



Multivariate Statistik

Lehr- und Handbuch der
angewandten Statistik

von

o. Prof. Dr. Joachim Hartung

Fachbereich Statistik der Universität Dortmund

und

Dr. Bärbel Elpelt

7., unveränderte Auflage

R. Oldenbourg Verlag München Wien

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

© 2007 Oldenbourg Wissenschaftsverlag GmbH
Rosenheimer Straße 145, D-81671 München
Telefon: (089) 45051-0
oldenbourg.de

Das Werk einschließlich aller Abbildungen ist urheberrechtlich geschützt. Jede Verwertung außerhalb der Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlages unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Bearbeitung in elektronischen Systemen.

Lektorat: Wirtschafts- und Sozialwissenschaften, wiso@oldenbourg.de
Herstellung: Anna Grosser
Satz: DTP-Vorlagen des Autors
Coverentwurf: Kochan & Partner, München
Coverausführung: Gerbert-Satz, Grasbrunn
Gedruckt auf säure- und chlorfreiem Papier
Gesamtherstellung: Druckhaus „Thomas Müntzer“ GmbH, Bad Langensalza

ISBN 3-486-58234-8
ISBN 978-3-486-58234-5

Kapitelverzeichnis

EINLEITUNG UND ÜBERBLICK	1
KAPITEL I: Einführung und Grundlagen	17
KAPITEL II: Die Regressionsanalyse	77
KAPITEL III: Die Korrelationsanalyse	143
KAPITEL IV: Multivariate Ein- und Zweistichprobenprobleme; Diskriminanzanalyse, Reduktion von Merkmalen	221
KAPITEL V: Aufbereitung und Auswertung qualitativer und gemischter Daten - Skalierung kategoriieller Merkmale (Skalierung in Kontingenztafeln)	269
KAPITEL VI: Die Multidimensionale Skalierung (MDS)	377
KAPITEL VII: Die Clusteranalyse	443
KAPITEL VIII: Die Faktorenanalyse	505
KAPITEL IX: Graphische Verfahren	593
KAPITEL X: Das Multivariate Lineare Modell (Multivariate Regressions-, Varianz-, Kovarianz- und Profil- analyse, Multivariate Varianzkomponentenmodelle, Präzisionsbestimmung bei MeBinstrumenten)	655
ANHANG	741
	Ende..... 815

Inhaltsverzeichnis

VORWORT	XIII
EINLEITUNG UND ÜBERBLICK	1
KAPITEL I: EINFÜHRUNG UND GRUNDLAGEN	17
1 Grundlegende Begriffe und elementare Datenbeschreibung	18
1.1 Merkmalstypen und Klassenbildung	18
1.2 Häufigkeiten, Summenhäufigkeiten und empirische Verteilungs- funktion	19
1.3 Empirische Lagemaße	21
1.4 Empirische Streuungsmaße	23
2 Grundlagen der Wahrscheinlichkeitsrechnung	25
2.1 Wahrscheinlichkeiten und bedingte Wahrscheinlichkeiten	26
2.2 Zufallsvariable und Verteilungen	28
2.3 Kenngrößen von Zufallsvariablen	32
3 Prinzipien des Schätzens und Testens; t-, χ^2 - und F-Verteilung	38
4 Vektor- und Matrizenrechnung	48
5 Mehrdimensionale und multivariate Verteilungen	64
6 Daten- und Distanzmatrix	70
KAPITEL II: DIE REGRESSIONSANALYSE	77
1 Multiple Regressionsanalyse für quantitative Daten	81
2 Das Gemischte Lineare Modell	118
3 Diskrete Regressionsanalyse für qualitative Daten; Lineares Wahr- scheinlichkeitsmodell, Probit-, Logitanalyse	128
KAPITEL III: DIE KORRELATIONSANALYSE	143
1 Die Korrelation normalverteilter Merkmale	144
1.1 Die Korrelation zweier normalverteilter Merkmale	144
1.1.1 Tests und Konfidenzintervalle für ρ	153
1.1.2 Vergleich von Korrelationen mehrerer Merkmalspaare	159
1.2 Zusammenhangsanalyse mehrerer Merkmale	162
1.3 Die multiple Korrelation	167
1.4 Die kanonische Korrelation	172
1.5 Die partielle Korrelation	181
1.6 Die bi-partielle Korrelation	186

2 Die Korrelation von nicht-normalverteilten Zufallsvariablen	190
2.1 Der Spearmansche Rangkorrelationskoeffizient	191
2.2 Der Kendallsche Korrelationskoeffizient	199
2.3 Korrelationskoeffizienten bei ordinalen Merkmalen	201
3 Assoziationsmaße und loglineares Modell für Kontingenztafeln	206
4 Ein zusammenfassendes Beispiel	212

KAPITEL IV: MULTIVARIATE EIN- UND ZWEISTICHPROBENPROBLEME;
DISKRIMINANZANALYSE, REDUKTION VON MERKMALEN 221

1 Das Multivariate Einstichprobenproblem	223
1.1 Schätzen des Mittelwertvektors μ und der Kovarianzmatrix Σ	223
1.2 Test über den Mittelwertvektor μ bei bekannter Kovarianzmatrix Σ ..	225
1.3 Test über den Mittelwertvektor μ bei unbekannter Kovarianz- matrix Σ	227
1.4 Ein Symmetrietest	228
2 Das Multivariate Zweistichprobenproblem	230
2.1 Mittelwertvergleich bei unverbundenen Stichproben	230
2.2 Mittelwertvergleich bei verbundenen Stichproben	232
3 Die Prüfung von Kovarianzhypothesen	234
3.1 Ein Test über die Struktur einer Kovarianzmatrix Σ	234
3.2 Ein Test auf Gleichheit mehrerer Kovarianzmatrizen	236
3.3 Ein simultaner Test über Mittelwertvektor und Kovarianzmatrix im Einstichprobenproblem	238
4 Die Diskriminanzanalyse (Identifikation von Objekten)	240
4.1 Der Zweigruppenfall	242
4.2 Der Mehrgruppenfall	245
4.3 Ein Beispiel	247
4.4 Ein Trennmaß und die Reduktion von Merkmalen	251
5 Ein zusammenfassendes Beispiel	258

KAPITEL V: AUFBEREITUNG UND AUSWERTUNG QUALITATIVER UND
GEMISCHTER DATEN - SKALIERUNG KATEGORIELLER
MERKMALE (SKALIERUNG IN KONTINGENZTAFELN)

1 Skalierung ordinaler und nominaler Merkmalsausprägungen	276
1.1 Skalierung ordinaler Merkmalsausprägungen	277
1.2 Skalierung nominaler Merkmalsausprägungen in zweidimensionalen Kontingenztafeln - kategorielle Skalierung, Lancaster - Ska- lierung	282

- 2 Multivariate Analyseverfahren in skalierten Kontingenztafeln mit einer Kriteriumsvariablen (Calibration Patterns) 290
 - 2.1 Beste Diskriminatoren zwischen den Stufen der Kriteriumsvariablen 296
 - 2.2 Methoden der Güteprüfung einer Skalierung 300
 - 2.2.1 Die Güteprüfung mittels Diskriminanzfunktionen 301
 - 2.2.2 Die Güteprüfung mittels Mahalanobisdistanzen 304
 - 2.3 Die Klassifizierung neuer Objekte 307
 - 2.4 Gewinnung einer Daten- und Distanzmatrix zur weiteren multivariaten Analyse 309
- 3 Ein Beispiel aus der Marktforschung zur Analyse multivariater kategoriemerkmale 313
- 4 Skalierung kategoriemerkmale von p Merkmalen 322
 - 4.1 Bestimmung der empirischen Korrelationsmatrix für p kategoriemerkmale 323
 - 4.2 Das Kriterium der maximalen Maximum-Exzentrizität und der minimalen Determinante 331
 - 4.3 Das Kriterium der maximalen multiplen Korrelation 334
 - 4.4 Das Kriterium der maximalen kanonischen Korrelation 347
- 5 Skalierung kategoriemerkmale bei gemischten Datentypen..... 350
- 6 Korrespondenzanalyse, Guttman'sche Skalierung und die ALS-Verfahren ... 369
- KAPITEL VI: DIE MULTIDIMENSIONALE SKALIERUNG (MDS) 377
 - 1 Nonlinear Mapping 384
 - 2 Die Haupt-Koordinaten-Methode 393
 - 3 Das Verfahren von Kruskal 405
 - 4 Die Unfolding-Technik 420
 - 4.1 Die Methode der Dreiecksanalyse 421
 - 4.2 Der Goode-Phillips-Algorithmus 426
- KAPITEL VII: DIE CLUSTERANALYSE 443
 - 1 Klassifikationstypen 447
 - 2 Bewertungskriterien für Klassifikationen 454
 - 2.1 Maße für die Homogenität einer Klasse 454
 - 2.2 Maße für die Heterogenität zwischen den Klassen 456
 - 2.3 Maße für die Güte einer Klassifikation 458

3 Konstruktionsverfahren für Überdeckungen	460
3.1 Ein exhaustives Verfahren für kleine Objektmengen	461
3.2 Ein iteratives Konstruktionsverfahren	463
4 Konstruktionsverfahren für Partitionen	465
4.1 Ein iteratives Verfahren	465
4.2 Ein rekursives Verfahren	469
5 Ein Verfahren zur Konstruktion einer Quasihierarchie	473
6 Ein Verfahren zur Konstruktion einer Hierarchie	478
7 Klassenzuordnung neuer Objekte - Diskrimination, Identifikation	489
8 Ein zusammenfassendes Beispiel	494
 KAPITEL VIII: DIE FAKTORENANALYSE	 505
1 Die Bestimmung der Faktorladungen	518
1.1 Die Maximum-Likelihood-Methode und ein Test über die Anzahl der Faktoren	519
1.2 Die kanonische Faktorenanalyse	525
1.3 Die Hauptkomponenten- und die Hauptfaktorenanalyse	527
1.4 Die Zentroidmethode	534
1.5 Die Jöreskog-Methode	541
2 Die Rotation der Faktoren	546
2.1 Die orthogonale Rotation der Faktoren	548
2.1.1 Die Varimax-Methode	551
2.1.2 Die Quartimax-Methode	559
2.2 Schiefwinkelige Rotation - Die Methode der Primärfaktoren	561
3 Schätzen von Faktorenwerten	568
4 Ein zusammenfassendes Beispiel	576
 KAPITEL IX: GRAPHISCHE VERFAHREN	 593
1 Gemeinsame Repräsentation von Objekten und (oder) Merkmalen	595
1.1 Graphische Darstellung ein- und zweidimensionaler Daten	596
1.1.1 Stem and Leaves und Box-Plot	597
1.1.2 Graphische Darstellung zweidimensionaler Daten am Beispiel eines Produkt-Markt-Portfolios	600
1.2 Die Probability-Plotting-Technik: Überprüfung auf multivariate Normalverteilung und multivariate Ausreißer (Q-Q-Plot)	602
1.3 Gleichzeitige Repräsentation von Merkmalen und Objekten: Der Bi-Plot	605
1.4 Weitere Graphische Repräsentationsformen für Objekte und Merkmale	608

2	Repräsentation einzelner Objekte oder Merkmale	610
2.1	Einfache Darstellungsformen bei Repräsentation von Merkmals- werten durch Strecken	612
2.1.1	Profile, Streifen	613
2.1.2	Polygonzüge	613
2.1.3	Sterne	614
2.1.4	Sonnen	614
2.1.5	Glyphs	616
2.2	Darstellung von Objekten vermittelt Diamanten	617
2.3	Darstellung von Objekten mittels Gesichtern	618
2.4	Darstellung von Objekten durch trigonometrische Funktionen	622
2.4.1	Andrews-Plots	622
2.4.2	Blumen	623
2.5	Darstellung von Objekten unter Berücksichtigung der Merkmals- ähnlichkeiten	626
2.5.1	Quader	628
2.5.2	Bäume	629
2.5.3	Burgen	633
2.6	Darstellung von Objekten unter Berücksichtigung der Diskrimi- nationsgüte der Merkmale: Facetten	636
2.7	Darstellung von Objekten unter Berücksichtigung der Merkmals- korrelationen: Bi-Plot-Sonnen	638
3	Bilanzkennzahlen der chemischen Industrie zwischen 1965 und 1980: Ein Beispiel für die Anwendung graphischer Verfahren zur Darstellung zeitlicher Entwicklungen	639
KAPITEL X: DAS MULTIVARIATE LINEARE MODELL (MULTIVARIATE REGRESSIONS-, VARIANZ-, KOVARIANZ- UND PROFIL- ANALYSE, MULTIVARIATE VARIANZKOMponentENMODELLE, PRÄZISIONSBESTIMMUNG BEI MEßINSTRUMENTEN)		
1	Das Multivariate Lineare Modell mit festen Effekten (Modell I)	656
1.1	Das allgemeine restringierte Multivariate Lineare Modell	659
1.2	Testverfahren im allgemeinen restringierten Multivariaten Linearen Modell	664
1.3	Multivariate Regressions- und Kovarianzanalyse	667
1.4	Einige Modelle der Multivariaten Varianzanalyse (MANOVA) mit festen Effekten	692
1.4.1	Die einfaktorielle multivariate Varianzanalyse (Vergleich von r unabhängigen Stichproben)	693
1.4.2	Die multivariate zweifache Kreuzklassifikation mit Wechselwirkungen	700

1.4.3 Die multivariate zweifache Kreuzklassifikation mit einer Beobachtung pro Zelle (Das einfache multivariate Blockexperiment)	705
1.4.4 Die multivariate zweifach hierarchische Klassifikation	707
1.5 Die Profilanalyse zur Untersuchung von Wachstums- und Verlaufskurven im Multivariaten Linearen Modell mit festen Effekten	710
1.5.1 Normalverteilungsverfahren	713
1.5.2 Ein nichtparametrisches Verfahren	717
2 Das Multivariate Lineare Modell mit zufälligen Effekten (MANOVA - Modelle II, Multivariate Varianzkomponentenmodelle)	719
2.1 Die balancierte multivariate Einfachklassifikation mit zufälligen Effekten	723
2.2 Das balancierte zweifach hierarchische Modell mit zufälligen Effekten	725
2.3 Das balancierte dreifach hierarchische Modell mit zufälligen Effekten	727
2.4 Die balancierte zweifache Kreuzklassifikation mit zufälligen Effekten	731
2.5 Ein Modell zur Präzisionsbestimmung von Meßinstrumenten bei zerstörenden Prüfungen	736
ANHANG	741
1 Tabellenanhang	741
- Verteilungsfunktion $\phi(x)$ der Standardnormalverteilung $N(0;1)$ [Tab.1]	742
- Quantile u_Y der Standardnormalverteilung $N(0;1)$ [Tab.2]	743
- Quantile $t_{n;Y}$ der t-Verteilung [Tab.3]	744
- Quantile $\chi_{n;Y}^2$ der χ^2 -Verteilung [Tab.4]	745
- Quantile $F_{n_1, n_2; Y}$ der F-Verteilung [Tab.5]	747
- Nomogramme von D.L. Heck zum Roy-Test [Chart I bis Chart XII]	754
2 Erläuterungen zu den multivariaten Testverfahren	766
2.1 Zum Roy-Test	766
2.2 Zum Wilks-Test	767
2.3 Zum Hotelling-Lawley-Test	768
2.4 Zum Pillai-Bartlett-Test	769
3 Griechisches Alphabet	770
4 Literaturverzeichnis	771
5 Stichwortverzeichnis	785
6 Symbolverzeichnis	807
Ende	815

Vorwort zur 6. Auflage

Nachdem auch die 5. Auflage dieses Buches recht schnell vergriffen war, liegt hier nun bereits die 6. Auflage vor. Da der Text in den Voraufgaben überarbeitet wurde, konnte er hier weitgehend unverändert bleiben. Auch weiterhin sind wir für Anregungen der Leser dankbar.

Joachim Hartung
Bärbel Elpelt

Aus dem Vorwort zur 1. Auflage

Die Statistik und insbesondere die Multivariate Statistik wird überall dort eingesetzt, wo es gilt, komplexes Datenmaterial auszuwerten. In allen Bereichen der Wissenschaft aber auch in Wirtschaft, Handel, Technik und Administration werden im Zuge der fortschreitenden Technisierung vielfältige Informationen erhoben, gemessen, beobachtet und registriert, aus denen es gilt, relevante Schlüsse zu ziehen.

Dieses Buch, in dem die wohl wichtigsten Verfahren der Multivariaten Statistik dargestellt werden, wendet sich sowohl an den im Beruf stehenden Praktiker als auch an Wissenschaftler und Studenten aller Fachrichtungen. Es ist somit nicht nur ein Lehrbuch sondern vornehmlich auch ein praktisches Handbuch und Nachschlagewerk für jeden, der mit der Auswertung umfangreicher Daten konfrontiert wird.

Um diesem Anspruch gerecht werden zu können, werden die einzelnen Verfahren bzw. die mit ihnen beantwortbaren Fragestellungen anhand von Beispielen aus den verschiedensten Anwendungsgebieten erläutert. An die Darstellung der Methoden respektive ihrer Voraussetzungen und ihrer Durchführung schließt sich stets ein konkretes, nachvollziehbares Zahlenbeispiel an; bis auf ganz wenige Ausnahmen wurden die zahlreichen Beispiele ausschließlich mit Taschenrechnern (TI 51111 und HP 15C) durchgerechnet und alle Zwischenschritte aufgeführt, so daß man einen tieferen Einblick in die Wirkungsweise der Verfahren erlangt.

Man kann sich fragen, wozu eine solche Darstellung im Zeitalter der Fertigprogramme überhaupt notwendig ist. Wir sind der Meinung, daß eine Anwendung von Fertigprogrammen nur dann sinnvoll erfolgen kann, wenn die implementierten Verfahren zumindest in ihren Grundzügen dem Benutzer bekannt sind.

Insbesondere ist nur dann eine sachgerechte Interpretation der Ergebnisse möglich.

Das Buch ist so weit wie möglich derart gehalten, daß es eigenständig und ohne große mathematische Vorkenntnisse gelesen und verstanden werden kann. Die unbedingt benötigten Grundlagen aus Statistik, Wahrscheinlichkeitsrechnung und Matrizenrechnung sind daher im ersten Kapitel des Buches noch einmal kurz dargestellt. Abgesehen von der erforderlichen Kenntnis der dort eingeführten Grundbegriffe können die einzelnen Kapitel weitgehend unabhängig voneinander gelesen und erarbeitet werden, was den Handbuchcharakter unterstreicht.

An dieser Stelle möchten wir nachstehenden Personen für ihre Unterstützung bei der Erstellung des Buches unseren Dank aussprechen. Frau Dipl.-Stat. *Barbara Heine* erstellte das gesamte Typoskript einschließlich der Tuschzeichnungen und war uns durch ihr sorgfältiges Mitlesen sowie das Nachrechnen einiger Beispiele sehr behilflich.

Das Programm zur Erstellung der Flury-Riedwyl-Faces in Kapitel IX wurde uns dankenswerter Weise von Herrn Dr. Bernhard Flury und Herrn Prof. Dr. Hans Riedwyl, Universität Bern, zur Verfügung gestellt und konnte unter Verwendung des Graphiksystems Disspla des Hochschulrechenzentrums der Universität Dortmund angewandt werden. In diesem Zusammenhang möchten wir auch die Herren cand. stat. Manfred Jutzi und cand. stat. Thomas Nawrath erwähnen, die sich bei der Implementierung und der Erstellung einiger Computer-Abbildungen, von denen 15 im Text aufgenommen wurden, einsetzten.

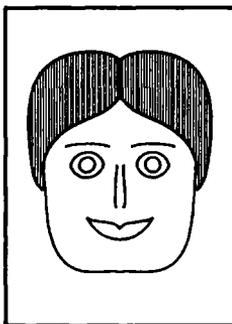
Mit Herrn Prof. Dr. Rolf E. Bargmann, University of Georgia, haben wir während seiner von der Deutschen Forschungsgemeinschaft unterstützten Gastprofessur im Sommersemester 1980 an der Universität Dortmund aufschlußreiche Gespräche geführt. Herr Priv.-Doz. Dr. Peter Pflaumer, z.Zt. Universität Dortmund, hat durch anregende Diskussionen während der Entstehungszeit des Buches dessen Ausrichtung beeinflußt.

Joachim Hartung
Bärbel Elpelt

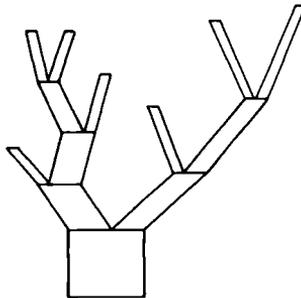
Einleitung und Überblick

In vielen Bereichen der *Wissenschaft*, z.B. in den Wirtschafts- und Sozialwissenschaften, den Ingenieurwissenschaften, der Psychologie, der Pädagogik, der Umweltforschung, den Agrarwissenschaften, der Biologie, der Medizin, der Chemie, der Archäologie, der Astronomie, der Physik, der Geographie, der Geodäsie, der Geologie oder der Informatik, spielen Auswertung und Interpretation großer Datenmengen eine entscheidende Rolle; in zunehmenden Maße der Fall ist dies aber auch in *Wirtschaft, Handel, Administration und Technik*.

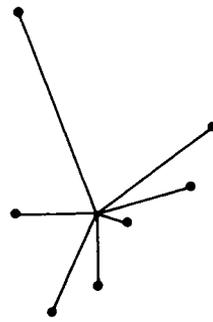
Die Statistik und insbesondere die *Multivariate Statistik* stellt Methoden und Verfahren zur Verfügung, die der *Aufbereitung, tabellarischen und graphischen Repräsentation und Auswertung* komplexer Datensituationen dienen. Zum Beispiel ermöglichen *graphische Repräsentationsformen* wie die in **Abb.1** abgebildeten Flury - Riedwyl - Faces, Kleiner - Hartigan - Trees und Biplot - Suns



Flury-Riedwyl-Face



Kleiner-Hartigan-Tree



Biplot-Sun

Abb.1: Drei Beispiele zur Graphischen Repräsentation komplexer Datensituationen

nicht nur einen schnellen und klaren Überblick über komplexe Strukturen sondern erlauben auch eine direkte Analyse und Interpretation der Daten;

da graphische Repräsentationen von Datenmengen oftmals erst Ergebnis anderer multivariater Verfahren sind bzw. andere Verfahren benutzt werden, um graphische Darstellungen zu gewinnen, werden wir uns erst im neunten Kapitel ausführlich mit ihnen beschäftigen.

Die *Gewinnung von Daten*, die der statistischen Analyse natürlich stets voraus geht, erfolgt in einem *Experiment* oder einer *Erhebung* durch *Beobachtungen* und *Messungen*, die an *Objekten*, *Untersuchungseinheiten* vorgenommen werden. Objekte können in diesem Zusammenhang etwa Firmen, Werkstücke, Personen, Tiere, Länder etc. sein. Beobachtet bzw. gemessen werden dann an den Objekten die *Ausprägungen* verschiedener interessierender *Merkmale*. Bei Personen können etwa die Ausprägungen von Merkmalen wie Größe, Gewicht, Familienstand, Alter, Blutdruck, Beruf, Parteizugehörigkeit interessieren; bei Firmen sind z.B. Bilanzkennzahlen wie Kapitalumschlag, Eigenkapitalanteil, dynamischer Verschuldungsgrad und Liquidität wichtig für die Beurteilung ihrer Kreditwürdigkeit; bei PKW's sind Hubraum, Leistung, Höchstgeschwindigkeit, Preis, Reparaturanfälligkeit, Kraftstoffverbrauch wesentlich zum Vergleich verschiedener Typen; Länder lassen sich bzgl. ihrer Einwohnerdichte, ihrer Fertilitätsrate, ihrem Altersaufbau, ihrem Industrialisierungsgrad, ihrem Prokopfeinkommen, ihrer landwirtschaftlichen Nutzfläche usw. untersuchen.

In einem Experiment oder einer Erhebung können nun oftmals nicht alle Objekte aus einer interessierenden *Grundgesamtheit* sondern nur einige stichprobenartig, zufällig ausgewählte Objekte berücksichtigt werden. Beispielsweise können nicht alle Werkstücke aus einer Produktion überprüft werden (insbesondere bei zerstörenden Prüfungen) und in einer Meinungsumfrage zur nächsten Wahl können nicht alle Wähler befragt werden. Man begnügt sich dann mit einer möglichst repräsentativen *Stichprobe* von Objekten aus der Grundgesamtheit, analysiert diese Stichprobe und möchte dann auch ausgehend von dieser Stichprobe "gültige" Rückschlüsse auf die interessierenden Merkmale in der Gesamtheit aller Objekte ziehen.

Bzgl. der Grundlagen von "vernünftigen" Experimenten und Erhebungen sowie der geschichtlichen und philosophischen Begründung des Einsatzes statistischer Analyseverfahren sei hier auf die ausführliche Einleitung in Hartung et al. (1982) hingewiesen.

Multivariate statistische Verfahren sind nun dadurch ausgezeichnet, daß sie die gemeinsame, gleichzeitige Analyse mehrerer Merkmale bzw. deren Ausprägungen erlaubt. Werden an Objekten (aus einer Grundgesamtheit) also die

Ausprägungen von mehreren Merkmalen beobachtet, so können alle Beobachtungsdaten mit Hilfe der Multivariaten Statistik gemeinsam ausgewertet werden. Der Vorteil gegenüber einzelnen, univariaten Analysen für jedes Merkmal besteht darin, daß auf diese Art die Abhängigkeiten zwischen den beobachteten Merkmalen berücksichtigt werden.

