

Springer-Lehrbuch

Springer-Verlag Berlin Heidelberg GmbH

Ludwig Fahrmeir · Rita Künstler
Iris Pigeot · Gerhard Tutz

Statistik

Der Weg zur Datenanalyse

Zweite, verbesserte Auflage

Mit 165 Abbildungen
und 34 Tabellen



Springer

Prof. Dr. Ludwig Fahrmeir
Dr. Rita Künstler
Prof. Dr. Iris Pigeot
Universität München
Institut für Statistik
Ludwigstraße 33
D-80539 München
Deutschland

Prof. Dr. Gerhard Tutz
Universität München
Institut für Statistik
Akademiestraße 1
D-80799 München
Deutschland

ISBN 978-3-540-65053-9 ISBN 978-3-662-10033-2 (eBook)
DOI 10.1007/978-3-662-10033-2

Die Deutsche Bibliothek – CIP-Einheitsaufnahme
Statistik: der Weg zur Datenanalyse / von Ludwig Fahrmeir ... – 2., verb.
Aufl. – Berlin; Heidelberg; New York; Barcelona; Budapest; Hongkong;
London; Mailand; Paris; Singapur; Tokio: Springer, 1999
(Springer-Lehrbuch)

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

© Springer-Verlag Berlin Heidelberg 1997, 1999

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

SPIN 10695093 42/2202-5 4 3 2 1 0 – Gedruckt auf säurefreiem Papier

Vorwort

Statistische Verfahren werden stets dann benötigt und eingesetzt, wenn im Rahmen empirischer Fragestellungen Daten erhoben, dargestellt und analysiert werden sollen. Dabei spiegelt die große Vielfalt statistischer Methoden die Breite praktischer Fragestellungen etwa aus den Wirtschafts- und Sozialwissenschaften, der Medizin und den Natur- und Ingenieurwissenschaften wider. Diese Verknüpfung der Statistik mit ihren Anwendungsdisziplinen zeigt sich unter anderem darin, daß die Vermittlung der Grundlagen statistischer Methodik integrierter Bestandteil vieler Studiengänge ist.

Dieses Buch wendet sich vorwiegend an Studierende der Wirtschafts- und Sozialwissenschaften, aber auch anderer Disziplinen wie Informatik oder Biometrie. Es ist zudem als Einführungstext für Studierende der Statistik geeignet.

Ausgehend von einer Reihe realer Fragestellungen, die typische Anwendungssituationen statistischer Verfahren veranschaulichen, werden in Kapitel 1 die Grundbegriffe der Datenanalyse und -erhebung dargestellt. Diese realen Daten werden – ergänzt durch weitere Beispiele – in den einzelnen Kapiteln erneut aufgegriffen, um die Auseinandersetzung mit statistischen Methoden inhaltlich zu motivieren, ihre Anwendung zu illustrieren und die gewonnenen Ergebnisse zu interpretieren. Damit zielt die Darstellung der Methoden eher darauf ab, das Verständnis für die statistischen Verfahren und die dahinterstehende Denkweise zu erhöhen, als die Methoden in mathematischer Ausführlichkeit zu diskutieren. Zahlreiche graphische Darstellungen sollen zu diesem Ziel beitragen.

Das Buch bietet insgesamt eine integrierte Einführung sowohl in die deskriptive Statistik und in moderne Methoden der explorativen Datenanalyse (Kapitel 2, 3) als auch in die induktive Statistik (Kapitel 9 bis 14). Letztere beinhaltet insbesondere auch Methoden der Regressions- (Kapitel 12) und der Varianzanalyse (Kapitel 13) sowie die Analyse von Zeitreihen (Kapitel 14). Eingeschlossen ist dabei eine ausführliche Beschreibung der Grundlagen der Stochastik (Kapitel 4 bis 8), die die Basis zum Verständnis der induktiven Statistik bildet.

Zur Erhöhung der Übersichtlichkeit und der Lesbarkeit werden die wesentlichen Aspekte der einzelnen Kapitel im Text und durch Einrahmungen hervorgehoben. Stichwörter am Rand weisen auf die jeweils behandelten Aspekte hin. Einige Abschnitte sind mit einem Stern versehen. Diese behandeln spezielle Fragestellungen, die für das weitere Verständnis nicht erforderlich sind und gegebenenfalls übersprungen werden können.

Am Ende eines jeden Kapitels werden zunächst die wichtigsten Aussagen dieses Kapitels noch einmal zusammengefaßt und Hinweise auf weiterführende Literatur gegeben. Abschließend dienen einige Aufgaben zur Vertiefung des jeweiligen Stoffes. Auf eine große Anzahl von Aufgaben und die Angabe von Lösungen wurde verzichtet, da uns die Bereitstellung einer eigenen, auf dieses Lehrbuch abgestimmten Aufgabensammlung zweckmäßiger erscheint.

