

Werner A. Stahel

Statistische Datenanalyse

Aus dem Programm
Mathematik/Statistik

Stochastik für Einsteiger

von N. Henze

**Einführung in die Wahrscheinlichkeitstheorie
und Statistik**

von U. Krengel

Stochastik

von G. Hübner

Stochastik mit Mathematica

von M. Overbeck-Larisch und W. Dolejsky

**Elementare Einführung in die
Wahrscheinlichkeitsrechnung**

von K. Bosch

**Elementare Einführung in die
angewandte Statistik**

von K. Bosch

Statistische Datenanalyse

von W. A. Stahel

vieweg

Werner A. Stahel

Statistische Datenanalyse

**Eine Einführung
für Naturwissenschaftler**

4., verbesserte Auflage



Bibliografische Information Der Deutschen Bibliothek

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

Dr. Werner Stahel
ETH Zürich
Seminar für Statistik
ETH Zentrum, SOL F 10
CH-8092 Zürich

E-Mail: stahel@stat.math.ethz.ch

1. Auflage 1995
- 2., überarbeitete Auflage 1999
- 3., durchgesehene Auflage 2000
- 4., verbesserte Auflage November 2002

Alle Rechte vorbehalten

© Friedr. Vieweg & Sohn Verlagsgesellschaft mbH, Braunschweig / Wiesbaden, 2002

Der Verlag Vieweg ist ein Unternehmen der Fachverlagsgruppe BertelsmannSpringer.
www.vieweg.de



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

Umschlaggestaltung: Ulrike Weigel, www.CorporateDesignGroup.de

Gedruckt auf säurefreiem und chlorfrei gebleichtem Papier.

ISBN 978-3-528-36653-7 ISBN 978-3-322-96962-0 (eBook)
DOI 10.1007/978-3-322-96962-0

Vorwort zur ersten Auflage

Statistik ist in vielen Gebieten der Wissenschaften, der Technik und des täglichen Lebens ein wichtiges Hilfsmittel. In den meisten höheren Ausbildungen werden deshalb einige Grundbegriffe der Statistik behandelt – mit unterschiedlichen Ansprüchen an ein tieferes Verständnis.

Entsprechend vielfältig ist das Angebot an Büchern, die in die Statistik einführen. Wieso also noch ein weiteres? Es war meine Absicht, auf ein intuitives Verständnis hinzuarbeiten, das je nach Begabungen und Ansprüchen auch durch theoretisches Wissen unterstützt wird. Ich habe mich deshalb bemüht, Begriffe immer anhand von Beispielen einzuführen und für übliche Stolpersteine Umgehungsmöglichkeiten zu zeigen, die wichtigen Zusammenhänge aber zusätzlich auch mathematisch darzustellen.

Was dieses Buch soll, steht für Leserinnen und Leser, die in dieses Gebiet einsteigen wollen oder sollen, in der Einleitung. Solche, die Statistik bereits kennen und sich fragen, was sie hier Besonderes finden können, verweise ich auf das Nachwort.

Dem geneigten Leser und der aufmerksamen Leserin wird nicht verborgen bleiben, dass der Autor aus der Schweiz stammt. Dies zeigt sich nur schon daran, dass das scharfe β durch ein Doppel-s ersetzt ist. Da das Manuskript druckreif in der Schweiz erstellt wurde, hoffe ich auf Toleranz bezüglich dieses „kulturellen“ Unterschiedes. Ebenso erlaubte ich mir, das Dezimal-Komma durch den in der übrigen Welt üblichen Punkt zu ersetzen.

Der vorliegende Text wurde aus einem Skript zu einer Einführungs-Vorlesung und einer daran anschliessenden Überblicks-Vorlesung über „Statistische Methoden“ entwickelt. Die ursprüngliche Vorlage für Teile des Skripts stammt von Prof. Frank Hampel, der dadurch Grundlegendes zu diesem Buch beigetragen hat. An einigen Stellen tauchen noch echte „Findlinge“ aus jenem Skript auf.

