

Herausgeber:

Prof. Dr. Holger Dette • Prof. Dr. Wolfgang Härdle

Statistik und ihre Anwendungen

- Azizi Ghanbari, S.
Einführung in die Statistik für Sozial- und Erziehungswissenschaftler 2002
- Brunner, E.; Munzel, U.
Nichtparametrische Datenanalyse 2003
- Dehling, H.; Haupt, B.
Einführung in die Wahrscheinlichkeitstheorie und Statistik 2. Auflage 2004
- Dümbgen, L.
Stochastik für Informatiker 2003
- Falk, M.; Becker, R.; Marohn, F.
Angewandte Statistik 2004
- Franke, J.; Härdle, W.; Hafner, C.
Statistik der Finanzmärkte 2. Auflage 2004
- Greiner, M.
Serodiagnostische Tests 2003
- Handl, A.
Multivariate Analysemethoden 2003
- Hilgers, R.-D.; Bauer, R.; Scheiber, V.
Einführung in die Medizinische Statistik 2003
- Kohn, W.
Statistik Datenanalyse und Wahrscheinlichkeitsrechnung 2005
- Ligges, U.
Programmieren mit R 2005
- Meintrup, D.; Schäffer, S.
Stochastik Theorie und Anwendungen 2005
- Plachky, D.
Mathematische Grundbegriffe der Stochastik 2002
- Schumacher, M.; Schulgen, G.
Methodik klinischer Versuche 2002
- Steland, A.
Mathematische Grundlagen der empirischen Forschung 2004

Wolfgang Kohn

Statistik

Datenanalyse und
Wahrscheinlichkeitsrechnung

 Springer

Wolfgang Kohn
Fachhochschule Bielefeld
Fachbereich Wirtschaft
Universitätsstraße 25
33615 Bielefeld
e-mail: wolfgang.kohn@fh-bielefeld.de

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.

Mathematics Subject Classification (2000): 62H17, 62H20, 62H25, 62H30, 62G10, 62J05

ISBN 3-540-21677-4 Springer Berlin Heidelberg New York

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

Springer ist ein Unternehmen von Springer Science+Business Media

springer.de

© Springer-Verlag Berlin Heidelberg 2005
Printed in Germany

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

Einbandgestaltung: *design & production*, Heidelberg
Datenerstellung durch den Autor unter Verwendung eines Springer \LaTeX -Makropakets
Herstellung: LE- \TeX Jelonek, Schmidt & Vöckler GbR, Leipzig
Gedruckt auf säurefreiem Papier 40/3142YL - 5 4 3 2 1 0

Für Coco, Hannah, Sophia, Leonhard
und meine Eltern

Vorwort

Schriftlichkeit ist nichts anderes als eine Gedächtnishilfe, die nicht von sich allein Weisheit vermitteln kann. Sokrates

Die Statistik ist ein enorm umfangreiches Wissensgebiet. Jede Wissenschaft entwickelt nicht nur eigene statistische Verfahren, sondern es sind auch viele Kenntnisse aus verschiedenen Wissensgebieten erforderlich, um diese Verfahren, ihre Anwendung und ihre Ergebnisse zu verstehen. So könnten bei einer Datenerhebung per Fragebogen oder Interview psychologische Kenntnisse benötigt werden. Für die Datenverarbeitung und Berechnung der statistischen Verfahren werden in der Regel mathematische und informatische Kenntnisse eingesetzt. Die Interpretation der Ergebnisse setzt wiederum sehr gute Fachkenntnisse des entsprechenden Wissensgebiets voraus. Von daher ist es schwierig, in einer Darstellung alle Aspekte gleichermaßen zu berücksichtigen. Ich habe mich bemüht, die Eigenschaften und Zusammenhänge statistischer Verfahren aufzuzeigen und ihre Anwendung durch zahlreiche Beispiele und Abbildungen verständlich zu machen.

