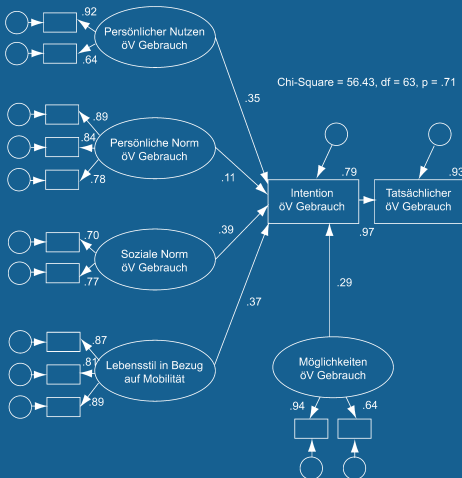


Huib Ernste

Angewandte Statistik in Geografie und Umwelt- wissenschaften



Huib Ernste

Angewandte Statistik in Geografie und Umwelt- wissenschaften

v/dlf

vdf Hochschulverlag AG an der ETH Zürich

Prof. Dr. Huib Ernste war langjähriger Mitarbeiter des Geografischen Instituts der ETH Zürich und ist derzeit Vorsteher des Instituts für Humangeografie an der Radboud University Nijmegen (Niederlande).

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.

ISBN: 978-3-7281-3700-5

DOI: 10.3218/3700-5

© 2011, vdf Hochschulverlag AG an der ETH Zürich

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

Internet: www.vdf.ethz.ch

E-Mail: verlag@vdf.ethz.ch

Vorwort

Die meisten Anwender von multivariaten statistischen Methoden sind keine professionellen Statistiker. Es sind anwendungsorientierte Forscher – Psychologen, Soziologen, Marktforscher, Management-Wissenschaftler, Politikwissenschaftler, Humangeographen und so weiter – die ab und zu statistische Analysemethoden benötigen, um ihre Arbeit besser durchführen zu können. Dieses Buch ist für sie und für Studenten aus diesen Disziplinen geschrieben, insbesondere richtet sich dieses Buch aber an Geographen und Umweltwissenschaftler, da wir unsere Beispiele vor allem aus diesen Fachbereiche genommen haben. Das Schwergewicht liegt darum auf der Datenanalyse und den Zielen jener Leute, die solche Datenanalysen durchführen. Die praktische Forschung prägt den Aufbau dieses Buches. Beim Schreiben dieses Buches habe ich darauf geachtet, dass es intuitiv zugänglich ist. Es setzt in erster Instanz nur wenig mathematische Kenntnisse voraus, ist aber andererseits auch nicht zu simpel. Eine allzu sehr vereinfachende Herangehensweise führt vielfach nicht wirklich zur Einsicht in die Methodik, sondern bleibt auf einen kochbuchartigen Zugang beschränkt. Anwender solcher Methodenkenntnisse wissen meistens nicht genau, was sie machen, sondern befolgen relativ mechanisch die Schritte, die das Kochrezept vorschreibt. In diesem Buch wollen wir deutlich mehr: zu wirklicher Einsicht gelangen. Das bedeutet, dass versucht wird, die Methoden durch Beispiele, graphische Darstellungen und die hinter der Methodik versteckten mathematischen Prinzipien zu erläutern. Auch ein Student mit nur beschränkten Kenntnissen der Matrix-Algebra wird nachvollziehen können, was sich eigentlich abspielt, wenn man eine statische Analysemethode anwendet. Nach eigener Erfahrung führt das nach Schritt-für-Schritt-Nachvollziehen der Analyse anhand eines vereinfachten Beispiels zur wirklichen Einsicht in der Methodik. Dafür ist es manchmal nötig, sich erst gewisse Notationsweisen und mathematische Rechenregeln wieder in Erinnerung zu rufen, um die Analyse schrittweise verfolgen zu können. Hat man sich damit einmal den Durchblick in die Prinzipien der Methodik erarbeitet, dann kann man meistens die mathematischen Einzelheiten auch schnell wieder vergessen und sich auf die Computerprogramme verlassen. Die Einsicht, das

Verständnis, bleibt dann aber. Wirkliche Einsicht führt, nach unserer Überzeugung, also nicht an gewissen mathematischen Einzelheiten vorbei. Diese Überzeugung basiert nicht nur auf langjähriger Erfahrung, sondern auch auf gesicherten didaktischen Kenntnissen. So hat man verschiedene Lernstile oder Aspekte des Lernens unterschieden.¹ Dabei ist bekannt, dass wir nur etwa 20% von dem behalten, was wir nur hören, also nur verbal angeboten bekommen. Dagegen steigert sich dieser Prozentsatz auf etwa 30%, wenn wir den Lernstoff visuell angeboten bekommen, und wir behalten sogar 50% von dem, was wir hören *und* sehen. Wenn wir zusätzlich auch noch darüber diskutieren, bleibt etwa 70% vom Lernstoff haften. Schliesslich kann man diesen Lernerfolg auf fast 90% steigern, wenn man das Gelernte auch noch praktisch anwendet, es also selber ‘tut’. Um also zu dauerhafter Einsicht in die Prinzipien einer statistischen Methode zu gelangen, soll man sie selbst mal praktisch durchgespielt haben und eventuell auch graphisch und mathematisch vorgeführt bekommen. Wenn man allen mathematischen und statistischen Einzelheiten aus dem Weg geht, führt das also lediglich zu ‘Black-Box’- oder ‘Kochbuch’-Wissen, das schnell und auch gern wieder vergessen wird und nicht zu wirklicher Einsicht führt. Bewusst haben wir darum dieses (Lehr-)Buch nicht zu einem ‘Tasten-Kurs’ reduzieren wollen, womit man höchstens lernt, welche Knöpfe man drucken muss, um eine bestimmtes Resultat zu erhalten, ohne dass man wirklich weiss warum. Wir haben uns die Mühe genommen, die Kerngedanken der verschiedenen statistischen Verfahren und die wichtigsten mathematischen und statistischen Prinzipien auch für ein Laienpublikum verständlich und nachvollziehbar zu machen.

