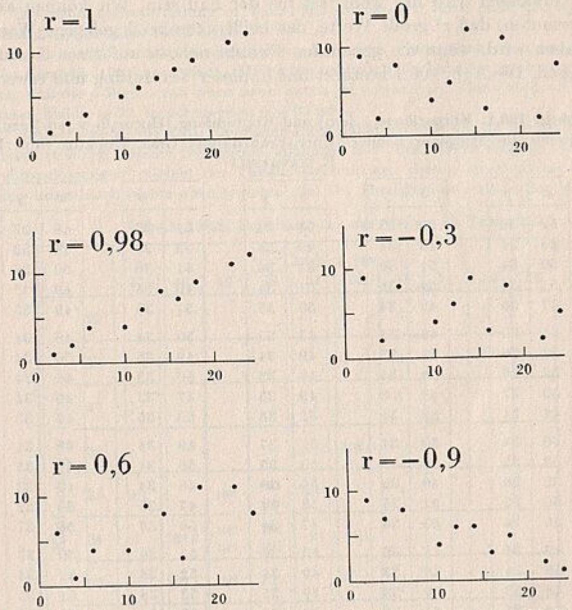


Abbildung 106.1. Graphische Darstellung von Stichproben mit verschiedenen Werten des Korrelationskoeffizienten r

[vgl. (94.5)] in (96.4) ein, so ergibt sich

$$a = (n-1)(1-r^2)\,s_2^{\,2}.$$

Da a eine Quadratsumme, also nicht negativ ist, muß auch die rechte Seite nicht negativ sein. Für $n > 1$ muß wegen $s_2^{\,2} > 0$ demnach

$$1 - r^2 \geqq 0 \qquad \text{oder} \qquad r^2 \leqq 1$$

sein, und (106.5) ist bewiesen.

Aus (106.4) ersieht man: Dann und nur dann wird $r^2 = 1$, wenn (96.5) gilt. So besagt nun die Aussage, in der (96.5) vorkommt:

Die Stichprobenwerte $(x_1, y_1), \ldots, (x_n, y_n)$ *liegen dann und nur dann genau auf einer Geraden, wenn der zugehörige Korrelationskoeffizient* r *den Wert 1 oder* -1 *hat.*

Praktisch wird das wohl fast nie der Fall sein. Wir können aber vermuten, daß r^2 große Werte, das heißt nahe bei 1 gelegene Werte, haben wird, wenn die genannten Punkte nahezu auf einer Geraden liegen. Die Abb. 106.1 bestärkt uns in dieser Vermutung und erweckt

Tabelle 106.1. Körperlänge x [cm] und Kopfumfang (Hutmaß) y [cm] neugeborener ausgetragener Kinder (Univ.-Frauenklinik Graz, Direktor Prof. Dr. E. NAVRATIL)

x	y	x	y	x	y	x	y	x	y
52	36	50	33	51	34	51	36	48	33
48	34	48	34	49	34	53	33	48	33
50	34	51	36	51	36	51	36	50	36
51	34	54	38	51	34	49	34	49	32
47	35	49	34	50	35	51	35	49	35
51	35	49	33	47	35	50	34	48	34
52	36	49	33	49	34	49	35	50	34
52	36	50	34	49	33	50	33	49	34
53	37	48	33	49	35	47	33	49	34
48	34	52	34	52	36	50	35	49	33
50	34	50	34	51	37	49	34	48	34
52	37	50	33	50	35	50	34	50	35
52	36	49	35	56	39	48	34	49	33
50	35	51	35	52	34	47	35	50	32
50	34	53	35	47	34	50	35	54	37
49	34	48	32	53	36	53	36	50	35
48	34	48	33	49	34	52	36	52	34
48	33	50	33	49	35	53	38	51	35
50	35	51	35	49	34	50	34	52	35
50	35	52	36	51	35	53	39	48	33

den Eindruck, als sei r ein Schätzwert eines Maßes für den Grad der *linearen* Abhängigkeit zwischen X und Y. Diesen Gedanken wollen wir im nächsten Abschnitt weiterverfolgen.

Im Augenblick betrachten wir nun ein Beispiel, das zugleich erläutern soll, wie man zweidimensionale Stichproben tabellarisch und graphisch darstellen kann.

Beispiel 106.1 (Körperlänge und Kopfumfang Neugeborener). Tab. 106.1 ist eine Urliste von Zahlenpaaren, die dem Geburtenbuch der Frauenklinik der Grazer Universität, Jahrgang 1962, entnommen wurde. Um einen Überblick über diese Stichprobe von $n = 100$ Wertepaaren zu erhalten, fassen wir gleiche Wertepaare zusammen. Wir durchlaufen hierzu die Urliste spaltenweise von oben nach unten und machen für jedes Wertepaar einen Strich in das zugehörige Feld der Tab. 106.2. Zum Schluß zählen wir die Striche in jedem Kästchen und erstellen die Tab. 106.3.

Die Häufigkeitsverteilung in Tab. 106.3 können wir auch noch graphisch darstellen, wie Abb. 106.2 zeigt.

In Tab. 106.2 und 106.3 wächst der y-Wert von unten nach oben zu, wie dies der Abb. 106.2 entspricht. Man kann natürlich die Tabellen auch so anlegen, daß die y-Werte von oben nach unten zu anwachsen. Dann steht in der ersten Zeile $y = 32$, in der zweiten $y = 33$ usw.

Betrachten wir die Tab. 106.2 oder 106.3 näher, so fällt uns an der Häufigkeitsverteilung unserer Stichprobe folgendes auf: Die Werte sind in einem „ellipsenförmigen" Gebiet der Tabelle konzentriert, dessen große Achse von links unten nach rechts oben gerichtet ist. Die Häufigkeiten sind in der Mitte

Tabelle 106.2. Strichliste zu der Urliste in Tab. 106.1

		Länge x [cm]									
		47	48	49	50	51	52	53	54	55	56
Kopfumfang y [cm]	39							𝍷			𝍷
	38							𝍷	𝍷		
	37					𝍷	𝍷	𝍷	𝍷		
	36				𝍷	𝍷𝍷𝍷𝍷	𝍸𝍷𝍷	𝍷𝍷			
	35	𝍷𝍷𝍷		𝍸	𝍸 𝍷𝍷𝍷𝍷	𝍸𝍷	𝍷	𝍷			
	34	𝍷	𝍸𝍷𝍷	𝍸 𝍸	𝍸 𝍷𝍷𝍷𝍷	𝍷𝍷𝍷	𝍷𝍷𝍷				
	33	𝍷	𝍸𝍷	𝍸	𝍷𝍷𝍷𝍷			𝍷			
	32		𝍷	𝍷	𝍷						

Tabelle 106.3. Häufigkeitsverteilung der Stichprobe in Tab. 106.1

		Länge x [cm]										Zeilen-summe
		47	48	49	50	51	52	53	54	55	56	
Kopfumfang y [cm]	39							1			1	2
	38							1	1			2
	37					1	1	1	1			4
	36				1	4	7	2				14
	35	3		5	9	6	1	1				25
	34	1	7	10	9	3	3					33
	33	1	6	5	4			1				17
	32		1	1	1							3
Spalten-summe		5	14	21	24	14	12	7	2	0	1	

des genannten Gebietes am größten und nehmen nach dem Rande zu ab. Sehr kurze Kinder mit sehr großem Schädel und sehr lange Kinder mit sehr kleinem Schädel kommen überhaupt nicht vor.

Den Korrelationskoeffizienten r unserer Stichprobe berechnen wir am besten nach der Formel

$$\text{(106.6)} \qquad \boxed{r = \frac{\sum x_j y_j - \frac{1}{n}\sum x_j \sum y_j}{\sqrt{\left[\sum x_j^2 - \frac{1}{n}(\sum x_j)^2\right]\left[\sum y_j^2 - \frac{1}{n}(\sum y_j)^2\right]}}}.$$

Diese ergibt sich unmittelbar aus (106.4), indem man (94.8) im Zähler und (94.6) und die entsprechende Formel für s_2 im Nenner einsetzt. Mit den Zahlenwerten

$$\sum x_j = 5 \cdot 47 + 14 \cdot 48 + \cdots + 1 \cdot 56 = 5009$$

$$\sum x_j^2 = 5 \cdot 47^2 + 14 \cdot 48^2 + \cdots + 1 \cdot 56^2 = 251\,215$$

$$\sum y_j = 3 \cdot 32 + 17 \cdot 33 + \cdots + 2 \cdot 39 = 3460$$

$$\sum y_j^2 = 3 \cdot 32^2 + 17 \cdot 33^2 + \cdots + 2 \cdot 39^2 = 119\,908$$

$$\sum x_j y_j = 47(33 + 34 + 3 \cdot 35) + \cdots + 56 \cdot 39 = 173\,477$$

erhalten wir

$$r = \frac{173477 - 0{,}01 \cdot 5009 \cdot 3460}{\sqrt{[251215 - 0{,}01 \cdot 5009^2]\,[119908 - 0{,}01 \cdot 3460^2]}} = 0{,}674.$$

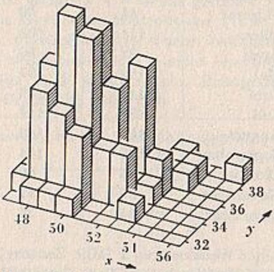


Abbildung 106.2. Histogramm zu Tab. 106.3

Aufgaben zu Abschnitt 106

106.1—106.6 Man stelle die folgenden Wertepaare durch Punkte graphisch dar und berechne den Korrelationskoeffizienten r.

106.1 Wellenlagerspiel x [mm] und Pleuellagerspiel y [mm] bei überholungsbedürftigen Schleppermotoren (E. Thum, Archiv f. Landtechnik 3, 1962, 108)

x	0,14	0,17	0,18	0,21	0,23	0,17	0,18
y	0,16	0,18	0,18	0,12	0,20	0,20	0,20

106.2 Nennmaß x [mm] und Gußschwindung y [%]. (Aus einer Untersuchung der Maßtoleranzen von Feingußteilen. H. Mann, Gießereitechnik 8, 1962, 308)

x	6	14	9	23	8	39	13	30	8	15
y	0,7	−0,2	−1,2	0,0	0,3	0,1	0,5	−0,4	0,8	−0,1

x	9	25	14	47	11	25	38	9	41	14
y	1,0	0,0	0,1	−0,4	−0,2	−0,5	−0,2	0,8	0,8	−0,1

x	15	11	35	18	10	23	30	13	15	12
y	0,5	0,2	0,1	−0,2	0,3	0,4	−0,6	0,5	1,1	−0,3

106.3 Monatliche Sonnenscheindauer vormittags (x) und nachmittags (y) (Wetterstation Univ. Graz, 1961)

Monat	Vormittag x [Std.]	Nachmittag y [Std.]
Januar	26	36
Februar	45	59
März	111	102
April	92	90
Mai	119	97
Juni	114	116
Juli	136	114
August	156	143
September	132	131
Oktober	55	59
November	30	41
Dezember	35	37

106.4 Österreichischer Weizenertrag x [Mill. Zentner] und Kartoffelertrag y [Mill. Zentner] (Statist. Handbuch f. d. Rep. Österreich, Wien 1956, S. 46)

Jahr	x	y
1952	4,0	25,7
1953	5,0	32,9
1954	4,5	27,9
1955	5,5	30,1

106.5 Klausurergebnis x [Punkte] während des Semesters und Ergebnis y [Punkte] der Semesterschlußklausur von 9 Mathematikstudenten (Ohio State University, Sommer 1962). Maximal erreichbare Punktezahl 100.

x	80	75	45	75	70	60	45	80	80
y	80	75	75	95	100	80	45	70	90

106.6 Einfuhr x [Milliarden DM] und Ausfuhr y [Milliarden DM] der Bundesrepublik Deutschland im Jahre 1960 [Statist. Jahrbuch f. d. Bundesrep. Deutschland 1961, Verl. Kohlhammer, Stuttgart, S. 316]

Land	x	y
Vereinigte Staaten	6,0	3,7
Frankreich	4,0	4,2
Niederlande	3,6	4,2
Italien	2,6	2,8
Belgien, Luxemburg	2,4	2,9
Großbritannien	2,0	2,1
Schweden	1,8	2,6
Schweiz	1,6	3,0
Dänemark	1,2	1,6
Österreich	1,2	2,4

106.7 In einem Hüttenwerk mit etwa 15000 Beschäftigten wurde eine neue Abteilung, die Stoffwirtschaft, eingerichtet. Der Psychologische Dienst des Werkes machte Eignungsuntersuchungen (Eignungsgrade 1 = gut geeignet, 2 = geeignet, 3 = noch geeignet, 4 = wenig geeignet, 5 = ungeeignet). Ein Jahr später wurde eine Bewährungskontrolle der 172 in der neuen Abteilung arbeitenden Personen durchgeführt (1 = sehr bewährt, 2 = bewährt, 3 = bedingt bewährt, 4 = wenig bewährt, 5 = nicht bewährt). Man berechne den Korrelationskoeffizienten und skizziere ein Histogramm der Stichprobe. (H. BECKER, Industr. Organ. **32**, 1963, 239)

Eignungsgrad	Bewährungsgrad y				
x	1	2	3	4	5
1	37	12			1
2	19	48	3	1	
3	1	28	15		1
4		1	3	2	

106.8 Bei einer Umfrage in gewissen Personengruppen wurde jeweils der Prozentsatz x [%] von Personen festgestellt, die im Jahre 1956 ein Auto auf Kredit kaufen wollten, und hinterher dann jeweils der Prozentsatz y [%] von Personen, die das tatsächlich getan hatten (M. L. LEE, Econometrica **30**, 1962, 185). Man stelle die Wertepaare graphisch dar und berechne den Korrelationskoeffizienten

x	5	15	25	35	45	55	65	75	85	95
y	0	19	21	41	43	48	71	81	89	97

107 Korrelationskoeffizient der Grundgesamtheit

Im vorigen Abschnitt haben wir eine Stichprobe von n Wertepaaren $(x_1, y_1), \ldots, (x_n, y_n)$ aus einer XY-Grundgesamtheit betrachtet. Den dabei vorkommenden Mittelwerten $\bar{x}$ und $\bar{y}$ entsprechen die Mittelwerte der beiden Variablen X und Y, die wir mit μ_1 bzw. μ_2 bezeichnen,

(107.1) $$\mu_1 = E(X), \qquad \mu_2 = E(Y).$$

Den Varianzen s_1^2 und s_2^2 im vorigen Abschnitt entsprechen die Varianzen

$$\sigma_1^2 = E([X - \mu_1]^2) \qquad \text{und} \qquad \sigma_2^2 = E([Y - \mu_2]^2)$$

der genannten Variablen. *Die Degenerationsfälle $\sigma_1 = 0$ oder $\sigma_2 = 0$, die kein Interesse besitzen, schließen wir stets aus.*

Die Größe

(107.2) $$\boxed{\sigma_{XY} = E([X - \mu_1]\,[Y - \mu_2])}$$

heißt die **Kovarianz** der Variablen X und Y. Wie in Abschn. 97 gezeigt wurde, ist

(107.3) $$\sigma_{XY} = E(XY) - E(X)\,E(Y).$$

Der Quotient

$$\boxed{\varrho = \frac{\sigma_{XY}}{\sigma_1 \sigma_2}} \tag{107.4}$$

heißt der **Korrelationskoeffizient** zwischen X und Y.

Ist $\varrho = 0$, so heißen X und Y **unkorreliert**.

Sind X und Y unabhängig, so ist σ_{XY} und damit auch ϱ wegen (58.6) gleich null. Es gilt also der

Satz 107.1. *Sind die Variablen X und Y unabhängig, so sind sie unkorreliert.*

Bei der zweidimensionalen Normalverteilung gilt auch die Umkehrung des Satzes (wie wir im nächsten Abschnitt sehen werden), im allgemeinen aber nicht. Dies zeigt das

Beispiel 107.1. Nimmt X jeden der Werte -1, 0, 1 mit der Wahrscheinlichkeit 1/3 an, so hat X den Mittelwert $E(X) = 0$. Setzen wir $Y = X^2$, so wird in (107.3)

$$E(XY) = E(X^3) = (-1)^3 \cdot \frac{1}{3} + 0^3 \cdot \frac{1}{3} + 1^3 \cdot \frac{1}{3} = 0.$$

Wegen $E(X) = 0$ ist also $\sigma_{XY} = 0$ und $\varrho = 0$. Die Variablen X und Y sind demnach unkorreliert. Sie sind aber sicher nicht unabhängig, weil ja zwischen ihnen sogar ein funktionaler Zusammenhang besteht.

Der Korrelationskoeffizient ist also *kein* Maß für Abhängigkeit schlechthin, wohl aber für *lineare* Abhängigkeit, wie sich ergeben wird.

Wir zeigen zuerst, daß stets

$$\boxed{-1 \leqq \varrho \leqq 1} \tag{107.5}$$

gilt.

Hierzu gehen wir zu den Variablen

$$X^* = \frac{X - \mu_1}{\sigma_1} \quad \text{und} \quad Y^* = \frac{Y - \mu_2}{\sigma_2}$$

über. Diese besitzen wegen Satz 34.2 den Mittelwert 0 und die Varianz

$$E(X^{*2}) = 1 \quad \text{bzw.} \quad E(Y^{*2}) = 1. \tag{107.6}$$

Folglich hat die Zufallsvariable

$$Z = tX^* + Y^* \qquad (t \text{ reell, konstant}) \tag{107.7}$$

den Mittelwert 0 und die Varianz

$$\begin{aligned} E(Z^2) &= E([tX^* + Y^*]^2) \\ &= t^2 E(X^{*2}) + 2tE(X^*Y^*) + E(Y^{*2}) \geqq 0. \end{aligned}$$

Hierbei ist, wie man durch Einsetzen sieht,

$$E(X^*Y^*) = \frac{1}{\sigma_1\sigma_2} E([X-\mu_1][Y-\mu_2]) = \frac{\sigma_{XY}}{\sigma_1\sigma_2} = \varrho.$$

So wird

$$E(Z^2) = t^2 + 2t\varrho + 1 \geqq 0.$$

Für jedes reelle t ist demnach

$$(t+\varrho)^2 + (1-\varrho^2) \geqq 0. \tag{107.8*}$$

Wäre $1-\varrho^2 < 0$, so wäre diese Ungleichung für $t = -\varrho$ nicht erfüllt. Ist $1-\varrho^2 \geqq 0$, so ist die Ungleichung wegen $(t+\varrho)^2 \geqq 0$ für alle t erfüllt. Es muß also $\varrho^2 \leqq 1$ sein. Dies ist gleichbedeutend mit (107.5).

Nun beweisen wir den grundlegenden

Satz 107.2. *Zwischen zwei Zufallsvariablen X und Y mit positiver Varianz besteht genau dann eine lineare Relation*

$$\text{(a)}\ Y = \beta X + \gamma \qquad [\textit{und}\quad \text{(b)}\ X = \beta^* Y + \gamma^*], \tag{107.9}$$

wenn der zugehörige Korrelationskoeffizient den Wert -1 oder 1 hat.

Beweis. **(a)** Es bestehe (107.9a). Wegen $\sigma_2 > 0$ muß $\beta \neq 0$ sein, und (107.9a) läßt sich in die Form (107.9b) überführen. Gemäß (34.2) und (34.4) hat Y den Mittelwert

$$\mu_2 = \beta\mu_1 + \gamma$$

und die Varianz

$$\sigma_2^2 = \beta^2\sigma_1^2.$$

Ist $\beta > 0$, so ist also $\sigma_2 = \beta\sigma_1$, und für $\beta < 0$ ist $\sigma_2 = -\beta\sigma_1 > 0$. Weiterhin ist demnach wegen (107.9a)

$$Y - \mu_2 = \beta X + \gamma - \mu_2 = \beta(X-\mu_1).$$

Damit gewinnt die Kovarianz die Form

$$\sigma_{XY} = E([Y-\mu_2][X-\mu_1]) = \beta E([X-\mu_1]^2) = \beta\sigma_1^2.$$

Für den Korrelationskoeffizienten erhalten wir also

$$\varrho = \frac{\sigma_{XY}}{\sigma_1\sigma_2} = \pm\frac{\beta\sigma_1^2}{\sigma_1\beta\sigma_1} = \pm 1.$$

(b) Umgekehrt sei $\varrho = 1$ bzw. $\varrho = -1$. Dann hat (107.7) mit $t = -1$ bzw. $t = 1$ die Varianz 0, wie man aus (107.8*) unmittelbar ersieht. Dies bedeutet, daß Z dann nur einen einzigen Wert (den Mittelwert 0) annimmt, und zwar mit der Wahrscheinlichkeit 1. Hinsichtlich der ursprünglichen Variablen X und Y, aus denen Z gebildet ist, bedeutet dies aber gerade das Bestehen einer Relation der Form (107.9). Dies war zu zeigen.

Der Korrelationskoeffizient erscheint damit als Maß für die lineare Abhängigkeit zwischen X und Y. Vor zwei Irrtümern müssen wir aber warnen:

1. Unsinnskorrelationen. Ein großer Wert von $|\varrho|$ beweist nicht, daß der dadurch ausgedrückte formale Zusammenhang auch ein ursächlicher Zusammenhang ist. Er kann es sein. (Beispiel: X = Anzahl der Verkehrsunfälle innerhalb von Ortschaften, Y = Anzahl der Verkehrsunfälle außerhalb von Ortschaften). Er braucht es aber nicht zu sein. (Beispiele: Von Norden nach Süden nimmt in Europa die durchschnittliche Körpergröße X ab, und die relative Anzahl Y der Katholiken nimmt zu. Nach dem 1. Weltkrieg nahmen in Deutschland die Anzahl X der Störche und die Geburtenziffer Y ab.) Dann spricht man von einer *Scheinkorrelation* oder *Unsinnskorrelation*. Dabei können beide Erscheinungen von einer gemeinsamen Ursache abhängen (bei den Störchen und Geburten von der Industrialisierung ?) oder auch nicht. Das alles ist nicht mathematisch sondern von der Sache her zu entscheiden. Umgekehrt kann man aber aus sachlichen Gründen einen Zusammenhang vermuten und dann zusehen, ob eine Stichprobe dafür oder dagegen spricht. Einen entsprechenden Test betrachten wir im übernächsten Abschnitt.

2. Fehlerhafte Vergleiche. Zu den Wertepaaren in Tab. 107.1 gehört der Korrelationskoeffizient 0,64. Vereinigt man jeweils die Absoluterträge von 2 Ländern, wie die geschweiften Klammern andeuten, und dividiert durch die Summe der beiden Anbauflächen, so erhält man

Tabelle 107.1. Relativer Roggenertrag x [dz/ha] und relativer Weizenertrag y [dz/ha] im Jahre 1960 (Statist. Jahrbuch Bundesrep. Deutschland 1961, S. 170)

Land	x	y
Baden-Württemberg	28	34
Bayern	27	34
Rheinland-Pfalz	29	37
Saarland	26	28
Hessen	32	38
Nordrhein-Westfalen	29	33
Niedersachsen	30	40
Schleswig-Holstein	27	38

statt der 8 Wertepaare in der Tabelle nur noch 4 Wertepaare. Zu diesen gehört der Korrelationskoeffizient 0,89. Vereinigt man nochmals paarweise, so bleiben 2 Paare übrig. Deren Korrelationskoeffizient ist gleich 1. (Warum?) Wie wir sehen, hängt der Wert des Korrelationskoeffizienten von der Anzahl der Wertepaare ab. Darauf muß man bei Vergleichen (etwa zwischen den Ernteerträgen mehrerer Jahre) achten. Sonst begeht man Irrtümer. Bei unserem Beispiel wäre weiter zu überlegen, ob man nicht günstigere Einheiten als die recht unterschiedlich großen Länder finden könnte.

Aufgaben zu Abschnitt 107

107.1 Die **Momente** einer zweidimensionalen Zufallsvariablen (X, Y) mit der Wahrscheinlichkeitsfunktion bzw. Dichte $f(x, y)$ sind durch

$$(107.10) \qquad m_{kl} = E(X^k Y^l) = \begin{cases} \sum\limits_i \sum\limits_j x_i^k y_j^l f(x_i, y_j) & \text{(Diskrete Verteilung)} \\ \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} x^k y^l f(x, y)\, dx\, dy & \text{(Stetige Verteilung)} \end{cases}$$

definiert und die **zentralen Momente** durch

$$(107.11) \qquad m_{kl}{}^* = E([X - \mu_1]^k [Y - \mu_2]^l).$$

Die (x_i, y_j) haben hierbei dieselbe Bedeutung wie in (53.1), und k und l sind nichtnegative ganze Zahlen. $k + l$ heißt die *Ordnung* des betreffenden Momentes. Man zeige: Die Momente

$$m_{10} = E(X) = \mu_1 \quad \text{und} \quad m_{01} = E(Y) = \mu_2$$

sind die beiden Mittelwerte. Die zentralen Momente

$$m_{20}{}^* = E([X - \mu_1]^2) = \sigma_1{}^2 \quad \text{und} \quad m_{02}{}^* = E([Y - \mu_2]^2) = \sigma_2{}^2$$

sind die beiden Varianzen. Das zentrale Moment

$$m_{11}{}^* = E([X - \mu_1][Y - \mu_2])$$

ist die Kovarianz der Zufallsvariablen (X, Y).

107.2 Man zeige: Gilt $f(x, y) = f(y, x)$ für alle (x, y), so ist $m_{kl} = m_{lk}$.

107.3—107.5 Man stelle die folgenden Dichten $f(x, y)$ als Fläche über der xy-Ebene graphisch dar und berechne die Momente m_{10}, m_{01}, $m_{20}{}^*$, $m_{02}{}^*$, $m_{11}{}^*$ und den Korrelationskoeffizienten ϱ.

107.3 $f(x, y) = x + y$ für $0 \leq x \leq 1$, $0 \leq y \leq 1$ und $f(x, y) = 0$ außerhalb dieses Quadrates.

107.4 $f(x, y) = (5 - x - y)/3$ im Rechteck $1 \leq x \leq 3$, $1 \leq y \leq 2$ und $f = 0$ außerhalb davon.

107.5 $f(x, y) = 4(1 - x)(1 - y)$ im Quadrat $0 \leq x \leq 1$, $0 \leq y \leq 1$ und $f = 0$ außerhalb davon.

107.6 Für die Verteilung in Aufgabe 107.3 berechne man die Wahrscheinlichkeit, daß (X,Y) irgendeinen Wert (x,y) annimmt, für den (a) $x > 1/2$, y beliebig, (b) $x > 1/3$, y beliebig, (c) $x + y < 1$ ist.

108 Zweidimensionale Normalverteilung

Ist X^* eine normalverteilte Zufallsvariable mit dem Mittelwert 0 und der Standardabweichung 1 und setzen wir

$$X^* = \frac{X - \mu}{\sigma}, \qquad (\sigma > 0)$$

so ist, wie wir wissen, die Variable X normalverteilt mit dem Mittelwert μ und der Standardabweichung σ. Man kann also im *eindimensionalen* Fall durch eine solche lineare Transformation von einem Spezialfall zum allgemeinen Fall übergehen. Entsprechend wollen wir nun bei der sogenannten *zweidimensionalen* Normalverteilung verfahren. Diese Verteilung spielt bei verschiedenartigen Anwendungen eine wichtige Rolle.

Den Ausgangspunkt bilden zwei unabhängige normalverteilte Variable X^* und Y^*, die beide den Mittelwert 0 und die Varianz 1 besitzen. Wegen der Unabhängigkeit hat die zweidimensionale Verteilung dieser Variablen die Dichte

$$f^*(x^*, y^*) = \frac{1}{2\pi} e^{-(x^{*2}+y^{*2})/2}. \tag{108.1}$$

Dies folgt unmittelbar aus (47.1) und (56.2).

Die Funktion (108.1) läßt sich graphisch als Fläche über der x^*y^*-Ebene darstellen (s. Abb. 108.1). Die Höhenlinien $f^* = konst.$ heißen

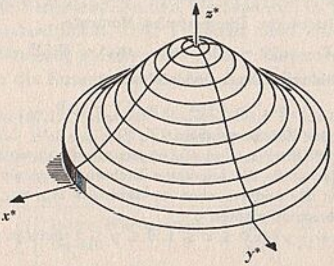


Abbildung 108.1. Dichte (108.1) als Fläche über der x^*y^*-Ebene (mittlerer Teil)

Kurven gleicher Wahrscheinlichkeit. Im vorliegenden Fall sind das Kreise. Unsere Fläche ist eine Drehfläche. Jede Schnittkurve der Fläche mit einer Ebene durch die Drehachse ist eine Glockenkurve.

Von X^* und Y^* gehen wir nun zu neuen Variablen X und Y über, indem wir

$$(108.2)\qquad \begin{aligned} X^* &= \frac{X-\mu_1}{\sigma_1} \\ Y^* &= \frac{1}{\sqrt{1-\varrho^2}}\left[\frac{Y-\mu_2}{\sigma_2} - \varrho\frac{X-\mu_1}{\sigma_1}\right] \end{aligned}$$

setzen. Hierbei sei $\varrho^2 < 1$, und σ_1 und σ_2 seien positiv. Daß wir diese lineare Transformation in einer so umständlichen Weise schreiben, hat einen guten Grund. Die Umkehrung lautet dann nämlich

$$\begin{aligned} X &= \mu_1 + \sigma_1 X^* \\ Y &= \mu_2 + \varrho\sigma_2 X^* + \sqrt{1-\varrho^2}\,\sigma_2 Y^*, \end{aligned}$$

und wir sehen: X und Y sind normalverteilt (vgl. Satz 71.1 und 71.2) mit dem Mittelwert μ_1 bzw. μ_2 und der Standardabweichung σ_1 bzw. σ_2, und der Korrelationskoeffizient dieser Variablen ist ϱ. Letzteres möge der Leser unter Benützung von (107.4) nachrechnen.

Wir wollen uns überlegen, wie die Dichte $f(x, y)$ der Variablen (X, Y) aussieht. Wir wissen, daß die Verteilungsfunktion der ursprünglichen Variablen (X^*, Y^*) ein Doppelintegral mit dem Integranden $f^*(x^*, y^*)$ ist. Beim Übergang zu den neuen Variablen ist nach der allgemeinen Regel für die Transformation von Doppelintegralen zu verfahren: Der Integrand des transformierten Integrals, also die gesuchte Dichte $f(x, y)$, ist gleich f^*, als Funktion von x und y ausgedrückt, multipliziert mit der zur Transformation (108.2) gehörigen Determinante

$$D = \begin{vmatrix} \dfrac{1}{\sigma_1} & 0 \\ -\dfrac{\varrho}{\sigma_1\sqrt{1-\varrho^2}} & \dfrac{1}{\sigma_2\sqrt{1-\varrho^2}} \end{vmatrix} = \frac{1}{\sigma_1\sigma_2\sqrt{1-\varrho^2}}.$$

f^* läßt sich vermöge (108.2) (geschrieben in x^*, y^*, x, y statt X^*, Y^*, X, Y) auf die neuen Variablen umrechnen. Das Ergebnis dieser elementaren Rechnung ist

$$f^*\bigl(x^*(x, y), y^*(x, y)\bigr) = \frac{1}{2\pi}e^{-h(x,y)/2}.$$

Hierbei wurde zur Abkürzung

$$\textbf{(108.3)}\quad h(x, y) = \frac{1}{1-\varrho^2}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\varrho\frac{x-\mu_1}{\sigma_1}\frac{y-\mu_2}{\sigma_2} + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]$$

gesetzt. Die gesuchte Dichte $f = f^* D$ hat also die Form

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\varrho^2}}\, e^{-h(x,y)/2} \tag{108.4}$$

mit h gemäß (108.3). Dies ist die allgemeine Form der Dichte der **zweidimensionalen Normalverteilung.**

Die zugehörigen Randverteilungen haben die Dichten

$$f_1(x) = \frac{1}{\sqrt{2\pi}\,\sigma_1}\, e^{-\frac{1}{2}\left(\frac{x-\mu_1}{\sigma_1}\right)^2} \tag{108.5}$$

und

$$f_2(y) = \frac{1}{\sqrt{2\pi}\,\sigma_2}\, e^{-\frac{1}{2}\left(\frac{y-\mu_2}{\sigma_2}\right)^2}, \tag{108.6}$$

denn X und Y sind normalverteilt mit dem Mittelwert μ_1 bzw. μ_2 und der Varianz σ_1^2 bzw. σ_2^2, wie wir schon wissen.

Die Kurven gleicher Wahrscheinlichkeit $f(x, y) = \mathit{konst.}$ oder $h(x, y) = \mathit{konst.}$ sind homothetische Ellipsen mit dem gemeinsamen Mittelpunkt (μ_1, μ_2).

Ist $\sigma_1 = \sigma_2$, so ist das Quadrat der Exzentrizität ε dieser Ellipsen durch

$$\varepsilon^2 = 2\,|\varrho|/(1 + |\varrho|)$$

gegeben. Die Ellipsen sind also desto kreisähnlicher (voller), je kleiner $|\varrho|$ ist. Für $\varrho = 0$ ergeben sich konzentrische Kreise.

Für $\varrho = 0$ ist weiterhin

$$f(x, y) = f_1(x)\, f_2(y),$$

wie man aus (108.3) erkennt. Während aus $\varrho = 0$ im allgemeinen keineswegs die Unabhängigkeit der betreffenden Variablen folgt, gilt also hier der bemerkenswerte

Satz 108.1. *Zwei normalverteilte Zufallsvariable sind dann und nur dann unabhängig, wenn der zugehörige Korrelationskoeffizient gleich null ist.*

Aufgaben zu Abschnitt 108

108.1 Man zeige, daß die im Text vorkommenden Variablen X und Y den Korrelationskoeffizienten ϱ haben.

108.2 Man verifiziere (108.3).

108.3 Man bestätige die Formel für die Exzentrizität der Ellipsen gleicher Wahrscheinlichkeit. Welchen Winkel bilden die Hauptachsen im Falle $\sigma_1 = \sigma_2$ mit der x-Achse ?

108.4 Man zeichne einige der Ellipsen gleicher Wahrscheinlichkeit für die zweidimensionale Normalverteilung mit $\mu_1 = \mu_2 = 0$, $\sigma_1 = \sigma_2 = 1$ und (a) $\varrho = 0{,}1$, (b) $\varrho = 0{,}5$, (c) $\varrho = 0{,}9$.

108.5 Wie sehen die Kurven gleicher Wahrscheinlichkeit für die zweidimensionale Normalverteilung im Falle $\varrho = 0$ bei beliebigem σ_1 und σ_2 aus?

108.6 Man gewinne eine Formel für die Exzentrizität der Ellipsen in Aufgabe 108.5.

108.7 Gegeben sei eine stetige zweidimensionale (X, Y)-Verteilung. Man betrachte das Ereignis $y < Y < y + \Delta y$ unter der Hypothese $x < X < x + \Delta x$ und zeige, daß für kleines Δx und Δy

$$P(y < Y < y + \Delta y \mid x < X < x + \Delta x) \approx \frac{f(x, y)}{f_1(x)} \Delta y$$

gilt. $f(x, y)$ ist die Dichte der (X, Y)-Verteilung und

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$$

die Dichte der einen Randverteilung. Man setzt

$$f(y \mid x) = \frac{f(x, y)}{f_1(x)} \qquad [f_1(x) \neq 0].$$

Diese Funktion heißt die **bedingte Dichte** *von Y unter der Hypothese $X = x$*. Die durch $f(y \mid x)$ gegebene Verteilung heißt die **bedingte Verteilung** von Y unter der genannten Hypothese.

108.8 Man zeige, daß im Falle der zweidimensionalen Normalverteilung

$$f(y \mid x) = \frac{1}{\sqrt{2\pi}\, \sigma_2 \sqrt{1 - \varrho^2}}\, e^{-\frac{h(x,y)}{2} + \frac{(x-\mu_1)^2}{2\sigma_1^2}}$$

gilt und daß dies die Dichte einer Normalverteilung mit dem Mittelwert

$$E(Y \mid X = x) = \int_{-\infty}^{\infty} y f(y \mid x)\, dy = \mu_2 + \frac{\varrho \sigma_2}{\sigma_1}(x - \mu_1)$$

und der Varianz $(1 - \varrho^2)\sigma_2^2$ ist. $E(Y \mid X = x)$ heißt der **bedingte Mittelwert** von Y unter der Hypothese $X = x$. Die Funktion $y(x) = E(Y \mid X = x)$ heißt die **Regressionsfunktion** von Y bezüglich X und die zugehörige Kurve die **Regressionskurve** des Mittelwertes von Y bezüglich X.

108.9 Man zeige: In Aufgabe 108.8 hat der Schwerpunkt der zur (X, Y)-Verteilung analogen Massenverteilung in einem schmalen vertikalen Streifen $x < X < x + \Delta x$ angenähert die Ordinate $y(x) = E(Y \mid X = x)$.

108.10 Man zeige: Die Regressionskurve in Aufgabe 108.8 ist unter allen möglichen Kurven $y = g(x)$ diejenige, für die der mittlere quadratische vertikale Abstand $E([Y - g(X)]^2)$ ein Minimum hat. Diese Kurve ist also eine Regressionskurve im Sinne des Prinzips der kleinsten Quadrate.

109 Tests und Konfidenzintervalle beim Korrelationskoeffizienten

Bei der zweidimensionalen Normalverteilung ist der Korrelationskoeffizient r einer Stichprobe ein Schätzwert für den Korrelationskoeffizienten ϱ der zugehörigen Grundgesamtheit [vgl. (106.4) und (107.4)]. In diesem Falle kann man die Hypothese $\varrho = 0$ gegen eine Alternative, etwa $\varrho > 0$, testen, wie Tab. 109.1 zeigt. Die Anweisung folgt daraus, daß t_0 einer Zufallsvariablen entspricht, die bei Richtigkeit der Hypothese eine t-Verteilung mit $n - 2$ Freiheitsgraden besitzt, wie R. A. Fisher (Biometrika **10**, 1915, 507) bewiesen hat. In diesem Falle sollte t_0 klein sein. Deshalb verwirft man bei zu großem t_0 die Hypothese.

Tabelle 109.1. Test der Hypothese $\varrho = 0$ gegen die Alternative $\varrho > 0$ bei der zweidimensionalen Normalverteilung

1. Schritt. Man wähle eine Signifikanzzahl α (5%, 1% oder dgl.).

2. Schritt. Man bestimme die Lösung c der Gleichung

$$P(T \leqq c) = 1 - \alpha \tag{109.1}$$

aus der Tafel 8 der t-Verteilung mit $n - 2$ Freiheitsgraden (siehe Anhang 5).

3. Schritt. Man berechne den Korrelationskoeffizienten r der gegebenen Stichprobe $(x_1, y_1), \ldots, (x_n, y_n)$.

4. Schritt. Man berechne

$$t_0 = r\sqrt{\frac{n-2}{1-r^2}}\,. \tag{109.2}$$

Ist $t_0 \leqq c$, so wird die Hypothese $\varrho = 0$ angenommen. Ist $t_0 > c$, so wird sie verworfen.

Beispiel 109.1. Unter der Annahme, daß die Grundgesamtheit, aus der die Stichprobe in Tab. 106.1 stammt, eine zweidimensionale Normalverteilung besitzt, teste man die Hypothese $\varrho = 0$ gegen die Alternative $\varrho > 0$.

1. Schritt. Wir wählen die Signifikanzzahl $\alpha = 5\%$.

2. Schritt. Es ist $n = 100$, also $n - 2 = 98$. Tafel 8 hat keine entsprechende Spalte. Für 100 Freiheitsgrade entnehmen wir der Tafel als Lösung der Gleichung (109.1) mit $1 - \alpha = 0{,}95$ den Wert $c = 1{,}66$. Diesen Wert können wir beim Test verwenden, wie ein Blick auf die in Tafel 8 links und rechts benachbarten Werte zeigt.

3. Schritt. Gemäß Beispiel 106.1 ist $r = 0{,}674$.

4. Schritt. So wird

$$t_0 = 0{,}674 \sqrt{\frac{98}{1 - 0{,}674^2}} = 9{,}03 .$$

Es ist $t_0 > c$. Die Hypothese wird also verworfen.

Aus (109.2) und Tafel 8 ergibt sich die Abb. 109.1, mit deren Hilfe man den betrachteten Test auch auf graphische Weise durchführen kann. Liegt der Punkt mit den Koordinaten n und r oberhalb der Kurve, so verwirft man die Hypothese. Anderenfalls nimmt man sie an.

Bemerkenswert ist in Abb. 109.1, daß sich der Annahmebereich bei kleinem Stichprobenumfang n bis zu relativ großen Werten von r hinauf erstreckt. Wäre im vorigen Beispiel n sehr klein, etwa $n = 5$, und $r = 0{,}674$ gewesen, so hätten wir die Hypothese $\varrho = 0$ trotz des relativ großen Wertes von r annehmen müssen. Dies erkennt man sofort aus Abb. 109.1.

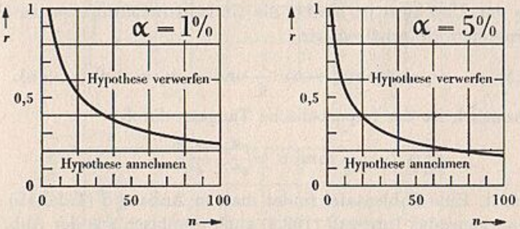


Abbildung 109.1. Test der Hypothese $\varrho = 0$ gegen die Alternative $\varrho > 0$

Wir zeigen nun, wie man bei der zweidimensionalen Normalverteilung **Konfidenzintervalle für den Korrelationskoeffizienten** ϱ bestimmt.

Aus dem Korrelationskoeffizienten r der Stichprobe berechnen wir den Hilfswert

$$z_0 = \frac{1}{2} \ln \frac{1 + r}{1 - r}. \tag{109.3}$$

Wie R. A. Fisher [Metron 1, 1921, 3] bewiesen hat, ist dies ein Schätzwert für den Mittelwert

$$\mu^* = \frac{1}{2} \ln \frac{1 + \varrho}{1 - \varrho}$$

einer asymptotisch normalverteilten Zufallsvariablen Z mit der Varianz

$$\sigma^{*2} = \frac{1}{n-3}.$$

Hierbei ist n der Stichprobenumfang. Praktisch bedeutet das: Wie in Abschn. 69 können wir ein Konfidenzintervall für μ^* bestimmen. Dabei übernimmt z_0 die Rolle von $\bar{x}$ in (69.2), und statt $\sigma/\sqrt{n}$ steht nun $\sigma^* = 1/\sqrt{n-3}$. So erhalten wir zunächst das Konfidenzintervall

$$\mathrm{KONF}\{z_0 - a \leqq \mu^* \leqq z_0 + a\} \quad \text{mit} \quad a = \frac{c}{\sqrt{n-3}}.$$

Hierbei ist c wie in Abschn. 69 aus Tab. 69.1 zu bestimmen. Dieses Intervall rechnen wir nun in ein Konfidenzintervall

$$\mathrm{KONF}\{r_1 \leqq \varrho \leqq r_2\} \tag{109.4}$$

für ϱ um. Und zwar ist hierbei die Umkehrtransformation zu (109.3) anzuwenden. So erhalten wir

$$r_1 = \tanh(z_0 - a) \qquad \text{und} \qquad r_2 = \tanh(z_0 + a). \tag{109.5}$$

Bekanntlich ist der hyperbolische Tangens durch

$$\tanh u = \frac{e^u - e^{-u}}{e^u + e^{-u}}$$

definiert. Eine Zahlentafel findet man in Anhang 5 (Tafel 4b).

Man kann das Intervall (109.4) auch graphisch aus der Abb. 109.2 bestimmen, wie wir nun an einem Beispiel erläutern wollen.

Beispiel 109.2. Unter der Annahme, daß die Stichprobe in Tab. 106.1 einer Grundgesamtheit entstammt, die eine zweidimensionale Normalverteilung besitzt, wollen wir ein 95%-Konfidenzintervall für den zugehörigen Korrelationskoeffizienten ϱ bestimmen. Es ist $r = 0{,}674$ (vgl. Beispiel 106.1). Mit Hilfe der ersten Kurve in Abb. 109.2 erhalten wir den zugehörigen Wert $z_0 = 0{,}8$. Aus der zweiten Kurve ersehen wir, daß zu $n = 100$ der Wert $a = 0{,}2$ gehört. Es ist also

$$z_0 - a = 0{,}6 \qquad \text{und} \qquad z_0 + a = 1{,}0.$$

Nun benutzen wir wieder die erste Kurve in Abb. 109.2 und finden

$$r_1 = \tanh 0{,}6 = 0{,}54, \quad r_2 = \tanh 1{,}0 = 0{,}76.$$

So ergibt sich das Konfidenzintervall

$$\mathrm{KONF}\{0{,}54 \leqq \varrho \leqq 0{,}76\}.$$

Graphische Verfahren haben den Vorteil der Schnelligkeit. Der Nachteil ihrer beschränkten Genauigkeit fällt hier praktisch kaum ins Gewicht. Auf rechnerischem Wege erhält man nämlich etwas genauer das Konfidenzintervall

$$\mathrm{KONF}\{0{,}550 \leqq \varrho \leqq 0{,}769\},$$

das sich nur sehr wenig von dem obigen unterscheidet.

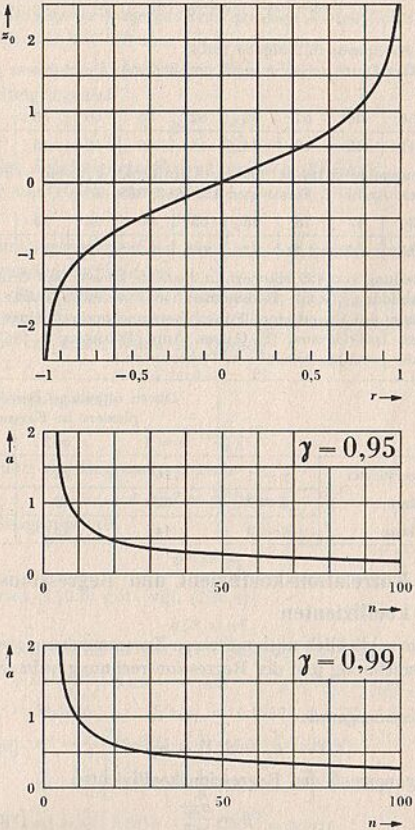


Abbildung 109.2. Zur graphischen Bestimmung des Konfidenzintervalles (109.4)

Aufgaben zu Abschnitt 109

109.1—109.5 Unter der Annahme, daß die jeweils gegebene Stichprobe einer zweidimensionalen Normalverteilung entstammt, bestimme man ein 99%-Konfidenzintervall für den Korrelationskoeffizienten ϱ.

109.1

x	1	3	4	6	8
y	2	4	1	2	5

109.2 Die Stichprobe in Aufgabe 106.1.

109.3 Größter Durchmesser x [mm] und kleinster Durchmesser y [mm] von Hühnereiern

x	58	56	57	52	62	55	60	55	61	58
y	41	43	42	44	44	40	42	45	45	44

109.4 Humusgehalt x [to/ha] und Stickstoffgehalt y [to/ha] von Waldböden (D. Heinsdorf, Archiv f. Forstwesen 12, 1963, 868—886)

x	30	43	75	55	55	60	66	73	80	80
y	0,7	1,5	1,5	1,4	1,3	1,7	2,9	3,1	2,5	2,1

109.5 Einteilung von 420 Männern in 3 soziale Klassen auf Grund gewisser Kriteria (Ausbildung, Beruf, Einkommen usw.) vermöge zweier Verfahren, nämlich objektiv auf Grund einer Punktbewertung und subjektiv unter Einschaltung eines Interviewers. (K. Gales, Appl. Statistics 6, 1957, 127. Die Tabelle gibt die jeweilige Anzahl von Personen an.)

		Durch objektive Bewertung plaziert in Klasse		
		$x = 1$	$x = 2$	$x = 3$
Durch Interviewer plaziert in Klasse	$y = 1$	116	20	8
	$y = 2$	33	48	60
	$y = 3$	14	13	108

110 Korrelationskoeffizient und Regressionskoeffizienten

Wir wollen schließlich noch auf einen Zusammenhang zwischen der Korrelationsrechnung und der Regressionsrechnung aufmerksam machen.

Die Regressionsgerade

(110.1) $$Y - \mu_2 = \beta(X - \mu_1)$$ [vgl. (97.5)]

hat als Steigungsmaß den Regressionskoeffizienten

(110.2) $$\beta = \frac{\sigma_{XY}}{\sigma_1^2}$$ [vgl. (97.4)].

Dabei wird X als unabhängige und Y als abhängige Variable angesehen, und man benutzt diese Regressionsgerade von Y bezüglich X, um zu gewählten Werten von X die entsprechenden Werte von Y abzuschätzen.

Vertauschen wir nun die Rollen von X und Y miteinander, so erhalten wir die *Regressionsgerade der Variablen X bezüglich der Variablen Y*. Diese Gerade hat offenbar die Darstellung

$$X - \mu_1 = \beta^*(Y - \mu_2) \tag{110.3}$$

mit dem Steigungsmaß

$$\beta^* = \frac{\sigma_{XY}}{\sigma_2{}^2}. \tag{110.4}$$

β^* heißt der *Regressionskoeffizient von X bezüglich Y.*

Wie wir aus (107.4) sehen, gilt

$$\beta\beta^* = \varrho^2. \tag{110.5}$$

Der Korrelationskoeffizient ist demnach das geometrische Mittel der beiden Regressionskoeffizienten.

Ist eine Stichprobe von n Wertepaaren $(x_1, y_1), \ldots, (x_n, y_n)$ gegeben, so können wir also nun *zwei* Regressionsgeraden berechnen, nämlich die uns schon bekannte Gerade

$$y - \bar{y} = b(x - \bar{x}) \tag{110.6}$$

[vgl. (94.3)]

mit dem Steigungsmaß

$$b = \frac{s_{xy}}{s_1{}^2} \tag{110.7}$$

und außerdem die Regressionsgerade

$$x - \bar{x} = b^*(y - \bar{y}) \tag{110.8}$$

mit dem Steigungsmaß

$$b^* = \frac{s_{xy}}{s_2{}^2}. \tag{110.9}$$

Entsprechend (110.5) gilt [vgl. (106.4)]

$$b\,b^* = r^2. \tag{110.10}$$

Beispiel 110.1. Für die in Tab. 110.1 gezeigten Werte wird

$$\bar{x} = \frac{18{,}55}{6} = 3{,}09 \qquad \text{und} \qquad \bar{y} = \frac{4{,}80}{6} = 0{,}80,$$

$$\sum x_k{}^2 = 59{,}4325 \qquad \sum x_k y_k = 15{,}1090 \qquad \sum y_k{}^2 = 3{,}8974$$

$$s_1{}^2 = \frac{1}{5}\left[59{,}4325 - \frac{1}{6}\cdot 18{,}55^2\right] = 0{,}416$$

$$s_{xy} = \frac{1}{5}\left[15{,}1090 - \frac{1}{6}\cdot 18{,}55\cdot 4{,}80\right] = 0{,}0538$$

$$s_2{}^2 = \frac{1}{5}\left[3{,}8974 - \frac{1}{6}\cdot 4{,}80^2\right] = 0{,}0115.$$

Tabelle 110.1. Jährliche Welternte von Kaffee und Tee [Statist. Jahrbuch f. d. Bundesrep. Deutschland 1961, p. 44*. Verlag Kohlhammer, Stuttgart]

Jahr	Kaffee x [Mill. to]	Tee y [Mill. to]
1954	2,40	0,66
1955	2,85	0,68
1956	2,50	0,82
1957	3,20	0,83
1958	3,50	0,90
1959	4,10	0,91
Summe	18,55	4,80

Hieraus folgt weiterhin

$$b = \frac{s_{xy}}{s_1^2} = 0{,}13 \qquad \text{und} \qquad b^* = \frac{s_{xy}}{s_2^2} = 4{,}7.$$

Die Regressionsgerade (110.6) hat also die Darstellung

$$y - 0{,}80 = 0{,}13\,(x - 3{,}09) \qquad \text{oder} \qquad y = 0{,}13x + 0{,}40,$$

und die Regressionsgerade (110.8) hat die Darstellung

$$x - 3{,}09 = 4{,}7\,(y - 0{,}80) \qquad \text{oder} \qquad x = 4{,}7y - 0{,}67.$$

Abb. 110.1 zeigt diese beiden Geraden.