Wir werden uns in den Kapiteln I bis X dieses Buches mit den verschiedenen multivariaten Verfahren beschäftigen, wobei insbesondere auch ihre konkrete Anwendung auf Beobachtungsdaten im Vordergrund steht, und die benötigten Grundlagen aus Statistik und Mathematik bereitstellen. Hier soll zunächst ein kurzer Überblick über die behandelten Methoden gegeben werden, wobei die jeweils zu beantwortenden Fragestellungen - also das inhaltliche Ziel der Verfahren - im Vordergrund stehen sollen. Der detaillierter an den Voraussetzungen und Möglichkeiten multivariater statistischer Verfahren interessierte Leser sei auf die ausführlichen *Einleitungen der einzelnen Kapitel bzw. Abschnitte* hingewiesen, die auch *zahlreiche Beispiele* aus den *verschiedenen Anwendungsgebieten* enthalten.

Im Kapitel I werden wir zunächst in knapper Form die wesentlichen Grundlagen der Multivariaten Statistik behandeln. Wir beschäftigen uns mit der Beschreibung von Datenmaterial durch Kenngrößen (deskriptive Statistik), mit Elementen der Wahrscheinlichkeitsrechnung, als da sind Wahrscheinlichkeiten, bedingte Wahrscheinlichkeiten, Zufallsvariablen, Verteilungen, Verteilungsfunktionen, Dichten und Kenngrößen von Verteilungen, und mit der induktiven; schließenden Statistik, d.h. mit den Prinzipien von Punkt- und Bereichsschätzungen für unbekannte Parameter einer Verteilung und mit statistischen Tests über solche Parameter; dabei werden insbesondere die Normal- und die Binomialverteilung berücksichtigt. Weiterhin werden wir uns mit der Vektor- und Matrizenrechnung auseinandersetzen, die ein wesentliches Hilfsmittel der Multivariaten Statistik ist. Sodann werden mehrdimensionale und multivariate Normalverteilungen eingeführt, die bei vielen statistischen Verfahren eine große Rolle spielen. Abschließend beschäftigen wir uns noch mit der Gewinnung von Daten- und Distanzmatrix; Datenmatrizen enthalten die an Objekten beobachteten Ausprägungen mehrerer Merkmale und Distanzmatrizen beschreiben die Ähnlichkeiten von Objekten.

Das Kapitel II ist der *Regressionsanalyse* gewidmet; die dort beschriebenen Vorgehensweisen sind zwar selbst nicht im eigentlichen Sinne multivariat, jedoch ist ihre Bedeutung in den Anwendungen und für andere multivariate Verfahren so groß, daß wir sie nicht vernachlässigen wollten. Die Regressionsanalyse untersucht den *funktionalen Zusammenhang* zwischen einem ein-

zelenes Merkmal, das an Objekten beobachtet wird, und einer Reihe weiterer von den Objekten getragener Merkmale. Wir beschäftigen uns mit der Spezifikation der funktionalen Beziehung und mit Untersuchungen über die *Einflüsse der Merkmale*: Welche der Merkmale sind wesentlich zur Erklärung des beobachteten Merkmals, welche der Merkmale können bei Berücksichtigung der übrigen vernachlässigt werden? Ein weiteres Ziel der Regressionsanalyse ist natürlich die *Prognose* zukünftiger Werte.

Beispiel: Bei der Angebotserstellung für Produkte ist die Kalkulation der Produktionskosten von entscheidender Bedeutung. Diese Kosten hängen von verschiedenen Einflußgrößen wie etwa den Rohmaterialkosten, der zur Produktion benötigten Arbeitszeit, der zu produzierenden Menge etc. ab. Bestimmt man aufgrund der bekannten Produktionskosten früherer Waren mittels Regressionsanalyse den funktionalen Zusammenhang zwischen Produktionskosten und solchen Einflußgrößen, so lassen sich die die Produktionskosten wesentlich bestimmenden Einflußgrößen ermitteln und die Produktionskosten für ein neues Produkt prognostizieren.

Bei der Regressionsanalyse werden nun verschiedene Fälle unterschieden. Ist wie im obigen Beispiel das interessierende Merkmal quantitativ, d.h. kann es beliebige Werte in einem Bereich annehmen, so spricht man von *quantitativer Regressionsanalyse*. Werden dann die Einflüsse der übrigen Merkmale durch feste Parameter beschrieben, so lassen sich die Methoden der *multiplen Regressionsanalyse* anwenden; sind hingegen zumindest einige der Einflüsse als zufällig anzunehmen, so kommt man zu *Gemischten Linearen Modellen*. Im Falle eines qualitativen, beobachteten Merkmals, d.h. eines diskreten Merkmals mit nur einigen möglichen Ausprägungen, kommt die *diskrete Regressionsanalyse* (*Lineares Wahrscheinlichkeitsmodell*, *Logit-*, *Probitanalyse*) zur Anwendung.

Wird in der Regressionsanalyse ein funktionaler Zusammenhang zwischen verschiedenen Merkmalen hergestellt, so dient die in Kapitel III dargestellte *Korrelationsanalyse* der Bestimmung einer Maßzahl für die *Stärke eines Zusammenhangs*. Hier werden solche Korrelationsmaße für verschiedene Merkmaltypen vorgestellt. Dabei kann der Zusammenhang zwischen zwei Merkmalen (*einfache Korrelation*), der zwischen einem Merkmal und einer Gruppe anderer Merkmale (*multiple Korrelation*) oder auch der zwischen zwei Gruppen von Merkmalen (*kanonische Korrelation*) von Interesse sein. Mitunter ist eine Korrelation z.B. zwischen zwei Merkmalen nur dadurch bedingt, daß beide Merkmale mit weiteren, noch gar nicht berücksichtigten Merkmalen korreliert sind; der Ausschaltung solcher Einflüsse dienen die *partiellen* und *bi-*par-

tielle Korrelationsanalyse. Neben Maßen für die Stärke eines Zusammenhangs werden in diesem Kapitel z.B. auch statistische Tests angegeben, mit denen sich etwa überprüfen läßt, ob solche Zusammenhänge auch signifikant vorhanden sind.

Beispiel: Um die Eignung eines Bewerbers für eine bestimmte Position zu prüfen, werden oftmals Eignungstests durchgeführt. Solche Tests müssen natürlich so gestaltet sein, daß sie die tatsächliche Eignung eines Bewerbers möglichst gut widerspiegeln. Aufgrund der Testergebnisse früherer, bereits im Betrieb arbeitender Personen, deren tatsächliche Eignung sich inzwischen erwiesen hat, läßt sich mit Hilfe der Korrelationsanalyse die Stärke des Zusammenhangs zwischen Test und tatsächlicher Eignung ermitteln.

Werden an jeweils einer Reihe von Objekten aus r verschiedenen, vergleichbaren Grundgesamtheiten (Bewohner verschiedener Länder, Betriebe aus verschiedenen Branchen oder Regionen, Tiere oder Pflanzen gleicher Gattung aber verschiedener Art, Produkte aus verschiedenen Produktionslosen usw.) p Merkmale beobachtet, so spricht man vom *multivariaten* (p -variaten) r -*Stichprobenproblem*. Im Kapitel IV beschäftigen wir uns zunächst mit *multivariaten Ein- und Zweistichprobenproblemen* im Falle gemeinsam in der jeweiligen Grundgesamtheit normalverteilter Merkmale. Von Interesse sind dann die *Schätzung der Parameter* in den ein bzw. zwei Grundgesamtheiten sowie *Tests über diese Parameter*. Im Einstichprobenfall überprüft man, ob die Parameter signifikant von vorgegebenen Werten verschieden sind, und im Falle zweier Stichproben testet man die Gleichheit der Parameter in beiden Grundgesamtheiten. Für den Fall von $r > 2$ *Stichproben* behandeln wir im Kapitel IV zudem den Vergleich der Streuungsmatrizen in den r zugrundeliegenden Grundgesamtheiten sowie die Zuordnung "neuer" Objekte zu einer der Grundgesamtheiten; der Mittelwertvergleich im r -Stichprobenproblem wird im Zusammenhang mit der multivariaten Varianzanalyse erst im Kapitel X behandelt. Die Zuordnung "neuer" Objekte, an denen die p Merkmale beobachtet werden, zu einer von r Grundgesamtheiten nennt man auch *Diskrimination* oder *Identifikation*. Wir werden uns im Rahmen der Diskriminanzanalyse nicht nur mit *Zuordnungsvorschriften für "neue" Objekte* sondern vielmehr auch mit der *Auswahl wesentlicher Merkmale* für die Diskrimination beschäftigen.

Beispiel: In einem metallverarbeitenden Betrieb werden Werkstücke aus verschiedenen Legierungen gefertigt. Beobachtet man an einigen Werkstücken, von denen man weiß, zu welchem von zwei unterschiedlichen Gefügen sie gehören, nun Merkmale wie Festigkeit, spezifisches Gewicht usw., so kann aufgrund der Verfahren für multivariate Zweistichprobenprobleme überprüft wer-

den, ob Unterschiede bzgl. Mittelwert und Streuung der Merkmale zwischen den Gefügen bestehen. Basierend auf den Zuordnungsvorschriften der Diskriminanzanalyse kann aufgrund der Beobachtungen der Merkmale an weiteren Werkstücken überprüft werden, welchem der Gefüge diese Werkstücke zuzuordnen sind. Außerdem können die für Unterschiede zwischen den Gefügen wesentlich verantwortlichen Merkmale bestimmt werden. Die nicht oder kaum zwischen den Gefügen trennenden Merkmale werden bei der Zuordnung "neuer" Werkstücke dann oft gar nicht mehr berücksichtigt.

Viele (multivariate) statistische Verfahren lassen sich nur anwenden, wenn die an einer Reihe von Objekten beobachteten Merkmale quantitativer Natur sind. Sind die Ausprägungen der interessierenden Merkmale (z.T.) qualitativer Art (z.B. Farben, Bewertungen, Nationalitäten, Berufe usw.) so müssen entweder qualitative Verfahren angewandt werden oder die Daten müssen für die statistische Analyse zunächst aufbereitet werden. Ein universelles Instrumentarium stellt hier die *Skalierung qualitativer Merkmalsausprägungen* dar, die im Kapitel V ausführlich behandelt wird; der Vorteil einer derartigen *Datenaufbereitung* gegenüber direkten Verfahren für qualitative Merkmale liegt insbesondere auch darin, daß die Anforderungen an die Datenbasis sehr viel geringer sind. Skalierungsverfahren ordnen den qualitativen, kategoriellen Merkmalsausprägungen Zahlen derart zu, daß die skalierten Daten dem vorgesehenen Auswertungsverfahren möglichst gut angepaßt sind. Wir beschäftigen uns zunächst mit der *Skalierung eines einzelnen Merkmals*, dessen qualitative Ausprägungen einer natürlichen Rangfolge unterliegen (*marginale Normalisierung*), und dann mit der *Skalierung zweier qualitativer Merkmale*, wobei die Ausprägungen des einen Merkmals derart skaliert werden, daß sie das andere Merkmal möglichst gut erklären, und umgekehrt, d.h. die Korrelation der skalierten Merkmale wird maximiert. Ausgehend von diesen Skalierungsverfahren werden einige spezielle statistische Verfahren wie z.B. die Güteprüfung solcher Skalierungen, die Diskrimination neuer Objekte zu den Ausprägungen eines der beiden Merkmale usw. behandelt. Werden *mehr als zwei qualitative Merkmale* beobachtet, so empfiehlt es sich, sie alle gemeinsam zu skalieren; ähnlich wie im Fall zweier Merkmale kann die gemeinsame Skalierung derart erfolgen, daß der Gesamtzusammenhang zwischen allen skalierten Merkmalen möglichst groß ist. Oftmals ist es nun so, daß eines oder mehrere der beobachteten Merkmale eine Sonderstellung einnehmen: Wie bei der Regressionsanalyse möchte man diese ausgezeichneten Merkmale, die auch *Kriteriumsvariablen* genannt werden, möglichst gut durch die übrigen Merkmale erklären. In diesem Fall kann eine Skalierung nach dem Kriterium der maximalen multiplen bzw. kanonischen Korrelation erfolgen. Schließlich werden wir im Kapitel V noch Skalierungsverfahren behandeln,

die gemeinsam mit den qualitativen Merkmalen beobachtete, quantitative Merkmale (*gemischte Datentypen*) berücksichtigen.

Beispiel:

(a) Im Abschnitt 3 des Kapitels V werden wir uns ausführlich mit einem Problem der Produktforschung beschäftigen. Von der "Consumers Union of the United States" wurde eine Befragung von 391 Autobesitzern über die Reparaturanfälligkeit verschiedener Teile ihrer PKW's durchgeführt, um Unterschiede zwischen den verschiedenen PKW-Herstellern aufzudecken. Wir berücksichtigen bei der Skalierung der insgesamt 14 Reparaturanfälligkeitsmerkmale (z.B. Bremsen, Automatikgetriebe, innere und äußere Karosserie, Heizungs- und Belüftungssystem) die Kriteriumsvariable "Hersteller" auf 5 Stufen (American Motors, Chrysler, Ford, General Motors, (bzgl. des US-Marktes) ausländische Hersteller), überprüfen die Güte der Skalierung, bestimmen die sieben für die Unterschiede zwischen den Herstellern wesentlich verantwortlichen Merkmale usw. Ausgehend von den sieben wesentlichen Merkmalen wird dann eine Datenmatrix und eine Distanzmatrix für die fünf Hersteller erstellt; die Datenmatrix enthält mittlere Skalenwerte der sieben Reparaturanfälligkeitsmerkmale bei jedem Hersteller und die Distanzmatrix beschreibt basierend auf diesen mittleren Skalenwerten die Ähnlichkeiten der Hersteller zueinander.

(b) Um zu einem optimalen Einsatz von Werbung zu gelangen, werden bei 10 Produkten die Werbekosten, die Art der Werbung (aggressiv, einschmeichelnd, erlebnisweckend) sowie der daraufhin zu verzeichnende Gewinn und der Geschäftstyp mit relativ höchstem Absatz (Kaufhaus, Supermarkt, Kleingeschäft) erhoben. Skaliert man bei diesen gemischt qualitativen und quantitativen Beobachtungsdaten die Merkmale Werbeart und Geschäftstyp mit relativ höchstem Absatz, so lassen sich hierzu die Methoden des Multivariaten Linearen Modells, das wir in Kapitel X behandeln, einsetzen. Natürlich wird man bei der Skalierung so vorgehen wollen, daß aufgrund der Merkmale Werbeart und Werbeart dann möglichst gut Gewinn und Geschäftstyp mit relativ höchstem Absatz für "neue" Produkte prognostiziert werden können.

Einen gänzlich anderen Zweck als die in Kapitel V dargestellte Skalierung von Ausprägungen qualitativer Merkmale verfolgt die sogenannte *Multidimensionale Skalierung (MDS)*, die wir im Kapitel VI behandeln. Ausgehend von Ähnlichkeitsinformationen über eine Reihe interessierender Objekte wird bei der Multidimensionalen Skalierung eine (q -dimensionale) Skala bestimmt, auf der die Objekte darstellbar sind; d.h. zu jedem der Objekte wird ein q -dimensionaler Datenvektor derart bestimmt, daß die Distanzen zwischen den Objekten durch die Abstände zwischen den Datenvektoren möglichst gut

approximiert werden. Ist die gewählte Dimension q des Repräsentationsraumes kleiner als vier, so lassen sich die Objekte als Ergebnis der Multidimensionalen Skalierung auch graphisch darstellen. Bei der *metrischen multidimensionalen Skalierung* wird ausgehend von einer Distanzmatrix für die interessierenden Objekte eine q -dimensionale Konfiguration so bestimmt, daß die ursprünglichen Distanzen zwischen je zwei Objekten möglichst gut durch die Distanzen der Skalenvektoren (zahlenmäßig) approximiert werden. Beim klassischen Verfahren, der *Haupt-Koordinaten-Methode* geschieht dies auf rein algebraische Art und Weise, wohingegen das *Nonlinear-Mapping-Verfahren*, dessen Zielsetzung die Erhaltung lokaler Strukturen in den Distanzen ist, iterativ arbeitet. Bei der *nichtmetrischen multidimensionalen Skalierung* werden konkrete Zahlenangaben für den Grad der Ähnlichkeit bzw. Verschiedenheit der interessierenden Objekte nicht benötigt. Das *Verfahren von Kruskal* geht von der Rangfolge der Ähnlichkeiten von je zwei Objekten aus und konstruiert eine Skala derart, daß diese Rangfolge möglichst exakt beibehalten wird. Die *Unfolding-Technik* hingegen benötigt sogenannte I -Skalen für jedes der Objekte; die I -Skala für ein Objekt gibt dabei die Rangfolge der Ähnlichkeiten der übrigen Objekte zu diesem Objekt an. Davon ausgehend wird dann eine (eindimensionale) Skala bestimmt, die möglichst gut die I -Skala wiedergibt.

Will man sich die *Ähnlichkeiten* der an Objekten beobachteten *Merkmale* veranschaulichen, so kann z.B. basierend auf den Korrelationen der Merkmale auch hierzu ein Verfahren der Multidimensionalen Skalierung verwandt werden.

Beispiel: Wir kommen hier noch einmal auf das Beispiel (a) zu Kapitel V zurück. Dort wurden Reparaturanfälligkeitsmerkmale von PKW's gegen die Kriteriumsvariable Hersteller skaliert und aufgrund der sieben wesentlichen Merkmale konnten eine Daten- und eine Distanzmatrix für die fünf betrachteten Hersteller bestimmt werden. Ausgehend von einer Distanzmatrix für die fünf Hersteller, die ja Informationen über die Ähnlichkeiten bzgl. der Reparaturanfälligkeitsmerkmale enthält, ist in Abb.2 eine dreidimensionale Konfiguration für die Hersteller graphisch dargestellt, die mittels eines MDS-Verfahrens gewonnen wurde. Umgekehrt können basierend auf den Korrelationen der Reparaturanfälligkeitsmerkmale auch diese repräsentiert werden.

Im Kapitel VII beschäftigen wir uns dann mit der *Clusteranalyse*, d.h. der *Klassifikation von Objekten (Taxonomie, pattern recognition)*. Ausgehend von einer Datenmatrix oder einer Distanzmatrix werden bei der Clusteranalyse die interessierenden Objekte in Klassen, Gruppen eingeteilt, und zwar derart, daß die Objekte, die zur selben Klasse gehören, einander möglichst

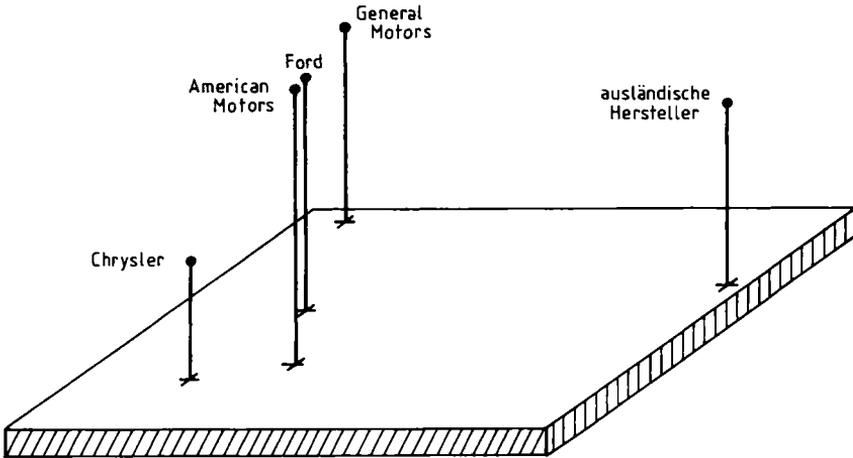


Abb.2: Dreidimensionale MDS - Konfiguration für fünf Autohersteller basierend auf Reparaturanfälligkeiten verschiedener PKW - Teile

ähnlich sind und Objekte aus verschiedenen Klassen sich möglichst stark unterscheiden, d.h. die Klassen sollen in sich homogen und untereinander heterogen sein. Wir werden Verfahren zur Konstruktion von vier verschiedenen *Klassifikationstypen* behandeln. Die Klassen einer *Überdeckung* können sich überschneiden, jedoch darf keine Klasse von Objekten vollständig in einer anderen Klasse enthalten sein. Eine *Partition* ist ein Spezialfall des Klassifikationstyps *Überdeckung*; hier werden die Objekte in disjunkte Klassen von Objekten, also Klassen, die sich auch nicht teilweise überschneiden, eingeteilt. *Quasihierarchien* und *Hierarchien* sind Verfeinerungen der beiden vorgenannten Klassifikationstypen. Eine *Quasihierarchie* wird gebildet durch eine Folge von *Überdeckungen*; ausgehend von einer feinsten *Überdeckung* (Überdeckung mit den meisten Klassen) werden auf jeder Stufe der *Quasihierarchie* Vergrößerungen vorgenommen, indem einander ähnliche Klassen von Objekten zu einer einzigen Klasse zusammengefaßt werden. Ebenso wird eine *Hierarchie* durch eine Folge von immer gröber werdenden *Partitionen* gebildet. *Quasihierarchien* und *Hierarchien* lassen sich in Form von "*Stammbäumen*", *Hierarchien* auch durch sogenannte *Dendrogramme* veranschaulichen. Möchte man auch *Partitionen* oder *Überdeckungen* graphisch darstellen, so kann man mittels *Multidimensionaler Skalierung*, vgl. Kapitel VI, zunächst eine etwa zweidimensionale Konfiguration für die interessierenden Objekte bestimmen und in diese dann die Klassen der *Partitionen* bzw. *Überdeckungen* einzeichnen. Wie schon bei der *Multidimensionalen Skalierung* sind auch die Verfahren der *Clusteranalyse* auf an Objekten beobachtete Merkmale anwendbar, wenn man etwa von den *Korrelationen* der Merkmale ausgeht. Betrachtet man

die mittels Clusteranalyse klassifizierte Objekte als Lernstichprobe aus einer größeren Gesamtheit, so können mittels spezieller auf Klassifikation zugeschnittener Diskriminationsverfahren auch weitere, "neue" Objekte den Klassen der Klassifikation zugeordnet werden.

Beispiel: Aufgrund der Evolution, d.h. der stammesgeschichtlichen Entwicklung von Lebewesen, sind heute existierende Tier- und Pflanzenarten mehr oder weniger verwandt. Basierend auf einer hierarchischen Clusteranalyse kann man versuchen, diese Entwicklung zu rekonstruieren ("Evolutions - Bäume"). Jede Stufe der Hierarchie entspricht dann einer Evolutionsphase und die Klassen der feinsten Partition werden gerade durch die heute existierenden, interessierenden Arten gebildet. Je gröber die Partition wird, desto weiter bewegt man sich in die Vergangenheit der Stammesgeschichte.

Die an Objekten beobachteten Merkmale sind in der Regel miteinander korreliert und lassen sich auf *latente*, "künstliche" Merkmale (*Faktoren*), die selbst nicht beobachtet werden können, zurückführen. Beispielsweise dienen Eignungstests für Bewerber zur Feststellung der charakterlichen Eignung, der Führungsqualitäten usw. Solche Eigenschaften sind nicht direkt meßbar oder beobachtbar und werden daher basierend auf speziellen Frage- bzw. Aufgabenstellungen überprüft. Die *Faktorenanalyse*, mit der wir uns im Kapitel VIII beschäftigen wollen, dient nun der Reduktion einer Vielzahl beobachteter Merkmale auf wenige latente, sie beschreibende Merkmale. Zunächst werden orthogonale, unkorrelierte Faktoren derart bestimmt, daß durch sie ein möglichst großer Teil der Korrelationen der Merkmale erklärt wird. Hier gibt es die verschiedensten Verfahren, die zu diesem Zwecke eine *Ladungsmatrix* konstruieren, welche die Korrelationen der beobachteten Merkmale mit den latenten Faktoren enthält; bei höchstens drei extrahierten Faktoren lassen sich, basierend auf einer solchen Ladungsmatrix, die beobachteten Merkmale auch *graphisch im Raum der Faktoren darstellen*. Es gibt unendlich viele in obigem Sinne optimale Ladungsmatrizen; daher stellen die verschiedenen *Konstruktionsverfahren*, von denen wir die Maximum-Likelihood-Methode, die kanonische Faktorenanalyse, die Hauptfaktorenanalyse, die Zentroidmethode und das Jöreskog-Verfahren vorstellen wollen, stets zusätzliche Bedingungen. Sie liefern keine im Sinne bestmöglicher Interpretierbarkeit der latenten Faktoren in Bezug auf die Merkmale optimale Lösung. Daher werden im Anschluß an die Konstruktion von Ladungsmatrizen zu meist noch *Faktorrotationen* durchgeführt, deren aller Ziel das Erreichen einer möglichst guten Interpretierbarkeit der Faktoren im Sinne des von Thurstone geprägten Begriffes der *Einfachstruktur* ist. Man unterscheidet bei den Rotationsverfahren zwischen orthogonalen und schiefwinkligen Rota-

tionen. Bei einer *Orthogonalrotation* wie z.B. der Varimax- und der Quartimax-Rotation bleibt die Orthogonalität der latenten Faktoren erhalten, wohingegen *schiefwinklige Rotationsverfahren*, von denen wir die Methode der Primärfaktoren behandeln, auf korrelierte, schiefwinklige Faktoren führen. Schließlich wird im Rahmen der Faktorenanalyse noch das Schätzen von sogenannten *Faktorenwerten* behandelt. Das Ziel hierbei ist, einem Objekt, an dem die zugrundegelegten Merkmale beobachtet wurden, einen niederdimensionalen Datenvektor bzgl. der Faktoren zuzuordnen. Natürlich lassen sich bei weniger als vier Faktoren dann die Objekte aufgrund dieser Datenvektoren im Koordinatensystem der Faktoren auch graphisch veranschaulichen. Grundvoraussetzung der Faktorenanalyse ist, daß die Zahl der betrachteten Objekte größer ist als die Zahl der an ihnen beobachteten Merkmale. Ist dies nicht der Fall, so wird die *Faktorenanalyse* oft auch *auf die Objekte* anstelle der Merkmale *angewandt*, was zwar im strengen Sinne nicht erlaubt ist, häufig jedoch zu gut interpretierbaren Ergebnissen führt, vgl. auch die Einleitung des Kapitels VIII.