Die Auswahl der statistischen Methoden von der Erhebung über beschreibende Techniken (Teil I) hin zur schließenden Analyse (Teil II) einschließlich einiger multivariater Verfahren (Teil III) erfolgte unter wirtschaftswissenschaftlicher Sicht. In einem Buch zur Einführung in die Statistik können außerdem nur die grundlegenden statistischen Verfahren berücksichtigt werden. Aber auch unter diesen ist nur eine Auswahl erfolgt. Dennoch sind über 500 Seiten zustande gekommen und es gäbe zu dieser Auswahl noch mehr zu schreiben!

Während der langen Zeit an der Arbeit an diesem Buch ist es mir immer wieder schwergefallen, bestimmte Themen und Aspekte nicht mehr aufzunehmen. Erst die Zeit wird zeigen, ob ich an den richtigen Stellen die Darstellung beendet habe.

Ich bedanke mich sehr herzlich bei Prof. Dr. Peter Naeve von der Universität Bielefeld für seine große Diskussionsbereitschaft und hilfreiche Kritik.

Bielefeld, im Mai 2004

Wolfgang Kohn

Inhaltsverzeichnis

1	Einleitung	1
----------	-------------------	----------

Teil I Deskriptive Statistik

2	Grundlagen	7
2.1	Einführung	7
2.2	Grundbegriffe der deskriptiven Statistik	8
2.3	Messbarkeitseigenschaften von Merkmalen	13
2.3.1	Nominalskala	13
2.3.2	Ordinalskala	13
2.3.3	Kardinalskala	14
2.4	Skalentransformation	15
2.5	Häufigkeitsfunktion	16
2.6	Klassierung von metrischen Merkmalswerten	18
2.7	Übungen	21
3	Datenerhebung und Erhebungsarten	23
3.1	Datenerhebung	23
3.1.1	Fragebogenerhebung	24
3.1.2	Interview	24
3.1.3	Gestaltung von Befragungen	25
3.2	Erhebungsarten	27
3.2.1	Zufallsstichproben	28
3.2.2	Nicht zufällige Stichproben	30
3.3	Datenschutz	32
4	Eindimensionale Datenanalyse	35
4.1	Einführung	36
4.2	Qualitative Merkmale	36
4.2.1	Modus	38

4.2.2	Informationsentropie	38
4.3	Komparative Merkmale	45
4.3.1	Empirische Verteilungsfunktion	45
4.3.2	Quantile	47
4.3.3	Boxplot	49
4.3.4	Summenhäufigkeitsentropie	51
4.4	Quantitative Merkmale	54
4.4.1	Stamm-Blatt Diagramm	54
4.4.2	Histogramm	57
4.4.3	Dichtespur	59
4.4.4	Arithmetisches Mittel	65
4.4.5	Harmonisches Mittel	67
4.4.6	Geometrisches Mittel	70
4.4.7	Spannweite	72
4.4.8	Median der absoluten Abweichung vom Median	73
4.4.9	Varianz und Standardabweichung	74
4.4.10	Variationskoeffizient	80
4.4.11	Relative Konzentrationsmessung	83
4.4.12	Absolute Konzentrationsmessung	92
4.5	Übungen	98
5	Zweidimensionale Datenanalyse	101
5.1	Zweidimensionale Daten und ihre Darstellung	101
5.2	Randverteilungen und bedingte Verteilungen	105
5.2.1	Randverteilung	105
5.2.2	Bedingte Verteilung	106
5.3	Zusammenhangsmaße qualitativer Merkmale	113
5.3.1	Quadratische Kontingenz	113
5.3.2	Informationsentropie bei zweidimensionalen Häufigkeitsverteilungen	116
5.3.3	Transinformation	118
5.3.4	Vergleich von C und T	119
5.4	Zusammenhangsmaße komparativer Merkmale	121
5.4.1	Kovarianz und Korrelationskoeffizient	121
5.4.2	Rangfolge und Rangzahlen	122
5.4.3	Rangkorrelationskoeffizient	123
5.5	Zusammenhangsmaße quantitativer Merkmale	126
5.5.1	Kovarianz	126
5.5.2	Korrelationskoeffizient	127
5.6	Interpretation von Korrelation	129
5.7	Simpson Paradoxon	131
5.8	Übungen	133