Unser besonderer Dank gilt Norbert Behrens und Thomas Fürniss, die mit außergewöhnlichem Einsatz den größten Teil des L^AT_EX-Manuskripts erstellt haben. Bedanken möchten wir uns auch bei Thomas Billenkamp, Stefan Lang, Lisa Pritscher, Evi Rainer, Andrea Schöpp und

Kurt Watzka für ihre Beiträge zu Beispielen, Grafiken, Aufgaben und Tabellen. Für Anregungen und Korrekturvorschläge sei zudem Artur Klinger, Leo Knorr-Held, Helmut Küchenhoff, Joachim Kunert, Nanny Wermuth und unseren Studenten gedankt sowie Joachim Hartung für das Überlassen der Tabellen zu den Wilcoxon-Tests. Schließlich gilt unser Dank dem Springer-Verlag für die stets gute Zusammenarbeit und für die Umsetzung all unserer Wünsche und besonders Herrn Werner Müller, der die Erstellung dieses Lehrbuches angeregt und durch seine Unterstützung erst ermöglicht hat.

München und Berlin
im Juni 1997

Ludwig Fahrmeir
Rita Künstler
Iris Pigeot
Gerhard Tutz

Bei der vorliegenden Auflage handelt es sich um eine durchgesehene und korrigierte Version der Erstauflage des Buches. Wir bedanken uns bei allen Kollegen, Freunden, Mitarbeitern und Studenten für Hinweise auf Fehler und für Verbesserungsvorschläge. Zu diesem Buch gibt es eine Homepage:

<http://www.stat.uni-muenchen.de/~fahrmeir/buchstat/>

Die Homepage enthält insbesondere einen Großteil der verwendeten Daten.

München
im Juli 1998

Ludwig Fahrmeir
Rita Künstler
Iris Pigeot
Gerhard Tutz

Inhaltsverzeichnis

Vorwort

1 Einführung

- 1.1 Wo braucht man Statistik?
- 1.2 Was macht man mit Statistik?
- 1.3 Was steht am Anfang?
 - 1.3.1 Statistische Einheiten, Merkmale und Gesamtheiten
 - 1.3.2 Merkmalstypen
- 1.4 Wie gewinnt man Daten?
 - 1.4.1 Elemente der Versuchsplanung
 - 1.4.2 Datengewinnung und Erhebungsarten
 - Einfache Zufallsstichproben
 - Geschichtete Zufallsstichproben
 - Klumpenstichprobe
 - Mehrstufige Auswahlverfahren
 - Bewußte Auswahlverfahren
- 1.5 Zusammenfassung und Bemerkungen
- 1.6 Aufgaben

2 Univariate Deskription und Exploration von Daten

- 2.1 Verteilungen und ihre Darstellungen
 - 2.1.1 Häufigkeiten
 - 2.1.2 Graphische Darstellungen
 - Stab- und Kreisdiagramme
 - Stamm-Blatt-Diagramme
 - Histogramme
 - Unimodale und multimodale Verteilungen
 - Symmetrie und Schiefe
 - 2.1.3 Kumulierte Häufigkeitsverteilung und empirische Verteilungsfunktion

2.2	Beschreibung von Verteilungen	51
2.2.1	Lagemaße	51
	Arithmetisches Mittel	51
	Median	53
	Modus	55
	Berechnung der Lagemaße bei gruppierten Daten	56
	Lageregeln	58
	Das geometrische Mittel	59
	Das harmonische Mittel	61
	Das getrimmte Mittel	62
2.2.2	Quantile und Box-Plot	62
2.2.3	Standardabweichung, Varianz und Variationskoeffizient	67
2.2.4	Maßzahlen für Schiefe und Wölbung	72
2.3	Konzentrationsmaße	74
2.3.1	Relative Konzentration: Lorenzkurve und Gini-Koeffizient	75
	Lorenzkurve aus den geordneten Daten	75
	Lorenzkurve bei gruppierten Daten	78
	Gini-Koeffizient	80
2.3.2	Alternative Konzentrationsmaße	82
	Konzentrationsrate CR_g	82
	Herfindahl-Index	83
2.4	Dichtekurven und Normalverteilung	84
2.4.1	Dichtekurven	84
2.4.2	Normalverteilungen	89
	*Normal-Quantil-Plots	93
*2.4.3	Approximation von Dichtekurven	98
2.5	Zusammenfassung und Bemerkungen	103
2.6	Aufgaben	104
3	Multivariate Deskription und Exploration	109
3.1	Diskrete und gruppierte Merkmale	109
3.1.1	Zweidimensionale Daten: Die Kontingenztabelle	109
3.1.2	Bedingte Häufigkeiten	115
3.2	Zusammenhangsanalyse in Kontingenztabelle	119
3.2.1	Chancen und relative Chancen	119
3.2.2	Kontingenz- und χ^2 -Koeffizient	122
3.3	Graphische Darstellungen quantitativer Merkmale	127
3.3.1	Streudiagramm	128
3.3.2	Zweidimensionale Histogramme und Dichten	130
3.3.3	Mehrdimensionale Darstellungen	132