Es wäre dabei aber zu viel des Guten gewesen, wenn wir alle denkbaren statistischen Verfahren behandelt hätten. Man würde vor lauter Bäumen den Wald nicht mehr sehen. Wir haben uns deshalb auf eine Auswahl von Methoden beschränkt, die — in Anbetracht der Forschungspraxis — in der Humangeographie und den Umweltwissenschaften relevant sind und gleichzeitig ein gewisses Spektrum unter-

¹ Kolb, D.A. (1983) *Experimental Learning, Experience as the Source of Learning and Development*. Prentice-Hall, Englewood Cliffs.

schiedliche Ansätzen repräsentieren. So werden Methoden sowohl für die Analyse kategorialer als auch metrischer Daten behandelt. Ebenso werden manifeste und auch latente Variablen betrachtet, wodurch auch die Messproblematik und die Verwendung von Indikatoren zur Messung nicht direkt messbarer Variablen berücksichtigt werden. In diesem Buch bleibt die statistische Behandlung der spezifischen Problematik von raum- und zeitabhängigen Daten unberücksichtigt. Dies entspricht weitgehend den neusten Erkenntnissen, dass Raum und Zeit vor allem als sozial konstruierte Aspekten aufgefasst werden und damit nicht mehr einfach mit dem euklidischen Raum verbunden werden können. Die entsprechenden Spezialverfahren für den Umgang z.B. mit räumlicher Autokorrelation und Ähnlichem haben damit wesentlich an Bedeutung eingebüsst und werden hier nicht angesprochen.² Das Methodenspektrum dieses Buches entspricht also weitgehend den heutigen Auffassungen der Geographie nach der kulturellen Wende. Auch wenn wir dabei aus Platzgründen nicht auf gewisse Methoden eingegangen sind, vermittelt dieses Buch trotzdem die Einblick in die allgemeine Problematik und statistischer Vorgehensweisen, so dass sich der Leser ohne Probleme selbst in weitere Aufgaben einarbeiten kann. Hier spielt eben die 'Einsicht' als Schlüsselgrösse eine wichtige Rolle.

Ein weiteres Prinzip dieses Buches ist es, dass es im Gegensatz zu anderen Einführungen nicht bei der klassischen Korrelations- und Regressionsanalyse stehenbleibt, sondern zu komplizierteren Methoden weiterführt (z.B. zur Strukturgleichungsanalyse). Damit wird der harmonische Übergang von der einführenden Statistik in die höhere Statistik gewährleistet. Schliesslich ist dieses Buch nicht einem bestimmten Computerprogramm verpflichtet. Inzwischen gibt es eine Vielzahl verschiedener Programmpakete, die für diese Methoden brauchbar sind. Diese sind inzwischen alle sehr benutzerfreundlich gestaltet und gut dokumentiert. Es ist vorgesehen, dass auf der dieses Buch begleitenden Website Beispiele und Übungsdatensätze zur

² Vgl. für eine Weiterführung z.B. Griffith, D.A. & Amrhein, C.G. (1997) *Multivariate Statistical Analysis for Geographers*. Prentice Hall, Upper Saddle River.

Verfügung gestellt werden. Zusammenfassend führt dies zu folgenden Zielsetzungen dieses Buches:

- praktische Anwendungsorientierung,
- intuitiver Zugang,
- Einsicht in statistische Prinzipien und Denkweisen,
- Orientierung an praktischen Forschungsfragen.

In diesem Band beschränken wir uns auf die gängigsten statistischen Methoden zur Analyse linearer Zusammenhänge und Abhängigkeiten. Besondere Aufmerksamkeit soll dabei der häufig vernachlässigten statistischen Modellierung des Messproblems gelten. In einem ersten Kapitel geben wir einen ausführlicheren Überblick über die behandelten Methoden. Kapitel 2 behandelt die einfache Zusammenhangsanalyse (einfache Korrelationsanalyse). In Kapitel 3 bis 7 wird die klassische Regressionsanalyse (Abhängigkeitsanalyse) behandelt, wobei Messfehler vorläufig noch ausser Betracht bleiben. In Kapitel 8 gehen wir dann auf die Erweiterung der Regressionsanalyse zu der simultanen Analyse mehrerer (verknüpfter) Abhängigkeiten (Pfadanalyse/Kausalmodelle) ein. Die Messproblematik wird in den Kapiteln 9 und 10 ausführlich behandelt, wo wir uns mit der sog. Faktorenanalyse beschäftigen. Schliesslich werden in Kapitel 11 die Pfadanalyse (Kausalmodellanalyse) und die Faktorenanalyse (Messmodellanalyse) in einem sog. Strukturgleichungsmodell vereinigt. Bei all diesen Methoden wird davon ausgegangen, dass die zu analysierenden Merkmale auf einer metrischen (kontinuierlichen) Messskala gemessen wurden, auch wenn wir in der Praxis mit diesen Methoden manchmal nicht metrisch skalierte Merkmale untersuchen werden. Schliesslich wenden wir uns in den Kapiteln 12 bis 14 der Analyse kategorialer Variablen zu. Im Kapitel 12 beschäftigen wir uns insbesondere mit der Analyse kategorialer abhängiger Variablen (Logit-Analyse). Im Kapitel 13 widmen wir uns der allgemeinen Analyse von Zusammenhängen zwischen kategorial skalierten Variablen in Form log-linearer Modelle. Schliesslich fragen wir uns im letzten Kapitel zur Latenten-Klassen-Analyse, wie wir anhand manifester kategorialer Variablen auf latente kategoriale Variablen schliessen können. Zur Ergänzung sind zusätzlich im Anhang zu den verbalen Erklärungen

im Haupttext die wichtigsten mathematischen Einzelheiten leicht verständlich aufgeführt. Auch die allgemeinen Prinzipien der statistischen Test-Theorie wird im Anhang zusammengefasst.