6	Lineare Regression	137
6.1	Einführung	137
6.2	Lineare Regressionsfunktion	138
6.3	Methode der Kleinsten Quadrate	141
6.3.1	Normalgleichungen	142
6.3.2	Kleinst-Quadrate Schätzung	143
6.3.3	Standardisierte Regressionskoeffizienten	146
6.4	Bestimmtheitsmaß	147
6.5	Spezielle Regressionsfunktionen	150
6.5.1	Trendfunktion	150
6.5.2	Linearisierung von Funktionen	150
6.5.3	Datentransformation	151
6.6	Lineares Modell	152
6.7	Prognose	153
6.8	Übungen	159
7	Verhältnis- und Indexzahlen	161
7.1	Einführung	161
7.2	Gliederungs-, Beziehungs- und Messzahlen	163
7.3	Umbasierung und Verkettung von Messzahlen	166
7.3.1	Umbasierung	166
7.3.2	Verkettung	167
7.4	Indexzahlen	167
7.4.1	Preisindex nach Laspeyres	169
7.4.2	Basiseffekt	172
7.4.3	Preisindex nach Paasche	173
7.4.4	Mengenindizes nach Laspeyres und Paasche	174
7.4.5	Umsatzindex	176
7.4.6	Deflationierung	176
7.4.7	Verkettung von Indexzahlen	177
7.4.8	Anforderungen an einen idealen Index	179
7.4.9	Preisindex nach Fischer	179
7.4.10	Kettenindex	180
7.5	Aktienindex DAX	181
7.6	Übungen	183

Teil II Schließende Statistik

8	Kombinatorik	187
8.1	Grundbegriffe	187
8.2	Permutation	189
8.2.1	Permutation ohne Wiederholung	189
8.2.2	Permutation mit Wiederholung	189
8.3	Variation	190

8.3.1	Variation ohne Wiederholung	191
8.3.2	Variation mit Wiederholung	192
8.4	Kombination	192
8.4.1	Kombination ohne Wiederholung	192
8.4.2	Kombination mit Wiederholung	193
8.5	Übungen	195
9	Grundzüge der Wahrscheinlichkeitsrechnung	197
9.1	Einführung	197
9.2	Zufallsexperiment	198
9.3	Ereignisoperationen	199
9.4	Wahrscheinlichkeitsbegriffe	204
9.4.1	Laplacescher Wahrscheinlichkeitsbegriff	205
9.4.2	Von Misesscher Wahrscheinlichkeitsbegriff	206
9.4.3	Subjektiver Wahrscheinlichkeitsbegriff	206
9.4.4	Axiomatische Definition der Wahrscheinlichkeit	207
9.4.5	Rechenregeln	209
9.5	Bedingte Wahrscheinlichkeiten	210
9.6	Satz von Bayes	216
9.7	Unabhängige Zufallsereignisse	220
9.8	Übungen	224
10	Zufallsvariablen und Wahrscheinlichkeitsfunktion	227
10.1	Zufallsvariablen	227
10.2	Wahrscheinlichkeitsfunktion	230
10.2.1	Wahrscheinlichkeitsfunktion einer diskreten Zufallsvariablen	230
10.2.2	Wahrscheinlichkeitsfunktion einer stetigen Zufallsvariablen	232
10.3	Verteilungen von transformierten Zufallsvariablen	238
10.4	Erwartungswert	240
10.5	Modus und Quantil	244
10.6	Varianz	244
10.7	Erwartungswert und Varianz linear transformierter Zufallsvariablen	246
10.7.1	Erwartungswert linear transformierter Zufallsvariablen	246
10.7.2	Varianz linear transformierter Zufallsvariablen	247
10.8	Momente und momenterzeugende Funktion	249
10.8.1	Momente	249
10.8.2	Momenterzeugende Funktion	250
10.9	Ungleichung von Chebyschew	251
10.10	Übungen	254
11	Gemeinsame Verteilung von Zufallsvariablen	257
11.1	Einführung	257
11.2	Bedingte Verteilung	261
11.3	Bedingter Erwartungswert	262
11.4	Kovarianz	264