Beispiel: An verschiedenen Klimastationen in Europa werden Merkmale wie mittlere Monatstemperatur, mittlere Jahrestemperatur, Temperaturschwankung, Zahl von Frost- und Sonnentagen, relative Luftfeuchtigkeit, Niederschlagsmenge, Zahl der Regen-, Gewitter-, Schneefalltage usw. erhoben. Führt man nun eine Faktorenanalyse durch, so läßt sich feststellen, daß alle diese Merkmale im wesentlichen durch vier latente Faktoren beschrieben werden können: die thermischen Verhältnisse, die Humidität, die thermische und hygrische Kontinentalität bzw. Ozeanität. Schätzt man nun die zugehörigen Faktorenwerte für die einzelnen Klimastationen, so ist es möglich, festzustellen, welche Regionen bzgl. dieser Faktoren einander ähnlich sind, also eine Klimaregion bilden.

Wie bereits eingangs dieser Einleitung erwähnt, dienen *graphische Verfahren* der Veranschaulichung komplexer Datensituationen bzw. der Darstellung von Ergebnissen (multivariater) statistischer Analysen; solche Verfahren werden im Kapitel IX ausführlich behandelt. Zunächst werden zwei spezielle Verfahren zur übersichtlichen Darstellung *eindimensionaler Daten*, "Stem and Leaves" und "Box-Plot", sowie eine Möglichkeit zur Repräsentation *zweidimensionaler Daten* am Beispiel eines *Produkt-Markt-Portfolios* vorgestellt. *Q-Q-Plots* dienen der *Überprüfung von Verteilungsannahmen* für mehrere Merkmale sowie dem Erkennen von *Ausreißern* in multivariaten Daten. In den übrigen Kapiteln des Buches werden immer wieder graphische Verfahren zur Repräsentation von Objekten oder den an ihnen beobachteten Merkmalen als Ergebnis einer multivariaten statistischen Analyse angegeben; im *Bi-Plot* lassen sich nun Merk-

male und Objekte gleichzeitig zweidimensional veranschaulichen, und zwar werden die Objekte basierend auf ihren Ähnlichkeiten und den konkret an ihnen beobachteten Merkmalswerten, die Merkmale basierend auf ihren Korrelationen und Streuungen repräsentiert. Bei den bisher angesprochenen Verfahren werden alle interessierenden Objekte und/oder an ihnen beobachtete Merkmale gleichzeitig in einem Bild dargestellt. Andere graphische Repräsentationsformen stellen *jedes Objekt in einem separaten Bild* dar, natürlich ausgehend von den an ihm beobachteten Merkmalsausprägungen. Die einfachsten Darstellungsformen sind die, bei denen nur die Ausprägungen jedes Merkmals durch Streckungen oder Winkel dargestellt werden: *Profile, Polygonzüge, Glyphs, Sterne, Sonnen, Diamanten*. Bei den *Flury - Riedwyl - Faces*, vgl. auch Abb.1, wird jedes Objekt durch ein Gesicht repräsentiert. Die *Form von Gesichtsteilen* wie Gesichtslinie, Mund, Nase, Augenbraue, Auge, Pupille, Haarschraffur usw. hängt dabei von den beobachteten Merkmalswerten ab. *Andrews - Plots* und *Blumen* dienen der Repräsentation von Objekten durch *trigonometrische Funktionen*, deren Koeffizienten durch die beobachteten Merkmalsausprägungen bestimmt sind. Bei den bisher angesprochenen graphischen Verfahren ist die Merkmalsanordnung beliebig. Dagegen erfolgt bei *Quadern, Bäumen und Burgen*, vgl. z.B. den in Abb.1 dargestellten *Kleiner - Hartigan - Tree*, die *Anordnung der Merkmale im Bild gemäß ihrer Ähnlichkeiten*. Für alle Objekte wird zunächst ausgehend von diesen Ähnlichkeiten eine gemeinsame Struktur bestimmt, die dann je nach beobachteten Merkmalsausprägungen für jedes Objekt variiert. Bei den *Facetten* erfolgt die *Merkmalsanordnung nach der Wichtigkeit* für die Diskrimination zwischen den Objekten. Die Wichtigkeit der Merkmale wird durch für alle Objekte gleiche Winkel, die konkreten Ausprägungen durch Strecken beschrieben. Schließlich werden bei den *Bi - Plot - Sonnen* die Korrelationen der Merkmale mit dargestellt. Ausgehend von der *Bi - Plot - Darstellung* der Merkmale werden die an den Objekten beobachteten Merkmalswerte durch Strecken repräsentiert, vgl. auch die Abb.1.

Beispiel: Die wirtschaftliche Entwicklung eines Betriebes oder einer Branche läßt sich anhand der Beobachtung von Bilanzkennzahlen über mehrere Jahre hinweg beurteilen. Im letzten Abschnitt des Kapitels IX wird mittels graphischer Verfahren die Entwicklung der chemischen Industrie der Bundesrepublik Deutschland in den Jahren 1965 bis 1980 dargestellt. Basierend auf Anlagenintensität, Eigenkapitalanteil, Eigenkapitalrendite (Return on Investment), Umsatzrendite, Liquidität, dynamischem Verschuldungsgrad und Kapitalumschlag werden die Jahre, die hier die Rolle der Objekte übernehmen, durch *Flury - Riedwyl - Faces, Andrews - Plots, Burgen, Bäume und Facetten* repräsentiert.

Im abschließenden Kapitel X beschäftigen wir uns schließlich noch mit dem *Multivariaten Linearen Modell*. Hier wird der Zusammenhang von m Einflußvariablen und p beobachteten Merkmalen untersucht. Grundsätzlich wird dabei zwischen *Modellen mit festen Effekten* und *Modellen mit zufälligen Effekten* unterschieden; bei ersteren werden die Einflußgrößen als fest, bei letzteren als zufällig angenommen. Wir beschäftigen uns zunächst mit den *allgemeinen Vorgehensweisen* bei festen Einflußvariablen, wie Parameterschätzungen, Vertrauensintervallen für die Parameter und für zukünftige Beobachtungen (*Prognosen*), Tests über die Parameter und entsprechende Testverfahren, usw. Sodann werden zwei spezielle Modellklassen behandelt. Bei der *multivariaten Regressionsanalyse* sind alle Einflußvariablen quantitativer Art, bei der *multivariaten Kovarianzanalyse* sind einige Einflußfaktoren quantitativ (Kovariablen), andere qualitativer Art (Faktoren auf endlich vielen Stufen, die mit 0 und 1 kodiert werden, je nachdem ob die entsprechende Stufe eines Faktors vorliegt oder nicht).

Beispiel: Wir haben im Beispiel (b) zum Kapitel V die Skalierung der Merkmale Gewinn und Geschäftstyp mit relativ höchstem Absatz gegen die Einflußvariablen Werbeetat und Werbeart angesprochen. Verwendet man die Skalenwerte für die Werbeart, so läßt sich der Einfluß der Werbung auf die beiden interessierenden Merkmale in einem multivariaten Regressionsmodell untersuchen. Betrachtet man hingegen die Werbeart als qualitativen Faktor auf den drei Stufen aggressiv, einschmeichelnd und erlebnisweckend, so wird die Analyse in einem Kovarianzanalysemodell mit der Kovariablen Werbeetat durchgeführt.

Weiterhin beschäftigen wir uns mit Modellen der *multivariaten Varianzanalyse* sowie der *Profilanalyse*, bei denen alle Einflußfaktoren qualitativer Art sind. Im Zusammenhang mit der multivariaten Varianzanalyse werden das Modell mit nur einem Einflußfaktor, dessen Wirkung auf die p interessierenden Merkmale untersucht wird, sowie verschiedene zweifaktorielle Versuchsanlagen (einfaches multivariates Blockexperiment, zweifache Kreuzklassifikation mit Wechselwirkungen, zweifach hierarchische Klassifikation) behandelt. Bei der Profilanalyse steht nicht der generelle Einfluß des qualitativen Faktors auf p Merkmale im Vordergrund sondern vielmehr die Untersuchung eines zeitlichen Verlaufs (*Verlaufskurven*); dazu wird nur ein einzelnes Merkmal, dieses aber zu p verschiedenen Zeitpunkten beobachtet und mit den Stufen eines qualitativen Einflußfaktors in Verbindung gebracht.

Beispiel: Um den Einfluß des Faktors "Region" auf den SO_2 -Gehalt in der Luft zu untersuchen, werden in m Gebieten (Stufen des Faktors "Region")

an verschiedenen Stellen (Wiederholungen) die SO_2 -Gehaltswerte gemessen und zwar am selben Tag jeweils alle zwei Stunden ($p = 12$). Mittels einfaktorier multivariater Varianzanalyse kann nun untersucht werden, ob der Faktor "Region" einen generellen Einfluß auf die SO_2 -Gehaltswerte zu verschiedenen Tageszeiten hat. Die Profilanalyse ermöglicht zudem die Untersuchung von Fragestellungen wie z.B.: Sind die Verläufe der Gehaltswerte in den Regionen wesentlich verschieden oder verlaufen sie weitgehend gleichläufig, nur auf unterschiedlichem Niveau; sind die Tagesmittelwerte des SO_2 -Gehalts in den Regionen als gleich anzusehen, d.h. sind die durchschnittlichen Umweltbelastungen durch SO_2 in den Regionen praktisch gleich?

Schließlich behandeln wir im Kapitel X dann spezielle Modelle mit zufälligen Effekten, wobei hier die Untersuchung des Einflusses qualitativer Variablen im Vordergrund steht. Erstrecken sich in Modellen mit festen Effekten die Aussagen nur auf die konkret im Experiment betrachteten Stufen qualitativer Einflußvariablen, so geht man nun davon aus, daß die konkreten Stufen nur eine Auswahl aus allen möglichen Faktorstufen darstellen, eine Aussage sich aber auf sämtliche Stufen beziehen soll. Die Beantwortung solcher Fragestellungen ermöglichen *multivariate Varianzkomponentenmodelle*, in denen die auf die Einflußfaktoren zurückzuführenden Streuungsanteile der p beobachteten Merkmale geschätzt werden. Ein hoher Streuungsanteil ist dann mit einem starken Einfluß des Faktors gleichzusetzen. Wir behandeln hier Modelle der einfach, zweifach und dreifach hierarchischen Klassifikation (ein, zwei bzw. drei hierarchisch angeordnete Faktoren) sowie Modelle der zweifachen Kreuzklassifikation (zwei Einflußfaktoren, deren Stufen kreuzweise kombiniert betrachtet werden). Außerdem beschäftigen wir uns noch mit einem speziellen *Meßmodell zur Präzisionsbestimmung von Meß- und Analyseverfahren bei zerstörenden Prüfungen bzw. Produktvariabilität*.

Beispiel: In einem Erzlager werden aus einigen Bohrlöchern Bodenproben entnommen und bzgl. ihres Gehalts an Metall und an taubem Gestein analysiert. Betrachtet man hier ein Modell mit festen Effekten, so kann sich eine Aussage nur über die speziellen Bohrlöcher erstrecken. Möchte man aber die Streuung von Metallgehalt und Gehalt an taubem Gestein im gesamten Erzlager beurteilen, so kommt man zu einem Modell mit zufälligen Effekten und betrachtet dann die konkreten Bohrlöcher als zufällig ausgewählt aus der Gesamtheit aller möglichen Bohrlöcher im Erzlager.

Am Ende dieser Einleitung wollen wir noch kurz auf zwei Unterscheidungsmerkmale hinweisen, die häufig zur Klassifikation (multivariater) statisti-

scher Verfahren verwandt werden. Zunächst einmal unterscheidet man zwischen sogenannten *dependenten* und *interdependenten Methoden* und dann noch zwischen *R-Techniken* und *Q-Techniken*.

Dependente Verfahren sind dadurch ausgezeichnet, daß der Einfluß einer Gruppe von Variablen auf eine Reihe interessierender, beobachteter Merkmale untersucht wird, wie dies bei der Regressions-, Varianz-, Kovarianzanalyse und der Profilanalyse der Fall ist, oder daß der Zusammenhang zwischen zwei Merkmalsgruppen im Vordergrund des Interesses steht, was bei den Verfahren der Korrelationsanalyse zutrifft. Bei den *interdependenten* (intra-dependenten) statistischen Verfahren hingegen sind alle an Objekten beobachteten bzw. erhobenen Merkmale gleichberechtigt in dem Sinne, daß sich Aussagen und Analyseergebnisse stets auf alle diese Merkmale in gleicher Art und Weise erstrecken; solche interdependenten Verfahren sind etwa die Multidimensionale Skalierung, die Clusteranalyse, die Faktorenanalyse und die meisten graphischen Methoden.

Die Aufteilung in R- und Q-Techniken bezieht sich nicht auf eine Unterscheidung in den Merkmalen. Vielmehr spricht man von *R-Techniken*, wenn Aussagen über die in einem Experiment beobachteten Merkmale getroffen werden, und von *Q-Techniken*, wenn sich Analyseergebnisse direkt auf die im Experiment berücksichtigten Objekte beziehen. Hier ist eine starre Einteilung der verschiedenen multivariaten statistischen Verfahren kaum möglich, da viele Methoden wahlweise als R- oder Q-Technik angewandt werden können. So ist z.B. die Clusteranalyse an und für sich eine Q-Technik für die Objekte; sie kann jedoch auch als R-Technik für die Merkmale eingesetzt werden, was z.B. bei den im Kapitel IX beschriebenen graphischen Verfahren teilweise der Fall ist (Burgen, Bäume, Quader). Die Faktorenanalyse ist eine R-Technik für die Merkmale, so lange das Interesse in der Bestimmung latenter Faktoren, welche die Merkmale beschreiben, liegt. Das sich anschließende Schätzen von Faktorenwerten hingegen ist eine Q-Technik, denn hier wird den Objekten je ein Datenvektor bzgl. der latenten Faktoren zugeordnet. Zudem wird - wie bereits erwähnt - die Faktorenanalyse häufig mit durchaus gut interpretierbaren Ergebnissen als Q-Technik angesetzt, wenn die Zahl der Objekte geringer ist als die Zahl der an ihnen beobachteten Merkmale.

Kapitel I: Einführung und Grundlagen

In diesem ersten Kapitel werden die für die multivariate Statistik unerlässlichen Grundlagen aus der Wahrscheinlichkeitsrechnung, Statistik und Matrizenrechnung kurz zusammengestellt, um es dem Leser zu ermöglichen, dieses Buch selbständig zu lesen.

Im Abschnitt 1 werden zunächst einige grundlegende Begriffe im Zusammenhang mit Datenmaterial und die wesentlichen Elemente der deskriptiven (beschreibenden) Statistik, wie Häufigkeiten, empirische Kenngrößen und empirische Verteilungsfunktion, dargestellt.

Der zweite Abschnitt ist der Wahrscheinlichkeitsrechnung gewidmet. Hier werden Wahrscheinlichkeiten, bedingte Wahrscheinlichkeiten, Zufallsvariablen, Verteilungen, Verteilungsfunktionen, Dichten und Kenngrößen von Verteilungen behandelt.

Im dritten Abschnitt beschäftigen wir uns dann mit den Prinzipien der induktiven (schließenden) Statistik. Die Begriffe Punktschätzer und Konfidenzintervallschätzung sowie statistischer Test werden eingeführt und Konstruktionsprinzipien für Punkt- und Intervallschätzer sowie statistische Tests werden erläutert.

Der vierte Abschnitt dieses Kapitels beschäftigt sich mit der Vektor- und Matrizenrechnung. Elemente dieser Gebiete, die für das Verständnis multivariater Verfahren unerlässlich sind, werden in einem kurzen Abriss behandelt.

Im fünften Abschnitt werden dann die Grundlagen der Wahrscheinlichkeitsrechnung für mehrdimensionale Zufallsvariablen (Zufallsvektoren, Zufallsmatrizen) bereitgestellt. Insbesondere werden die mehrdimensionale und die multivariate Normalverteilung sowie die Wishartverteilung betrachtet.

Schließlich werden wir uns im sechsten Abschnitt mit einigen elementaren Begriffen im Zusammenhang mit Daten- und Distanzmatrizen für mehrdimensionale Beobachtungsvektoren auseinandersetzen. Es werden Distanzmaße für

beliebig skalierte Merkmale untersucht.

1 GRUNDLEGENDE BEGRIFFE UND ELEMENTARE DATENBESCHREIBUNG

Die Statistik beschäftigt sich mit der Beschreibung und Analyse umfangreicher Datenmengen, die z.B. durch Befragungen oder Messungen gewonnen werden. Die *Objekte*, an denen Messungen vorgenommen werden, bzw. die Personen, die befragt werden, nennt man auch *Untersuchungseinheiten*. Die Größen, die gemessen bzw. nach denen gefragt wird, heißen *Merkmale*, und die an den Untersuchungseinheiten festgestellten Werte der Merkmale heißen *Merkmalswerte* oder *Merkmalsausprägungen*.

Beispiel: Um die Altersstruktur in einer Gemeinde zu erfassen, werden alle Einwohner (Objekte, Untersuchungseinheiten) nach ihrem Alter (Merkmal) befragt. Die Altersangaben der Personen sind dann die Merkmalsausprägungen.]

Oft kann man nicht alle interessierenden Objekte untersuchen, sondern muß sich mit einem Teil (z.B. n Objekten) von ihnen begnügen. Man spricht dann von einer *Stichprobe* von n Objekten aus der *Grundgesamtheit* der interessierenden Objekte. Wie man Stichprobenuntersuchungen anlegt, damit sie ein möglichst genaues Bild der Grundgesamtheit widerspiegeln, wird ausführlich in Hartung et al. (1982, Kap.V) dargestellt.

Beispiel: In einem Land ohne Meldepflicht ist es kaum möglich alle Einwohner (Grundgesamtheit) nach ihrem Alter zu befragen. Man muß sich mit einer Stichprobe von Einwohnern begnügen.]

1.1 MERKMALSTYPEN UND KLASSENBILDUNG

Merkmale werden nach der Art ihrer Ausprägungen klassifiziert. Zum einen unterscheidet man quantitative und qualitative Merkmale und zum anderen stetige und diskrete Merkmale.

Die Ausprägungen *quantitativer Merkmale* (z.B. Alter und Gewicht von Personen, Lebensdauern von Bauelementen) unterscheiden sich in ihrer Größe und die Ausprägungen *qualitativer Merkmale* (z.B. Farbe, Rasse, Fabrikat, Beruf) in ihrer Art.

Geht man von der Zahl der möglichen Ausprägungen eines Merkmals aus, so

gelangt man zur Aufteilung in diskrete und stetige Merkmale. *Diskrete Merkmale* (z.B. Geschlecht, Anzahlen) können nur abzählbar viele Werte annehmen, *stetige Merkmale* (z.B. Längenmessungen, Lebensdauern) können beliebige Werte in einem bestimmten Bereich annehmen.

Eine weitere Unterscheidung von Merkmalstypen erhält man dadurch, daß man die zugrundeliegende Meß - Skala betrachtet. Die Werte einer *Nominalskala* unterliegen keiner Reihenfolge und sind nicht vergleichbar, die einer *Ordinalskala* unterscheiden sich in ihrer Intensität und unterliegen einer Rangfolge und die einer *metrischen Skala* sind derart, daß sich zudem Abstände zwischen den Werten interpretieren lassen. *Nominale Merkmale* sind also z.B. Beruf oder Rasse, *ordinale Merkmale* haben z.B. Ausprägungen wie gut, mittel, schlecht (etwa Zensuren) und *metrische Merkmale* sind z.B. Größe oder Gewicht.

Mitunter werden auch, wenn ein Merkmal viele Ausprägungen besitzt, mehrere Merkmalsausprägungen in Klassen zusammengefaßt. Dadurch erreicht man etwa bei der graphischen Darstellung von Datenmaterial, daß die Darstellung übersichtlich ist. Durch eine solche *Klassenbildung* gelangt man dann z.B. zur Diskretisierung eines stetigen Merkmals (z.B. Gewichtsklassen bei Hühnereiern).

In den folgenden Teilen des Abschnitts 1 werden einige wesentliche Möglichkeiten der Beschreibung von Datenmaterial dargestellt. Der stärker an Methoden der beschreibenden (deskriptiven) Statistik interessierte Leser sei auf Hartung et al. (1982, Kap.I) hingewiesen.

1.2 HÄUFIGKEITEN, SUMMENHÄUFIGKEITEN UND EMPIRISCHE VERTEILUNGSFUNKTION

Bezeichnet man die möglichen Ausprägungen eines Merkmals X mit a_1, \dots, a_k und beobachtet das Merkmal X dann an n Untersuchungseinheiten, so bezeichnet man mit

$$H_n(a_j) = \text{"Anzahl der Fälle, in denen } a_j \text{ auftritt"}$$

für $j=1, \dots, k$ die *absolute Häufigkeit* des Auftretens der Ausprägung a_j und mit

$$h_n(a_j) = \frac{1}{n} H_n(a_j)$$

für $j=1, \dots, k$ die *relative Häufigkeit* der Ausprägung a_j . Die relative Häufigkeit von a_j ist also der prozentuale Anteil der Untersuchungseinheiten

ten, die die Ausprägung a_j tragen. Daher nimmt $h_n(a_j)$ unabhängig von n stets Werte zwischen 0 und 1 an und es gilt:

$$\sum_{j=1}^k h_n(a_j) = h_n(a_1) + \dots + h_n(a_k) = 1 \quad ;$$

natürlich ist

$$\sum_{j=1}^k H_n(a_j) = H_n(a_1) + \dots + H_n(a_k) = n \quad .$$

Bei ordinalen und metrischen Merkmalen lassen sich die Ausprägungen der Größe nach ordnen: $a_1 < a_2 < \dots < a_k$, und man betrachtet dann oft sogenannte Summenhäufigkeiten. Die *absolute Summenhäufigkeit* der Ausprägung a_j ist

$$\sum_{i=1}^j h_n(a_i) = H_n(a_1) + \dots + H_n(a_j) \quad ,$$

also die Anzahl der Untersuchungseinheiten, an denen eine der Ausprägungen a_1, a_2, \dots, a_j beobachtet wurde. Entsprechend ist

$$\sum_{i=1}^j h_n(a_i) = h_n(a_1) + \dots + h_n(a_j)$$

die *relative Summenhäufigkeit* der Ausprägung a_j .

Durch die Folge der relativen Summenhäufigkeiten wird nun die *empirische Verteilungsfunktion* (*Summenhäufigkeitsfunktion*) $S_n(x)$ des Merkmals X bestimmt. Und zwar ist für alle Zahlen x mit $a_j \leq x < a_{j+1}$

$$S_n(x) = \sum_{i=1}^j h_n(a_i) = \frac{1}{n} \sum_{i=1}^j H_n(a_i) \quad .$$

Die empirische Verteilungsfunktion ist also eine sogenannte Treppenfunktion mit

$$S_n(x) = 0 \quad \text{für } x < a_1 \quad , \quad S_n(x) = 1 \quad \text{für } x \geq a_k \quad ,$$

und Sprungstellen bei a_1, a_2, \dots, a_k . Die Höhe der Sprungstellen ist für $i=1, \dots, k$ gerade $h_n(a_i)$, also die relative Häufigkeit der Ausprägung a_i .

Beispiel: In einer Klausur mit 54 Teilnehmern wurde 3 mal die Note 1, 6 mal die Note 2, 18 mal die Note 3, 18 mal die Note 4 und 9 mal die Note 5 erreicht. In der *Tab.1* sind nun absolute und relative Häufigkeiten bzw. Summenhäufigkeiten der Ausprägungen (Noten) $a_1 = 1$, $a_2 = 2$, $a_3 = 3$, $a_4 = 4$ und $a_5 = 5$ des Merkmals $X = \text{Klausurnote}$ angegeben. Aus den relativen Häufigkeiten bzw. Summenhäufigkeiten ergibt sich die in *Abb.1* dargestellte

empirische Verteilungsfunktion $S_n(x)$ der Klausurnoten.

Tab.1: Klausurnoten von $n = 54$ Teilnehmern

Ausprägung (Klausur- note)	absolute Häufigkeit	relative Häufigkeit	absolute Summen- häufigkeit	relative Summen- häufigkeit
a_j	$H_n(a_j)$	$h_n(a_j)$	$\sum_{i=1}^j H_n(a_i)$	$\sum_{i=1}^j h_n(a_i)$
$a_1 = 1$	3	$3/54 = 1/18$	3	$3/54 = 1/18$
$a_2 = 2$	6	$6/54 = 1/9$	$3+6 = 9$	$9/54 = 1/6$
$a_3 = 3$	18	$18/54 = 1/3$	$3+6+18 = 27$	$27/54 = 1/2$
$a_4 = 4$	18	$18/54 = 1/3$	$3+6+18+18 = 45$	$45/54 = 5/6$
$a_5 = 5$	9	$9/54 = 1/6$	$3+6+18+18+9 = 54$	$54/54 = 1$

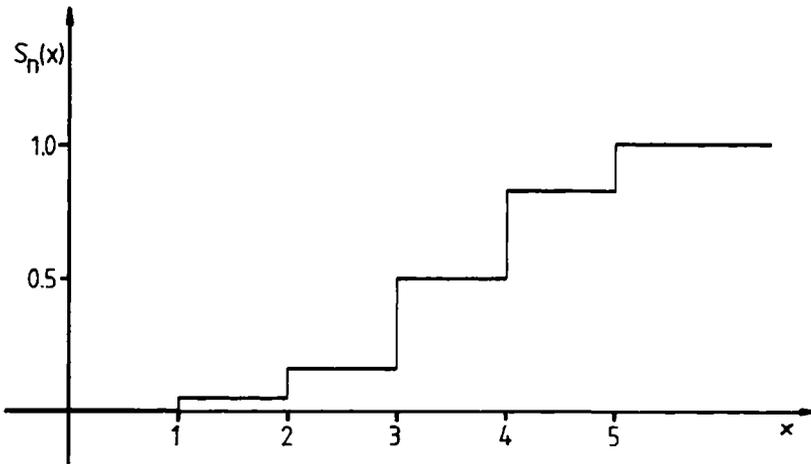


Abb.1: Empirische Verteilungsfunktion $S_n(x)$ der Klausurnoten von 54 Teilnehmern

1.3 EMPIRISCHE LAGEMASSE

Empirische Lagemaße geben ein "Zentrum" von n beobachteten Merkmalswerten x_1, \dots, x_n eines Merkmals X mit Ausprägungen a_1, \dots, a_k an.