3.4	Zusammenhangsmaße bei metrischen Merkmalen	135
3.4.1	Empirischer Korrelationskoeffizient nach Bravais-Pearson	135
3.4.2	Spearman's Korrelationskoeffizient	141
3.4.3	Invarianzeigenschaften	146
3.5	Korrelation und Kausalität	147
3.6	Regression	152
3.6.1	Das lineare Regressionsmodell	152
3.6.2	Die Berechnung der Ausgleichsgeraden	153
3.6.3	Bestimmtheitsmaß und Residualanalyse	158
*3.6.4	Nichtlineare Regression	165
3.7	Zusammenfassung und Bemerkungen	167
3.8	Aufgaben	169
4	Wahrscheinlichkeitsrechnung	171
4.1	Definition und Begriff der Wahrscheinlichkeit	172
4.1.1	Mengen und Mengenoperationen	173
4.1.2	Zufallsereignisse	177
4.1.3	Wahrscheinlichkeiten	179
4.2	Zur empirischen Interpretation von Wahrscheinlichkeiten	185
4.2.1	Die Laplace-Wahrscheinlichkeit	186
4.2.2	Objektive Wahrscheinlichkeiten als Grenzwert relativer Häufigkeiten	189
4.2.3	Subjektive Wahrscheinlichkeiten	191
4.3	Zufallsstichproben und Kombinatorik	192
4.3.1	Modell mit Zurücklegen	193
4.3.2	Modell ohne Zurücklegen	194
4.3.3	Permutationen	195
4.3.4	Modell ohne Zurücklegen und ohne Berücksichtigung der Reihenfolge	196
4.4	Bedingte Wahrscheinlichkeiten	199
4.5	Unabhängigkeit von zwei Ereignissen	203
4.6	Totale Wahrscheinlichkeit	206
4.7	Der Satz von Bayes	209
4.8	Unendliche Grundgesamtheiten	213
4.9	Zusammenfassung und Bemerkungen	215
4.10	Aufgaben	217
5	Diskrete Zufallsvariablen	221
5.1	Zufallsvariablen	221
5.2	Verteilungen und Parameter von diskreten Zufallsvariablen	225
5.2.1	Definition und Verteilung	225
5.2.2	Unabhängigkeit von diskreten Zufallsvariablen	236

5.2.3	Lageparameter, Quantile und Streuungsparameter einer diskreten Verteilung	239
5.3	Spezielle diskrete Verteilungsmodelle	250
5.3.1	Die Binomialverteilung	250
5.3.2	Die hypergeometrische Verteilung	256
5.3.3	Die Poisson-Verteilung	258
5.4	Zusammenfassung und Bemerkungen	262
5.5	Aufgaben	264
6	Stetige Zufallsvariablen	267
6.1	Definition und Verteilung	267
6.2	Lageparameter, Quantile und Varianz von stetigen Zufallsvariablen	279
6.3	Spezielle stetige Verteilungsmodelle	289
6.3.1	Die Normalverteilung	289
6.3.2	Die logarithmische Normalverteilung	297
6.3.3	Chi-Quadrat-, Student- und Fisher-Verteilung	298
6.4	Zusammenfassung und Bemerkungen	302
6.5	Aufgaben	304
7	Mehr über Zufallsvariablen und Verteilungen	307
7.1	Gesetz der großen Zahlen und Grenzwertsätze	307
7.1.1	Das Gesetz der großen Zahlen und der Hauptsatz der Statistik	309
7.1.2	Der zentrale Grenzwertsatz	312
7.2	Approximation von Verteilungen	315
*7.3	Zufallszahlen und Simulation	318
*7.4	Einige Ergänzungen	321
7.4.1	Zufallsvariablen als Abbildungen	321
7.4.2	Verteilungsfunktion und ihre Eigenschaften	323
7.4.3	Ungleichung von Tschebyscheff	325
7.4.4	Maßzahlen für Schiefe und Wölbung	327
7.5	Zusammenfassung und Bemerkungen	328
7.6	Aufgaben	329
8	Mehrdimensionale Zufallsvariablen	331
8.1	Begriff mehrdimensionaler Zufallsvariablen	331
8.2	Zweidimensionale diskrete Zufallsvariablen	334
8.3	Zweidimensionale stetige Zufallsvariablen	339
8.4	Unabhängigkeit von Zufallsvariablen	342
8.5	Kovarianz und Korrelation	345
8.6	Die zweidimensionale Normalverteilung	353