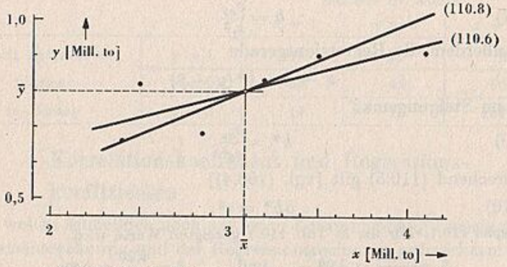


Abbildung 110.1. Regressionsgeraden in Beispiel 110.1

Wie unser Beispiel illustriert, sind die beiden Regressionsgeraden (110.6) und (110.8) im allgemeinen voneinander verschieden. Sie bilden, wie man anschaulich sagt, eine „*Regressionsschere*". Sie fallen dann und nur dann zusammen, wenn alle Stichprobenwerte genau auf einer Geraden liegen.

Übrigens kommt es praktisch fast nie vor, daß man bei einem Problem beide Regressionsgeraden *gleichzeitig* benötigt. Denn bei Regressionsbetrachtungen ist die übliche Sachlage, daß man zu gewählten Werten der einen Variablen, etwa x, die entsprechenden Werte der anderen abschätzen will. Dann bedient man sich der Geraden (110.6), und die Gerade (110.8) braucht man nicht.

Aufgaben zu Abschnitt 110

110.1 Man beweise die Aussage über das Zusammenfallen der beiden Regressionsgeraden am Schluß des Abschnittes.

110.2 Man zeige: Ist $r = 0$, so sind die beiden Regressionsgeraden achsparallel, stehen also aufeinander senkrecht.

110.3—110.6 Für die folgenden Stichproben bestimme man jeweils beide Regressionsgeraden und den Korrelationskoeffizienten r.

110.3

x	1	2	3	4	5
y	1	1	2	3	3

110.4

x	−2	0,5	1	3
y	2,4	1,65	1,5	0,9

110.5

x	0,3	1,3	1,3	1,3	2,3
y	0,4	1,4	0,4	−0,6	0,4

110.6 Die Stichprobe in Aufgabe 106.1.

Kapitel 19

Theorie der Meßfehler, Ausgleichsrechnung

Statistische Überlegungen spielen auch im Zusammenhang mit physikalischen Messungen eine Rolle. Jedes Meßergebnis ist nämlich mit einem Fehler behaftet. Dieser wird zum Teil durch zufällige kleine Störungen bedingt, die auch bei aller erdenklichen Sorgfalt nicht völlig vermieden werden können. Die Fehlertheorie, der das vorliegende Kapitel gewidmet ist, hat die Aufgabe, Verfahren zu entwickeln, mit deren Hilfe man aus gemessenen Werten möglichst „genaue" Näherungen der unbekannten physikalischen Größe (Länge, Masse, Stromstärke oder was man sonst gemessen hat) und auch ein Urteil über die Genauigkeit der Näherung erhalten kann. Die genannten Rechenverfahren werden unter dem Begriff der **Ausgleichsrechnung** zusammengefaßt.

111 Arten von Meßfehlern, Genauigkeitsmaße

Die Summe der Winkel eines ebenen Dreiecks beträgt bekanntlich 180°. *Mißt* man aber die Winkel, so beträgt die Summe der gemessenen Werte im allgemeinen nicht genau 180°. Vielmehr ergibt sich eine kleine Abweichung. Mißt man mehrmals, so stimmen die Ergebnisse meist nicht genau überein, sondern variieren von Messung zu Messung. Dieselbe Erfahrung macht man bei der Messung von Längen, Gewichten, Stromstärken, Drucken, Temperaturen und anderen physikalischen Größen. Man sagt, daß jede Messung mit einem **Meßfehler** oder *Beobachtungsfehler* behaftet sei. Dieser ist durch das Wesen des Meßprozesses bedingt: Der Beobachter, die Instrumente und das zu messende System sind verschiedenartigen Störeinflüssen unterworfen. Die Größe der Meßfehler läßt sich zwar verringern, indem man von „gröberen" zu „präziseren" Meßmethoden übergeht, aber niemals ganz ausschalten.

Bei jeder Messung kann man zwischen verschiedenen Arten von Fehlern unterscheiden, wie wir uns nun überlegen wollen.

Mißt man z. B. die Winkel eines Dreiecks mit einem Theodoliten, der 5'' Ablesegenauigkeit besitzt, und ergibt sich daraus als Winkelsumme 180° 01' statt 180° 00', so liegt ein **grober Fehler** vor. Bei jeder Meßmethode muß man sich durch entsprechende Proben unbedingt gegen derartige Fehler sichern.

Differiert eine 5 m lange Meßlatte infolge Abnutzung um 1/2 mm von ihrem Sollwert, so wirkt sich dies bei einer 1 km langen Strecke bereits um 10 cm aus. Eine solche Differenz ist bei der Basismessung in der Landestriangulation untragbar. Man muß dann die Lattenmessung trotz des geringen Einzelfehlers durch ein genaueres Meßverfahren ersetzen. Derartige kleine Abweichungen, die die Ergebnisse aller Messungen einer bestimmten Größe mit einer bestimmten Meßmethode in der gleichen Weise verfälschen, werden als *konstante*, *regelmäßige* oder **systematische Fehler** bezeichnet.

Diese Fehler interessieren uns im folgenden ebenso wenig wie die groben Fehler, weil beide Fehlerarten nichts mit Wahrscheinlichkeitstheorie zu tun haben. Übrigens sieht man systematische Fehler als im wesentlichen — wenn auch nie völlig — vermeidbar an. Man begegnet ihnen durch Sorgfalt bei der Wahl und Eichung der Meßinstrumente, beim Aufbau der Versuchsapparatur und durch andere Maßnahmen, die in das Gebiet der physikalischen Meßtechnik gehören.

Von den beiden genannten Fehlerarten sehen wir also ab. Übrig bleiben dann noch Fehler infolge einer Vielzahl kleiner Störeinflüsse, die die Meßergebnisse in ganz unkontrollierbarer und von Messung zu Messung wechselnder Weise verfälschen. Diese Fehler werden als *zufällige* oder **statistische Fehler** bezeichnet. Sie interessieren uns hier ausschließlich. Da sich ihre Ursachen unserer Kontrolle entziehen, sind diese Fehler unvermeidbar. Ihre theoretische Behandlung wurde von Gauss und Laplace begründet. Eine solche Theorie der statistischen Meßfehler hat den Zweck, Methoden zu entwickeln, mit deren Hilfe man einen brauchbaren Näherungswert der zu messenden Größe und ein Urteil über die Genauigkeit der Näherung gewinnt.

Ein solches Urteil läßt sich natürlich nicht aus einer einzelnen Messung erhalten. Aus diesem Grunde mißt man eine physikalische Größe mehrmals hintereinander. So erhält man eine Stichprobe von Meßwerten $x_1, \ldots, x_n$. In der großen Mehrzahl der Fälle — aber nicht bei *allen* physikalischen Messungen — macht man dann eine erstaunliche Entdeckung: Die Stichprobe erweckt den Anschein, als ob sie einer normalverteilten Grundgesamtheit entstammt. Dies ist deshalb überraschend, weil ja physikalische Messungen ihrer Art und Aufgabe nach ganz verschieden sein können.

Wie kann man erklären, daß der statistische Fehler in den allermeisten Fällen eine normalverteilte oder annähernd normalverteilte Zufallsvariable ist? Am einleuchtendsten ist es, sich vorzustellen, der genannte Fehler entstünde durch das Zusammenwirken sehr vieler sehr kleiner Einzelfehler („*Elementarfehler*"). Sehen wir die zugehörigen Zufallsvariablen als unabhängig an, so ist deren Summe, die dem statistischen Gesamtfehler entspricht, auf Grund des zentralen Grenzwertsatzes (s. Abschn. 76) annähernd normalverteilt.

So ähnlich hat schon Gauss argumentiert. In diesem Zusammenhang hat er übrigens die Normalverteilung oder Gauss-Verteilung eingeführt.

Die genannte Erklärung hat viel für sich. In der Tat unterliegt jede Messung einer großen Zahl unkontrollierbarer kleiner Einflüsse. Je nach Art des Experimentes stören geringfügige Temperatur-, Druck-, Feuchtigkeits- oder Spannungsschwankungen, Luftbewegungen, Staubteilchen, Lichteinwirkungen, mechanische Erschütterungen, elektrische oder magnetische Felder und andere Faktoren das zu messende Objekt und auch die Meßapparatur in der genannten Weise. Hinzu kommen Ablesefehler beim Schätzen von Bruchteilen von Skaleneinheiten. In den meisten Fällen ist es gar nicht möglich, alle diese Störursachen auch nur aufzuzählen, geschweige denn die Größe der zugehörigen, zum Teil winzigen Elementarfehler vorherzusagen, die sich von Messung zu Messung ändern.

Soviel zur Natur der Meßfehler. Nun wollen wir Maßzahlen („Genauigkeitsmaße") einführen, die uns eine Vorstellung davon geben, welche Genauigkeit ein gewisses Meßergebnis etwa hat.

Haben wir eine physikalische Größe n-mal gemessen, so liegen n Meßwerte

$$x_1, \ldots, x_n$$

vor. Dann können wir den Mittelwert

$$\bar{x} = \frac{1}{n}(x_1 + \cdots + x_n)$$

dieser Stichprobe bilden und als Näherungswert für die unbekannte Größe μ, die wir messen, ansehen. Entsprechend ist die Standardabweichung

$$s = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(x_j - \bar{x})^2} \tag{111.1}$$

als Schätzung der Standardabweichung σ der Grundgesamtheit aller denkbaren Messungen und demnach als Maß für die Präzision der Messungen aufzufassen. s wird in der Fehlertheorie als der **mittlere Fehler** oder der *mittlere quadratische Fehler* bezeichnet. Oftmals schreibt man das Ergebnis einer Meßreihe in der Form

$$\boxed{\bar{x} \pm s.}$$

Ist der Meßfehler annähernd normalverteilt, so können wir erwarten, daß bei einer hinreichend großen Stichprobe ungefähr 2/3 aller Meßwerte zwischen $\bar{x} - s$ und $\bar{x} + s$ liegen (vgl. Abschn. 48).

Entsprechend liegen bei einer hinreichend großen Stichprobe etwa 50% aller Meßwerte zwischen

$$\bar{x} - 0{,}67\, s \qquad \text{und} \qquad \bar{x} + 0{,}67\, s,$$

wie man sich an Hand der Tafel 3a in Anhang 5 leicht überlegt. $0{,}67\, s$ wird als *wahrscheinlicher Fehler* bezeichnet. Dieses Fehlermaß ist veraltet.

Übrigens verwendet man statt (111.1) oft auch die Maximum-Likelihood-Schätzung (vgl. Beispiel 68.3)

$$s^* = \sqrt{\frac{1}{n} \sum_{j=1}^{n} (x_j - \bar{x})^2} \tag{111.2}$$

und bezeichnet diese als *mittleren Fehler* (auch wenn es sich nicht um eine Normalverteilung handelt).

Weiterhin besteht natürlich die Möglichkeit, Konfidenzintervalle für μ und σ zu bestimmen, sobald man den Verteilungstyp des statistischen Fehlers kennt.

Schließlich sei noch ein weiteres, ebenfalls weniger wichtiges Genauigkeitsmaß erwähnt, nämlich der sogenannte *Durchschnittsfehler*

$$\delta = \frac{1}{n} \sum_{j=1}^{n} |x_j - \bar{x}|. \tag{111.3}$$

Tabelle 111.1. Flächeninhaltsbestimmung mit einem Planimeter

	x_j	$\lvert x_j - \bar{x}\rvert$	$(x_j - \bar{x})^2$
	402	0,2	0,04
	398	4,2	17,64
	405	2,8	7,84
	406	3,8	14,44
	400	2,2	4,84
Summe	2011	13,2	44,80

Beispiel 111.1. Der Inhalt eines Flächenstückes wurde mit einem Planimeter (Fabrikat A. Ott, Kempten) bestimmt, und zwar fünfmal nacheinander. Dabei ergaben sich die Werte x_j in Tab. 111.1. Aus diesen erhalten wir den Mittelwert

$$\bar{x} = \frac{2011}{5} = 402{,}2\ [\mathrm{mm}^2]$$

und den mittleren Fehler

$$s = \sqrt{\frac{44{,}8}{4}} = 3{,}3\ [\mathrm{mm}^2].$$

Dieses Ergebnis schreiben wir in der Form

$$x = 402{,}2 \pm 3{,}3\ [\mathrm{mm}^2].$$

Für die übrigen im Text genannten Größen ergeben sich im vorliegenden Falle die Werte

$$0{,}67s = 2{,}2\ [\mathrm{mm}^2], \quad s^* = \sqrt{\frac{44{,}8}{5}} = 3{,}0\ [\mathrm{mm}^2], \quad \delta = \frac{13{,}2}{5} = 2{,}6\ [\mathrm{mm}^2].$$

Wir wollen erwähnen, daß man in der Fehlerrechnung bisweilen eine Schreibweise verwendet, die von GAUSS herrührt: Man setzt $v_j = x_j - \bar{x}$ und schreibt statt des Summenzeichens eckige Klammern, wobei der Summationsindex auch noch weggelassen wird. Dann ergibt sich

$$\bar{x} = \frac{1}{n} \sum (x_j - \bar{x}) = \frac{1}{n} \sum v_j = \frac{[v]}{n}.$$

Entsprechend schreibt man (111.1) nun in der Form

$$s = \sqrt{\frac{[v^2]}{n-1}}$$

oder auch, indem man den Exponenten vermeidet und vv statt v^2 schreibt,

$$s = \sqrt{\frac{[vv]}{n-1}}.$$

Formel (111.3) gewinnt nun die Gestalt

$$\delta = \frac{[|v|]}{n}.$$

Wir machen von dieser Schreibweise keinen Gebrauch.

Aufgaben zu Abschnitt 111

111.1—111.3 Für die gegebenen Meßwerte bestimme man den Mittelwert, den mittleren Fehler und den Durchschnittsfehler.

111.1 Kalorimeterkonstante [BTU/pound/°Fahrenheit] (G. WERNIMONT, Industr. Quality Control **3**, 1947, 5—11)

2435,6   2433,6   2428,8   2428,6   2435,9   2441,7   2433,7   2437,8

111.2 Zugfestigkeit [kg/mm²] von Blechen (A. Slattenschek u. G. Schneeweiss, Maschinenbau u. Wärmewirtsch. 12, 1957, 277)

44,0 42,8 40,8 41,4 44,4 43,9 42,8 44,0 42,2 44,8

111.3 Schwingungsdauer [sec/50 Schwingungen] eines Pendels (Praktikumsversuch, Physik. Inst. Techn. Hochschule Graz, Direktor Prof. Dr. R. Gebauer)

80,1 80,1 79,8 80,0 80,0 80,1 80,1 80,0 79,8 80,2

111.4 Man berechne s^*/s [vgl. (111.1) und (111.2)] für $n = 2, 3, \ldots, 10$ und stelle die Werte graphisch dar.

111.5 Unter der Annahme, daß die Stichprobe in Aufgabe 111.2 einer normalverteilten Grundgesamtheit entstammt, bestimme man ein 95%-Konfidenzintervall für das Quadrat des mittleren Fehlers.

112 Gewogener Mittelwert

Oftmals besitzen die Werte einer Stichprobe $x_1, \ldots, x_n$ eine unterschiedliche Zuverlässigkeit. Dies wollen wir uns zuerst an einem sehr einfachen Beispiel klarmachen.

Die Schwingungsdauer eines Pendels sei zu ermitteln. Hierzu messen wir die Zeiten t_1, t_2, t_3, t_4 vier aufeinanderfolgender Durchgänge durch die Gleichgewichtslage in derselben Richtung. Daraus bilden wir die beiden unabhängigen Näherungswerte

$$x_1 = \frac{1}{3}(t_4 - t_1) \quad \text{und} \quad x_2 = t_3 - t_2$$

für die Schwingungsdauer. Auf Grund der Natur des Experimentes können wir annehmen, daß den beiden Zeitdifferenzen zwei Zufallsvariable mit derselben (unbekannten) Varianz σ^2 entsprechen. Wegen (34.4) sind dann x_1 und x_2 beobachtete Werte zweier unabhängiger Zufallsvariablen X_1 und X_2 mit der Varianz $\sigma_1^2 = \sigma^2/9$ und $\sigma_2^2 = \sigma^2$. Es handelt sich also um Messungen mit verschiedener Genauigkeit. x_1 sollte bei der Mittelbildung stärker ins Gewicht fallen als x_2. An die Stelle des üblichen Mittelwertes

$$\bar{x} = \frac{1}{2}(x_1 + x_2) = \frac{1}{2}x_1 + \frac{1}{2}x_2$$

sollten wir also einen Ausdruck der Form

$$\bar{x}^* = g_1 x_1 + g_2 x_2 \qquad (g_1 + g_2 = 1)$$

setzen und dabei die positiven Zahlen g_1 und g_2 so wählen, daß die Standardabweichung der entsprechenden Zufallsvariablen

$$\bar{X}^* = g_1 X_1 + g_2 X_2$$

möglichst klein ausfällt. Denn dies bedeutet, daß $\bar{x}^*$ einen möglichst kleinen mittleren Fehler besitzt. Gemäß (34.4) und Satz 59.1 hat $\bar{X}^*$ die Varianz

$$\sigma^{*2} = g_1{}^2\sigma_1{}^2 + g_2{}^2\sigma_2{}^2.$$

Wegen $g_2 = 1 - g_1$ wird dies

$$\sigma^{*2} = g_1{}^2\sigma_1{}^2 + (1 - g_1)^2\sigma_2{}^2.$$

Durch Nullsetzen der Ableitung nach g_1 erhalten wir

(112.1) $$g_1\sigma_1{}^2 = g_2\sigma_2{}^2.$$

Hieraus folgt, daß die Standardabweichung σ^* ein Minimum hat, wenn wir die Zahlen g_1, g_2 umgekehrt proportional zu den Varianzen der Variablen X_1, X_2 wählen. Wegen $\sigma_1{}^2 = \sigma^2/9$ und $\sigma_2{}^2 = \sigma^2$ erhalten wir aus (112.1) nun $g_1 = 9g_2$, also $g_1 = 0{,}9$ und $g_2 = 0{,}1$, denn $g_1 + g_2$ muß gleich 1 sein. Es ergibt sich also

$$\bar{x}^* = 0{,}9x_1 + 0{,}1x_2.$$

Diesen Ausdruck bezeichnen wir als einen *gewogenen Mittelwert* von x_1 und x_2, und die positiven Zahlen 0,9 und 0,1 heißen die *Gewichte*.

Unser Beispiel ist typisch für Probleme, wie sie praktisch häufig auftreten, etwa in der Geodäsie. Die Lage ist die folgende:

Gegeben ist eine Stichprobe $x_1, \ldots, x_n$. Die entsprechenden unabhängigen Zufallsvariablen $X_1, \ldots, X_n$ haben die Varianzen $\sigma_1{}^2, \ldots, \sigma_n{}^2$, die nicht alle gleich sind. Es handelt sich also um Messungen verschiedenen Genauigkeitsgrades. *Dann wird der mittlere Fehler des Ausdruckes*

(112.2) $$\boxed{\bar{x}^* = g_1x_1 + \cdots + g_nx_n} \qquad (g_1 + \cdots + g_n = 1)$$

ein Minimum, wenn wir die Zahlen $g_1, \ldots, g_n$ umgekehrt proportional zu den Varianzen $\sigma_1{}^2, \ldots, \sigma_n{}^2$ wählen. Diese Zahlen heißen **Gewichte,** *und* (112.2) *heißt der* **gewogene Mittelwert** *der n Stichprobenwerte.*

Sind die genannten Varianzen gleich, so wird $\bar{x}^*$ mit dem üblichen Mittelwert identisch.

Man beachte, daß man für die Bestimmung der Gewichte die genannten Varianzen nicht zu kennen braucht sondern nur die Verhältnisse zwischen je zwei dieser Varianzen. Schreiben wir

(112.3) $$\sigma_j{}^2 = \frac{\sigma^2}{c_j},$$

so kann σ^2 eine unbekannte Konstante sein, während $c_1, \ldots, c_n$ bekannt sein muß.

Übrigens ist es rechentechnisch günstiger, in (112.2) die Bedingung $g_1 + \cdots + g_n = 1$ entfallen zu lassen. Dann wird

$$\bar{x}^* = \frac{g_1 x_1 + \cdots + g_n x_n}{g_1 + \cdots + g_n}. \tag{112.4}$$

Aufgaben zu Abschnitt 112

112.1 Unter der Annahme, daß $X_1, \ldots, X_n$ unabhängige normalverteilte Variable mit dem Mittelwert μ und der Varianz (112.3) sind, gewinne man (112.4) mit Hilfe der Maximum-Likelihood-Methode.

112.2 Die Aussage, in der (112.2) vorkommt, wurde im Text für $n = 2$ bewiesen. Man beweise sie für beliebiges n.

112.3 Vereinigt man zwei Stichproben $x_1, \ldots, x_n$ und $\tilde{x}_1, \ldots, \tilde{x}_m$, die aus derselben Grundgesamtheit stammen und den Mittelwert $\bar{x}$ bzw. $\bar{\bar{x}}$ besitzen, zu einer einzigen Stichprobe, so hat letztere den Mittelwert

$$\bar{x}^* = g_1 \bar{x} + g_2 \bar{\bar{x}} \quad \text{mit} \quad g_1 = \frac{n}{n+m} \quad \text{und} \quad g_2 = \frac{m}{n+m}.$$

(a) Man bestätige diese Formel unter Benutzung der Definitionsformel für den gewöhnlichen Mittelwert. (b) Man zeige, daß die Wahl der Gewichte mit der im Text genannten Minimumsforderung im Einklang steht.

112.4 Man vereinige die beiden Stichproben

2,3 2,1 1,9 und 2,0 2,2 2,1 2,0 1,8

zu einer einzigen und berechne den Mittelwert der letzteren (a) aus den Mittelwerten der ersteren, (b) direkt.

112.5 Ein Winkel wurde mit zwei Theodoliten jeweils mehrmals gemessen. Daraus ergab sich $73°2'7'' \pm 10''$ bzw. $73°2'12'' \pm 20''$. Man bilde das gewogene Mittel beider Werte.

112.6 Die Höchstlast von Lötstellen (Fläche 20 mm²) zwischen Titan- und Aluminiumblechen wurde gemessen. Zwei Stichproben von 10 bzw. 4 Werten hatten den Mittelwert 448 kg bzw. 418 kg (G. E. Metzger, Entwicklung eines Hartlötverfahrens für die Paarung Titan—Aluminium. Diss. Aachen 1961). Wie groß ist der Mittelwert der vereinigten Stichprobe aller 14 Werte?

113 Vermittelnde Beobachtungen. Fehlerfortpflanzungsgesetz

Werden die physikalischen Größen, für die man sich in einem konkreten Falle interessiert, direkt gemessen, so spricht man von **direkten Beobachtungen.** Dagegen spricht man von *indirekten* oder **vermittelnden Beobachtungen**, wenn man die genannten Größen nicht direkt mißt, sondern aus anderen gemessenen Größen errechnet. Das kommt häufig vor: Um die Fläche F eines Rechtecks zu bestimmen, mißt man dessen Seitenlängen a und b und berechnet dann daraus den gesuchten Wert $F = ab$. Weitere Beispiele kann sich der Leser leicht bilden.

Dabei ist es wichtig zu wissen, in welcher Weise sich die Fehler der gemessenen Größen auf die berechnete Größe auswirken oder „fortpflanzen“. Dieser Frage wollen wir nachgehen.

Wir nehmen an, daß zwei unabhängige Größen X und Y gemessen werden und daraus die Größe

$$Z = h(X, Y)$$

berechnet wird. Und zwar werde X genau n mal und Y genau m mal gemessen. Die Meßwerte seien

$$x_1, \ldots, x_n \quad \text{bzw.} \quad y_1, \ldots, y_m.$$

Hieraus ergeben sich die Mittelwerte

$$\bar{x} = \frac{1}{n}(x_1 + \cdots + x_n) \quad \text{bzw.} \quad \bar{y} = \frac{1}{m}(y_1 + \cdots + y_m).$$

Zu X gehört der mittlere Fehler (111.1), den wir jetzt mit s_1 statt s bezeichnen. Setzen wir zur Abkürzung

$$x_i = \bar{x} + u_i, \quad \text{also} \quad x_i - \bar{x} = u_i,$$

so folgt aus (111.1) der Ausdruck

(113.1) $$s_1 = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n} u_i^2}.$$

Ganz entsprechend erhalten wir, indem wir

$$y_j = \bar{y} + v_j, \quad \text{also} \quad y_j - \bar{y} = v_j$$

setzen, den zu Y gehörigen mittleren Fehler s_2 in der Form

(113.2) $$s_2 = \sqrt{\frac{1}{m-1}\sum_{j=1}^{m} v_j^2}.$$

Greifen wir irgendeinen Meßwert x_i und irgendeinen Meßwert y_j heraus, so können wir den zugehörigen Wert

$$z_{ij} = h(x_i, y_j)$$

berechnen. Auf Grund der TAYLOR-Formel gilt

(113.3) $$\begin{aligned} z_{ij} &= h(\bar{x} + u_i, \bar{y} + v_j) \\ &= h(\bar{x}, \bar{y}) + u_i \frac{\partial h}{\partial x} + v_j \frac{\partial h}{\partial y} + \cdots. \end{aligned}$$

Die durch Punkte angedeuteten Glieder enthalten höhere Potenzen von u_i und v_j. Wir können annehmen, daß $|u_i|$ und $|v_j|$ klein sind, und wollen die genannten Glieder deshalb vernachlässigen. Summieren wir über i von 1 bis n und über j von 1 bis m, so tritt auf der rechten Seite

$$\Sigma u_i = 0 \quad \text{und} \quad \Sigma v_j = 0 \tag{113.4}$$

auf. Übrig bleibt rechts $n m h(\bar{x}, \bar{y})$. Links steht dann die Doppelsumme über alle z_{ij}. Division durch $n m$ liefert links den Mittelwert $\bar{z}$, und rechts ergibt sich gerade $h(\bar{x}, \bar{y})$. Es gilt also einfach

$$\bar{z} = \frac{1}{mn} \sum_{i=1}^{n} \sum_{j=1}^{m} z_{ij} = h(\bar{x}, \bar{y}). \tag{113.5}$$

Wir wollen nun den zugehörigen mittleren Fehler

$$s = \sqrt{\frac{1}{nm-1} \sum_{i=1}^{n} \sum_{j=1}^{m} (z_{ij} - \bar{z})^2} \tag{113.6}$$

durch die mittleren Fehler s_1 und s_2 ausdrücken. Aus (113.3) und (113.5) folgt

$$z_{ij} - \bar{z} \approx u_i \frac{\partial h}{\partial x} + v_j \frac{\partial h}{\partial y}.$$

Wir quadrieren auf beiden Seiten, multiplizieren rechts aus und summieren dann über i und j. Dies ergibt links die Summe in (113.6) und rechts wegen (113.4) den Ausdruck

$$\sum_{i=1}^{n} \sum_{j=1}^{m} u_i^2 \left(\frac{\partial h}{\partial x}\right)^2 + \sum_{i=1}^{n} \sum_{j=1}^{m} v_j^2 \left(\frac{\partial h}{\partial y}\right)^2.$$

Dies ist offenbar gleich

$$\left(\frac{\partial h}{\partial x}\right)^2 m \sum_{i=1}^{n} u_i^2 + \left(\frac{\partial h}{\partial y}\right)^2 n \sum_{j=1}^{m} v_j^2.$$

Wegen (113.1) und (113.2) ist nun aber

$$\sum_{i=1}^{n} u_i^2 = (n-1)\, s_1^2 \quad \text{und} \quad \sum_{j=1}^{m} v_j^2 = (m-1)\, s_2^2.$$

Indem wir dies alles in (113.6) einsetzen, folgt

$$s = \sqrt{\frac{1}{nm-1} \left\{(n-1)\, m \left(\frac{\partial h}{\partial x}\right)^2 s_1^2 + (m-1)\, n \left(\frac{\partial h}{\partial y}\right)^2 s_2^2\right\}}. \tag{113.7}$$

Für großes n und m heben sich die Zahlenfaktoren nahezu weg, und es gilt dann einfach

$$\boxed{s \approx \sqrt{\left(\frac{\partial h}{\partial x}\right)^2 s_1^2 + \left(\frac{\partial h}{\partial y}\right)^2 s_2^2}.} \tag{113.8}$$

Dies ist das sogenannte **Gaußsche Fehlerfortpflanzungsgesetz.** Es gibt an, in welcher Weise sich die Meßfehler durch die Rechnung in das

Resultat hinein fortpflanzen. Die partiellen Ableitungen sind dabei an der Stelle $(\bar{x}, \bar{y})$ zu nehmen.

Beispiel 113.1. Zwei elektrische Widerstände werden mehrmals gemessen, und es ergeben sich dabei die Widerstandswerte

$$R_1 = 100 \pm 0{,}8 \text{ [Ohm]} \qquad \text{und} \qquad R_2 = 200 \pm 1{,}0 \text{ [Ohm]}.$$

Man berechne den Widerstandswert (a) bei Hintereinanderschaltung, (b) bei Parallelschaltung und gebe den zugehörigen mittleren Fehler an.

(a) Der gesuchte Widerstand ist

$$R = R_1 + R_2 = 100 + 200 = 300.$$

In (113.8) brauchen wir

$$\frac{\partial R}{\partial R_1} = 1 \qquad \text{und} \qquad \frac{\partial R}{\partial R_2} = 1.$$

So ergibt sich wegen $s_1 = 0{,}8$ und $s_2 = 1{,}0$ bei Reihenschaltung gemäß (113.8) der mittlere Fehler

$$s = \sqrt{\left(\frac{\partial R}{\partial R_1}\right)^2 s_1^2 + \left(\frac{\partial R}{\partial R_2}\right)^2 s_2^2} = \sqrt{0{,}8^2 + 1{,}0^2} = 1{,}3. \tag{113.9}$$

Unser Ergebnis lautet demnach

$$R = 300{,}0 \pm 1{,}3 \text{ [Ohm]}.$$

(b) Im Falle der Parallelschaltung gilt für den gesuchten Widerstand R das Gesetz

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2}.$$

Indem wir dies nach R auflösen und die obigen Zahlenwerte einsetzen, folgt

$$R = \frac{R_1 R_2}{R_1 + R_2} = \frac{100 \cdot 200}{100 + 200} = 66{,}67 \text{ [Ohm]}.$$

Für die Berechnung des zugehörigen Fehlers brauchen wir die partiellen Ableitungen

$$\frac{\partial R}{\partial R_1} = \frac{R_2^2}{(R_1 + R_2)^2} = \frac{200^2}{(100 + 200)^2} = 0{,}44$$

und

$$\frac{\partial R}{\partial R_2} = \frac{R_1^2}{(R_1 + R_2)^2} = \frac{100^2}{(100 + 200)^2} = 0{,}11.$$

Setzen wir diese Zahlenwerte in (113.8) ein, so erhalten wir mit den gegebenen Werten $s_1 = 0{,}8$ und $s_2 = 1{,}0$ nun den mittleren Fehler

$$s = \sqrt{0{,}44^2 \cdot 0{,}8^2 + 0{,}11^2 \cdot 1{,}0^2} = 0{,}37.$$

Unser Resultat lautet also

$$R = 66{,}67 \pm 0{,}37 \text{ [Ohm]}.$$

Aufgaben zu Abschnitt 113

113.1—113.5 Aus den jeweils gegebenen Meßergebnissen bestimme man die gesuchte Größe und deren mittleren Fehler.

113.1 Man berechne den Flächeninhalt F eines Rechtecks mit den Seiten $a = 150 \pm 0{,}1$ [cm] und $b = 80 \pm 0{,}1$ [cm].

113.2 Gesucht ist der Winkel γ eines ebenen Dreiecks, dessen andere Winkel $\alpha = 40° \pm 30''$ und $\beta = 45° \pm 30''$ sind.

113.3 Eine Strecke L wurde mit 5-m-Meßlatten ausgemessen. Es ergaben sich 9 Lattenlagen. Der mittlere Fehler einer Lattenlage betrug 4,0 mm. Wie groß ist L?

113.4 Gesucht sind die Seiten b und c eines ebenen Dreiecks mit der Seite $a = 210{,}12 \pm 0{,}05$ [m] und den Winkeln $\beta = 52°10' \pm 1'$ und $\gamma = 61°23' \pm 1'$, die zu den Endpunkten von a gehören.

113.5 Wie groß ist das spezifische Gewicht γ einer Eisenkugel mit dem Gewicht $G = 1000 \pm 0{,}1$ [g] und dem Durchmesser $D = 6{,}2 \pm 0{,}01$ [cm]?

113.6 Man zeige: Für die spezielle Funktion $h(x, y) = x\,y$ hat (113.8), geschrieben unter Verwendung der „relativen mittleren Fehler“

$$m = \frac{s}{h(\bar{x}, \bar{y})}, \qquad m_1 = \frac{s_1}{\bar{x}}, \qquad m_2 = \frac{s_2}{\bar{y}}$$

die Form

$$m \approx \sqrt{m_1^2 + m_2^2}.$$

113.7 Welche Form hat (113.8), geschrieben unter Verwendung der relativen mittleren Fehler (s. Aufgabe 113.6), für $h = k\,x^a y^b$? Hierbei seien a, b und k konstant.

113.8 Man löse Aufgabe 113.5 unter Benutzung der Fehlerformel in Aufgabe 113.7.

114 Ausgleichsgeraden und -kurven

Es ist wohl fast selbstverständlich, daß die Bestimmung von Ausgleichsgeraden und -kurven nach dem Prinzip der kleinsten Quadrate, wie wir sie in Kap. 17 kennengelernt haben, in der Fehler- und Ausgleichsrechnung wichtige Anwendungen hat und deshalb in diesem Gebiet eine grundlegende Bedeutung besitzt. Wegen der Ausführlichkeit der Betrachtungen in Kap. 17 können wir uns hier auf einige typische Aufgaben beschränken.

Aufgaben zu Abschnitt 114

114.1—114.3 Man bestimme jeweils die Ausgleichsgerade.

114.1 Spannung y [mV] eines Thermoelementes in Abhängigkeit von der Temperaturdifferenz x [°].

x	20	30	40	50	60	70	80	90
y	0,50	0,77	1,06	1,35	2,02	2,40	2,80	3,15

114.2 Ohmscher Widerstand R [Ohm] in Abhängigkeit von der Temperatur T [°C]

T	15	25	35	50	70	90
R	5,43	5,60	5,81	6,10	6,52	6,95

114.3 Relative Volumenabnahme y des Leders als Funktion des Druckes x [in Vielfachen von 1000 Atm.] (C. E. WEIR, Journ. Res. Nat. Bureau of Standards **45**, 1950, 468)

x	2	4	6	8	10
y	0,000	0,023	0,041	0,057	0,069

114.4—114.6 Man bestimme jeweils die Ausgleichsparabel der Form (104.4).

114.4 Widerstand R [Ohm] einer Kohlefadenlampe in Abhängigkeit von der Stromstärke I [mA]

I	20	30	40	50	60	70	80	90	100
R	1,45	1,33	1,30	1,24	1,20	1,17	1,15	1,12	1,10

114.5

x	−1	0	1	2
y	1	0	1	2

114.6 Frequenz y [Hertz] einer Stahlfederzunge in Abhängigkeit vom Ausschlagswinkel x [Grad] (R. HARTMANN-KEMPF, Dissertation Würzburg 1902; vgl. W. WIEN u. F. HARMS, Handbuch der Experimentalphysik **17**, 105. Leipzig 1934)

x	1,5	2,0	2,5	3,0	4,0	5,0	6,5	9,0	12,5
y	98,98	98,97	98,97	98,96	98,96	98,98	99,00	99,07	99,19

Kapitel 20

Verteilungsunabhängige Verfahren

Verteilungsunabhängige Verfahren (auch **nichtparametrische Verfahren** genannt) sind Verfahren, deren Anwendbarkeit nicht an die Voraussetzung gebunden ist, daß eine ganz bestimmte Verteilung (z. B. eine Normalverteilung) vorliegt. Man wendet sie an, wenn man über die Verteilung der Grundgesamtheit, aus der eine Stichprobe stammt, gar nichts weiß. Sie sind rechnerisch einfach.

Man kann von diesen Methoden im Einzelfalle nicht so viel erwarten, wie ein Verfahren, das für eine bestimmte Verteilung zugeschnitten ist, beim Vorliegen der betreffenden Verteilung liefert. Aber der Unterschied ist manchmal gar nicht so groß. Zum Beispiel braucht man im Falle großer Stichproben aus normalverteilten Grundgesamtheiten beim Test in Abschn. 118 etwa 5% mehr Stichprobenwerte, um dieselbe Information wie beim Test in Abschn. 79 zu erhalten. Bei kleinen Stichproben ist das Verhältnis etwas ungünstiger. Schätzt man, daß man mit 9 Werten bei dem Test in Abschn. 79 so viel erreicht wie mit 10 Werten bei dem Rangtest in Abschn. 118 und spart man für 10 Pfennige Rechenarbeit beim Rangtest, so lohnt sich dieser, falls die Einzelmessung weniger als 10 Pfennige kostet. Ist hingegen die Einzelmessung teuer und die Grundgesamtheit angenähert normalverteilt, so sollte man den Test in Abschn. 79 bevorzugen.

Beim Vorzeichentest in Abschn. 115 braucht man im Falle großer Stichproben aus normalverteilten Grundgesamtheiten etwa den anderthalbfachen Umfang, verglichen mit dem Test in Abschn. 79.

Im vorliegenden Kap. 20 setzen wir voraus, daß die vorkommenden Verteilungen **stetig** sind.

115 Vorzeichentest

Den sehr einfachen Grundgedanken und die praktische Durchführung dieses Tests wollen wir an einem Beispiel erläutern.

Beispiel 115.1. Um eine Spezialsämaschine mit einer gewöhnlichen zu vergleichen, wurden 10 Parzellen je zur Hälfte mit der einen und mit der anderen Maschine besät. Die mit der Spezialmaschine besäten Hälften zeigten die folgenden Mehrerträge [Bushels] gegenüber den jeweils zugehörigen anderen Hälften:

$$2{,}4 \quad 1{,}0 \quad 0{,}7 \quad 0{,}0 \quad 1{,}1 \quad 1{,}6 \quad 1{,}1 \quad -0{,}4 \quad 0{,}1 \quad 0{,}7$$

(J. WISHART, Suppl. J. Royal Statist. Soc. **1**, 1934, 26—51). Man teste die Hypothese, daß beide Maschinen hinsichtlich des erzielten Ertrages gleich gut sind, gegen die Alternative, daß man mit der Spezialsämaschine höhere Erträge erzielt. Als Signifikanzzahl wähle man $\alpha = 5\%$.

Träfe die Hypothese zu, so wäre die Wahrscheinlichkeit p eines positiven Wertes, also eines Mehrertrages, ebenso groß wie die eines negativen Wertes. Demnach gälte dann $p = 0{,}5$, und die Zufallsvariable

$$X = \textit{Anzahl der positiven Werte unter } n \textit{ Werten}$$

wäre binomialverteilt mit $p = 0{,}5$. Die Stichprobe hat 10 Werte. Den Wert 0 scheiden wir aus, weil er nichts zur Entscheidung beiträgt. Dann bleiben noch 9 Werte übrig, davon 8 positive. Die Wahrscheinlichkeit, unter 9 Werten wenigstens 8 positive zu erhalten, beträgt, wenn die Hypothese zutrifft,

$$P(X \geqq 8) = 1 - P(X \leqq 7) = 1 - 0{,}9805 = 1{,}95\%$$

(siehe Tafel 1d in Anhang 5 und Aufgabe 41.5). Da dies kleiner als 5% ist, verwerfen wir die Hypothese und nehmen an, daß man mit der Spezialmaschine größere Erträge erhält.

Aufgaben zu Abschnitt 115

115.1 Man nehme an, die Stichprobe in Beispiel 115.1 entstamme einer normalverteilten Grundgesamtheit, und führe den Test nach der Methode in Abschn. 79 durch.

115.2 Zwei Schlafmittel A und B wurden jeweils an denselben 10 Patienten ausprobiert („STUDENT", Biometrika **6**, 1908, 1—25). Dabei ergaben sich die folgenden Werte für die Schlafverlängerung [Stunden]:

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
Diff.	1,2	2,4	1,3	1,3	0,0	1,0	1,8	0,8	4,6	1,4

Ist der Unterschied signifikant?

115.3 Man bearbeite die Aufgabe 115.2 nach der Methode in Abschn. 79.

115.4 Man bearbeite Aufgabe 79.7 mittels des Vorzeichentests. Die hierbei notwendigen Werte der Binomialverteilung berechne man selbst.

115.5 Rindspastetchen wurden 8 Monate aufbewahrt, zum Teil bei −18 °C, zum Teil bei zwischen −9 °C und −18 °C schwankender Temperatur. 7 von 8 befragten Personen empfanden das Aroma der erstgenannten Pastetchen als angenehmer, eine Person urteilte umgekehrt. Ist der Unterschied signifikant? (F. EHRENKRANTZ u. H. ROBERTS, Journ. of Home Econom. **44**, 1952, 441.)

116 Test für beliebigen Trend

Auch diesen Test erläutern wir am besten durch ein typisches Beispiel. Sind die in Kap. 17 gemachten Voraussetzungen erfüllt, so wende man die dortigen Methoden an, die dann bessere Ergebnisse liefern.

Beispiel 116.1. Über die Wirkung kleiner Strahlendosen bei Saatgutbestrahlung weiß man noch nicht viel. A. Süss (Atompraxis **9**, 1963, 91) bestrahlte im Jahre 1959 Sommergerste Stanka und erhielt im Jahre 1960 die folgenden Kornerträge [g/Gefäß]:

Strahlendosis x	0	$2\,r$	$4\,r$	$40\,r$	$400\,r$
Ertrag y	29,2	30,2	28,2	29,7	30,4

Man teste die Hypothese, daß die angewendete schwache Bestrahlung keinen Einfluß hat, gegen die Alternative, daß positiver Trend vorliegt, d. h. daß sich der Ertrag mit wachsender Strahlung vergrößert.

Die x-Werte sind der Größe nach geordnet. Bestünde positiver Trend, so hätten die y-Werte auch die Tendenz anzuwachsen. Bestünde kein Trend, so sollten die y-Werte ungeordnet aufeinanderfolgen. Wir zählen, wie oft ein größerer Wert vor einem kleineren steht:

29,2 steht vor 28,2 (1 „Fehlordnung")
30,2 steht vor 28,2 und 29,7 (2 „Fehlordnungen").

Die restlichen Werte (28,2, 29,7, 30,4) stehen der Größe nach. Wir haben also insgesamt 3 „Fehlordnungen" (Inversionen). Die Verteilung der Variablen

$$T = \textit{Anzahl der Fehlordnungen}$$

unter der Hypothese, daß kein Trend besteht, können wir leicht aufstellen: Besteht kein Trend, so sind alle $5! = 120$ Permutationen von 5 Elementen 1 2 3 4 5 gleichmöglich (vgl. Abschn. 35), und wir ordnen diese nach der Anzahl ihrer Fehlordnungen an:

$T = 0$	$T = 1$	$T = 2$	$T = 3$
1 2 3 4 5	1 2 3 5 4	1 2 4 5 3	1 2 5 4 3
	1 2 4 3 5	1 2 5 3 4	1 3 4 5 2
	1 3 2 4 5	1 3 2 5 4	1 3 5 2 4
	2 1 3 4 5	1 3 4 2 5	1 4 2 5 3
		1 4 2 3 5	1 4 3 2 5
		2 1 3 5 4	1 5 2 3 4
		2 1 4 3 5	2 1 4 5 3
		2 3 1 4 5	2 1 5 3 4 usw.
		3 1 2 4 5	2 3 1 5 4
			2 3 4 1 5
			2 4 1 3 5
			3 1 2 5 4
			3 1 4 2 5
			3 2 1 4 5
			4 1 2 3 5

Wie wir sehen, ist

$$P(T \leqq 3) = \frac{1}{120} + \frac{4}{120} + \frac{9}{120} + \frac{15}{120} = \frac{29}{120} = 24\%.$$

Wir nehmen deshalb die Hypothese an.

Für diesen Test verwende man die Tafel 10a in Anhang 5. Unsere Überlegung und die Tafel gelten für den Fall einer stetig verteilten Grundgesamtheit. Theoretisch sind dann nämlich lauter verschiedene Stichprobenwerte zu erwarten. Praktisch können trotzdem mehrere Werte gleich sein, weil man abrundet. Sind m Werte gleich, so zähle man zu den sich sowieso ergebenden Fehlordnungen noch $m(m-1)/4$ (= Mittelwert der Fehlordnungen bei den Permutationen von m Dingen) dazu, also je 1/2 für je 2 gleiche Werte, je 3/2 für jedes Tripel gleicher Werte usw. Ist die Verteilung der Grundgesamtheit diskret, so muß man daran denken, daß auch theoretisch derselbe Meßwert mehrmals auftreten kann, und die obige Überlegung entsprechend abändern.

Aufgaben zu Abschnitt 116

116.1 Man stelle die in Beispiel 116.1 angefangene Tafel fertig.

116.2 Man stelle die der Tafel in Beispiel 116.1 entsprechende Tafel für $n = 4$ auf.

116.3—116.5 Man wende den im Text besprochenen Test auf die angegebene Stichprobe an.

116.3 Die Stichprobe in Aufgabe 95.4.

116.4 Die Stichprobe in Aufgabe 95.5.

116.5 Die Stichprobe in Aufgabe 95.10.

116.6 Wie ist die Überlegung im Text im Falle diskreter Verteilungen der Grundgesamtheit zu modifizieren?

117 Testen der Zufälligkeit in Stichproben

Wir werfen eine Münze zehnmal und schreiben die Ergebnisse nacheinander auf (K = „Kopf", W = „Wappen"). Dann fassen wir gleichartige nebeneinanderstehende Buchstaben zusammen, indem wir sie unterklammern:

$$\underbrace{K\;K}\quad \underbrace{W\;W\;W\;W}\quad \underbrace{K\;K}\quad \underbrace{W}\quad \underbrace{K}$$

Damit haben wir unsere Stichprobe in 5 Teile unterteilt. Diese heißen **Iterationen.** Zuwenige Iterationen, wie z. B. im Falle

$$\underbrace{K\;K\;K\;K\;K\;K}\quad \underbrace{W\;W\;W\;W} \qquad \text{(2 Iterationen),}$$

oder zuviele Iterationen, wie z. B. im Falle

$$\underbrace{K}\;\underbrace{W}\;\underbrace{K}\;\underbrace{W}\;\underbrace{K}\;\underbrace{W}\;\underbrace{K}\;\underbrace{W}\;\underbrace{K}\;\underbrace{W} \qquad \text{(10 Iterationen)}$$

(und außerdem die Regelmäßigkeit), lassen den Verdacht aufkommen, daß die Zufälligkeit gestört ist.

Wie definiert und bestimmt man Iterationen, wenn es sich um eine Stichprobe von Zahlenwerten handelt? Dann zeichnet man diese Werte der Reihe nach auf und zieht eine horizontale Gerade M („*Medianlinie*") derart, daß die Hälfte der Stichprobenwerte oberhalb und die Hälfte unterhalb liegt. Unten hin schreibt man ein Pluszeichen bzw. ein Minuszeichen je nachdem, ob der betreffende Stichprobenwert oberhalb oder unterhalb der Medianlinie liegt (siehe Abb. 117.1).

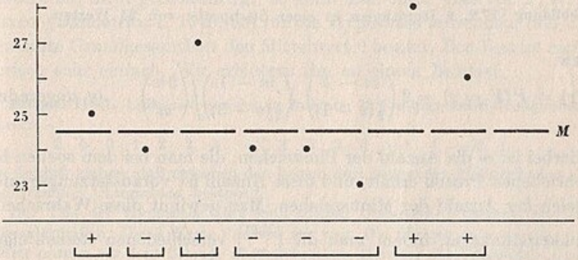


Abbildung 117.1. Zur Definition der Iterationen (Stichprobe von Schalldämmzahlen in Aufgabe 4.4)

Es kann sein, daß es gar keine solche Gerade M gibt und daß auf der horizontalen Geraden M^*, die die Stichprobe wenigstens annähernd in zwei gleich umfangreiche Teile teilt, zwei oder drei oder noch mehr Stichprobenwerte liegen. Dann ordne man einen oder mehrere Stichprobenwerte auf M^* demjenigen Teil zu, der weniger Punkte hat, bis beide Teile gleichviele Werte umfassen. Dabei ist so zu verfahren, daß möglichst viele Iterationen entstehen. In Abb. 117.2 können wir z. B. den mit einem Pfeil versehenen Wert dem unteren Teil zurechnen, der dann aus 8 Werten besteht.

Wie wir ohne Beweis erwähnen, hat im Falle einer stetigen Verteilung die Zufallsvariable

$$V = \textit{Anzahl der Iterationen}$$

bei Zutreffen der Hypothese, daß Zufälligkeit der Reihenfolge vorliegt, die Wahrscheinlichkeitsfunktion

$$f(v) = P(V = v) = 2\binom{m-1}{\frac{1}{2}(v-2)}^2 \Big/ \binom{2m}{m} \qquad (v \text{ gerade})$$

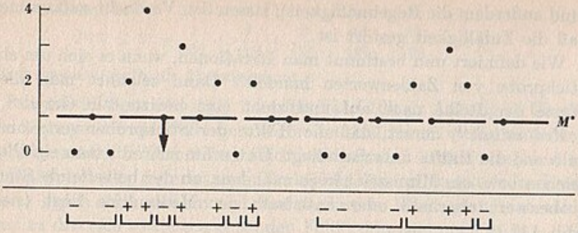


Abbildung 117.2. 9 Iterationen in einer Stichprobe von 25 Werten

bzw.

$$f(v) = P(V = v) = 2\binom{m-1}{\frac{1}{2}(v-1)}\binom{m-1}{\frac{1}{2}(v-3)}\Big/\binom{2m}{m} \qquad (v \text{ ungerade}).$$

Hierbei ist m die Anzahl der Pluszeichen, die man bei dem soeben beschriebenen Prozeß erhält, und diese Anzahl ist voraussetzungsgemäß gleich der Anzahl der Minuszeichen. Man gewinnt diese Wahrscheinlichkeitsfunktion, indem man die $\binom{2m}{m}$ verschiedenen Reihenfolgen von m Pluszeichen und m Minuszeichen als gleichwahrscheinlich ansieht und jeweils die Anzahl von Iterationen bestimmt.

Ist die Anzahl der Iterationen zu klein oder zu groß, so verwerfen wir die genannte Hypothese. Die kritischen Werte, von denen ab verworfen wird, sind in Tafel 10b für die Signifikanzzahlen 5% und 1% angegeben.

In Abb. 117.2 ist $m = 8$, und es liegen $v = 9$ Iterationen vor. Wir wählen die Signifikanzzahl 5%. Aus Tafel 10b entnehmen wir dazu die kritischen Werte 5 und 12. Da $v = 9$ zwischen diesen Werten liegt, wird die Hypothese, daß Zufälligkeit der Reihenfolge vorliegt, angenommen.