Danksagung

Wenn man auf die beendete Arbeit zurückblickt, ist es nicht nur Befriedigung, was man empfindet, sondern auch grosse Dankbarkeit. So möchte ich hier Prof. Dr. Dieter Steiner von der ETH Zürich dafür danken, dass er mir die Gelegenheit und die Freiheit gelassen hat, mich in die Materie einzuarbeiten und meinen eigene Ideen dazu zu entfalten, die ich auch später als Professor für Humangeographie habe weiter ausbauen und anwenden können. Die Fachgruppe für Quantitative Geographie und Humanökologie des Geographischen Institutes der ETH Zürich war für mich der Kern einer herausfordernden und stimulierenden Arbeitsumgebung, ohne die diese Arbeit unmöglich gewesen wäre. Aber auch meine jetzige Fachgruppe für Humangeographie an der Universität Nijmegen in den Niederlanden ist für ein wichtiges Arbeitsfeld, wo ich meine Erfahrung weiter ausbauen und meine Lehrerfahrung auf dem Gebiet der Forschungsmethodik vertiefen konnte. Insbesondere ist hier auch Kollege Dr. Martin van der Velde hervorzuheben, mit dem ich viele Lehrveranstaltungen durchgeführt habe. Nicht zuletzt sei hier auch der vdf Hochschulverlag erwähnt, der mit viel Geduld und grosser Zuwendung die Publikation dieses Buches begleitet hat. Ich bin ihnen allen zu grossen Dank verpflichtet. Meiner Frau Claudia, Frau Uta Thun und die Lektorinnen des Verlags, die der Text auf orthographische Fehler durchgesehen haben, möchte ich ebenfalls herzlich danken. Für alle verbleibende Fehler zeichne ich selbst die Verantwortung. Selbstverständlich bin ich für Hinweise auf noch verbliebene Fehler dankbar.

*Prof. Dr. Huib Ernste
Nijmegen, Juli 2011*

Inhaltsverzeichnis

1. Überblick	1
1.1 Einleitung	1
1.2 Grundbegriffe	1
1.3 Skalentypen	6
1.4 Relationen: Kausalität und Kovariation	12
1.5 Statistische Methoden: Ein erster Überblick	19
2. Einfache Korrelationsanalyse	27
2.1 Einleitung	27
2.2 Das Messen von einfachen Zusammenhängen	28
2.3 Der einfache Korrelationskoeffizient	32
2.3.1 Die bi-variate Normalverteilung	32
2.3.2 Der Pearson-Produkt-Moment-Korrelationskoeffizient	40
2.4 Rückschlüsse auf die Grundgesamtheit	48
3. Einfache Regressionsanalyse	51
3.1 Einführung	51
3.2 Kausalität und Geschlossenheit	53
3.3 Regressionsanalyse und Geschlossenheit	56
3.4 Die Schätzung der Parameter der Regressionsgleichung	61
3.4.1 Entscheidungskriterien für die Schätzung	62
3.4.2 Die Schätzung der Koeffizienten	66
3.5 Die Interpretation der Resultate	70
3.6 Die Güte des Regressionsmodells	71
3.6.1 Die Zerlegung der Variation	71
3.6.2 Die Anzahl der Freiheitsgrade	75
4. Multiple Regression und multiple Korrelation	79
4.1 Einführung	79
4.2 Die Aufnahme zusätzlicher unabhängiger Variablen ins Modell	81
4.3 Die graphische Darstellung der multiplen Regressionsgleichung	84

4.4	Die Schätzung der Koeffizienten der multiplen Regressionsgleichung	86
4.5	Die Interpretation der Koeffizienten	88
4.6	Der multiple Korrelationskoeffizient	91
4.7	Der partielle Korrelationskoeffizient	94
5.	Das Schliessen auf die Grundgesamtheit bei der Regressionsanalyse	99
5.1	Einleitung	99
5.2	Test für das Bestimmtheitsmass oder Test der 'Güte' des Gesamtmodells	100
5.3	Test für den Regressionskoeffizienten	102
5.4	Test für die Regressionskonstante	106
5.5	Verallgemeinertes Testverfahren für allgemeine lineare Hypothesen	107
5.6	Vertrauensintervalle für Regressionskoeffizienten und -konstante	109
5.7	Vertrauensintervalle für Vorhersagen	110
6.	Regressionsanalyse mit kategorialen unabhängigen Variablen	111
6.1	Einleitung	111
6.2	Regression mit kategorialen unabhängigen Variablen	112
6.3	Regression mit metrischen und kategorialen unabhängigen Variablen	117
6.4	Interaktionseffekte zwischen metrischen und kategorialen unabh. Variablen	118
6.5	Wie erkennt man die Wirkung einer kategorialen Variablen?	122
6.6	Ein Beispiel	125
7.	Überprüfung der Anwendungsbedingungen der Regressionsanalyse	129
7.1	Einleitung	129
7.2	Bedingungen der gewöhnlichen Kleinst-Quadrat-Schätzung	130
7.2.1	Erwartungswert der Residualwerte beträgt Null	132
7.2.2	Keine Autokorrelation	133
7.2.3	Homoskedastizität	136
7.2.4	Kein Zusammenhang zwischen der Störvariablen und den unabhängigen Variablen	139
7.2.5	Keine Kollinearität	140
7.2.6	Residualwerte sind normalverteilt	145
7.3	Überprüfung der Bedingungen	146
7.3.1	Der Erwartungswert der Residualwerte beträgt Null	147
7.3.2	Keine Autokorrelation	147