Die auffälligsten „Zutaten“, die Karikaturen, stammen von dipl. math. Hanspeter Endres.

Viele Studierende, Assistentinnen, Assistenten und Dozenten haben mich immer wieder auf Fehler und auf Verbesserungsmöglichkeiten aufmerksam gemacht. Einige möchte ich namentlich erwähnen: Prof. Frank Hampel, Prof. Hans-Rudolf Künsch, Dr. Robert Aebi, Prof. Paul Embrechts, Dr. Helmut Glemser, Dr. Beat Hulliger, Dr. Markus Hürzeler, Otto Kaufmann, Dr. Martin Mächler, Dr. Marianne Müller, Prisca Risold-Durrer, Dr. Hans-Rudolf Roth, Dr. Andreas Ruckstuhl, Caterina Savi. An der Schreibearbeit waren Frau Margrit Karrer und Frau Christina Künzli beteiligt. Ihnen allen möchte ich für ihre wertvollen Beiträge bestens danken.

Zürich, 10. Januar 1995

Werner Stahel

Vorwort zur zweiten Auflage

Die zweite Auflage bringt gegenüber der ersten einige Verbesserungen und kleinere Ergänzungen, die sich aus Anregungen aus der Leserschaft, von Kolleginnen und Kollegen und aus Erfahrungen im Unterricht herleiten:

- Dass die Motivation in einer Einführungsvorlesung in die Statistik aufrechtzuerhalten nicht einfach ist, belegt die Erfahrung vieler Dozentinnen und Dozenten. Ein alternativer Aufbau, der in 1.3.c erläutert wird, kann dabei helfen. Damit er ohne Schwierigkeiten möglich wird, waren Umstellungen in Kapitel 4 nötig.
- Einige Ergänzungen wurden angefügt, besonders in den Ausblicks-Abschnitten der Regression und in der Multivariaten Statistik.
- In einem Anhang wurde, entsprechend einem häufig geäußerten Wunsch, der wichtigste Stoff der Grundlagenteile zusammengefasst.
- Einige weitere Beispiele und verbesserte Formulierungen wurden aufgrund von Anregungen und didaktischen Überlegungen angebracht.
- Der ganze Text wurde durchgesehen und an zahlreichen Orten stilistisch verbessert. Die neue deutsche Rechtschreibung wurde eingearbeitet, soweit dies ohne entsprechendes Korrekturprogramm mit begrenztem Aufwand möglich war.
- Es waren leider recht viele Druckfehler zu korrigieren. Sicher werden sich noch einige weitere zeigen. Sie werden auf meiner Internet-Seite

www.stat.math.ethz.ch/~stahel/publications

nachgetragen.

Für alle Anregungen bedanke ich mich herzlich. Ich hoffe, dass der Text durch sie noch um einiges brauchbarer geworden ist, und wünsche den Leserinnen und Lesern eine angenehme und lehrreiche Lektüre.

Zürich, 11. Januar 1999

Werner Stahel

In der **dritten und vierten Auflage** wurden Druckfehler korrigiert und einige kleinere Änderungen eingearbeitet.

Zürich, 11. Februar 2000 und 11. Oktober 2002

Werner Stahel

*Meiner Familie und
meinen Kolleginnen und Kollegen
vom Seminar für Statistik
gewidmet*