Das wohl wichtigste Lagemaß ist das arithmetische Mittel (durchschnittlicher Wert, Mittelwert)

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{n} \sum_{j=1}^k a_j H_n(a_j) \quad ;$$

es ist insbesondere sinnvoll bei metrisch skalierten Größen.

Auch bei ordinal skalierten Merkmalen läßt sich der *Median* oder *Zentralwert* als Lagemaß verwenden, der dadurch charakterisiert ist, daß mindestens 50% der Beobachtungswerte x_1, \dots, x_n größer oder gleich und 50% kleiner oder gleich dem Median sind. Sind $x_{(1)} \leq \dots \leq x_{(n)}$ die der Größe nach geordneten Beobachtungswerte (bei Gleichheit mehrerer Werte spielt deren Reihenfolge keine Rolle), so ist

$$\bar{x}_{0.5} = \begin{cases} x_{((n+1)/2)} & , \text{ falls } n \text{ ungerade} \\ \frac{1}{2}(x_{(n/2)} + x_{((n+2)/2)}) & , \text{ falls } n \text{ gerade} \end{cases}$$

ein Median der Beobachtungsreihe.

Der *Modalwert* ist der häufigste Wert in einer Beobachtungsreihe. Damit gilt für den Modalwert x_{mod} , der sich auch noch bei nominalen Merkmalen als Lagemaß verwenden läßt,

$$h_n(x_{\text{mod}}) \geq h_n(a_j)$$

für alle Merkmalsausprägungen $a_j, j=1, \dots, k$. Ist diese Bedingung für mehrere Ausprägungen erfüllt, so ist es nicht sinnvoll, den Modalwert als Lagemaß zu verwenden.

Beispiel: Bei $n = 35$ Glühbirnen wurden die in **Tab.2** angegebenen Lebensdauern gemessen. Die beobachteten Werte x_1, \dots, x_{35} sind der Größe nach geordnet und stimmen somit direkt mit den Größen $x_{(1)}, \dots, x_{(35)}$ überein.

Tab.2: Lebensdauern x_i (in Stunden) von $n = 35$ Glühbirnen

i	x_i	i	x_i	i	x_i	i	x_i	i	x_i
1	430	8	611	15	647	22	705	29	798
2	580	9	612	16	647	23	730	30	815
3	581	10	617	17	648	24	732	31	832
4	582	11	617	18	683	25	743	32	832
5	595	12	628	19	696	26	789	33	856
6	596	13	629	20	703	27	789	34	857
7	608	14	631	21	705	28	789	35	938

Für diese Beobachtungswerte ergeben sich das arithmetische Mittel zu

$$\bar{x} = \frac{1}{35}(430 + 580 + \dots + 938) = \frac{1}{35} \cdot 24251 = 692.886 \quad ,$$

der Median, da $n = 35$ ungerade ist, zu

$$\bar{x}_{0.5} = x_{((n+1)/2)} = x_{(36/2)} = x_{18} = 683$$

und der Modalwert (der hier bei so vielen und dünn besetzten Klassen jedoch wenig Information enthält) zu

$$x_{\text{mod}} = 789 \quad .$$

Weitere empirische Lagemaße findet man bei Hartung et al. (1982, Kap.I, Abschnitt 4).

1.4 EMPIRISCHE STREUUNGSMAßE

Beschreiben Lagemaße ein Zentrum einer Beobachtungsreihe, so geben Streuungsmaße Aufschluß darüber, wieweit ein Beobachtungswert von einem solchen Zentrum abweichen kann.

Das wohl wichtigste Streuungsmaß ist die (empirische) *Varianz* s^2 einer Beobachtungsreihe x_1, \dots, x_n für $n \geq 2$:

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

(Mitunter wird bei der empirischen Varianz statt des Faktors $\frac{1}{n-1}$ auch der Faktor $\frac{1}{n}$ verwandt.) Die *Standardabweichung* s ist gerade die Wurzel aus der Varianz, also

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

und hat gegenüber der Varianz den Vorteil, daß bei ihr die gleiche Maßeinheit zugrundeliegt, wie bei den Beobachtungswerten.

Der *Variationskoeffizient* v ist zudem vom arithmetischen Mittel \bar{x} bereinigt, d.h. ohne Nennung von \bar{x} interpretierbar:

$$v = \frac{s}{\bar{x}} \quad , \text{ definiert für positive Meßwerte } x_1, \dots, x_n \quad .$$

Allen bisher erwähnten Maßen ist gemeinsam, daß als Bezugsgröße das arith-

metische Mittel \bar{x} verwandt wird. Dies ist durch die Minimaleigenschaft des arithmetischen Mittels begründet:

$$\sum_{i=1}^n (x_i - \bar{x})^2 \leq \sum_{i=1}^n (x_i - c)^2 \quad \text{für jede beliebige Zahl } c.$$

Dagegen wird bei der *mittleren absoluten (betragsmäßigen) Abweichung* (vom Median)

$$d = \frac{1}{n} \sum_{i=1}^n |x_i - \tilde{x}_{0.5}|$$

aufgrund der Minimaleigenschaft des Medians

$$\sum_{i=1}^n |x_i - \tilde{x}_{0.5}| \leq \sum_{i=1}^n |x_i - c| \quad \text{für jede beliebige Zahl } c$$

der Median $\tilde{x}_{0.5}$ als Bezugsgröße des Streuungsmaßes verwandt.

Ein grober Anhaltspunkt für die Lage der Beobachtungswerte ist der sogenannte *Strebereich* $[x_{(1)}, x_{(n)}]$, also das Intervall, dessen Grenzen durch kleinsten und größten Beobachtungswert gegeben sind. Das zugehörige Streuungsmaß ist die *Spannweite* (*range*) der Beobachtungswerte:

$$R = x_{(n)} - x_{(1)} \quad .$$

Beispiel: Alle angegebenen Streuungsmaße sollen nun für die Daten aus Tab.2 im Abschnitt 1.3, wo auch schon die empirischen Lagemaße angegeben sind, berechnet werden. Es ergibt sich für die Varianz

$$s^2 = \frac{1}{35-1} \sum_{i=1}^{35} (x_i - 692.886)^2 = \frac{1}{34} \cdot 387875.54 = 11408.104 \quad ,$$

für die Standardabweichung

$$s = \sqrt{11408.104} = 106.809 \quad ,$$

für den Variationskoeffizienten

$$v = \frac{106.809}{692.886} = 0.154 \quad ,$$

für die mittlere absolute Abweichung vom Median

$$d = \frac{1}{35} \sum_{i=1}^{35} |x_i - 683| = \frac{1}{35} \cdot 3050 = 87.143 \quad ;$$

der Streubereich ist das Intervall

$$[x_{(1)}; x_{(35)}] = [430; 938] \quad ,$$

und somit ergibt sich die Spannweite der Beobachtungswerte zu

$$R = 938 - 430 = 508 \quad .$$

Weitere Streuungsmaße sowie nähere Erläuterungen zu den hier erwähnten findet man bei Hartung et al. (1982, Kap.I, Abschnitt 5).

2 GRUNDLAGEN DER WAHRSCHEINLICHKEITSRECHNUNG

Sind Vorgänge zufälliger Natur oder derart komplex, daß sie nicht quantitativ erfaßt werden können, so spricht man von *zufälligen Ereignissen*. Mit der Untersuchung solcher Ereignisse beschäftigt sich die Wahrscheinlichkeitsrechnung und Statistik.

Ein *Zufallsexperiment* ist die Ausführung eines Vorgangs, der (beliebig oft) wiederholt werden kann und dessen Ausgang zufällig ist. Die Menge der möglichen Ausgänge eines Zufallsexperimentes nennt man auch *Grundraum* oder *Ereignisraum* und bezeichnet sie mit Ω . Die zufälligen Ereignisse sind dann Teilmengen dieses Grundraums, wobei einelementige Ereignisse auch *Elementarereignisse* genannt werden.

Beispiel: Durch die zufällige Auswahl eines Fernsehgerätes aus einem Produktionslos ist ein Zufallsexperiment gegeben. Da als Lebensdauer des Fernsehgerätes jede Zahl $x \geq 0$ in Frage kommt, ist der Grundraum durch diese Zahlen gegeben. Das Ereignis "das Fernsehgerät hat eine Lebensdauer von mindestens 300 Tagen und höchstens 800 Tagen" ist das Intervall $[300; 800]$.

Das zu einem Ereignis A *komplementäre Ereignis* \bar{A} , ist das Ereignis "A tritt nicht ein". Das Ereignis $A \cup B$ (Vereinigung) ist das Ereignis, daß A oder B oder beide Ereignisse eintreten. $A \cap B$ (Durchschnitt) ist das Ereignis, daß sowohl A als auch B eintreten. Ist der Durchschnitt $A \cap B$ zweier Ereignisse A und B leer, so heißen A und B *disjunkte Ereignisse*. Die leere Menge \emptyset heißt *unmögliches Ereignis* und der Grundraum Ω heißt *sicheres Ereignis*.

2.1 WAHRSCHEINLICHKEITEN UND BEDINGTE WAHRSCHEINLICHKEITEN

Bei einem Zufallsexperiment ist das Eintreten bestimmter Ereignisse zwar nicht vorhersehbar, jedoch mehr oder weniger wahrscheinlich. Zum Beispiel ist die Wahrscheinlichkeit dafür, daß beim Würfeln eine bestimmte Zahl fällt, nicht sehr groß. Man weiß aber, daß beim Würfeln jede Zahl zwischen 1 und 6 die gleiche Chance hat. Führt man das Zufallsexperiment "Würfeln" nun immer wieder durch, so wird die relative Häufigkeit des Werfens z.B. einer 4 sich etwa bei $1/6$ stabilisieren. Dieser Wert $1/6$ wird als *Wahrscheinlichkeit* des Ereignisses "Werfen einer 4" interpretiert und man schreibt kurz

$$P(\text{Werfen einer 4}) = \frac{1}{6} \quad ,$$

dabei steht P für *Probability* (Wahrscheinlichkeit).

Allgemein lassen sich Wahrscheinlichkeiten durch die drei Eigenschaften

$$P(A) \geq 0 \quad \text{für jedes Ereignis } A \quad (\text{Positivität})$$

$$P(\Omega) = 1 \quad (\text{Normiertheit})$$

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \quad \text{für jede Folge paarweise disjunkter Ereignisse } A_i \subset \Omega \quad (\sigma\text{-Additivität}) \quad ,$$

die auch *Kolmogoroffsche Axiome* genannt werden, vollständig charakterisieren. Aus diesen Eigenschaften lassen sich alle *Rechenregeln für Wahrscheinlichkeiten* folgern, wie z.B.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad \text{für beliebige Ereignisse } A \text{ und } B \quad ,$$

$$P(\bar{A}) = 1 - P(A) \quad ,$$

$$P(A - B) = P(A) - P(B) \quad , \quad \text{falls } B \subset A \quad .$$

Vielfach ist man auch daran interessiert, die Wahrscheinlichkeit zu kennen, mit der ein Ereignis A eintritt, falls ein anderes Ereignis B bereits eingetreten ist. Wie groß ist etwa die Wahrscheinlichkeit, daß ein Fernsehgerät 7 Jahre hält (Ereignis A), wenn es bereits 5 Jahre (Ereignis B) überlebt hat. Eine solche Wahrscheinlichkeit $P(A | B)$ nennt man eine *bedingte Wahrscheinlichkeit* des Ereignisses A unter der Bedingung B und legt sie fest als

$$P(A | B) = \frac{P(A \cap B)}{P(B)} \quad \text{für } P(B) > 0 \quad .$$

Mit B und \bar{B} als komplementäre Ereignisse sind auch $A \cap B$ und $A \cap \bar{B}$ disjunkt, und es gilt natürlich

$$\begin{aligned} P(A) &= P(A \cap B) \cup P(A \cap \bar{B}) = P(A \cap B) + P(A \cap \bar{B}) \\ &= \frac{P(A \cap B)}{P(B)} \cdot P(B) + \frac{P(A \cap \bar{B})}{P(\bar{B})} \cdot P(\bar{B}) \\ &= P(A | B) \cdot P(B) + P(A | \bar{B}) \cdot P(\bar{B}) \quad . \end{aligned}$$

Diese Beziehung ist ein Spezialfall des Satzes von der totalen Wahrscheinlichkeit: Sind B_1, \dots, B_k paarweise disjunkte Ereignisse mit $\bigcup_{i=1}^k B_i = \Omega$, so gilt für ein beliebiges Ereignis A

$$P(A) = \sum_{i=1}^k P(A | B_i) P(B_i) \quad .$$

Die Berechnung solcher bedingter Wahrscheinlichkeiten vereinfacht sich bei "unabhängigen" Ereignissen, wie z.B. den einzelnen Ereignissen beim mehrmaligen Würfeln. Wir sagen zwei Ereignisse A und B sind (*stochastisch unabhängig*), wenn gilt

$$P(A \cap B) = P(A) \cdot P(B) \quad ,$$

was gleichbedeutend ist mit

$$P(A | B) = P(A) \quad \text{bzw.} \quad P(B | A) = P(B) \quad .$$

Aus der Definition der bedingten Wahrscheinlichkeit und dem Satz von der totalen Wahrscheinlichkeit ergibt sich die sogenannte *Bayessche Formel*, die sehr nützlich bei der Berechnung bedingter Wahrscheinlichkeiten ist: Für paarweise disjunkte Ereignisse B_1, \dots, B_k mit $\bigcup_{i=1}^k B_i = \Omega$ und ein beliebiges Ereignis A gilt

$$P(B_i | A) = \frac{P(A | B_i) \cdot P(B_i)}{\sum_{j=1}^k P(A | B_j) \cdot P(B_j)} \quad .$$

Beispiel: Durch einen zu spät erkannten Fabrikationsfehler sind in einer Automobilproduktion genau zwanzig defekte Lenkgetriebe eingebaut worden. In einer Rückrufaktion werden alle 200000 Wagen dieser Serie überprüft und ein als fehlerhaft eingestuftes Lenkgetriebe durch ein neues ersetzt. Dabei wird mit 99% Sicherheit die Überprüfung zu einem korrekten Ergebnis führen. Wie groß ist die Wahrscheinlichkeit, daß ein ausgewechseltes Lenkgetriebe auch defekt ist? Bezeichnet nun B das Ereignis eines defekten Lenkgetriebes und A das eines ausgewechselten, so lassen sich die gegebenen Informationen schreiben als

$$P(B) = \frac{20}{200000} = 0.0001, \quad P(A | B) = 0.99 \quad \text{und} \quad P(A | \bar{B}) = 1 - 0.99 = 0.01.$$

Gesucht ist $P(B | A)$, und es ergibt sich mit der Bayesschen Formel ($B = B_1$, $\bar{B} = B_2$)

$$P(B | A) = \frac{P(A | B) \cdot P(B)}{P(A | B) \cdot P(B) + P(A | \bar{B}) \cdot P(\bar{B})} = \frac{0.99 \cdot 0.0001}{0.99 \cdot 0.0001 + 0.01 \cdot 0.9999} = 0.01.$$

Fast alle ausgewechselten Lenkgetriebe waren demnach nicht defekt. └

2.2 ZUFALLSVARIABLE UND VERTEILUNGEN

Würfelt man zweimal hintereinander, so kann man sich etwa für die Summe $X(i, j)$, kurz X , der beiden Augenzahlen i und j interessieren. Mögliche Ereignisse sind dann zum Beispiel $X = 4$ oder $10 \leq X \leq 12$. Eine solche Funktion X , die den Ergebnissen eines Zufallsexperiments reelle Zahlen so zuordnet, daß für die mit Hilfe von X beschriebenen Ereignisse Wahrscheinlichkeiten angebar sind, heißt *Zufallsvariable*. Eine *Realisation* x von X ist dann der Wert, den die Zufallsvariable X bei Durchführung des Zufallsexperimentes annimmt.

Jedes Zufallsexperiment kann durch eine (oder mehrere) Zufallsvariable X beschrieben werden und jedes Ereignis kann als $X \in B$, wobei B eine Teilmenge des Wertebereichs von X ist, angegeben werden.

Gibt man zu einer Zufallsvariablen X die Wahrscheinlichkeiten der Ereignisse $X \leq x$ an, so lassen sich daraus die Wahrscheinlichkeiten aller weiteren Ereignisse bestimmen, d.h. man kennt die *Wahrscheinlichkeitsverteilung* von X . Die Funktion

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

heißt dann auch die *Verteilungsfunktion* der Zufallsvariablen X .

Beispiel:

(a) Eine Zufallsvariable X mit der Verteilungsfunktion

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(\xi - \mu)^2}{2\sigma^2}} d\xi, \quad \text{vgl. Abb. 2},$$

heißt normalverteilt mit Parametern μ und σ^2 (kurz: $X \sim N(\mu; \sigma^2)$).

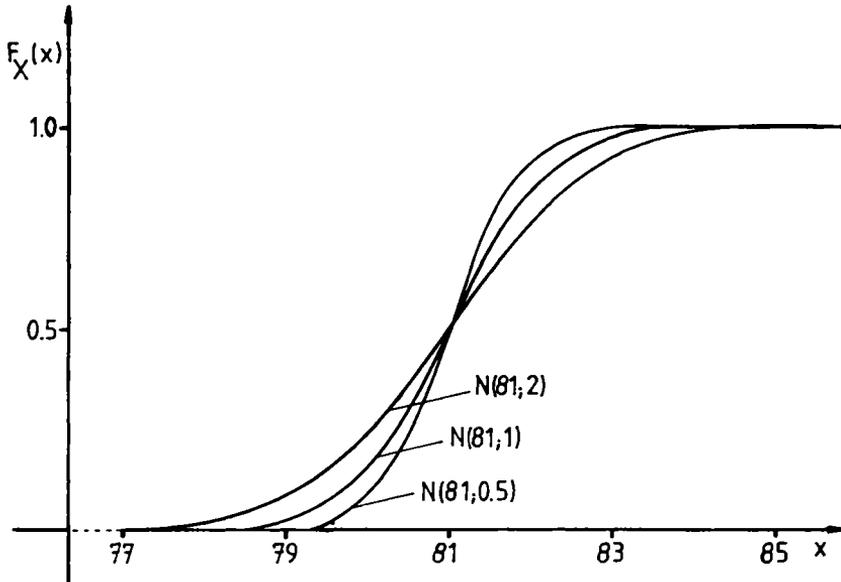


Abb.2: Verteilungsfunktion der $N(81;0.5)$ -, $N(81;1)$ - und $N(81;2)$ - Verteilung

Speziell nennt man eine $N(0;1)$ - Verteilung auch *Standardnormalverteilung* und bezeichnet die zugehörige Verteilungsfunktion mit

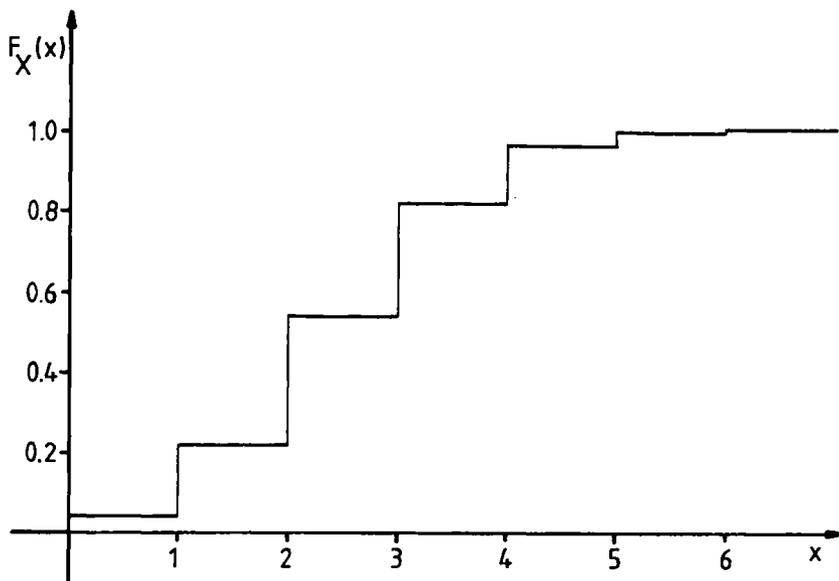
$$\Phi(x) = F_X(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{\xi^2}{2}} d\xi \quad \text{für } X \sim N(0;1) \quad .$$

(b) Eine Zufallsvariable X , deren Realisationen nur die Zahlen $0, 1, 2, \dots, n$ sein können und für die gilt:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad ,$$

heißt *binomialverteilt* mit Parametern n und p (kurz: $X \sim B(n, p)$). Die Verteilungsfunktion einer $B(n, p)$ - verteilten Zufallsvariablen ist somit

$$F_X(x) = P(X \leq x) = \sum_{i=0}^k \binom{n}{i} p^i (1-p)^{n-i} \quad \text{für } k \leq x < k+1, \\ \text{vgl. Abb.3} \quad .$$

Abb.3: Verteilungsfunktion einer $B(6,0.4)$ - Verteilung

Allgemein wird die Verteilungsfunktion einer Zufallsvariablen X durch die folgenden drei Eigenschaften charakterisiert:

F_X ist monoton nicht fallend, d.h. $F_X(x_1) \leq F_X(x_2)$ für $x_1 < x_2$,

F_X ist rechtsseitig stetig, d.h. $\lim_{\substack{h \rightarrow 0 \\ h > 0}} F_X(x+h) = F_X(x)$,

$\lim_{x \rightarrow \infty} F_X(x) = 1$ und $\lim_{x \rightarrow -\infty} F_X(x) = 0$.

Wie bereits im obigen Beispiel sichtbar, gibt es zwei verschiedene Arten von Verteilungsfunktionen. Im Beispiel (b) kann X lediglich die Werte $0, 1, 2, \dots, n$ annehmen, und die Verteilungsfunktion wird durch die Angabe der Wahrscheinlichkeiten dieser $n+1$ Ereignisse vollständig beschrieben. Eine Zufallsvariable X mit endlich oder abzählbar unendlich vielen möglichen Realisationen x_1, x_2, \dots heißt auch *diskret verteilt*; die Gesamtheit der Werte x_1, x_2, \dots heißt dann die *Trägermenge* von X . Eine normalverteilte Zufallsvariable kann dagegen unendlich viele Werte annehmen und jede Realisation x von X hat dabei die Wahrscheinlichkeit 0, denn

$$P(X = x) = P(X \leq x) - P(X < x) = F_X(x) - \lim_{\substack{h \rightarrow 0 \\ h > 0}} P(X \leq x+h)$$

$$= F_X(x) - \lim_{\substack{h \rightarrow 0 \\ h > 0}} F_X(x+h) = 0$$

Die Verteilungsfunktion einer solchen Zufallsvariablen X kann oft in der Form

$$F_X(x) = \int_{-\infty}^x f_X(\xi) d\xi$$

angegeben werden, wobei die nichtnegative Funktion f_X *Dichtefunktion* oder *Dichte* genannt wird. Die Dichte der $N(\mu; \sigma^2)$ - Verteilung ist z.B.

$$f_X(x) = \frac{1}{\sqrt{2\pi} \sigma} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad \text{vgl. Abb.4}$$

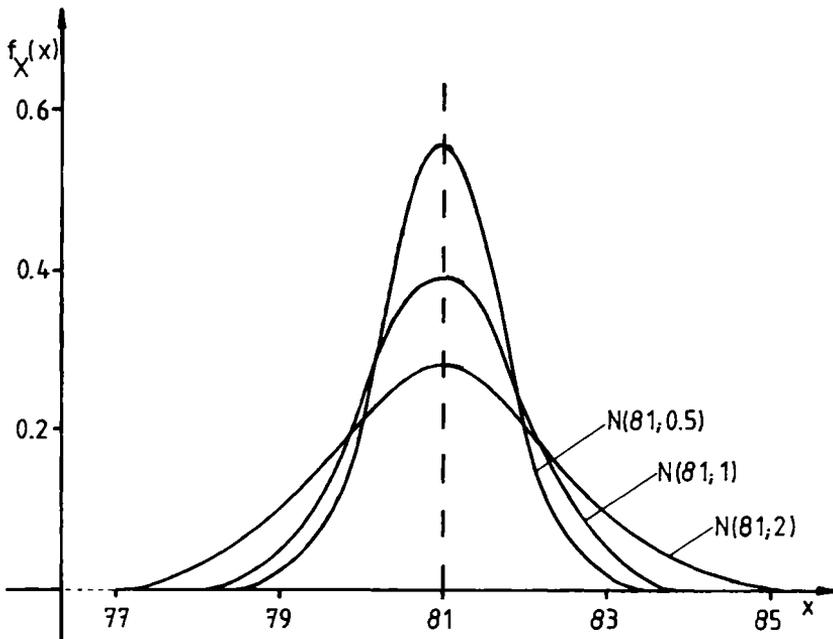


Abb.4: Dichte der $N(81;0.5)$ -, $N(81;1)$ - und $N(81;2)$ - Verteilung

Jede Wahrscheinlichkeitsverteilung einer Zufallsvariablen X , für die eine Dichte existiert, heißt eine *stetige Verteilung*. Der Zusammenhang zwischen Dichte und Verteilungsfunktion läßt sich folglich so beschreiben: Der Wert $F_X(x)$ gibt die Fläche oberhalb der x -Achse und unterhalb der Dichte

f_X zwischen $-\infty$ und x an. Die Gesamtfläche unterhalb von f_X (und oberhalb der x -Achse) ist natürlich Eins.