7.3.3	Homoskedastizität	150
7.3.4	Kein Zusammenhang zwischen den Residualwerten und den unabhängigen Variablen	152
7.3.5	Keine Kollinearität	153
7.3.6	Die Residualwerte sind normalverteilt	161
7.4	Ausreisser	169
8.	Pfadanalyse	185
8.1	Einleitung	185
8.2	Transformation der Variablen	190
8.2.1	Pfadanalyse mit zentrierten Va- riablen.....	190
8.2.2	Pfadanalyse mit standardisierten Variablen	191
8.3	Notation und Begriffe	194
8.3.1	Endogene und exogene Variablen.....	194
8.3.2	Regressions- und Pfadkoeffizienten ...	195
8.3.3	Strukturgleichung	195
8.3.4	Rekursive und nicht-rekursive Pfadmodelle.....	199
8.4	Die Beziehung zwischen den (Ko-)Va- rianzen und den Parametern	201
8.5	Das Schätzen der Parameter	204
8.5.1	Identifikation	208
8.5.2	Das Prinzip der Parameter- Schätzung bei der klassischen Regressionsanalyse	211
8.5.3	Das Prinzip der Parameter- Schätzung bei der Pfadanalyse.....	212
8.5.4	Maximum-Likelihood-Schätz- funktion	213
8.6	Die Interpretation der Resultate	218
8.6.1	Interpretation der standardisier- ten und der unstandardisierten Lösung	218
8.6.2	Direkte, indirekte und totale Ef- fekte	219
8.7	Modellevaluation	220
8.7.1	Die Güte des gesamten Modells	221
8.7.2	Die Beurteilung der einzelnen Komponenten des Modells	228
8.8	Vergleich und Verbesserung von Modellen....	231
8.8.1	Log-Likelihood-Ratio-Chi-Qua- drat-Test für den Unterschied zwischen Modellen	231
8.8.2	Lagrange-Multiplier-Test für mögliche Erweiterungen des Modells ..	233
8.8.3	Reformulierung des Modells	235
8.9	Spezielle Aspekte	236

9.	Konfirmatorische Faktorenanalyse	239
9.1	Einleitung	239
9.1.1	Das Messproblem	240
9.1.2	Das Spezifikationsproblem	242
9.2	Das Messmodell	243
9.3	Die konfirmatorische Faktorenanalyse	247
9.3.1	Die Festlegung einer Skala für die latenten Variablen	248
9.3.2	Die Identifikation des Messmodells	253
9.3.3	Die Validität des Messmodells	255
9.3.4	Die Zuverlässigkeit des Messmodells	261
10.	Explorative Faktorenanalyse	267
10.1	Einleitung	267
10.2	Ziele der Faktorenanalyse	270
10.3	Algebraische Formulierung des Grundproblems	273
10.4	Arten von Faktoren	278
10.5	Ablauf der Faktorenanalyse	283
10.6	Das Faktorenproblem	284
10.7	Geometrische Grundbegriffe	285
10.7.1	Punktdarstellung im Merkmalsraum	285
10.7.2	Vektordarstellung im Objektraum	290
10.7.3	Zur Geometrie der Hauptkomponentenanalyse	299
10.7.4	Algebraische Formulierung der Hauptkomponentenmethode	304
10.7.5	Bestimmung der Anzahl zu extrahierender gemeinsamer Faktoren	305
10.7.6	Die Maximum-Likelihood-Methode	309
10.8	Das Kommunalitätenproblem	313
10.9	Das Rotationsproblem	315
10.9.1	Orthogonale Rotation nach der Varimax-Methode	317
10.9.2	Schiefwinklige Rotation nach der Promax-Methode	319
10.10	Das Faktorenwertproblem	323
10.11	Vergleich mit Regression	326
11.	Strukturgleichungsmodelle	329
11.1	Einleitung	329
11.2	Die Teile des Strukturgleichungsmodells	330
11.2.1	Die Identifikation des Gesamtmodells	333
11.2.2	Spezifikationsprobleme und Interpretationsprobleme des Gesamtmodells	334
11.3	Fazit	339