Aufgaben zu Abschnitt 117

117.1—117.3 Man wende den im Text betrachteten Test auf die folgenden Stichproben an:

117.1 Die Stichprobe in Aufgabe 4.3.

117.2 Die Stichprobe in Aufgabe 4.5.

117.3

97	3	16	12	55	16	84	63	33	57
18	26	23	52	37	70	56	99	16	31

118 Ein Rangtest

In Abschn. 115 haben wir ein Beispiel eines Tests kennengelernt, bei dem man nur die Vorzeichen der Stichprobenwerte berücksichtigt. Es gibt andere Tests, bei denen man zwar auch Information verschenkt, aber nicht gar so viel, indem man nämlich außer dem Vorzeichen noch die Größe der Stichprobenwerte in gewisser Weise mitberücksichtigt. Einen solchen Test wollen wir jetzt betrachten.

Liegen Differenzen von Meßwerten vor, wie z. B. in Aufgabe 79.7, und ist die Annahme, daß die zugehörige Grundgesamtheit normalverteilt ist, nicht gerechtfertigt, so kann man einen Test von F. Wilcoxon (Biometrics 1, 1945, 80) für die Hypothese anwenden, daß die genannte Grundgesamtheit den Mittelwert 0 besitzt. Der Test ist rechnerisch sehr einfach. Wir erläutern ihn an einem Beispiel.

Beispiel 118.1. Läßt sich aus den in Aufgabe 79.7 angegebenen Stichprobenwerten

$$2 \quad 0 \quad 0 \quad 1 \quad 2 \quad 2 \quad 3 \quad -3 \quad 1 \quad 2 \quad 3 \quad 0 \quad -1 \quad 1 \quad -2 \quad 1$$

der Schluß ziehen, daß zwischen den beiden dort genannten Meßmethoden ein signifikanter Unterschied besteht?

Der Test der Hypothese, daß kein solcher Unterschied besteht, verläuft folgendermaßen: Die 3 Werte 0 lassen wir weg. Die übrigen $m = n - 3 = 13$ Werte ordnen wir dem Betrage nach, schreiben darunter das Vorzeichen, dann die Platznummer und schließlich den Rang. Letzterer ist die mit dem genannten Vorzeichen versehene Platznummer bzw. das arithmetische Mittel der Platznummern betragsgleicher Stichprobenwerte. So ergibt sich:

Betrag	1	1	1	1	1	2	2	2	2	2	3	3	3
Vorzeichen	−	+	+	+	+	−	+	+	+	+	−	+	+
Platz Nr.	1	2	3	4	5	6	7	8	9	10	11	12	13
Rang	−3	3	3	3	3	−8	8	8	8	8	−12	12	12

In der Tat sind die ersten 5 Stichprobenwerte betragsgleich, und das arithmetische Mittel der Platznummern ist $(1 + 2 + 3 + 4 + 5)/5 = 3$ usw. Die Summe der negativen Ränge ist $-3 - 8 - 12 = -23$. Deren Betrag $u_0 = 23$ ist unsere Testgröße. Überwiegen die positiven oder die negativen Ränge zu sehr, so ist u_0 zu klein bzw. zu groß, und die Hypothese wird verworfen. Den unteren kritischen Wert in diesem zweiseitigen Test zeigt Tafel 10c. Zur Signifikanzzahl $\alpha = 5\%$ gehört 17, und zu $\alpha = 1\%$ gehört 10, denn es ist ja $m = n - 3 = 13$. Den oberen kritischen Wert brauchen wir gar nicht, sobald wir uns vergewissert haben, daß u_0 nicht größer als die Summe der positiven Ränge ist. Wäre dies der Fall, hätten wir einfach die letztere als Testgröße u_0 zu nehmen. In unserem Falle ist $u_0 = 23$, und die Hypothese wird bei Wahl von $\alpha = 1\%$ und auch $\alpha = 5\%$ angenommen.

Die Tafel 10c erhält man durch kombinatorische Überlegungen. Zum Beispiel gibt es für $m = 8$ gemäß Satz 37.1 genau $2^8 = 256$ Stichproben verschiedener Vorzeichenkombinationen, die wir nach steigendem u_0 geordnet hinschreiben:

Platz Nr. 1	2	3	4	5	6	7	8	u_0
+	+	+	+	+	+	+	+	0
−	+	+	+	+	+	+	+	1
+	−	+	+	+	+	+	+	2
+	+	−	+	+	+	+	+	3
−	−	+	+	+	+	+	+	3
+	+	+	−	+	+	+	+	4
−	+	−	+	+	+	+	+	4

und so weiter. Die Stichprobenwerte sind ja jeweils Differenzen aus zugehörigen Werten zweier Stichproben. Deren Grundgesamtheiten setzen wir als stetig verteilt voraus. Im Falle der Richtigkeit der Hypothese sind diese identisch. Dann sind die obigen 256 Stichprobentypen gleichwahrscheinlich, wie man beweisen kann. Also bilden die ersten 6 ($\approx 2{,}5\%$ von 256) und die letzten 6 Stichproben den Verwerfungsbereich beim zweiseitigen Test mit der Konfidenzzahl $\alpha = 5\%$. Die zugehörigen kritischen Werte sind $u_0 = 4$, in Übereinstimmung mit Tafel 10c, und ferner $u_0 = 36 - 4 = 32$.

Aufgaben zu Abschnitt 118

118.1—118.2 Man wende den besprochenen Test auf die folgenden Stichproben an:

118.1 Die Stichprobe in Aufgabe 79.5.

118.2 Die Stichprobe in Beispiel 115.1.

118.3 Man vergleiche die in Tab. 10c angegebene Näherungsformel für $m = 10$, 15 und 20 mit den genauen Werten.

118.4 Man berechne die kritischen Werte für $m = 25$, 30, 40 mittels der in Tafel 10c angegebenen Näherungsformel.

Kapitel 21

Entscheidungstheorie

Die statistische Entscheidungstheorie wurde von A. Wald begründet; vgl. [15] in Anhang 3. Sie behandelt statistische Verfahren unter Benutzung von Ideen und Methoden aus der sog. *mathematischen Spieltheorie*, die auf J. v. Neumann [11] zurückgeht. Dies führt zu einer einheitlicheren Behandlungsweise und Gütebeurteilung statistischer Verfahren.

Das vorliegende Kapitel bringt eine Einführung in die Grundbegriffe und -prinzipien der Entscheidungstheorie. Weitere Einzelheiten und Literaturangaben findet man in [1] und [15] (s. Anhang 3).

Die Entscheidungstheorie umfaßt die Theorie der Punktschätzungen für Parameter (Kap. 12), Konfidenzintervalle (Kap. 13) und Tests (Kap. 14) als Sonderfälle. Zu entscheiden hat man sich jeweils für eine von mehreren möglichen Aktionen. Die Entscheidung wird getroffen unter Berücksichtigung des Ausganges eines Zufallsexperimentes. Daß sich die Spieltheorie dabei als brauchbares Werkzeug erweist, hat folgenden Grund:

Ein *Zwei-Personen-Spiel* ist ein Spiel für 2 Spieler. Genauer interessiert uns hier der Begriff des *Zwei-Personen-Nullsummen-Spiels*, das ist ein Zwei-Personen-Spiel, bei dem der Gewinn des einen Spielers jeweils gleich dem Verlust des anderen ist, also insgesamt die Summe null herauskommt. Jeder der beiden Spieler hat das Bestreben, seinen Gewinn maximal (oder seinen Verlust minimal) zu machen. In der Entscheidungstheorie wird ein statistisches Verfahren nun aufgefaßt als ein Zwei-Personen-Spiel: Der eine Spieler ist fiktiv — wir nennen ihn die „*Natur*" (oder „*Umwelt*") —, der andere ist der Statistiker. Um diesen Grundgedanken verständlich zu machen, betrachten wir ein Beispiel:

Nehmen wir an, der Statistiker interessiere sich für eine normalverteilte Zufallsvariable X mit bekannter Varianz und unbekanntem Mittelwert μ. Wäre dem Statistiker der Zahlenwert μ bekannt (und könnte der Statistiker die Folgen aller Aktionen, die für ihn möglich sind, genau übersehen), so wüßte er genau, für welche Aktion er sich vernünftigerweise entscheiden sollte. Da er aber den Wert μ nicht kennt, kann er sich vorstellen, er spiele ein Spiel gegen die „Natur", das sich so beschreiben läßt: Die „Natur" wählt einen Zahlenwert μ (den der Statistiker nicht kennt). Der Statistiker „spioniert" dann etwas (macht ein Zufallsexperiment), um Anhaltspunkte über das ihm unbekannte μ zu erhalten. Er entscheidet sich schließlich für eine günstige der für ihn möglichen Aktionen. Bei dieser Entscheidung berücksichtigt der Statistiker das Ergebnis des Zufallsexperimentes und bewertet jede mögliche Aktion nach den (noch von μ abhängigen) Folgen, die sie für ihn haben würde.

Entsprechend kann man andere statistische Verfahren als Spiele gegen die Natur ansehen. Diese Auffassung hat zu wichtigen Fortschritten in der Statistik geführt.

119 Entscheidungsproblem. Verlust. Risiko

Wir geben nun eine allgemeine Kennzeichnung von Entscheidungsproblemen als Spielen des Statistikers gegen die Natur und führen dabei einige Grundbegriffe ein.

1. Die Natur wählt ein Element θ aus einer gegebenen Menge Ω. Diese Menge Ω heißt **Parameterraum**, und jedes Element θ von Ω heißt ein **Zustand** der Natur. (In unserem obigen Beispiel war $\theta = \mu$ ein Parameterwert.)

2. Der Statistiker wählt ein Element a (*Aktion* oder *Entscheidung* genannt) aus einer gegebenen Menge A. Die Menge A heißt der **Raum der Aktionen** des Statistikers.

3. Bei diesem Spiel hat der Statistiker einen Gewinn oder Verlust, der davon abhängt, was die Natur tut und was er selber tut. Der Verlust ist also eine Funktion L von θ und a. Diese heißt **Verlustfunktion** (*loss function* im Englischen). Und jedem Paar (θ, a) entspricht eine gewisse Zahl $L(\theta, a)$ als zugehöriger *Verlust*. L darf auch negativ sein; ein solcher negativer Verlust $L(\theta, a)$ bedeutet dann einen Gewinn für den Statistiker.

$L(\theta, a)$ ist also der Verlust für den Statistiker, wenn er Aktion a wählt und sich die Natur im Zustand θ befindet.

Würde der Statistiker den Zustand der Natur kennen, so würde er aus den Werten der Verlustfunktion L ersehen, welche Aktion für ihn am günstigsten ist, und ein Zufallsexperiment wäre dann ganz überflüssig. Da er den Zustand der Natur aber nicht kennt, sammelt er zuerst Informationen darüber, indem er eine Zufallsvariable X (oder (X, Y) usw.) beobachtet, deren Verteilung vom Zustand θ der Natur abhängt. Diese Information benutzt der Statistiker, wenn er schließlich entscheidet, welche Aktion a aus A er wählt.

Der Statistiker spielt also ein Spiel gegen die Natur, das durch Ω, A, L gekennzeichnet ist und bei dem er im Rahmen eines Zufallsexperimentes eine Zufallsvariable X beobachtet, deren Verteilungsfunktion $F(x, \theta)$ vom Zustand θ der Natur abhängt. Ein solches Spiel, das in dieser Weise mit einem Zufallsexperiment gekoppelt ist, heißt ein **statistisches Entscheidungsproblem** oder **statistisches Spiel.**

Grob gesprochen werden also statistische Verfahren (samt möglichen Konsequenzen) in der Entscheidungstheorie als Entscheidungsprobleme aufgefaßt. Typische Beispiele folgen im nächsten Abschnitt.

Wie gesagt, hängt die Wahl einer Aktion a von dem beobachteten Wert x der Zufallsvariablen X ab, ist also eine Funktion von x, die (nichtrandomisierte) **Entscheidungsfunktion** heißt (im Englischen (nonrandomized) *decision function*) und mit d bezeichnet wird:

$$a = d(x).$$

(In der Sprache der Spieltheorie heißt d eine *reine Strategie.*) Entsprechend hat man bei Benutzung einer Stichprobe $x_1, \ldots, x_n$ die Beziehung

$$a = d(x_1, \ldots, x_n). \tag{119.1}$$

Setzen wir $a = d(x)$ in die Verlustfunktion $L(\theta, a)$ ein, so erhalten wir eine Funktion

$$L(\theta, d(x)),$$

die von x abhängt. x ist dabei, wie gesagt, ein beobachteter Wert einer Zufallsvariablen X. Also ist $d(x)$ ein beobachteter Wert der Zufallsvariablen $d(X)$. Demnach ist auch

$$L(\theta, d(X))$$

eine Zufallsvariable, und wir können deren Mittelwert bilden. Dieser wird mit R bezeichnet; er hängt von θ und d ab und heißt **Risikofunktion.** Wir haben also, in Formeln geschrieben,

$$R(\theta, d) = E(L(\theta, d(X))). \tag{119.2}$$

Dies ist der durchschnittlich zu erwartende Verlust des Statistikers, wenn er eine gewisse Entscheidungsfunktion d wählt und die Natur sich im Zustand θ befindet. Vorausgesetzt wird dabei, daß d so beschaffen ist, daß R für jeden Zustand θ aus Ω einen endlichen Wert hat.

Aufgaben zu Abschnitt 119

119.1 Jeder von zwei Spielern wählt eine ganze Zahl. Sind beide Zahlen ungerade, zahlt S_1 DM 1,— an S_2. Sind beide Zahlen gerade, zahlt S_1 DM 3,— an S_2. In den beiden übrigen Fällen zahlt S_2 DM 2,— an S_1. Man gebe Ω, A und alle Werte von L (= Verlust des Spielers S_2) an.

119.2 θ sei die Wahrscheinlichkeit für „Kopf" beim Wurf einer Münze, und es sei entweder $\theta = \theta_1 = 0{,}4$ oder $\theta = \theta_2 = 0{,}6$. Es sei A die Zahlengerade. Die Entscheidungsfunktion $d(x)$ sei gegeben durch die Werte $d(0) = a_0$, $d(1) = a_1$; dabei seien a_0 und a_1 Zahlen, und x sei ein beobachteter Wert der Zufallsvariablen X = *Anzahl „Köpfe" beim einzelnen Wurf*. Die Verlustfunktion sei $L(\theta, a) = (\theta - a)^2$, also $L(\theta_1, d(0)) = (0{,}4 - a_0)^2$ usw. Man zeige, daß die Risikofunktion die folgenden Werte hat:

$$R(\theta_1, d) = 0{,}6(0{,}4 - a_0)^2 + 0{,}4(0{,}4 - a_1)^2,$$
$$R(\theta_2, d) = 0{,}4(0{,}6 - a_0)^2 + 0{,}6(0{,}6 - a_1)^2.$$

119.3 Eine Urne enthalte 2 Kugeln. θ sei die Anzahl weißer Kugeln in der Urne, so daß also $\Omega = \{0, 1, 2\}$ ist. Es sei $A = \{a_0, a_1, a_2\}$ mit $a_0 = 0$, $a_1 = 1$, $a_2 = 2$. Der Statistiker zieht eine Kugel aus der Urne und beobachtet X = *Anzahl weißer Kugeln beim einzelnen Zug*. Wie viele (nichtrandomisierte) Entscheidungsfunktionen gibt es bei diesem Problem? Einige davon sind sinnlos. Die übrigens gebe man an. (Die „Aktion" a_0 des Statistikers bestehe also hier lediglich darin, daß er die Feststellung trifft, daß $a_0 = 0$ weiße Kugeln in der Urne sind, usw.)

119.4 In Aufgabe 119.3 sei die Verlustfunktion

$$L(\theta, a) = |\theta - a|.$$

Man bestimme die Werte der Risikofunktionen, die zu den folgenden Entscheidungsfunktionen gehören:

	$x = 0$	$x = 1$
$d_1(x)$	a_0	a_1
$d_2(x)$	a_0	a_2
$d_3(x)$	a_1	a_1
$d_4(x)$	a_1	a_2

119.5 Man zeige, daß in Aufgabe 119.3 im Fall der Verlustfunktion $L(\theta, a) = (\theta - a)^2$ die Risikofunktion R die folgenden Werte hat:

	$\theta = 0$	$\theta = 1$	$\theta = 2$
d_1	0	0,5	1
d_2	0	1	0
d_3	1	0	1
d_4	1	0,5	0

119.6 In Aufgabe 119.3 nehme man an, der Statistiker ziehe 2 Kugeln mit Zurücklegen und beobachte $X =$ *Anzahl weißer Kugeln unter den gezogenen*. Man gebe alle (nichtrandomisierten) Entscheidungsfunktionen an, die sinnvoll sind.

119.7 In Aufgabe 119.3 nehme man an, der Statistiker ziehe 2 Kugeln mit Zurücklegen und beobachte $X =$ *Anzahl weißer Kugeln unter den gezogenen*. Die Verlustfunktion sei $L(\theta, a) = |\theta - a|$. Man bestimme die Risikofunktion $R(\theta, d_j)$ für die folgenden Entscheidungsfunktionen.

x	$d_1(x)$	$d_2(x)$	$d_3(x)$	$d_4(x)$
0	0	0	1	1
1	1	1	1	1
2	1	2	1	2

120 Beispiele

Diese Beispiele sollen zeigen, wie sich das Testen von Hypothesen (vgl. Kap. 14), die Punktschätzung von Parametern (vgl. Kap. 12) und die Angabe von Konfidenzintervallen (vgl. Kap. 13) als spezielle Entscheidungsprobleme im Rahmen der Entscheidungstheorie formulieren lassen.

Beispiel 120.1. Beim Testen von Hypothesen gibt es 2 mögliche Aktionen:

a_1: Annehmen (Nichtverwerfen) der Hypothese

a_2: Verwerfen der Hypothese

Also ist hier $A = \{a_1, a_2\}$. Der Einfachheit halber nehmen wir an, daß die Natur auch nur 2 mögliche Zustände $\theta = \theta_0$ und $\theta = \theta_1$ hat, und wir wollen die Hypothese $H_0: \theta = \theta_0$ gegen die Alternative $H_1: \theta = \theta_1$ testen. Als Verlustfunktion können wir dann $L(\theta, a)$ mit den folgenden Werten benutzen:

	$\theta = \theta_0$	$\theta = \theta_1$
$a = a_1$	0	L_{II}
$a = a_2$	L_I	0

L_I und L_{II} sind dabei positive Konstanten. Ist die Natur im Zustand θ_0 und wählen wir Aktion a_1 (Annehmen der Hypothese), so ist der Verlust 0. Dasselbe gilt, wenn die Natur den Zustand θ_1 hat und wir uns für a_2 (Verwerfen der Hypothese) entscheiden. In den beiden anderen noch möglichen Fällen begehen wir einen Fehler 1. Art (zugehöriger Verlust L_I) oder 2. Art (zugehöriger Verlust L_{II}); vgl. Abschn. 78. Die zugehörigen Wahrscheinlichkeiten sind

$$\alpha = P(d(X_1, \ldots, X_n) = a_2;\ \theta = \theta_0 \text{ ist richtig})$$

bzw.

$$1 - \beta = P(d(X_1, \ldots, X_n) = a_1;\ \theta = \theta_1 \text{ ist richtig}).$$

Diese Wahrscheinlichkeiten hängen von der Wahl der Entscheidungsfunktion d ab. Die Risikofunktion

$$R(\theta, d) = E(L(\theta, d(X_1, \ldots, X_n)))$$

hat die Werte

$$R(\theta_0, d) = (1-\alpha) \cdot 0 + \alpha L_I = \alpha L_I$$

und

$$R(\theta_1, d) = (1-\beta) L_{II} + \beta \cdot 0 = (1-\beta) L_{II}.$$

Wir sollten darauf hinweisen, daß die Theorie des Testens statistischer Hypothesen wohl noch nicht ihre endgültige Form gefunden hat, da man bei der Begründung vieler Tests gewisse Konventionen aus der Praxis mitberücksichtigt.

Oft kann man von mehreren Tests für dieselbe Hypothese einen günstigen Test dadurch auswählen, daß man zuerst eine Verlustfunktion bestimmt, die (durch ihre Werte) die relative Wichtigkeit des Fehlers angibt (und den Schaden, der durch das Nichtverwerfen einer falschen Hypothese entsteht). Man wählt dann einen Test aus, bei dem der zu erwartende Verlust möglichst klein wird. Allerdings gibt es bei diesem Vorgehen gewisse Schwierigkeiten, auf die wir später eingehen.

Beispiel 120.2. Die in Kap. 12 behandelte Aufgabe der Punktschätzung von Parametern läßt sich ebenfalls als ein statistisches Entscheidungsproblem auffassen. Beispielsweise interessiere uns eine Zufallsvariable, deren Varianz wir kennen und deren Mittelwert μ wir abschätzen möchten. μ ist dann der Parameter θ. Es möge μ irgendwelcher Werte fähig sein. Dann nehmen wir als Parameterraum Ω die ganze Zahlengerade. Für jedes μ gebe es eine verschiedene Aktion a. Als Raum A der Aktionen können wir dann ebenfalls die Zahlengerade nehmen, und jeder Schätzwert $\hat{\mu}$ für μ ist eine Aktion. (Die „Aktion" des Statistikers besteht also darin, daß er den Näherungswert $\hat{\mu}$ für wahr hält; der Begriff *Aktion* ist demnach hier etwas allgemeiner als in der Umgangssprache.)

Eine Entscheidungsfunktion d, die solche Schätzwerte liefert, ist

$$a = d(x_1, \ldots, x_n) = \bar{x} = \frac{1}{n}(x_1 + \cdots + x_n).$$

Wir könnten auch andere Entscheidungsfunktionen benutzen und stehen damit wiederum vor dem Problem der Wahl einer möglichst „guten" Entscheidungsfunktion.

Als Verlustfunktion wollen wir die **quadratische Verlustfunktion**

$$L(\theta, a) = (\theta - a)^2 \tag{120.1}$$

verwenden. Wir können nun versuchen, die Entscheidungsfunktion $d(x_1, \ldots, x_n)$ so zu wählen, daß

$$R(\theta, d) = E((\theta - \hat{\Theta})^2)$$

minimal wird; hierbei ist

$$\hat{\Theta} = d(X_1, \ldots, X_n)$$

eine Schätzfunktion für θ. Und wenn wir annehmen, daß $\hat{\Theta}$ erwartungstreu ist, so zeigt Abschn. 64, daß Minimisierung von $R(\theta, d)$ bedeutet, daß d wirksam sein soll.

Nehmen wir $d(x_1, \ldots, x_n) = \bar{x}$, so ist die Risikofunktion

$$R(\theta, d) = E((\bar{X} - \theta)^2) = \frac{\sigma^2}{n};$$

dabei ist σ^2 die Varianz von X; vgl. Satz 59.1 und Formel (34.4).

Andere sinnvolle Verlustfunktionen wären z.B.

$$L(\theta, a) = |\theta - a|, \quad L(\theta, a) = (\theta - a)^4$$

usw.

Beispiel 120.3. Es sei X eine Zufallsvariable mit bekanntem Mittelwert μ und unbekannter Varianz σ^2, und wir stellen uns die Aufgabe, Schätzwerte für σ^2 zu gewinnen. σ^2 sei irgendwelcher positiver Werte fähig. Dann ist Ω die positive Halbgerade, und wenn jeder Schätzwert von σ^2 eine Aktion darstellt, ist A ebenfalls die positive Halbgerade. Dies ist so ähnlich wie im vorigen Beispiel.

Als Entscheidungsfunktion können wir z.B.

$$d(x_1, \ldots, x_n) = s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

verwenden und als Verlustfunktion z.B.

$$L(\sigma^2, d) = (d - \sigma^2)^2 \tag{120.2}$$

mit dem soeben angegebenen d. Etwas allgemeiner könnten wir

$$L(\sigma^2, d) = k(\sigma)(d - \sigma^2)^2 \tag{120.3}$$

ansetzen und $k(\sigma)$ so bestimmen, daß sich Verlustwerte ergeben, die aus der Sicht des betreffenden Problems vernünftig erscheinen. Zugunsten des Ansatzes (120.3) spricht vor allem die Tatsache, daß sich viele Funktionen von σ, die für $\sigma = \sigma$ ein Minimum haben, in der Umgebung von σ durch quadratische Funktionen brauchbar annähern lassen.

Beispiel 120.4. Die in Kap. 13 behandelte Intervallschätzung von Parametern θ (Angabe von Konfidenzintervallen) läßt sich ebenfalls als ein statistisches Entscheidungsproblem auffassen. Wir können dann

$$a = d(x_1, \ldots, x_n) = i(x_1, \ldots, x_n)$$

setzen; dabei ist $i(x_1, \ldots, x_n)$ eine Intervallfunktion, die jeder Stichprobe $x_1, \ldots, x_n$ ein Intervall zuordnet. Eine einfache Verlustfunktion ist

$$L(\theta, a) = L(\theta, i) = \begin{cases} 0 \text{ für } \theta \text{ im Intervall } i, \\ 1 \text{ für } \theta \text{ nicht in } i. \end{cases}$$

Dann ist das Risiko

$$R(\theta, d) = R(\theta, i) = \gamma \cdot 0 + (1 - \gamma) \cdot 1 = 1 - \gamma$$

(γ = Konfidenzzahl), also gleich der Wahrscheinlichkeit, ein Konfidenzintervall zu erhalten, das den unbekannten wahren Parameterwert nicht überdeckt.

Aufgaben zu Abschnitt 120

120.1 Es sei x_1, x_2 eine Stichprobe aus einer Grundgesamtheit mit dem unbekannten Mittelwert μ und der Varianz 1, wobei μ irgendeinen Wert haben kann. Für die Entscheidungsfunktion

$$d(x_1, x_2) = \bar{x} = \frac{1}{2}(x_1 + x_2)$$

und die quadratische Verlustfunktion

$$L(\mu, d) = (\mu - d)^2 = (\mu - \bar{x})^2$$

bestimme man den Wert der Risikofunktion $R(\mu, d)$.

120.2 In Aufgabe 120.1 ersetze man d durch

$$d_1(x_1, x_2) = \frac{1}{4}x_1 + \frac{3}{4}x_2,$$

bestimme den Wert der Risikofunktion und vergleiche mit Aufgabe 120.1.

120.3 In Aufgabe 120.1 ersetze man d durch

$$d_2(x_1, x_2) = k x_1 + (1 - k) x_2$$

und bestimme den Wert k, für den die Risikofunktion einen minimalen Wert hat.

120.4 Es sei x_1, x_2, x_3 eine Stichprobe aus einer Grundgesamtheit mit unbekanntem Mittelwert μ und bekannter Varianz σ^2, wobei μ irgendeinen Zahlenwert haben kann. Man bestimme die Risikofunktion $R(\mu, d)$ im Falle der Entscheidungsfunktion

$$d(x_1, x_2, x_3) = k_1 x_1 + k_2 x_2 + (1 - k_1 - k_2) x_3$$

und der Verlustfunktion

$$L(\mu, d) = (\mu - d)^2.$$

Für welche Werte von k_1 und k_2 wird $R(\mu, d)$ minimal?

120.5 Man verallgemeinere Aufgabe 120.4 für eine Stichprobe $x_1, \ldots, x_n$ unter Benutzung von

$$d(x_1, \ldots, x_n) = k_1 x_1 + \cdots + k_n x_n \qquad (k_1 + \cdots + k_n = 1)$$

und L wie zuvor.

120.6 Es sei θ die Wahrscheinlichkeit von „Kopf" beim Wurf einer gegebenen Münze, wobei θ irgendeinen unbekannten Wert zwischen 0 und 1 einschließlich habe. Man zeige: Bei der Schätzung von θ vermöge eines einzelnen Wurfs ist A das Intervall $0 \leq a \leq 1$, und die Menge D aller (nichtrandomisierten) Entscheidungsfunktionen besteht aus allen $d(x)$ mit $d(0) = a_0$, $d(1) = a_1$, $0 \leq a_0 \leq 1$, $0 \leq a_1 \leq 1$, wobei x ein beobachteter Wert von X = *Anzahl „Köpfe" beim einzelnen Wurf* ist. Die Verlustfunktion sei $L(\theta, a) = (\theta - a)^2$. Man bestimme die Risikofunktion.

120.7 Man zeige, daß $d(X_1, \ldots, X_n)$ mit $d(x_1, \ldots, x_n)$ wie in Aufgabe 120.5 eine erwartungstreue Schätzfunktion für den Mittelwert μ ist.

120.8 Welche Verlustwerte muß man in Beispiel 120.1 wählen, damit die Risiken mit den in Abschn. 78 angegebenen übereinstimmen? Warum ist der Ausdruck „Risiko" hier besser gerechtfertigt als in Abschn. 78?

120.9 Man beschreibe den Unterschied der drei Verlustfunktionen in Beispiel 120.2 anschaulich. Alle drei Funktionen bewerten gleiche Abweichungen nach oben und unten gleichstark. Man überlege sich Anwendungsbeispiele, bei denen dies nicht sachgerecht wäre.

121 Allgemeine Bemerkungen zur Entscheidungstheorie

Wir machen nun einige allgemeine Bemerkungen zur entscheidungstheoretischen Behandlung statistischer Probleme und zu den Begriffen, die wir dafür bisher eingeführt haben.

Zwischen der Spieltheorie im eigentlichen Sinne und der Entscheidungstheorie besteht ein grundlegender Unterschied. Dieser kommt daher, daß in der Entscheidungstheorie einer der beiden Spieler — nämlich die Natur — nicht darauf bedacht ist, den anderen Spieler möglichst weitgehend zu schädigen (um möglichst viel zu gewinnen), sondern in einer „neutralen", d. h. nicht auf den Verlust des Statistikers ausgehenden Weise vorgeht (seinen Zustand θ wählt). Dies beeinflußt die Bewertung der Entscheidungen für den Statistiker und erweitert die Klasse der „vernünftigen" Entscheidungsfunktionen gegenüber der Situation in der eigentlichen Theorie der Spiele zweier intelligenter und auf ihren Vorteil bedachter Spieler. Zur Verdeutlichung ein Beispiel:

Bei einem gewissen Spiel habe die Verlustfunktion $L(\theta, a)$ die folgenden Werte:

	θ_1	θ_2
a_1	3	1
a_2	5	—2

Dies ist der Verlust des Spielers S_2, der a wählt, also der Gewinn des Spielers S_1, der θ wählt. In der Spieltheorie wird S_1 als intelligent vorausgesetzt. Vernünftig ist für ihn dann nur die Wahl von θ_1, weil er dabei mehr gewinnt als bei der Wahl von θ_2. Also sollte Spieler S_2 die Wahl a_1 (Verlust 3) treffen und nicht etwa a_2 (Verlust 5) wählen.

Die Sache ändert sich, wenn Spieler S_1 die Natur ist, die keine besonderen Absichten hat und demnach auch θ_2 wählen könnte. Dann ist a_1 nicht mehr die einzige vernünftige Aktion für den Spieler S_2 (den Statistiker): Wählt die Natur nämlich θ_2, so ist a_2 die bessere Aktion für S_2, weil er dann einen Gewinn (2) erzielt.

Welcher grundsätzliche Unterschied besteht zwischen unserem klassischen Vorgehen in den bisherigen Kapiteln und der entscheidungstheoretischen Behandlung statistischer Probleme? Hierzu läßt sich folgendes sagen:

Klassisch zerfällt das Aktionsproblem in 2 getrennte Teile; der erste Teil ist z.B. die Bestimmung eines Konfidenzintervalles und der zweite Teil dann die Aktion, die sich daraus ergibt.

Dagegen macht man in der Entscheidungstheorie den formalen Versuch, die abschließenden Aktionen und die wirtschaftlichen und psychologischen Verluste bei falschen Aktionen direkt mit in das entscheidungstheoretische Modell hineinzunehmen.

Klassisch tut man letzteres nur indirekt, vermöge von Begriffen wie Konfidenzzahlen (Kap. 13), Längen von Konfidenzintervallen, Signifikanzzahlen (Abschn. 78) und dergleichen. Dies braucht übrigens kein Nachteil zu sein, sondern bringt oft Vereinfachungen. Zum Beispiel kann ein Konfidenzintervall möglicherweise dazu führen, daß man gewisse Aktionen, die zunächst mit in Betracht kamen, außer acht lassen kann und sich damit die heikle und zeitraubende Bestimmung zugehöriger Verlustwerte erspart.

Dazu ist folgendes zu sagen: Das Festlegen der Verluste ist kein mathematisches Problem (so wenig wie das Festlegen von γ; vgl. S. 183). Sondern es muß vom Standpunkt der Anwendung her gelöst werden, indem sich der Statistiker in das Anwendungsgebiet hineinzudenken versucht und zusammen mit dem Anwender eine vernünftige Lösung findet. Das erfordert oft Zeit und Geduld. Und kann manchmal eine Hauptschwierigkeit bei der Anwendung entscheidungstheoretischer Methoden sein:

Wie soll man beispielsweise bei einer Forschungsarbeit die Verluste realistisch veranschlagen, die sich aus dem fälschlichen Verwerfen oder Annehmen einer Arbeitshypothese ergeben? Oder die Verluste, die zu erwarten sind, wenn man ein schädliches Medikament fälschlicherweise als gut ansieht und auf den Markt bringt? Oder die Verluste beim Parameterschätzen, wenn die Schätzwerte unter Umständen ganz verschiedenen Zwecken dienen sollen, die der Statistiker vielleicht nicht einmal alle kennt? Und so weiter.

Diese Beispiele illustrieren Grenzen der Entscheidungstheorie oder zum mindesten praktische Schwierigkeiten. Dagegen ist positiv zu sagen, daß man bei vielen praktischen Problemen geeignete Verlustfunktionen ohne zu große Mühe finden kann. Kleine Ungenauigkeiten in den Verlustwerten fallen dabei übrigens kaum ins Gewicht, denn die Erfahrung lehrt, daß sehr viele Entscheidungsverfahren gegenüber kleinen Abweichungen nicht empfindlich sind. Weiterhin liefert die Spieltheorie Methoden, die helfen, das Festlegen von Verlustwerten zu systematisieren und zu rationalisieren; hierauf gehen wir im nächsten Abschnitt ein.

Schließlich sollten wir noch etwas zu der Frage sagen, wie sich die Verwendung der Risikofunktion motivieren läßt. Hat man einmal eine Verlustfunktion eingeführt, so kann der Statistiker die Wirkung einer Entscheidungsfunktion $d(x)$ bei den verschiedenen Zuständen θ der Natur feststellen. Leider hängt der Verlust, den er tatsächlich erfährt, nicht nur von θ und der speziell gewählten Funktion d ab, sondern auch von dem Wert der Zufallsvariablen X (oder $X_1, \ldots, X_n$), die zu beobachten ist. Dies legt es nahe, den zu erwartenden oder durchschnittlichen Verlust $R(\theta, d)$ einzuführen und zu untersuchen, der nur von θ und d abhängt.

122 Verlust und Nutzen

Die bisherigen Überlegungen zeigen, daß man bei einem Entscheidungsproblem nun zwei Hauptaufgaben zu bewältigen hat, nämlich erstens die Angabe einer brauchbaren Verlustfunktion und zweitens die Wahl einer geeigneten Entscheidungsfunktion. Kriterien für die zweite Aufgabe gewinnen wir später (in den nächsten beiden Abschnitten); dies ist ein mathematisches Problem, bei der wir die Verlustfunktion brauchen. Wir sagen nochmals: Ob eine Verlustfunktion vernünftig ist, ist keine mathematische Frage, sondern eine, die von der Sache her beantwortet werden muß. Praktisch bedeutet dies, der Statistiker versucht, sich in das Entscheidungsproblem von der Seite der Anwendung her — Wirtschaft, Technik, Biologie oder was es immer sei — hineinzudenken und im Gespräch mit dem Anwender (Wirtschaftler, Ingenieur, Arzt o. dgl.) eine problemgerechte Verlustfunktion festzulegen, die der Wirklichkeit einigermaßen sinnvoll Rechnung trägt.

Obwohl hier also ein Problem der Anwendung und nicht der Mathematik vorliegt (ähnlich wie bei der Wahl einer Signifikanzzahl oder Konfidenzzahl im klassischen Fall), kann man doch mathematische Hilfsmittel der Spieltheorie verwenden, um die Aufgabe, Verlustwerte anzugeben, zu vereinfachen. In diese Möglichkeit und zugehörige Begriffe wollen wir jetzt einführen.

Wir verwenden dazu die **Nutzenfunktion** $u(\theta, a)$ (*utility function* im Englischen), die man in der Spieltheorie neben der Verlustfunktion benutzt. Für unsere Zwecke genügt es, einfach

$$u(\theta, a) = -L(\theta, a)$$

zu setzen. $u(\theta, a)$ ist also der Gewinn für den Statistiker, wenn er die Aktion a wählt und die Natur sich im Zustand θ befindet. Wir geben im vorliegenden Abschnitt einige allgemeine Eigenschaften dieser Funktion an, die in der Entscheidungstheorie für den genannten Zweck von Interesse sind. Dazu definieren wir zuerst einige weitere Bezeichnungen und Begriffe.

Die Wahrscheinlichkeitsverteilung einer Zufallsvariablen, die Werte in einer Menge S annehmen kann, wird auch *Wahrscheinlichkeitsverteilung auf* S genannt. Der Einfachheit halber bezeichnen wir eine Wahrscheinlichkeitsverteilung im folgenden durch die zugehörige Wahrscheinlichkeitsfunktion bzw. Dichte. Dann bezeichnen also p_1 und p_2 Verteilungen mit der Wahrscheinlichkeitsfunktion (oder Dichte) p_1 bzw. p_2, und $\varkappa p_1 + \lambda p_2$ ($\varkappa \geqq 0$, $\lambda \geqq 0$, $\varkappa + \lambda = 1$) bezeichnet die Verteilung mit der Wahrscheinlichkeitsfunktion (oder Dichte) $\varkappa p_1 + \lambda p_2$.

Wir betrachten nun ein beliebiges Entscheidungsproblem. Das Resultat einer ganz bestimmten Aktion des Statistikers bei diesem Problem nennen wir ein **Ergebnis** des Entscheidungsproblems. Es sei K die Menge aller Ergebnisse des Problems. Dann nennen wir eine Wahrscheinlichkeitsverteilung auf K einen **Prospekt.**

Warum betrachtet man Wahrscheinlichkeitsverteilungen auf K? Der Grund ist der folgende:

Der einfachste denkbare Fall wäre der, daß ein Entscheidungsproblem mit Sicherheit zu einem Ergebnis r_0 führt. Dieser Fall ist aber uninteressant, da man eine Zufallsvariable X beobachtet, also vermöge $d(X)$ Wahrscheinlichkeitsverteilungen in das Problem einführt.

Noch einen weiteren Grund gibt es: Anstatt eine feste Entscheidungsfunktion d zu wählen, kann der Statistiker eine Wahrscheinlichkeitsverteilung auf der Menge D aller Entscheidungsfunktionen des betreffenden Problems zugrundelegen. Dies bedeutet, der Statistiker macht ein *Hilfsexperiment* zwecks Zufallsauswahl einer Entscheidungsfunktion d aus D; bei diesem Experiment beobachtet er eine Zufallsvariable, die gerade die genannte Wahrscheinlichkeitsverteilung hat.

Eine derartige „Zufallsmischung von Entscheidungsfunktionen" heißt eine **randomisierte Entscheidungsfunktion** (oder **gemischte**

Strategie in der Sprache der Spieltheorie), im Gegensatz zu einer *nichtrandomisierten Entscheidungsfunktion* (oder **reinen Strategie**), wie wir sie in Abschn. 119 eingeführt haben.

Dem Praktiker kommt diese Randomisierung vielleicht seltsam vor, und er mag sich unbehaglich fühlen. Eine Methode, bei der man nichts verliert, wenn man auf gemischte Strategien verzichtet, besprechen wir in Abschn. 124.

Es ist klar, daß der genannte Zufallsprozeß (das Hilfsexperiment) Wahrscheinlichkeitsverteilungen auf der Menge K der Ergebnisse des Entscheidungsproblems zur Folge hat.

Ein Prospekt des Statistikers hängt also nun ab von dem Zustand θ der Natur und der Aktion $a = \delta(x)$, wobei δ eine (nichtrandomisierte oder randomisierte) Entscheidungsfunktion bezeichnet. Damit wird die Nutzenfunktion

$$u(\theta, a) = -L(\theta, a),$$

die den Gewinn des Statistikers bedeutet, eine Funktion der Prospekte; wir schreiben

$$u = u(p).$$

u existiert, wenn wir einige einfache und natürliche Annahmen [(P 1)—(P 4); s. unten] machen.

Wir legen wieder ein beliebiges Entscheidungsproblem fest zugrunde. Wir nehmen an, der Statistiker ist sich über seine Ziele und Absichten so weit im klaren, daß er von irgend zwei Ergebnissen stets weiß, welches er bevorzugt (bzw. ob sie ihm beide gleichwertig sind). Dies ist die erste Annahme, die wir formal so fassen können:

(P1) Bei Vorgabe von irgend zwei Prospekten p_1 und p_2 ist der Statistiker fähig zu entscheiden, ob er p_1 dem p_2 vorzieht oder ob ihm p_1 mit p_2 gleichwertig ist oder ob er p_2 dem p_1 vorzieht; in Formeln:

$p_1 \succ p_2$ (p_1 ist dem p_2 vorzuziehen)
$p_1 \sim p_2$ (Gleichwertigkeit)
$p_2 \succ p_1$ (p_2 ist dem p_1 vorzuziehen)

Nun folgt eine vereinfachte Sprech- und Schreibweise: Ist p_1 dem Statitiker gleichwertig mit p_2 oder zieht er p_1 dem p_2 sogar vor, so sagen wir kurz, daß ihm p_1 *wenigstens gleichwertig* mit p_2 ist, und schreiben kurz

$$p_1 \succeq p_2,$$

wenn also $p_1 \sim p_2$ oder $p_1 \succ p_2$ gilt.

Wir nehmen weiterhin an:

(P2) Ist dem Statistiker p_1 wenigstens gleichwertig mit p_2 und p_2 wenigstens gleichwertig mit p_3, dann ist ihm p_1 wenigstens gleichwertig mit p_3. In Formeln:

$$\text{Aus } p_1 \succeq p_2 \text{ und } p_2 \succeq p_3 \quad \text{folgt} \quad p_1 \succeq p_3.$$

(P1) und (P2) definieren eine lineare Ordnung auf den Prospekten; diese heißt ein **Vorzugsschema.**

Einen Prospekt p, den man durch ein Zufallsexperiment aus einer Menge von Prospekten auswählt, nennen wir kurz eine **Zufallsmischung** *dieser Prospekte.* Über Zufallsmischungen von Prospekten machen wir zwei ebenfalls naheliegende Annahmen:

(P3) Zieht der Statistiker p_1 einem Prospekt p_2 vor und p_2 einem Prospekt p_3, so gibt es eine Zufallsmischung von p_1 und p_3, die der Statistiker dem p_2 vorzieht, und eine andere Zufallsmischung von p_1 und p_3 derart, daß er p_2 dieser Mischung vorzieht; in Formeln:

Gilt $p_1 \succ p_2 \succ p_3$, so gibt es Zahlen k und l, $0 < k < 1$, $0 < l < 1$ derart, daß gilt:

$$k p_1 + (1-k) p_3 \succ p_2, \quad p_2 \succ l p_1 + (1-l) p_3.$$

(P4) Zieht der Statistiker p_1 einem Prospekt p_2 vor und ist p_3 ein beliebiger anderer Prospekt, so zieht der Statistiker auch eine Zufallsmischung von p_1 und p_3 derselben Mischung von p_2 und p_3 vor; in Formeln:

Ist $p_1 \succ p_2$, dann gilt für jeden Prospekt p_3 und jede Zahl k, $0 < k < 1$,

$$k p_1 + (1-k) p_3 \succ k p_2 + (1-k) p_3.$$

Man kann beweisen, daß diese Eigenschaften der Prospekte hinreichen für die Existenz einer Nutzenfunktion $u(p)$, die jedem Prospekt p auf K eine Zahl $u(p)$, den **Nutzen** des Prospektes p (für den Statistiker), zuordnet und die folgenden Eigenschaften hat:

(U1) Es gilt $u(p_1) > u(p_2)$ dann und nur dann, wenn der Statistiker p_1 dem p_2 vorzieht.

(U2) Ist $p = k p_1 + (1-k) p_2$, so gilt

$$u(p) = k u(p_1) + (1-k) u(p_2).$$

Übrigens garantieren die Annahmen (P1)—(P4) nicht nur die Existenz, sondern sogar die Eindeutigkeit (bis auf lineare Transformationen) einer solchen Funktion u für ein gegebenes Vorzugsschema.

Die Eigenschaft (U2) ist sehr wichtig, weil sie es ermöglicht, aus bekannten Werten der Nutzenfunktion zahlreiche weitere Werte zu berechnen. Überhaupt helfen Überlegungen wie die des vorliegenden Abschnittes, gedankliche Arbeit im Zusammenhang mit Vorzugsschemata durch einfache Rechnungen zu ersetzen, die auch weniger zeitraubend sind.

123 Minimax-Prinzip

Wir wenden uns nun dem Problem zu, wie man Entscheidungsfunktionen auswählt. Als erste Methode betrachten wir das sog. Minimax-Prinzip.

Dabei verwenden wir die Verlustfunktion und die zugehörige Risikofunktion $R(\theta, d)$, die in Abschn. 119 eingeführt wurde. Für jede Entscheidungsfunktion bei dem betreffenden Entscheidungsproblem kümmern wir uns um das maximal mögliche Risiko (also um den schlimmsten denkbaren Fall für den Statistiker) und wählen dann diejenige Entscheidungsfunktion aus, für die dieses maximale Risiko möglichst klein (also minimal) ausfällt. Hat man nur endlich viele Entscheidungsfunktionen zur Auswahl und hat die Natur nur endlich viele Zustände, so gibt es wenigstens eine solche Entscheidungsfunktion. Diese heißt **Minimax-Entscheidungsfunktion.** Und das genannte Auswahlprinzip heißt **Minimax-Prinzip.** Das Risiko, das zu einer Minimax-Entscheidungsfunktion gehört, läßt sich im Falle endlich vieler Funktionen und Zustände in der Form

$$\min_d [\max_\theta R(\theta, d)]$$

schreiben. In Worten bedeutet dies: Für jede Funktion d stelle man zuerst den maximalen Wert von $R(\theta, d)$ fest. (Dieser Wert entspricht jeweils einem gewissen Zustand θ). Dann bestimme man die kleinste dieser so erhaltenen Zahlen. Eine zu dieser kleinsten Zahl gehörige Funktion d ist dann eine Minimax-Entscheidungsfunktion des betreffenden Entscheidungsproblems.

Hat man bei einem Problem unendlich viele Entscheidungsfunktionen zur Auswahl, kompliziert sich die Lage theoretisch, und es braucht dann keine Minimax-Entscheidungsfunktion zu existieren.

Praktisch ist dies aber ziemlich unwesentlich, weil dann eine „fast so gute" Entscheidungsfunktion existiert. Für eine präzise Beschreibung des Sachverhaltes und weitere Einzelheiten siehe z. B. [1] in Anhang 3.

Praktische Nachteile des Minimax-Prinzips liegen in einer anderen Richtung: Erstens ist der Rechenaufwand oft ziemlich erheblich (, selbst wenn man ihn vermindert vermöge gewisser Methoden der Spieltheorie, auf die wir hier nicht eingehen können). Zweitens ist das Prinzip im allgemeinen zu pessimistisch, wie wir gegen Ende dieses Abschnittes sehen werden. Zuvor wollen wir das Verfahren an einem einfachen Beispiel erläutern.

Beispiel 123.1. Wir betrachten ein Entscheidungsproblem mit 2 Zuständen der Natur und 2 Aktionen des Statistikers. Es sei $\Omega = \{\theta_1, \theta_2\}$ mit $\theta_1 = 0{,}25$, $\theta_2 = 0{,}5$ und $A = \{a_1, a_2\}$. Die Verlustfunktion habe die folgenden Werte:

	θ_1	θ_2
a_1	0	2
a_2	3	1

Im zugehörigen Zufallsexperiment beobachte der Statistiker eine diskrete Zufallsvariable X, die nur die Werte 0 und 1 annimmt, und zwar mit den folgenden Wahrscheinlichkeiten:

$$f(0) = 0{,}75, \quad f(1) = 0{,}25 \quad \text{wenn} \quad \theta = \theta_1 = 0{,}25 \text{ ist,}$$
$$f(0) = 0{,}5, \quad f(1) = 0{,}5 \quad \text{wenn} \quad \theta = \theta_2 = 0{,}5 \quad \text{ist.}$$

Man bestimme die Minimax-Entscheidungsfunktion in der Menge D aller möglichen nichtrandomisierten Entscheidungsfunktionen.

D besteht aus 4 Funktionen $d(x)$, die die folgenden Werte haben:

Entscheidungsfunktion	Wert an der Stelle $x = 0$	$x = 1$
d_1	a_1	a_1
d_2	a_1	a_2
d_3	a_2	a_1
d_4	a_2	a_2

Man beachte, daß sich d_1 und d_4 gar nicht um den Ausgang des Zufallsexperimentes, bei dem X beobachtet wird, kümmern. Wir berechnen nun die Werte der Risikofunktion, zuerst für d_1,

$$R(\theta_1, d_1) = 0{,}75 \cdot 0 + 0{,}25 \cdot 0 = 0$$
$$R(\theta_2, d_1) = 0{,}50 \cdot 2 + 0{,}50 \cdot 2 = 2$$

dann für d_2,

$$R(\theta_1, d_2) = 0{,}75 \cdot 0 + 0{,}25 \cdot 3 = 0{,}75$$
$$R(\theta_2, d_2) = 0{,}50 \cdot 2 + 0{,}50 \cdot 1 = 1{,}50$$

dann für d_3,

$$R(\theta_1, d_3) = 0{,}75 \cdot 3 + 0{,}25 \cdot 0 = 2{,}25$$
$$R(\theta_2, d_3) = 0{,}50 \cdot 1 + 0{,}50 \cdot 2 = 1{,}50$$

und schließlich für d_4,

$$R(\theta_1, d_4) = 0{,}75 \cdot 3 + 0{,}25 \cdot 3 = 3$$
$$R(\theta_2, d_4) = 0{,}50 \cdot 1 + 0{,}50 \cdot 1 = 1$$

Wie wir sehen, haben die maximalen Risiken die folgenden Werte:

Entscheidungsfunktion	d_1	d_2	d_3	d_4
$\max R(\theta, d_j)$	2	1,5	2,25	3

Das Minimum ist 1,5. Also ist d_2 die Minimax-Entscheidungsfunktion.

Gegen das Minimax-Prinzip gibt es ernsthafte Einwände, und der Grund ist der folgende:

S_1 und S_2 mögen gegeneinander spielen, S_1 wähle θ und S_2 wähle d. Ist S_1 intelligent, wählt er θ so, daß S_2 möglichst viel verliert. Und S_2 wählt eine Strategie, die ein d mit minimalem Maximalrisiko ergibt. Dies ist genau das Minimax-Prinzip.