In diesem Zusammenhang sei noch erwähnt, daß n Zufallsvariablen X_1, \dots, X_n (stochastisch) unabhängig heißen, falls für alle Zahlen x_1, \dots, x_n gilt:

$$P(X_1 \leq x_1, \dots, X_n \leq x_n) = P(X_1 \leq x_1) \cdot \dots \cdot P(X_n \leq x_n) \quad .$$

Bei diskret verteilten Zufallsvariablen X_1, \dots, X_n läßt sich natürlich schon aus

$$P(X_1 = x_1, \dots, X_n = x_n) = P(X_1 = x_1) \cdot \dots \cdot P(X_n = x_n)$$

auf die Unabhängigkeit von X_1, \dots, X_n schließen.

2.3 KENNGRÖßEN VON ZUFALLSVARIABLEN

So wie im Abschnitt 1 Kenngrößen empirischer Verteilungen betrachtet wurden, werden hier Kenngrößen zur Charakterisierung der Wahrscheinlichkeitsverteilung von Zufallsvariablen eingeführt.

Der wichtigste Lageparameter ist der *Erwartungswert* oder *Mittelwert* einer Zufallsvariablen bzw. ihrer Verteilung. Für eine stetige Zufallsvariable X mit der Dichte f_X ergibt er sich im Falle seiner Existenz zu

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx$$

und für eine diskrete Zufallsvariable X mit Trägermenge x_1, x_2, \dots (das sind die Werte, die mit positiver Wahrscheinlichkeit angenommen werden) ist

$$E(X) = \sum_{i=1}^{\infty} x_i P(X = x_i) \quad .$$

Für die Berechnung von Erwartungswerten sind folgende Regeln nützlich. Ist $Y = \alpha E(X) + \beta$, α und β feste Zahlen, so gilt

$$E(Y) = \alpha E(X) + \beta \quad .$$

Sind X_1, \dots, X_n Zufallsvariablen, so ist

$$E(X_1 + \dots + X_n) = E(X_1) + \dots + E(X_n) \quad .$$

Beispiel:

(a) Für die Standardnormalverteilung $N(0;1)$ ergibt sich mit $c = \sqrt{2\pi}$

$$E(X) = \frac{1}{c} \int_{-\infty}^{\infty} x \cdot e^{-\frac{1}{2}x^2} dx = \frac{1}{c} \int_0^{\infty} |x| \cdot e^{-\frac{1}{2}x^2} dx - \frac{1}{c} \int_{-\infty}^0 |x| \cdot e^{-\frac{1}{2}x^2} dx = 0 \quad .$$

(b) Der Erwartungswert einer $B(n,p)$ -verteilten Zufallsvariablen X ergibt sich zu

$$\begin{aligned} E(X) &= \sum_{i=0}^n i \binom{n}{i} p^i (1-p)^{n-i} = \sum_{i=1}^n i \binom{n}{i} p^i (1-p)^{n-i} \\ &= \sum_{i=1}^n i \cdot \frac{n!}{i!(n-i)!} p^i (1-p)^{n-i} = np \cdot \sum_{i=1}^n \frac{(n-1)!}{(i-1)!(n-i)!} p^{i-1} (1-p)^{n-i} \\ &= np \cdot \sum_{i=0}^{n-1} \frac{(n-1)!}{i!(n-1-i)!} p^i (1-p)^{n-1-i} = np \cdot \sum_{i=0}^{n-1} \binom{n-1}{i} p^i (1-p)^{n-1-i} \\ &= np \quad . \end{aligned}$$

Ein anderes Lagemaß ist der Median $\xi_{0.5}$, für den gilt

$$P(X \leq \xi_{0.5}) \geq \frac{1}{2} \quad \text{und} \quad P(X \geq \xi_{0.5}) \geq \frac{1}{2} \quad .$$

Der Median läßt sich mit Hilfe der verallgemeinerten inversen Verteilungsfunktion

$$F_X^{-1}(y) = \inf\{x \mid F_X(x) \geq y\} \quad ,$$

die mit der gewöhnlichen Inversen übereinstimmt, falls es sie gibt, auch schreiben als

$$\xi_{0.5} = F_X^{-1}(0.5) \quad .$$

Der Median ist ein spezielles α -Quantil $\xi_\alpha = F_X^{-1}(\alpha)$, nämlich gerade das 0.5-Quantil. Besitzt nun X eine stetige Verteilung, so gilt für das α -Quantil ξ_α

$$F_X(\xi_\alpha) = \int_{-\infty}^{\xi_\alpha} f_X(x) dx = \alpha \quad ;$$

besitzt X eine diskrete Verteilung, so ist

$$F_X(\xi_\alpha) \geq \alpha \quad \text{und} \quad F_X(x) < \alpha \quad \text{für jedes } x < \xi_\alpha \quad .$$

Es sei noch erwähnt, daß das α -Quantil gelegentlich $(1-\alpha)$ -Fraktile genannt wird.

Beispiel:

(a) Wie bei jeder symmetrischen Verteilung, stimmt bei der Standardnormalverteilung der Median mit dem Erwartungswert überein, denn

$$F_X(0) = \int_{-\infty}^0 \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}x^2} dx = 0.5 \quad .$$

Die Quantile der Standardnormalverteilung werden mit u_α bezeichnet. Aus der Tabelle im Anhang ergibt sich z.B. $u_{0.95} = 1.6449$ und wegen der Symmetrie der Dichte der Standardnormalverteilung ergibt sich $u_\alpha = -u_{1-\alpha}$, also $u_{0.05} = -1.6449$. vgl. Abb.5.

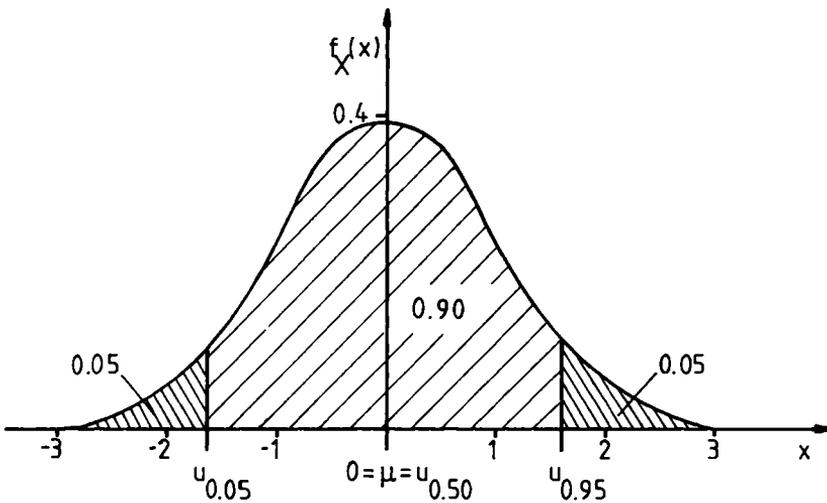


Abb.5: Dichte der $N(0;1)$ -Verteilung mit einigen Quantilen

(b) Für eine $B(4,0.2)$ -verteilte Zufallsvariable X gilt:

$$F_X(0) = \binom{4}{0} 0.2^0 (1-0.2)^{4-0} = 0.4096 \quad ,$$

$$F_X(1) = F_X(0) + \binom{4}{1} 0.2^1 (1-0.2)^{4-1} = 0.4096 + 0.4096 = 0.8192 \quad ,$$

$$F_X(2) = F_X(1) + \binom{4}{2} 0.2^2 (1-0.2)^{4-2} = 0.8192 + 0.1536 = 0.9728 \quad ,$$

$$F_X(3) = F_X(2) + \binom{4}{3} 0.2^3 (1-0.2)^{4-3} = 0.9728 + 0.0256 = 0.9984 \quad ,$$

$$F_X(4) = F_X(3) + \binom{4}{4} 0.2^4 (1-0.2)^{4-4} = 0.9984 + 0.0016 = 1 \quad ,$$

so daß sich z.B. der Median zu 1 und das 0.99 - Quantil zu 3 ergeben, vgl. Abb.6.

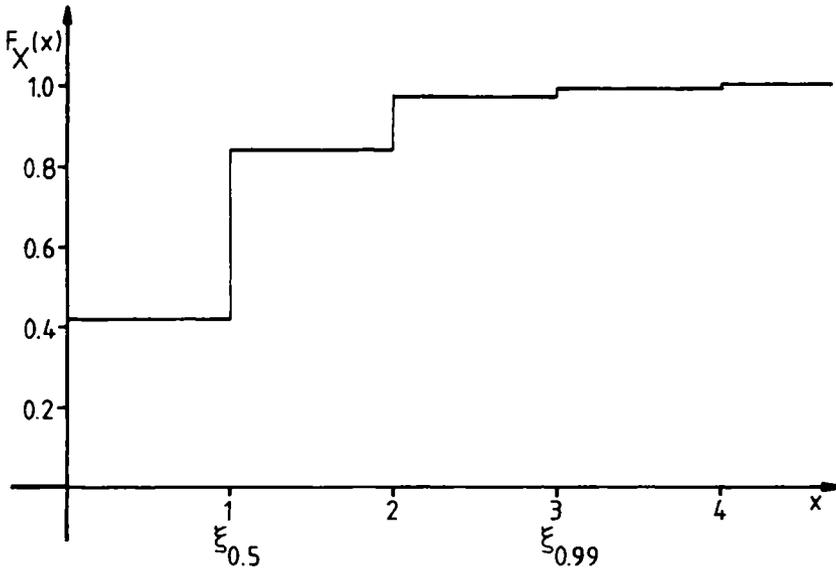


Abb.6: Verteilungsfunktion der $B(4,0.2)$ - Verteilung mit einigen Quantilen

Der *Modalwert* oder *Modus* ist ein weiteres Lagemaß. Bei stetigen Verteilungen ist er der Wert, an dem die Dichte maximal ist, bei diskreten Verteilungen der Wert mit der größten Wahrscheinlichkeit.

Wir wollen nun noch einige Streuungsmaße betrachten, von denen Varianz und Standardabweichung die wichtigsten sind.

Die *Varianz* einer Zufallsvariablen X ist - ihre Existenz vorausgesetzt -

$$\text{Var}(X) = E(X - E(X))^2 \quad (\text{Symbol: } \sigma^2)$$

und die Quadratwurzel aus der Varianz $\sqrt{\text{Var}(X)}$ ist die *Standardabweichung* von X . Die Bedeutung dieser Größen wird durch die *Tschebyscheffsche Ungleichung* deutlich. Für jede positive Zahl c gilt nämlich mit $\sigma^2 = \text{Var}(X)$

$$P(|X - E(X)| \geq c) \leq \frac{\text{Var}(X)}{c^2} = \frac{\sigma^2}{c^2} ;$$

wählt man speziell $c = k\sigma$, so ergibt sich, daß eine Zufallsvariable X höchstens mit Wahrscheinlichkeit $1/k^2$ Werte annehmen kann, die weiter als $k\sigma$ von $E(X)$ entfernt sind; speziell für $k = 3$ (3σ -Regel, s. auch weiter unten) ergibt sich $1/9 = 0.111 = 11.1\%$. Bei der Berechnung der Varianz sind noch der Satz von Steiner (Verschiebungssatz)

$$\text{Var}(X) = E(X^2) - (E(X))^2$$

und die folgende Rechenregel von Nutzen:

$$\text{Var}(\alpha X + \beta) = \alpha^2 \text{Var}(X) .$$

Ist allgemein X eine Zufallsvariable mit Erwartungswert μ und Varianz σ^2 , so gilt für die *standardisierte Zufallsvariable* $Z = (X - \mu)/\sigma$:

$$E(Z) = 0 \quad \text{und} \quad \text{Var}(Z) = 1 .$$

Varianz bzw. Standardabweichung sind Maße für die absolute Größe der Streuung um den Erwartungswert. Dagegen mißt der *Variationskoeffizient* VK (für positive Zufallsvariablen) die Streuung relativ zum Erwartungswert:

$$\text{VK} = \sqrt{\text{Var}(X)}/E(X) .$$

Beispiel:

(a) Die Varianz der Standardnormalverteilung, $Z \sim N(0;1)$, ergibt sich mit Hilfe des Satzes von Steiner wegen $E(Z) = 0$ und

$$E(Z^2) = \int_{-\infty}^{\infty} z^2 \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}z^2} dz = \frac{2}{\sqrt{2\pi}} \int_0^{\infty} z^2 \cdot e^{-\frac{1}{2}z^2} dz = 1$$

zu 1. Bei der $N(\mu; \sigma^2)$ -Verteilung sind μ der Erwartungswert und σ^2 die Varianz der Verteilung und nach der 3σ -Regel für $X \sim N(\mu; \sigma^2)$ ist, vgl. Hartung et al. (1982),

$$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = 0.9973 = 99.73\% .$$

(b) Bei einer $B(n,p)$ -verteilten Zufallsvariablen X ergibt sich zunächst

$$E(X^2) = \sum_{i=0}^n i^2 \binom{n}{i} p^i (1-p)^{n-i} = \sum_{i=0}^n (i(i-1)+i) \binom{n}{i} p^i (1-p)^{n-i}$$

$$\begin{aligned}
&= \sum_{i=0}^n i(i-1) \binom{n}{i} p^i (1-p)^{n-i} + \sum_{i=0}^n i \binom{n}{i} p^i (1-p)^{n-i} \\
&= \sum_{i=2}^n i(i-1) \binom{n}{i} p^i (1-p)^{n-i} + E(X) \\
&= \sum_{i=2}^n i(i-1) \frac{n!}{i!(n-i)!} p^i (1-p)^{n-i} + np \\
&= n(n-1)p^2 \sum_{i=2}^n \frac{(n-2)!}{(i-2)!(n-i)!} p^{i-2} (1-p)^{n-i} + np \\
&= n(n-1)p^2 \cdot \sum_{i=0}^{n-2} \frac{(n-2)!}{i!(n-i-2)!} p^i (1-p)^{n-i-2} + np \\
&= n(n-1)p^2 \cdot \sum_{i=0}^{n-2} \binom{n-2}{i} p^i (1-p)^{n-2-i} + np \\
&= n(n-1)p^2 + np
\end{aligned}$$

und somit nach dem Satz von Steiner

$$\begin{aligned}
\text{Var}(X) &= E(X^2) - (E(X))^2 = n(n-1)p^2 + np - n^2p^2 = n^2p^2 - np^2 + np - n^2p^2 \\
&= np - np^2 = np(1-p) \quad .
\end{aligned}$$

Es sei hier noch erwähnt, daß die Größen $E(X^k)$ *k-te Momente* und die Größen $E(X - E(X))^k$ *k-te zentrale Momente* der Zufallsvariablen X genannt werden, die im folgenden bei Verwendung auch existieren mögen.

Abschließend seien hier noch zwei Größen eingeführt, die den Zusammenhang zwischen 2 Zufallsvariablen X und Y kennzeichnen. Die *Kovarianz* von X und Y ist

$$\text{Cov}(X, Y) = E(X - E(X))(Y - E(Y)) \quad ;$$

sie ändert sich durch eine lineare Transformation in folgender Weise

$$\text{Cov}(\alpha X + \beta, \gamma Y + \delta) = \alpha \cdot \gamma \text{Cov}(X, Y) \quad .$$

Ein Zusammenhangsmaß, das zu keiner derartigen Änderung führt, ist die *Korrelation* von X und Y

$$\rho = \rho_{XY} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)} \cdot \sqrt{\text{Var}(Y)}} \quad .$$

Diese normierte Größe nimmt stets Werte zwischen -1 und $+1$ an.

Ist die Korrelation ρ zwischen X und Y Null, so sagt man auch "die Zufallsvariablen X und Y sind unkorreliert". Das heißt jedoch nicht, daß es keinen

Zusammenhang zwischen X und Y gibt, denn ρ mißt nur die Stärke des linearen Zusammenhangs zwischen zwei Zufallsvariablen. Ist $\rho = 1$ bzw. $\rho = -1$, so besteht zwischen den Zufallsvariablen X und Y ein direkter positiver bzw. negativer linearer Zusammenhang:

$$X = aY + b \quad ,$$

wobei b eine beliebige Zahl und a im Falle $\rho = 1$ ($\rho = -1$) eine beliebige positive (negative) Zahl ist.

Sind X und Y stochastisch unabhängige Zufallsvariablen, so sind sie auch unkorreliert. Die Umkehrung gilt im allgemeinen nicht. Jedoch beinhaltet die Unkorreliertheit normalverteilter Zufallsvariablen auch deren stochastische Unabhängigkeit.

Schließlich sei noch bemerkt, daß für die Varianz der Summe von n beliebigen Zufallsvariablen X_1, \dots, X_n gilt:

$$\text{Var} \left(\sum_{i=1}^n X_i \right) = \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{Cov}(X_i, X_j) \quad .$$

Das heißt natürlich für stochastisch unabhängige Zufallsvariablen X_1, \dots, X_n

$$\text{Var} \left(\sum_{i=1}^n X_i \right) = \sum_{i=1}^n \text{Var}(X_i) \quad (\text{Gleichung von Bienaymé}) \quad .$$

3. PRINZIPIEN DES SCHÄTZENS UND TESTENS; t -, χ^2 -UND F -VERTEILUNG

Wie wir in Abschnitt 2 gesehen haben, lassen sich Zufallsexperimente mit Hilfe von Zufallsvariablen und zugehörigen Wahrscheinlichkeitsverteilungen beschreiben. Dabei ergeben sich die Zufallsvariablen meist in kanonischer Weise aus der Beschreibung des Experiments und der Angabe der interessierenden Ereignisse. Dagegen ist die Bestimmung der Wahrscheinlichkeitsverteilung oft mit Schwierigkeiten verbunden. Läßt sich der Typ einer Verteilung, z.B. X ist binomialverteilt, meist noch recht einfach festlegen, so ist man bei der Bestimmung der Parameter dieser Verteilung auf mehr oder weniger exakte Schätzungen angewiesen, die sich mittels mehrmaliger Durchführung des betreffenden Experiments bestimmen lassen.

Wir wollen nun einige Verfahren zur hier angesprochenen *statistischen Schätzung* von Parametern einer Verteilung behandeln. Aufgrund der Realisa-

tionen x_1, \dots, x_n von n unabhängigen, identisch verteilten Zufallsvariablen X_1, \dots, X_n , sollen dabei die Parameter der Verteilung der X_i möglichst "gut" geschätzt werden. Wir suchen also eine Schätzfunktion $\hat{\theta}_n(x_1, \dots, x_n)$ die der Zufallsstichprobe x_1, \dots, x_n einen Wert zuordnet, der möglichst nahe dem wahren Parameter θ , für den wir uns interessieren, ist.

Eine Schätzfunktion, die wenigstens im Mittel den richtigen Wert liefert, d.h.

$$E(\hat{\theta}_n(X_1, \dots, X_n)) = \theta \quad ,$$

nennt man eine *erwartungstreue* Schätzfunktion.

Beispiel: Sind x_1, \dots, x_n Realisationen von unabhängigen, identisch $N(\mu; \sigma^2)$ -verteilten Zufallsvariablen X_1, \dots, X_n , so sind

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

erwartungstreue Schätzungen für den Erwartungswert μ bzw. die Varianz σ^2 , denn

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} \sum_{i=1}^n \mu = \frac{1}{n} \cdot n \cdot \mu = \mu$$

und

$$\begin{aligned} E(S^2) &= E\left(\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2\right) = E\left(\frac{1}{n-1} \left(\sum_{i=1}^n X_i^2 - n\bar{X}^2\right)\right) \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n E(X_i^2) - nE(\bar{X}^2)\right) \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n [\text{Var}(X_i) + (E(X_i))^2] - n[\text{Var}(\bar{X}) + (E(\bar{X}))^2]\right) \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n [\sigma^2 + \mu^2] - n[\text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) + \mu^2]\right) \\ &= \frac{1}{n-1} \left(n[\sigma^2 + \mu^2] - n\mu^2 - \frac{n}{n-1} \sum_{i=1}^n \text{Var}(X_i)\right) \\ &= \frac{1}{n-1} \left(n\sigma^2 + \mu^2 - \frac{n}{n-1} \sum_{i=1}^n \sigma^2\right) \\ &= \frac{1}{n-1} \left(n\sigma^2 - \frac{1}{n-1} \cdot n\sigma^2\right) = \frac{1}{n-1} (n-1)\sigma^2 \\ &= \sigma^2 \quad . \end{aligned}$$

Eine andere wünschenswerte Eigenschaft einer Schätzfunktion ist, daß mit wachsendem Stichprobenumfang n die Schätzung genauer wird:

$$P(|\hat{\theta}_n(X_1, \dots, X_n) - \theta| > \epsilon) \xrightarrow{n \rightarrow \infty} 0 \quad \text{für jedes } \epsilon > 0$$

Eine solche Schätzfunktion, die *asymptotisch* richtige Werte liefert, heißt *konsistent*.

Ein Kriterium für die Güte eines Schätzers $\hat{\theta}$ für θ ist der sogenannte *mittlere quadratische Fehler* (*mean squared error*, MSE):

$$\text{MSE}(\theta, \hat{\theta}) = E(\hat{\theta} - \theta)^2 = \text{Var}(\hat{\theta}) + (E(\hat{\theta}) - \theta)^2 .$$

Die Größe $E(\hat{\theta}) - \theta$ nennt man auch *Bias* oder *Verzerrung* des Schätzers $\hat{\theta}$. Bei einem erwartungstreuen Schätzer $\hat{\theta}$ für θ ist der MSE natürlich gerade die Varianz des Schätzers. Bei konkreten Schätzungen ist man dann natürlich bemüht, den MSE möglichst klein zu halten.

Kommen wir nun zu den Schätzverfahren. Eine einfache Vorgehensweise ist die sogenannte *Momentenmethode*. Dabei wird der zu schätzende Parameter θ durch die Momente der Verteilung ausgedrückt. Diese werden dann durch die entsprechenden empirischen Momente geschätzt: Das *k-te empirische Moment* der Stichprobe x_1, \dots, x_n ist gegeben durch

$$m_k = \frac{1}{n} \sum_{i=1}^n x_i^k$$

Beispiel: Ein Schätzer für die Varianz σ^2 der $N(\mu; \sigma^2)$ -Verteilung nach der Momentenmethode ist

$$m_2 - m_1^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \left(\frac{1}{n} \sum_{i=1}^n x_i \right)^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 . \quad \square$$

Sind x_1, \dots, x_n Realisationen von n unabhängig identisch verteilten Zufallsvariablen X_1, \dots, X_n , so ist der Schätzer für den Parameter θ der Verteilung nach der *Maximum-Likelihood-Methode* gerade derjenige Wert $\hat{\theta}$, der die *Likelihood*

$$L(x_1, \dots, x_n) = \begin{cases} P(X_1 = x_1) \cdot \dots \cdot P(X_n = x_n) & , \text{ bei diskreten Verteilungen} \\ f_{X_1}(x_1) \cdot \dots \cdot f_{X_n}(x_n) & , \text{ bei stetigen Verteilungen} \end{cases}$$

in Abhängigkeit von θ maximiert.

Beispiel: Seien X_1, \dots, X_n unabhängig identisch $N(\mu; 1)$ -verteilt. Dann ist

$$L(x_1, \dots, x_n) = f_{X_1}(x_1) \cdot \dots \cdot f_{X_n}(x_n) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}(x_i - \mu)^2}.$$

Diese Likelihood wird maximal, wenn es die zugehörige *log*-Likelihood wird:

$$\begin{aligned} \ln L(x_1, \dots, x_n) &= \ln \left(\prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x_i - \mu)^2} \right) \\ &= \sum_{i=1}^n \ln \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x_i - \mu)^2} \right) \\ &= \sum_{i=1}^n \left(\ln \frac{1}{\sqrt{2\pi}} - \frac{1}{2}(x_i - \mu)^2 \right), \end{aligned}$$

d.h. wenn $-\sum_{i=1}^n (x_i - \mu)^2$ maximal bzw. $\sum_{i=1}^n (x_i - \mu)^2$ minimal in Abhängigkeit von μ wird. Wegen der Minimumeigenschaft des arithmetischen Mittels \bar{x} , vgl. Abschnitt 1, ergibt sich \bar{x} als Maximum-Likelihood-Schätzer für den Parameter μ . ┘

Allgemein ist die Likelihood als die Wahrscheinlichkeit für das Vorliegen der Stichprobe bzw. im stetigen Fall als der Wert der gemeinsamen Dichte für die vorliegende Stichprobe definiert, vgl. Abschnitt 5.

Als letztes Schätzverfahren wollen wir hier die *Methode der kleinsten Quadrate* vorstellen. Dabei geht man davon aus, daß der Erwartungswert der zu beobachtenden unabhängig identisch verteilten Zufallsvariablen X_1, \dots, X_n in bekannter Weise vom interessierenden Parameter θ abhängt:

$$E(X_i) = g(\theta),$$

und bestimmt die Schätzung $\hat{\theta} = \hat{\theta}(x_1, \dots, x_n)$ für θ so, daß gilt:

$$\sum_{i=1}^n (x_i - g(\hat{\theta}))^2 \leq \sum_{i=1}^n (x_i - g(\theta))^2$$

für alle möglichen Parameterwerte θ . Dieses Verfahren läßt sich auch auf beliebige Zufallsvariablen X_1, \dots, X_n übertragen, vgl. Hartung et al. (1982, Kap.III).

Beispiel: Aufgrund der Minimumeigenschaft des arithmetischen Mittels \bar{x} , vgl. Abschnitt 1, ist \bar{x} ein Schätzer für den Parameter μ einer $N(\mu; \sigma^2)$ -Verteilung nach der Methode der kleinsten Quadrate.