12. Logit-Analyse	347
12.1 Einleitung	347
12.2 Basis-Form	349
12.3 Der konventionelle regressionsanalytische Ansatz	352
12.4 Alternative Ansätze	358
12.5 Erweiterung auf mehrere unabhängige Variablen	362
12.6 Die Kodierungsformen und Interpretati- on der Parameter	362
12.6.1 Dummy-Kodierung	365
12.6.2 Effekt-Kodierung	368
12.7 Das Schätzen der Parameter	370
12.8 Das gewichtete Kleinste-Quadraten-Ver- fahren	371
12.9 Das Maximum-Likelihood-Verfahren	374
12.10 Die Güte des gesamten Modells	375
12.11 Das Prüfen von Hypothesen über die Pa- rameter	378
13. Log-lineare Modelle	383
13.1 Einleitung	383
13.2 Darstellung der Zusammenhänge in Form einer Kreuztabelle	383
13.3 Formen der Datenerhebung	386
13.3.1 Multinomiales Erhebungsschema	386
13.3.2 Produktnomiales Erhebungsschema	387
13.3.3 Poisson-Erhebungsschema	388
13.4 Zusammenhänge zwischen kategorialen Variablen	388
13.5 Der Chi-Quadrat-Test für die Un- abhängigkeit zweier Variablen	392
13.6 Vergleich relativer Häufigkeitsverteilungen	399
13.6.1 Differenzen relativer Häufigkeiten	400
13.6.2 Relatives Risiko bzw. Verhältnis relativer Häufigkeiten	400
13.6.3 Odds-Ratio	401
13.7 Masse für die Stärke des Zusammenhangs	412
13.7.1 Kontingenzkoeffizient	412
13.7.2 Cramers V	413
13.7.3 Proportionale Fehler-Reduktion	413
13.7.4 Goodman und Kruskals Lambda (asymmetrisch)	415
13.7.5 Goodman und Kruskals Lambda (symmetrisch)	415
13.8 Ein log-lineares Modell mit zwei Variablen	416
13.9 Interpretation der Parameter	419
13.10 Weitere mögliche log-lineare Modelle	421
13.11 Verallgemeinerte Schreibweise für log- lineare Modelle	424

13.12	Schätzung der Parameter und der erwarteten Häufigkeiten	427
13.13	Test für die Güte des Modells	428
13.14	Vergleich verschiedener Modelle	430
13.15	Test für einzelne Parameter	431
13.16	Test für den Einfluss der Haupt- und Interaktionseffekte	432
13.17	Die Suche nach einem geeigneten Modell	432
13.18	Leere Zellen	434
13.19	Die Verwendung log-linearer Modelle für die Analyse von Logit-Modellen	436
14.	Latente-Klassen-Analyse	443
14.1	Einleitung	443
14.2	Lokale Unabhängigkeit	448
14.3	Formale Darstellung der Latenten-Klassen-Analyse	452
14.3.1	Latente Klassenwahrscheinlichkeiten	453
14.3.2	Konditionale Wahrscheinlichkeiten ...	454
14.4	Die Schätzung der Parameter	455
14.5	Die Identifikation	462
14.6	Die Zuordnung der Objekte zu latenten Klassen	466
14.7	In wie weit stimmt unser Modell mit der Wirklichkeit überein?	470
14.8	Anwendungen der Latenten-Klassen-Analyse	475
14.8.1	Exkurs: Die (wirtschafts-)geographische Bedeutung der Arbeitsmoral und Berufsethik	475
14.8.2	Anwendung am Beispiel zur Arbeitsmoral und Berufsethik	477
14.8.3	Modelle mit einer latenten Variablen .	479
14.8.4	Modelle mit mehreren latenten Variablen	491
14.8.5	Vergleich zwischen Gruppen	502
14.9	Probleme der Latenten-Klassen-Analyse	512

Anhang

A.	Repetitorium: Matrix-Algebra	517
A.1	Einleitung	517
A.2	Allgemeines	517
A.3	Definitionen	519
A.4	Matrizenoperationen	523
A.4.1	Addition	523
A.4.2	Subtraktion	524

A.4.3	Multiplikation	524
A.4.4	Multiplikation von Vektoren	526
A.4.5	Exkurs: Vektoren geometrisch betrachtet	527
A.4.6	Division (Inversion)	532
A.4.7	Eigenwerte und Eigenvektoren	535
A.4.8	Wichtigste elementare Rechenregeln für Matrizen	536
A.5	Beispiele für die Verwendung von Matrizen-Algebra	537
A.5.1	Berechnung der Spalten- und Zeilen-Summen und der Summe aller Matricelemente	537
A.5.2	Berechnung von Mittelwerten, Kovarianz und Korrelation mittels Matrizenrechnung	538
A.5.3	Anwendung von Inversen bei der Lösung eines linearen Gleichungssystems	540
A.5.4	Diagonalisierung symmetrischer Matrizen	544
A.5.5	Hauptachsentransformation	545
A.6	Vektor und Matrixdifferentiation	548
A.7	Ermittlung von Extrema ohne Nebenbedingungen	551
A.8	Ermittlung von Extrema mit Nebenbedingungen	552
B.	Grundbegriffe der Testtheorie	555
B.1	Einleitung	555
B.2	Was wollen wir testen?	555
B.3	Testverfahren	557
B.4	Bemerkungen zum Gebrauch und Missbrauch statistischer Tests	568

Abbildungsverzeichnis

1.1	Ein mögliches Ablaufschema des empirischen Forschungsprozesses	2
1.2	Operationalisierung	4
1.3	Pfaddiagramm einer einfachen kausalen Beziehung	15
1.4	Pfaddiagramm einer Kovariation	15
1.5	Pfaddiagramm einer Scheinrelation	16
1.6	Pfaddiagramm einer partiellen Scheinrelation	16
1.7	Pfaddiagramme mit einer konditionalen Variablen (Interaktion)	17
1.8	Beispiel einer konditionalen Variablen	17
1.9	Verkettung von Abhängigkeitsbeziehungen	17
1.10	Reziproke Beziehungen	18
1.11	Rekursives Modell	19
1.12	Zusammenfassung mehrerer nominaler Variablen zu einer nominalen Variablen	23
1.13	Beispiel eines Pfaddiagramms einer Pfadanalyse	24
1.14	Beispiel eines Pfaddiagramms eines Strukturgleichungsmodells	24
2.1	Zusammenhang zwischen ‘Ausbildungsdauer’ und ‘Wissen über Umweltprobleme’	30
2.2	Zusammenhang zwischen ‘Anzahl Fahrten mit öffentl. Verkehr’ und ‘Einkommen’	30
2.3	Bedingte Verteilungen von X und Y	33
2.4	Das Brett von Galton	34
2.5	Das Dreieck von Pascal für 10 Niveaus	36
2.6	Häufigkeitshistogramm und Häufigkeitspolygon ...	36
2.7	Iso-Wahrscheinlichkeitsellipse einer bivariaten Normalverteilung	39
2.8	Iso-Wahrscheinlichkeitsellipse	41
2.9	Transformation der ursprünglichen Werte zu Abweichungen vom Mittelwert	43
2.10	Verteilung der Punkte über die verschiedenen Quadranten	44
2.11	Einfluss eines Ausreissers auf den Korrelationskoeffizienten	47