Für ein statistisches Entscheidungsproblem ist aber diese Haltung des Statistikers zu pessimistisch, denn die Natur trifft ihre Wahl ohne feindliche Absichten gegenüber dem Statistiker. In der Tat besagt ja das Minimax-Prinzip, man solle so vorgehen, als ob die Natur auf ihren Zuständen eine für den Statistiker möglichst ungünstige Wahrscheinlichkeitsverteilung wählt. Das ist aber viel zu konservativ. Denn aus welchem Grunde sollte die Natur ausgerechnet eine derartige Verteilung wählen? Das Minimax-Prinzip kann unter Umständen sogar zu völlig unsinnigen Entscheidungen führen, wie wir uns an einem einfachen Beispiel klarmachen wollen:

Beispiel 123.2. Bei einem Produktionsprozeß von Dichtungen sei die Wahrscheinlichkeit θ, eine defekte Dichtung zu produzieren, unbekannt, aber konstant. Für eine Menge von 1000 solcher Dichtungen betrachte man die beiden Aktionen:

a_1: *Diese* 1000 *Dichtungen werden nicht verkauft, sondern weggeworfen.*

a_2: *Die Dichtungen werden für DM* 0,20/*Stück verkauft, und als Garantie erhält der Käufer für jede fehlerhafte Dichtung das Doppelte dieses Betrages zurück.*

Wählen wir a_1, ist der Verlust 0. Also ist $L(\theta, a_1) = 0$. Wir betrachten a_2. Da unter den 1000 Dichtungen 1000 θ fehlerhafte zu erwarten sind, ist ein Verlust $1000\,\theta \cdot 0{,}4 = 400\,\theta$ DM zu erwarten, und außerdem haben wir durch den Verkauf einen Gewinn (negativen Verlust) von $1000 \cdot 0{,}2 = 200$ DM. Also wird $L(\theta, a_2) = 400\,\theta - 200$ DM. Wählen wir a_1 mit einer Wahrscheinlichkeit

k und a_2 mit der Wahrscheinlichkeit $1 - k$, so ist dies eine randomisierte Entscheidungsfunktion d, die noch von k abhängt. Zu d gehört die Risikofunktion

$$R(\theta, d) = k \cdot 0 + (1 - k)(400\,\theta - 200).$$

Bei festem k ist diese maximal für $\theta = 1$, wie wir sehen, und dann hat man den Wert

$$\max R(\theta, d) = (1 - k) \cdot 200.$$

Dieser Ausdruck hängt von k ab und wird minimal für $k = 1$. Auf Grund des Minimax-Prinzipes bedeutet dies: Besteht eine — wenn auch geringe — Möglichkeit, daß $\theta > 0{,}5$ sein könnte, so sollte sich der Statistiker für Aktion a_1 (Wegwerfen der Dichtungen) entscheiden.

Dieses pessimistische Ergebnis folgt daraus, daß ein feindlicher Gegenspieler des Statistikers, wenn er θ frei wählen könnte, dieses so groß wie möglich wählen würde, also $\theta = 1$.

Dieses Ergebnis des Minimax-Prinzipes ändert sich nicht einmal dann, wenn wir eine Stichprobenentnahme durchführen. Zum Beispiel könnten wir eine Stichprobe vom Umfange n mit Zurücklegen ziehen und uns für Aktion a_1 entscheiden, wenn die Stichprobe wenigstens eine defekte Dichtung enthält, und für Aktion a_2 nur dann, wenn wir lauter einwandfreie Dichtungen ziehen. Aktion a_2 würde dann mit der Wahrscheinlichkeit $(1 - \theta)^n$ gewählt werden, also a_1 mit der Wahrscheinlichkeit $1 - (1 - \theta)^n$. Der zu erwartende Verlust wäre dann

$$[1 - (1 - \theta)^n] \cdot 0 + (1 - \theta)^n (400\,\theta - 200)$$

und ist positiv für jedes $\theta > 0{,}5$. Also sollten wir die Dichtungen wegwerfen, sogar, wenn wir keine defekte Dichtungen erhalten, sofern nur ein Wert $\theta > 0{,}5$ nicht ganz und gar auszuschließen ist.

Rückblickend mache man sich die prinzipielle Lage nochmals klar: Die Minimax-Entscheidungsfunktion ist optimal bezüglich eines Kriteriums. Das Kriterium benutzt die Risikofunktion; es beurteilt die Entscheidungsfunktionen nach der Größe des Maximums der Risikofunktion. Die Risikofunktion hängt ihrerseits ab von der Verlustfunktion (s. Abschn. 119). Deren Gewinnung ist eine Aufgabe des Anwenders (, bei der die Mathematik nur Hilfestellung leisten kann; vgl. Abschn. 122). Unsere Überlegung zeigt: Ehe der Statistiker etwas über die Güte von Entscheidungsfunktionen (statistischen Verfahren) sagen konnte, mußte der Anwender Verlustwerte angeben. Ohne Verlustfunktion kein Gütekriterium. Und dieser Gütemaßstab hängt ab von den Annahmen, die der Anwender macht, wenn er Verlustwerte festsetzt.

Genau dieselbe Situation liegt vor bei anderen Kriterien, zum Beispiel beim Bayes-Kriterium im nächsten Abschnitt.

Aufgaben zu Abschnitt 123

123.1 Man bestimme die Minimax-Entscheidungsfunktion in der Menge $D = \{d_1, \ldots, d_4\}$ in Beispiel 123.1, wenn die Verlustfunktion $L(\theta, a)$ die Werte

	θ_1	θ_2
a_1	0	1
a_2	1	0

hat und die übrigen Annahmen wie in Beispiel 123.1 sind.

123.2 Es sei $\Omega = \{\theta_1, \theta_2, \theta_3\}$ mit $\theta_1 = 0{,}2$, $\theta_2 = 0{,}5$, $\theta_3 = 0{,}8$ und $A = \{a_1, a_2\}$, und es habe die Verlustfunktion $L(\theta, a)$ die folgenden Werte:

	θ_1	θ_2	θ_3
a_1	0	1	2
a_2	1	0	1

X sei eine diskrete Zufallsvariable mit 0 und 1 als möglichen Werten, und die zugehörige Wahrscheinlichkeitsfunktion $f(x \mid \theta)$ habe die folgenden Werte:

	$x = 0$	$x = 1$
θ_1	0,1	0,9
θ_2	0,5	0,5
θ_3	0,9	0,1

Man bestimme die Minimax-Entscheidungsfunktion in der in Beispiel 123.1 gegebenen Menge $D = \{d_1, d_2, d_3, d_4\}$.

123.3 Eine Entscheidungsfunktion d^* heißt **besser** als eine Entscheidungsfunktion d, wenn gilt

$$R(\theta, d^*) \leq R(\theta, d) \quad \text{für alle } \theta \text{ in } \Omega$$

und

$$R(\theta, d^*) < R(\theta, d) \quad \text{für wenigstens ein } \theta \text{ in } \Omega.$$

Man wende diesen Begriff auf die Entscheidungsfunktionen in Beispiel 123.1 an.

123.4 Eine Entscheidungsfunktion d in D heißt **zulässig,** wenn es in D keine Entscheidungsfunktion gibt, die besser als d ist. d heißt *unzulässig,* wenn es in D eine Entscheidungsfunktion gibt, die besser als d ist. Man bestimme die zulässigen Entscheidungsfunktionen in der Menge D in Beispiel 123.1.

123.5 Eine Klasse C von Entscheidungsfunktionen in D heißt **vollständig** wenn gilt: Zu jeder Entscheidungsfunktion in D, die nicht zu C gehört, gibt es in C eine Entscheidungsfunktion, die besser ist. Man zeige, daß eine zulässige Entscheidungsfunktion d in jeder vollständigen Klasse enthalten ist.

123.6 Warum ist der Begriff der vollständigen Klasse von Wichtigkeit?

123.7 Es sei $\Omega = \{1, 2\}$, $A = \{1, 2\}$, und $L(\theta, a)$ habe die folgenden Werte:

	$\theta_1 = 1$	$\theta_2 = 2$
$a_1 = 1$	-1	2
$a_2 = 2$	2	-3

Die Menge D bestehe aus den Entscheidungsfunktionen

d_k: *Wähle a_1 mit der Wahrscheinlichkeit k und a_2 mit der Wahrscheinlichkeit $1-k$.*

Dabei sei k eine Zahl zwischen 0 und 1 einschließlich. Man bestimme die Minimax-Entscheidungsfunktion in D und berechne das zugehörige Risiko. Man stelle das maximale Risiko $R(\theta, d_k)$ als Funktion von k graphisch dar.

123.8 Man modifiziere Aufgabe 123.7, indem man eine andere Verlustfunktion nimmt, mit den Werten

	$\theta_1 = 1$	$\theta_2 = 2$
$a_1 = 1$	$-c$	1
$a_2 = 2$	1	$-2+c$

Dabei sei c eine Konstante, $-1 \leqq c \leqq 3$. Man bestimme die Minimax-Entscheidungsfunktion in D (D wie zuvor).

123.9 Welche der Entscheidungsfunktionen in Aufgabe 119.7 ist eine Minimax-Entscheidungsfunktion?

123.10 Welche der Entscheidungsfunktionen in Aufgabe 119.4 ist eine Minimax-Entscheidungsfunktion?

123.11 Man zeichne die zu $d_1, \ldots, d_4$ in Beispiel 123.1 gehörigen Risiken als Punkte in der Ebene mit den Koordinaten

$$z_{1j} = R(\theta_1, d_j), \quad z_{2j} = R(\theta_2, d_j), \quad j = 1, \ldots, 4.$$

Es sei δ die randomisierte Entscheidungsfunktion, die durch eine Wahrscheinlichkeitsverteilung $p(d)$ auf $D = \{d_1, \ldots, d_4\}$ gegeben ist. Man zeige, daß das zugehörige Risiko einem Punkt entspricht, der zu dem Viereck mit den Ecken (z_{1j}, z_{2j}), $j = 1, \ldots, 4$ gehört (also auf den Seiten oder innerhalb des Vierecks liegt); vgl. Abb. 123.1.

123.12 In Abb. 123.1 begründe man die Lage der Minimax-Entscheidungsfunktion in der Menge aller Entscheidungsfunktionen, die man durch Randomisierung von $d_1, \ldots, d_4$ erhält.

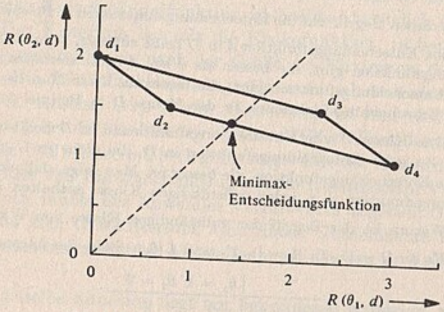


Abb. 123.1. Aufgabe 123.11 und 123.12

124 Bayes-Prinzip

Wir erläutern nun eine weitere Methode zur Wahl von Entscheidungsfunktionen, das sog. Bayes-Prinzip. Der Grundgedanke ist einfach: Benutzt wird die Tatsache, daß es oft möglich sein wird, schon vorher Anhaltspunkte darüber zu gewinnen, wie häufig die Natur die einzelnen Zustände etwa annimmt. Mit anderen Worten: Wir nehmen an, der Statistiker habe schon vorher Information über θ, die er durch eine Wahrscheinlichkeitsverteilung auf dem Parameterraum Ω ausdrücken kann. Es sei $\Omega = \{\theta_1, \ldots, \theta_m\}$, und die genannte Verteilung habe die Wahrscheinlichkeitsfunktion

$$q(\theta) \qquad (\theta = \theta_1, \ldots \theta_m). \tag{124.1}$$

Es sei d eine Entscheidungsfunktion aus der Menge D der Entscheidungsfunktionen des betreffenden Problems, und $R(\theta, d)$ sei die zu d gehörige Risikofunktion. Dann können wir die Werte von R „mitteln" bezüglich der Verteilung $q(\theta)$. Dies ergibt das **Durchschnittsrisiko** oder **Bayes-Risiko**

$$\overline{R}(q, d) = \sum_{j=1}^{m} q(\theta_j) R(\theta_j, d).$$

Anschaulich bedeutet dies, wir berücksichtigen jeden Wert $R(\theta_j, d)$ nach Maßgabe der Wahrscheinlichkeit $q(\theta_j)$.

Eine Entscheidungsfunktion aus D, die $\overline{R}(q, d)$ zum Minimum macht, heißt eine **Bayes-Lösung** oder **Bayes-Entscheidungsfunktion** des betreffenden Entscheidungsproblems *in D bezüglich der Verteilung* $q(\theta)$. Die Verteilung $q(\theta)$ heißt eine **a-priori-Verteilung.**

Offenbar ist ein Vorteil dieser Methode, daß ein Zustand der Natur desto stärker ins Gewicht fällt, je wahrscheinlicher sein Auftreten ist, so daß also praktisch nicht vorkommende Zustände auch praktisch nicht berücksichtigt werden. Aber ein Nachteil ist eine gewisse Subjektivität bei der Wahl einer a-priori-Verteilung für ein gegebenes Problem; verschiedene Bearbeiter eines Problems werden unter Umständen (je nach Sachkenntnis, Erfahrung, Veranlagung usw.) verschiedene solche Verteilungen zugrundelegen und dann möglicherweise zu verschiedenen Ergebnissen gelangen, obgleich sie dieselben experimentellen Daten benutzen.

Wir erläutern nun die Methode an einem einfachen Beispiel.

Beispiel 124.1. Wir greifen zurück auf die vier Entscheidungsfunktionen $d_1, \ldots, d_4$ in Beispiel 123.1. Die Berechnung der Risikofunktion ergab die Werte

$$0,\ 2;\quad 0{,}75,\ 1{,}5;\quad 2{,}25,\ 1{,}5;\quad 3,\ 1.$$

Mögliche Zustände der Natur waren

$$\theta = \theta_1 = 0{,}25 \quad \text{und} \quad \theta = \theta_2 = 0{,}5.$$

Wir nehmen nun eine a-priori-Verteilung an, sagen wir

$$q(\theta_1) = 0{,}6, \quad q(\theta_2) = 0{,}4. \tag{124.2}$$

Als zugehörige Durchschnittsrisiken erhalten wir dann

$$\begin{aligned}
\bar{R}(q, d_1) &= 0{,}6 \cdot 0 + 0{,}4 \cdot 2 = 0{,}8\\
\bar{R}(q, d_2) &= 0{,}6 \cdot 0{,}75 + 0{,}4 \cdot 1{,}5 = 1{,}05\\
\bar{R}(q, d_3) &= 0{,}6 \cdot 2{,}25 + 0{,}4 \cdot 1{,}5 = 1{,}95\\
\bar{R}(q, d_4) &= 0{,}6 \cdot 3 + 0{,}4 \cdot 1 = 2{,}2.
\end{aligned}$$

Wie wir sehen, ist d_1 die Bayes-Entscheidungsfunktion in der Menge $D = \{d_1, \ldots, d_4\}$ bezüglich der gegebenen a-priori-Verteilung.

Im Falle endlich vieler Zustände der Natur können wir das Bayes-Prinzip geometrisch diskutieren, wie folgt. Es sei $\Omega = \{\theta_1, \ldots, \theta_m\}$, wie zuvor. Eine a-priori-Verteilung $q(\theta)$ ist dann durch m nichtnegative Zahlen

$$q_1 = q(\theta_1), \ldots, q_m = q(\theta_m)$$

bestimmt. Hierbei ist $q_j = q(\theta_j)$ die Wahrscheinlichkeit, daß die Natur den Zustand θ_j wählt, und die Summe dieser m Zahlen ist 1 (warum?). Zu einer Entscheidungsfunktion d gehört eine Risikofunktion $R(\theta, d)$ mit den m Werten $R(\theta_1, d), \ldots, R(\theta_m, d)$. Diese Risikofunktion läßt sich nun im m-dimensionalen Raum als Punkt mit den m Koordinaten

$$z_1 = R(\theta_1, d), \quad z_2 = R(\theta_2, d), \ldots, \quad z_m = R(\theta_m, d)$$

darstellen. Es sei Q die Menge aller Punkte im m-dimensionalen Raum, die der Menge D aller (nichtrandomisierten und randomisierten) Entscheidungsfunktionen des betreffenden Problems entsprechen. Zu der gegebenen Verteilung q und einer einzelnen Entscheidungsfunktion d in D gehört das Durchschnittsrisiko

$$\bar{R}(q, d) = \sum_{j=1}^{m} q_j R(\theta_j, d) = \sum_{j=1}^{m} q_j z_j.$$

Also erhalten wir für alle Entscheidungsfunktionen d, für die

$$(124.3) \qquad \bar{R}(q, d) = \sum_{j=1}^{m} q_j z_j = k = \text{konst}$$

gilt, denselben Wert des Durchschnittsrisikos bezüglich der gegebenen a-priori-Verteilung. Formel (124.3) stellt eine Hyperebene H im m-dimensionalen Raum dar, und $(q_1, \ldots, q_m)$ ist der Normalenvektor von H, d. h. H ist senkrecht zu diesem Vektor. Und eine Bayes-Entscheidungsfunktion entspricht dem Minimum der Werte k, für die die durch (124.3) gegebene Hyperebene H wenigstens einen Punkt mit Q gemeinsam hat. (Eine Bayes-Entscheidungsfunktion braucht nicht zu existieren, wenn der Rand von Q nicht zu Q gehört.)

Ist $m = 2$, so wird alles ganz einfach: Abb. 124.1 zeigt die Funktionen $d_1, \ldots, d_4$ in Beispiel 124.1. Jede Hyperebene H, die durch (124.3) dargestellt wird, hat dann die Dimension 1, ist also eine Gerade in der Ebene. Abb. 124.1 zeigt weiter den gemeinsamen Normalenvektor q dieser Hyperebenen (parallelen Geraden) im Falle der a-priori-Verteilung (124.2) und die Gerade H_{min}, für die k in (124.3) in dem eben erklärten Sinne minimal wird. d_1 liegt auf H_{min}, und dies bedeutet, d_1 ist die Bayes-Entscheidungsfunktion in der Menge $D = \{d_1, \ldots, d_4\}$. Weiterhin sehen wir aus der Abbildung, daß kein anderer Punkt des Vierecks mit den Ecken $d_1, \ldots, d_4$ auf H_{min} liegt. Diese Punkte in dem Viereck und auf seinem Rand entsprechen Entscheidungsfunktionen, die man durch Randomisierung aus $d_1, \ldots, d_4$ erhält. Die Situation ist typisch. In der Tat kann man nämlich zeigen: Existiert für ein Problem eine Bayes-Entscheidungsfunktion bezüglich einer a-priori-Verteilung q, so existiert stets auch eine nichtrandomisierte Bayes-Entscheidungsfunktion des Problems bezüglich q. Das ist ein großer rechnerischer Vorteil des Bayes-Prinzips gegenüber dem Minimax-Prinzip, indem man jetzt nur noch die nichtrandomisierten Entscheidungsfunktionen zu untersuchen braucht. Daß dies für das Minimax-Prinzip nicht gilt, wird durch Abb. 123.1 belegt, denn die Minimax-Entscheidungsfunktion ist dort randomisiert.

Interessant ist auch noch folgendes: Wir können eine a-priori-Verteilung wählen, für die (124.3) parallel zur Strecke $d_2 d_4$ in Abb. 124.1 verläuft. Dann erhalten wir unendlich viele zugehörige Bayes-Entscheidungsfunktionen. Dazu gehören auch d_2 und d_4. Und es gehört auch die Minimax-Entscheidungsfunktion dazu (vgl. Abb.

123.1). Dies bedeutet, es gibt eine a-priori-Verteilung, die so ungünstig ist, daß die Minimax-Entscheidungsfunktion eine Bayes-Entscheidungsfunktion bezüglich dieser Verteilung wird.

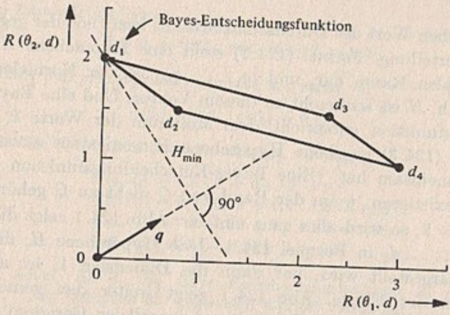


Abb. 124.1. Bayes-Entscheidungsfunktion in Beispiel 124.1

Aufgaben zu Abschnitt 124

124.1 Man bestimme eine Bayes-Entscheidungsfunktion in Beispiel 124.1 für die a-priori-Verteilung

(a) $q(\theta_1) = 0{,}1, \quad q(\theta_2) = 0{,}9$
(b) $q(\theta_1) = 0{,}3, \quad q(\theta_2) = 0{,}7$
(c) $q(\theta_1) = 0{,}5, \quad q(\theta_2) = 0{,}5.$

124.2 In Beispiel 124.1 bestimme man die vier Durchschnittsrisiken, wenn die a-priori-Verteilung durch $q(\theta_1) = q_1$, $q(\theta_2) = 1 - q_1$ mit $0 \leqq q_1 \leqq 1$ gegeben ist. Man stelle diese Risiken als Funktionen von q_1 graphisch dar. Man bestimme die Werte von q_1, für die (a) d_1, (b) d_2, (c) d_4 eine Bayes-Entscheidungsfunktion ist.

124.3 In Aufgabe 124.2 ist d_3 für keinen Wert von q_1 eine Bayes-Entscheidungsfunktion. Kann man das aus den in Beispiel 124.1 gegebenen Werten der Risikofunktion unmittelbar ersehen?

124.4 Es sei $\Omega = \{1, 2\}$ und $A = \{1, 2\}$. Es habe $L(\theta, a)$ die folgenden Werte:

	$\theta_1 = 1$	$\theta_2 = 2$
$a_1 = 1$	-1	2
$a_2 = 2$	2	-3

Die Menge D bestehe aus den Entscheidungsfunktionen

d_k: *Wähle a_1 mit der Wahrscheinlichkeit k und a_2 mit der Wahrscheinlichkeit $1 - k$.*

In D bestimme man eine Bayes-Entscheidungsfunktion bezüglich der a-priori-Verteilung

$$q(\theta_1) = 0{,}4, \quad q(\theta_2) = 0{,}6.$$

124.5 Man löse Aufgabe 124.4 im Falle der a-priori-Verteilung

$$q(\theta_1) = \frac{5}{8}, \quad q(\theta_2) = \frac{3}{8}$$

(die übrigen Angaben unverändert wie oben). Man zeige, daß die Bayes-Entscheidungsfunktion eine Minimax-Entscheidungsfunktion ist, und begründe diese Tatsache.

124.6 In Aufgabe 124.4 bestimme man eine Bayes-Entscheidungsfunktion in D bezüglich der a-priori-Verteilung

$$q(\theta_1) = q_1, \quad q(\theta_2) = 1 - q_1$$

und zeige, daß $q_1 = \frac{5}{8}$ der einzige Wert ist, für den diese Bayes-Entscheidungsfunktion eine Minimax-Entscheidungsfunktion in D ist.

124.7 In Aufgabe 124.6 zeige man, daß das Durchschnittsrisiko der Bayes-Entscheidungsfunktion für jeden Wert $q_1 \neq \frac{5}{8}$ kleiner als das zu der Minimax-Entscheidungsfunktion gehörige Risiko ausfällt.

124.8 Es sei $\Omega = \{\theta_1, \theta_2\}$, $A = \{a_1, a_2\}$, und $L(\theta, a)$ habe die Werte

	θ_1	θ_2
a_1	$-c$	1
a_2	1	$-2+c$

wobei c eine Konstante, $-1 \leq c \leq 3$, ist. In der Menge D der Entscheidungsfunktionen der Form

d_k: *Wähle a_1 mit der Wahrscheinlichkeit k und a_2 mit der Wahrscheinlichkeit* $1 - k$

bestimme man eine Bayes-Entscheidungsfunktion bezüglich der a-priori-Verteilung

$$q(\theta_1) = q_1, \quad q(\theta_2) = 1 - q_1 \quad (q_1 \text{ konstant}, \ 0 \leq q_1 \leq 1).$$

124.9 Man kann beweisen: Existiert eine Bayes-Entscheidungsfunktion bezüglich einer a-priori-Verteilung q, so existiert auch eine nichtrandomisierte Bayes-Entscheidungsfunktion bezüglich q. Man überlege sich, daß die Aufgabe 124.8 ein Beispiel für diesen Sachverhalt liefert.

124.10 In Aufgabe 119.4 bestimme man eine Bayes-Entscheidungsfunktion bezüglich der a-priori-Verteilung $q(0) = 0{,}2$, $q(1) = 0{,}6$, $q(2) = 0{,}2$.

124.11 In Aufgabe 119.4 bestimme man eine Bayes-Entscheidungsfunktion bezüglich der a-priori-Verteilung $q(0) = 0{,}5$, $q(1) = q_1$, $q(2) = 0{,}5 - q_1$, wobei $0 \leq q_1 \leq 0{,}5$ sei.

124.12 In Aufgabe 119.7 bestimme man eine Bayes-Entscheidungsfunktion bezüglich der a-priori-Verteilung $q(0) = 0{,}1$, $q(1) = 0{,}3$, $q(2) = 0{,}6$.

124.13 In Aufgabe 119.7 bestimme man eine Bayes-Entscheidungsfunktion bezüglich der a-priori-Verteilung $q(0) = 0{,}25$, $q(1) = q_1$, $q(2) = 0{,}75 - q_1$, wobei $0 \leqq q_1 \leqq 0{,}75$ gelte.

124.14 In Aufgabe 119.2 bestimme man eine Bayes-Entscheidungsfunktion in der Menge D der nichtrandomisierten Entscheidungsfunktionen d von der Form

$$d(0) = a_0, \quad d(1) = a_1 \quad (a_0, a_1 \text{ beliebig}),$$

wobei die a-priori-Verteilung die Form $q(\theta_1) = 0{,}5$, $q(\theta_2) = 0{,}5$ habe.

124.15 Man bearbeite Aufgabe 124.14 für den Fall, daß $q(\theta_1) = q_1$ und $q(\theta_2) = 1 - q_1$ ist (die übrigen Angaben wie bisher).

124.16 In Beispiel 124.1 bestimme man eine a-priori-Verteilung derart, daß jede beliebige Zufallsmischung von d_1 und d_2 eine Bayes-Entscheidungsfunktion bezüglich dieser Verteilung ist.

Anhang 1

ZUSÄTZE

Zu Abschnitt 13. Begriffe aus der Mengenlehre

Für den Leser, der die Grundbegriffe der Mengenlehre etwas kennt, merken wir an, daß sich Abschn. 13 in der Sprache der Mengenlehre formulieren läßt. Dazu benötigen wir zuerst einige Begriffe und Bezeichnungen:

Eine **Menge** besteht aus **Elementen.** (Nach G. CANTOR, dem Begründer der Mengenlehre, versteht man unter einer *Menge* eine Zusammenfassung bestimmter wohlunterschiedener Objekte der Anschauung oder des Denkens zu einem Ganzen; diese Objekte heißen die *Elemente* der Menge.)

M sei eine gegebene Menge. Ist a ein Element von M, schreibt man $a \in M$. Ist a kein Element von M, schreibt man $a \notin M$.

Eine Menge A heißt eine **Teilmenge** von M, wenn jedes Element von A auch Element von M ist. Ist A eine Teilmenge von M, schreibt man $A \subset M$ (früher auch $A \subseteq M$).

Nach dieser Definition ist M eine Teilmenge von sich selbst. Es erweist sich als praktisch, die Definition so zu fassen.

Die **leere Menge** (= Menge, die kein Element enthält) wird mit $\emptyset$ bezeichnet. Sie ist Teilmenge jeder Menge.

A und B seien Teilmengen von M. Alle Elemente, die zu A oder zu B oder zu beiden gehören, bilden eine Menge; diese heißt **Vereinigung** von A und B und wird mit

$$A \cup B$$

bezeichnet. Alle Elemente, die gleichzeitig zu A und B gehören, bilden eine Menge; diese heißt **Durchschnitt** von A und B und wird mit

$$A \cap B$$

bezeichnet.

Nun sind wir so weit, daß wir Abschn. 13 in der Sprache der Mengenlehre formulieren können:

Zu einem Zufallsexperiment gehört eine ganz bestimmte Menge, nämlich die Menge aller Ergebnisse des Experimentes. Diese heißt **Grundmenge** oder *Ergebnismenge*. Wir bezeichnen sie mit S. Ihre Elemente heißen **Elementarereignisse** oder *Ergebnisse*.

Beispiel. Beim Würfeln (S. 51) können wir 1, 2, 3, 4, 5, 6 als Elementarereignisse ansehen. Die Grundmenge ist dann

$$S = \{1, 2, 3, 4, 5, 6\}.$$

Ereignisse $A, B, C, \ldots$ bei einem Zufallsexperiment sind Teilmengen der Grundmenge S des Experimentes. Das Ereignis $A + B$ (Summe von A und B) ist die Vereinigung der Mengen A und B. Dieses Ereignis heißt deshalb auch **Vereinigung** der Ereignisse A und B und wird auch mit $A \cup B$ bezeichnet.

Das Ereignis AB (Produkt von A und B) ist der Durchschnitt der Mengen A und B. Dieses Ereignis heißt deshalb auch **Durchschnitt** der Ereignisse A und B und wird auch mit $A \cap B$ bezeichnet.

Also sind einander ausschließende Ereignisse genau diejenigen, deren Durchschnitt die leere Menge $\emptyset$ ist.

In Beispiel 13.2 können wir nun schreiben:

$$\begin{aligned}
S &= \{1, 2, 3, 4, 5, 6\}\\
A &= \{2, 4, 6\}\\
B &= \{3, 6\}\\
C &= \{1\}\\
A + B &= A \cup B = \{2, 3, 4, 6\}\\
AB &= A \cap B = \{6\}\\
AC &= A \cap C = \emptyset.
\end{aligned}$$

Zu Abschnitt 15 und 16. Aus der Mengenlehre

Auch die Überlegungen dieser beiden Abschnitte lassen sich in der Sprache der Mengenlehre formulieren. Hierzu einige Bemerkungen:

Zu Abschn. 13 haben wir in diesem Anhang soeben gesagt, daß einem Zufallsexperiment eine Grundmenge oder Ergebnismenge S zugeordnet ist, die aus allen Elementarereignissen (Ergebnissen) besteht. Laut Axiom 2 ist $P(S) = 1$.

B und C seien Teilmengen von S. Diese Mengen heißen **gleich**, wenn jedes Element von B auch zu C gehört und umgekehrt. Ist B gleich C, schreibt man $B = C$.

Also sind zwei Ereignisse B und C dann und nur dann äquivalent, wenn die Mengen B und C gleich sind.

A sei eine Teilmenge von S. Die Menge aller Elemente von S, die nicht Elemente von A sind, heißt das **Komplement** von A (in S); dieses wird oft mit A^c bezeichnet.

Komplementäre Ereignisse A und $\bar{A}$ (s. Abschn. 16) sind also komplementäre Mengen. A und $\bar{A}$ schließen einander aus; man hat $A \cap A^c = \emptyset$. Weiter ist $A \cup A^c = S$.

Das Ereignis $A_1 + A_2 + \cdots + A_m$ in Formel (16.3) ist die Vereinigung $A_1 \cup A_2 \cup \cdots \cup A_m$ der Mengen $A_1, \ldots, A_m$.

Das Ereignis $A_1 + A_2 + \cdots$ in (16.6) ist die Vereinigung $A_1 \cup A_2 \cup \cdots$ der abzählbar unendlich vielen Mengen $A_1, A_2, \ldots$ Definitionsgemäß ist diese Vereinigung die Menge aller Elemente von S, die Element von wenigstens einer der Mengen $A_1, A_2, \ldots$ sind.

Das unmögliche Ereignis ist die leere Menge $\emptyset$, und es ist $P(\emptyset) = 0$. Aus $P(A) = 0$ folgt aber nicht notwendig $A = \emptyset$.

Es ist $P(S) = 1$. Aus $P(A) = 1$ folgt aber nicht notwendig $A = S$. Vgl. Abschn. 16.

Zu Abschnitt 21. Begriff des Intervalls

Ein **endliches Intervall** I ist eine Strecke auf der Zahlengeraden. Sieht man die beiden Endpunkte von I als zu I gehörig an, so spricht man von einem **abgeschlossenen Intervall.** Sind a und b diese Endpunkte, so umfaßt also I alle Zahlen x, für die

$$a \leqq x \leqq b$$

ist. Entsprechend heißt I **offen,** wenn die Endpunkte nicht als zu I gehörig angesehen werden. Dann umfaßt also I alle Zahlen x, für die

$$a < x < b$$

ist. Schließlich heißt I **halboffen,** wenn man nur einen der beiden Endpunkte als zu I gehörig ansieht. I umfaßt dann alle Zahlen x, für die

$$a \leqq x < b \quad \text{oder} \quad a < x \leqq b$$

ist je nachdem, welchen der beiden Endpunkte man zu I rechnet.

Unendlich heißt ein Intervall, wenn es die ganze Zahlengerade

$$-\infty < x < \infty$$

umfaßt oder sich nach links bzw. rechts hin ins Unendliche erstreckt. In den letzten beiden Fällen umfaßt I also alle Zahlen x, für die

$$-\infty < x \leqq a \qquad (\text{oder } -\infty < x < a)$$

bzw.

$$b \leqq x < \infty \qquad (\text{oder } b < x < \infty)$$

ist. Stattdessen schreibt man oft einfach

$$x \leqq a \qquad (\text{oder } x < a)$$

bzw.

$$x \geqq b \qquad (\text{oder } x > b).$$

Zu Abschnitt 26. Integral und Ableitung

Es folgen einige (bewußt naiv gehaltene) Bemerkungen, die sich an Leser richten, die mit der Differential- und Integralrechnung wenig vertraut sind, und ihnen zu einem besseren Verständnis des Abschnittes verhelfen sollen.

a und b seien zwei Zahlen, und es sei $a < b$. Diese können wir als Punkte auf der x-Achse darstellen (s. Abb. 26.1). Sie sind Endpunkte eines Intervalles (einer Strecke auf der x-Achse).

$f(x)$ sei eine gegebene Funktion, die für alle x in dem genannten Intervall erklärt und stetig ist. (Stetigkeit bedeutet, grob gesprochen, daß die Kurve von $f(x)$ keine Sprünge hat.) Dann können wir uns die Aufgabe stellen, den „*Flächeninhalt unter der Kurve von $f(x)$ von a bis b*" zu berechnen; gemeint ist damit der Flächeninhalt des (in Abb. 26.1 schraffierten) Stückes der Ebene, das nach unten durch einen Teil der x-Achse begrenzt wird, nach links und rechts ebenfalls geradlinig durch zwei vertikale Strecken und nach oben krummlinig durch ein Stück der Kurve von $f(x)$ (die auch viel komplizierter aussehen kann).

Eine solche Inhaltsbestimmung krummlinig begrenzter Ebenenstücke ist Grundaufgabe der Integralrechnung. Die Lösung ist das *bestimmte Integral*

$$\int_a^b f(x)\,dx. \tag{26.5}$$

Dieses ist also geometrisch der Flächeninhalt des genannten Ebenenstückes. $f(x)$ heißt *Integrand* des Integrals, a und b heißen *Integrationsgrenzen*, die Strecke von a bis b heißt *Integrationsintervall*, und x heißt *Integrationsvariable*. Auf die Bezeichnung der letzteren kommt es nicht an. Wir können auch v schreiben — wie in (26.3) — oder irgendeinen Buchstaben wählen, der sonst noch nicht vorkommt. Das ist ähnlich wie beim Summationsbuchstaben in Summen.

Definieren kann man diesen Begriff des Integrals dadurch, daß man die Kurve von $f(x)$ zwischen a und b durch kleine horizontale Strecken treppenstufenartig annähert, dadurch zu schmalen Rechtecken gelangt, deren Inhalt man angeben kann (Höhe mal Breite) und dabei die Genauigkeit der Approximation ständig vergrößert, indem man zu immer schmäleren Rechtecken übergeht (also in geeigneter Weise einen „Grenzübergang" ausführt).

Berechnen kann man solche Integrale oft vermöge der Tatsache, daß, grob gesprochen, die Differentiation die Umkehrung der Integration ist. Damit hat es folgende Bewandtnis:

Es sei $F(x)$ eine gegebene Funktion, die in einer Umgebung einer Stelle x definiert ist. Existiert dann der Grenzwert

$$\lim_{h \to 0} \frac{F(x+h) - F(x)}{h},$$

so heißt $F(x)$ an der betreffenden Stelle x *differenzierbar*, und der Grenzwert heißt die *Ableitung* von $F(x)$ an der Stelle x und wird mit $F'(x)$ bezeichnet:

$$F'(x) = \lim_{h \to 0} \frac{F(x+h) - F(x)}{h}. \tag{26.6}$$

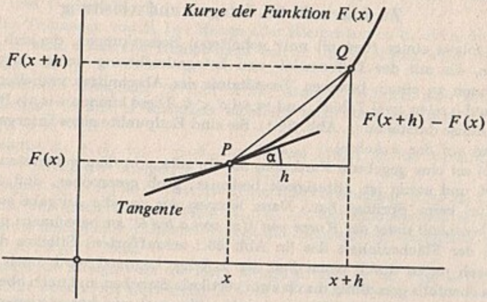


Abbildung 26.3. Zum Begriff der Ableitung einer Funktion $F(x)$

Geometrisch ist das die Steigung der Tangente an die Kurve der Funktion $F(x)$ im Punkte P mit der Abszisse x und der Ordinate $F(x)$. (Diese Steigung ist definiert als der Tangens des Winkels α zwischen der Tangente und der x-Achse.) Der Bruch auf der rechten Seite von (26.6) ist nämlich geometrisch die Steigung der Geraden durch P und einen Punkt Q der Kurve von $F(x)$ mit der Abszisse $x + h$, also der Ordinate $F(x + h)$. Und wenn h gegen 0 geht, läuft Q auf der Kurve gegen P, und die genannte Gerade dreht sich gegen eine Grenzlage, sie strebt gegen die Tangente der Kurve von $F(x)$ im Punkte P. Vgl. Abb. 26.3.

Nun zurück zu unserem Problem der Berechnung des Integrals (26.5): Es sei also $f(x)$ eine im Intervall von a bis b gegebene stetige Funktion, wie zuvor. Können wir (durch Erinnerung an Formeln der Differentialrechnung) eine Funktion $F(x)$ finden, deren Ableitung für jedes x in dem genannten Intervall gleich $f(x)$ ist, also

$$F'(x) = f(x), \tag{26.7}$$

so können wir das Integral (26.5) sofort berechnen vermöge der Formel

$$\int_a^b f(x)\,dx = F(b) - F(a) \qquad (F'(x) = f(x)). \tag{26.8}$$

Beispiel. Man berechne

$$\int_1^4 x^2\,dx.$$

Lösung: Es ist $f(x) = x^2$. Eine Funktion, die (26.7) erfüllt, ist $F(x) = x^3/3$, denn durch Differentiation erhalten wir $F'(x) = 3x^2/3 = x^2$. In (26.8) brauchen wir $F(4) = 4^3/3 = 64/3$ und $F(1) = 1^3/3 = 1/3$. Also ergibt (26.8)

$$\int_1^4 x^2\,dx = F(4) - F(1) = \tfrac{64}{3} - \tfrac{1}{3} = \tfrac{63}{3} = 21.$$

Abschließend weisen wir noch auf folgendes hin: In (26.1) ist die untere Integrationsgrenze $-\infty$; man hat also ein unendliches Integrationsintervall. Ein solches Integral heißt *uneigentlich*. Begrifflich bedeutet es

$$\int_{-\infty}^{x} f(v)\,dv = \lim_{k \to -\infty} \int_k^x f(v)\,dv,$$

in Worten: Man berechne erst das Integral über ein endliches Intervall von k bis x und lasse dann in dem Ergebnis k gegen $-\infty$ streben.

Ferner haben wir in Abschn. 26 zugelassen, daß $f(x)$ an endlich vielen Stellen unstetig sein kann; an einer solchen Stelle gilt (26.7) dann nicht mehr. Und man integriert dann über jedes Teilintervall zwischen benachbarten Unstetigkeitsstellen und addiert zum Schluß diese Teilergebnisse.

Zu Abschnitt 50. Beweis des Satzes 50.1

Aus der STIRLING-Formel

$$k! = \sqrt{2\pi k}\left(\frac{k}{e}\right)^k e^{\theta/12k} \qquad (0 < \theta < 1)$$

[Beweis siehe z. B. in K. KNOPP, Theorie und Anwendung der unendlichen Reihen. 5. Aufl. Berlin: Springer, 1964] folgt

$$f(x) = \frac{n!}{x!(n-x)!} p^x q^{n-x} = \frac{1}{\sqrt{2\pi npq}} \left(\frac{np}{x}\right)^{x+\frac{1}{2}} \left(\frac{nq}{n-x}\right)^{n-x+\frac{1}{2}} e^{\tau}$$

$$= \frac{1}{\sqrt{2\pi npq}} e^{-h} e^{\tau}$$

mit

$$\tau = \frac{1}{12}\left(\frac{\theta_1}{n} - \frac{\theta_2}{x} - \frac{\theta_3}{n-x}\right) \qquad (0 < \theta_i < 1)$$

und

$$h = \left(x + \frac{1}{2}\right) \ln \frac{x}{np} + \left(n - x + \frac{1}{2}\right) \ln \frac{n-x}{nq}$$

$$= \left(np + \frac{1}{2} + z\sqrt{npq}\right) \ln\left(1 + z\sqrt{\frac{q}{np}}\right)$$

$$+ \left(nq + \frac{1}{2} - z\sqrt{npq}\right) \ln\left(1 - z\sqrt{\frac{p}{nq}}\right),$$

wobei z in (50.4) definiert ist. $|z|$ ist nach Voraussetzung beschränkt. Für hinreichend große n ist also

$$\left|z\sqrt{\frac{q}{np}}\right| < 1, \qquad \left|z\sqrt{\frac{p}{nq}}\right| < 1,$$

und dann gelten wegen

$$\ln(1+u) = u - \frac{u^2}{2} + \frac{u^3}{3} - \frac{u^4}{4} + - \cdots \qquad (|u| < 1)$$

die Entwicklungen

$$\ln\left(1 + z\sqrt{\frac{q}{np}}\right) = z\sqrt{\frac{q}{np}} - \frac{1}{2} z^2 \frac{q}{np} + \cdots$$

$$\ln\left(1 - z\sqrt{\frac{p}{nq}}\right) = -z\sqrt{\frac{p}{nq}} - \frac{1}{2} z^2 \frac{p}{nq} - \cdots.$$

Dies setzen wir in den Ausdruck für h ein, multiplizieren aus und benutzen $p + q = 1$. Dann ergibt sich

$$h = \frac{z^2}{2} + \frac{1}{\sqrt{n}} R.$$

Hierbei ist R eine Funktion, die für $n \to \infty$ dem Betrage nach beschränkt bleibt. Also gilt in jedem beliebigen endlichen Intervall $\alpha \leq z \leq \beta$ gleichmäßig

$$e^{-h} \to e^{-z^2/2} \qquad \text{für} \qquad n \to \infty.$$

Aus $\alpha \leq z \leq \beta$ und $1 - p = q$ folgt

$$x \geq np\left(1 + \alpha\sqrt{\frac{q}{np}}\right), \qquad n - x \geq nq\left(1 - \beta\sqrt{\frac{p}{nq}}\right)$$

und demnach

$$|\tau| < \frac{1}{12}\left(\frac{1}{n} + \frac{1}{x} + \frac{1}{n-x}\right) \leq \frac{1}{12n}\left(1 + \frac{1}{p\left(1 + \alpha\sqrt{\frac{q}{np}}\right)} + \frac{1}{q\left(1 - \beta\sqrt{\frac{p}{nq}}\right)}\right).$$

Also gilt in $\alpha \leq z \leq \beta$ gleichmäßig

$$\tau \to 0 \quad \text{oder} \quad e^{\tau} \to 1 \quad \text{für} \quad n \to \infty$$

und damit nun insgesamt in $\alpha \leq z \leqq \beta$ gleichmäßig

$$f(x) \to \frac{1}{\sqrt{2\pi npq}} e^{-z^2/2}.$$

Zu Abschnitt 50. Beweis des Satzes 50.2

Man setze

$$Z = \frac{X - np}{\sqrt{npq}}, \qquad \gamma = \frac{a - np}{\sqrt{npq}}, \qquad \delta = \frac{b - np}{\sqrt{npq}}.$$

Vorausgesetzt wird, daß γ und δ auch für $n \to \infty$ endlich bleiben. Wegen Satz 50.1 gilt dann

$$P(a \leqq X \leqq b) = P(\gamma \leqq Z \leqq \delta)$$

$$= \sum_{x=a}^{b} \binom{n}{x} p^x q^{n-x} = \frac{1}{\sqrt{2\pi npq}} \sum e^{-z_j^2/2} + \sum \frac{R_j}{n} \tag{50.8}$$

(Summation über alle z_j, für die x ganzzahlig ist und in dem Intervalle $a \leq x \leq b$ liegt). Hierbei bleibt $|R_j|$ für $n \to \infty$ beschränkt. Wendet man auf das Integral

$$I_j = \frac{1}{\sqrt{2\pi}} \int\limits_{z_j - \frac{1}{2\sqrt{npq}}}^{z_j + \frac{1}{2\sqrt{npq}}} e^{-u^2/2}\, du$$

den 1. Mittelwertsatz an, so ergibt sich

$$I_j = \frac{1}{\sqrt{2\pi}\sqrt{npq}} e^{-(z_j+\theta)^2/2} \qquad \left(|\theta| < \frac{1}{2\sqrt{npq}}\right)$$

$$= \frac{1}{\sqrt{2\pi npq}} e^{-z_j^2/2} + \frac{R_j^*}{n}$$

wobei $|R_j^*|$ für $n \to \infty$ beschränkt bleibt. Also lassen sich die Exponentialfunktionen in (50.8) durch Integrale ersetzen, und es wird

$$P(\gamma \leqq Z \leqq \delta) = \frac{1}{\sqrt{2\pi}} \int\limits_{\gamma - \frac{1}{2\sqrt{npq}}}^{\delta + \frac{1}{2\sqrt{npq}}} e^{-u^2/2}\, du + \sum \frac{\tilde{R}_j}{n}$$

$$= \Phi(\beta) - \Phi(\alpha) + \sum \frac{\tilde{R}_j}{n}$$

mit α und β wie in (50.5). Hierbei bleibt $|\tilde{R}_j|$ für $n \to \infty$ beschränkt. Die letzte Summe hat

$$b - a + 1 = (\delta - \gamma)\sqrt{npq} + 1$$

Glieder. Ihr Wert strebt also für $n \to \infty$ gegen null.

Zu Abschnitt 60. Herleitung der Dichte der Chi-Quadrat-Verteilung

In (60.1) stehen lauter Quadrate. χ^2 kann also keine negativen Werte annehmen. Demnach ist $f(x) = 0$ für $x < 0$. Da nun dem Intervall $0 \leq X_j^2 \leq x$ das Intervall $-\sqrt{x} \leq X_j \leq \sqrt{x}$ entspricht, so gilt also

$$P(X_j^2 \leq x) = P(0 \leq X_j^2 \leq x) = P\left(-\sqrt{x} \leq X_j \leq \sqrt{x}\right).$$

Und da X_j normalverteilt ist mit dem Mittelwert 0 und der Varianz 1, so sehen wir aus dem letzten Ausdruck, daß X_j^2 die Verteilungsfunktion

$$F_1(x) = P(X_j^2 \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{x}}^{\sqrt{x}} e^{-u^2/2}\, du$$

besitzt. Da der Integrand symmetrisch ist, so wird

$$F_1(x) = \frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{x}} e^{-u^2/2}\, du.$$

Wir setzen $u^2 = v$. Dann ist über v von 0 bis x zu integrieren, es ist $u = \sqrt{v}$, $du = dv/2\sqrt{v}$, und wir erhalten

$$F_1(x) = \frac{1}{\sqrt{2\pi}} \int_0^x \frac{e^{-v/2}}{\sqrt{v}}\, dv.$$

Differentiation nach x ergibt die Dichte

$$f_1(x) = \frac{dF_1}{dx} = \frac{1}{\sqrt{2\pi}}\, x^{-1/2}\, e^{-x/2} \qquad (x > 0). \tag{60.9}$$

Dies ist die Dichte von X_j^2 und damit zugleich auch die Dichte von χ^2 im Falle $n = 1$, denn dann ist ja $\chi^2 = X_1^2$. In der Tat sind (60.9) und (60.2) mit $n = 1$ wegen (60.8) identisch.

Damit ist (60.2) für $n = 1$ bewiesen. Wir beenden den Beweis für beliebiges n durch Induktion. Statt f schreiben wir dabei f_n und statt χ^2 vorübergehend χ_n^2. Wir zeigen:

Gilt (60.2) für $n - 1$, so daß also

$$f_{n-1}(x) = \frac{1}{2^{(n-1)/2}\, \Gamma\left(\frac{n-1}{2}\right)}\, x^{(n-3)/2}\, e^{-x/2} \qquad (x > 0)$$

ist, so gilt (60.2) auch für n Freiheitsgrade.

Es ist

$$\chi_n^2 = \chi_{n-1}^2 + X_n^2.$$

Hierbei hat χ_{n-1}^2 gemäß Induktionsvoraussetzung die Dichte $f_{n-1}(x)$, und X_n^2 hat die Dichte $f_1(x)$, wie gezeigt wurde. Diese beiden Variablen sind unabhängig, weil dies für $X_1, \ldots, X_n$ gilt. Wir dürfen also (57.5a) (siehe Aufgabe 57.3) anwenden und erhalten für die Dichte von χ_n^2 den Ausdruck

$$f_n(x) = \int_{-\infty}^{\infty} f_1(x - y)\, f_{n-1}(y)\, dy.$$

Wir brauchen nur von 0 bis x zu integrieren, denn $f_{n-1}(y)$ ist für negative y gleich null, und $f_1(x-y)$ ist null für $x-y<0$, also $y-x>0$ oder $y>x$. Indem wir f_1 und f_{n-1} einsetzen, ergibt sich demnach zunächst

$$f_n(x) = \frac{1}{\sqrt{2\pi}} \frac{1}{2^{\frac{n-1}{2}} \Gamma\left(\frac{n-1}{2}\right)} \int_0^x (x-y)^{-\frac{1}{2}} e^{-\frac{x-y}{2}} y^{\frac{n-3}{2}} e^{-\frac{y}{2}} dy.$$

Durch Vereinfachung erhalten wir hieraus

$$f_n(x) = \frac{1}{2^{\frac{n}{2}} \sqrt{\pi}\, \Gamma\left(\frac{n-1}{2}\right)} e^{-\frac{x}{2}} \int_0^x y^{\frac{n-3}{2}} (x-y)^{-\frac{1}{2}} dy. \tag{60.10}$$

Das Integral bezeichnen wir mit I. Wir setzen $y = xu$. Dann läuft u von 0 bis 1, es ist $x - y = x - xu = x(1-u)$, $dy = x\,du$. So wird

$$I = x^{\frac{n-2}{2}} \int_0^1 u^{\frac{n-3}{2}} (1-u)^{-\frac{1}{2}} du.$$

Nun benutzen wir die bekannte Formel

$$\int_0^1 u^{a-1} (1-u)^{b-1} du = \frac{\Gamma(a)\,\Gamma(b)}{\Gamma(a+b)} \qquad (a>0,\ b>0), \tag{60.11}$$

die wir zum Schluß beweisen. Dann wird

$$I = x^{\frac{n-2}{2}} \frac{\Gamma\left(\frac{n-1}{2}\right) \Gamma\left(\frac{1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}.$$

Dies setzen wir in (60.10) ein. Wegen $\Gamma(1/2) = \sqrt{\pi}$ ergibt sich dann gerade (60.2). Damit ist der Beweis beendet. Die dabei benutzte Formel (60.11) kann man folgendermaßen gewinnen:

Gemäß (60.5) ist

$$\begin{aligned} \Gamma(a)\,\Gamma(b) &= \int_0^\infty e^{-t}\, t^{a-1} dt \int_0^\infty e^{-v}\, v^{b-1}\, dv \\ &= \int_0^\infty \int_0^\infty e^{-(t+v)}\, t^{a-1}\, v^{b-1}\, dt\, dv. \end{aligned}$$

Wir führen neue Integrationsvariable r und u ein, indem wir

$$t = ru \qquad \text{und} \qquad v = r(1-u)$$

setzen. Dann ist $dt\,dv$ durch $dr\,du$ mal dem Absolutbetrag der Jacobischen Determinante

$$\frac{\partial(t,v)}{\partial(r,u)} = \begin{vmatrix} \frac{\partial t}{\partial r} & \frac{\partial t}{\partial u} \\ \frac{\partial v}{\partial r} & \frac{\partial v}{\partial u} \end{vmatrix} = \begin{vmatrix} u & r \\ 1-u & -r \end{vmatrix} = -r$$

ersetzen. So folgt

$$\begin{aligned}\Gamma(a)\,\Gamma(b) &= \int_0^1\int_0^\infty e^{-r}\,(ru)^{a-1}\,[r(1-u)]^{b-1}\,r\,dr\,du\\ &= \int_0^\infty e^{-r}\,r^{a+b-1}\,dr\int_0^1 u^{a-1}\,(1-u)^{b-1}\,du\\ &= \Gamma(a+b)\int_0^1 u^{a-1}\,(1-u)^{b-1}\,du\,.\end{aligned}$$

Division durch $\Gamma(a+b)$ liefert (60.11).