Sind x_1, \dots, x_n Realisationen aus einer $N(\mu; \sigma^2)$ -Verteilung, so werden die Parameter μ und σ^2 in der Regel vermittels des arithmetischen Mittels

$$\bar{x} = \frac{1}{n}(x_1 + \dots + x_n)$$

bzw. der empirischen Varianz (für $n \geq 2$)

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

geschätzt. (Diese Schätzer für erste und zweite Momente einer Zufallsvariablen werden zumeist auch dann benutzt, wenn nichts über den Verteilungstyp bekannt ist.) Wie wir gesehen haben ist

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

ein Schätzer für σ^2 nach der Momentenmethode; der gleiche Schätzer ergibt sich auch nach der Maximum-Likelihood-Methode. Weitere Schätzer für σ^2 findet man in Hartung et al. (1982, Kap. IV, Abschnitt 1.3).

Sind x_1, \dots, x_n und y_1, \dots, y_n Realisationen aus Normalverteilungen $N(\mu_X; \sigma_X^2)$ bzw. $N(\mu_Y; \sigma_Y^2)$, so wird deren Kovarianz σ_{XY} (für $n \geq 2$) durch

$$s_{XY} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

geschätzt, und deren Korrelation $\rho = \rho_{XY}$ durch

$$r_{XY} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}}$$

geschätzt.

Fehlen in Fällen anderer stetiger Verteilungen weitere Informationen, so werden diese Schätzer häufig ebenfalls verwandt, vgl. jedoch für Korrelationschätzer auch Abschnitt 2 in Kap. III.

Anstatt den Parameter θ möglichst "gut" durch einen einzigen Wert $\hat{\theta}_n = \hat{\theta}_n(x_1, \dots, x_n)$ zu schätzen, möchte man oft einen "kleinen" Bereich angeben, in dem θ liegt. Da man Informationen über θ aus einem Zufallsexperiment gewinnt, ist dies i.a. nicht möglich. Wohl aber gibt es Methoden, die nur mit geringer Wahrscheinlichkeit Bereiche liefern, die den Parame-

ter θ nicht enthalten. Ist diese *Irrtumswahrscheinlichkeit* höchstens α , so ergibt sich ein Bereich, der θ mit Wahrscheinlichkeit $1-\alpha$ enthält; ein solcher Bereich heißt auch *Vertrauens- oder Konfidenzbereich zum Niveau $1-\alpha$* . Handelt es sich hierbei um ein Intervall auf der Zahlengerade, so spricht man auch von einem *Konfidenzintervall zum Niveau $1-\alpha$* oder kurz *$(1-\alpha)$ -Konfidenzintervall*.

Bei der Konstruktion von Konfidenzintervallen geht man häufig so vor, daß man sich zunächst einen Punktschätzer $\hat{\theta}_n$ für den interessierenden Parameter θ beschafft und dessen Erwartungswert $\mu(\theta)$ und Varianz $\sigma^2(\theta)$, die meist vom Parameter θ abhängen, bestimmt. Dann bildet man die standardisierte Zufallsvariable

$$\frac{\hat{\theta}_n - \mu(\theta)}{\sigma(\theta)} .$$

Hängt deren Verteilung nicht mehr von θ ab, so bestimmt man Zahlen u_1 und u_2 derart, daß gilt

$$P \left(u_1 \leq \frac{\hat{\theta}_n - \mu(\theta)}{\sigma(\theta)} \leq u_2 \right) \geq 1 - \alpha \quad \text{für alle } \theta .$$

Dabei sollten u_1 und u_2 so gewählt werden, daß die Wahrscheinlichkeit möglichst nahe bei $1-\alpha$ liegt. Diejenigen Werte θ , für die gilt

$$u_1 \leq \frac{\hat{\theta}_n - \mu(\theta)}{\sigma(\theta)} \leq u_2 ,$$

bilden dann das $(1-\alpha)$ -Konfidenzintervall für θ . Ist die Verteilung der standardisierten Zufallsvariablen abhängig von θ , so kann man - zunächst für große Stichprobenumfänge n - die Verteilung der Größe

$$\sqrt{n} \cdot \frac{\hat{\theta}_n - \mu(\theta)}{\sigma} \quad \text{mit} \quad \sigma(\theta) = \frac{\sigma}{\sqrt{n}}$$

im Falle

$$P \left(\sqrt{n} \cdot \frac{\hat{\theta}_n - \mu(\theta)}{\sigma} \leq z \right) \xrightarrow{n \rightarrow \infty} \Phi(z)$$

durch die Standardnormalverteilung approximieren und erhält mit

$$\left[\hat{\theta}_n - \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2} ; \hat{\theta}_n + \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2} \right]$$

ein *asymptotisches Konfidenzintervall zum Niveau $1-\alpha$* . Hier konvergiert mit wachsendem n die Wahrscheinlichkeit, daß man ein Intervall erhält, in dem der Parameter θ liegt, gegen $1-\alpha$.

Beispiel: Sind x_1, \dots, x_n Realisationen von n unabhängigen $N(\mu; \sigma^2)$ -verteilten Zufallsvariablen mit bekannter Varianz σ^2 , so hat die Schätzfunktion $\bar{X} = \frac{1}{n}(X_1 + \dots + X_n)$ den Erwartungswert μ und die Varianz $\frac{1}{n}\sigma^2$. Damit ist

$$\sqrt{n} \cdot \frac{\bar{X} - \mu}{\sigma}$$

$N(0;1)$ -verteilt und es ist

$$P\left(-u_{1-\alpha/2} \leq \sqrt{n} \frac{\bar{X} - \mu}{\sigma} \leq u_{1-\alpha/2}\right) = 1 - \alpha \quad ,$$

wobei u_α das α -Quantil der $N(0;1)$ -Verteilung bezeichnet. Das Konfidenzintervall zum Niveau $1-\alpha$ für den Parameter θ ergibt sich somit zu

$$\left[\bar{x} - \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2} ; \bar{x} + \frac{\sigma}{\sqrt{n}} u_{1-\alpha/2}\right] \quad .$$

Bevor wir einige weitere Konfidenzintervalle für die Parameter einer $N(\mu; \sigma^2)$ -Verteilung angeben, wollen wir noch einige sogenannte *Prüfverteilungen* einführen.

Sind X_1, \dots, X_n unabhängige $N(0;1)$ -verteilte Zufallsvariablen, so heißt die Verteilung von $\sum_{i=1}^n X_i^2$ eine (zentrale) χ^2 -Verteilung mit n Freiheitsgraden, in Zeichen

$$\sum_{i=1}^n X_i^2 \sim \chi_n^2 \quad .$$

Die Quantile $\chi_{n;\alpha}^2$ der χ^2 -Verteilung sind für einige n und α im Anhang vertafelt.

Sind X_0, X_1, \dots, X_n unabhängige $N(0;1)$ -verteilte Zufallsvariablen, so ist die Zufallsvariable

$$t = \frac{X_0}{\sqrt{\sum_{i=1}^n X_i^2}}$$

(zentral) t -verteilt mit n Freiheitsgraden, kurz

$$t \sim t_n \quad .$$

Die Quantile $t_{n;\alpha}$ der t -Verteilung sind ebenfalls für einige n und α im Anhang vertafelt.

Schließlich heißt eine Zufallsvariable

$$F = \left(\frac{1}{m} \sum_{i=1}^m X_i^2 \right) / \left(\frac{1}{n} \sum_{i=1}^n Y_i^2 \right)$$

(zentral) F -verteilt mit m und n Freiheitsgraden, kurz

$$F \sim F_{m,n}$$

falls $X_1, \dots, X_m, Y_1, \dots, Y_n$ unabhängig standardnormalverteilte Zufallsvariablen sind. Auch die Quantile $F_{m,n;\alpha}$ der F -Verteilung sind für einige Kombinationen von m , n und α im Anhang vertafelt.

Nun noch einige weitere Konfidenzintervalle für die Parameter μ und σ^2 einer $N(\mu; \sigma^2)$ -Verteilung. Bei unbekannter Varianz σ^2 ist

$$\left[\bar{x} - t_{n-1; 1-\alpha/2} \cdot \frac{s}{\sqrt{n}}; \bar{x} + t_{n-1; 1-\alpha/2} \cdot \frac{s}{\sqrt{n}} \right]$$

ein $(1-\alpha)$ -Konfidenzintervall für den Erwartungswert μ . Ist der Erwartungswert μ einer $N(\mu; \sigma^2)$ -Verteilung bekannt, so ist

$$\left[\frac{\sum_{i=1}^n (x_i - \mu)^2}{\chi_{n; 1-\alpha/2}^2}; \frac{\sum_{i=1}^n (x_i - \mu)^2}{\chi_{n; \alpha/2}^2} \right]$$

ein $(1-\alpha)$ -Konfidenzintervall für die Varianz σ^2 . Ein solches Intervall ist bei unbekanntem μ gegeben durch

$$\left[\frac{(n-1)s^2}{\chi_{n-1; 1-\alpha/2}^2}; \frac{(n-1)s^2}{\chi_{n-1; \alpha/2}^2} \right]$$

Abschließend wollen wir uns nun mit *statistischen Tests* beschäftigen. Das sind Verfahren, die die Richtigkeit von Aussagen über statistische Verteilungen überprüfen: Ist ein Parameter größer oder kleiner als ein bestimmter Wert? Liegt er in einem bestimmten Intervall? Ist eine Zufallsvariable normalverteilt? Mit Fragestellungen wie den beiden ersten wollen wir uns hier beschäftigen (*Parameter-Tests*). Die dritte Fragestellung kann mit Hilfe sogenannter *Anpassungstests* beantwortet werden, man vgl. hierzu Hartung et al. (1982, Kap.IV).

Solche *Testverfahren* oder *Prüfverfahren* können, da sie auf der Grundlage zufälliger Experimente basieren, natürlich nicht immer richtige Entscheidungen liefern. Prüft man z.B., ob der Parameter μ der $N(\mu; \sigma^2)$ -Verteilung größer oder kleiner als ein vorgegebener Wert μ_0 ist, so könnte etwa $\mu \leq \mu_0$

die sogenannte (Null)hypothese H_0 und $\mu > \mu_0$ die Alternativhypothese H_1 sein. Man beobachtet nun eine Stichprobe x_1, \dots, x_n aus dieser Verteilung und will sich aufgrund dessen für H_0 oder H_1 entscheiden. Dabei kann man zwei Fehler machen: Die Nullhypothese trifft zu, und man entscheidet sich fälschlicherweise für die Alternativhypothese, oder die Alternativhypothese trifft zu, und man entscheidet sich für die Nullhypothese. Im ersten Fall spricht man vom Fehler 1. Art, im zweiten vom Fehler 2. Art. Bei einem statistischen Testverfahren geht man nun so vor, daß man den Fehler 1. Art beschränkt, er soll höchstens mit einer Wahrscheinlichkeit α auftreten (Irrtumswahrscheinlichkeit, Signifikanzniveau). Einen solchen Test nennt man Test zum Niveau α oder auch Niveau- α -Test, egal mit welcher Wahrscheinlichkeit β der Fehler 2. Art auftritt. Natürlich wird man bei der Konstruktion von Tests bemüht sein, diesen möglichst gering zu halten.

Bei der Bestimmung eines solchen Tests geht man wie folgt vor. Nehmen wir an, daß getestet werden soll

$$H_0: \theta \leq \theta_0 \quad \text{gegen} \quad H_1: \theta > \theta_0 \quad ,$$

d.h. man will wissen, ob der Parameter θ einer Verteilung kleiner oder größer als ein vorgegebener Wert θ_0 ist. Man ordnet dann mit Hilfe einer Teststatistik oder Prüfgröße T der Stichprobe x_1, \dots, x_n eine Zahl $T(x_1, \dots, x_n)$ zu, bei der z.B. ein großer Wert eher für das Vorliegen von H_1 , ein kleiner Wert eher für das Vorliegen von H_0 spricht. Dann sucht man eine Zahl $c_{1-\alpha}$ derart, daß gilt

$$P_{\theta}(T > c_{1-\alpha}) \leq \alpha \quad \text{für alle } \theta \in H_0 \quad ;$$

der Index θ soll hierbei andeuten, daß die Wahrscheinlichkeit bei Vorliegen von θ gemeint ist. Entscheidet man sich nun für H_1 , falls $T > c_{1-\alpha}$, und für H_0 , falls $T \leq c_{1-\alpha}$, so handelt es sich bei dieser Entscheidungsregel um einen Test zum Niveau α . Die Größe $c_{1-\alpha}$ heißt kritischer Wert, der Bereich $T \leq c_{1-\alpha}$ Annahmereich und der Bereich $T > c_{1-\alpha}$ Ablehnungsbereich des Tests. Entscheidet man sich nun konkret für H_0 , so sagt man " H_0 kann nicht verworfen werden" oder " H_0 wird angenommen", entscheidet man sich für H_1 , so sagt man " H_1 ist signifikant (zum Niveau α)" oder auch " H_1 ist zum Niveau α statistisch gesichert".

Ein Hypothesenpaar der obigen Form

$$H_0: \theta \leq \theta_0 \quad \text{gegen} \quad H_1: \theta > \theta_0$$

nennt man auch einseitige Hypothese, den entsprechenden Test auch einseitigen Test. Von einer zweiseitigen Hypothese bzw. einem zweiseitigen Test

spricht man bei Hypothesenpaaren der Gestalt

$$H_0: \theta = \theta_0 \quad \text{gegen} \quad H_1: \theta \neq \theta_0 \quad .$$

Beispiel: Eine Maschine produziert Schrauben, deren tatsächliche Länge als $N(\mu; \sigma^2)$ -verteilt angenommen werden kann. Die Standardabweichung σ der Maschine möge vom Hersteller mit 0.3 mm angegeben sein. Aufgrund der in Tab.3 angegebenen Längen x_1, \dots, x_{14} einer Stichprobe vom Umfang $n = 14$ aus der Produktion möchte man zum Niveau $\alpha = 0.05$ testen, ob die mittlere Schraubenlänge μ der Produktion kleiner oder größer als $\mu_0 = 30$ mm ist.

Tab.3: Längen x_i von $n = 14$ Schrauben in mm

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14
x_i	30.6	30.9	29.8	30.1	29.5	29.7	30.6	30.5	29.4	29.9	30.0	30.2	29.7	30.1

Das Testproblem hier ist

$$H_0: \mu \leq 30 \quad \text{gegen} \quad H_1: \mu > 30 \quad .$$

Da ein kleiner Mittelwert \bar{x} eher für H_0 , ein großer eher für H_1 spricht, verwenden wir die im Falle $\mu = \mu_0$ standardnormalverteilte Größe

$$T = \sqrt{n} \frac{\bar{X} - \mu_0}{\sigma}$$

als Teststatistik. Falls $\mu = \mu_0$ ist, gilt also

$$P_{\mu} (T > u_{1-\alpha}) = \alpha \quad ,$$

und falls $\mu < \mu_0$ ist, gilt

$$P_{\mu} (T > u_{1-\alpha}) < \alpha \quad .$$

Entscheidet man sich nun im Falle $T(x_1, \dots, x_n) \leq u_{1-\alpha}$ für H_0 und im Falle $T(x_1, \dots, x_n) > u_{1-\alpha}$ für H_1 , so hat man einen Niveau- α -Test vorliegen. Dieser Test heißt auch (einseitiger) *Einstichprobengaußtest*.

In unserem Beispiel ergibt sich aus Tab.3 $\bar{x} = 30.071$, und es ist

$u_{1-\alpha} = u_{1-0.05} = u_{0.95} = 1.6449$, so daß gilt:

$$T(x_1, \dots, x_n) = \sqrt{14} \frac{30.071 - 30}{0.3} = 0.886 < 1.6449 = u_{0.95} \quad .$$

Wir können die Hypothese H_0 also nicht verwerfen, d.h. es kann nicht zum 5% Niveau gesichert werden, daß μ größer als 30 ist, obwohl die mittlere

Schraubenlänge in der Stichprobe dies ist.

An dieser Stelle sei noch etwas zur *Notation von Zufallsvariablen* in den weiteren Kapiteln bemerkt. Wir werden aus Notationsrunden in der Bezeichnungsweise nicht immer streng zwischen Zufallsvariablen und ihren Realisationen unterscheiden. Ist von Wahrscheinlichkeiten, Verteilungen, Erwartungswerten etc. die Rede, so ist klar, daß sich dies auf die den Realisationen zugrundeliegenden Zufallsvariablen bezieht.

4. VEKTOR- UND MATRIZENRECHNUNG

In der multivariaten Statistik ist die Verwendung von Vektoren und Matrizen unumgänglich. Daher werden in diesem Abschnitt die Begriffe Vektor und Matrix eingeführt und wichtige Rechenregeln sowie Funktionen von Matrizen betrachtet.

Ein *n-dimensionalen Vektor* a ist ein als Spalte geschriebenes n -Tupel von Zahlen

$$a = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} .$$

Der zugehörige *transponierte Vektor* $a^T = (a_1, a_2, \dots, a_n)$ ist der zum Spaltenvektor a gehörige Zeilenvektor. Für zwei n -dimensionale Vektoren a und b ist eine *Addition*

$$a + b = \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \\ \vdots \\ a_n + b_n \end{pmatrix}$$

und eine *Multiplikation*

$$a^T \cdot b = a_1 b_1 + a_2 b_2 + \dots + a_n b_n$$

erklärt. Weiterhin läßt sich ein Vektor a mit einer Zahl z multiplizieren (*skalare Multiplikation*):

$$z \cdot a = \begin{pmatrix} za_1 \\ za_2 \\ \vdots \\ za_n \end{pmatrix} .$$

Von einem n -dimensionalen Vektor a der Länge q , spricht man, wenn gilt

$$a^T a = q^2 .$$

Möchte man einen beliebigen n -dimensionalen Vektor $b = (b_1, \dots, b_n)^T$ so normieren, daß der normierte Vektor eine vorgegebene Länge q besitzt, so kann man wie folgt vorgehen. Man berechnet für $i=1, \dots, n$

$$\tilde{b}_i = b_i / \sqrt{\sum_{i=1}^n b_i^2}$$

(dann hat der Vektor $\tilde{b} = (\tilde{b}_1, \dots, \tilde{b}_n)^T$ die Länge 1) und multipliziert die Größen \tilde{b}_i mit q :

$$a_i = q \cdot \tilde{b}_i \quad \text{für } i=1, \dots, n .$$

Mit $a = (a_1, \dots, a_n)^T$ ist dann der normierte Vektor der Länge q gefunden.

Multipliziert man zwei n -dimensionale Vektoren a und b in der Form ab^T , so ergibt sich ein quadratisches Zahlenschema:

$$a \cdot b^T = \begin{pmatrix} a_1 b_1 & a_1 b_2 & \dots & a_1 b_n \\ a_2 b_1 & a_2 b_2 & \dots & a_2 b_n \\ \vdots & \vdots & & \vdots \\ a_n b_1 & a_n b_2 & \dots & a_n b_n \end{pmatrix} .$$

Ein solches Zahlenschema nennen wir eine $n \times n$ -Matrix. Bevor wir auf allgemeine Matrizen eingehen zunächst ein Beispiel.

Beispiel: Es seien a und b 4-dimensionale Vektoren, nämlich $a^T = (4, 3, 7, 10)$ und $b^T = (3, 6, 1, 0)$, und es sei $z = 5$. Dann ist

$$a + b = \begin{pmatrix} 4 + 3 \\ 3 + 6 \\ 7 + 1 \\ 10 + 0 \end{pmatrix} = \begin{pmatrix} 7 \\ 9 \\ 8 \\ 10 \end{pmatrix} ,$$

$$a^T b = 4 \cdot 3 + 3 \cdot 6 + 7 \cdot 1 + 10 \cdot 0 = 37 ,$$

$$z \cdot a = 5a = \begin{pmatrix} 20 \\ 15 \\ 35 \\ 50 \end{pmatrix} \quad \text{und}$$

$$ab^T = \begin{pmatrix} 4 \cdot 3 & 4 \cdot 6 & 4 \cdot 1 & 4 \cdot 0 \\ 3 \cdot 3 & 3 \cdot 6 & 3 \cdot 1 & 3 \cdot 0 \\ 7 \cdot 3 & 7 \cdot 6 & 7 \cdot 1 & 7 \cdot 0 \\ 10 \cdot 3 & 10 \cdot 6 & 10 \cdot 1 & 10 \cdot 0 \end{pmatrix} = \begin{pmatrix} 12 & 24 & 4 & 0 \\ 9 & 18 & 3 & 0 \\ 21 & 42 & 7 & 0 \\ 30 & 60 & 10 & 0 \end{pmatrix} .$$

Allgemein ist eine $n \times m$ -Matrix A ein Zahlenschema bestehend aus n Zeilen und m Spalten:

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & & a_{nm} \end{pmatrix} .$$

Die *Transponierte* A^T von A entsteht durch Spiegelung von A und ist eine $m \times n$ -Matrix:

$$A^T = \begin{pmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \vdots & \vdots & & \vdots \\ a_{1m} & a_{2m} & & a_{nm} \end{pmatrix} .$$

Eine $n \times m$ -Matrix A kann mit einer Zahl z elementweise multipliziert werden (*skalare Multiplikation*)

$$z \cdot A = \begin{pmatrix} za_{11} & \dots & za_{1m} \\ \vdots & & \vdots \\ za_{n1} & \dots & za_{nm} \end{pmatrix} ,$$

und für 2 $n \times m$ -Matrizen A und B wird in folgender Weise eine *Addition* erklärt:

$$A + B = \begin{pmatrix} a_{11} & \dots & a_{1m} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nm} \end{pmatrix} + \begin{pmatrix} b_{11} & \dots & b_{1m} \\ \vdots & & \vdots \\ b_{n1} & \dots & b_{nm} \end{pmatrix} = \begin{pmatrix} a_{11} + b_{11} & \dots & a_{1m} + b_{1m} \\ \vdots & & \vdots \\ a_{n1} + b_{n1} & \dots & a_{nm} + b_{nm} \end{pmatrix} .$$

Weiterhin gilt natürlich für 2 $n \times m$ -Matrizen A und B und eine Zahl z

$$(A + B)^T = A^T + B^T \quad , \quad (zA)^T = zA^T .$$

Falls gilt $n = m$, so heißt eine $n \times m = n \times n$ -Matrix A *quadratisch* und falls zusätzlich $A = A^T$ gilt, so heißt A *symmetrisch*.

Für einige symmetrische Matrizen gibt es spezielle Bezeichnungen. So ist I_n die $n \times n$ -dimensionale *Einheitsmatrix*

$$I_n = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}$$

und J_n die $n \times n$ -dimensionale Einsermatrix

$$J_n = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \dots & 1 \end{pmatrix} .$$

Beispiel: Es seien

$$A = \begin{pmatrix} 2 & 4 & 7 \\ 3 & 0 & 1 \end{pmatrix} \quad \text{und} \quad B = \begin{pmatrix} 1 & 6 & 9 \\ 8 & 3 & 2 \end{pmatrix}$$

2×3 -Matrizen und es sei $z = 4$. Dann ist z.B.

$$A^T = \begin{pmatrix} 2 & 3 \\ 4 & 0 \\ 7 & 1 \end{pmatrix}$$

eine 3×2 -Matrix,

$$zA = 4 \begin{pmatrix} 2 & 4 & 7 \\ 3 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 8 & 16 & 28 \\ 12 & 0 & 4 \end{pmatrix}$$

und

$$A + B = \begin{pmatrix} 2 & 4 & 7 \\ 3 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 6 & 9 \\ 8 & 3 & 2 \end{pmatrix} = \begin{pmatrix} 3 & 10 & 16 \\ 11 & 3 & 3 \end{pmatrix} .$$

Kommen wir nun noch zur *Multiplikation von Matrizen*. Soll das Produkt $A \cdot B$ zweier Matrizen A und B bestimmt werden, so muß folgende Dimensionsbedingung erfüllt sein: Die Anzahl der Spalten von A muß mit der Anzahl der Zeilen von B übereinstimmen. Ist also z.B. A eine $n \times m$ -Matrix, so muß B eine $m \times k$ -Matrix sein, und es ist dann

$$A \cdot B = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{pmatrix} \cdot \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1k} \\ \vdots & \vdots & & \vdots \\ b_{m1} & b_{m2} & \dots & b_{mk} \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{i=1}^m a_{1i} b_{i1} & \sum_{i=1}^m a_{1i} b_{i2} & \dots & \sum_{i=1}^m a_{1i} b_{ik} \\ \vdots & \vdots & & \vdots \\ \sum_{i=1}^m a_{ni} b_{i1} & \sum_{i=1}^m a_{ni} b_{i2} & \dots & \sum_{i=1}^m a_{ni} b_{ik} \end{pmatrix}$$

eine $n \times k$ -Matrix.