3.1	(a) Intermediäre, (b) gemeinsame und (c) reziproke Verursachung	59
3.2	Streuungsdiagramm umweltverantwortliches Handeln und Problembewusstsein	61
3.3	Verschiedene Entfernungen zur Regressionsgeraden	63
3.4	Vertikale Abweichung oder Residualwert	63
3.5	Zwei Regressionsgeraden, die beide gleich gut zu den beobachteten Punkten passen	65
3.6	Beispiel zweier Regressionslinien	65
3.7	Regressionssschere	69
3.8	Variationszerlegung graphisch dargestellt	72
4.1	Streuungsdiagramm der Residualwerte e und der potenziellen zusätzlichen unabhängigen Variablen X_2	81
4.2	Kausalschemata für korrelierte und unkorrelierte unabhängige Variablen	83
4.3	Darstellung in einem drei-dimensionalen Raum ...	84
4.4	Regressionsebene bei zwei unabhängigen Variablen	85
4.5	Die Regressionsebene und die Ebene $X_2 = c$	88
4.6	Einfaches Pfaddiagramm für eine multiple Regression	95
4.7	Pfaddiagramm einer Scheinkorrelation	96
6.1	Regression mit nur einer Dummy-Variablen	115
6.2	Regression mit einer metrisch skalierten und einer Dummy-Variablen	117
6.3	Interaktionseffekt zwischen einer Dummy-Variablen und einer metrisch skalierten Variablen ...	120
6.4	Kombination von einfachen Dummy-Variablen und einer Interaktion zwischen einer Dummy-Variablen und einer metrisch skalierten Variablen	121
6.5	Typische Muster, die auf den Einfluss einer diskreten unabhängigen Variablen in Form einer additiven Dummy-Variablen hinweisen	121
6.6	Typische Muster, die auf den Einfluss einer diskreten unabhängigen Variablen in Form einer Interaktionsvariablen hinweisen	122
6.7	Typische Muster, die auf den Einfluss einer diskreten unabhängigen Variablen in Form einer additiven Dummy-Variablen und einer Interaktionsvariablen hinweisen	123
6.8	Typische Muster, die auf den Einfluss einer wesentlichen Strukturveränderung hinweisen	123
6.9	Regression von umweltverantwortlichem Handeln auf (Umwelt-)Problembewusstsein unter Berücksichtigung der Altersklassen	126
7.1	Beispiel für Situation ohne Kollinearität	142
7.2	Beispiel für Situation mit perfekter Kollinearität ..	143

7.3	Beispiel für Situation mit starker, aber nicht perfekter Kollinearität	143
7.4	Beispiel für Situation mit starker Kollinearität, wobei nur ein Regressionskoeffizient genau bestimmbar ist	144
7.5	Beispiel für Situation mit starker Kollinearität, wobei nur die Regressionskonstante gut bestimmbar ist	145
7.6	Beispiel für Situation mit starker Kollinearität, die durch einen Ausreisser verdeckt wird . . .	145
7.7	Durbin-Watson-Entscheidungsregel	149
7.8	Beispiele von Heteroskeastizität anhand des Streudiagramms der Residualwerte	151
7.9	a) Beispiel eines Quantils; b) Quantil-Diagramm . .	167
7.10	Quantil-Diagramm der Standardnormalverteilung .	169
7.11	Quantil-Quantil-Diagramm für den Vergleich der beobachteten Verteilung mit der Standardnormalverteilung	170
7.12	Beispiele verschiedener Quantil-Quantil-Diagramme	171
7.13	Beispiele für Streudiagramme mit Ausreisser .	172
8.1	Einfaches deterministische Regressionsmodell	186
8.2	Einfaches probabilistische Regressionsmodell	187
8.3	Beispiel eines multiplen Regressionsmodells	188
8.4	Beispiel eines komplexen Kausalmodells	189
8.5	Beispiel eines Pfadmodells	194
8.6	Beispiel eines komplexen Pfadmodells	196
8.7	Beispiel für eine direkte reziproke Beziehung	200
8.8	Beispiel für eine indirekte reziproke Beziehung . . .	200
8.9	Das Prinzip der Maximum-Likelihood-Schätzung . .	215
9.1	Beispiel eines einfachen Messmodells für die konfirmatorische Faktorenanalyse	245
9.2	Beispiel eines Messmodells einer explorativen Faktorenanalyse	246
9.3	Beispiel eines Messmodells	252
9.4	Beispiel eines Modells zur Überprüfung der Kriteriumsvalidität	257
10.1	Beispiel eines Pfaddiagramms für explorative Faktorenanalyse	268
10.2	Streudiagramm der Objekte im Merkmalsraum	286
10.3	Streudiagramm der Daten zu Beispiel a	287
10.4	Streudiagramm der Daten zu Beispiel b	288
10.5	Streudiagramm der Daten zu Beispiel c	289
10.6	Merkmalsvektoren im Objektraum (nur die ersten zwei Objektachsen wurden dargestellt)	290
10.7	Winkel zwischen zwei Merkmalsvektoren	292
10.8	Rechtwinkliges Dreieck	292