Zu Abschnitt 62. Herleitung der Dichte der t-Verteilung

Nach Voraussetzung hat X in (62.1) die Dichte

$$f_1(x) = \frac{1}{\sqrt{2\pi}}\,e^{-x^2/2},$$

und Y hat die Dichte [vgl. (60.2)]

$$f_2(y) = \frac{1}{2^{n/2}\,\Gamma\left(\frac{n}{2}\right)}\,y^{\frac{n-2}{2}}\,e^{-\frac{y}{2}} \qquad (y > 0)$$

und $f_2(y) = 0$ für $y < 0$. Da X und Y unabhängig sind, hat die zugehörige zweidimensionale Verteilung die Dichte

$$f(x, y) = f_1(x)\,f_2(y).$$

Für die Verteilungsfunktion

$$F(z) = P(T \leqq z) = P\left(\frac{X}{\sqrt{Y/n}} \leqq z\right) = P\left(X \leqq z\sqrt{Y/n}\right)$$

der t-Verteilung erhalten wir damit

$$F(z) = \iint\limits_{x \leqq z\sqrt{y/n}} f(x, y)\,dx\,dy\,.$$

Setzen wir $f = f_1 f_2$ ein, so ergibt sich zunächst

$$F(z) = C_n \iint\limits_{\substack{x \leqq z\sqrt{y/n}\\ y>0}} e^{-\frac{x^2}{2}-\frac{y}{2}}\,y^{\frac{n-2}{2}}\,dx\,dy\,.$$

Hierbei ist

$$C_n = \frac{1}{\sqrt{2\pi}\,2^{n/2}\,\Gamma\left(\frac{n}{2}\right)}\,.$$

In diesem Doppelintegral können wir zuerst bei festem y über x von $-\infty$ bis $z\sqrt{y/n}$ integrieren und dann über y von 0 bis ∞. Demnach wird

$$F(z) = C_n \int_0^\infty \left[\int_{-\infty}^{z\sqrt{y/n}} e^{-\frac{x^2}{2}-\frac{y}{2}}\,y^{\frac{n-2}{2}}\,dx\right]dy\,.$$

Setzen wir $x = u\sqrt{y/n}$, so ist $dx = du\sqrt{y/n}$ und

$$\frac{x^2}{2} + \frac{y}{2} = \frac{Hy}{2} \quad \text{mit} \quad H = 1 + \frac{u^2}{n}.$$

Über u ist dann von $-\infty$ bis z zu integrieren. Da diese neuen Grenzen nicht von y abhängen, können wir die Integrationsreihenfolge vertauschen. So ergibt sich

$$F(z) = \frac{C_n}{\sqrt{n}} \int_{-\infty}^{z} \int_{0}^{\infty} e^{-\frac{Hy}{2}} y^{\frac{n-1}{2}} \, dy \, du.$$

Das innere Integral läßt sich durch die Gammafunktion ausdrücken, indem wir $Hy = 2v$ setzen. Dann ist $y = 2v/H$, $dy = 2dv/H$, und wir erhalten

$$F(z) = 2^{\frac{n+1}{2}} \frac{C_n}{\sqrt{n}} \int_{-\infty}^{z} \frac{du}{H^{(n+1)/2}} \int_{0}^{\infty} e^{-v} v^{\frac{n-1}{2}} \, dv.$$

Wegen (60.5) ist das letzte Integral gleich $\Gamma\left(\frac{n+1}{2}\right)$. Indem wir H und C_n wieder ausführlich schreiben, erhalten wir (62.3) und daraus durch Differentiation (62.2).

Zu Abschnitt 73. Beweis des Satzes 73.1

Ersetzen wir $X_1, \ldots, X_n$ durch $X_1 - \mu, \ldots, X_n - \mu$, so hat das auf Y keinen Einfluß, weil $\bar{X}$ dann auch durch $\bar{X} - \mu$ zu ersetzen ist. Im vorliegenden Beweis können wir deshalb ohne Beschränkung der Allgemeinheit annehmen, es sei $\mu = 0$. Die n Variablen $(X_j - \bar{X})^2$ in Y sind *nicht* unabhängig, denn es ist

$$\sum_{j=1}^{n} (X_j - \bar{X}) = \sum_{j=1}^{n} X_j - n\bar{X} = 0.$$

Wir wollen zeigen, daß die Quadratsumme

$$Q_n = \sum_{j=1}^{n} (X_j - \bar{X})^2$$

in der Form

$$Q_n = Y_1^2 + \cdots + Y_{n-1}^2 \tag{73.8}$$

darstellbar ist, wobei $Y_1, \ldots, Y_{n-1}$ unabhängige normalverteilte Variable mit dem Mittelwert 0 und der Varianz σ^2 sind. Wir setzen

$$\begin{aligned}
Y_1 &= \frac{1}{\sqrt{1 \cdot 2}} (X_1 - X_2) \\
Y_2 &= \frac{1}{\sqrt{2 \cdot 3}} (X_1 + X_2 - 2X_3) \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
Y_{n-1} &= \frac{1}{\sqrt{(n-1)\,n}} (X_1 + X_2 + \cdots + X_{n-1} - (n-1)\,X_n) \\
Y_n &= \frac{1}{\sqrt{n}} (X_1 + X_2 + \cdots + X_{n-1} + X_n).
\end{aligned} \tag{73.9}$$

Diese sogenannte HELMERT-Transformation ist eine lineare Transformation der Form

$$
\begin{array}{ll}
 & Y_1 = a_{11} X_1 + a_{12} X_2 + \cdots + a_{1n} X_n \\
(73.10) & Y_2 = a_{21} X_1 + a_{22} X_2 + \cdots + a_{2n} X_n \\
 & \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
 & Y_n = a_{n1} X_1 + a_{n2} X_2 + \cdots + a_{nn} X_n
\end{array}
$$

Wie man aus (73.9) erkennt, genügen die Koeffizienten den „Orthogonalitätsbedingungen"

$$a_{j1} a_{k1} + a_{j2} a_{k2} + \cdots + a_{jn} a_{kn} = \begin{cases} 0 & (j \neq k) \\ 1 & (j = k) \end{cases}$$

Für $n = 3$ sind dies die bekannten Bedingungen dafür, daß (73.10) eine (möglicherweise mit einer Spiegelung an einer Ebene durch den Nullpunkt verbundene) Rotation eines rechtwinkligen Koordinatensystems um den Nullpunkt darstellt. Dabei bleibt das Abstandsquadrat jedes Punktes (X_1, X_2, X_3) vom Nullpunkt unverändert,

$$X_1^2 + X_2^2 + X_3^2 = Y_1^2 + Y_2^2 + Y_3^2.$$

Dasselbe gilt auch für beliebiges n, also

$$X_1^2 + X_2^2 + \cdots + X_n^2 = Y_1^2 + Y_2^2 + \cdots + Y_n^2.$$

Aus (73.9) sehen wir unmittelbar, daß

$$Y_n^2 = \frac{1}{n} (n X)^2 = n \bar{X}^2$$

ist. Demnach wird

$$Q_n = \sum_{j=1}^{n} (X_j - \bar{X})^2 = \sum_{j=1}^{n} X_j^2 - n \bar{X}^2 = \sum_{j=1}^{n} Y_j^2 - Y_n^2,$$

und (73.8) ist bewiesen.

Aus Satz 71.2 folgt, daß die Zufallsvariablen $-X_2, -2X_3, -3X_4, \ldots$ in (73.9) normalverteilt sind mit dem Mittelwert 0 und der Varianz $\sigma^2, 4\sigma^2, 9\sigma^2, \ldots$ Außerdem sind die jeweils in einer Zeile von (73.9) auf der rechten Seite stehenden Summanden unabhängig. Damit ergibt sich aus Satz 71.1 unter Benutzung von (34.2) und (34.4), daß $Y_1/\sigma, \ldots, Y_n/\sigma$ normalverteilte Variable mit dem Mittelwert 0 und der Varianz 1 sind. Daß diese Variablen auch unabhängig sind, ergibt sich durch die Betrachtung der n-dimensionalen Normalverteilung, auf die wir hier nicht eingehen; vgl. H. CRAMER [2]. Wegen

$$Y = \frac{1}{\sigma^2} Q_n = \frac{Y_1^2}{\sigma^2} + \cdots + \frac{Y_{n-1}^2}{\sigma^2}$$

ist also Y nun durch eine Summe von Quadraten von $n - 1$ unabhängigen normalverteilten Variablen mit dem Mittelwert 0 und der Varianz 1 dargestellt, hat also eine χ^2-Verteilung mit $n - 1$ Freiheitsgraden (vgl. Abschn. 60).

Weiterhin sind die Zufallsvariablen $Y_1, \ldots, Y_{n-1}$ alle auch unabhängig von $Y_n = \sqrt{n}\, \bar{X}$. Hieraus folgt, daß auch Y und Z unabhängig sind.

Anhang 2

AUFGABENLÖSUNGEN

6.9 Weil die Werte nicht genau sondern auf die angegebene Ziffernzahl abgerundet sind.

7.1 $\tilde{F}(x) = 0$ für $x < 0$, $\tilde{F} = 0{,}4$ für $0 \leq x < 1$, $\tilde{F} = 1$ für $x \geq 1$.

7.3 $n = 30$, $\tilde{f}(41) = 4/30$, $\tilde{f}(42) = 3/30$ usw., also $\tilde{F}(x) = 0$ für $x < 41$, $\tilde{F} = 2/15$ für $41 \leq x < 42$, $\tilde{F} = 7/30$ für $42 \leq x < 43$ usw.

7.5 Klassenmitten 65, 85, 105, ... Relative Klassenhäufigkeiten 2/125, 0, 4/125, ... Also $\tilde{F}(x) = 0$ für $x < 65$, $\tilde{F}(x) = 2/125$ für $65 \leq x < 105$, $\tilde{F}(x) = 6/125$ für $105 \leq x < 125$ usw.

8.3 $\bar{x} = 721{,}8/8 = 90{,}23$, $s^2 = 9{,}075/7 = 1{,}296$, $s = 1{,}14$.

8.5 $-1/\sqrt{2}$, $1/\sqrt{2}$.

8.7 Die Spannweite ist einfach zu berechnen. Sie hat aber den Nachteil, daß sie nur durch 2 Stichprobenwerte bestimmt ist, während die übrigen ganz unberücksichtigt bleiben.

8.9 $dH/dt = 0$ liefert $t = \bar{x}$.

9.1
$$\begin{aligned}\sum (x_j - \bar{x})^2 &= \sum (x_j^2 - 2x_j\,\bar{x} + \bar{x}^2)\\ &= \sum x_j^2 - 2\bar{x} \sum x_j + n\,\bar{x}^2\\ &= \sum x_j^2 - 2\frac{1}{n}\left(\sum x_j\right)^2 + n\frac{1}{n^2}\left(\sum x_j\right)^2\\ &= \sum x_j^2 - \frac{1}{n}\left(\sum x_j\right)^2.\end{aligned}$$

9.3 $\bar{x} = \frac{1}{n}\sum\,(c + x_j^*) = \frac{1}{n}(n\,c + \sum x_j^*) = c + \frac{1}{n}\sum x_j^*$.

9.5 $\bar{x} = 5$, $s^2 = 20/3$.

9.9 $\bar{x} = \frac{1}{n}\sum (c_1\,x_j^* + c_2) = \frac{c_1}{n}\sum x_j^* + \frac{n\,c_2}{n} = c_1\,\bar{x}^* + c_2$,

$$x_j - \bar{x} = c_1\,x_j^* + c_2 - (c_1\,\bar{x}^* + c_2) = c_1(x_j^* - \bar{x}^*).$$

Also wird $\sum (x_j - \bar{x})^2 = c_1^2 \sum (x_j^* - \bar{x}^*)^2$. Daraus folgt (9.7).

9.11 79,0 und $(2 + 4)/2 = 3$ bzw. 3,05.

10.1 $\bar{x} = \frac{1}{n}\sum_{j=1}^{m} x_j A_j = \sum_{j=1}^{m} x_j(A_j/n)$, A_j = absolute Häufigkeit von x_j, $A_j/n = \tilde{f}(x_j)$.

$$s^2 = \frac{1}{n-1}\sum_{j=1}^{m}(x_j - \bar{x})^2 A_j = \frac{n}{n-1}\sum_{j=1}^{m}(x_j - \bar{x})^2 (A_j/n).$$

10.3 $\bar{x} = 18{,}37$, $s^2 = 8{,}68$.

10.5 $\bar{x} = 1503$, $s^2 = 41\,665$.

10.7 (a) 3,205, (b) 3,260, (c) 3,225, ursprünglich 3,235.

13.1 $h(A) = 3/50$, $h(B) = 4/50$, $h(AB) = 1/50$, $h(A + B) = 6/50$.

13.5 $h(A) = 62{,}2\%$, $h(B) = 61{,}9\%$, $h(AB) = 47{,}1\%$, $h(A + B) = 77{,}0\%$, $h(B \mid A) = 75{,}7\%$.

14.1 50%. **14.3** 1/8. **14.5** 1/6.

16.1 $\bar{A}$ ist ein sicheres Ereignis; also ist $P(\bar{A}) = 1$, und $P(A) = 0$ folgt dann aus (16.1).

18.1 $P(A + B + C) = P(A) + P(B) + P(C)$
$$- P(AB) - P(AC) - P(BC) + P(ABC)$$

19.1 (7/10) (6/9) (5/8) = 29%.

19.3 $BBBB$, $BBBD$, $BBDB$, $BDBB$, $DBBB$ sind die sich gegenseitig ausschließenden Formen des betrachteten Ereignisses (B = brauchbare Kappe, D = defekte Kappe). Summation der zugehörigen Wahrscheinlichkeiten ergibt die Antwort $(1/6) + 4(1/8) = 2/3$.

19.5 Aus (19.4) folgt der recht beträchtliche Wert
$$P(A_1 A_2 \cdots A_5) = \frac{40}{50} \cdot \frac{39}{49} \cdot \frac{38}{48} \cdot \frac{37}{47} \cdot \frac{36}{46} = 31\%.$$

20.1 1/16.

20.3 $P(A) = P(B) = P(C) = 2/4 = 1/2$, $P(AB) = P(AC) = P(BC) = 1/4$, also $P(A)P(B) = P(AB)$ usw., d. h. die Ereignisse sind paarweise unabhängig. Wären alle drei Ereignisse unabhängig, so müßte $P(ABC) = P(A)P(B)P(C) = 1/8$ sein. Es ist aber $P(ABC) = 0$, weil es kein Los Nr. 111 gibt.

21.1 1/4, 1/2, 1/4, 0, 3/4, 3/4, 1/4, 3/4.

21.3 Das Ereignis $X \le c$ ist die Summe der einander ausschließenden Ereignisse $X \le b$ und $b < X \le c$. Gemäß Axiom 3 gilt also
$$P(X \le c) = P(X \le b) + P(b < X \le c).$$
Gemäß Axiom 1 ist der letzte Summand nicht negativ. Daraus folgt die Behauptung.

23.3 $\sum_{x=1}^{\infty} \frac{1}{2^x} = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots = 1$ (geometrische Reihe).

23.5 13/18, 1/3.

23.7 $f(0) = f(4) = 1/16$, $f(1) = f(3) = 1/4$, $f(2) = 3/8$.

25.5 $f(0) = 1/3$, $f(1) = 8/15$, $f(2) = 2/15$.

26.1 Die erste Beziehung folgt daraus, daß Wahrscheinlichkeiten nichtnegative Zahlen sind, die zweite aus (24.1) zusammen mit der Tatsache, daß $-\infty < X < \infty$ ein sicheres Ereignis ist, also die Wahrscheinlichkeit 1 hat.

26.3 $f(x) = 1/2$ für $-1 < x < 1$, $F(x) = 0$ für $x < -1$, $F(x) = (x + 1)/2$ für $-1 < x < 1$, $F(x) = 1$ für $x > 1$, $P(X \ge 0) = 1/2$, $P(0 \le X \le 0{,}3) = 0{,}15$, $P(X = 0{,}5) = 0$.

26.5 Weil $f(x) \ge 0$ sein muß.

27.1 Je ein Massenpunkt der Masse 1/6 in den Punkten $x = 1, 2, \ldots, 6$.

27.3 Verteilung im Intervall $0 < x < 2\pi$ mit der konstanten Dichte $1/2\pi$.

28.1 7.

28.3 $\mu = \int_0^{\infty} x^2 e^{-x}\, dx = 2$ (partielle Integration).

28.5 $\mu = \sum_{x=1}^{\infty} \frac{x}{2^x}$. Es ist $\sum_{x=0}^{\infty} y^x = \frac{1}{1-y}$ $(|y| < 1)$. Differentiation nach y

ergibt

$$\sum_{x=0}^{\infty} x y^{x-1} = \sum_{x=1}^{\infty} x y^{x-1} = \frac{1}{(1-y)^2}.$$

Für $y = 1/2$ folgt

$$\sum_{x=1}^{\infty} \frac{x}{2^{x-1}} = 2 \sum_{x=1}^{\infty} \frac{x}{2^x} = \frac{1}{\left(1 - \frac{1}{2}\right)^2} = 4, \quad \text{also} \quad \mu = 2.$$

28.9 $\mu = \frac{1}{6} \sum_{x=1}^{\infty} x \left(\frac{5}{6}\right)^{x-1} = \frac{1}{6} \cdot \frac{1}{(1/6)^2} = 6$. Der Wert der Reihe ergibt sich dabei, indem man $y = 5/6$ in die differenzierte Reihe der Lösung der Aufgabe 28.5 einsetzt.

29.1 $\mu = 4/3$, $\sigma^2 = 2/9$.

29.3 $k = 750$, $\mu = 1$, $\sigma^2 = 0{,}002$.

29.5 $\int_{1-a}^{1+a} f(x)\, dx = 0{,}9$. $a = 0{,}073$, Grenzen $0{,}927$ und $1{,}073$.

29.7 0,8 cbm.

30.1 Im diskreten Falle ist

$$E\big(ag(X) + bh(X)\big) = \sum [a g(x_j) + b h(x_j)] f(x_j)$$
$$= a \sum g(x_j) f(x_j) + b \sum h(x_j) f(x_j) = a E\big(g(X)\big) + b E\big(h(X)\big).$$

30.3 1,50 Mark.

31.1 $E(X - \mu) = E(X) - \mu E(1) = E(X) - \mu = 0$.

31.3 $E([X - c]^2) = E([(X - \mu) - (c - \mu)]^2)$

$$= E([X - \mu]^2 - 2(c - \mu)(X - \mu) + [c - \mu]^2)$$
$$= E([X - \mu]^2) - 2(c - \mu)\, E(X - \mu) + [c - \mu]^2$$
$$= \sigma^2 - 2(c - \mu) \cdot 0 + [c - \mu]^2 \geq \sigma^2.$$

Das letzte Gleichheitszeichen gilt genau dann, wenn $c = \mu$ ist.

31.5 $\sigma^2 = E([X - \mu]^2) = E(X[X - 1] + X - 2\mu X + \mu^2)$ usw.

32.1 Mit $x - c = t$ wird wegen $\mu = c$ und $f(c - t) = f(c + t)$ im stetigen Falle

$$E([X - \mu]^3) = \int_{-\infty}^{\infty} t^3 f(c + t)\, dt$$
$$= \int_{-\infty}^{0} \tau^3 f(c + \tau)\, d\tau + \int_{0}^{\infty} t^3 f(c + t)\, dt$$
$$= \int_{\infty}^{0} (-t)^3 f(c - t)\,(-dt) + \int_{0}^{\infty} t^3 f(c + t)\, dt = 0.$$

Hierbei wurde $\tau = -t$ gesetzt.

32.3 $2\sqrt{2}/5$.

33.1 $\sum_{x=0}^{\infty} e^{tx} q^x p = p \sum_{x=0}^{\infty} (e^t q)^x = \frac{p}{1 - e^t q}$

33.5 $|E(e^{itX})| = \left|\sum_j e^{itx_j} f(x_j)\right| \leq \sum_j |e^{itx_j}|\, f(x_j) = \sum_j f(x_j) = 1$, denn

$$|e^{itx_j}| = \sqrt{\cos^2(tx_j) + \sin^2(tx_j)} = 1.$$

33.7 $\mu = 1/a$, $\sigma^2 = 1/a^2$.

34.1 $G^*(t) = E(e^{tX^*}) = E(e^{t(c_1X + c_2)}) = E(e^{c_1tX}\, e^{tc_2})$

$$= e^{c_2t}\, E(e^{c_1tX}) = e^{c_2t}\, G(c_1t).$$

35.1 Sind die Elemente alle verschieden, so gibt es $n!$ Permutationen. Fallen nun die unterscheidenden Merkmale der n_1 Elemente der 1. Klasse fort, so fallen je $n_1!$ Permutationen in eine einzige zusammen, nämlich jedesmal diejenigen, die durch Vertauschung der n_1 Elemente untereinander ohne Änderung der übrigen Elemente auseinander hervorgehen. Übrig bleiben nur noch $n!/n_1!$ Permutationen. Entsprechend schließt man, daß die Gleichheit der n_2 Elemente der 2. Klasse das Zusammenfallen von je $n_2!$ Permutationen in eine einzige zur Folge hat, so daß dann nur noch $n!/n_1!\,n_2!$ Permutationen übrigbleiben, usw.

35.3 1/1260, denn gemäß Satz 35.2 gibt es $9!/2!\,3!\,4! = 1260$ Permutationen der Kugeln, also 1260 gleichmögliche Fälle beim Ziehen.

36.1 Aus (19.4) oder Satz 36.1 folgt, daß die Wahrscheinlichkeit, ein Spiel zu gewinnen, 1/60 beträgt. Der Einsatz ist 1/50 des Gewinns, das Spiel also auf lange Dauer ungünstig für den Spieler.

36.3 Aus Satz 36.2 folgt, daß die Gewinnaussicht pro Spiel nun 1/10 beträgt und das Spiel sehr günstig für den Spieler ist.

37.1 Für $k = 1$ trifft die Behauptung zu. Wir nehmen an, sei sei für eine beliebige positive ganze Zahl k richtig, und zeigen, daß dann die Anzahl der Kombinationen $(k+1)$ter Ordnung $\binom{n+k}{k+1}$ beträgt. Hierzu vereinigen wir jeweils alle die Kombinationen $(k+1)$ter Ordnung, die mit dem gleichen Element beginnen, zu einer Klasse. Dies ergibt $\binom{n+k-1}{k}$ mit dem Element 1 beginnende Kombinationen, da man auf das Element 1 alle Kombinationen k-ter Ordnung der Elemente $1, 2, \ldots, n$ mit Wiederholung und ohne Berücksichtigung der Anordnung folgen lassen darf. Weiterhin hat man $\binom{n+k-2}{k}$ mit dem Element 2 beginnende Kombinationen $(k+1)$ter Ordnung, denn da man die Anordnung nicht berücksichtigt, dürfen auf das Anfangselement 2 nur noch die Kombinationen k-ter Ordnung der $n-1$ Elemente $2, 3, \ldots, n$ mit Wiederholung und ohne Berücksichtigung der Anordnung folgen. Entsprechend ergeben sich $\binom{n+k-3}{k}$ mit dem Element 3 beginnende Kombinationen usw. Schließlich ergibt sich $\binom{n+k-n}{k} = 1$ mit dem Element n beginnende Kombination. Also erhält man insgesamt

$$\binom{k}{k} + \binom{k+1}{k} + \cdots + \binom{n+k-1}{k} = \sum_{s=0}^{n-1} \binom{k+s}{k} = \binom{n+k}{k+1}$$

Kombinationen $(k+1)$ter Ordnung der betrachteten Art, wie aus (38.10) im nächsten Abschnitt folgt. Damit ist der Satz bewiesen.

37.3 $26^5 = 11\,881\,376$.

37.5 $f(1) = f(4) = 1/8$, $f(2) = f(3) = 3/8$.

38.1 $$\sum_k \binom{p}{k} x^k \sum_s \binom{q}{s} x^s = \sum_k \sum_s \binom{p}{k}\binom{q}{s} x^{k+s} = \sum_r \binom{p+q}{r} x^r.$$

Damit dies identisch in x gilt, müssen die Koeffizienten entsprechender Potenzen von x auf beiden Seiten übereinstimmen. Durch Vergleich der Koeffizienten von x^r ergibt sich (38.8).

38.3 Die Formel gilt für $n = 1$. Wir nehmen an, sie gelte für ein $n = r \geq 1$,

$$\sum_{s=0}^{r-1} \binom{k+s}{k} = \binom{r+k}{k+1},$$

und zeigen, daß sie dann auch für $n = r + 1$ gilt. Letzteres folgt durch Addition von $\binom{k+r}{k}$ wegen (38.5).

39.3 Es ist $(n-x)p/(x+1)q > 1$ bzw. < 1 für $x < np-q$ bzw. $> np-q$. Also wächst $f(x)$ im allgemeinen (d. h. von trivialen Ausnahmefällen abgesehen) zunächst an und nimmt dann wieder ab. Das Maximum liegt bei der kleinsten ganzen Zahl $x \geq np-q$. Ist $np-q$ ganzzahlig, so gilt außerdem $f(np-q+1) = f(np-q)$.

41.1 (a) 3/8, (b) 1/4, (c) 1/16.

41.3 $\frac{1}{2^{15}}\left[\binom{15}{0} + \binom{15}{1}\right] = \frac{1}{2048}$, also sehr gering.

41.5 2%.

41.9 Die Wahrscheinlichkeit, daß eine einzelne Zwiebel nicht keimt, ist $p = 0{,}05$. Also ist die Wahrscheinlichkeit, daß 10 Zwiebeln die Garantiebedingung erfüllen,

$$\sum_{x=0}^{1} \binom{10}{x} (0{,}05)^x (0{,}95)^{10-x} = 91\%.$$

Die Antwort lautet demnach 9%.

41.11 Die Wahrscheinlichkeit, daß ein Arbeiter Strom benötigt, beträgt $p = 10/60 = 1/6$. Die Wahrscheinlichkeit, daß 4 oder 5 Arbeiter gleichzeitig Strom benötigen, ist demnach gleich

$$\sum_{x=4}^{5} \binom{5}{x} \left(\frac{1}{6}\right)^x \left(\frac{5}{6}\right)^{5-x} = 0{,}3\%,$$

also recht gering.

41.13 $\binom{10}{x} (0{,}05)^x (0{,}95)^{10-x}$.

43.1

$$\sum_{x=11}^{1000} \binom{1000}{x} 0{,}01^x \, 0{,}99^{1000-x}$$

$$\approx e^{-10} \sum_{x=11}^{\infty} \frac{10^x}{x!} = 1 - e^{-10} \sum_{x=0}^{10} \frac{10^x}{x!} = 42\%.$$

43.3 Wir nehmen an, daß die Geburtstage ganz zufällig verteilt sind, und können dann das Modell eines BERNOULLI-Experimentes mit $p = 1/365$ und $n = 500$ anwenden. Aus (39.9) in Aufgabe 39.5 folgt

$$P = 1 - (364/365)^{500} = 1 - 0{,}2537 = 74{,}63\%.$$

Die POISSON-Näherung mit $\mu = np = 500/365$ für 0,2537 ist 0,2541.

43.5 $P(X \leq 4) = \sum_{x=0}^{4} \binom{100}{x} 0{,}1^x \, 0{,}9^{100-x} \approx e^{-10} \sum_{x=0}^{4} \frac{10^x}{x!} = 3\%$

43.7 X = Zahl der Anrufe je Minute. $\mu = 300/60 = 5$ = durchschnittliche Zahl der Anrufe je Minute. $P(X \leq 10) = F(10) = 0{,}9863$ (s. Tafel 2 in Anhang 5). Antwort 1,4%.

44.1 $G(t) = \sum_{x=0}^{\infty} e^{tx}\, e^{-\mu} \frac{\mu^x}{x!} = e^{-\mu} \sum_{x=0}^{\infty} \frac{(\mu e^t)^x}{x!} = e^{-\mu}\, e^{\mu e^t}$.

44.3 $G^*(t) = e^{\mu(e^t - t - 1)}$, $G^*(0) = 1$,
$G^{*\prime}(t) = \mu(e^t - 1)\, G^*(t)$, $G^{*\prime}(0) = 0$,
$G^{*\prime\prime}(t) = \mu\, e^t\, G^*(t) + \mu(e^t - 1)\, G^{*\prime}(t)$, $\sigma^2 = G^{*\prime\prime}(0) = \mu$.

45.1 Gemäß Satz 36.2 kann man aus N Schrauben $\binom{N}{n}$ verschiedene Kombinationen n-ter Ordnung bilden. Das ist die Anzahl gleichmöglicher sich ausschließender Fälle. x defekte Schrauben lassen sich auf $\binom{M}{x}$ verschiedene Weisen aus den M defekten auswählen und $n - x$ brauchbare auf $\binom{N-M}{n-x}$ Weisen aus den $N - M$ brauchbaren. So ergibt sich (45.1).

45.3 $\mu = \frac{1}{\binom{N}{n}} \sum_x x \binom{M}{x} \binom{N-M}{n-x}$.

Durch Nachrechnen bestätigt man

$$x\binom{M}{x} = M\binom{M-1}{x-1}.$$

Wegen (38.8) mit $p = M - 1$, $k = x - 1$, $q = N - M$, $r - k = n - x$ wird damit

$$\mu = \frac{M}{\binom{N}{n}} \sum_x \binom{M-1}{x-1}\binom{N-M}{n-x} = \frac{M}{\binom{N}{n}}\binom{N-1}{n-1} = n\frac{M}{N}.$$

45.5 $\left[4\binom{28}{3} + \binom{28}{2}\right] \Big/ \binom{32}{6} = \frac{13\,104 + 378}{906\,192} = 1{,}5\%$.

45.7 Enthält eine Packung von $N = 100$ Stück genau $M = 10$ fehlerhafte Stücke, so beträgt die Wahrscheinlichkeit, daß unter $n = 10$ Stücken kein fehlerhaftes Stück ist,

$$f(0) = \binom{10}{0}\binom{90}{10}\Big/\binom{100}{10} = \frac{90 \cdot 89 \cdots 81}{100 \cdot 99 \cdots 91} = 33\%.$$

Antwort 67%. Das Prüfverfahren ist also unzulänglich.

45.9 $P = f(0) = \binom{M}{0}\binom{100-M}{3}\Big/\binom{100}{3} = \frac{(100-M)(99-M)(98-M)}{100 \cdot 99 \cdot 98}$

M/N [%]	0	10	20	30	40	50	60	70	80
P [%]	100	73	51	34	21	12	6	3	1

45.11 $N = 90$, $n = 5$, $M = x = 1, 2, \ldots, 5$,

$$P = f(x) = \binom{90-x}{5-x}\Big/\binom{90}{5} = \frac{5 \cdot 4 \cdots (5 - x + 1)}{90 \cdot 89 \cdots (90 - x + 1)}$$

Antwort 1/18, 2/801, 1/11 748, 1/511 038, 1/43 949 268.

46.1 $npq = \frac{nM(N-M)}{N^2} > \frac{nM(N-M)}{N^2}\left(\frac{N-n}{N-1}\right) = \sigma^2$,
vgl. (40.4) und (45.3).

46.3 Schreibt man die Binomialkoeffizienten in (45.1) ausführlich, so sieht man, daß $f(x) = \binom{n}{x} AB/C$ mit

$$A = M(M-1)\cdots(M-x+1) \leqq M^x,$$
$$B = (N-M)(N-M-1)\cdots(N-M-n+x+1) \leqq (N-M)^{n-x}$$
$$C = N(N-1)\cdots(N-n+1) > (N-n)^n$$

ist. (Für $x = 0$ bzw. n ist $A = 1$ bzw. $B = 1$ zu setzen.) Wie behauptet wurde, gilt also

$$f(x) < \binom{n}{x} \frac{M^x (N-M)^{n-x}}{(N-n)^n} = \binom{n}{x} \left(\frac{M}{N}\right)^x \left(\frac{N-M}{N}\right)^{n-x} \Big/ \left(\frac{N-n}{N}\right)^n.$$

46.5 22%, 43%, 29%, 6% bzw. 17%, 50%, 30%, 3%, $\mu = 1{,}2$, $\sigma^2 = 0{,}72$ bzw. 0,56.

46.7 $\mu = 3p$. Die Näherung ist nur für kleine p brauchbar.

48.1 $$\Phi^2(\infty) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-u^2/2}\, du \int_{-\infty}^{\infty} e^{-v^2/2}\, dv$$
$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(u^2+v^2)/2}\, du\, dv = \frac{1}{2\pi} \int_0^{2\pi} \int_0^{\infty} e^{-r^2/2}\, r\, dr\, d\varphi$$
$$= \frac{1}{2\pi} \int_0^{2\pi} d\varphi \int_0^{\infty} e^{-r^2/2}\, r\, dr = \frac{1}{2\pi} \cdot 2\pi \cdot 1 = 1$$

($u = r\cos\varphi$, $v = r\sin\varphi$, $du\, dv = r\, dr\, d\varphi$).

48.3 Wir schreiben in (47.1) zunächst β statt σ. Dann ist

$$\sigma^2 = \frac{1}{\sqrt{2\pi}\,\beta} \int_{-\infty}^{\infty} (x-\mu)^2 e^{-\frac{1}{2}\left(\frac{x-\mu}{\beta}\right)^2} dx \qquad \left(u = \frac{x-\mu}{\beta}\right)$$
$$= \frac{1}{\sqrt{2\pi}\,\beta} \int_{-\infty}^{\infty} \beta^2 u^2 e^{-u^2/2} \beta\, du = \frac{\beta^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u^2 e^{-u^2/2}\, du$$
$$= \frac{\beta^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (-u)(-u e^{-u^2/2})\, du$$
$$= \beta^2 \left[\frac{1}{\sqrt{2\pi}} (-u)\, e^{-u^2/2} \Big|_{-\infty}^{\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-u^2/2}\, du \right] = \beta^2.$$

48.5 $P = \Phi(c) - \Phi(-c) = 2\Phi(c) - 1 = \gamma$.

49.5 $P(9{,}97 \leq X \leq 10{,}03) = \Phi\left(\frac{0{,}03}{\sigma}\right) - \Phi\left(\frac{-0{,}03}{\sigma}\right) = D\left(\frac{0{,}03}{\sigma}\right) = 99\%$, $0{,}03/\sigma = 2{,}576$ (aus Tafel 3b), $\sigma = 0{,}012$.

49.7 $\sigma = 0{,}608$. **49.9** 16.

49.11 $c = 13{,}29$, $\beta \approx 1$. β wächst.

50.1 $$P = \sum_{x=0}^{10} \binom{1000}{x} 0{,}01^x\, 0{,}99^{1000-x}$$
$$\approx \Phi\left(\frac{10 - 10 + 0{,}5}{\sqrt{9{,}9}}\right) - \Phi\left(\frac{0 - 10 - 0{,}5}{\sqrt{9{,}9}}\right) = 0{,}564$$

(exakt 0,583).

50.3 $$\Phi\left(\frac{3000 - 1500 + 0{,}5}{\sqrt{750}}\right) - \Phi\left(\frac{1578 - 1500 - 0{,}5}{\sqrt{750}}\right) = 0{,}23\%.$$

also sehr gering. Unsere Annahme dürfte deshalb falsch sein.

50.5 $\Phi\left(\frac{4040 - 2020 + 0{,}5}{\sqrt{1010}}\right) - \Phi\left(\frac{2048 - 2020 - 0{,}5}{\sqrt{1010}}\right) = 19{,}3\,\%.$

50.7 $\Phi\left(\frac{200 - 10 + 0{,}5}{\sqrt{9{,}5}}\right) - \Phi\left(\frac{10 - 10 - 0{,}5}{\sqrt{9{,}5}}\right) = 56\,\%.$

52.1 Rechts stehen die Wahrscheinlichkeiten

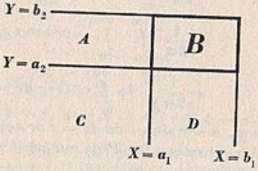
$F(b_1, b_2)$, mit der (X, Y) in A, B, C oder D liegt (d. h. irgendein Wertepaar annimmt, das einem Punkt in A, B, C oder D entspricht),

$F(a_1, b_2)$, mit der (X, Y) in A oder C liegt,

$F(b_1, a_2)$, mit der (X, Y) in C oder D liegt, und

$F(a_1, a_2)$, mit der (X, Y) in C liegt.

Wird der zweite und dritte Ausdruck von der Summe des ersten und vierten abgezogen, so bleibt gerade die zu B gehörige Wahrscheinlichkeit übrig, und (52.2) ist bewiesen.



53.5 $$f(x_1, x_2) = \frac{20!}{x_1!\,x_2!\,(20 - x_1 - x_2)!}\,0{,}03^{x_1}\,0{,}05^{x_2}\,0{,}92^{20-x_1-x_2}.$$

54.1 $$F(x, y) = \begin{cases} 0 & \text{für } x \leq \alpha_1 \text{ oder } y \leq \alpha_2, \\ (a - \alpha_1)(b - \alpha_2)/k & \text{für } x > \alpha_1 \text{ und } y > \alpha_2 \end{cases}$$

mit

$$a = \begin{cases} x & \text{für } \alpha_1 < x \leq \beta_1 \\ \beta_1 & \text{für } x > \beta_1 \end{cases} \quad \text{und} \quad b = \begin{cases} y & \text{für } \alpha_2 < y \leq \beta_2 \\ \beta_2 & \text{für } y > \beta_2. \end{cases}$$

55.3 $$f_1(x) = 1/(\beta_1 - \alpha_1) \quad \text{für} \quad \alpha_1 < x < \beta_1,$$
$$f_2(y) = 1/(\beta_2 - \alpha_2) \quad \text{für} \quad \alpha_2 < y < \beta_2.$$

56.1 Aus (24.3) folgt

$$P(a_1 < X \leqq b_1)\,P(a_2 < Y \leqq b_2) = [F_1(b_1) - F_1(a_1)]\,[F_2(b_2) - F_2(a_2)].$$

Rechts erhält man im Falle der Unabhängigkeit wegen (56.1) durch Ausmultiplizieren

$$F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2)$$

und hieraus (56.3) wegen (52.2). Gilt umgekehrt (56.3) für jedes Paar der genannten Ereignisse, so können wir a_1 und a_2 gegen $-\infty$ streben lassen und statt b_1 bzw. b_2 auch x bzw. y schreiben. Dann wird (56.3) mit (56.1) identisch.

56.3 $f(0, 0) = \frac{N - M}{N} \cdot \frac{N - M - 1}{N - 1}$, $f(0, 1) = \frac{N - M}{N} \cdot \frac{M}{N - 1} = f(1, 0)$,

$f_1(0) = f(0, 0) + f(0, 1) = \frac{N - M}{N} = f_2(0)$, $f_1(0)\,f_2(0) \neq f(0, 0)$.

57.1 $f(x, y) = 1$ in Q, $F(z) = \iint\limits_{x+y \leq z} dx\, dy =$ Fläche des Stückes von Q „unterhalb" der Geraden $x + y = z$, also $F = 0$ für $z \leq 0$, $F = z^2/2$ für $0 < z \leq 1$, $F = 1 - 2^{-1}(2 - z)^2$ für $1 < z \leq 2$, $F = 1$ für $z > 2$. Differentiation ergibt $f = 0$ für $z < 0$ oder $z > 2$, $f = z$ für $0 < z < 1$, $f = 2 - z$ für $1 < z < 2$.

57.3 $$F(z) = \iint\limits_{x+y \leq z} f(x, y)\, dx\, dy = \iint\limits_{x \leq z-y} f_1(x)\, f_2(y)\, dx\, dy$$
$$= \int\limits_{-\infty}^{\infty} f_2(y) \left[\int\limits_{-\infty}^{z-y} f_1(x)\, dx\right] dy.$$

Hieraus folgt (57.5a) durch Differentiation unter dem Integralzeichen.

58.1 $G_j(t) = E(e^{tX_j})$. Zu $X_1 + \cdots + X_n$ gehört
$$G(t) = E(e^{t(X_1+\cdots+X_n)}) = E(e^{tX_1} \cdots e^{tX_n}).$$
$$= E(e^{tX_1}) \cdots E(e^{tX_n}) = G_1(t) \cdots G_n(t).$$

58.3 $Z = X_1 + \cdots + X_n$, $X_j =$ Anzahl der roten Kugeln beim j-ten Zug, $E(X_j) = r/(r + s)$, Antwort $nr/(r + s)$.

60.1 $$\Gamma(\alpha + 1) = \int\limits_0^{\infty} e^{-t}\, t^{\alpha}\, dt = -e^{-t}\, t^{\alpha}\Big|_0^{\infty} + \alpha \int\limits_0^{\infty} e^{-t}\, t^{\alpha-1}\, dt = 0 + \alpha\, \Gamma(\alpha)$$

60.3 $$\Gamma\left(\frac{1}{2}\right) = \int\limits_0^{\infty} e^{-\tau}\, \tau^{-1/2}\, d\tau = \sqrt{2} \int\limits_0^{\infty} e^{-t^2/2}\, dt = \frac{\sqrt{2}}{2} \int\limits_{-\infty}^{\infty} e^{-t^2/2}\, dt = \frac{1}{\sqrt{2}} \sqrt{2\pi}$$

61.1 $$\mu = \frac{1}{2^{n/2}\, \Gamma\left(\frac{n}{2}\right)} \int\limits_0^{\infty} x^{n/2}\, e^{-x/2}\, dx$$
$$= \frac{2^{(n+2)/2}}{2^{n/2}\, \Gamma\left(\frac{n}{2}\right)} \int\limits_0^{\infty} e^{-t}\, t^{n/2}\, dt = \frac{2\, \Gamma\left(\frac{n}{2} + 1\right)}{\Gamma\left(\frac{n}{2}\right)} = n.$$

62.1 $\sigma^2 = I\, \Gamma\left(\frac{n+1}{2}\right) \Big/ \sqrt{n\pi}\, \Gamma\left(\frac{n}{2}\right)$ mit
$$I = 2 \int\limits_0^{\infty} z^2 \left(1 + \frac{z^2}{n}\right)^{-\frac{n+1}{2}} dz = n^{\frac{3}{2}} \int\limits_0^{1} t^{\frac{n}{2}-2}\, (1 - t)^{\frac{1}{2}}\, dt$$
$$= n^{\frac{3}{2}}\, \frac{\Gamma\left(\frac{n}{2} - 1\right) \Gamma\left(\frac{3}{2}\right)}{\Gamma\left(\frac{n+1}{2}\right)}.$$

Hierbei ist $1 + \frac{z^2}{n} = \frac{1}{t}$. So wird
$$\sigma^2 = n\, \Gamma\left(\frac{n}{2} - 1\right) \Gamma\left(\frac{3}{2}\right) \Big/ \sqrt{\pi}\, \Gamma\left(\frac{n}{2}\right), \qquad \Gamma\left(\frac{3}{2}\right) = \frac{1}{2}\, \Gamma\left(\frac{1}{2}\right) = \frac{\sqrt{\pi}}{2},$$
$$\Gamma\left(\frac{n}{2}\right) = \left(\frac{n}{2} - 1\right) \Gamma\left(\frac{n}{2} - 1\right).$$

64.1

x	1	1,5	2	2,5	3	3,5	4	4,5	5	5,5	6
$f(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

64.3 $S^2 = \frac{1}{n-1} \sum [(X_j - \mu) - (\overline{X} - \mu)]^2$

$$= \frac{1}{n-1} \sum (X_j - \mu)^2 - \frac{2}{n-1} (\overline{X} - \mu) \sum (X_j - \mu) + \frac{n}{n-1} (\overline{X} - \mu)^2$$

$$= \frac{1}{n-1} \sum (X_j - \mu)^2 - \frac{n}{n-1} (\overline{X} - \mu)^2,$$

denn aus (64.2) folgt $\sum (X_j - \mu) = n(\overline{X} - \mu)$. So wird

$$E(S^2) = \frac{1}{n-1} \sum E([X_j - \mu]^2) - \frac{n}{n-1} E([\overline{X} - \mu]^2)$$

$$= \frac{n}{n-1} \sigma^2 - \frac{n}{n-1} \left(\frac{\sigma^2}{n}\right) = \sigma^2,$$

denn $\overline{X}$ hat wegen Satz 59.1 und (34.4) die Varianz σ^2/n.

66.1 $\bar{x} \approx 50$ cm, $s \approx 2$ cm.

66.3 $\bar{x} \approx 165$ cm, $s \approx 6$ cm.

68.1 $l = 1/(b-a)^n$ wird am größten, wenn $b - a$ so klein wie möglich ist, also für a = kleinster Stichprobenwert, b = größter Stichprobenwert.

68.3 Setzen wir die Inverse $u = \psi(\varphi)$ von $\varphi(u)$ in die Likelihood-Funktion $l(u)$ ein, so erhalten wir $l(\psi(\varphi))$. Da l sein Maximum für $u = \tilde{u}$ hat, so muß für das $\varphi = \tilde{\varphi}$, das l zum Maximum macht, $\tilde{u} = \psi(\tilde{\varphi})$ gelten. Hieraus folgt $\tilde{\varphi} = \varphi(\tilde{u})$.

69.1 $5 \pm 1{,}86$, $5 \pm 0{,}186$.

69.3 $a = 2{,}576 \cdot 2{,}4/15 = 0{,}41$.

69.5 $2 \cdot 2{,}576 \cdot 3/\sqrt{n} = 0{,}2$, $n \approx 6000$.

70.3 KONF $\{4{,}48 \leq \mu \leq 5{,}75\}$.

70.5 Bei festem γ, σ und n wird $(c_1 + c_2)\, \sigma/\sqrt{n}$ am kleinsten, wenn $c_1 + c_2$ am kleinsten ist. Letzteres gilt wegen (70.6) für $c_1 = c_2$.

71.1 Normalverteilung mit $\mu = 8$, -4 bzw. 0 und $\sigma^2 = 100$, 25 bzw. 225.

71.3 Das Gewicht $Z = X + Y$ der Packungen ist gemäß Satz 71.1 normalverteilt mit $\mu = 105$ g und $\sigma^2 = 1{,}25\ \text{g}^2$. Antwort $P(104 \leq Z \leq 106) = 63\%$.

71.5 $X_2 - X_1$ ist normalverteilt mit $\mu = \mu_2 - 1$ und $\sigma^2 = 0{,}2$. Also wird

$$P(X_2 - X_1 \leq 0) = \Phi\left(\frac{0 - (\mu_2 - 1)}{\sqrt{0{,}2}}\right) = \frac{1}{1000}, \quad \frac{\mu_2 - 1}{\sqrt{0{,}2}} = 3{,}090, \quad \mu_2 = 2{,}38.$$

72.1 KONF $\{0{,}726 \leq \mu \leq 0{,}751\}$.

72.3 $c/\sqrt{n} = 0{,}715$, $\bar{x} = 4{,}37$, $s = 0{,}157$, KONF $\{4{,}25 \leq \mu \leq 4{,}49\}$.

72.5 $c/\sqrt{n} = 0{,}198$, $\bar{x} = 165{,}05$, $s = 5{,}858$ (vgl. Beispiel 10.2), KONF $\{163{,}8 \leq \mu \leq 166{,}3\}$.

72.7 $c = 1{,}98$, $\bar{x} = 0{,}445$, $s = 0{,}035$, KONF $\{0{,}438 \leq \mu \leq 0{,}453\}$.

72.9 $n = 8$.

74.1 $c_1 = 0{,}41$, $c_2 = 16{,}7$, KONF $\{0{,}00004 \leq \sigma^2 \leq 0{,}00167\}$.

74.3 $c_1 = (\sqrt{253} - 2{,}58)^2/2 = 88{,}79$, $c_2 = 170{,}87$, KONF $\{1{,}42 \leq \sigma^2 \leq 2{,}75\}$.

74.5 $(n-1)\, s^2 = 0{,}01301$, KONF $\{0{,}0005 \leq \sigma^2 \leq 0{,}0076\}$.

74.7 $(n-1)\, s^2 = 3396{,}75$, KONF $\{24{,}4 \leq \sigma^2 \leq 51{,}1\}$.

75.1 KONF $\{0{,}507 \leq p \leq 0{,}516\}$. Der Würfel war also offenbar nicht völlig regelmäßig.

75.3 KONF $\{0{,}491 \leq p \leq 0{,}523\}$, KONF $\{0{,}492 \leq p \leq 0{,}511\}$, KONF $\{0{,}494 \leq p \leq 0{,}507\}$.

76.1 $\bar{x} = 1502{,}7$, $s^2 = 41\,665$, KONF $\{1479 \leq \mu \leq 1526\}$, KONF $\{35\,766 \leq \sigma^2 \leq 49\,333\}$.

76.3 $\bar{x} = 154{,}1$, $s^2 = 353{,}2$, KONF $\{151{,}7 \leq \mu \leq 156{,}5\}$. KONF $\{298 \leq \sigma^2 \leq 426\}$.

76.5 $\bar{x} = 4{,}197$, $s^2 = 0{,}0024$, KONF $\{4{,}18 \leq \mu \leq 4{,}21\}$, KONF $\{0{,}0018 \leq \sigma^2 \leq 0{,}0033\}$.

77.1 Hypothese $p = 0{,}5$, Alternative $p > 0{,}5$, p = Wahrscheinlichkeit des Ereignisses „Kopf" beim einzelnen Wurf. Überlegung wie im Text: Ist die Hypothese richtig, so ist X (= Anzahl der „Köpfe" bei 4040 Würfen) annähernd normalverteilt mit $\mu = np = 2020$ und $\sigma^2 = npq = 1010$, $q = 1 - p$. Im Text stand $P(X \leq c) = 0{,}05$, weil die Alternativwerte kleiner als das hypothetische p waren. Nun ergibt sich c aus $P(X \geq c) = 0{,}05$ [also $P(X \leq c) = 1 - 0{,}05 = 0{,}95$], weil die Alternativwerte größer als das hypothetische p sind. So wird

$$P(X \leq c) = \Phi\left(\frac{c - 2020}{\sqrt{1010}}\right) = 0{,}95, \qquad \frac{c - 2020}{\sqrt{1010}} = 1{,}645,$$

$$c = 2020 + 52{,}28 \approx 2072 > 2048.$$

Die Hypothese wird angenommen.

77.3 $c = 6090 > 6019$, $c = 12127 > 12012$. Hypothese angenommen.

79.1 Alternative $\mu > 20$, $t = \sqrt{10}\,(21{,}1 - 20)/0{,}738 = 4{,}71 > c = 1{,}83$ ($\alpha = 5\%$), Hypothese verworfen.

79.3 $t = \sqrt{10}\,(1{,}193 - 1{,}200)/0{,}038 = -0{,}58 > c = -1{,}83$ ($\alpha = 5\%$). Hypothese angenommen.

79.5 Alternative $\mu > 0$, $t = \sqrt{7}\,(0{,}286 - 0)/4{,}31 = 0{,}18 < c = 1{,}94$ ($\alpha = 5\%$), Hypothese angenommen.