Inhaltsverzeichnis

1	Einleitung	1
1.1	Was ist Statistische Datenanalyse?	1
1.2	Ziele	6
1.3	Hinweise	6
	Literatur zur angewandten Statistik	8
I	Beschreibende Statistik	9
2	Beschreibung eindimensionaler Stichproben	11
2.1	Einfache grafische Darstellungen	11
2.2	Einige Bezeichnungen und Begriffe	14
2.3	Kennzahlen für eine quantitative Stichprobe	16
2.4	Klassierte Daten	21
2.5	Mehrere Stichproben	24
2.6	Transformationen von Beobachtungen	27
2.7	Wertebereiche, Datensorten	30
2.8	* Transformationen und Unterschiede zwischen Beobachtungen	32
3	Beschreibende Statistik mehrdimensionaler Daten	33
3.1	Grafische Darstellungen für zwei zusammenhängende Grössen	33
3.2	Die Produktmomenten-Korrelation	36
3.3	Rangkorrelationen	40
3.4	Zur Interpretation von Korrelationen	42
3.5	Regression	43
3.6	Multivariate Beobachtungen	47
3.7	Zeitreihen und räumliche Daten	52
3.8	Allgemeines zu grafischen Darstellungen	56
3.9	Wie weiter?	57
	Literatur zu Teil I	57
II	Wahrscheinlichkeitsrechnung	59
4	Wahrscheinlichkeit	61
4.1	Einleitung	61
4.2	Grundbegriffe und Grundeigenschaften	62
4.3	Zufallsvariable	67
4.4	Zufallszahlen	71
4.5	Zwei Zufallsvariable, gemeinsame Verteilung	74
4.6	Unabhängige Ereignisse und Zufallsvariable	76
4.7	Bedingte Wahrscheinlichkeit	78
4.8	Bedingte Verteilung	84
4.9	Der Satz von Bayes	86
4.10	* Was ist eine Wahrscheinlichkeit?	88
4.11	Wie weiter?	90

5	Diskrete Verteilungen	92
5.1	Bernoulli- und Binomial-Verteilung	92
5.2	Die Poisson-Verteilung	96
5.3	Kennzahlen	99
5.4	Verteilungsfamilien	103
5.5	Die multinomiale Verteilung	105
5.6	Summen von Zufallsvariablen	107
5.7	Zufalls-Stichproben	111
5.8	* Gesetze der grossen Zahl	114
5.9	* Stochastische Prozesse	118
6	Stetige Verteilungen	120
6.1	Grundlagen	120
6.2	Grundbegriffe, Exponential- und uniforme Verteilung	121
6.3	Kennzahlen für stetige Verteilungen	126
6.4	Transformationen von Zufallsvariablen	127
6.5	Die Normalverteilung	132
6.6	* Weitere stetige Verteilungsfamilien	136
6.7	Gemeinsame und bedingte Verteilung	139
6.8	Unabhängige Zufallsvariable und Korrelation	142
6.9	Funktionen von mehreren Zufallsvariablen	144
6.10	Gaussche Fehler-Fortpflanzung	149
6.11	Der Zentrale Grenzwertsatz	152
6.12	Rückblick	155
	Literatur zu Teil II	156
III	Schliessende Statistik	157
7	Schätzungen	159
7.1	Drei Grundfragen der schliessenden Statistik	159
7.2	Schätzungen für B , \mathcal{P} und \mathcal{N}	161
7.3	Eigenschaften von Schätzungen	162
7.4	Die Maximum-Likelihood-Methode	167
7.5	Robuste Schätzungen	171
8	Tests	174
8.1	Einführende Beispiele und Begriffe	174
8.2	Test für eine Wahrscheinlichkeit	178
8.3	Die Teststatistik	181
8.4	Vorgehen bei einem statistischen Test	183
8.5	Tests für eine Stichprobe oder zwei gepaarte Stichproben	185
8.6	Interpretation von Testergebnissen	191
8.7	Bemerkungen zum P-Wert	193
8.8	Vergleich von zwei quantitativen Stichproben	195
8.9	Macht	200
8.10	* Asymptotische Tests und Randomisierungs-Tests	202
8.11	Sinn und Unsinn statistischer Tests	204