10.9	Einige Spezialfälle	294
10.10	Operationen mit Vektoren und Basisvektoren	297
10.11	Geometrische Bedeutung der Spalten der Faktorladungsmatrix	299
10.12	Geometrische Bedeutung der Zeilen der Faktorladungsmatrix	300
10.13	Hauptkomponenten einer bi-variaten Normalverteilung	301
10.14	Hauptkomponenten einer drei-dimensionalen Normalverteilung	301
10.15	Projektion eines Punktes auf eine Fläche	302
10.16	Standardabweichung der Streuung auf den Hauptachsen	303
10.17	Scree-Diagramm von Zufallsdaten und realen Daten	307
10.18	Beispiele für eine Einfachstruktur	316
10.19	Primärfaktoren einer schiefwinkligen Rotation	320
10.20	Faktorenmuster und Faktorenstruktur der Primärfaktoren einer schiefwinkligen Rotation	320
10.21	Referenzfaktoren und Referenzfaktorenstruktur einer schiefwinkligen Rotation	322
11.1	Beispiel eines Pfaddiagramms eines Strukturgleichungsmodells	332
11.2	Beispiel für direkte Beziehungen zwischen zwei Messmodellen	336
12.1	Lineares Regressionsmodell für eine dichotome abhängige Variable und eine metrisch skalierte unabhängige Variable	353
12.2	Problem der funktionalen Form bei Anwendung klassischer linearer Regression	357
12.3	Die kumulative logistische Wahrscheinlichkeitsfunktion	360
12.4	Verschiedene Varianten der logistischen Kurve für die Wahrscheinlichkeit der Bewertung 'nützlich' in Bezug auf das neue Landwirtschaftsgesetz	365
13.1	Graphische Darstellung einer diskreten Wahrscheinlichkeitsverteilung	395
13.2	Graphische Darstellung einer stetigen Wahrscheinlichkeitsverteilung	396
13.3	Dichtekurve einer stetigen Wahrscheinlichkeitsverteilung	397
13.4	Dichtekurven verschiedener Chi-Quadrat-Wahrscheinlichkeitsverteilungen	398
13.5	Graphische Darstellung des Vergleichs von Odds	407

13.6	Beispiele für die Ermittlung der Anzahl der Freiheitsgrade	430
13.7	Drei-dimensionale Kreuztabelle $A*B*C$	436
13.8	marginale Kreuztabelle $B*C$	437
14.1	Pfaddiagramm der Latenten-Klassen-Analyse	444
14.2	Beispiele für latente Variablen und Indikatoren ...	445
14.3	Beispiel mehrerer latenter Ursachen für die gleichen manifesten Folgen	446
14.4	Eine Klassifikation der verschiedenen Ansätze in der 'Latente-Struktur-Analyse'	447
14.5	Beispiel einer Likelihood-Funktion mit nur einem Parameter	456
14.6	Beispiel einer Likelihood-Funktion mit einem lokalen Maximum	458
14.7	Entscheidung bei einem Chi-Quadrat-Test für die Güte des Modells	474
14.8	Pfaddiagramm der hypothetischen kausalen Relationen	478
14.9	Pfaddiagramm eines Modells mit einer latenten Variablen (Beispiel bezüglich der Einstellung zur Arbeit)	479
14.10	Pfaddiagramm für Modell mit zwei korrelierten latenten Variablen	492
14.11	Das Zusammenklappen zweier latenter Variablen zu einer	493
14.12	Pfaddiagramme für Modelle mit unkorrelierten latenten Variablen	500
14.13	Pfaddiagramm eines Modells mit einer Gruppenvariablen (Beispiel bezüglich der Einstellung zur Arbeit)	504
14.14	Pfaddiagramm eines Modells mit einer latenten Gruppenvariablen (Beispiel bezüglich der Einstellung zur Arbeit)	505
14.15	Pfaddiagramm einer simultanen Gruppenanalyse, wobei die Gruppenvariable und die latente Variable unabhängig sind (Beispiel bezüglich der Einstellung zur Arbeit)	508
14.16	Pfaddiagramm einer simultanen Gruppenanalyse mit gleich bedeutenden latenten Klassen (Beispiel bezüglich der Einstellung zur Arbeit)	508
14.17	Pfaddiagramm einer simultanen Gruppenanalyse ohne irgendeinen Einfluss der Gruppenvariablen (Beispiel bezüglich der Einstellung zur Arbeit)	511
14.18	Pfaddiagramm einer simultanen Gruppenanalyse ohne Interaktionseffekt zwischen latenten Variablen und Gruppenvariablen bezüglich der manifesten Variablen (Beispiel bezüglich der Einstellung zur Arbeit)	512

B.1	Wahrscheinlichkeitsdichtekurve bei $\rho = 0$	559
B.2	Lineare Transformation verschiedener Verteilungen in eine generalisierte Verteilung	562
B.3	Wirkliche und geschätzte Stichprobenverteilung . . .	563
B.4	Beispiele verschiedener (generalisierter) Stichprobenverteilungen	565
B.5	Ein- und zweiseitige Tests	566
B.6	Verschiedene Fehlerwahrscheinlichkeiten	567