79.7 $n = 16$, $\bar{x} = 0{,}75$, $s^2 = 2{,}87$, $t = 1{,}77 < c = 2{,}13$ [$\alpha = 5\%$, Alternative $\mu \neq 0$, c aus $P(-c \leq T \leq c) = 1 - \alpha$], Hypothese angenommen.

81.1 $\bar{x} = 376{,}3$, $\bar{y} = 335{,}3$, $s_1^2 = 1009{,}3$, $s_2^2 = 869{,}3$, $t_0 = 1{,}64 < c = 2{,}13$ ($\alpha = 5\%$, 4 Freiheitsgrade), Hypothese angenommen (Alternative $\mu_1 > \mu_2$).

81.3 Hypothese $\mu_1 = \mu_2$, Alternative $\mu_1 \neq \mu_2$, $\bar{x} = 106{,}22$, $\bar{y} = 106{,}34$, $s_1^2 = 0{,}717$, $s_2^2 = 0{,}243$, $t_0 = -0{,}27$ liegt zwischen $c_1 = -2{,}31$ und $c_2 = 2{,}31$ ($\alpha = 5\%$, 8 Freiheitsgrade), Hypothese angenommen.

81.5 2. Art: $\bar{x} = 260$, $s_1^2 = 796{,}4$. 1. Art: $\bar{y} = 236$, $s_2^2 = 607{,}1$. $t_0 = 1{,}9 > c = 1{,}75$ ($\alpha = 5\%$), Hypothese verworfen. Wir nehmen also an, daß die 2. Art des Lesens besser ist.

81.7 $\bar{x} = 19{,}84$, $\bar{y} = 19{,}02$, $s_1^2 = 63{,}9$, $s_2^2 = 65{,}2$, $t_0 = 0{,}23 < c = 1{,}73$ ($\alpha = 5\%$), Hypothese angenommen.

81.9 Nein.

83.1 $v_0 = 679/254 = 2{,}67$. Wir wählen $\alpha = 5\%$. Tafel 9a enthält keinen Wert für (11, 5) Freiheitsgrade. Dieser Wert ist aber sicher größer als v_0, wie ein Blick auf die in der Tafel benachbarten Werte zeigt. Wir nehmen also die Hypothese an, daß kein Unterschied zwischen den Varianzen besteht.

83.3 $v_0 = 2{,}4 < c = 3{,}18$ [$\alpha = 5\%$, (9, 9) Freiheitsgrade].

83.5 $\bar{x} = 219$, $\bar{y} = 183$, $s_1^2 = 111$, $s_2^2 = 133$, $t_0 = 6{,}8 > c = 1{,}75$ ($\alpha = 5\%$). Antwort: Ja.

83.7 $v_0 = 13{,}61/2{,}36 = 5{,}77 > c = 3{,}79$. Antwort: Ja.

85.1 $n = 100$, $K = 6$, $e_j = 100/6$, $b_1 = 16$, $|b_1 - e_1| = 2/3$ usw., $\chi_0^2 = 1{,}28 < c = 11{,}07$ ($\alpha = 5\%$). Die Hypothese wird angenommen.

85.3 $a \geq 32$mal, denn es ist

$$\chi_0^2 = \frac{1}{25}\left[(a-25)^2 + (50-a-25)^2\right] > c = 3{,}84$$

für $a \geq 32$ (oder natürlich auch für $a \leq 18$).

85.5 $e_j = 78\,700$, $\chi_0^2 = 2000 > c = 19{,}68$ ($\alpha = 5\%$, 11 Freiheitsgrade).

85.7 Für die Theorie, denn es ist

$$\chi_0^2 = 0{,}47 < c = 7{,}81 \ (\alpha = 5\%).$$

85.9 $n = 200$, $\bar{x} = 122/200 = 0{,}61$, $f(x) = e^{-0{,}61} 0{,}61^x/x!$

x_j	0	1	2	3 und mehr
b_j	109	65	22	4
e_j	108,67	66,29	20,22	4,74

$\chi_0^2 = 0{,}3 < c = 5{,}99$ ($\alpha = 5\%$, 2 Freiheitsgrade). Die Hypothese, daß die Grundgesamtheit nach POISSON verteilt ist, wird angenommen.

85.11 $p = \bar{x}/8 = 0{,}5147$, $e_1 = 165$, $e_2 = 1402$, $e_3 = 5203$, $e_4 = 11\,035$, $e_5 = 14\,628$, $e_6 = 12\,410$, $e_7 = 6580$, $e_8 = 1994$, $e_9 = 264$, $\chi_0^2 = 92 > c = 14{,}07$ ($\alpha = 5\%$, 7 Freiheitsgrade), Hypothese verworfen. Offenbar ist also das zugrunde gelegte Modell etwas zu einfach. Große Abweichungen treten an den Enden auf (nur zum Teil durch die Neigung zur Gleichgeschlechtlichkeit bei Zwillingsgeburten erklärbar) und in der Mitte.

87.1 $n = 100$, $\bar{x} = 1{,}944$, $s = 0{,}053$, $a = 0{,}072 < c = 0{,}134$ ($\alpha = 5\%$), Hypothese angenommen.

87.3 $n = 247$, $\bar{x} = 154{,}1$, $s = 18{,}79$, $a = 0{,}14 > 1{,}36/\sqrt{247} = 0{,}09$, Hypothese verworfen. Das Ergebnis ist typisch: Auch andere (umfangreichere) Stichproben haben eine starke Schiefe.

87.5 $n = 291$, $\bar{x} = 112{,}96$, $s = 4{,}61$, $a = 0{,}14 > c = 1{,}36/\sqrt{291} = 0{,}08$ ($\alpha = 5\%$). Hypothese verworfen.

89.1 $n = 6$, $r = 3$, $n_i = 2$, $q_1 = 0{,}028$, $q_2 = 0{,}471$, $v_0 = 0{,}09 < c = 9{,}55$ ($\alpha = 5\%$). Antwort: Zufallsbedingt.

89.3 $n = 27$, $n_i = 9$, $r = 3$, $\bar{x}_1 = 259{,}9$, $\bar{x}_2 = 255{,}8$, $\bar{x}_3 = 246{,}3$, $q_1 = 869{,}6$, $q_2 = 15\,260$, $v_0 = 0{,}68 < c = 3{,}40$ ($\alpha = 5\%$). Die Hypothese, daß zwischen den Standorten kein Unterschied besteht, wird also angenommen.

89.5 $n = 12$, $r = 3$, $n_i = 4$, $q_1 = 0{,}505$, $q_2 = 4{,}70$, $v_0 = 0{,}48 < c = 4{,}26$ ($\alpha = 5\%$). Die Hypothese, daß das Pflanzdatum den Ertrag nicht beeinflußt, wird angenommen.

89.7 $n = 30$, $r = 3$, $q_1 = 9{,}89$, $q_2 = 36{,}3$, $v_0 = 3{,}68 > c = 3{,}35$ ($\alpha = 5\%$). Antwort: Ja.

89.9 $n = 12$, $r = 3$, $q_1 = 5$, $q_2 = 29$, $v_0 = 0{,}78 < c = 4{,}26$ ($\alpha = 5\%$). Antwort: Nein.

93.1 Zusätzlich zu der Rechnung in Aufgabe 89.3 sind nur die Spaltensummen und daraus ein neues $q_2 = 12051$ sowie $q_3 = q - q_1 - q_2 = 3209$ zu berechnen. Dann folgt $v_1 = 2{,}17 < c_1 = 3{,}63$ [$\alpha = 5\%$, (2, 16) Freiheitsgrade], $v_2 = 7{,}51 > c_2 = 2{,}59$ [(8, 16) Freiheitsgrade]. Demgemäß wird angenommen, daß zwischen den Zeilen (= Standplätzen) kein signifikanter Unterschied besteht, wohl aber zwischen den Spalten (= den 9 Teilnehmern).

93.3 $q_1 = 0{,}0188$, $q_2 = 0{,}0991$, $q_3 = 0{,}0238$, $v_1 = 1{,}58 < c_1 = 6{,}94$ [$\alpha = 5\%$, (2, 4) Freiheitsgrade], $v_2 = 8{,}31 > c_2 = 6{,}94$ [(2, 4) Freiheitsgrade]. Demgemäß wird angenommen, daß die Gewichte keinen signifikanten Einfluß haben, wohl aber die Bewegungsarten.

93.5 $q_1 = 565$, $q_2 = 259$, $q_3 = 93{,}1$, $v_1 = 24{,}28 > c_1 = 4{,}46$ [$\alpha = 5\%$, (2, 8) Freiheitsgrade], $v_2 = 5{,}56 > c_2 = 3{,}84$ [(4, 8) Freiheitsgrade]. Also wird angenommen, daß hinsichtlich des Ertrages zwischen den Sorten und auch zwischen den Orten ein signifikanter Unterschied besteht.

93.7 $q_1 = 2{,}72$, $q_2 = 109{,}23$, $q_3 = 0{,}20$, $v_1 = 13{,}4 < c_1 = 19{,}0$ [$\alpha = 5\%$, (2, 2) Freiheitsgrade], $v_2 = 1074{,}4 > c_2 = 18{,}5$ [(1, 2) Freiheitsgrade]. Also wird angenommen, daß zwischen den Gruppen kein signifikanter Unterschied besteht, wohl aber zwischen den beiden Methoden.

95.1 $(n-1)\, s_{xy} = \sum (x_j - \bar{x})(y_j - \bar{y}) = \sum c_1 (x_j{}^* - \bar{x}^*)\, c_2 (y_j{}^* - \bar{y}^*)$ $= c_1 c_2 s_{xy}{}^*$.

95.3 $y - 281{,}16 = 1{,}93(x - 50)$.

95.5 $y - 40{,}9 = -0{,}17(x - 47{,}5)$.

95.7 $y - 14{,}0 = 0{,}84(x - 29{,}5)$.

95.9 $y - 21{,}8 = 1{,}07(x - 9{,}8)$.

97.1 $\mu_1 = \int\limits_0^1 \int\limits_0^1 x(x+y)\,dx\,dy = \frac{7}{12}$, $\mu_2 = \frac{7}{12}$,

$$\sigma_1{}^2 = \int\limits_0^1 \int\limits_0^1 (x - \mu_1)^2 (x + y)\,dx\,dy = \frac{11}{144}$$

$$\sigma_{XY} = \int\limits_0^1 \int\limits_0^1 (x - \mu_1)(y - \mu_2)(x + y)\,dx\,dy = -\frac{1}{144}$$

$$Y - \frac{7}{12} = -\frac{1}{11}\left(X - \frac{7}{12}\right).$$

97.3 $\sigma_{XY} = 0$, $Y = \mu_2 = 1/3$.

98.1 $\gamma = 95\%$, $c = 3{,}18$, $s_1{}^2 = 20500$, $s_2{}^2 = 382520$, $b = 4{,}3$, $a = 973$, $l = 0{,}2$, KONF $\{4{,}1 \leq \beta \leq 4{,}5\}$.

98.3 $\gamma = 0{,}95$, $c = 2{,}10$, $s_1{}^2 = 17{,}74$, $s_2{}^2 = 608$, $b = 5{,}39$, $a = 1748$, $l = 1{,}13$, KONF $\{4{,}26 \leq \beta \leq 6{,}52\}$.

99.1 $\gamma = 95\%$, $c = 2{,}37$, $s_1{}^2 = 0{,}037$, $s_2{}^2 = 11{,}25$, $b = 12{,}07$, $a = 47$, $y = 12{,}07x - 8{,}01$, $l = 6{,}1h$,

$$h^2 = \frac{1}{9} + \frac{(x - 3{,}1)^2}{0{,}3}.$$

99.3 $\gamma = 95\%$, $c = 2{,}31$, $s_1{}^2 = 4{,}9$, $s_2{}^2 = 2{,}4$, $b = 0{,}55$, $a = 8{,}4$, $y = 0{,}55x + 7{,}7$, $l = 2{,}4h$,

$$h^2 = \frac{1}{10} + \frac{(x - 50{,}4)^2}{44{,}4}.$$

100.1 $\alpha^* = 5\%$, $c = 1{,}86$, $s_1{}^2 = 9{,}6$, $s_2{}^2 = 140$, $b = 3{,}5$, $t_0 = 6{,}34 > c$, Hypothese verworfen.

100.3 $\alpha^* = 5\%$, $c = 6{,}31$, $s_1{}^2 = 1$, $s_2{}^2 = (3 + p^2)/3$, $b = 1$, $a = 2p^2/3$, $t_0 = \sqrt{3}/|p|$, also $t_0 = c$ für $|p| = \sqrt{3}/c = 0{,}27$, $t_0 < c$ für $|p| > 0{,}27$. Sobald demnach die Punkte hinreichend stark streuen, wird die Hypothese $\beta = 0$ angenommen, obwohl b relativ groß ist.

102.1 $q = 4$, $q_1 = 2$, $q_2 = 2$, $v_0 = 2 < c = 18{,}5$ $(\alpha = 5\%)$, Hypothese angenommen.

102.3 $n = 10$, $q = 40{,}1$, $q_1 = 14{,}9$, $q_2 = 25{,}2$, $v_0 = 4{,}73 < c = 5{,}32$ $(\alpha = 5\%)$, Hypothese angenommen.

102.5 $n = 10$, $q = 1259$, $q_1 = 1050$, $q_2 = 209$, $v_0 = 40{,}25 > c = 5{,}32$ $(\alpha = 5\%)$, Hypothese verworfen.

103.1 $$\sum_{j=1}^{n_i} (y_{ij} - \eta_i)^2 = \sum_{j=1}^{n_i} [(\bar{y}_i - \eta_i) + (y_{ij} - \bar{y}_i)]^2$$

$$= n_i(\bar{y}_i - \eta_i)^2 + \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 + 2(\bar{y}_i - \eta_i) \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i).$$

Die letzte Summe ist null [vgl. (103.1)]. Summation über i von 1 bis r ergibt (103.4).

103.3 $n = 10$, $r = 5$, $q = 31{,}8$, $q_1 = 9{,}8$, $q_2 = 22{,}0$, $v_0 = 0{,}75 < c = 5{,}41$ $(\alpha = 5\%)$, Hypothese angenommen.

103.5 $n = 12$, $r = 6$, $q = 9697$, $q_1 = 9623$, $q_2 = 74$, $v_0 = 196 > c = 4{,}53$ $(\alpha = 5\%)$, Hypothese verworfen.

104.1 $y = 0{,}5 + 1{,}5x^2$.

104.3 $y = 6642 + 762{,}3x - 156{,}1x^2$.

105.1 $y = 29{,}12 - 0{,}17x + 0{,}00028x^2$, $q_1 = 2{,}56$, $q_2 = 0{,}12$, $v_0 = 32 > c = 3{,}37$ [$\alpha = 5\%$, (6, 9) Freiheitsgrade], Hypothese verworfen.

105.3 Man berechne $y_{ij}{}^* = \ln y_{ij}$, dann die Regressionsgerade

$$y^* - 2{,}59 = -0{,}007(x - 116{,}7)$$

der $(x_i, y_{ij}{}^*)$-Werte und schließlich

$$y = e^{y^*} = 30e^{-0{,}007x}.$$

106.1 $-0{,}028$. **106.3** 0,98. **106.5** 0,55. **106.7** 0,6.

107.3 $m_{10} = m_{01} = 7/12 = 0{,}58$, $m_{20}{}^* = m_{02}{}^* = 11/144 = 0{,}076$, $m_{11}{}^* = -1/144 = -0{,}0069$, $\varrho = -1/11 = 0{,}091$.

107.5 $m_{10} = m_{01} = 1/3$, $m_{20}{}^* = m_{02}{}^* = 1/18$, $m_{11}{}^* = 0$, $\varrho = 0$.

108.1 $\sigma_{XY} = E([X - \mu_1][Y - \mu_2]) = E\left(\sigma_1 X^* [\varrho\sigma_2 X^* + \sqrt{1-\varrho^2}\, \sigma_2 Y^*]\right)$, $E(X^{*2}) = 1$, $E(X^* Y^*) = E(X^*)\, E(Y^*) = 0$, also $\sigma_{XY} = \sigma_1 \varrho \sigma_2$, $\sigma_{XY}/\sigma_1 \sigma_2 = \varrho$, vgl. (107.4).

108.3 45°.

108.5 Ellipsen mit Hauptachsen parallel zu den Koordinatenachsen.

108.7 Man benutze (19.1).

109.1 $c = 2{,}576$, $r = 0{,}47$, $z_0 = 0{,}51$, $a = 1{,}82$, KONF $\{-0{,}87 \leq \varrho \leq 0{,}99\}$.

109.3 $c = 2{,}576$, $r = 0{,}128$, $z_0 = 0{,}129$, $a = 0{,}974$, KONF $\{-0{,}69 \leq \varrho \leq 0{,}81\}$.

109.5 $c = 2{,}576$, $r = 0{,}66$, $z_0 = 0{,}79$, $a = 0{,}13$, KONF $\{0{,}58 \leq \varrho \leq 0{,}73\}$.

110.1 Ist $b = 0$, so ist $s_{xy} = 0$, also $b^* = 0$, und die Geraden stehen aufeinander senkrecht. Ist $b \neq 0$, so folgt aus (110.6) sofort $x - \bar{x} = (1/b)\,(y - \bar{y})$. Dies stimmt genau dann mit (110.8) überein, wenn $b^* = 1/b$, also $b b^* = 1$ ist. Gemäß (110.10) ist dann $r^2 = 1$. Hieraus folgt die zu beweisende Aussage (vgl. Abschn. 106).

110.3 $y = 0{,}6x + 0{,}2$, $x = 1{,}5y$, $r = 0{,}95$.

110.5 $y = 0{,}4$, $x = 1{,}3$, $r = 0$.

111.1 $2434{,}46 \pm 4{,}39$, $\delta = 3{,}29$.

111.3 $80{,}02 \pm 0{,}13$, $\delta = 0{,}10$.

111.5 KONF $\{0{,}8 \leq \sigma^2 \leq 6{,}0\}$.

112.1 X_j hat die Dichte

$$f_j(x_j) = \frac{\sqrt{c_j}}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{c_j}{2}\left(\frac{x_j - \mu}{\sigma}\right)^2}.$$

Man bilde l gemäß (67.1), dann $\ln l$. Dann setze man

$$\frac{\partial \ln l}{\partial \mu} = \frac{1}{\sigma^2}\left[c_1\,(x_1 - \mu) + \cdots + c_n\,(x_n - \mu)\right]$$

gleich null. Auflösung nach μ ergibt die Abschätzung

$$\tilde{\mu} = \frac{c_1 x_1 + \cdots + c_n x_n}{c_1 + \cdots + c_n}.$$

Wegen (112.3) sind die c_j umgekehrt proportional zu den Varianzen. Wir können also $c_j = g_j$ setzen. Dann wird $\tilde{\mu} = \bar{x}^*$.

112.3 (a) $\bar{x}^* = \dfrac{1}{n+m}(x_1 + \cdots + x_n + \tilde{x}_1 + \cdots + \tilde{x}_m) = \dfrac{n\,\bar{x}}{n+m} + \dfrac{m\,\bar{\bar{x}}}{n+m}$
wegen $x_1 + \cdots + x_n = n\,\bar{x}$ und $\tilde{x}_1 + \cdots + \tilde{x}_m = m\,\bar{\bar{x}}$.

(b) $\bar{X}$ bzw. $\bar{\bar{X}} = \dfrac{1}{m}(\tilde{X}_1 + \cdots + \tilde{X}_m)$ hat wegen Satz 59.8 und Formel (34.4) die Varianz
σ^2/n bzw. σ^2/m. Also hat $\bar{X}^* = g_1\,\bar{X} + g_2\,\bar{\bar{X}}$ wegen (34.4) und Satz 59.1 die Varianz

$$\sigma^{*2} = \left(\frac{g_1^2}{n} + \frac{g_2^2}{m}\right)\sigma^2.$$

Man setze $g_2 = 1 - g_1$ ein und $\partial\sigma^{*2}/\partial g_1$ gleich null.

112.5 $73°2'8''$.

113.1 $F = ab = 12\,000 \pm 17$ [cm²].

113.3 $45 \pm 0{,}012$ [m].

113.5 $\gamma = 8{,}014 \pm 0{,}039$ [g/cm³].

113.7 $m = \sqrt{a^2 m_1^2 + b^2 m_2^2}$.

114.1 $y - 1{,}76 = 0{,}04\,(x - 55)$.

114.3 $y - 0{,}038 = 0{,}0086\,(x - 6)$.

114.5

$$\begin{aligned} 4b_0 + 2b_1 + 6b_2 &= 4 \\ 2b_0 + 6b_1 + 8b_2 &= 4 \\ 6b_0 + 8b_1 + 18b_2 &= 10. \end{aligned}$$

$y = 0{,}3 - 0{,}1x + 0{,}5x^2$.

115.1 Hypothese $\mu = 0$, Alternative $\mu > 0$, $\bar{x} = 0{,}83$, $t = \sqrt{10}\cdot 0{,}83/0{,}82 = 3{,}21 > c = 1{,}83$ ($\alpha = 5\%$, 9 Freiheitsgrade, Tafel 8). Hypothese verworfen (auch bei $\alpha = 1\%$).

115.3 Hypothese $\mu = 0$, Alternative $\mu > 0$, $\bar{x} = 1{,}58$, $t = \sqrt{10} \cdot 1{,}58/1{,}23 = 4{,}06 > c = 1{,}83$ ($\alpha = 5\%$), Hypothese verworfen.

115.5 $n = 8$, $P(X \geq 7) = 1 - P(X \leq 6) = 1 - 0{,}9648 = 3{,}5\%$ (siehe Tafel 1d). Hypothese, daß kein Unterschied besteht, bei Wahl der Signifikanzzahl $\alpha = 5\%$ verworfen.

116.3 $n = 7$. Hypothese: Kein Trend. Alternative: Negativer Trend. 5,6 und 5,4 stehen vor 6,0 (also 2 Fehlordnungen). Im übrigen sind die Werte fallend angeordnet. $P(T \leq 2) = 0{,}5\%$ (s. Tafel 10a), und die Hypothese wird verworfen.

116.5 Hypothese: Kein Trend. Alternative: Positiver Trend. $P(T \leq 10) = 1{,}4\%$ (s. Tafel 10a), Hypothese verworfen (Signifikanzzahl 5%).

117.1 $n = 20$, $M^* = 84$, $m = 9$, $v = 4 < c_1 = 6$ ($\alpha = 5\%$), Hypothese, daß die Reihenfolge zufällig ist, verworfen.

117.3 $n = 20$, $M = 35$, $m = 10$, $v = 10 > c_1 = 6$ ($\alpha = 5\%$), Hypothese, daß die Reihenfolge zufällig ist, angenommen.

118.1 $n = 7$, $m = 6$, $u_0 = 8$, Hypothese, daß die Grundgesamtheit den Mittelwert 0 besitzt, angenommen (s. Tafel 10c).

119.1 $\Omega = \{1, 2\}$, $A = \{1, 2\}$, wobei $1 =$ Wahl einer ungeraden Zahl und $2 =$ Wahl einer geraden Zahl bedeutet.

	$\theta_1 = 1$	$\theta_2 = 2$
$a_1 = 1$	-1	2
$a_2 = 2$	2	-3

119.3 9 Entscheidungsfunktionen. 5 davon sind sinnlos, denn $x = 0$ macht $\theta = 2$ unmöglich, usw. Die anderen 4 sind in Aufgabe 119.4 angegeben.

119.7 $R(\theta, d_1) = 0,\ 0{,}25,\ 1$; $R(\theta, d_2) = 0,\ 0{,}5,\ 0$; $R(\theta, d_3) = 1,\ 0,\ 1$; $R(\theta, d_4) = 1,\ 0{,}25,\ 0$.

120.1 0,5.

120.3 $R(\mu, d_2) = k^2 + (1 - k)^2$, $k = 0{,}5$.

120.5 $R(u, d) = (k_1^2 + \cdots + k_n^2)\,\sigma^2$, $k_1 = \cdots = k_n = 1/n$.

123.1 d_2.

123.3 d_2 ist besser als d_3.

123.5 Es gebe eine vollständige Klasse, die d nicht enthält. Dann enthält diese Klasse ein d^*, das besser als d ist, im Widerspruch zu der vorausgesetzten Zulässigkeit.

123.7 $k = 5/8$, $R_{min} = 1/8$. Maximales Risiko $R(\theta_1, d_k) = 2 - 3k$ für $k \leqq 5/8$ und $R(\theta_2, d_k) = -3 + 5k$ für $k \geqq 5/8$.

123.9 d_2.

124.1 (a) d_4, (b) d_2, (c) d_1.

124.5 k beliebig; $\bar{R}(q, d_k) = 1/8$, dies ist der Wert von $R(\theta_1, d_k)$ und $R(\theta_2, d_k)$ im Falle der Minimax-Entscheidungsfunktion.

124.11 d_2 für $0 \leqq q_1 \leqq 1/3$; d_1 für $1/3 \leqq q_1 \leqq 1/2$.

124.13 d_2 für $0 \leqq q_1 \leqq 0{,}6$; d_1 für $0{,}6 \leqq q_1 \leqq 0{,}75$.

124.15 Die Bayes-Entscheidungsfunktion entspricht den Werten

$$a_0 = \frac{1{,}2}{2 + q_1}, \qquad a_1 = \frac{1{,}8 - q_1}{3 - q_1}.$$

LITERATUR

[1] D. Blackwell and M. A. Girshick, Theory of Games and Statistical Decisions. New York: Wiley, 1954.

[2] H. Cramér, Mathematical Methods of Statistics. 8th printing. Princeton: Princeton University Press, 1958.

[3] W. Feller, An Introduction to Probability Theory and Its Applications. Vol. I. 2nd ed. 6th printing. New York: Wiley, 1961.

[4] R. A. Fisher, Statistical Methods for Research Workers. 14th ed. Edinburgh: Oliver and Boyd, 1970.

[5] B. W. Gnedenko, Lehrbuch der Wahrscheinlichkeitsrechnung. 6. Aufl. Berlin: Akademie-Verlag, 1970.

[6] W. Grossmann, Grundzüge der Ausgleichsrechnung. 2. Aufl. Berlin: Springer, 1961.

[7] D. Huff, How to Lie with Statistics. New York: Norton, 1954.

[8] A. Kolmogoroff, Grundbegriffe der Wahrscheinlichkeitsrechnung. Berlin: Springer, 1933.

[9] K. Krickeberg, Wahrscheinlichkeitstheorie. Stuttgart: Teubner, 1963.

[10] A. Linder, Statistische Methoden für Naturwissenschafter, Mediziner und Ingenieure. 3. Aufl. Basel: Birkhäuser, 1960.

[11] J. v. Neumann and O. Morgenstern, Theory of Games and Economic Behavior. 3rd. ed. Princeton, N. J.: Princeton University Press, 1953.

[12] L. Schmetterer, Einführung in die mathematische Statistik. 2. Aufl. Wien: Springer, 1966.

[13] B. L. van der Waerden, Mathematische Statistik. Berlin: Springer, 1957.

[14] E. Wagemann, Narrenspiegel der Statistik. 3. Aufl. Bern: Francke, 1950.

[15] A. Wald, Statistical Decision Functions. 2nd ed. Bronx, N. Y.: Chelsea, 1971.

[16] S. S. Wilks, Mathematical Statistics. New York: Wiley, 1962.

Tafeln

[17] R. A. Fisher and F. Yates, Statistical Tables for Biological, Agricultural, and Medical Research. 5th ed. London: Oliver and Boyd, 1957.

[18] A. Hald, Statistical Tables and Formulas. 4th printing. New York: Wiley, 1962.

[19] Harvard Computation Laboratory, Tables of the Cumulative Binomial Probability Distribution. Cambridge, Mass.: Harvard University Press, 1955.

[20] National Bureau of Standards, Tables of Normal Probability Functions. Washington: US Gov. Printing Office, 1953.

[21] E. S. Pearson and H. O. Hartley, Biometrika Tables for Statisticians. Vol. 1. 2nd ed. Cambridge: University Press, 1958.

[22] Rand Corporation, A Million Random Digits With 100,000 Normal Deviates. Glencoe: Free Press, 1955.

Anhang 4

ENGLISCHE FACHAUSDRÜCKE

Deutsch—Englisch

A

Abbrechen (einer Erhebung)	Cut-off
Abhängig	Dependent
Abhängigkeit	Dependence
Ablehnen	Reject
Ablehnungsbereich	Rejection region
Abnahmeprüfung	Acceptance inspection
Absolute Häufigkeit	Absolute frequency
Absolute Klassenhäufigkeit	Absolute class frequency
Abweichung	Deviation
Additionssatz der Wahrscheinlichkeit	Theorem of total probability, addition rule
Alternative	Alternative, alternative hypothesis
Alternativhypothese	Alternative, alternative hypothesis
Annahmebereich	Acceptance region
Annehmen	Accept
Anordnungsstatistik	Order statistics
Anpassung, beste	Best fit
Anzahl in einer Klasse	Absolute class frequency, absolute cell frequency
Arithmetischer Mittelwert	Arithmetic mean
Asymmetrie	Dissymmetry
Aufspaltung	Decomposition
Ausführung eines Experimentes	Experimentation
Ausschußanteil	Fraction defective
Ausschußzahl, zulässige	Tolerance number of defects
Auswahleinheit	Sampling unit
Auswahlverfahren, systematisches	Systematic sampling
Autokorrelation	Autocorrelation
Autoregression	Autoregression

B

Balkendiagramm, Stabdiagramm	Bar chart
Bedingte Wahrscheinlichkeit	Conditional probability
Behandlung	Treatment

Belegungsproblem	Occupancy problem
Beobachtung	Observation
Beobachtungsfehler	Error of observation
Beobachtungsmaterial	Data
Bereich	Region
BERNOULLI-Experiment	BERNOULLI trial
BERNOULLI-Verteilung	BERNOULLI distribution, binomial distribution
Besetzungsproblem	Occupancy problem
Beta-Verteilung	Beta distribution
Bevölkerung	Population
Binomialverteilung	Binomial distribution, BERNOULLI distribution
Block	Block
BOREL-Menge	BOREL set

C

CAUCHY-Verteilung	CAUCHY distribution
Charakteristische Funktion	Characteristic function
Chi-Quadrat-Test	Chi-square test
Chi-Quadrat-Verteilung	Chi-square distribution

D

Daten	Data
Dezil	Decile
Dichte	Density
Diskrete Verteilung	Discrete distribution
Diskriminanzanalyse	Discriminant analysis
Dosis	Dose
Durchschnitt	Arithmetic mean, mean

E

Eigenkorrelation	Autocorrelation
Eigenregression	Autoregression
Einander ausschließende Ereignisse	Mutually exclusive events, incompatible events
Eindimensionale Verteilung	One-dimensional distribution, univariate distribution
Einfacher Stichprobenplan	Single-sampling plan
Einfach-Klassifikation	One-way classification
Eingipflig	Unimodal
Einteilung in Schichten	Stratification
Endliche Gesamtheit	Finite population
Entgegengesetztes Ereignis	Complementary event
Entscheidung	Decision
Entscheidungsfunktion	Decision function
Ereignis	Event

Erfolg	Success
Erfolgswahrscheinlichkeit	Probability of success
Erhebung	Census, inquiry, survey
Erwartung	Expectation
Erwartungstreu	Unbiased
Erwartungswert	Expectation, expected value
Erzeuger-Risiko	Producer's risk
Experiment	Experiment
Exzeß	Excess

F

Faktorenanalyse	Factor analysis
Faltung	Convolution
Fast sicher	Almost certain
Fehler	Error
Fehler 1. Art	Type I error
Fehler 2. Art	Type II error
Fragebogen	Questionnaire
Freiheitsgrade	Degrees of freedom
Frequenzfunktion	Frequency function
Frequenzfunktion der Stichprobe	Sample frequency function
F-Verteilung	F-distribution, variance-ratio distribution

G

GAUSS-Verteilung	Normal distribution, GAUSS distribution
Geburtenziffer	Birth rate
Gegenhypothese	Alternative, alternative hypothesis
Geometrische Verteilung	Geometric distribution
Gerechtes Spiel	Fair game
Gesamtheit	Population
Gesetz der großen Zahlen	Law of large numbers
Gewicht	Weight
Gewogenes Mittel	Weighted average
Gitterplan	Lattice design
Gleichförmige Verteilung	Uniform distribution
Gleichverteilung	Uniform distribution
Glockenkurve	Bell-shaped curve
Grenzverteilung	Asymptotic distribution
Grenzwertsatz	Limit theorem
Grundgesamtheit	Population, parent population, universe
Gruppe	Group
Gruppierung	Grouping
Güte der Anpassung	Goodness of fit
Gütefunktion	Power function
Güteüberwachung, statistische	Quality control, statistical

H

Häufigkeit	Frequency
Häufigkeitsfunktion	Frequency function
Häufigkeitspolygon	Frequency polygon
Häufigkeitstabelle	Frequency table
Häufigkeitsverteilung	Frequency distribution
Häufigkeitsverteilung einer Stichprobe	Sample distribution
Häufigster Wert	Mode
Herstellerrisiko	Producer's risk
Herstellungslos	Lot
Herstellungsposten	Lot
Histogramm	Histogram
Hypergeometrische Verteilung	Hypergeometric distribution
Hypothese	Hypothesis

I

Intervallschätzung	Interval estimation
Iterationen	Runs

K

Klasse	Class
Klasseneinteilung	Grouping
Klassenhäufigkeit	Class frequency, cell frequency
Klassenintervall	Class interval, cell
Klassenmitte	Class midpoint, class mark
Klassifizierung	Grouping
Klumpen	Cluster
Komplementäres Ereignis	Complementary event
Konfidenzbereich	Confidence region, confidence belt
Konfidenzgrenzen	Confidence limits
Konfidenzgürtel	Confidence belt
Konfidenzintervall	Confidence interval
Konfidenzzahl	Confidence coefficient
Konsistente Schätzfunktion	Consistent estimator
Konsumentenrisiko	Consumer's risk
Kontrolle	Control
Kontrollkarte	Control chart
Korrelation	Correlation
Korrelationskoeffizient	Correlation coefficient
Korrelierte Variable	Correlated variables
Kovarianz	Covariance
Kovarianzanalyse	Analysis of covariance
Kreisdiagramm	Pie chart
Kumulante	Cumulant
Kurven gleicher Wahrscheinlichkeit	Equiprobability curves

L

Langfristiger Trend	Secular trend
LAPLACEscher Grenzwertsatz	Limit theorem of LAPLACE
Lateinisches Quadrat	Latin square
Lieferposten	Lot
Likelihood-Funktion	Likelihood function
Likelihood-Gleichung	Likelihood equation
Likelihood-Quotiententest	Likelihood ratio test
Lineare Korrelation	Linear correlation
Lineare Regression	Linear regression
Linearer Trend	Linear trend
Logarithmische Normalverteilung	Log-normal distribution

M

Macht	Power
Maßzahl	Statistic
Maßzahl der Stichprobe	Sample statistic
Mathematische Erwartung	Mathematical expectation
Maximum-Likelihood-Methode	Maximum likelihood method
Medianwert	Median
Mehrdimensional	Multivariate
Mehrstufen-Stichprobenverfahren	Multi-stage sampling
Meßfehler	Error of observation
Methode der kleinsten Quadrate	Method of least squares
Methode der maximalen Mutmaßlichkeit	Maximum likelihood method
Mißerfolg	Failure
Mittel	Mean, average
Mittelwert	Mean, mean value
Mittlerer Fehler	Standard error
Modalwert	Mode
Moment	Moment
Momentenmethode	Method of moments
Momenterzeugende Funktion	Moment generating function
Multinomialverteilung	Multinomial distribution
Multiplikationssatz	Theorem of compound probability, multiplication rule
Mutmaßlichkeit	Likelihood
Mutungsgrenzen	Confidence limits

N

Nebenbedingung	Constraint
Negative Korrelation	Inverse correlation
Negative Schiefe	Negative skewness
Neuerkrankungsziffer	Attack rate
Nichtbeantwortung	Non-response
Nicht erwartungstreu	Biased

Nichtlineare Korrelation	Curvilinear correlation
Nichtparametrische Methoden	Nonparametric methods
Normalgleichungen	Normal equations
Normalverteilung	Normal distribution, GAUSS distribution
Normalverteilung, mehrdimensionale	Multivariate normal distribution
Normalverteilung, zweidimensionale	Bivariate normal distribution
Nullhypothese	Null hypothesis

O

OC-Kurve	OC curve
Operations-Charakteristik	Operating characteristic
Orthogonale Regression	Orthogonal regression

P

Parameter	Parameter
Parameterfrei	Nonparametric
Parzelle	Plot
Periode	Cycle, period
Permutation	Permutation
Planung des Experimentes	Design of experiment
POISSON-Verteilung	POISSON distribution
Positive Korrelation	Direct correlation
Positive Schiefe	Positive skewness
Posten	Lot
Prinzip der kleinsten Quadrate	Principle of least squares
Probeerhebung	Exploratory survey, pilot survey
Produkt von Ereignissen	Product of events
Produzentenrisiko	Producer's risk
Provisorisches Mittel	Assumed mean
Prozeß	Process
Prüfverfahren	Test (of significance)
Punktdiagramm	Dot frequency diagram
Punktschätzung	Point estimation

Q

Qualitätskontrolle	Quality control
Quartil	Quartile

R

Randverteilung	Marginal distribution
Rangkorrelation	Rank correlation
Reaktion	Response
Rechteckverteilung	Uniform distribution
Regression	Regression
Regressionskoeffizient	Regression coefficient
Relative Häufigkeit	Relative frequency, frequency ratio

Relative Klassenhäufigkeit	Relative class frequency
Relative Summenhäufigkeit	Cumulative relative frequency
Risiko 1. Art (2. Art)	Risk of Type I (of Type II)
Rückweisungsbereich	Rejection region

S

Saisonschwankung	Seasonal variation
Schätzfehler	Error of estimation
Schätzfunktion	Estimator
Schätzung der Parameter	Estimation of parameters
Schätzwert	Estimate
Schicht	Stratum
Schichtung	Stratification
Schiefe	Skewness, asymmetry
Schluß von der Stichprobe auf die Grundgesamtheit	Statistical inference
Schwankung	Variation
Sequenzanalyse	Sequential analysis
Sicheres Ereignis	Certain event
Sich gegenseitig ausschließende Ereignisse	Mutually exclusive events, incompatible events
Signifikanzzahl	Significance level
Spalten	Columns
Spannweite	Range
Spieltheorie	Theory of games
Stabdiagramm	Bar chart, bar diagram
Staffelbild	Histogram
Standardabweichung	Standard deviation
Statistik	Statistics
Statistische Maßzahl	Statistic
Statistischer Fehler	Statistical error
Statistisches Material	Data
Sterbetafel	Life table
Sterbeziffer	Death rate
Stetige Verteilung	Continuous distribution
Stichprobe	Sample
Stichprobenentnahme	Sampling
Stichprobenumfang	Sample size
Stichprobenverteilung	Sample distribution
Stichprobenwerte	Sample values
Stochastische Abhängigkeit	Stochastic dependence
Stochastische Unabhängigkeit	Stochastic independence
Stochastische Konvergenz	Convergence in probability
Stochastische Variable	Random variable, variate, stochastic variable
Streuung	(siehe Abschnitt 8)
Streuungszerlegung	Analysis of variance
Strichliste	Tally chart

STUDENTsche Verteilung	STUDENT's t-distribution
Summe von Ereignissen	Sum of events
Summenhäufigkeit	Cumulative frequency
Symmetrische Verteilung	Symmetric distribution
Systematischer Fehler	Systematic error, procedural bias

T

Tabelle	Table
Test	Test
Toleranzgrenzen	Tolerance limits
Tortendiagramm	Pie chart
Trend	Trend
Treppendiagramm	Histogram
t-Verteilung	t-distribution

U

Überwachung	Control
Umfang einer Stichprobe	Size of a sample
Unabhängige Ereignisse	Independent events
Unabhängigkeit	Independence
Unendliche Grundgesamtheit	Infinite population
Unkorreliert	Uncorrelated
Unmögliches Ereignis	Impossible event
Untersuchungseinheit, kleinste	Elementary unit
Unvollständig	Incomplete
Urne	Urn

V

Varianz	Variance
Varianzanalyse	Analysis of variance
Verbraucherrisiko	Consumer's risk
Verlaufsrichtung	Trend
Verschlüsselung	Coding (vgl. Abschnitt 9)
Versuch	Trial, experiment
Versuchsfehler	Experimental error
Verteilung	Distribution
Verteilungsfunktion	(siehe Abschnitt 24)
Verteilungsunabhängige Verfahren	Nonparametric methods
Vertrauensbereich	Confidence region, confidence interval, confidence belt
Verwerfen	Reject
Verwerfungsbereich	Critical region
Vierfeldertafel	Fourfold table
Voraussage	Prediction

W

Wahrscheinlichkeit	Probability
Wahrscheinlichkeitsdichte	Probability density function

Wahrscheinlichkeitsfunktion	Probability function
Wahrscheinlichkeitsmasse	Probability mass
Wahrscheinlichkeitsnetz	Probability paper
Wahrscheinlichkeitspapier	Probability paper
Wahrscheinlichkeitsverteilung	Probability distribution
Wartezeitproblem	Congestion problem
Wirksamkeit	Efficiency

Z

Zeilen	Rows
Zentraler Grenzwertsatz	Central limit theorem
Zentrales Moment	Central moment
Zentralwert	Median
Ziehen mit (ohne) Zurücklegen	Drawing with (without) replacement
Zufälliger Fehler	Statistical error, random error
Zufallsbeobachtung	Random observation
Zufallsereignis	Random event
Zufallsexperiment	Random experiment
Zufallsprozeß	Random process
Zufallsvariable	Random variable, stochastic variable, variate
Zusammengesetztes Stabdiagramm	Component bar chart
Zweidimensionale Verteilung	Bivariate distribution, joint distribution
Zweifache Einteilung	Two-way classification

Englisch — Deutsch

A

Absolute cell frequency	Absolute Klassenhäufigkeit
Absolute class frequency	Absolute Klassenhäufigkeit
Absolute error	Absoluter Fehler
Absolute frequency	Absolute Häufigkeit
Accept	Annehmen, nicht verwerfen
Acceptance inspection	Abnahmeprüfung
Addition rule	Additionssatz
Aleatory variable	Zufallsvariable
Alternative	Alternative, Alternativhypothese
Alternative hypothesis	Alternative, Alternativhypothese
Analysis of covariance	Kovarianzanalyse
Analysis of variance	Varianzanalyse, Streuungszerlegung
Arithmetic mean	Mittelwert
Asymmetry	Schiefe
Asymptotically normal	Asymptotisch normal
Asymptotic distribution	Grenzverteilung
At random	Zufällig
Autocorrelation	Autokorrelation

Autoregression	Autoregression
Average	Mittelwert (auch Medianwert)
Average quality protection	Gewährleistung der Durchschnittsqualität

B

Bar chart	Stabdiagramm
Bar diagram	Stabdiagramm
Bell-shaped curve	Glockenkurve
BERNOULLI distribution	BERNOULLI-Verteilung, Binomialverteilung
BERNOULLI trial	BERNOULLI-Experiment
Beta distribution	Beta-Verteilung
Beta error	Fehler 2. Art
Bias	Bias, systematischer Fehler
Biased	Mit einem systematischen Fehler behaftet
Bimodal distribution	Zweigipflige Verteilung
Binomial distribution	Binomialverteilung, BERNOULLI-Verteilung
Birth rate	Geburtenziffer
Bivariate distribution	Zweidimensionale Verteilung
Block	Block
BOREL set	BOREL-Menge

C

CAUCHY distribution	CAUCHY-Verteilung
Cell	Klassenintervall
Cell frequency	Klassenhäufigkeit
Census	Totalerhebung
Central limit theorem	Zentraler Grenzwertsatz
Central moment	Zentrales Moment
Certain event	Sicheres Ereignis
Chance, by	Zufällig
Characteristic function	Charakteristische Funktion
Chi-square distribution	Chi-Quadrat-Verteilung
Chi-square test	Chi-Quadrat-Test
Class	Klasse
Class frequency	Klassenhäufigkeit
Class interval	Klassenintervall
Class mark	Klassenmittelpunkt
Class midpoint	Klassenmittelpunkt
Closed-ended question	Frage mit vorgegebenen Antwortmöglichkeiten
Cluster	Klumpen
Cluster sampling	Klumpenauswahlverfahren
Coding	Verschlüsselung

Columns	Spalten
Combination	(siehe Abschnitt 36)
Complementary event	Entgegengesetztes Ereignis
Component bar chart	Zusammengesetztes Stabdiagramm
Composite hypothesis	Zusammengesetzte Hypothese
Conditional probability	Bedingte Wahrscheinlichkeit
Confidence belt	Konfidenzbereich, Konfidenzgürtel, Vertrauensbereich
Confidence interval	Konfidenzintervall, Vertrauensintervall
Confidence level	Konfidenzzahl, Vertrauenswahrscheinlichkeit
Confidence limits	Konfidenzgrenzen, Vertrauensgrenzen
Confidence region	Konfidenzbereich
Congestion Problem	Wartezeitproblem
Consistent estimator	Konsistente Schätzfunktion
Consumer's risk	Verbraucherrisiko
Continuous distribution	Stetige Verteilung
Continuous variate	Stetige Zufallsvariable
Convergence in probability	Konvergenz im Sinne der Wahrscheinlichkeit, stochastische Konvergenz
Convolution	Faltung
Correlation	Korrelation
Correlation coefficient	Korrelationskoeffizient
Covariance	Kovarianz
Critical region	Kritischer Bereich, Verwerfungsbereich
Cumulative relative frequency	Relative Summenhäufigkeit
Cumulative distribution function	(Siehe Abschnitt 24)
Curvilinear correlation	Nichtlineare Korrelation
Curvilinear regression	Nichtlineare Regression
Cut-off	Abbrechen einer Erhebung

D

Data	Daten
Death rate	Sterbeziffer
Decision function	Entscheidungsfunktion
Degrees of freedom	Freiheitsgrade
Density	Dichte
Dependence	Abhängigkeit
Dependent	Abhängig
Design of experiment	Planung eines Experimentes
Discontinuous variate	Diskrete Zufallsvariable
Discrete variate	Diskrete Zufallsvariable
Distribution	Verteilung
Distribution curve	Kurve der Verteilungsfunktion
Distribution function	(Siehe Abschnitt 24)
Dot frequency diagram	Punktdiagramm

Drawing without replacement	Ziehen ohne Zurücklegen
Drawing with replacement	Ziehen mit Zurücklegen
E	
Efficient estimator	Wirksame Schätzfunktion
Equiprobability curves	Kurven gleicher Wahrscheinlichkeit
Error	Fehler
Error of observation	Meßfehler, Beobachtungsfehler
Error of first (second) kind	Fehler 1. (2.) Art
Estimate	Schätzung, Schätzwert, auch Schätzfunktion
Estimation of parameter	Parameterschätzung
Estimator	Schätzfunktion
Event	Ereignis
Expectation	Erwartung
Experiment	Experiment
Experimentation	Ausführung eines Experimentes
Exponential distribution	Exponentialverteilung
F	
Factorial	Fakultät
Failure	Mißerfolg
Fair game	Gerechtes Spiel
F-distribution	F-Verteilung
Fertility rate	Fruchtbarkeitsziffer
Finite population	Endliche Grundgesamtheit
Freehand method	Graphische Ausgleichung nach Augenmaß
Frequency	Häufigkeit
Frequency function	Häufigkeitsfunktion
Frequency polygon	Häufigkeitspolygon
Frequency ratio	Relative Häufigkeit
Frequency table	Häufigkeitstabelle
F-test	F-Test
G	
Gamma distribution	Gamma-Verteilung
GAUSS distribution	GAUSS-Verteilung, Normalverteilung
Generating function	Erzeugende Funktion
Geometric distribution	Geometrische Verteilung
Goodness of fit	Güte der Anpassung
Group	Gruppe
Grouping	Klassifizieren, gruppieren
H	
Histogram	Histogramm, Staffelbild
Hypergeometric distribution	Hypergeometrische Verteilung
Hypothesis	Hypothese

I

Illusory correlation	Scheinkorrelation, Unsinnskorrelation
Impossible event	Unmögliches Ereignis
Incompatible	Unverträglich, einander ausschließend
Independent	Unabhängig
Infinite population	Unendliche Grundgesamtheit
Information	Information
Inquiry	Erhebung
Interviewer bias	Ergebnisbeeinflussung durch den Befrager

J

Joint distribution	Mehrdimensionale, insbesondere zweidimensionale Verteilung

L

Latin square	Lateinisches Quadrat
Law of large numbers	Gesetz der großen Zahlen
Least squares method	Methode der kleinsten Quadrate
Life table	Sterbetafel
Likelihood	Likelihood, Mutmaßlichkeit
Likelihood function	Likelihood-Funktion
Limit theorem	Grenzwertsatz
Linear regression	Lineare Regression
Loss of information	Informationsverlust
Lot	Herstellungsposten, Herstellungslos
Lot quality protection	Gewähr der Qualität eines Herstellungspostens

M

Marginal distribution	Randverteilung
Mathematical expectation	Erwartung, mathematische Erwartung
Maverick	Ausreißer
Maximum likelihood method	Maximum-Likelihood-Methode, Methode der größten Mutmaßlichkeit
Mean	Mittelwert
Mean value	Mittelwert
Median	Medianwert, Zentralwert
Method of moments	Momentenmethode
Mode	Häufigster Wert
Moment	Moment
Moment generating function	Momenterzeugende Funktion
Moment of inertia	Trägheitsmoment

Most powerful test	Mächtigster Test
Multinomial distribution	Multinomialverteilung
Multiplication rule	Multiplikationssatz der Wahrscheinlichkeit
Multi-stage sampling	Mehrstufenstichprobenverfahren
Mutually exclusive events	Sich gegenseitig ausschließende Ereignisse

N

Negative binomial distribution	Negative Binomialverteilung
Nonlinear regression	Nichtlineare Regression
Nonparametric methods	Nichtparametrische Methoden, verteilungsunabhängige Methoden
Nonsense correlation	Unsinnskorrelation, Scheinkorrelation
Normal distribution	Normalverteilung, GAUSS-Verteilung
Normal equations	Normalgleichungen
Null hypothesis	Nullhypothese

O

Observation	Beobachtung
Occupancy problem	Besetzungsproblem
One-way classification	Klassifizierung nach einem einzigen Merkmal
Open-ended question	Frage ohne vorgegebene Antwortmöglichkeiten
Operating characteristic	Operations-Charakteristik

P

Parameter	Parameter
Parent population	Grundgesamtheit
PASCAL distribution	PASCAL-Verteilung
Permutation	(siehe Abschnitt 35, 36)
Pie chart	Tortendiagramm, Kreisdiagramm
POISSON distribution	POISSON-Verteilung
Population	Grundgesamtheit
Power	Macht
Powerful	Mächtig
Principle of least squares	Methode der kleinsten Quadrate
Probability	Wahrscheinlichkeit
Probability density	Wahrscheinlichkeitsdichte
Probability distribution	Wahrscheinlichkeitsverteilung
Probability function	Wahrscheinlichkeitsfunktion
Probability mass	Wahrscheinlichkeitsmasse
Probability paper	Wahrscheinlichkeitsnetz
Producer's risk	Produzentenrisiko
Product of events	Produkt von Ereignissen

Q

Quality control	Qualitätskontrolle
Questionnaire	Fragebogen

R

Random error	Zufallsfehler, statistischer Fehler
Random event	Zufallsereignis
Random experiment	Zufallsexperiment
Random observation	Zufallsbeobachtung
Random sample	Stichprobe, zufällige Stichprobe
Random variable	Zufallsvariable
Range	Spannweite, Bereich
Rank correlation	Rangkorrelation
Rectangular distribution	Rechteckverteilung
Regression	Regression
Regression coefficient	Regressionskoeffizient
Reject	Verwerfen
Relative class frequency	Relative Klassenhäufigkeit
Relative frequency	Relative Häufigkeit
Replacement	Zurücklegen (beim Ziehen)
Risk of Type I (Type II)	Risiko 1. Art (2. Art)
Rows	Zeilen
Runs	Iterationen