11.5	Übungen	268
12	Normalverteilung	271
12.1	Einführung	271
12.2	Standardnormalverteilung	276
12.3	Lognormalverteilung	281
12.4	Übungen	284
13	Bernoulli-verwandte Zufallsvariablen	285
13.1	Einführung	285
13.2	Bernoulliverteilung	285
13.3	Binomialverteilung	286
13.4	Hypergeometrische Verteilung	292
13.5	Geometrische Verteilung	297
13.6	Negative Binominalverteilung	300
13.7	Negative hypergeometrische Verteilung	303
13.8	Poissonverteilung	304
13.9	Exponentialverteilung	311
13.10	Übungen	314
14	Stichproben	315
14.1	Einführung	315
14.2	Identisch verteilte unabhängige Stichproben	316
14.3	Schwaches Gesetz der großen Zahlen	317
14.4	Starkes Gesetz der großen Zahlen	320
14.5	Hauptsatz der Statistik	321
14.6	Zentraler Grenzwertsatz	322
14.6.1	Approximation der Binomialverteilung durch die Normalverteilung	326
14.6.2	Approximation der hypergeometrischen Verteilung durch die Normalverteilung	329
14.6.3	Approximation der Poissonverteilung durch die Normalverteilung	329
14.7	Stichprobenverteilungen aus normalverteilten Grundgesamtheiten ..	330
14.7.1	χ^2 -Quadrat Verteilung	330
14.7.2	t -Verteilung	332
14.7.3	F -Verteilung	334
14.8	Hauptsatz der Stichprobentheorie	335
14.9	Stichproben aus bernoulli-, exponential- und poissonverteilten Grundgesamtheiten	338
14.9.1	Stichproben aus bernoulli- und binomialverteilten Grundgesamtheiten	338
14.9.2	Stichproben aus poissonverteilten Grundgesamtheiten	339
14.9.3	Stichproben aus exponentialverteilten Grundgesamtheiten ..	339
14.10	Übungen	341

15 Parameterschätzung	343
15.1 Einführung	343
15.2 Punktschätzung	344
15.2.1 Methode der Momente	345
15.2.2 Maximum-Likelihood Schätzung	346
15.2.3 Methode der Kleinsten Quadrate	349
15.3 Intervallschätzung	351
15.3.1 Konfidenzintervall für μ_X einer Normalverteilung bei bekanntem σ_X^2	352
15.3.2 Konfidenzintervall für μ_X einer Normalverteilung bei unbekanntem σ_X^2	353
15.3.3 Konfidenzintervall für Regressionskoeffizienten	354
15.3.4 Konfidenzintervall für eine ex post Prognose	357
15.3.5 Approximatives Konfidenzintervall für μ_X	358
15.3.6 Approximatives Konfidenzintervall für den Anteilswert θ ..	358
15.3.7 Konfidenzintervall für σ_X^2 bei normalverteilter Grundgesamtheit	358
15.4 Eigenschaften von Schätzstatistiken	359
15.4.1 Erwartungstreue	359
15.4.2 Mittlerer quadratischer Fehler	363
15.4.3 Konsistenz	367
15.4.4 Effizienz	369
15.5 Übungen	372
16 Statistische Tests	375
16.1 Einführung	375
16.2 Klassische Testtheorie	376
16.3 Parametertests bei normalverteilten Grundgesamtheiten	380
16.3.1 Gauss-Test	380
16.3.2 t -Test	383
16.3.3 Parametertest im linearen Regressionsmodell	384
16.4 Binomialtest	388
16.5 Testentscheidung	391
16.6 Gütefunktion	397
16.7 Operationscharakteristik	403
16.8 Test auf Gleichheit von zwei Mittelwerten	409
16.8.1 Vergleich zweier unabhängiger Stichproben	410
16.8.2 Vergleich von zwei verbundenen Stichproben	418
16.8.3 Unterschied zwischen verbundenen und unabhängigen Stichproben	420
16.9 Übungen	421