Beispiel: Es seien

$$A = \begin{pmatrix} 3 & 7 & 10 \\ 4 & 1 & 6 \\ 5 & 2 & 3 \\ 0 & 6 & 9 \end{pmatrix} \quad \text{und} \quad B = \begin{pmatrix} 4 & 3 \\ 5 & 7 \\ 1 & 6 \end{pmatrix}$$

4x3- bzw. 3x2-Matrizen. Dann ist

$$A \cdot B = \begin{pmatrix} 3 \cdot 4 + 7 \cdot 5 + 10 \cdot 1 & 3 \cdot 3 + 7 \cdot 7 + 10 \cdot 6 \\ 4 \cdot 4 + 1 \cdot 5 + 6 \cdot 1 & 4 \cdot 3 + 1 \cdot 7 + 6 \cdot 6 \\ 5 \cdot 4 + 2 \cdot 5 + 3 \cdot 1 & 5 \cdot 3 + 2 \cdot 7 + 3 \cdot 6 \\ 0 \cdot 4 + 6 \cdot 5 + 9 \cdot 1 & 0 \cdot 3 + 6 \cdot 7 + 9 \cdot 6 \end{pmatrix} = \begin{pmatrix} 57 & 118 \\ 27 & 55 \\ 33 & 47 \\ 39 & 96 \end{pmatrix}$$

eine 4x2-Matrix. ┌

Im Zusammenhang mit der Multiplikation von Matrizen gelten die folgenden Rechenregeln:

$$\begin{aligned} (A+B) \cdot C &= A \cdot C + B \cdot C, & \text{falls } A, B, C \text{ } n \times m\text{-}, n \times m\text{- bzw. } m \times k\text{- Matrizen,} \\ A \cdot (B+C) &= A \cdot B + A \cdot C, & \text{falls } A, B, C \text{ } n \times m\text{-}, m \times k\text{- bzw. } m \times k\text{- Matrizen,} \\ (A \cdot B) \cdot C &= A \cdot (B \cdot C), & \text{falls } A, B, C \text{ } n \times m\text{-}, m \times k\text{- bzw. } k \times q\text{- Matrizen,} \\ (A \cdot B)^T &= B^T \cdot A^T, & \text{falls } A, B \text{ } n \times m\text{- bzw. } m \times k\text{- Matrizen sind.} \end{aligned}$$

Ein anderes häufig verwandtes Matrizenprodukt ist das sogenannte *Kroneckerprodukt* " \otimes ". Ist A eine $n \times m$ -Matrix und B eine $k \times q$ -Matrix, so ist das Kroneckerprodukt von A und B eine $nk \times mq$ -Matrix, nämlich

$$A \otimes B = \begin{pmatrix} a_{11} & \dots & a_{1m} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nm} \end{pmatrix} \cdot B = \begin{pmatrix} a_{11}B & \dots & a_{1m}B \\ \vdots & & \vdots \\ a_{n1}B & \dots & a_{nm}B \end{pmatrix}.$$

Beispiel: Das Kroneckerprodukt der Matrizen

$$A = \begin{pmatrix} 3 & 2 \\ 7 & 4 \end{pmatrix} \quad \text{und} \quad B = \begin{pmatrix} 3 & 7 & 6 & 5 \\ 2 & 9 & 8 & 1 \\ 3 & 0 & 4 & 4 \end{pmatrix}$$

ist gegeben als

$$A \otimes B = \begin{pmatrix} 3B & 2B \\ 7B & 4B \end{pmatrix} = \begin{pmatrix} 9 & 21 & 18 & 15 & | & 6 & 14 & 12 & 10 \\ 6 & 27 & 24 & 3 & | & 4 & 18 & 16 & 2 \\ 9 & 0 & 12 & 12 & | & 9 & 0 & 8 & 8 \\ \hline 21 & 49 & 42 & 35 & | & 12 & 28 & 24 & 20 \\ 14 & 63 & 56 & 7 & | & 8 & 36 & 32 & 4 \\ 21 & 0 & 28 & 28 & | & 12 & 0 & 16 & 16 \end{pmatrix}.$$
┌

Für das Kroneckerprodukt gelten die folgenden Rechenregeln:

$$(A+B) \otimes C = (A \otimes C) + (B \otimes C), \text{ falls } A, B, C \text{ } n \times m\text{-}, n \times m\text{- bzw. } k \times q\text{- Matrizen,}$$

$$\begin{aligned}
 A \otimes (B + C) &= (A \otimes B) + (A \otimes C), \text{ falls } A, B, C \text{ } n \times m \text{-}, k \times q \text{- bzw. } k \times q \text{- Matrizen,} \\
 z_1 A \otimes z_2 B &= z_1 z_2 A \otimes B \quad , \text{ falls } A, B \text{ } n \times m \text{- bzw. } k \times q \text{- Matrizen und } z_1, \\
 &\quad z_2 \text{ Zahlen,} \\
 A \cdot B \otimes C \cdot D &= (A \otimes C) \cdot (B \otimes D) \quad , \text{ falls } A, B, C, D \text{ } n \times m \text{-}, m \times k \text{-}, p \times q \text{- bzw. } q \times \ell \text{-} \\
 &\quad \text{Matrizen,} \\
 (A \otimes B)^T &= A^T \otimes B^T \quad , \text{ falls } A, B \text{ } n \times m \text{- bzw. } k \times q \text{- Matrizen sind.}
 \end{aligned}$$

Rechentechnisch ist es oft günstiger mit Vektoren als mit Matrizen zu arbeiten. Daher führen wir hier den Operator "vec" ein, der einer beliebigen $n \times m$ -Matrix A einen nm -dimensionalen Spaltenvektor zuordnet:

$$\text{vec } A = \text{vec} \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{pmatrix} = \begin{pmatrix} a_{11} \\ a_{12} \\ \vdots \\ a_{1m} \\ a_{21} \\ \vdots \\ a_{n1} \\ \vdots \\ a_{nm} \end{pmatrix} ;$$

vec A entsteht also aus A , indem man die Zeilen von A transponiert und untereinander schreibt.

Beispiel: Es sei

$$A = \begin{pmatrix} 3 & 4 & 1 & 2 \\ 7 & 6 & 5 & 3 \\ 2 & 4 & 1 & 0 \end{pmatrix} .$$

Dann ist

$$\text{vec } A = (3, 4, 1, 2, 7, 6, 5, 3, 2, 4, 1, 0)^T$$

Wir werden uns nun mit einigen Funktionen quadratischer Matrizen beschäftigen. Hier sei zunächst die Spur (trace) einer $n \times n$ -Matrix

$$A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix} \quad \leftarrow \text{Hauptdiagonale}$$

erwähnt. Die Spur $\text{tr} A$ von A ist die Summe der Elemente von A , die auf der Hauptdiagonale liegen, also

$$\text{tr } A = a_{11} + a_{22} + \dots + a_{nn} = \sum_{i=1}^n a_{ii} .$$

Die Determinante $\det A$ einer $n \times n$ -Matrix gibt das "Spaltenvolumen" von A an. Für eine 2×2 -Matrix A ist

$$\det A = \det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = a_{11}a_{22} - a_{21}a_{12} \quad ,$$

für eine 3×3 -Matrix A ist

$$\det A = \det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \\ = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{21}a_{32}a_{13} - a_{31}a_{22}a_{13} - a_{32}a_{23}a_{11} - a_{21}a_{12}a_{33} .$$

Für höherdimensionale Matrizen nimmt man Entwicklungen nach einer Spalte oder Zeile gemäß dem *Laplaceschen Entwicklungssatz* vor und reduziert dadurch die Dimension der Matrizen, deren Determinanten berechnet werden müssen, bis auf 3×3 -Matrizen oder auch 2×2 -Matrizen. Bezeichnet A_{ij} die Matrix, die dadurch entsteht, daß man in einer Matrix A die i -te Zeile und die j -te Spalte streicht, so ergibt sich für die Entwicklung nach der i -ten Zeile ($i=1, \dots, n$)

$$\det A = \det \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix} = \sum_{j=1}^n (-1)^{i+j} a_{ij} \det A_{ij}$$

und für die Entwicklung nach der j -ten Spalte ($j=1, \dots, n$)

$$\det A = \det \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix} = \sum_{i=1}^n (-1)^{i+j} a_{ij} \det A_{ij} \quad .$$

Natürlich wird man bei einer Matrix, in der einige Elemente a_{ij} den Wert Null haben, nach derjenigen Spalte oder Zeile entwickeln, die möglichst viele Nullen enthält.

Beispiel: Wir wollen die Determinante der 4×4 -Matrix

$$A = \begin{pmatrix} 7 & -1 & 3 & 2 \\ 4 & 0 & 9 & 8 \\ 7 & 6 & 3 & 10 \\ 6 & 4 & 1 & 5 \end{pmatrix}$$

durch Entwicklung nach der 2. Spalte ($j=2$) berechnen. Es ergibt sich

$$\det A = (-1)^{1+2} a_{12} \det A_{12} + (-1)^{2+2} a_{22} \det A_{22} + (-1)^{3+2} a_{32} \det A_{32} \\ + (-1)^{4+2} a_{42} \det A_{42}$$

$$\begin{aligned}
&= (-1)^3 \cdot (-1) \cdot \det \begin{pmatrix} 4 & 9 & 8 \\ 7 & 3 & 10 \\ 6 & 1 & 5 \end{pmatrix} + (-1)^4 \cdot 0 \cdot \det \begin{pmatrix} 7 & 3 & 2 \\ 7 & 3 & 10 \\ 6 & 1 & 5 \end{pmatrix} \\
&\quad + (-1)^5 \cdot 6 \cdot \det \begin{pmatrix} 7 & 3 & 2 \\ 4 & 9 & 8 \\ 6 & 1 & 5 \end{pmatrix} + (-1)^6 \cdot 4 \cdot \det \begin{pmatrix} 7 & 3 & 2 \\ 4 & 9 & 8 \\ 7 & 3 & 10 \end{pmatrix} \\
&= 1(4 \cdot 3 \cdot 5 + 9 \cdot 10 \cdot 6 + 7 \cdot 1 \cdot 8 - 6 \cdot 3 \cdot 8 - 1 \cdot 10 \cdot 4 - 7 \cdot 9 \cdot 5) + 0 \\
&\quad - 6(7 \cdot 9 \cdot 5 + 3 \cdot 8 \cdot 6 + 4 \cdot 1 \cdot 2 - 6 \cdot 9 \cdot 2 - 1 \cdot 8 \cdot 7 - 4 \cdot 3 \cdot 5) \\
&\quad + 4(7 \cdot 9 \cdot 10 + 3 \cdot 8 \cdot 7 + 4 \cdot 3 \cdot 2 - 7 \cdot 9 \cdot 2 - 3 \cdot 8 \cdot 7 - 4 \cdot 3 \cdot 10) \\
&= 1 \cdot 157 + 0 - 6 \cdot 243 + 4 \cdot 408 \\
&= 331 \quad .
\end{aligned}$$

Eine $n \times n$ -Matrix, für die $\det A \neq 0$ gilt, heißt *reguläre Matrix*.

Für eine $n \times n$ -Matrix D , die auf der Hauptdiagonale Elemente d_1, \dots, d_n und sonst nur Nullen hat, schreibt man häufig

$$D = \text{diag}(d_1, \dots, d_n) = \begin{pmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & & d_n \end{pmatrix} .$$

Ein wichtiges Hilfsmittel und gleichzeitig Anwendungsgebiet der Matrizenrechnung ist das Lösen von *linearen Gleichungssystemen*. Betrachten wir etwa das Gleichungssystem

$$\begin{aligned}
2x_1 + 5x_2 &= 4 \\
-6x_1 + 10x_2 &= -2 \quad ,
\end{aligned}$$

dessen Lösung durch $x_1 = 1$ und $x_2 = 2/5$ gegeben ist. In Matrixschreibweise $Ax = b$ mit $x = (x_1, x_2)^T$ läßt sich dieses System auch schreiben als

$$\begin{pmatrix} 2 & 5 \\ -6 & 10 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 4 \\ -2 \end{pmatrix} .$$

Natürlich ist ein solches Gleichungssystem nicht immer lösbar. Eine Lösung existiert nur dann, wenn der *Rang* $\text{rg} A$ der Matrix A , d.h. die Anzahl der *linear unabhängigen* Zeilen (oder Spalten) von A gleich dem Rang der Matrix (A, b) ist. Eine Zeile (oder Spalte) einer Matrix heißt *linear unabhängig* von den übrigen Zeilen (oder Spalten), wenn man sie nicht als Summe von Vielfachen (Linearkombinationen) der übrigen Zeilen (oder Spalten) darstellen kann. Den Rang einer Matrix bestimmt man, indem man sie durch Multiplikation von Zeilen mit Zahlen und Addition von Zeilen auf *obere Dreiecks-*

gestalt, d.h. unterhalb der vom oberen Eckelement ausgehenden Diagonalen sollen nur noch Nullen stehen, bringt. (Ferner gilt: $\text{rg } A = \text{tr } AA^+$; s. unten.)

Beispiel: Wir betrachten

$$A = \begin{pmatrix} 3 & 7 & 6 & 5 \\ 1 & 0 & 7 & 2 \\ 6 & 7 & 27 & 11 \end{pmatrix} .$$

Multipliziert man die zweite Zeile von A mit (-3) und addiert sie zur ersten, so ergibt sich die Matrix

$$\begin{pmatrix} 3 & 7 & 6 & 5 \\ 0 & 7 & -15 & -1 \\ 6 & 7 & 27 & 11 \end{pmatrix} .$$

Addiert man dann die dritte Zeile zum (-2) - fachen der ersten, so erhält man die Matrix

$$\begin{pmatrix} 3 & 7 & 6 & 5 \\ 0 & 7 & -15 & -1 \\ 0 & -7 & 15 & 1 \end{pmatrix} ,$$

und man sieht, daß die dritte Zeile gerade das (-1) - fache der zweiten ist, d.h. es ergibt sich die Matrix

$$\begin{pmatrix} 3 & 7 & 6 & 5 \\ 0 & 7 & -15 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix} .$$

Wir sehen also, daß gilt $\text{rg } A = 2$, denn A hat nur zwei linear unabhängige Zeilen. ┌

Weiterhin ist es möglich, daß ein Gleichungssystem mehrere Lösungen besitzt. Dies ist zum Beispiel für das System

$$\begin{aligned} 6x_1 - x_2 - x_3 &= 0 \\ 8x_1 - x_2 &= 6 \end{aligned} ,$$

das in Matrixschreibweise die Gestalt

$$\begin{pmatrix} 6 & -1 & -1 \\ 8 & -1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 6 \end{pmatrix}$$

besitzt, der Fall. Für jede Zahl x_1 ist $(x_1, x_2, x_3) = (x_1, 8x_1 - 6, -2x_1 + 6)$ eine Lösung des Systems; z.B. $(x_1, x_2, x_3) = (4, 26, -2)$.

Ein Verfahren zur Lösung linearer Gleichungssysteme ist das Gauß'sche Eliminationsverfahren. Dabei macht man sich zunutzen, daß die Multiplikation einer beliebigen Zeile eines Gleichungssystems das System nicht inhaltlich

verändert und daß sich auch durch Addition zweier beliebiger Zeilen das System nicht ändert. Bei diesem Verfahren bringt man die *Koeffizientenmatrix* A des Gleichungssystems auf obere Dreiecksgestalt und führt die dabei verwandten Multiplikationen und Additionen parallel beim *Lösungsvektor* b durch.

Beispiel: Mittels des Gauß'schen Eliminationsverfahrens wollen wir die Lösungen des Gleichungssystems

$$\begin{pmatrix} 3 & 4 & 2 & 0 \\ 2 & 1 & 4 & 2 \\ 1 & 0 & 6 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix}$$

berechnen. Dabei benutzen wir das Schema aus Tab.4, in dem rechts die ausgeführten Multiplikationen und Additionen angegeben sind.

Tab.4: Gauß'sches Eliminationsverfahren im Beispiel

x_1	x_2	x_3	x_4		
3	4	2	0	2	
2	1	4	2	1	$\cdot 2$
1	0	6	1	3	$\cdot (-3) \rightarrow +$ $\cdot (-3) \rightarrow +$
3	4	2	0	2	
0	5	-8	-6	1	$\cdot 4$
0	4	-16	-3	-7	$\cdot (-5) \rightarrow +$
3	4	2	0	2	
0	5	-8	-6	1	
0	0	48	-9	39	

Es ergibt sich

$$48x_3 - 9x_4 = 39 \quad , \text{ d.h. } x_4 = \frac{48x_3 - 39}{9} = \frac{16}{3}x_3 - \frac{13}{3}$$

und somit

$$5x_2 - 8x_3 - 6x_4 = 5x_2 - 8x_3 - 48x_3 + 39 = 1 \quad ,$$

$$\text{d.h. } x_2 = \frac{56x_3 - 38}{5} = \frac{56}{5}x_3 - \frac{38}{5}$$

und schließlich

$$3x_1 + 4x_2 + 2x_3 = 3x_1 + \frac{224}{5}x_3 - \frac{152}{5} + 2x_3 = 3x_1 + \frac{234}{5}x_3 - \frac{152}{5} = 2$$

$$\text{d.h. } x_1 = \frac{-234x_3 + 162}{15} = -\frac{78}{5}x_3 + \frac{54}{5} .$$

Für jede Wahl von x_3 ist also

$$(x_1, x_2, x_3, x_4) = \left(-\frac{78}{5}x_3 + \frac{54}{5}, \frac{56}{5}x_3 - \frac{38}{5}, x_3, \frac{16}{3}x_3 - \frac{13}{3}\right)$$

eine Lösung unseres Gleichungssystems. └

Ein wichtiger Teil der Matrizenrechnung ist die Berechnung von Eigenwerten und Eigenvektoren. Die *Eigenwerte* einer quadratischen $n \times n$ -Matrix A werden bestimmt als die Nullstellen des *charakteristischen Polynoms* $\det(A - \lambda I_n)$, d.h. die n Eigenwerte $\lambda_1, \dots, \lambda_n$ von A sind die Lösungen der Gleichung

$$\det(A - \lambda I_n) = 0 .$$

Die zum Eigenwert λ_i gehörigen *Eigenvektoren* x sind dann die Lösungen des Gleichungssystems ($x \neq 0$)

$$Ax = \lambda_i x .$$

Es gilt $\lambda_1 + \dots + \lambda_n = \text{tr} A$ und für symmetrisches A auch $\lambda_1 \cdot \dots \cdot \lambda_n = \det A$.

Beispiel: Wir wollen die Eigenwerte und die zugehörigen Eigenvektoren der Matrix

$$A = \begin{pmatrix} 7 & 3 \\ 1 & 5 \end{pmatrix}$$

berechnen. Das charakteristische Polynom von A ergibt sich zu

$$\begin{aligned} \det(A - \lambda I_2) &= \det \left(\begin{pmatrix} 7 & 3 \\ 1 & 5 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right) = \det \begin{pmatrix} 7-\lambda & 3 \\ 1 & 5-\lambda \end{pmatrix} \\ &= (7-\lambda)(5-\lambda) - 1 \cdot 3 = 35 - 7\lambda - 5\lambda + \lambda^2 - 3 \\ &= \lambda^2 - 12\lambda + 32 = \lambda^2 - 12\lambda + 36 - 36 + 32 \\ &= (\lambda - 6)^2 - 4 = (\lambda - 6 - 2)(\lambda - 6 + 2) \\ &= (\lambda - 8)(\lambda - 4) . \end{aligned}$$

Da für $\lambda = 8$ oder $\lambda = 4$ gilt

$$(\lambda - 8)(\lambda - 4) = 0$$

sind $\lambda_1 = 8$ und $\lambda_2 = 4$ die Eigenwerte von A . Die zu $\lambda_1 = 8$ gehörigen Eigenvektoren sind die Lösungen des Systems

$$Ax = 8x \quad , \text{ d.h. } \begin{cases} 7x_1 + 3x_2 = 8x_1 \\ x_1 + 5x_2 = 8x_2 \end{cases} .$$

Damit ergeben sich die Eigenvektoren zum Eigenwert $\lambda_1 = 8$ zu

$(x_1, x_2)^T = (3x_2, x_2)^T$, und genauso erhalten wir als Eigenvektoren zum Eigenwert $\lambda_2 = 4$ die Vektoren $(x_1, x_2)^T = (-x_2, x_2)^T$ mit $x_2 \neq 0$ jeweils beliebig. ┌

Ein weiteres wichtiges Element der Matrizenrechnung ist die Bestimmung von inversen Matrizen. Für eine quadratische $n \times n$ -Matrix A mit $\text{rg } A = n$ (äquivalent dazu ist $\det A \neq 0$; A reguläre Matrix) ist die zu A inverse Matrix A^{-1} bestimmt durch

$$AA^{-1} = I_n \quad .$$

Ist A eine reguläre Matrix, so ist jedes Gleichungssystem der Art $Ax = b$ lösbar, und zwar eindeutig. Die Lösung läßt sich angeben als $x = A^{-1}b$.

Im folgenden werden wir bei der Verwendung einer Inversen immer implizit unterstellen, daß eine solche auch existiert, ohne dies in den Voraussetzungen immer explizit zu erwähnen.

Bei der Bestimmung von inversen Matrizen kann man sich wiederum das Gauß'sche Eliminationsverfahren zunutze machen. Man notiert die reguläre Matrix A und daneben die n -dimensionale Einheitsmatrix I_n und multipliziert dann einzelne Zeilen mit Zahlen und addiert Zeilen von A solange bis man A in eine Einheitsmatrix verwandelt hat. Parallel dazu führt man die gleichen Multiplikationen und Additionen ausgehend von I_n durch. Ist A dann in eine Einheitsmatrix verwandelt, so ist I_n in die inverse Matrix A^{-1} verwandelt.

Beispiel: Wir wollen die zur regulären 3×3 -Matrix

$$A = \begin{pmatrix} 4 & 7 & -3 \\ 2 & 6 & 1 \\ 1 & 5 & 9 \end{pmatrix}$$

inverse Matrix A^{-1} bestimmen. Dazu verwenden wir das Schema aus Tab.5, in dem rechts die durchgeführten Multiplikationen und Additionen angegeben sind. Die zu A inverse Matrix ergibt sich also zu

$$A^{-1} = \frac{1}{65} \begin{pmatrix} 49 & -78 & 25 \\ -17 & 39 & -10 \\ 4 & -13 & 10 \end{pmatrix} \quad .$$

Dieses Ergebnis kann dadurch kontrolliert werden, daß man AA^{-1} ausrechnet. Hat man korrekt gerechnet, so muß gelten

$$AA^{-1} = I_3 \quad .$$

Hier ergibt sich

$$AA^{-1} = \frac{1}{65} \begin{pmatrix} 4 & 7 & -3 \\ 2 & 6 & 1 \\ 1 & 5 & 9 \end{pmatrix} \begin{pmatrix} 49 & -78 & 25 \\ -17 & 39 & -10 \\ 4 & -13 & 10 \end{pmatrix} = \frac{1}{65} \begin{pmatrix} 65 & 0 & 0 \\ 0 & 65 & 0 \\ 0 & 0 & 65 \end{pmatrix} = I_3 ;$$

wir haben also richtig gerechnet.

Tab.5: Berechnung einer inversen Matrix

A =	$\begin{pmatrix} 4 & 7 & -3 \\ 2 & 6 & 1 \\ 1 & 5 & 9 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	$\begin{matrix} \cdot (-2) \downarrow^+ \\ \\ (-4) \downarrow^+ \end{matrix}$
	$\begin{pmatrix} 4 & 7 & -3 \\ 0 & -5 & -5 \\ 0 & -13 & -39 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 \\ 1 & -2 & 0 \\ 1 & 0 & -4 \end{pmatrix}$	$\begin{matrix} \cdot 13 \downarrow^+ \\ \cdot (-5) \downarrow^+ \end{matrix}$
	$\begin{pmatrix} 4 & 7 & -3 \\ 0 & -5 & -5 \\ 0 & 0 & 130 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 \\ 1 & -2 & 0 \\ 8 & -26 & 20 \end{pmatrix}$	$\begin{matrix} \cdot 26 \downarrow^+ & \cdot 130 \downarrow^+ \\ & \cdot 3 \downarrow^+ \end{matrix}$
	$\begin{pmatrix} 520 & 910 & 0 \\ 0 & -130 & 0 \\ 0 & 0 & 130 \end{pmatrix}$	$\begin{pmatrix} 154 & -78 & 60 \\ 34 & -78 & 20 \\ 8 & -26 & 20 \end{pmatrix}$	$\cdot 7 \downarrow^+$
	$\begin{pmatrix} 520 & 0 & 0 \\ 0 & -130 & 0 \\ 0 & 0 & 130 \end{pmatrix}$	$\begin{pmatrix} 392 & -624 & 200 \\ 34 & -78 & 20 \\ 8 & -26 & 20 \end{pmatrix}$	$\begin{matrix} \cdot 1/520 \\ \cdot (-1/130) \\ \cdot 1/130 \end{matrix}$
	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 49/65 & -78/65 & 25/65 \\ -17/65 & 39/65 & -10/65 \\ 4/65 & -13/65 & 10/65 \end{pmatrix}$	$= A^{-1}$

Ist eine Matrix A nicht regulär, so läßt sich keine Inverse bestimmen. Stattdessen muß man sich mit sogenannten generalisierten Inversen von A begnügen. A^- heißt eine *generalisierte Inverse* zur $n \times m$ -Matrix A, falls gilt

$$AA^-A = A$$

Hat man ein lösbares Gleichungssystem $Ax = b$ gegeben, so ist eine Lösung angebar als $x = A^-b$. Die $n \times m$ -Matrix A^- ist allerdings i.a. nicht eindeutig. Um Eindeutigkeit zu erreichen, wird daher häufig die sogenannte *Pseudoinverse* oder *Moore-Penrose-Inverse* A^+ von A verwandt. A^+ ist durch die folgenden 4 Bedingungen bestimmt:

$$AA^+A = A$$

$$A^+AA^+ = A^+$$

$$(A^+A) = (A^+A)^T$$

$$(AA^+) = (AA^+)^T$$

Da die erste Bedingung gerade der für generalisierte Inverse entspricht, ist A^+ eine spezielle generalisierte Inverse, die in Fällen regulärer Matrizen A mit A^{-1} übereinstimmt. Zur Berechnung von A^+ gibt es viele Verfahren, vgl. z.B. Albert (1972), Ben-Israel/Greville (1980); ein spezielles wollen wir hier angeben. (Dies eignet sich im Falle der Existenz auch zur Berechnung einer Inversen.) Wir bezeichnen die Spaltenvektoren einer $n \times m$ -Matrix A mit a_1, \dots, a_m und die Matrix der ersten k Spalten von A mit A_k , d.h. $A = (a_1, \dots, a_m)$ und $A_k = (a_1, \dots, a_k)$ für $k=1, \dots, m$. Für $k=2, \dots, m$ definieren wir weiter

$$d_k = A_{k-1}^+ a_k \quad , \quad c_k = a_k - A_{k-1} d_k \quad .$$

Dann gilt für $k=2, \dots, m$

$$A_k^+ = \begin{pmatrix} A_{k-1}^+ & -d_k b_k^T \\ & b_k^T \end{pmatrix}$$

mit

$$b_k^T = \begin{cases} c_k^+ & , \text{ falls } c_k \neq 0 \\ (1 + d_k^T d_k)^{-1} d_k^T A_{k-1}^+ & , \text{ falls } c_k = 0 \end{cases} .$$

Benutzt man, daß die Pseudoinverse eines Vektors x durch

$$x^+ = \frac{1}{x^T x} x^T \quad , \text{ falls } x \neq 0 \quad , \text{ und } \quad x^+ = 0 \quad , \text{ falls } x = 0 \quad ,$$

gegeben ist, so kann man ausgehend von

$$A_1^+ = a_1^+$$

sukzessive $A_2^+, A_3^+, \dots, A_m^+ = A^+$ bestimmen.