S

Sample	Stichprobe
Sample distribution	Häufigkeitsverteilung der Stichprobe
Sample frequency function	Häufigkeitsfunktion der Stichprobe
Sample mean	Mittelwert der Stichprobe
Sample regression line	Regressionsgerade der Stichprobe
Sample space	Stichprobenraum
Sample statistic	Maßzahl einer Stichprobe
Sample values	Stichprobenwerte
Sampling	Stichprobenentnahme, Stichprobenverfahren
Seasonal variation	Jahreszeitliche Schwankung
Sequential analysis	Sequenzanalyse
Sign test	Zeichentest
Significance level	Signifikanzzahl
Size of sample	Stichprobenumfang
Skewness	Schiefe
Standard deviation	Standardabweichung
Standard error	Mittlerer Fehler
Standardized variable	Zufallsvariable in Standardform (vgl. Abschnitt 34)
Statistic	Maßzahl
Statistical error	Zufälliger Fehler, statistischer Fehler

Statistical inference	Schluß von der Stichprobe auf die Grundgesamtheit
Stochastically independent	Stochastisch unabhängig
Stochastic process	Stochastischer Prozeß
Stochastic variable	Zufallsvariable
STUDENT's distribution	t-Verteilung von STUDENT
Subsample	Gruppe (bei der Varianzanalyse)
Success	Erfolg
Sum of events	Summe von Ereignissen
Symmetric distribution	Symmetrische Verteilung
Systematic error	Systematischer Fehler

T

Tally chart	Strichliste
t-distribution	t-Verteilung von STUDENT
Test	Test
Theorem of total probability	Additionssatz der Wahrscheinlichkeit
Theorem of compound probability	Multiplikationssatz der Wahrscheinlichkeit
Trend	Trend
Trial	Zufallsexperiment
Type I error	Fehler 1. Art
Type II error	Fehler 2. Art

U

Unbiased	Frei von systematischen Fehlern, erwartungstreu
Uncorrelated	Unkorreliert
Uniform distribution	Rechteckverteilung, gleichförmige Verteilung
Universe	Grundgesamtheit

V

Variance	Varianz
Variance ratio distribution	F-Verteilung
Variate	Zufallsvariable

W

Weight	Gewicht
Weight function	Gewichtsfunktion
Weighted average	Gewogenes Mittel

Anhang 5

ZAHLENTAFELN

1 Fakultät. Binomialkoeffizienten. Binomialverteilung. Gammafunktion

Tafel 1a. Fakultät

n	$n!$
1	1
2	2
3	6
4	24
5	120

n	$n!$
6	720
7	5 040
8	40 320
9	362 880
10	3 628 800

n	$n!$
11	39 916 800
12	479 001 600
13	6 227 020 800
14	87 178 291 200
15	1 307 674 368 000

n	$n!$
16	20 922 789 888 000
17	355 687 428 096 000
18	6 402 373 705 728 000
19	121 645 100 408 832 000
20	2 432 902 008 176 640 000

Tafel 1b. Logarithmus der Fakultät

n	$\log(n!)$
1	0,000 000
2	0,301 030
3	0,778 151
4	1,380 211
5	2,079 181

n	$\log(n!)$
6	2,857 332
7	3,702 431
8	4,605 521
9	5,559 763
10	6,559 763

n	$\log(n!)$
11	7,601 156
12	8,680 337
13	9,794 280
14	10,940 408
15	12,116 500

n	$\log(n!)$
16	13,320 620
17	14,551 069
18	15,806 341
19	17,085 095
20	18,386 125

Tafel 1e. Binomialkoeffizienten

k	$\binom{1}{k}$	$\binom{2}{k}$	$\binom{3}{k}$	$\binom{4}{k}$	$\binom{5}{k}$	$\binom{6}{k}$	$\binom{7}{k}$	$\binom{8}{k}$	$\binom{9}{k}$	$\binom{10}{k}$	$\binom{11}{k}$	$\binom{12}{k}$	$\binom{13}{k}$
0	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	2	3	4	5	6	7	8	9	10	11	12	13
2		1	3	6	10	15	21	28	36	45	55	66	78
3			1	4	10	20	35	56	84	120	165	220	286
4				1	5	15	35	70	126	210	330	495	715
5					1	6	21	56	126	252	462	792	1287
6						1	7	28	84	210	462	924	1716
7							1	8	36	120	330	792	1716
8								1	9	45	165	495	1287
9									1	10	55	220	715
10										1	11	66	286
11											1	12	78
12												1	13
13													1

k	$\binom{14}{k}$	$\binom{15}{k}$	$\binom{16}{k}$	$\binom{17}{k}$	$\binom{18}{k}$	$\binom{19}{k}$	$\binom{20}{k}$
0	1	1	1	1	1	1	1
1	14	15	16	17	18	19	20
2	91	105	120	136	153	171	190
3	364	455	560	680	816	969	1140
4	1001	1365	1820	2380	3060	3876	4845
5	2002	3003	4368	6188	8568	11628	15504
6	3003	5005	8008	12376	18564	27132	38760
7	3432	6435	11440	19448	31824	50388	77520
8	3003	6435	12870	24310	43758	75582	125970
9	2002	5005	11440	24310	48620	92378	167960
10	1001	3003	8008	19448	43758	92378	184756
11	364	1365	4368	12376	31824	75582	167960
12	91	455	1820	6188	18564	50388	125970
13	14	105	560	2380	8568	27132	77520
14	1	15	120	680	3060	11628	38760
15		1	16	136	816	3876	15504
16			1	17	153	969	4845
17				1	18	171	1140
18					1	19	190
19						1	20
20							1

Tafel 1d. Binomialverteilung

Vergleiche (39.4) und (39.7)

n	x	$p = 0,1$		$p = 0,2$		$p = 0,3$		$p = 0,4$		$p = 0,5$	
		$f(x)$	$F(x)$	$f(x)$	$F(x)$	$f(x)$	$F(x)$	$f(x)$	$F(x)$	$f(x)$	$F(x)$
		0,		0,		0,		0,		0,	
1	0	9000	0,9000	8000	0,8000	7000	0,7000	6000	0,6000	5000	0,5000
	1	1000	1,0000	2000	1,0000	3000	1,0000	4000	1,0000	5000	1,0000
2	0	8100	0,8100	6400	0,6400	4900	0,4900	3600	0,3600	2500	0,2500
	1	1800	0,9900	3200	0,9600	4200	0,9100	4800	0,8400	5000	0,7500
	2	0100	1,0000	0400	1,0000	0900	1,0000	1600	1,0000	2500	1,0000
3	0	7290	0,7290	5120	0,5120	3430	0,3430	2160	0,2160	1250	0,1250
	1	2430	0,9720	3840	0,8960	4410	0,7840	4320	0,6480	3750	0,5000
	2	0270	0,9990	0960	0,9920	1890	0,9730	2880	0,9360	3750	0,8750
	3	0010	1,0000	0080	1,0000	0270	1,0000	0640	1,0000	1250	1,0000
4	0	6561	0,6561	4096	0,4096	2401	0,2401	1296	0,1296	0625	0,0625
	1	2916	0,9477	4096	0,8192	4116	0,6517	3456	0,4752	2500	0,3125
	2	0486	0,9963	1536	0,9728	2646	0,9163	3456	0,8208	3750	0,6875
	3	0036	0,9999	0256	0,9984	0756	0,9919	1536	0,9744	2500	0,9375
	4	0001	1,0000	0016	1,0000	0081	1,0000	0256	1,0000	0625	1,0000
5	0	5905	0,5905	3277	0,3277	1681	0,1681	0778	0,0778	0313	0,0313
	1	3281	0,9185	4096	0,7373	3602	0,5282	2592	0,3370	1563	0,1875
	2	0729	0,9914	2048	0,9421	3087	0,8369	3456	0,6826	3125	0,5000
	3	0081	0,9995	0512	0,9933	1323	0,9692	2304	0,9130	3125	0,8125
	4	0005	1,0000	0064	0,9997	0284	0,9976	0768	0,9898	1563	0,9688
	5	0000	1,0000	0003	1,0000	0024	1,0000	0102	1,0000	0313	1,0000
6	0	5314	0,5314	2621	0,2621	1176	0,1176	0467	0,0467	0156	0,0156
	1	3543	0,8857	3932	0,6554	3025	0,4202	1866	0,2333	0938	0,1094
	2	0984	0,9841	2458	0,9011	3241	0,7443	3110	0,5443	2344	0,3438
	3	0146	0,9987	0819	0,9830	1852	0,9295	2765	0,8208	3125	0,6563
	4	0012	0,9999	0154	0,9984	0595	0,9891	1382	0,9590	2344	0,8906
	5	0001	1,0000	0015	0,9999	0102	0,9993	0369	0,9959	0938	0,9844
	6	0000	1,0000	0001	1,0000	0007	1,0000	0041	1,0000	0156	1,0000
7	0	4783	0,4783	2097	0,2097	0824	0,0824	0280	0,0280	0078	0,0078
	1	3720	0,8503	3670	0,5767	2471	0,3294	1306	0,1586	0547	0,0625
	2	1240	0,9743	2753	0,8520	3177	0,6471	2613	0,4199	1641	0,2266
	3	0230	0,9973	1147	0,9667	2269	0,8740	2903	0,7102	2734	0,5000
	4	0026	0,9998	0287	0,9953	0972	0,9712	1935	0,9037	2734	0,7734
	5	0002	1,0000	0043	0,9996	0250	0,9962	0774	0,9812	1641	0,9375
	6	0000	1,0000	0004	1,0000	0036	0,9998	0172	0,9984	0547	0,9922
	7	0000	1,0000	0000	1,0000	0002	1,0000	0016	1,0000	0078	1,0000
8	0	4305	0,4305	1678	0,1678	0576	0,0576	0168	0,0168	0039	0,0039
	1	3826	0,8131	3355	0,5033	1977	0,2553	0896	0,1064	0313	0,0352
	2	1488	0,9619	2936	0,7969	2965	0,5518	2090	0,3154	1094	0,1445
	3	0331	0,9950	1468	0,9437	2541	0,8059	2787	0,5941	2188	0,3633
	4	0046	0,9996	0459	0,9896	1361	0,9420	2322	0,8263	2734	0,6367
	5	0004	1,0000	0092	0,9988	0467	0,9887	1239	0,9502	2188	0,8555
	6	0000	1,0000	0011	0,9999	0100	0,9987	0413	0,9915	1094	0,9648
	7	0000	1,0000	0001	1,0000	0012	0,9999	0079	0,9993	0313	0,9961
	8	0000	1,0000	0000	1,0000	0001	1,0000	0007	1,0000	0039	1,0000

Tafel 1d. Binomialverteilung (Fortsetzung)

$F(x)$ für $p = 0{,}5$ [vgl. (39.7)]

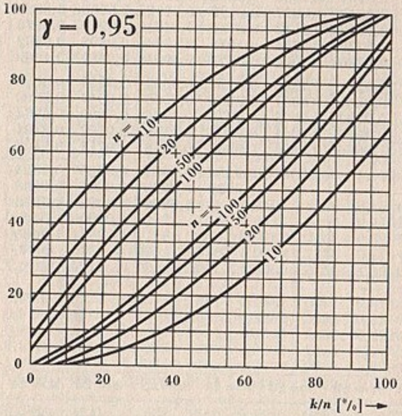
x	$n=9$	$n=10$	$n=11$	$n=12$	$n=13$	$n=14$	$n=15$	$n=16$	$n=17$
0	0,0020	0,0010	0,0005	0,0002	0,0001	0,0001	0,0000	0,0000	0,0000
1	0,0195	0,0107	0,0059	0,0032	0,0017	0,0009	0,0005	0,0003	0,0001
2	0,0898	0,0547	0,0327	0,0193	0,0112	0,0065	0,0037	0,0021	0,0012
3	0,2539	0,1719	0,1133	0,0730	0,0461	0,0287	0,0176	0,0106	0,0064
4	0,5000	0,3770	0,2744	0,1938	0,1334	0,0898	0,0592	0,0384	0,0245
5	0,7461	0,6230	0,5000	0,3872	0,2905	0,2120	0,1509	0,1051	0,0717
6	0,9102	0,8281	0,7256	0,6128	0,5000	0,3953	0,3036	0,2272	0,1662
7	0,9805	0,9453	0,8867	0,8062	0,7095	0,6047	0,5000	0,4018	0,3145
8	0,9980	0,9893	0,9673	0,9270	0,8666	0,7880	0,6964	0,5982	0,5000
9	1,0000	0,9990	0,9941	0,9807	0,9539	0,9102	0,8491	0,7728	0,6855
10		1,0000	0,9995	0,9968	0,9888	0,9713	0,9408	0,8949	0,8338
11			1,0000	0,9998	0,9983	0,9935	0,9824	0,9616	0,9283
12				1,0000	0,9999	0,9991	0,9963	0,9894	0,9755
13					1,0000	0,9999	0,9995	0,9979	0,9936
14						1,0000	1,0000	0,9997	0,9988
15							1,0000	1,0000	0,9999
16								1,0000	1,0000

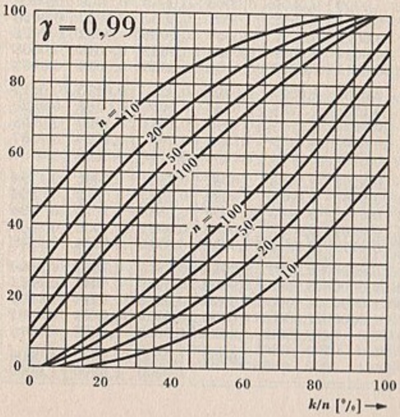
x	$n=18$	$n=19$	$n=20$	$n=21$	$n=22$	$n=23$	$n=24$	$n=25$	$n=26$
0	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
1	0,0001	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
2	0,0007	0,0004	0,0002	0,0001	0,0001	0,0000	0,0000	0,0000	0,0000
3	0,0038	0,0022	0,0013	0,0007	0,0004	0,0002	0,0001	0,0001	0,0000
4	0,0154	0,0096	0,0059	0,0036	0,0022	0,0013	0,0008	0,0005	0,0003
5	0,0481	0,0318	0,0207	0,0133	0,0085	0,0053	0,0033	0,0020	0,0012
6	0,1189	0,0835	0,0577	0,0392	0,0262	0,0173	0,0113	0,0073	0,0047
7	0,2403	0,1796	0,1316	0,0946	0,0669	0,0466	0,0320	0,0216	0,0145
8	0,4073	0,3238	0,2517	0,1917	0,1431	0,1050	0,0758	0,0539	0,0378
9	0,5927	0,5000	0,4119	0,3318	0,2617	0,2024	0,1537	0,1148	0,0843
10	0,7597	0,6762	0,5881	0,5000	0,4159	0,3388	0,2706	0,2122	0,1635
11	0,8811	0,8204	0,7483	0,6682	0,5841	0,5000	0,4194	0,3450	0,2786
12	0,9519	0,9165	0,8684	0,8083	0,7383	0,6612	0,5806	0,5000	0,4225
13	0,9846	0,9682	0,9423	0,9054	0,8569	0,7976	0,7294	0,6550	0,5775
14	0,9962	0,9904	0,9793	0,9608	0,9331	0,8950	0,8463	0,7878	0,7214
15	0,9993	0,9978	0,9941	0,9867	0,9738	0,9534	0,9242	0,8852	0,8365
16	0,9999	0,9996	0,9987	0,9964	0,9915	0,9827	0,9680	0,9461	0,9157
17	1,0000	1,0000	0,9998	0,9993	0,9978	0,9947	0,9887	0,9784	0,9622
18	1,0000	1,0000	1,0000	0,9999	0,9996	0,9987	0,9967	0,9927	0,9855
19		1,0000	1,0000	1,0000	0,9999	0,9998	0,9992	0,9980	0,9953
20			1,0000	1,0000	1,0000	1,0000	0,9999	0,9995	0,9988
21				1,0000	1,0000	1,0000	1,0000	0,9999	0,9997
22					1,0000	1,0000	1,0000	1,0000	1,0000

Tafel 1e. Konfidenzintervall für den Parameter p der Binomialverteilung

γ = Konfidenzzahl, n = Stichprobenumfang,
k = beobachtete Anzahl in der Stichprobe.
Vergleiche Abschnitt 75.

Beispiel: Für $n = 20$, $k = 8$ wird $k/n = 40\%$, und ein 95%-Konfidenzintervall ist KONF $\{0{,}19 \leq p \leq 0{,}64\}$.





Tafel 1f. Gammafunktion

Definition siehe (60.5). Werte außerhalb des Intervalls von 1 bis 2 berechnet man mittels (60.6). *Beispiel*: $\Gamma(3{,}4) = 2{,}4 \cdot \Gamma(2{,}4) = 2{,}4 \cdot 1{,}4 \cdot \Gamma(1{,}4) = 2{,}4 \cdot 1{,}4 \cdot 0{,}8873$

α	$\Gamma(\alpha)$	α	$\Gamma(\alpha)$	α	$\Gamma(\alpha)$	α	$\Gamma(\alpha)$
	0,		0,		0,		0,
1,01	9943	1,26	9044	1,51	8866	1,76	9214
1,02	9888	1,27	9025	1,52	8870	1,77	9238
1,03	9835	1,28	9007	1,53	8876	1,78	9262
1,04	9784	1,29	8990	1,54	8882	1,79	9288
1,05	9735	1,30	8975	1,55	8889	1,80	9314
1,06	9687	1,31	8960	1,56	8896	1,81	9341
1,07	9642	1,32	8946	1,57	8905	1,82	9368
1,08	9597	1,33	8934	1,58	8914	1,83	9397
1,09	9555	1,34	8922	1,59	8924	1,84	9426
1,10	9514	1,35	8912	1,60	8935	1,85	9456
1,11	9474	1,36	8902	1,61	8947	1,86	9487
1,12	9436	1,37	8893	1,62	8959	1,87	9518
1,13	9399	1,38	8885	1,63	8972	1,88	9551
1,14	9364	1,39	8879	1,64	8986	1,89	9584
1,15	9330	1,40	8873	1,65	9001	1,90	9618
1,16	9298	1,41	8868	1,66	9017	1,91	9652
1,17	9267	1,42	8864	1,67	9033	1,92	9688
1,18	9237	1,43	8860	1,68	9050	1,93	9724
1,19	9209	1,44	8858	1,69	9068	1,94	9761
1,20	9182	1,45	8857	1,70	9086	1,95	9799
1,21	9156	1,46	8856	1,71	9106	1,96	9837
1,22	9131	1,47	8856	1,72	9126	1,97	9877
1,23	9108	1,48	8857	1,73	9147	1,98	9917
1,24	9085	1,49	8859	1,74	9168	1,99	9958
1,25	9064	1,50	8862	1,75	9191	2,00	*0000

2 Poisson-Verteilung

Tafel 2. Wahrscheinlichkeitsfunktion $f(x)$ und Verteilungsfunktion $F(x)$ der Poisson-Verteilung

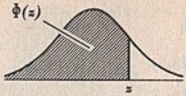
Vergleiche (42.2) und (42.3)

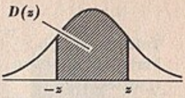
x	$\mu = 0,1$		$\mu = 0,2$		$\mu = 0,3$		$\mu = 0,4$		$\mu = 0,5$	
	$f(x)$	$F(x)$	$f(x)$	$F(x)$	$f(x)$	$F(x)$	$f(x)$	$F(x)$	$f(x)$	$F(x)$
	0,		0,		0,		0,		0,	
0	9048	0,9048	8187	0,8187	7408	0,7408	6703	0,6703	6065	0,6065
1	0905	0,9953	1637	0,9825	2222	0,9631	2681	0,9384	3033	0,9098
2	0045	0,9998	0164	0,9989	0333	0,9964	0536	0,9921	0758	0,9856
3	0002	1,0000	0011	0,9999	0033	0,9997	0072	0,9992	0126	0,9982
4	0000	1,0000	0001	1,0000	0003	1,0000	0007	0,9999	0016	0,9998
5							0001	1,0000	0002	1,0000

x	$\mu = 0,6$		$\mu = 0,7$		$\mu = 0,8$		$\mu = 0,9$		$\mu = 1$	
	$f(x)$	$F(x)$	$f(x)$	$F(x)$	$f(x)$	$F(x)$	$f(x)$	$F(x)$	$f(x)$	$F(x)$
	0,		0,		0,		0,		0,	
0	5488	0,5488	4966	0,4966	4493	0,4493	4066	0,4066	3679	0,3679
1	3293	0,8781	3476	0,8442	3595	0,8088	3659	0,7725	3679	0,7358
2	0988	0,9769	1217	0,9659	1438	0,9526	1647	0,9371	1839	0,9197
3	0198	0,9966	0284	0,9942	0383	0,9909	0494	0,9865	0613	0,9810
4	0030	0,9996	0050	0,9992	0077	0,9986	0111	0,9977	0153	0,9963
5	0004	1,0000	0007	0,9999	0012	0,9998	0020	0,9997	0031	0,9994
6			0001	1,0000	0002	1,0000	0003	1,0000	0005	0,9999
7									0001	1,0000

x	$\mu = 1,5$		$\mu = 2$		$\mu = 3$		$\mu = 4$		$\mu = 5$	
	$f(x)$	$F(x)$	$f(x)$	$F(x)$	$f(x)$	$F(x)$	$f(x)$	$F(x)$	$f(x)$	$F(x)$
	0,		0,		0,		0,		0,	
0	2231	0,2231	1353	0,1353	0498	0,0498	0183	0,0183	0067	0,0067
1	3347	0,5578	2707	0,4060	1494	0,1991	0733	0,0916	0337	0,0404
2	2510	0,8088	2707	0,6767	2240	0,4232	1465	0,2381	0842	0,1247
3	1255	0,9344	1804	0,8571	2240	0,6472	1954	0,4335	1404	0,2650
4	0471	0,9814	0902	0,9473	1680	0,8153	1954	0,6288	1755	0,4405
5	0141	0,9955	0361	0,9834	1008	0,9161	1563	0,7851	1755	0,6160
6	0035	0,9991	0120	0,9955	0504	0,9665	1042	0,8893	1462	0,7622
7	0008	0,9998	0034	0,9989	0216	0,9881	0595	0,9489	1044	0,8666
8	0001	1,0000	0009	0,9998	0081	0,9962	0298	0,9786	0653	0,9319
9			0002	1,0000	0027	0,9989	0132	0,9919	0363	0,9682
10					0008	0,9997	0053	0,9972	0181	0,9863
11					0002	0,9999	0019	0,9991	0082	0,9945
12					0001	1,0000	0006	0,9997	0034	0,9980
13							0002	0,9999	0013	0,9993
14							0001	1,0000	0005	0,9998
15									0002	0,9999
16									0000	1,0000

3 Normalverteilung





Tafel 3a.

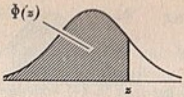
Verteilungsfunktion (48.3)

$$D(z) = \Phi(z) - \Phi(-z).$$

$$\Phi(-z) = 1 - \Phi(z), \quad \Phi(0) = 0{,}5$$

z	$\Phi(-z)$	$\Phi(z)$	$D(z)$	z	$\Phi(-z)$	$\Phi(z)$	$D(z)$	z	$\Phi(-z)$	$\Phi(z)$	$D(z)$
	0,	0,	0,		0,	0,	0,		0,	0,	0,
0,01	4960	5040	0080	0,51	3050	6950	3899	1,01	1562	8438	6875
0,02	4920	5080	0160	0,52	3015	6985	3969	1,02	1539	8461	6923
0,03	4880	5120	0239	0,53	2981	7019	4039	1,03	1515	8485	6970
0,04	4840	5160	0319	0,54	2946	7054	4108	1,04	1492	8508	7017
0,05	4801	5199	0399	0,55	2912	7088	4177	1,05	1469	8531	7063
0,06	4761	5239	0478	0,56	2877	7123	4245	1,06	1446	8554	7109
0,07	4721	5279	0558	0,57	2843	7157	4313	1,07	1423	8577	7154
0,08	4681	5319	0638	0,58	2810	7190	4381	1,08	1401	8599	7199
0,09	4641	5359	0717	0,59	2776	7224	4448	1,09	1379	8621	7243
0,10	4602	5398	0797	0,60	2743	7257	4515	1,10	1357	8643	7287
0,11	4562	5438	0876	0,61	2709	7291	4581	1,11	1335	8665	7330
0,12	4522	5478	0955	0,62	2676	7324	4647	1,12	1314	8686	7373
0,13	4483	5517	1034	0,63	2643	7357	4713	1,13	1292	8708	7415
0,14	4443	5557	1113	0,64	2611	7389	4778	1,14	1271	8729	7457
0,15	4404	5596	1192	0,65	2578	7422	4843	1,15	1251	8749	7499
0,16	4364	5636	1271	0,66	2546	7454	4907	1,16	1230	8770	7540
0,17	4325	5675	1350	0,67	2514	7486	4971	1,17	1210	8790	7580
0,18	4286	5714	1428	0,68	2483	7517	5035	1,18	1190	8810	7620
0,19	4247	5753	1507	0,69	2451	7549	5098	1,19	1170	8830	7660
0,20	4207	5793	1585	0,70	2420	7580	5161	1,20	1151	8849	7699
0,21	4168	5832	1663	0,71	2389	7611	5223	1,21	1131	8869	7737
0,22	4129	5871	1741	0,72	2358	7642	5285	1,22	1112	8888	7775
0,23	4090	5910	1819	0,73	2327	7673	5346	1,23	1093	8907	7813
0,24	4052	5948	1897	0,74	2296	7704	5407	1,24	1075	8925	7850
0,25	4013	5987	1974	0,75	2266	7734	5467	1,25	1056	8944	7887
0,26	3974	6026	2051	0,76	2236	7764	5527	1,26	1038	8962	7923
0,27	3936	6064	2128	0,77	2206	7794	5587	1,27	1020	8980	7959
0,28	3897	6103	2205	0,78	2177	7823	5646	1,28	1003	8997	7995
0,29	3859	6141	2282	0,79	2148	7852	5705	1,29	0985	9015	8029
0,30	3821	6179	2358	0,80	2119	7881	5763	1,30	0968	9032	8064
0,31	3783	6217	2434	0,81	2090	7910	5821	1,31	0951	9049	8098
0,32	3745	6255	2510	0,82	2061	7939	5878	1,32	0934	9066	8132
0,33	3707	6293	2586	0,83	2033	7967	5935	1,33	0918	9082	8165
0,34	3669	6331	2661	0,84	2005	7995	5991	1,34	0901	9099	8198
0,35	3632	6368	2737	0,85	1977	8023	6047	1,35	0885	9115	8230
0,36	3594	6406	2812	0,86	1949	8051	6102	1,36	0869	9131	8262
0,37	3557	6443	2886	0,87	1922	8078	6157	1,37	0853	9147	8293
0,38	3520	6480	2961	0,88	1894	8106	6211	1,38	0838	9162	8324
0,39	3483	6517	3035	0,89	1867	8133	6265	1,39	0823	9177	8355
0,40	3446	6554	3108	0,90	1841	8159	6319	1,40	0808	9192	8385
0,41	3409	6591	3182	0,91	1814	8186	6372	1,41	0793	9207	8415
0,42	3372	6628	3255	0,92	1788	8212	6424	1,42	0778	9222	8444
0,43	3336	6664	3328	0,93	1762	8238	6476	1,43	0764	9236	8473
0,44	3300	6700	3401	0,94	1736	8264	6528	1,44	0749	9251	8501
0,45	3264	6736	3473	0,95	1711	8289	6579	1,45	0735	9265	8529
0,46	3228	6772	3545	0,96	1685	8315	6629	1,46	0721	9279	8557
0,47	3192	6808	3616	0,97	1660	8340	6680	1,47	0708	9292	8584
0,48	3156	6844	3688	0,98	1635	8365	6729	1,48	0694	9306	8611
0,49	3121	6879	3759	0,99	1611	8389	6778	1,49	0681	9319	8638
0,50	3085	6915	3829	1,00	1587	8413	6827	1,50	0668	9332	8664

Tafel 3a.
Verteilungsfunktion (48.3)
(Fortsetzung)



z	$\Phi(-z)$	$\Phi(z)$	$D(z)$
	0,	0,	0,
1,51	0655	9345	8690
1,52	0643	9357	8715
1,53	0630	9370	8740
1,54	0618	9382	8764
1,55	0606	9394	8789
1,56	0594	9406	8812
1,57	0582	9418	8836
1,58	0571	9429	8859
1,59	0559	9441	8882
1,60	0548	9452	8904
1,61	0537	9463	8926
1,62	0526	9474	8948
1,63	0516	9484	8969
1,64	0505	9495	8990
1,65	0495	9505	9011
1,66	0485	9515	9031
1,67	0475	9525	9051
1,68	0465	9535	9070
1,69	0455	9545	9090
1,70	0446	9554	9109
1,71	0436	9564	9127
1,72	0427	9573	9146
1,73	0418	9582	9164
1,74	0409	9591	9181
1,75	0401	9599	9199
1,76	0392	9608	9216
1,77	0384	9616	9233
1,78	0375	9625	9249
1,79	0367	9633	9265
1,80	0359	9641	9281
1,81	0351	9649	9297
1,82	0344	9656	9312
1,83	0336	9664	9328
1,84	0329	9671	9342
1,85	0322	9678	9357
1,86	0314	9686	9371
1,87	0307	9693	9385
1,88	0301	9699	9399
1,89	0294	9706	9412
1,90	0287	9713	9426
1,91	0281	9719	9439
1,92	0274	9726	9451
1,93	0268	9732	9464
1,94	0262	9738	9476
1,95	0256	9744	9488
1,96	0250	9750	9500
1,97	0244	9756	9512
1,98	0239	9761	9523
1,99	0233	9767	9534
2,00	0228	9772	9545

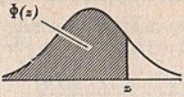
z	$\Phi(-z)$	$\Phi(z)$	$D(z)$
	0,	0,	0,
2,01	0222	9778	9556
2,02	0217	9783	9566
2,03	0212	9788	9576
2,04	0207	9793	9586
2,05	0202	9798	9596
2,06	0197	9803	9606
2,07	0192	9808	9615
2,08	0188	9812	9625
2,09	0183	9817	9634
2,10	0179	9821	9643
2,11	0174	9826	9651
2,12	0170	9830	9660
2,13	0166	9834	9668
2,14	0162	9838	9676
2,15	0158	9842	9684
2,16	0154	9846	9692
2,17	0150	9850	9700
2,18	0146	9854	9707
2,19	0143	9857	9715
2,20	0139	9861	9722
2,21	0136	9864	9729
2,22	0132	9868	9736
2,23	0129	9871	9743
2,24	0125	9875	9749
2,25	0122	9878	9756
2,26	0119	9881	9762
2,27	0116	9884	9768
2,28	0113	9887	9774
2,29	0110	9890	9780
2,30	0107	9893	9786
2,31	0104	9896	9791
2,32	0102	9898	9797
2,33	0099	9901	9802
2,34	0096	9904	9807
2,35	0094	9906	9812
2,36	0091	9909	9817
2,37	0089	9911	9822
2,38	0087	9913	9827
2,39	0084	9916	9832
2,40	0082	9918	9836
2,41	0080	9920	9840
2,42	0078	9922	9845
2,43	0075	9925	9849
2,44	0073	9927	9853
2,45	0071	9929	9857
2,46	0069	9931	9861
2,47	0068	9932	9865
2,48	0066	9934	9869
2,49	0064	9936	9872
2,50	0062	9938	9876

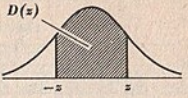
z	$\Phi(-z)$	$\Phi(z)$	$D(z)$
	0,	0,	0,
2,51	0060	9940	9879
2,52	0059	9941	9883
2,53	0057	9943	9886
2,54	0055	9945	9889
2,55	0054	9946	9892
2,56	0052	9948	9895
2,57	0051	9949	9898
2,58	0049	9951	9901
2,59	0048	9952	9904
2,60	0047	9953	9907
2,61	0045	9955	9909
2,62	0044	9956	9912
2,63	0043	9957	9915
2,64	0041	9959	9917
2,65	0040	9960	9920
2,66	0039	9961	9922
2,67	0038	9962	9924
2,68	0037	9963	9926
2,69	0036	9964	9929
2,70	0035	9965	9931
2,71	0034	9966	9933
2,72	0033	9967	9935
2,73	0032	9968	9937
2,74	0031	9969	9939
2,75	0030	9970	9940
2,76	0029	9971	9942
2,77	0028	9972	9944
2,78	0027	9973	9946
2,79	0026	9974	9947
2,80	0026	9974	9949
2,81	0025	9975	9950
2,82	0024	9976	9952
2,83	0023	9977	9953
2,84	0023	9977	9955
2,85	0022	9978	9956
2,86	0021	9979	9958
2,87	0021	9979	9959
2,88	0020	9980	9960
2,89	0019	9981	9961
2,90	0019	9981	9963
2,91	0018	9982	9964
2,92	0018	9982	9965
2,93	0017	9983	9966
2,94	0016	9984	9967
2,95	0016	9984	9968
2,96	0015	9985	9969
2,97	0015	9985	9970
2,98	0014	9986	9971
2,99	0014	9986	9972
3,00	0013	9987	9973

Tafel 3b. Normalverteilung. Werte von z zu gegebenen Werten von (48.3) und $D(z)$

$D(z) = \Phi(z) - \Phi(-z)$

Beispiel: $\Phi(z) = 61\%$ für $z = 0{,}279$, $D(z) = 61\%$ für $z = 0{,}860$





%	$z(\Phi)$	$z(D)$
1	−2,326	0,013
2	−2,054	0,025
3	−1,881	0,038
4	−1,751	0,050
5	−1,645	0,063
6	−1,555	0,075
7	−1,476	0,088
8	−1,405	0,100
9	−1,341	0,113
10	−1,282	0,126
11	−1,227	0,138
12	−1,175	0,151
13	−1,126	0,164
14	−1,080	0,176
15	−1,036	0,189
16	−0,994	0,202
17	−0,954	0,215
18	−0,915	0,228
19	−0,878	0,240
20	−0,842	0,253
21	−0,806	0,266
22	−0,772	0,279
23	−0,739	0,292
24	−0,706	0,305
25	−0,674	0,319
26	−0,643	0,332
27	−0,613	0,345
28	−0,583	0,358
29	−0,553	0,372
30	−0,524	0,385
31	−0,496	0,399
32	−0,468	0,412
33	−0,440	0,426
34	−0,412	0,440
35	−0,385	0,454
36	−0,358	0,468
37	−0,332	0,482
38	−0,305	0,496
39	−0,279	0,510
40	−0,253	0,524

%	$z(\Phi)$	$z(D)$
41	−0,228	0,539
42	−0,202	0,553
43	−0,176	0,568
44	−0,151	0,583
45	−0,126	0,598
46	−0,100	0,613
47	−0,075	0,628
48	−0,050	0,643
49	−0,025	0,659
50	0,000	0,674
51	0,025	0,690
52	0,050	0,706
53	0,075	0,722
54	0,100	0,739
55	0,126	0,755
56	0,151	0,772
57	0,176	0,789
58	0,202	0,806
59	0,228	0,824
60	0,253	0,842
61	0,279	0,860
62	0,305	0,878
63	0,332	0,896
64	0,358	0,915
65	0,385	0,935
66	0,412	0,954
67	0,440	0,974
68	0,468	0,994
69	0,496	1,015
70	0,524	1,036
71	0,553	1,058
72	0,583	1,080
73	0,613	1,103
74	0,643	1,126
75	0,674	1,150
76	0,706	1,175
77	0,739	1,200
78	0,772	1,227
79	0,806	1,254
80	0,842	1,282

%	$z(\Phi)$	$z(D)$
81	0,878	1,311
82	0,915	1,341
83	0,954	1,372
84	0,994	1,405
85	1,036	1,440
86	1,080	1,476
87	1,126	1,514
88	1,175	1,555
89	1,227	1,598
90	1,282	1,645
91	1,341	1,695
92	1,405	1,751
93	1,476	1,812
94	1,555	1,881
95	1,645	1,960
96	1,751	2,054
97	1,881	2,170
98	2,054	2,326
99	2,326	2,576
99,1	2,366	2,612
99,2	2,409	2,652
99,3	2,457	2,697
99,4	2,512	2,748
99,5	2,576	2,807
99,6	2,652	2,878
99,7	2,748	2,968
99,8	2,878	3,090
99,9	3,090	3,291
99,91	3,121	3,320
99,92	3,156	3,353
99,93	3,195	3,390
99,94	3,239	3,432
99,95	3,291	3,481
99,96	3,353	3,540
99,97	3,432	3,615
99,98	3,540	3,719
99,99	3,719	3,891

Tafel 3c. Wahrscheinlichkeitsdichte $f(z) = e^{-z^2/2}/\sqrt{2\pi}$

Vergleiche (48.3).

$$f(-z) = f(z), \quad f(0) = 1/\sqrt{2\pi} = 0{,}3989$$

z	$f(z)$	z	$f(z)$	z	$f(z)$	z	$f(z)$	z	$f(z)$
	0,		0,		0,		0,		0,
0,01	3989	0,41	3668	0,81	2874	1,21	1919	1,61	1092
0,02	3989	0,42	3653	0,82	2850	1,22	1895	1,62	1074
0,03	3988	0,43	3637	0,83	2827	1,23	1872	1,63	1057
0,04	3986	0,44	3621	0,84	2803	1,24	1849	1,64	1040
0,05	3984	0,45	3605	0,85	2780	1,25	1826	1,65	1023
0,06	3982	0,46	3589	0,86	2756	1,26	1804	1,66	1006
0,07	3980	0,47	3572	0,87	2732	1,27	1781	1,67	0989
0,08	3977	0,48	3555	0,88	2709	1,28	1758	1,68	0973
0,09	3973	0,49	3538	0,89	2685	1,29	1736	1,69	0957
0,10	3970	0,50	3521	0,90	2661	1,30	1714	1,70	0940
0,11	3965	0,51	3503	0,91	2637	1,31	1691	1,71	0925
0,12	3961	0,52	3485	0,92	2613	1,32	1669	1,72	0909
0,13	3956	0,53	3467	0,93	2589	1,33	1647	1,73	0893
0,14	3951	0,54	3448	0,94	2565	1,34	1626	1,74	0878
0,15	3945	0,55	3429	0,95	2541	1,35	1604	1,75	0863
0,16	3939	0,56	3410	0,96	2516	1,36	1582	1,76	0848
0,17	3932	0,57	3391	0,97	2492	1,37	1561	1,77	0833
0,18	3925	0,58	3372	0,98	2468	1,38	1539	1,78	0818
0,19	3918	0,59	3352	0,99	2444	1,39	1518	1,79	0804
0,20	3910	0,60	3332	1,00	2420	1,40	1497	1,80	0790
0,21	3902	0,61	3312	1,01	2396	1,41	1476	1,81	0775
0,22	3894	0,62	3292	1,02	2371	1,42	1456	1,82	0761
0,23	3885	0,63	3271	1,03	2347	1,43	1435	1,83	0748
0,24	3876	0,64	3251	1,04	2323	1,44	1415	1,84	0734
0,25	3867	0,65	3230	1,05	2299	1,45	1394	1,85	0721
0,26	3857	0,66	3209	1,06	2275	1,46	1374	1,86	0707
0,27	3847	0,67	3187	1,07	2251	1,47	1354	1,87	0694
0,28	3836	0,68	3166	1,08	2227	1,48	1334	1,88	0681
0,29	3825	0,69	3144	1,09	2203	1,49	1315	1,89	0669
0,30	3814	0,70	3123	1,10	2179	1,50	1295	1,90	0656
0,31	3802	0,71	3101	1,11	2155	1,51	1276	1,91	0644
0,32	3790	0,72	3079	1,12	2131	1,52	1257	1,92	0632
0,33	3778	0,73	3056	1,13	2107	1,53	1238	1,93	0620
0,34	3765	0,74	3034	1,14	2083	1,54	1219	1,94	0608
0,35	3752	0,75	3011	1,15	2059	1,55	1200	1,95	0596
0,36	3739	0,76	2989	1,16	2036	1,56	1182	1,96	0584
0,37	3725	0,77	2966	1,17	2012	1,57	1163	1,97	0573
0,38	3712	0,78	2943	1,18	1989	1,58	1145	1,98	0562
0,39	3697	0,79	2920	1,19	1965	1,59	1127	1,99	0551
0,40	3683	0,80	2897	1,20	1942	1,60	1109	2,00	0540

4 Quadratwurzeln. Quadrate. Exponentialfunktion. Konstanten

Tafel 4a. Quadratwurzeln und Quadrate

n	$\sqrt{n}$	$\sqrt{10n}$	n^2	n	$\sqrt{n}$	$\sqrt{10n}$	n^2
1	1,000 000	3,162 278	1	51	7,141 428	22,583 180	2601
2	1,414 214	4,472 136	4	52	7,211 103	22,803 509	2704
3	1,732 051	5,477 226	9	53	7,280 110	23,021 729	2809
4	2,000 000	6,324 555	16	54	7,348 469	23,237 900	2916
5	2,236 068	7,071 068	25	55	7,416 198	23,452 079	3025
6	2,449 490	7,745 967	36	56	7,483 315	23,664 319	3136
7	2,645 751	8,366 600	49	57	7,549 834	23,874 673	3249
8	2,828 427	8,944 272	64	58	7,615 773	24,083 189	3364
9	3,000 000	9,486 833	81	59	7,681 146	24,289 916	3481
10	3,162 278	10,000 000	100	60	7,745 967	24,494 897	3600
11	3,316 625	10,488 088	121	61	7,810 250	24,698 178	3721
12	3,464 102	10,954 451	144	62	7,874 008	24,899 799	3844
13	3,605 551	11,401 754	169	63	7,937 254	25,099 801	3969
14	3,741 657	11,832 160	196	64	8,000 000	25,298 221	4096
15	3,872 983	12,247 449	225	65	8,062 258	25,495 098	4225
16	4,000 000	12,649 111	256	66	8,124 038	25,690 465	4356
17	4,123 106	13,038 405	289	67	8,185 353	25,884 358	4489
18	4,242 641	13,416 408	324	68	8,246 211	26,076 810	4624
19	4,358 899	13,784 049	361	69	8,306 624	26,267 851	4761
20	4,472 136	14,142 136	400	70	8,366 600	26,457 513	4900
21	4,582 576	14,491 377	441	71	8,426 150	26,645 825	5041
22	4,690 416	14,832 397	484	72	8,485 281	26,832 816	5184
23	4,795 832	15,165 751	529	73	8,544 004	27,018 512	5329
24	4,898 979	15,491 933	576	74	8,602 325	27,202 941	5476
25	5,000 000	15,811 388	625	75	8,660 254	27,386 128	5625
26	5,099 020	16,124 515	676	76	8,717 798	27,568 098	5776
27	5,196 152	16,431 677	729	77	8,774 964	27,748 874	5929
28	5,291 503	16,733 201	784	78	8,831 761	27,928 480	6084
29	5,385 165	17,029 386	841	79	8,888 194	28,106 939	6241
30	5,477 226	17,320 508	900	80	8,944 272	28,284 271	6400
31	5,567 764	17,606 817	961	81	9,000 000	28,460 499	6561
32	5,656 854	17,888 544	1024	82	9,055 385	28,635 642	6724
33	5,744 563	18,165 902	1089	83	9,110 434	28,809 721	6889
34	5,830 952	18,439 089	1156	84	9,165 151	28,982 753	7056
35	5,916 080	18,708 287	1225	85	9,219 544	29,154 759	7225
36	6,000 000	18,973 666	1296	86	9,273 618	29,325 757	7396
37	6,082 763	19,235 384	1369	87	9,327 379	29,495 762	7569
38	6,164 414	19,493 589	1444	88	9,380 832	29,664 794	7744
39	6,244 998	19,748 418	1521	89	9,433 981	29,832 868	7921
40	6,324 555	20,000 000	1600	90	9,486 833	30,000 000	8100
41	6,403 124	20,248 457	1681	91	9,539 392	30,166 206	8281
42	6,480 741	20,493 902	1764	92	9,591 663	30,331 502	8464
43	6,557 439	20,736 441	1849	93	9,643 651	30,495 901	8649
44	6,633 250	20,976 177	1936	94	9,695 360	30,659 419	8836
45	6,708 204	21,213 203	2025	95	9,746 794	30,822 070	9025
46	6,782 330	21,447 611	2116	96	9,797 959	30,983 867	9216
47	6,855 655	21,679 483	2209	97	9,848 858	31,144 823	9409
48	6,928 203	21,908 902	2304	98	9,899 495	31,304 952	9604
49	7,000 000	22,135 944	2401	99	9,949 874	31,464 265	9801
50	7,071 068	22,360 680	2500	100	10,000 000	31,622 777	10000

Tafel 4b. Exponentialfunktion und hyperbolischer Tangens

x	e^x	$\tanh x$	x	e^x	$\tanh x$	x	e^x	$\tanh x$
0,00	1,0000	0,00000	0,50	1,6487	0,46212	1,0	2,7183	0,76159
0,01	1,0101	0,01000	0,51	1,6653	0,46995	1,1	3,0042	0,80050
0,02	1,0202	0,02000	0,52	1,6820	0,47770	1,2	3,3201	0,83365
0,03	1,0305	0,02999	0,53	1,6989	0,48538	1,3	3,6693	0,86172
0,04	1,0408	0,03998	0,54	1,7160	0,49299	1,4	4,0552	0,88535
0,05	1,0513	0,04996	0,55	1,7333	0,50052	1,5	4,4817	0,90515
0,06	1,0618	0,05993	0,56	1,7507	0,50798	1,6	4,9530	0,92167
0,07	1,0725	0,06989	0,57	1,7683	0,51536	1,7	5,4739	0,93541
0,08	1,0833	0,07983	0,58	1,7860	0,52267	1,8	6,0496	0,94681
0,09	1,0942	0,08976	0,59	1,8040	0,52990	1,9	6,6859	0,95624
0,10	1,1052	0,09967	0,60	1,8221	0,53705	2,0	7,3891	0,96403
0,11	1,1163	0,10956	0,61	1,8404	0,54413	2,1	8,1662	0,97045
0,12	1,1275	0,11943	0,62	1,8589	0,55113	2,2	9,0250	0,97574
0,13	1,1388	0,12927	0,63	1,8776	0,55805	2,3	9,9742	0,98010
0,14	1,1503	0,13909	0,64	1,8965	0,56490	2,4	11,023	0,98367
0,15	1,1618	0,14889	0,65	1,9155	0,57167	2,5	12,182	0,98661
0,16	1,1735	0,15865	0,66	1,9348	0,57836	2,6	13,464	0,98903
0,17	1,1853	0,16838	0,67	1,9542	0,58498	2,7	14,880	0,99101
0,18	1,1972	0,17808	0,68	1,9739	0,59152	2,8	16,445	0,99263
0,19	1,2092	0,18775	0,69	1,9937	0,59798	2,9	18,174	0,99396
0,20	1,2214	0,19738	0,70	2,0138	0,60437	3,0	20,086	0,99505
0,21	1,2337	0,20697	0,71	2,0340	0,61068	3,1	22,198	0,99595
0,22	1,2461	0,21652	0,72	2,0544	0,61691	3,2	24,533	0,99668
0,23	1,2586	0,22603	0,73	2,0751	0,62307	3,3	27,113	0,99728
0,24	1,2712	0,23550	0,74	2,0959	0,62915	3,4	29,964	0,99777
0,25	1,2840	0,24492	0,75	2,1170	0,63515	3,5	33,115	0,99818
0,26	1,2969	0,25430	0,76	2,1383	0,64108	3,6	36,598	0,99851
0,27	1,3100	0,26362	0,77	2,1598	0,64693	3,7	40,447	0,99878
0,28	1,3231	0,27291	0,78	2,1815	0,65271	3,8	44,701	0,99900
0,29	1,3364	0,28213	0,79	2,2034	0,65841	3,9	49,402	0,99918
0,30	1,3499	0,29131	0,80	2,2255	0,66404	4,0	54,598	0,99933
0,31	1,3634	0,30044	0,81	2,2479	0,66959	4,1	60,340	0,99945
0,32	1,3771	0,30951	0,82	2,2705	0,67507	4,2	66,686	0,99955
0,33	1,3910	0,31852	0,83	2,2933	0,68048	4,3	73,700	0,99963
0,34	1,4049	0,32748	0,84	2,3164	0,68581	4,4	81,451	0,99970
0,35	1,4191	0,33638	0,85	2,3396	0,69107	4,5	90,017	0,99975
0,36	1,4333	0,34521	0,86	2,3632	0,69626	4,6	99,484	0,99980
0,37	1,4477	0,35399	0,87	2,3869	0,70137	4,7	109,95	0,99983
0,38	1,4623	0,36271	0,88	2,4109	0,70642	4,8	121,51	0,99986
0,39	1,4770	0,37136	0,89	2,4351	0,71139	4,9	134,29	0,99989
0,40	1,4918	0,37995	0,90	2,4596	0,71630	5,0	148,41	0,99991
0,41	1,5068	0,38847	0,91	2,4843	0,72113	5,1	164,02	0,99993
0,42	1,5220	0,39693	0,92	2,5093	0,72590	5,2	181,27	0,99994
0,43	1,5373	0,40532	0,93	2,5345	0,73059	5,3	200,34	0,99995
0,44	1,5527	0,41364	0,94	2,5600	0,73522	5,4	221,41	0,99996
0,45	1,5683	0,42190	0,95	2,5857	0,73978	5,5	244,69	0,99997
0,46	1,5841	0,43008	0,96	2,6117	0,74428	5,6	270,43	0,99997
0,47	1,6000	0,43820	0,97	2,6379	0,74870	5,7	298,87	0,99998
0,48	1,6161	0,44624	0,98	2,6645	0,75307	5,8	330,30	0,99998
0,49	1,6323	0,45422	0,99	2,6912	0,75736	5,9	365,04	0,99998
0,50	1,6487	0,46212	1,00	2,7183	0,76159	6,0	403,43	0,99999