17 Statistische Tests für kategoriale Merkmale 423

17.1 Einführung 423

17.2 χ^2 -Anpassungstest 423

17.3 χ^2 -Homogenitätstest 433

17.4 χ^2 -Unabhängigkeitstest 439

17.5 Übungen 440

Teil III Einführung in die multivariaten Verfahren

18 Überblick über verschiedene multivariate Verfahren 445

18.1 Einführung 445

18.2 Asymmetrische Modelle 446

18.2.1 Regressionsanalyse 446

18.2.2 Varianzanalyse 447

18.2.3 Diskriminanzanalyse 447

18.3 Symmetrische Modelle 448

18.3.1 Kontingenzanalyse 448

18.3.2 Faktorenanalyse 449

18.3.3 Clusteranalyse 449

18.3.4 Multidimensionale Skalierung 450

18.3.5 Conjoint Analyse 451

19 Varianzanalyse 453

19.1 Einführung 453

19.2 Einfaktorielle Varianzanalyse 454

19.3 Zweifaktorielle Varianzanalyse 462

19.4 Andere Versuchspläne 474

19.5 Multivariate Varianzanalyse 475

20 Diskriminanzanalyse 481

20.1 Problemstellung der Diskriminanzanalyse 481

20.2 Klassische Diskriminanzanalyse nach Fisher 483

20.2.1 Klassifikationsregel 486

20.2.2 Spezialfall zwei Gruppen 487

20.2.3 Weitere Klassifikationsregeln 488

20.3 Überprüfung der Diskriminanzfunktion 489

20.4 Überprüfung der Merkmalsvariablen 490

20.5 Anwendungsbeispiel 491

21 Grundlagen der hierarchischen Clusteranalyse 503

21.1 Problemstellung der Clusteranalyse 504

21.2 Ähnlichkeits- und Distanzmaße 505

21.2.1 Nominalskalierte binäre Merkmale 506

21.2.2 Nominalskalierte mehrstufige Merkmale 510

21.2.3	Ordinalskalierte Merkmale	512
21.2.4	Quantitative Merkmale	513
21.2.5	Merkmale mit unterschiedlichem Skalenniveau	518
21.3	Hierarchische Clusteranalyse	519
21.4	Agglomerative Verfahren	519
21.4.1	Nearest Neighbour Verfahren	520
21.4.2	Furthest Neighbour Verfahren	523
21.4.3	Centroid Verfahren	525
21.4.4	Median Cluster Verfahren	527
21.4.5	Average Linkage Verfahren	528
21.4.6	Ward's Verfahren	533
21.4.7	Entropieanalyse	537
21.5	Fusionseigenschaften agglomerativer Verfahren	539
21.6	Divisive Verfahren	540
21.6.1	Ein polythetisches Verfahren	541
21.6.2	Ein monothetisches Verfahren	542
21.7	Probleme von Clusterverfahren	545
21.7.1	Definition von Clustern	545
21.7.2	Entscheidung über die Anzahl der Klassen	546
21.7.3	Beurteilung der Klassen	547
21.7.4	Anwendungsempfehlungen	549
21.8	Anmerkung	551
Lösungen zu den Übungen		553
A.1	Deskriptive Statistik	553
A.2	Induktive Statistik	570
Tabellen		595
B.1	Binomialverteilung	595
B.2	Poissonverteilung	600
B.3	Standardnormalverteilung	604
B.4	χ^2 -Verteilung	605
B.5	t -Verteilung	606
B.6	F -Verteilung	607
Literaturverzeichnis		611
Sachverzeichnis		617