Ist A eine $n \times m$ -Matrix mit $n < m$, so wird man den Grevilleschen Algorithmus nicht auf A sondern auf A^T anwenden, da dann die Schrittzahl geringer ist. Ist dann $(A^T)^+$ die Pseudoinverse von A^T , so ergibt sich die Pseudoinverse von A zu

$$A^+ = \left((A^T)^+ \right)^T \quad ,$$

denn es gilt stets

$$(A^+)^T = (A^T)^+ \quad .$$

Beispiel: Wir wollen die Pseudoinverse A^+ zur 4×3 -Matrix

$$A = \begin{pmatrix} 3 & 2 & 7 \\ -5 & 4 & 1 \\ 9 & 3 & -2 \\ 8 & 6 & 10 \end{pmatrix}$$

bestimmen. Es ist zunächst

$$A_1^+ = \begin{pmatrix} 3 \\ -5 \\ 9 \\ 8 \end{pmatrix}^+ = \frac{1}{3^2 + (-5)^2 + 9^2 + 8^2} (3, -5, 9, 8) = \frac{1}{179} (3, -5, 9, 8) \quad ,$$

und somit ergibt sich für $k=2$

$$d_2 = A_1^+ a_2 = \frac{1}{179} (3, -5, 9, 8) \begin{pmatrix} 2 \\ 4 \\ 3 \\ 6 \end{pmatrix} = \frac{1}{179} (3 \cdot 2 - 5 \cdot 4 + 9 \cdot 3 + 8 \cdot 6) = \frac{61}{179} \quad ,$$

$$c_2 = a_2 - A_1 d_2 = a_2 - a_1 d_2 = \begin{pmatrix} 2 \\ 4 \\ 3 \\ 6 \end{pmatrix} - \begin{pmatrix} 3 \\ -5 \\ 9 \\ 8 \end{pmatrix} \cdot \frac{61}{179} = \frac{1}{179} \begin{pmatrix} 358 - 183 \\ 716 + 305 \\ 537 - 549 \\ 1074 - 488 \end{pmatrix} = \frac{1}{179} \begin{pmatrix} 175 \\ 1021 \\ -12 \\ 586 \end{pmatrix} \quad ,$$

d.h. es ist

$$b_2^T = c_2^+ = \frac{1}{c_2^T c_2} c_2^T = \frac{179^2}{1416606} \cdot \frac{1}{179} (175, 1021, -12, 586) \\ = \frac{1}{7914} (175, 1021, -12, 586) \quad .$$

Damit ergibt sich dann

$$A_2^+ = \begin{pmatrix} A_1^+ - d_2 b_2^T \\ b_2^T \end{pmatrix} = \begin{pmatrix} \frac{1}{179} (3, -5, 9, 8) - \frac{61}{179} \frac{1}{7914} (175, 1021, -12, 586) \\ \frac{1}{7914} (175, 1021, -12, 586) \end{pmatrix} \\ = \frac{1}{7914} \begin{pmatrix} 73 & -569 & 402 & 154 \\ 175 & 1021 & -12 & 586 \end{pmatrix} \quad .$$

Für $k=3$ erhalten wir nun

$$d_3 = A_2^+ a_3 = \frac{1}{7914} \begin{pmatrix} 73 & -569 & 402 & 154 \\ 175 & 1021 & -12 & 586 \end{pmatrix} \begin{pmatrix} 7 \\ 1 \\ -2 \\ 10 \end{pmatrix} = \frac{1}{7914} \begin{pmatrix} 678 \\ 8130 \end{pmatrix} \quad ,$$

$$c_3 = a_3 - A_2 d_3 = \begin{pmatrix} 7 \\ 1 \\ -2 \\ 10 \end{pmatrix} - \begin{pmatrix} 3 & 2 \\ -5 & 4 \\ 9 & 3 \\ 8 & 6 \end{pmatrix} \cdot \frac{1}{7914} \begin{pmatrix} 678 \\ 8130 \end{pmatrix} = \frac{1}{7914} \begin{pmatrix} 55398 \\ 7914 \\ -15828 \\ 79140 \end{pmatrix} - \begin{pmatrix} 18294 \\ 29130 \\ 30492 \\ 54204 \end{pmatrix}$$

$$= \frac{1}{7914} \begin{pmatrix} 37104 \\ -21216 \\ -46320 \\ 24936 \end{pmatrix}$$

und daraus dann

$$\begin{aligned} b_3^T &= c_3^+ = \frac{1}{c_3^T c_3} c_3^T = \frac{7914^2}{4594171968} \cdot \frac{1}{7914} (37104, -21216, -46320, 24936) \\ &= \frac{1}{580512} (37104, -21216, -46320, 24936) \quad , \end{aligned}$$

so daß sich

$$\begin{aligned} A_3^+ &= A^+ = \begin{pmatrix} A_2^+ - d_3 b_3^T \\ b_3^T \end{pmatrix} \\ &= \frac{1}{580512} \begin{pmatrix} 2176 & -39920 & 33456 & 9160 \\ -25280 & 96688 & 46704 & 17368 \\ 37104 & -21216 & -46320 & 24936 \end{pmatrix} \\ &= \frac{1}{72564} \begin{pmatrix} 272 & -4990 & 4182 & 1145 \\ -3160 & 12086 & 5838 & 2171 \\ 4638 & -2652 & -5790 & 3117 \end{pmatrix} \end{aligned}$$

ergibt.

Nützlich sind häufig auch die Beziehungen

$$A^+ = (A^T A)^+ A^T \quad \text{und} \quad A^+ = A^T (A A^T)^+ \quad ,$$

womit man z.B. für den Fall, daß $A^T A$ invertierbar ist, erhält: $A^+ = (A^T A)^{-1} A^T$.

[Betrachtet man das Gleichungssystem

$$Ax = b \quad ,$$

bzw., falls dies nicht lösbar ist, das allgemeine Problem

$$\text{minimiere } (Ax - b)^T (Ax - b) \text{ bzgl. } x \quad ,$$

mit A $n \times m$ -Matrix, b n -dimensionaler Vektor, x m -dimensionaler (unbekannter) Vektor, so sind die Lösungen x darstellbar als

$$x = A^+ b + (I_m - A^+ A) w \quad ,$$

wobei w ein beliebiger n -dimensionaler Vektor ist. Auch im Falle eines lösbaren Gleichungssystems $Ax = b$ sind dessen Lösungen in dieser Form angebar, und es gilt in diesem Falle insbesondere $Ax_0 = b$ für $x_0 = A^+ b$ ($x = x_0$, $w = 0$).

Dabei haben die Vektoren $v = (I_m - A^+A)w$ die Eigenschaft, daß $Av = 0$ gilt, d.h. sie gehören zum Kern (nullspace) $N(A)$ von A ; das sind gerade alle Vektoren x , für die $Ax = 0$ ist, und jedes solche x läßt sich in der Form $x = (I_m - A^+A)w$, mit w m -dimensionaler Vektor, darstellen.

Die ausgezeichnete Lösung $x = x_0 = A^+b$ ($w = 0$) hat die Eigenschaft, daß sie unter allen Lösungen obiger Aufgabe den kleinsten Betrag bzw. die kleinste Norm $\|x\| = (x^T x)^{1/2}$ besitzt, und daß sie orthogonal zum Kern $N(A)$ ist, d.h. für alle Vektoren v mit $Av = 0$ gilt: $v^T A^+ b = 0$. Ferner gilt: $\text{rg } A = \text{tr } AA^+$.]

Allgemein nennt man zwei Vektoren x und y *orthogonale* Vektoren, wenn $x^T y = 0$ gilt, und schreibt dann $x \perp y$. Sie heißen *orthonormale* Vektoren, wenn sie orthogonal und zudem noch normiert (auf Betrag 1) sind, d.h. $x^T y = 0$ und $\|x\| = \|y\| = 1$.

Verwenden wir das Zeichen $\|\cdot\|$ ohne weitere Kennzeichnung, so verstehen wir darunter stets den *euklidischen Betrag* bzw. die *euklidische Norm*

$$\|x\| = (x^T x)^{1/2} = \sqrt{\sum_{i=1}^n x_i^2} \quad \text{für } x = (x_1, \dots, x_n)^T \quad .$$

Zum Abschluß dieses Abschnitts über Matrizenrechnung wollen wir noch den Begriff der *Definitheit* einer symmetrischen $n \times n$ -Matrix A einführen. Dazu bezeichne x einen n -dimensionalen Vektor und A eine $n \times n$ -dimensionale Matrix. A heißt *positiv (negativ) definit*, falls für alle $x \neq 0$ gilt

$$x^T A x > 0 \quad (< 0) \quad [\hat{=} \text{alle Eigenwerte von } A > 0 \quad (< 0)] \quad ,$$

und A heißt *positiv (negativ) semidefinit*, falls für alle x gilt

$$x^T A x \geq 0 \quad (\leq 0) \quad [\hat{=} \text{alle Eigenwerte von } A \geq 0 \quad (\leq 0)] \quad .$$

In allen anderen Fällen heißt die Matrix A indefinit. Ist A semidefinit und invertierbar, so ist A definit.

5. MEHRDIMENSIONALE UND MULTIVARIATE VERTEILUNGEN

Im Abschnitt 3 haben wir uns mit Zufallsvariablen und ihren Verteilungen beschäftigt. Mitunter ist man aber gleichzeitig an mehreren Zufallsvariablen X_1, \dots, X_n , d.h. an einem *zufälligen Vektor* oder *Zufallsvektor* $(X_1, \dots, X_n)^T$, interessiert. Zum Beispiel kann man sich bei Menschen für ihre Größe X_1 und ihr Gewicht X_2 interessieren. Zur Beschreibung der *Wahrscheinlichkeits-*

Verteilung eines Zufallsvektors $X = (X_1, \dots, X_n)^T$ genügt in der Regel nicht die Kenntnis der Verteilung der einzelnen Komponenten X_1, \dots, X_n ; dies ist nur dann der Fall, wenn X_1, \dots, X_n stochastisch unabhängige Zufallsvariablen sind. Vielmehr muß die *mehrdimensionale (gemeinsame) Verteilungsfunktion* zur gemeinsamen Verteilung des Zufallsvektors X

$$F_X(x_1, \dots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n)$$

bekannt sein. Die Verteilung einer einzelnen Zufallsvariablen X_i bezeichnet man als die *i-te Randverteilung* von X und ihre Verteilungsfunktion berechnet sich als

$$\begin{aligned} F_{X_i}(x_i) &= P(X_i \leq x_i) = P(X_1 \leq \infty, \dots, X_i \leq x_i, \dots, X_n \leq \infty) \\ &= \lim_{\substack{x_j \rightarrow \infty \\ j \neq i}} F_X(x_1, \dots, x_n) \quad . \end{aligned}$$

Wie im eindimensionalen spricht man von einem *diskret verteilten* Zufallsvektor X , wenn die Anzahl der möglichen Realisationen $(x_1, \dots, x_n)^T$ abzählbar ist, d.h. wenn die Verteilung von X durch die Angabe der Wahrscheinlichkeiten für ihre Realisationen eindeutig festgelegt werden kann. Dagegen spricht man von einem *stetig verteilten* Zufallsvektor X , wenn sich seine Verteilungsfunktion in der Form

$$F_X(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f_X(\xi_1, \dots, \xi_n) d\xi_1 \dots d\xi_n$$

darstellen läßt; die nichtnegative Funktion $f_X(x_1, \dots, x_n)$ nennt man auch die *gemeinsame Dichte* der X_i . Sind die Zufallsvariablen X_i stochastisch unabhängig, so entspricht ihre gemeinsame Dichte gerade dem Produkt der Dichten der X_i :

$$f_X(x_1, \dots, x_n) = f_{X_1}(x_1) \cdot \dots \cdot f_{X_n}(x_n) \quad .$$

Im folgenden werden wir uns mit der mehr- und multidimensionalen Normalverteilung beschäftigen. Dabei nennen wir einen n -dimensionalen Zufallsvektor X *mehrdimensional normalverteilt*, wenn er sich darstellen läßt als

$$X = AY + b \quad ,$$

wobei A eine $n \times p$ -Matrix, b ein n -dimensionaler Vektor und $Y = (Y_1, \dots, Y_p)$ ein zufälliger Vektor aus unabhängigen $N(0;1)$ -verteilten Zufallsvariablen Y_1, \dots, Y_p ist. Der Erwartungswert des Zufallsvektors X ist dann wegen $E(Y_i) = 0$ für $i=1, \dots, p$ gerade

$$E(X) = E \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} = \begin{pmatrix} E(X_1) \\ \vdots \\ E(X_n) \end{pmatrix} = b$$

und die Kovarianzmatrix von X ist gegeben durch

$$\text{Cov}(X) = \mathbb{I}_X = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{pmatrix} = AA^T \quad .$$

Man schreibt auch kurz X ist $N(b; \mathbb{I}_X)$ -verteilt oder $X \sim N(b; \mathbb{I}_X)$, denn die mehrdimensionale Normalverteilung ist durch Erwartungswert und Kovarianzmatrix eindeutig charakterisiert. [Entsprechend läßt sich $E(X)$ und \mathbb{I}_X auch bei nichtnormalverteiltem X bzw. Y_i für $i=1, \dots, p$ definieren.]

Aus der Kovarianzmatrix \mathbb{I}_X von X läßt sich direkt die zugehörige Korrelationsmatrix $\text{Corr}(X)$ bestimmen. Es ist

$$\text{Corr}(X) = \begin{pmatrix} 1 & \frac{\sigma_{12}}{\sigma_1 \sigma_2} & \cdots & \frac{\sigma_{1n}}{\sigma_1 \sigma_n} \\ \frac{\sigma_{21}}{\sigma_2 \sigma_1} & 1 & \cdots & \frac{\sigma_{2n}}{\sigma_2 \sigma_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sigma_{n1}}{\sigma_n \sigma_1} & \frac{\sigma_{n2}}{\sigma_n \sigma_2} & \cdots & 1 \end{pmatrix} = \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{21} & 1 & \cdots & \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \cdots & 1 \end{pmatrix} \quad .$$

Natürlich sind Kovarianz- und Korrelationsmatrix von X stets symmetrische $n \times n$ -Matrizen, d.h. $\sigma_{ij} = \sigma_{ji}$ und $\rho_{ij} = \rho_{ji}$ und positiv semidefinit bzw., falls sie invertierbar sind, sogar positiv definit.

Sind $x_1 = (x_{11}, \dots, x_{1n})^T, \dots, x_l = (x_{l1}, \dots, x_{ln})^T$ Realisationen aus einer n -dimensionalen $N(b; \mathbb{I}_X)$ -Verteilung, so wird b vermittels

$$\bar{x} = \begin{pmatrix} \bar{x}_{.1} \\ \vdots \\ \bar{x}_{.n} \end{pmatrix} = \frac{1}{l} \begin{pmatrix} 1 \\ \sum_{i=1}^l x_{i1} \\ \vdots \\ 1 \\ \sum_{i=1}^l x_{in} \end{pmatrix}$$

und \mathbb{I}_X im Falle $l \geq 2$ vermittels

$$S = \frac{1}{l-1} \sum_{i=1}^l (x_i - \bar{x})(x_i - \bar{x})^T$$

$$= \begin{pmatrix} \frac{1}{T-1} \sum_{i=1}^1 (x_{i1} - \bar{x}_{.1})^2 & \dots & \frac{1}{T-1} \sum_{i=1}^1 (x_{i1} - \bar{x}_{.1})(x_{j1} - \bar{x}_{.1}) \\ \vdots & & \vdots \\ \frac{1}{T-1} \sum_{i=1}^1 (x_{i1} - \bar{x}_{.1})(x_{i1} - \bar{x}_{.1}) & \dots & \frac{1}{T-1} \sum_{i=1}^1 (x_{i1} - \bar{x}_{.1})^2 \end{pmatrix}$$

$$= \begin{pmatrix} s_1^2 & s_{12} & \dots & s_{11} \\ s_{21} & s_2^2 & \dots & s_{21} \\ \vdots & \vdots & & \vdots \\ s_{11} & s_{12} & \dots & s_1^2 \end{pmatrix}$$

geschätzt. Ein Schätzer für $\text{Corr}(X)$ ist dann

$$R = \begin{pmatrix} 1 & r_{12} & \dots & r_{11} \\ r_{21} & 1 & \dots & r_{21} \\ \vdots & \vdots & & \vdots \\ r_{11} & r_{12} & \dots & 1 \end{pmatrix} \quad \text{mit}$$

$$r_{jk} = \frac{s_{jk}}{s_j \cdot s_k} = \frac{\sum_{i=1}^1 (x_{ij} - \bar{x}_{.j})(x_{ik} - \bar{x}_{.k})}{\sqrt{\sum_{i=1}^1 (x_{ij} - \bar{x}_{.j})^2 \cdot \sum_{i=1}^1 (x_{ik} - \bar{x}_{.k})^2}} \quad .$$

Ist \mathbb{f}_X regulär, d.h. \mathbb{f}_X^{-1} existiert, so besitzt die Verteilung von X eine Dichte; diese hat mit $x = (x_1, \dots, x_n)^T$ die Gestalt

$$f_X(x_1, \dots, x_n) = \frac{1}{\sqrt{(2\pi)^n \cdot \det \mathbb{f}_X}} \cdot e^{-\frac{1}{2}(x-b)^T \mathbb{f}_X^{-1}(x-b)} \quad ,$$

vgl. Abb.7- 10, wo die Dichten einiger bivariater Normalverteilungen dargestellt sind; natürlich haben diese Dichten eigentlich eine glatte Oberfläche, was sich aus technischen Gründen in den Abbildungen jedoch nicht realisieren ließ. Die Verteilungsfunktion von X ist dann natürlich

$$F_X(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f_X(\xi_1, \dots, \xi_n) d\xi_1 \dots d\xi_n \quad .$$

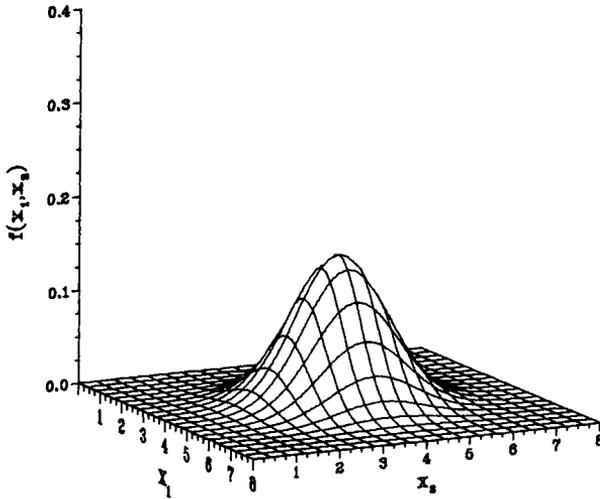


Abb.7: Dichte der bivariaten Normalverteilung $N\left(\begin{pmatrix} 4 \\ 4 \end{pmatrix}; \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$

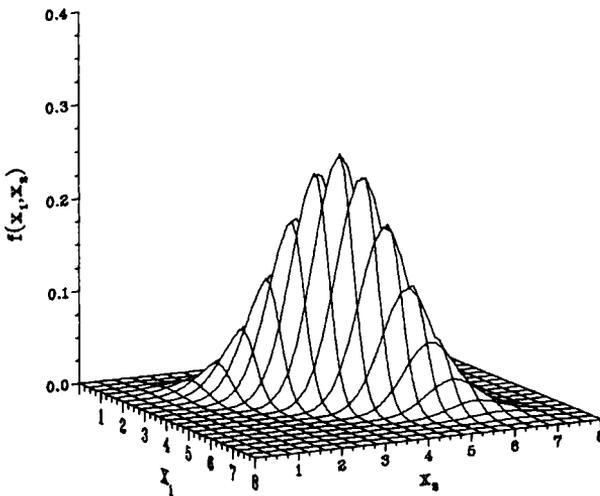


Abb.8: Dichte der bivariaten Normalverteilung $N\left(\begin{pmatrix} 4 \\ 4 \end{pmatrix}; \begin{pmatrix} 1 & 0.8 \\ 0.8 & 1 \end{pmatrix}\right)$

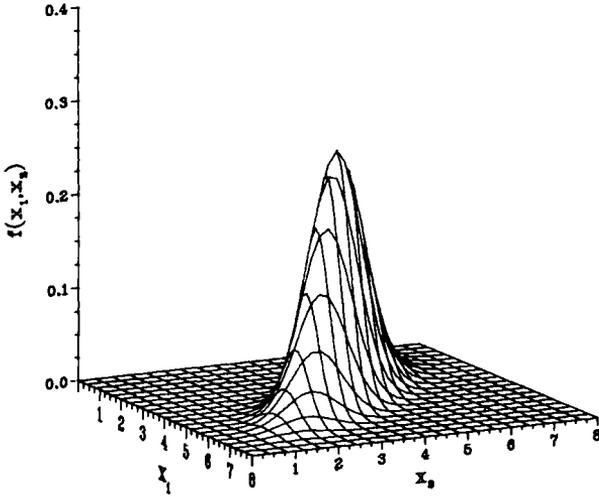


Abb.9: Dichte der bivariaten Normalverteilung $N\left(\begin{pmatrix} 4 \\ 4 \end{pmatrix}; \begin{pmatrix} 1 & -0.8 \\ -0.8 & 1 \end{pmatrix}\right)$

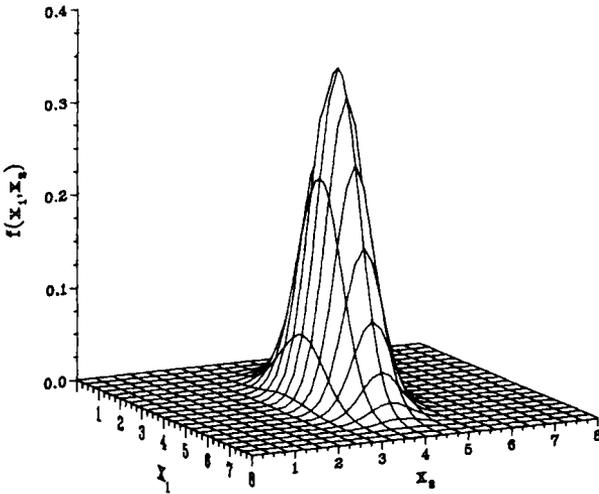


Abb.10: Dichte der bivariaten Normalverteilung $N\left(\begin{pmatrix} 4 \\ 4 \end{pmatrix}; \begin{pmatrix} 1 & 0 \\ 0 & 0.2 \end{pmatrix}\right)$

Die mehrdimensionale Normalverteilung hat einige schöne *Eigenschaften*. Gilt für einen n -dimensionalen Zufallsvektor X , daß $X \sim N(b; \Sigma_X)$, und ist B eine $m \times n$ -Matrix, c ein m -dimensionaler Vektor, so ist

$$Z = BX + c \sim N(Bb + c; B\Sigma_X B^T) \quad .$$

Insbesondere heißt dies, daß alle Randverteilungen einer mehrdimensionalen Normalverteilung wieder Normalverteilungen sind. Weiterhin weiß man, daß die Komponenten von X stochastisch unabhängig sind, wenn sie paarweise unkorreliert, d.h. $\text{Corr}(X) = I_n$, sind.

Es sei an dieser Stelle noch bemerkt, daß man auf einen normalverteilten n -dimensionalen Zufallsvektor X schließen kann, wenn für alle n -dimensionalen Vektoren d die Zufallsvariable $d^T X$ normalverteilt ist. Bei konkret vorliegenden mehrdimensionalen Daten (Beobachtungsvektoren) läßt sich mittels eines sogenannten Q - Q -Plots, vgl. Kap. IX, überprüfen, ob diese einer mehrdimensionalen Normalverteilung entstammen.

Kommen wir nun zur *multivariaten Normalverteilung*. Eine zufällige $n \times p$ -Matrix Y heißt multivariat normalverteilt mit $n \times p$ -Erwartungswertmatrix A und $n \times n$ -Kovarianzmatrix Σ , kurz $Y \sim N(A; \Sigma)$, falls mit dem im Abschnitt 4 eingeführten vec -Operator gilt:

$$\text{vec } Y \sim N(\text{vec } A; \Sigma) \quad ,$$

d.h. $\text{vec } A$ ist der Erwartungswert und Σ die Kovarianzmatrix von $\text{vec } Y$.

Eine in engem Zusammenhang zur mehrdimensionalen Normalverteilung stehende Verteilung ist die *Wishartverteilung*; sie ist eine Verallgemeinerung der in Abschnitt 3 eingeführten χ^2 -Verteilung.

Sind X_1, \dots, X_m m unabhängige n -dimensionale $N(0; \Sigma)$ -verteilte Zufallsvektoren, so besitzt, falls Σ positiv definit ist,

$$Y = X_1 X_1^T + \dots + X_m X_m^T$$

eine (zentrale) Wishartverteilung $W_m(\Sigma)$ mit m Freiheitsgraden und Parameter Σ . Der Erwartungswert von Y ist gerade $E(Y) = m\Sigma$.

6. DATENMATRIX UND DISTANZMATRIX

Multivariate statistische Verfahren werden ausgehend von einer Datenmatrix Y oder einer Distanzmatrix D verwandt. Daher wollen wir uns zunächst mit die-