Tafel 4c. Einige wichtige Zahlenwerte

$e = 2{,}718\ 281\ 828 = 1/0{,}367\ 879\ 441$

$e^2 = 7{,}389\ 056\ 099 = 1/0{,}135\ 335\ 283$

$\sqrt{e} = 1{,}648\ 721\ 271 = 1/0{,}606\ 530\ 660$

$\pi = 3{,}141\ 592\ 654 = 1/0{,}318\ 309\ 886$

$\pi^2 = 9{,}869\ 604\ 401 = 1/0{,}101\ 321\ 184$

$\sqrt{\pi} = 1{,}772\ 453\ 851 = 1/0{,}564\ 189\ 584$

$\sqrt{2\pi} = 2{,}506\ 628\ 275 = 1/0{,}398\ 942\ 280$

$\sqrt{\pi/2} = 1{,}253\ 314\ 137 = 1/0{,}797\ 884\ 561$

$\sqrt[3]{\pi} = 1{,}464\ 591\ 888 = 1/0{,}682\ 784\ 063$

$e^{\pi} = 23{,}140\ 692\ 633 = 1/0{,}043\ 213\ 918$

$e^{\pi/2} = 4{,}810\ 477\ 381 = 1/0{,}207\ 879\ 576$

$e^{\pi/4} = 2{,}193\ 280\ 051 = 1/0{,}455\ 938\ 128$

$\log_{10} \pi = 0{,}497\ 149\ 873$

$\log_e \pi = 1{,}144\ 729\ 886$

$\sqrt{2} = 1{,}414\ 213\ 562 = 1/0{,}707\ 106\ 781$

$\sqrt{3} = 1{,}732\ 050\ 808 = 1/0{,}577\ 350\ 269$

$\sqrt{10} = 3{,}162\ 277\ 660 = 1/0{,}316\ 227\ 766$

$\sqrt[3]{2} = 1{,}259\ 921\ 050 = 1/0{,}793\ 700\ 526$

$\sqrt[3]{10} = 2{,}154\ 434\ 690 = 1/0{,}464\ 158\ 883$

$\log_e 2 = 0{,}693\ 147\ 181$

$\log_e 3 = 1{,}098\ 612\ 289$

$\log_e 10 = 2{,}302\ 585\ 093 = 1/0{,}434\ 294\ 482$

5 Hilfstafel für die Zufallsauswahl

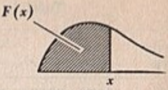
Tafel 5. Zufällig angeordnete Ziffern

Zeile Nr.	Spalte Nr. 0	1	2	3	4	5	6	7	8	9
0	87331	82442	28104	26432	83640	17323	68764	84728	37995	96106
1	33628	17364	01409	87803	65641	33433	48944	64299	79066	31777
2	54680	13427	72496	16967	16195	96593	55040	53729	62035	66717
3	51199	49794	49407	10774	98140	83891	37195	24066	61140	65144
4	78702	98067	61313	91661	59861	54437	77739	19892	54817	88645
5	55672	16014	24892	13089	00410	81458	76156	28189	40595	21500
6	18880	58497	03862	32368	59320	24807	63392	79793	63043	09425
7	10242	62548	62330	05703	33535	49128	66298	16193	55301	01306
8	54993	17182	94618	23228	83895	73251	68199	64639	83178	70521
9	22686	50885	16006	04041	08077	33065	35237	02502	94755	72062
10	42349	03145	15770	70665	53291	32288	41568	66079	98705	31029
11	18093	09553	39428	75464	71329	86344	80729	40916	18860	51780
12	11535	03924	84252	74795	40193	84597	42497	21918	91384	84721
13	35066	73848	65351	53270	67341	70177	92373	17604	42204	60476
14	57477	22809	73558	96182	96779	01604	25748	59553	64876	94611
15	48647	33850	52956	45410	88212	05120	99391	32276	55961	41775
16	86857	81154	22223	74950	53296	67767	55866	49061	66937	81818
17	20182	36907	94644	99122	09774	29189	27212	79000	50217	71077
18	83687	31231	01133	41432	54542	60204	81618	09586	34481	87683
19	81315	12390	46074	47810	90171	36313	95440	77583	28506	38808
20	87026	52826	58341	76549	04105	66191	12914	55348	07907	06978
21	34301	76733	07251	90524	21931	83695	41340	53581	64582	60210
22	70734	24337	32674	49508	49751	90489	63202	24380	77943	09942
23	94710	31527	73445	32839	68176	53580	85250	53243	03350	00128
24	76462	16987	07775	43162	11777	16810	75158	13894	88945	15539
25	14348	28403	79245	69023	34196	46398	05964	64715	11330	17515
26	74618	89317	30146	25606	94507	98104	04239	44973	37636	88866
27	99442	19200	85406	45358	86253	60638	38858	44964	54103	57287
28	26869	44399	89452	06652	31271	00647	46551	83050	92058	83814
29	80988	08149	50499	98584	28385	63680	44638	91864	96002	87802
30	07511	79047	89289	17774	67194	37362	85684	55505	97809	67056
31	49779	12138	05048	03535	27502	63308	10218	53296	48687	61340
32	47938	55945	24003	19635	17471	65997	85906	98694	56420	78357
33	15604	06626	14360	79542	13512	87595	08542	03800	35443	52823
34	12307	27726	21864	00045	16075	03770	86978	52718	02693	09096
35	02450	28053	66134	99445	91316	25727	89399	85272	67148	78358
36	57623	54382	35236	89244	27245	90500	75430	96762	71968	65838
37	91762	78849	93105	40481	99431	03304	21079	86459	21287	76566
38	87373	31137	31128	67050	34309	44914	80711	61738	61498	24288
39	67094	41485	54149	86088	10192	21174	39948	67268	29938	32476
40	94456	66747	76922	87627	71834	57688	04878	78348	68970	60048
41	68359	75292	27710	86889	81678	79798	58360	39175	75667	65782
42	52393	31404	32584	06837	79762	13168	76055	54833	22841	98889
43	59565	91254	11847	20672	37625	41454	86861	55824	79793	74575
44	48185	11066	20162	38230	16043	48409	47421	21195	98008	57305
45	19230	12187	86659	12971	52204	76546	63272	19312	81662	96557
46	84327	21942	81727	68735	89190	58491	55329	96875	19465	89687
47	77430	71210	00591	50124	12030	50280	12358	76174	48353	09682
48	12462	19108	70512	53926	25595	97085	03833	59806	12351	64253
49	11684	06644	57816	10078	45021	47751	38285	73520	08434	65627

Tafel 5. Zufällig angeordnete Ziffern (Fortsetzung)

Zeile Nr.	Spalte Nr. 0	1	2	3	4	5	6	7	8	9
50	12896	36576	68686	08462	65652	76571	70891	09007	04581	01684
51	59090	05111	27587	90349	30789	50304	70650	06646	70126	15284
52	42486	67483	65282	19037	80588	73076	41820	46651	40442	40718
53	88662	03928	03249	85910	97533	88643	29829	21557	47328	36724
54	69403	03626	92678	53460	15465	83516	54012	80509	55976	46115
55	56434	70543	38696	98502	32092	95505	62091	39549	30117	98209
56	58227	62694	42837	29183	11393	68463	25150	86338	95620	39836
57	41272	94927	15413	40505	33123	63218	72940	98349	57249	40170
58	36819	01162	30425	15546	16065	68459	35776	64276	92868	07372
59	31700	66711	26115	55755	33584	18091	38709	57276	74660	90392
60	69855	63699	36839	90531	97125	87875	62824	03889	12538	24740
61	44322	17569	45439	41455	34324	90902	07978	26268	04279	76816
62	62226	36661	87011	66267	78777	78044	40819	49496	39814	73867
63	27284	19737	98741	72531	52741	26699	98755	19657	08665	16818
64	88341	21652	94743	77268	79525	44769	66583	30621	90534	62050
65	53266	18783	51903	56711	38060	69513	61963	80470	88018	86510
66	50527	49330	24839	42529	03944	95219	88724	37247	84166	23023
67	15655	07852	77206	35944	71446	30573	19405	57824	23576	23301
68	62057	22206	03314	83465	57466	10465	19891	32308	01900	67484
69	41769	56091	19892	96253	92808	45785	52774	49674	68103	65032
70	25993	72416	44473	41299	93095	17338	69802	98548	02429	85238
71	22842	57871	04470	37373	34516	04042	04078	35336	34393	97573
72	55704	31982	05234	22664	22181	40358	28089	15790	33340	18852
73	94258	18706	09437	96041	90052	80862	20420	24323	11635	91677
74	74145	20453	29657	98868	56695	53483	87449	35060	98942	62697
75	88881	12673	73961	89884	73247	97670	69570	88888	58560	72580
76	01508	56780	52223	35632	73347	71317	46541	88023	36656	76332
77	92069	43000	23233	06058	82527	25250	27555	20426	60361	63525
78	53366	35249	02117	68620	39388	69795	73215	01846	16983	78560
79	88057	54097	49511	74867	32192	90071	04147	46094	63519	07199
80	85492	82238	02668	91854	86149	28590	77853	81035	45561	16032
81	39453	62123	69611	53017	34964	09786	24614	49514	01056	18700
82	82627	98111	93870	56969	69566	62662	07353	84838	14570	14508
83	61142	51743	38209	31474	96095	15163	54380	77849	20465	03142
84	12031	32528	61311	53730	89032	16124	58844	35386	45521	59368
85	31313	59838	29147	76882	74328	09955	63673	96651	53264	29871
86	50767	41056	97409	44376	62219	35439	70102	99248	71179	26052
87	30522	95699	84966	26554	24768	72247	84993	85375	92518	16334
88	74176	19870	89874	64799	03792	57006	57225	36677	46825	14087
89	17114	93248	37065	91346	04657	93763	92210	43676	44944	75798
90	53005	11825	64608	87587	05742	31914	55044	41818	29667	77424
91	31985	81539	79942	49471	46200	27639	94099	42085	79231	03932
92	63499	60508	77522	15624	15088	78519	52279	79214	43623	69166
93	30506	42444	99047	66010	91657	37160	37408	85714	21420	80996
94	78248	16841	92357	10130	68990	38307	61022	56806	81016	38511
95	64996	84789	50185	32200	64382	29752	11876	00664	54547	62597
96	11963	13157	09136	01769	30117	71486	80111	09161	08371	71749
97	44335	91450	43456	90449	18338	19787	31339	60473	06606	89788
98	42277	11868	44520	01113	11341	11743	97949	49718	99176	42006
99	77562	18863	58515	90166	78508	14864	19111	57183	85808	59385

6 Chi-Quadrat-Verteilung

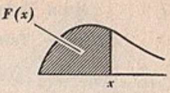


Tafel 6. Werte von x zu gegebenen Werten der Verteilungsfunktion (60.3)

Beispiel: Bei 3 Freiheitsgraden ist $F = 0,99$ für $x = 11,34$.

$F(x)$	Anzahl der Freiheitsgrade									
	1	2	3	4	5	6	7	8	9	10
0,001	0,00	0,00	0,02	0,09	0,21	0,38	0,60	0,86	1,15	1,48
0,005	0,00	0,01	0,07	0,21	0,41	0,68	0,99	1,34	1,73	2,16
0,01	0,00	0,02	0,11	0,30	0,55	0,87	1,24	1,65	2,09	2,56
0,025	0,00	0,05	0,22	0,48	0,83	1,24	1,69	2,18	2,70	3,25
0,05	0,00	0,10	0,35	0,71	1,15	1,64	2,17	2,73	3,33	3,94
0,1	0,02	0,21	0,58	1,06	1,61	2,20	2,83	3,49	4,17	4,87
0,25	0,10	0,58	1,21	1,92	2,67	3,45	4,25	5,07	5,90	6,74
0,5	0,45	1,39	2,37	3,36	4,35	5,35	6,35	7,34	8,34	9,34
0,75	1,32	2,77	4,11	5,39	6,63	7,84	9,04	10,22	11,39	12,55
0,9	2,71	4,61	6,25	7,78	9,24	10,64	12,02	13,36	14,68	15,99
0,95	3,84	5,99	7,81	9,49	11,07	12,59	14,07	15,51	16,92	18,31
0,975	5,02	7,38	9,35	11,14	12,83	14,45	16,01	17,53	19,02	20,48
0,99	6,63	9,21	11,34	13,28	15,09	16,81	18,48	20,09	21,67	23,21
0,995	7,88	10,60	12,84	14,86	16,75	18,55	20,28	21,96	23,59	25,19
0,999	10,83	13,82	16,27	18,47	20,52	22,46	24,32	26,13	27,88	29,59

$F(x)$	Anzahl der Freiheitsgrade									
	11	12	13	14	15	16	17	18	19	20
0,001	1,83	2,21	2,62	3,04	3,48	3,94	4,42	4,90	5,41	5,92
0,005	2,60	3,07	3,57	4,07	4,60	5,14	5,70	6,26	6,84	7,43
0,01	3,05	3,57	4,11	4,66	5,23	5,81	6,41	7,01	7,63	8,26
0,025	3,82	4,40	5,01	5,63	6,26	6,91	7,56	8,23	8,91	9,59
0,05	4,57	5,23	5,89	6,57	7,26	7,96	8,67	9,39	10,12	10,85
0,1	5,58	6,30	7,04	7,79	8,55	9,31	10,09	10,86	11,65	12,44
0,25	7,58	8,44	9,30	10,17	11,04	11,91	12,79	13,68	14,56	15,45
0,5	10,34	11,34	12,34	13,34	14,34	15,34	16,34	17,34	18,34	19,34
0,75	13,70	14,85	15,98	17,12	18,25	19,37	20,49	21,60	22,72	23,83
0,9	17,28	18,55	19,81	21,06	22,31	23,54	24,77	25,99	27,20	28,41
0,95	19,68	21,03	22,36	23,68	25,00	26,30	27,59	28,87	30,14	31,41
0,975	21,92	23,34	24,74	26,12	27,49	28,85	30,19	31,53	32,85	34,17
0,99	24,73	26,22	27,69	29,14	30,58	32,00	33,41	34,81	36,19	37,57
0,995	26,76	28,30	29,82	31,32	32,80	34,27	35,72	37,16	38,58	40,00
0,999	31,26	32,91	34,53	36,12	37,70	39,25	40,79	42,31	43,82	45,32



Tafel 6. Werte von x zu gegebenen Werten der Verteilungsfunktion (60.3)

(Fortsetzung)

$F(x)$	Anzahl der Freiheitsgrade									
	21	22	23	24	25	26	27	28	29	30
0,001	6,4	7,0	7,5	8,1	8,7	9,2	9,8	10,4	11,0	11,6
0,005	8,0	8,6	9,3	9,9	10,5	11,2	11,8	12,5	13,1	13,8
0,01	8,9	9,5	10,2	10,9	11,5	12,2	12,9	13,6	14,3	15,0
0,025	10,3	11,0	11,7	12,4	13,1	13,8	14,6	15,3	16,0	16,8
0,05	11,6	12,3	13,1	13,8	14,6	15,4	16,2	16,9	17,7	18,5
0,1	13,2	14,0	14,8	15,7	16,5	17,3	18,1	18,9	19,8	20,6
0,25	16,3	17,2	18,1	19,0	19,9	20,8	21,7	22,7	23,6	24,5
0,5	20,3	21,3	22,3	23,3	24,3	25,3	26,3	27,3	28,3	29,3
0,75	24,9	26,0	27,1	28,2	29,3	30,4	31,5	32,6	33,7	34,8
0,9	29,6	30,8	32,0	33,2	34,4	35,6	36,7	37,9	39,1	40,3
0,95	32,7	33,9	35,2	36,4	37,7	38,9	40,1	41,3	42,6	43,8
0,975	35,5	36,8	38,1	39,4	40,6	41,9	43,2	44,5	45,7	47,0
0,99	38,9	40,3	41,6	43,0	44,3	45,6	47,0	48,3	49,6	50,9
0,995	41,4	42,8	44,2	45,6	46,9	48,3	49,6	51,0	52,3	53,7
0,999	46,8	48,3	49,7	51,2	52,6	54,1	55,5	56,9	58,3	59,7

$F(x)$	Anzahl der Freiheitsgrade							
	40	50	60	70	80	90	100	> 100 (Näherung)
0,001	17,9	24,7	31,7	39,0	46,5	54,2	61,9	$\frac{1}{2}(h - 3{,}09)^2$
0,005	20,7	28,0	35,5	43,3	51,2	59,2	67,3	$\frac{1}{2}(h - 2{,}58)^2$
0,01	22,2	29,7	37,5	45,4	53,5	61,8	70,1	$\frac{1}{2}(h - 2{,}33)^2$
0,025	24,4	32,4	40,5	48,8	57,2	65,6	74,2	$\frac{1}{2}(h - 1{,}96)^2$
0,05	26,5	34,8	43,2	51,7	60,4	69,1	77,9	$\frac{1}{2}(h - 1{,}64)^2$
0,1	29,1	37,7	46,5	55,3	64,3	73,3	82,4	$\frac{1}{2}(h - 1{,}28)^2$
0,25	33,7	42,9	52,3	61,7	71,1	80,6	90,1	$\frac{1}{2}(h - 0{,}67)^2$
0,5	39,3	49,3	59,3	69,3	79,3	89,3	99,3	$\frac{1}{2}h^2$
0,75	45,6	56,3	67,0	77,6	88,1	98,6	109,1	$\frac{1}{2}(h + 0{,}67)^2$
0,9	51,8	63,2	74,4	85,5	96,6	107,6	118,5	$\frac{1}{2}(h + 1{,}28)^2$
0,95	55,8	67,5	79,1	90,5	101,9	113,1	124,3	$\frac{1}{2}(h + 1{,}64)^2$
0,975	59,3	71,4	83,3	95,0	106,6	118,1	129,6	$\frac{1}{2}(h + 1{,}96)^2$
0,99	63,7	76,2	88,4	100,4	112,3	124,1	135,8	$\frac{1}{2}(h + 2{,}33)^2$
0,995	66,8	79,5	92,0	104,2	116,3	128,3	140,2	$\frac{1}{2}(h + 2{,}58)^2$
0,999	73,4	86,7	99,6	112,3	124,8	137,2	149,4	$\frac{1}{2}(h + 3{,}09)^2$

In der letzten Spalte ist $h = \sqrt{2m - 1}$

(m = Anzahl der Freiheitsgrade)

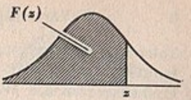
7 Zum Kolmogorov-Smirnow-Test

Tafel 7. Lösungen c der Gleichung (86.1)

n = Stichprobenumfang

n	$\alpha = 20\%$	$\alpha = 10\%$	$\alpha = 5\%$	$\alpha = 2\%$	$\alpha = 1\%$
	0,	0,	0,	0,	0,
1	900	950	975	990	995
2	684	776	842	900	929
3	565	636	708	785	829
4	493	565	624	689	734
5	447	509	563	627	669
6	410	468	519	577	617
7	381	436	483	538	576
8	359	410	454	507	542
9	339	387	430	480	513
10	323	369	409	457	486
11	308	352	391	437	468
12	296	338	375	419	449
13	285	325	361	404	432
14	275	314	349	390	418
15	266	304	338	377	404
16	258	295	327	366	392
17	250	286	318	355	381
18	244	279	309	346	371
19	237	271	301	337	361
20	232	265	294	329	352
21	226	259	287	321	344
22	221	253	281	314	337
23	216	247	275	307	330
24	212	242	269	301	323
25	208	238	264	295	317
26	204	233	259	290	311
27	200	229	254	284	305
28	197	225	250	279	300
29	193	221	246	275	295
30	190	218	242	270	290
35	177	202	224	251	269
40	165	189	210	235	252
45	156	179	198	222	238
50	148	170	188	211	226
55	142	162	180	201	216
60	136	155	172	193	207
65	131	149	166	185	199
70	126	144	160	179	192
75	122	139	154	173	185
80	118	135	150	167	179
85	114	131	145	162	174
90	111	127	141	158	169
95	108	124	137	154	165
100	106	121	134	150	161
Näherung für große n	$1{,}07/\sqrt{n}$	$1{,}22/\sqrt{n}$	$1{,}36/\sqrt{n}$	$1{,}52/\sqrt{n}$	$1{,}63/\sqrt{n}$

8 Students t-Verteilung



Tafel 8. Werte von z zu gegebenen Werten der Verteilungsfunktion (62.3)

Beispiel: Bei 9 Freiheitsgraden ist $F(z) = 0{,}95$ für $z = 1{,}83$.

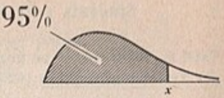
$$F(-z) = 1 - F(z)$$

$F(z)$	Anzahl der Freiheitsgrade									
	1	2	3	4	5	6	7	8	9	10
0,5	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
0,6	0,33	0,29	0,28	0,27	0,27	0,27	0,26	0,26	0,26	0,26
0,7	0,73	0,62	0,58	0,57	0,56	0,55	0,55	0,55	0,54	0,54
0,8	1,38	1,06	0,98	0,94	0,92	0,91	0,90	0,89	0,88	0,88
0,9	3,08	1,89	1,64	1,53	1,48	1,44	1,42	1,40	1,38	1,37
0,95	6,31	2,92	2,35	2,13	2,02	1,94	1,90	1,86	1,83	1,81
0,975	12,7	4,30	3,18	2,78	2,57	2,45	2,37	2,31	2,26	2,23
0,99	31,8	6,97	4,54	3,75	3,37	3,14	3,00	2,90	2,82	2,76
0,995	63,7	9,93	5,84	4,60	4,03	3,71	3,50	3,36	3,25	3,17
0,999	318,3	22,3	10,2	7,17	5,89	5,21	4,79	4,50	4,30	4,14

$F(z)$	Anzahl der Freiheitsgrade									
	11	12	13	14	15	16	17	18	19	20
0,5	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
0,6	0,26	0,26	0,26	0,26	0,26	0,26	0,26	0,26	0,26	0,26
0,7	0,54	0,54	0,54	0,54	0,54	0,54	0,53	0,53	0,53	0,53
0,8	0,88	0,87	0,87	0,87	0,87	0,87	0,86	0,86	0,86	0,86
0,9	1,36	1,36	1,35	1,35	1,34	1,34	1,33	1,33	1,33	1,33
0,95	1,80	1,78	1,77	1,76	1,75	1,75	1,74	1,73	1,73	1,73
0,975	2,20	2,18	2,16	2,15	2,13	2,12	2,11	2,10	2,09	2,09
0,99	2,72	2,68	2,65	2,62	2,60	2,58	2,57	2,55	2,54	2,53
0,995	3,11	3,06	3,01	2,98	2,95	2,92	2,90	2,88	2,86	2,85
0,999	4,03	3,93	3,85	3,79	3,73	3,69	3,65	3,61	3,58	3,55

$F(z)$	Anzahl der Freiheitsgrade									
	22	24	26	28	30	40	50	100	200	∞
0,5	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
0,6	0,26	0,26	0,26	0,26	0,26	0,26	0,26	0,25	0,25	0,25
0,7	0,53	0,53	0,53	0,53	0,53	0,53	0,53	0,53	0,53	0,52
0,8	0,86	0,86	0,86	0,86	0,85	0,85	0,85	0,85	0,84	0,84
0,9	1,32	1,32	1,32	1,31	1,31	1,30	1,30	1,29	1,29	1,28
0,95	1,72	1,71	1,71	1,70	1,70	1,68	1,68	1,66	1,65	1,65
0,975	2,07	2,06	2,06	2,05	2,04	2,02	2,01	1,98	1,97	1,96
0,99	2,51	2,49	2,48	2,47	2,46	2,42	2,40	2,37	2,35	2,33
0,995	2,82	2,80	2,78	2,76	2,75	2,70	2,68	2,63	2,60	2,58
0,999	3,51	3,47	3,44	3,41	3,39	3,31	3,26	3,17	3,13	3,09

9 *F*-Verteilung



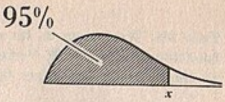
Tafel 9a. Werte von x, für die die Verteilungsfunktion (83.2) der F-Verteilung mit (m, n) Freiheitsgraden den Wert 0,95 hat

Beispiel: Bei (7, 4) Freiheitsgraden ist $F = 0{,}95$ für $x = 6{,}09$.

n	$m=1$	$m=2$	$m=3$	$m=4$	$m=5$	$m=6$	$m=7$	$m=8$	$m=9$
1	161	200	216	225	230	234	237	239	241
2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4
3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21
32	4,15	3,30	2,90	2,67	2,51	2,40	2,31	2,24	2,19
34	4,13	3,28	2,88	2,65	2,49	2,38	2,29	2,23	2,17
36	4,11	3,26	2,87	2,63	2,48	2,36	2,28	2,21	2,15
38	4,10	3,24	2,85	2,62	2,46	2,35	2,26	2,19	2,14
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04
70	3,98	3,13	2,74	2,50	2,35	2,23	2,14	2,07	2,02
80	3,96	3,11	2,72	2,49	2,33	2,21	2,13	2,06	2,00
90	3,95	3,10	2,71	2,47	2,32	2,20	2,11	2,04	1,99
100	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,97
150	3,90	3,06	2,66	2,43	2,27	2,16	2,07	2,00	1,94
200	3,89	3,04	2,65	2,42	2,26	2,14	2,06	1,98	1,93
1000	3,85	3,00	2,61	2,38	2,22	2,11	2,02	1,95	1,89
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88

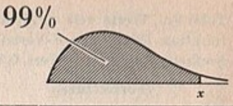
Tafel 9a. Werte von x, für die die Verteilungsfunktion (83.2) der F-Verteilung mit (m, n) Freiheitsgraden den Wert 0,95 hat

(Fortsetzung)



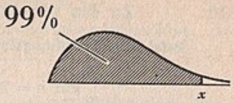
n	$m=10$	$m=15$	$m=20$	$m=30$	$m=40$	$m=50$	$m=100$	∞
1	242	246	248	250	251	252	253	254
2	19,4	19,4	19,4	19,5	19,5	19,5	19,5	19,5
3	8,79	8,70	8,66	8,62	8,59	8,58	8,55	8,53
4	5,96	5,86	5,80	5,75	5,72	5,70	5,66	5,63
5	4,74	4,62	4,56	4,50	4,46	4,44	4,41	4,37
6	4,06	3,94	3,87	3,81	3,77	3,75	3,71	3,67
7	3,64	3,51	3,44	3,38	3,34	3,32	3,27	3,23
8	3,35	3,22	3,15	3,08	3,04	3,02	2,97	2,93
9	3,14	3,01	2,94	2,86	2,83	2,80	2,76	2,71
10	2,98	2,85	2,77	2,70	2,66	2,64	2,59	2,54
11	2,85	2,72	2,65	2,57	2,53	2,51	2,46	2,40
12	2,75	2,62	2,54	2,47	2,43	2,40	2,35	2,30
13	2,67	2,53	2,46	2,38	2,34	2,31	2,26	2,21
14	2,60	2,46	2,39	2,31	2,27	2,24	2,19	2,13
15	2,54	2,40	2,33	2,25	2,20	2,18	2,12	2,07
16	2,49	2,35	2,28	2,19	2,15	2,12	2,07	2,01
17	2,45	2,31	2,23	2,15	2,10	2,08	2,02	1,96
18	2,41	2,27	2,19	2,11	2,06	2,04	1,98	1,92
19	2,38	2,23	2,16	2,07	2,03	2,00	1,94	1,88
20	2,35	2,20	2,12	2,04	1,99	1,97	1,91	1,84
22	2,30	2,15	2,07	1,98	1,94	1,91	1,85	1,78
24	2,25	2,11	2,03	1,94	1,89	1,86	1,80	1,73
26	2,22	2,07	1,99	1,90	1,85	1,82	1,76	1,69
28	2,19	2,04	1,96	1,87	1,82	1,79	1,73	1,65
30	2,16	2,01	1,93	1,84	1,79	1,76	1,70	1,62
32	2,14	1,99	1,91	1,82	1,77	1,74	1,67	1,59
34	2,12	1,97	1,89	1,80	1,75	1,71	1,65	1,57
36	2,11	1,95	1,87	1,78	1,73	1,69	1,62	1,55
38	2,09	1,94	1,85	1,76	1,71	1,68	1,61	1,53
40	2,08	1,92	1,84	1,74	1,69	1,66	1,59	1,51
50	2,03	1,87	1,78	1,69	1,63	1,60	1,52	1,44
60	1,99	1,84	1,75	1,65	1,59	1,56	1,48	1,39
70	1,97	1,81	1,72	1,62	1,57	1,53	1,45	1,35
80	1,95	1,79	1,70	1,60	1,54	1,51	1,43	1,32
90	1,94	1,78	1,69	1,59	1,53	1,49	1,41	1,30
100	1,93	1,77	1,68	1,57	1,52	1,48	1,39	1,28
150	1,89	1,73	1,64	1,53	1,48	1,44	1,34	1,22
200	1,88	1,72	1,62	1,52	1,46	1,41	1,32	1,19
1000	1,84	1,68	1,58	1,47	1,41	1,36	1,26	1,08
∞	1,83	1,67	1,57	1,46	1,39	1,35	1,24	1,00

Tafel 9b. Werte von x, für die die Verteilungsfunktion (83.2) der F-Verteilung mit (m, n) Freiheitsgraden den Wert 0,99 hat



n	$m=1$	$m=2$	$m=3$	$m=4$	$m=5$	$m=6$	$m=7$	$m=8$	$m=9$
1	4052	4999	5403	5625	5764	5859	5928	5982	6022
2	98,5	99,0	99,2	99,3	99,3	99,3	99,4	99,4	99,4
3	34,1	30,8	29,5	28,7	28,2	27,9	27,7	27,5	27,3
4	21,2	18,0	16,7	16,0	15,5	15,2	15,0	14,8	14,7
5	16,3	13,3	12,1	11,4	11,0	10,7	10,5	10,3	10,2
6	13,7	10,9	9,78	9,15	8,75	8,47	8,26	8,10	7,98
7	12,2	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72
8	11,3	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91
9	10,6	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35
10	10,0	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19
14	8,86	6,51	5,56	5,04	4,70	4,46	4,28	4,14	4,03
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46
22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,26
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,18
28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	3,12
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07
32	7,50	5,34	4,46	3,97	3,65	3,43	3,26	3,13	3,02
34	7,44	5,29	4,42	3,93	3,61	3,39	3,22	3,09	2,98
36	7,40	5,25	4,38	3,89	3,57	3,35	3,18	3,05	2,95
38	7,35	5,21	4,34	3,86	3,54	3,32	3,15	3,02	2,92
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89
50	7,17	5,06	4,20	3,72	3,41	3,19	3,02	2,89	2,79
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72
70	7,01	4,92	4,08	3,60	3,29	3,07	2,91	2,78	2,67
80	6,96	4,88	4,04	3,56	3,26	3,04	2,87	2,74	2,64
90	6,93	4,85	4,01	3,54	3,23	3,01	2,84	2,72	2,61
100	6,90	4,82	3,98	3,51	3,21	2,99	2,82	2,69	2,59
150	6,81	4,75	3,92	3,45	3,14	2,92	2,76	2,63	2,53
200	6,76	4,71	3,88	3,41	3,11	2,89	2,73	2,60	2,50
1000	6,66	4,63	3,80	3,34	3,04	2,82	2,66	2,53	2,43
∞	6,63	4,61	3,78	3,32	3,02	2,80	2,64	2,51	2,41

Tafel 9b. Werte von x, für die die Verteilungsfunktion (83.2) der F-Verteilung mit (m, n) Freiheitsgraden den Wert 0,99 hat



(Fortsetzung)

n	$m=10$	$m=15$	$m=20$	$m=30$	$m=40$	$m=50$	$m=100$	∞
1	6056	6157	6209	6261	6287	6300	6330	6366
2	99,4	99,4	99,4	99,5	99,5	99,5	99,5	99,5
3	27,2	26,9	26,7	26,5	26,4	26,4	26,2	26,1
4	14,5	14,2	14,0	13,8	13,7	13,7	13,6	13,5
5	10,1	9,72	9,55	9,38	9,29	9,24	9,13	9,02
6	7,87	7,56	7,40	7,23	7,14	7,09	6,99	6,88
7	6,62	6,31	6,16	5,99	5,91	5,86	5,75	5,65
8	5,81	5,52	5,36	5,20	5,12	5,07	4,96	4,86
9	5,26	4,96	4,81	4,65	4,57	4,52	4,42	4,31
10	4,85	4,56	4,41	4,25	4,17	4,12	4,01	3,91
11	4,54	4,25	4,10	3,94	3,86	3,81	3,71	3,60
12	4,30	4,01	3,86	3,70	3,62	3,57	3,47	3,36
13	4,10	3,82	3,66	3,51	3,43	3,38	3,27	3,17
14	3,94	3,66	3,51	3,35	3,27	3,22	3,11	3,00
15	3,80	3,52	3,37	3,21	3,13	3,08	2,98	2,87
16	3,69	3,41	3,26	3,10	3,02	2,97	2,86	2,75
17	3,59	3,31	3,16	3,00	2,92	2,87	2,76	2,65
18	3,51	3,23	3,08	2,92	2,84	2,78	2,68	2,57
19	3,43	3,15	3,00	2,84	2,76	2,71	2,60	2,49
20	3,37	3,09	2,94	2,78	2,69	2,64	2,54	2,42
22	3,26	2,98	2,83	2,67	2,58	2,53	2,42	2,31
24	3,17	2,89	2,74	2,58	2,49	2,44	2,33	2,21
26	3,09	2,82	2,66	2,50	2,42	2,36	2,25	2,13
28	3,03	2,75	2,60	2,44	2,35	2,30	2,19	2,06
30	2,98	2,70	2,55	2,39	2,30	2,25	2,13	2,01
32	2,93	2,66	2,50	2,34	2,25	2,20	2,08	1,96
34	2,89	2,62	2,46	2,30	2,21	2,16	2,04	1,91
36	2,86	2,58	2,43	2,26	2,17	2,12	2,00	1,87
38	2,83	2,55	2,40	2,23	2,14	2,09	1,97	1,84
40	2,80	2,52	2,37	2,20	2,11	2,06	1,94	1,80
50	2,70	2,42	2,27	2,10	2,01	1,95	1,82	1,68
60	2,63	2,35	2,20	2,03	1,94	1,88	1,75	1,60
70	2,59	2,31	2,15	1,98	1,89	1,83	1,70	1,54
80	2,55	2,27	2,12	1,94	1,85	1,79	1,66	1,49
90	2,52	2,24	2,09	1,92	1,82	1,76	1,62	1,46
100	2,50	2,22	2,07	1,89	1,80	1,73	1,60	1,43
150	2,44	2,16	2,00	1,83	1,73	1,66	1,52	1,33
200	2,41	2,13	1,97	1,79	1,69	1,63	1,48	1,28
1000	2,34	2,06	1,90	1,72	1,61	1,54	1,38	1,11
∞	2,32	2,04	1,88	1,70	1,59	1,52	1,36	1,00

10 Zu den verteilungsunabhängigen Verfahren

Tafel 10a. Verteilungsfunktion $F(x) = P(T \leq x)$ der Variablen T in Abschnitt 116

Für $n = 3$ ist $F(2) = 1 - 0{,}167 = 0{,}833$.
Für $n = 4$ ist $F(3) = 1 - 0{,}375 = 0{,}625$, $F(4) = 1 - 0{,}167 = 0{,}833$ usw.

x	$n = 3$
	0,
0	167
1	500

x	$n = 4$
	0,
0	042
1	167
2	375

x	$n = 5$
	0,
0	008
1	042
2	117
3	242
4	408

x	$n = 6$
	0,
0	001
1	008
2	028
3	068
4	136
5	235
6	360
7	500

x	$n = 7$
	0,
1	001
2	005
3	015
4	035
5	068
6	119
7	191
8	281
9	386
10	500

x	$n = 8$
	0,
2	001
3	003
4	007
5	016
6	031
7	054
8	089
9	138
10	199
11	274
12	360
13	452

x	$n = 9$
	0,
4	001
5	003
6	006
7	012
8	022
9	038
10	060
11	090
12	130
13	179
14	238
15	306
16	381
17	460

x	$n = 10$
	0,
6	001
7	002
8	005
9	008
10	014
11	023
12	036
13	054
14	078
15	108
16	146
17	190
18	242
19	300
20	364
21	431
22	500

x	$n = 11$
	0,
8	001
9	002
10	003
11	005
12	008
13	013
14	020
15	030
16	043
17	060
18	082
19	109
20	141
21	179
22	223
23	271
24	324
25	381
26	440
27	500

x	$n = 20$
	0,
50	001
51	002
52	002
53	003
54	004
55	005
56	006
57	007
58	008
59	010
60	012
61	014
62	017
63	020
64	023
65	027
66	032
67	037
68	043
69	049
70	056
71	064
72	073
73	082
74	093
75	104
76	117
77	130
78	144
79	159
80	176
81	193
82	211
83	230
84	250
85	271
86	293
87	315
88	339
89	362
90	387
91	411
92	436
93	462
94	487

x	$n = 19$
	0,
43	001
44	002
45	002
46	003
47	003
48	004
49	005
50	006
51	008
52	010
53	012
54	014
55	017
56	021
57	025
58	029
59	034
60	040
61	047
62	054
63	062
64	072
65	082
66	093
67	105
68	119
69	133
70	149
71	166
72	184
73	203
74	223
75	245
76	267
77	290
78	314
79	339
80	365
81	391
82	418
83	445
84	473
85	500

x	$n = 18$
	0,
38	001
39	002
40	003
41	003
42	004
43	005
44	007
45	009
46	011
47	013
48	016
49	020
50	024
51	029
52	034
53	041
54	048
55	056
56	066
57	076
58	088
59	100
60	115
61	130
62	147
63	165
64	184
65	205
66	227
67	250
68	275
69	300
70	327
71	354
72	383
73	411
74	441
75	470
76	500

x	$n = 17$
	0,
32	001
33	002
34	002
35	003
36	004
37	005
38	007
39	009
40	011
41	014
42	017
43	021
44	026
45	032
46	038
47	046
48	054
49	064
50	076
51	088
52	102
53	118
54	135
55	154
56	174
57	196
58	220
59	245
60	271
61	299
62	328
63	358
64	388
65	420
66	452
67	484

x	$n = 16$
	0,
27	001
28	002
29	002
30	003
31	004
32	006
33	008
34	010
35	013
36	016
37	021
38	026
39	032
40	039
41	048
42	058
43	070
44	083
45	097
46	114
47	133
48	153
49	175
50	199
51	225
52	253
53	282
54	313
55	345
56	378
57	412
58	447
59	482

x	$n = 15$
	0,
23	001
24	002
25	003
26	004
27	006
28	008
29	010
30	014
31	018
32	023
33	029
34	037
35	046
36	057
37	070
38	084
39	101
40	120
41	141
42	164
43	190
44	218
45	248
46	279
47	313
48	349
49	385
50	423
51	461
52	500

x	$n = 14$
	0,
18	001
19	002
20	002
21	003
22	005
23	007
24	010
25	013
26	018
27	024
28	031
29	040
30	051
31	063
32	079
33	096
34	117
35	140
36	165
37	194
38	225
39	259
40	295
41	334
42	374
43	415
44	457
45	500

x	$n = 13$
	0,
14	001
15	001
16	002
17	003
18	005
19	007
20	011
21	015
22	021
23	029
24	038
25	050
26	064
27	082
28	102
29	126
30	153
31	184
32	218
33	255
34	295
35	338
36	383
37	429
38	476

x	$n = 12$
	0,
11	001
12	002
13	003
14	004
15	007
16	010
17	016
18	022
19	031
20	043
21	058
22	076
23	098
24	125
25	155
26	190
27	230
28	273
29	319
30	369
31	420
32	473

Tafel 10b. Kritische Werte zu Abschnitt 117

c_1 und c_2 sind der größte ganzzahlige bzw. kleinste ganzzahlige Wert, für den $P(V \leqq c_1) \leqq \alpha$ bzw. $P(V \geqq c_2) \leqq \alpha$ ist.

m	c_1	
	$\alpha = 5\%$	$\alpha = 1\%$
5	3	2
6	3	2
7	4	3
8	5	4
9	6	4
10	6	5
11	7	6
12	8	7
13	9	7
14	10	8
15	11	9
16	11	10
17	12	10
18	13	11
19	14	12
20	15	13
21	16	14
22	17	14
23	17	15
24	18	16
25	19	17
26	20	18
27	21	19
28	22	19
29	23	20
30	24	21

m	c_2	
	$\alpha = 5\%$	$\alpha = 1\%$
5	8	9
6	10	11
7	11	12
8	12	13
9	13	15
10	15	16
11	16	17
12	17	18
13	18	20
14	19	21
15	20	22
16	22	23
17	23	25
18	24	26
19	25	27
20	26	28
21	27	29
22	28	31
23	30	32
24	31	33
25	32	34
26	33	35
27	34	36
28	35	38
29	36	39
30	37	40

Tafel 10c. Zum Rangtest in Abschnitt 118

Zweiseitiger Test. Verwerfung, wenn die Testgröße u_0 (s. Abschn. 118) gleich der in der Tabelle angegebenen Zahl oder kleiner als diese ist

m	5%	1%
6	0	0
7	2	0
8	4	0
9	6	1
10	8	3

m	5%	1%
11	11	5
12	14	7
13	17	10
14	21	13
15	25	16

m	5%	1%
16	30	19
17	35	23
18	40	28
19	46	32
20	52	38

Für große m ist der kritische Wert ungefähr gleich

$$M - 1{,}96\sqrt{N} \quad (5\%) \qquad \text{bzw.} \qquad M - 2{,}58\sqrt{N} \quad (1\%)$$

mit

$$M = \frac{m(m+1)}{4} \qquad \text{und} \qquad N = \frac{(2m+1)(m+1)m}{24}.$$

Anhang 6

FORMELZUSAMMENSTELLUNG

Einige Grundbegriffe

Klassische Wahrscheinlichkeitsdefinition von LAPLACE

(14.1) $$P(A) = g/m$$

g = Anzahl der Fälle, bei denen A eintrifft, m = Anzahl aller gleichmöglichen Fälle

Axiome der Wahrscheinlichkeit siehe Abschn. 15.

Komplementäres Ereignis. Es ist

(16.1) $$P(A) = 1 - P(\bar{A})$$

(21.1) $$P(X > c) = 1 - P(X \leqq c).$$

Additionsformel für beliebige Ereignisse

(18.1) $$P(A + B) = P(A) + P(B) - P(AB).$$

Bedingte Wahrscheinlichkeit

(19.1) $$P(B \mid A) = P(AB)/P(A) \qquad [P(A) \neq 0]$$

(19.3) $$P(AB) = P(A)\,P(B \mid A) = P(B)\,P(A \mid B).$$

Unabhängigkeit. Zwei Ereignisse A und B heißen unabhängig, wenn

(20.1) $$P(AB) = P(A)\,P(B)$$

gilt. Zwei Zufallsvariable X und Y heißen unabhängig, wenn für alle (x, y) die Beziehung

(56.1) $$F(x, y) = F_1(x)\,F_2(y)$$

gilt.

Häufigkeitsfunktion $\tilde{f}(x)$ und *Summenhäufigkeitsfunktion* $\tilde{F}(x)$ *einer Stichprobe* siehe Abschn. 4 bzw. 7.

Wahrscheinlichkeitsfunktion $f(x)$ bzw. *Dichte* $f(x)$ einer Zufallsvariablen siehe Abschn. 22 bzw. 26.

Verteilungsfunktion

(24.1) $$F(x) = P(X \leqq x)$$

(24.3) $$P(a < X \leqq b) = F(b) - F(a)$$

Im diskreten bzw. stetigen Fall ist

(25.1), (26.1) $$F(x) = \sum_{x_j \leqq x} f(x_j) \quad \text{bzw.} \quad \int_{-\infty}^{x} f(v)\, dv.$$

Mittelwert

Mittelwert einer Stichprobe $x_1, x_2, \ldots, x_n$:

(8.1) $$\bar{x} = \sum_{j=1}^{n} x_j = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$$

Häufigster Wert, Medianwert und *Zentralwert* einer Stichprobe siehe Abschn. 9.

Mittelwert einer diskreten bzw. stetigen *Verteilung* mit der Wahrscheinlichkeitsfunktion bzw. Dichte $f(x)$:

(28.1) $$\mu = \sum_j x_j f(x_j) \quad \text{bzw.} \quad \int_{-\infty}^{\infty} x f(x)\, dx.$$

Der Mittelwert einer Summe von Zufallsvariablen ist gleich der Summe der Mittelwerte dieser Variablen (Satz 58.1). Der Mittelwert des Produktes *unabhängiger* Zufallsvariablen ist gleich dem Produkt der Mittelwerte dieser Variablen (Satz 58.2).

Lineare Skalentransformation. Hat die Zufallsvariable X den Mittelwert μ, so hat die Zufallsvariable $X^* = c_1 X + c_2$ den Mittelwert

(71.2) $$\mu^* = c_1 \mu + c_2.$$

Varianz

Varianz einer Stichprobe $x_1, x_2, \ldots, x_n$:

(8.3) $$s^2 = \frac{1}{n-1} \sum_{j=1}^{n} (x_j - x)^2$$

(9.1) $$s^2 = \frac{1}{n-1}\left[\sum_{j=1}^{n} x_j^2 - \frac{1}{n}\left(\sum_{j=1}^{n} x_j\right)^2\right]$$

s heißt *Standardabweichung der Stichprobe* (in der Fehlerrechnung auch *mittlerer Fehler*; vgl. Abschn. 111).

Spannweite einer Stichprobe siehe Abschn. 8.

Varianz einer diskreten bzw. stetigen *Verteilung* mit der Wahrscheinlichkeitsfunktion bzw. Dichte $f(x)$:

(29.1) $$\sigma^2 = \sum_j (x_j - \mu)^2 f(x_j) \quad \text{bzw.} \quad \int_{-\infty}^{\infty} (x - \mu)^2 f(x)\, dx$$

(31.5) $$\sigma^2 = E(X^2) - \mu^2$$

Die Summe $Z = X + Y$ hat die Varianz

(59.5) $$\sigma^2 = \sigma_1{}^2 + \sigma_2{}^2 + 2\sigma_{XY}.$$

$\sigma_1{}^2$ bzw. $\sigma_2{}^2$ ist die Varianz von X bzw. Y, und σ_{XY} ist die *Kovarianz*,

(59.4) $$\sigma_{XY} = E(XY) - E(X)\,E(Y).$$

Die Varianz einer Summe unabhängiger Zufallsvariablen ist gleich der Summe der Varianzen dieser Variablen (Satz 59.1).

Lineare Skalentransformation. Hat die Zufallsvariable X die Varianz σ^2, so hat die Zufallsvariable $X^* = c_1 X + c_2$ die Varianz

(71.3) $$\sigma^{*2} = c_1{}^2 \sigma^2.$$

Momente

k-tes Moment

(31.1) $$E(X^k) = \sum_j x_j^k f(x_j) \quad \text{bzw.} \quad \int_{-\infty}^{\infty} x^k f(x)\, dx$$

k-tes zentrales Moment

(31.3) $$E([X - \mu]^k) = \sum_j (x_j - \mu)^k f(x_j) \quad \text{bzw.} \quad \int_{-\infty}^{\infty} (x - \mu)^k f(x)\, dx$$

Schiefe

(32.1) $$\gamma = \sigma^{-3} E([X - \mu]^3)$$

Momenterzeugende Funktion

(33.1) $$G(t) = E(e^{tX})$$

Charakteristische Funktion

(33.4) $$\Psi(t) = E(e^{itX})$$

Momente und zentrale Momente zweidimensionaler Verteilungen siehe Aufgabe 107.1.

Spezielle Verteilungen

Binomialverteilung. Wahrscheinlichkeitsfunktion:

(39.4) $$f(x) = \binom{n}{x} p^x q^{n-x} \qquad (x = 0, 1, \ldots, n)$$

Mittelwert np, Varianz npq (siehe Abschn. 40). Approximation durch die Normalverteilung siehe Abschn. 50. *Multinomialverteilung* siehe Aufgabe 53.3. BERNOULLI-Experimente mit variabler Erfolgswahrscheinlichkeit siehe Beispiel 59.2.

PASCAL-*Verteilung*. Wahrscheinlichkeitsfunktion (siehe Aufgabe 41.15)

(41.2) $$f(x) = \binom{k+x-1}{x} p^k q^x \qquad (x = 0, 1, \ldots)$$

POISSON-*Verteilung*. Wahrscheinlichkeitsfunktion:

(42.2) $$f(x) = \frac{\mu^x}{x!} e^{-\mu} \qquad (x = 0, 1, \ldots)$$

Mittelwert μ, Varianz μ, Schiefe $1/\sqrt{\mu}$ (siehe Abschn. 44).

Hypergeometrische Verteilung. Wahrscheinlichkeitsfunktion:

(45.1) $$f(x) = \binom{M}{x}\binom{N-M}{n-x} \Big/ \binom{N}{n} \qquad (x = 0, 1, \ldots, n)$$

Mittelwert nM/N, Varianz $nM(N-M)(N-n)/N^2(N-1)$ (siehe Abschn. 45).

Normalverteilung. Dichte:

(47.1) $$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad (-\infty < x < \infty, \sigma > 0)$$

Mittelwert μ, Varianz σ^2. Es ist

(48.5) $$P(a < X \leqq b) = F(b) - F(a) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

mit Φ gemäß (48.3); Zahlenwerte in Anhang 5.

Logarithmische Normalverteilung. Dichte (siehe Aufgabe 49.12)

(49.1) $$f(x) = \frac{1}{x\lambda\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln x - \varkappa}{\lambda}\right)^2} \qquad (x > 0)$$

Zweidimensionale Normalverteilung siehe Abschnitt 108. Die Summe unabhängiger normalverteilter Zufallsvariablen ist normalverteilt (Satz 71.1).

Chi-Quadrat-Verteilung. Dichte:

(60.2), (60.4) $$f(x) = x^{(n-2)/2} e^{-x/2} / [2^{n/2} \Gamma(n/2)] \qquad (x > 0)$$

Mittelwert n, Varianz $2n$ (siehe Abschn. 61).

Gamma-Verteilung und *Beta-Verteilung* siehe Aufgabe 60.4 bzw. 60.5.

t-Verteilung von STUDENT. Dichte:

(62.2) $$f(z) = \Gamma\left(\frac{n+1}{2}\right) \Big/ \left[\sqrt{n\pi}\, \Gamma\left(\frac{n}{2}\right) \left(1 + \frac{z^2}{n}\right)^{(n+1)/2}\right]$$

Mittelwert 0, Varianz $n/(n-2)$ (siehe Abschn. 62).

F-Verteilung. Dichte (vgl. Abschn. 83):

$$f(x) = \Gamma\left(\frac{m+n}{2}\right) m^{m/2}\, n^{n/2}\, x^{(m-2)/2} \Big/ \left[\Gamma\left(\frac{m}{2}\right) \Gamma\left(\frac{n}{2}\right) (mx+n)^{(m+n)/2}\right] \quad (x > 0)$$

Regression und Korrelation

Regressionsgrade einer Stichprobe $(x_1, y_1), \ldots, (x_n, y_n)$

(94.3) $$y - \bar{y} = b(x - \bar{x})$$

$\bar{x}$ bzw. $\bar{y}$ ist der Mittelwert der x- bzw. y-Werte.

(94.5) $$b = s_{xy}/s_1^2$$

ist der *Regressionskoeffizient der Stichprobe*. s_1^2 ist die Varianz der x-Werte, und

(94.7) $$s_{xy} = \frac{1}{n-1} \sum_{j=1}^{n} (x_j - \bar{x})(y_j - \bar{y}),$$

(94.8) $$s_{xy} = \frac{1}{n-1} \left[\sum_{j=1}^{n} x_j y_j - \frac{1}{n} \left(\sum_{j=1}^{n} x_j\right) \left(\sum_{j=1}^{n} y_j\right)\right]$$

ist die *Kovarianz der Stichprobe*.

Regressionsgerade der Grundgesamtheit

(97.5) $$Y - \mu_2 = \beta(X - \mu_1)$$

μ_1 bzw. μ_2 ist der Mittelwert von X bzw. Y.

(97.4) $$\beta = \sigma_{XY}/\sigma_1^2$$

ist der *Regressionskoeffizient*. σ_1^2 ist die Varianz von X, und

(97.2) $$\sigma_{XY} = E([X - \mu_1][Y - \mu_2]),$$

(97.3) $$\sigma_{XY} = E(XY) - E(X)E(Y)$$

ist die *Kovarianz* der Variablen X und Y.

Regressionsparabel einer Stichprobe:

(104.4) $$y = b_0 + b_1 x + b_2 x^2.$$

Die Koeffizienten b_0, b_1 und b_2 ergeben sich als Lösungen der „*Normalgleichungen*“ (104.5).

Korrelationskoeffizient einer Stichprobe $(x_1, y_1), \ldots, (x_n, y_n)$:

(106.4) $$r = s_{xy}/s_1 s_2$$

s_1^2 bzw. s_2^2 ist die Varianz der x- bzw. y-Werte. Stets ist

(106.5) $$-1 \leqq r \leqq 1.$$

Korrelationskoeffizient der Grundgesamtheit

(107.4) $$\varrho = \sigma_{XY}/\sigma_1 \sigma_2$$

σ_1^2 bzw. σ_2^2 ist die Varianz von X bzw. Y, und σ_{XY} ist die *Kovarianz* (siehe oben). Stets ist

(107.5) $$-1 \leqq \varrho \leqq 1.$$

Zwei *normalverteilte* Zufallsvariable sind dann und nur dann unabhängig, wenn $\varrho = 0$ ist.

Sachverzeichnis