9	Vertrauensintervalle	205
9.1	Vertrauensintervalle für Binomial- und Poisson-Verteilung	205
9.2	Eigenschaften von Vertrauensintervallen	207
9.3	Vertrauensintervalle für Lageparameter	209
9.4	Bootstrap und andere Resampling-Methoden	211
9.5	Vertrauens- und andere Intervalle	214
9.6	Schätzungen, Tests und Vertrauensintervalle im Vergleich	216
9.7	Wo stehen wir?	217
	Literatur zu Teil III	217
IV	Methoden der Datenanalyse	219
10	Nominale Daten	221
10.1	Multinomiale Verteilung und Chiquadrat-Test	221
10.2	Der Chiquadrat-Anpassungstest	226
10.3	Der Chiquadrat-Test in Kontingenztafeln	230
10.4	Die häufigsten Fehler beim Chiquadrat-Test	233
11	Überprüfung von Voraussetzungen	235
11.1	Problemstellung	235
11.2	Quantil-Quantil-Diagramme	236
11.3	Anpassungstests	239
11.4	Bedeutung von Tests zur Prüfung von Voraussetzungen	240
11.5	Unabhängigkeit	241
12	Varianzanalyse	243
12.1	Vergleich mehrerer Stichproben, einfache Varianzanalyse	243
12.2	Multiple Vergleiche, multiple Tests	247
12.3	Mehrere verbundene Stichproben	249
12.4	Zweiweg-Varianzanalyse	251
12.5	Zufällige Effekte, Varianz-Komponenten	256
12.6	Ausblick	258
	Literatur	259
13	Regression	260
13.1	Das Modell der einfachen linearen Regression	260
13.2	Schätzung der Parameter	261
13.3	Tests und Vertrauensintervalle für die Parameter	264
13.4	Vertrauens- und Vorhersage-Bereiche	266
13.5	Multiple lineare Regression	268
13.6	Vielfalt der Modelle der multiplen linearen Regression	272
13.7	Residuen-Analyse	275
13.8	Einflussreiche Beobachtungen	281
13.9	Modellwahl	283
13.10	Allgemeinere Modelle für stetige Zielgrößen	287
13.11	Verallgemeinerte lineare Modelle	292
	Literatur	294

14 Versuchsplanung	295
14.1 Einleitung	295
14.2 Allgemeine Überlegungen	295
14.3 Versuchspläne	299
14.4 Eine Checkliste	301
Literatur	303
15 Multivariate Statistik	304
15.1 Mehrdimensionale Zufallsvariable	304
15.2 Schätzung von Erwartungswert und Kovarianz-Matrix	309
15.3 Die mehrdimensionale Normalverteilung	311
15.4 Statistik der Normalverteilung	313
15.5 Hauptkomponenten	315
15.6 Diskriminanz-Analyse	318
Literatur	321
16 Zeitreihen	322
16.1 Fragestellungen	322
16.2 Auto-Korrelation	323
16.3 ARMA-Modelle	324
16.4 Statistik von Zeitreihen	325
16.5 Vorhersage	325
16.6 Zustandsraum-Modelle	327
16.7 Spektralanalyse	328
16.8 Räumliche Korrelation	329
16.9 Regression mit Zeitreihen	332
Literatur	333
17 Stichproben-Erhebungen	335
17.1 Einleitung	335
17.2 Einfache Zufalls-Stichprobe	336
17.3 Geschichtete Stichproben	339
17.4 Weitere Stichproben-Pläne	340
17.5 Weitere Schätzmethoden	342
17.6 Auswertung von Umfragen	342
17.7 Eine Checkliste	344
Literatur	347
18 Ausblick	348
18.1 Was ist erreicht?	348
18.2 Grosse Datensätze, beschreibende Modelle	349
18.3 Die Statistik und ihre Anwendungen	351
A Anhang: Kurzfassung des wichtigsten Stoffes	354
A.1 Beschreibende Statistik	354
A.2 Wahrscheinlichkeitsrechnung	355
A.3 Schliessende Statistik	360
Literaturverzeichnis	363
Sachwortverzeichnis	367