8.7	Zusammenfassung und Bemerkungen	357
8.8	Aufgaben	357
9	Parameterschätzung	359
9.1	Punktschätzung	360
9.2	Eigenschaften von Schätzstatistiken	362
9.2.1	Erwartungstreue	362
9.2.2	Erwartete mittlere quadratische Abweichung und Konsistenz	366
9.2.3	Wirksamste Schätzstatistiken	369
9.3	Konstruktion von Schätzfunktionen	371
9.3.1	Maximum Likelihood-Schätzung	372
9.3.2	Kleinste-Quadrate-Schätzung	375
9.4	Intervallschätzung	376
9.4.1	Konfidenzintervalle für Erwartungswert und Varianz	378
9.4.2	Konfidenzintervalle für den Anteilswert	382
9.5	Zusammenfassung und Bemerkungen	383
9.6	Aufgaben	384
10	Testen von Hypothesen	387
10.1	Der Binomial- und der Gauß-Test	387
10.1.1	Der exakte Binomialtest	391
10.1.2	Der approximative Binomialtest	394
10.1.3	Der Gauß-Test	397
10.2	Prinzipien des Testens	401
	Fehlentscheidungen	404
	Statistische Tests und Konfidenzintervalle	407
	Überschreitungswahrscheinlichkeit	408
	Gütefunktion	409
	*Multiple Testprobleme	416
10.3	Zusammenfassung und Bemerkungen	417
10.4	Aufgaben	419
11	Spezielle Testprobleme	421
11.1	Ein-Stichproben-Fall	422
11.1.1	Tests zu Lagealternativen	423
11.1.2	Anpassungstests	432
11.2	Vergleiche aus unabhängigen Stichproben	441
11.2.1	Tests zu Lagealternativen	442
11.2.2	χ^2 -Homogenitätstest	448
11.3	Vergleiche aus verbundenen Stichproben	450

11.4	Zusammenhangsanalyse	452
11.4.1	χ^2 -Unabhängigkeitstest	452
11.4.2	Korrelation bei metrischen Merkmalen	454
11.5	Zusammenfassung und Bemerkungen	456
11.6	Aufgaben	457
12	Regressionsanalyse	459
12.1	Lineare Einfachregression	460
12.1.1	Das Modell der linearen Einfachregression	460
12.1.2	Schätzen, Testen und Prognose	463
12.1.3	Residualanalyse	473
12.2	Multiple lineare Regression	476
12.2.1	Das multiple lineare Regressionsmodell	478
12.2.2	Schätzen, Testen und Prognose	479
*12.2.3	Multiple lineare Regression in Matrixnotation	488
*12.3	Nichtlineare und nichtparametrische Regression	490
12.4	Zusammenfassung und Bemerkungen	494
12.5	Aufgaben	495
13	Varianzanalyse	497
13.1	Einfaktorielle Varianzanalyse	498
13.2	Zweifaktorielle Varianzanalyse mit festen Effekten	507
13.3	Zusammenfassung und Bemerkungen	521
13.4	Aufgaben	522
14	Zeitreihen	525
14.1	Indizes	528
14.2	Komponentenmodelle	532
14.3	Globale Regressionsansätze	534
14.3.1	Trendbestimmung	534
14.3.2	Bestimmung der Saisonkomponente	536
14.4	Lokale Ansätze	537
14.4.1	Trendbestimmung	538
	Gleitende Durchschnitte	538
	Lokale Regression	539
	*Spline-Glättung	542
14.4.2	Bestimmung der Saisonkomponente	544
	Gleitende Durchschnitte und lokale Regression	544
	*Spline-Glättung	546
14.5	Zusammenfassung und Bemerkungen	547

14.6 Aufgaben	547
Tabellen	551
A Standardnormalverteilung	551
B Binomialverteilung	552
C χ^2 -Verteilung	567
D Students t -Verteilung	568
E F -Verteilung	569
F Wilcoxon-Vorzeichen-Rang-Test	578
G Wilcoxon-Rangsummen-Test	578
Literatur	581
Sachverzeichnis	585