Tabellenverzeichnis

1.1	Zusammenhangsanalyse	20
1.2	Abhängigkeitsanalyse	21
1.3	Messmodellanalyse oder Ähnlichkeitsanalyse	22
2.1	Zusammenhang zwischen kategorialen Variablen ..	31
3.1	Variationszerlegung	77
6.1	Beispiele für Dummy-Variablen	113
6.2	Daten für die Analyse des Einflusses des Geschlechts und der auf das Einkommen	114
6.3	Beispiel mit zwei Dummy-Variablen	116
7.1	Matrix mit den Anteilen der Varianzkomponenten.	159
7.2	Links ursprüngliche Daten mit Rangnummern; rechts gleiche Daten in der Reihenfolge ihrer Rangnummern	167
7.3	Ober- und Untergrenzen für den Durbin-Watson-Test mit nur einer unabhängigen Variablen	183
8.1	Zerlegung der Effekte	220
10.1	Verschiedene statistische Methoden zur Bestimmung latenter Variablen	269
10.2	Rotationsmethoden	318
12.1	Rohdatenmatrix	350
12.2	Kompakte Darstellung der Rohdaten mittels Subpopulationen	351
12.3	Relative Häufigkeiten der Antwortkategorien innerhalb der Subpopulationen	351
12.4	Dummy-Kodierung zweier dichotomer Variablen ..	366
12.5	Dummy-Kodierung einer kategorialen Variablen mit vier Kategorien	367
12.6	Effekt-Kodierung zweier dichotomer Variablen	369
12.7	Effekt-Kodierung einer kategorialen Variablen mit vier Kategorien	369

13.1	Zwei-dimensionale (2×2)-Kreuztabelle mit den absoluten Häufigkeiten der Variablen A und B	384
13.2	Verallgemeinerte Schreibweise für eine zwei-dimensionale Kreuztabelle mit absoluten Häufigkeiten	384
13.3	Verallgemeinerte Schreibweise für eine zwei-dimensionale Kreuztabelle mit relativen Häufigkeiten	385
13.4	Integrierte Darstellung der Kreuztabelle mittels der Prozedur FREQ aus dem SAS-Statistik-Programm-Paket	386
13.5	Kreuztabelle mit den erwarteten Häufigkeiten unter der Annahme unabhängiger Merkmale	390
13.6	Pearson-Chi-Quadrat für Unabhängigkeit von A und B	391
13.7	Beispiel einer diskreten Wahrscheinlichkeitsverteilung	394
13.8	Beispiel einer stetigen Wahrscheinlichkeitsverteilung	395
13.9	Tabelle 13.4 wiederholt	399
13.10	Kreuztabelle der Häufigkeit von Autofahrten und das freiwillige Einhalten von Geschwindigkeitsbeschränkungen im Interesse der Umwelt	405
13.11	Differenz zwischen den Zeilenanteilen in der ersten Zeile und den entsprechenden Zeilenanteilen in den anderen Zeilen der Kreuztabelle 13.10	406
13.12	Bedingte Odds für die Personen, die sehr selten von der Autobahn Gebrauch machen	406
13.13	Bedingte Odds für die verschiedenen Kategorien der Spaltenvariablen in Tabelle 13.10	408
13.14	Odds-Ratios und Log-Odds-Ratios für den Vergleich der bedingten Odds der Personen, die sehr selten von der Autobahn Gebrauch machen, mit den bedingten Odds der Personen, die sehr oft von der Autobahn Gebrauch machen	410
13.15	<i>Lokale</i> Odds-Ratios und <i>lokale</i> Log-Odds-Ratios der Kreuztabelle 13.10	411
14.1	Zwei-dimensionale (2×2)-Kreuztabelle mit den absoluten Häufigkeiten der Variablen A und B	448
14.2	Die gesamte (manifeste) Kreuztabelle für A und B und die latenten Kreuztabellen für die Variablen A und B innerhalb der Kategorien von X	449

14.3	Listenweise Darstellung der latenten Kreuzta- belle für die Variablen A , B und X	449
14.4	Ausgabe des Computerprogramms LCAG bezüglich des Beispiels zur Einstellung zur Arbeit .	455
14.5	Zuordnung der Objekte zu latenten Klassen am Beispiel zur Einstellung zur Arbeit	469
14.6	Ergebnisse der Modellevaluation des Beispiels zur Einstellung zur Arbeit	474
14.7	Residualwerte für das Beispiel zur Einstellung zur Arbeit	476
14.8	Resultate einer Latenten-Klassen-Analyse mit einer latenten Variablen mit zwei latenten Klassen (Beispiel bezüglich der Einstellung zur Arbeit)	481
14.9	Resultate einer Latenten-Klassen-Analyse mit einer latenten Variablen und drei latenten Klassen (Beispiel bezüglich der Einstellung zur Arbeit)	482
14.10	Entwicklung der Evaluationsstatistiken bei Zunahme der Anzahl latenter Klassen (Bei- spiel bezüglich der Einstellung zur Arbeit)	483
14.11	Mit einem 'fixed value constraint' restringier- tes Modell (Beispiel bezüglich der Einstellung zur Arbeit)	484
14.12	Teststatistiken für die Überprüfung der einzel- nen Hypothesen (Beispiel bezüglich der Ein- stellung zur Arbeit)	487
14.13	Mit zwei 'equality constraints' restringiertes Modell (Beispiel bezüglich der Einstellung zur Arbeit)	488
14.14	Restringiertes Modell für ordinale Variablen (Beispiel bezüglich der Einstellung zur Arbeit)	491
14.15	Arbeitsmoral und Berufsethik als zwei korre- lierte latente Dimensionen	495
14.16	Latente Klassenwahrscheinlichkeiten der bei- den latenten Variablen (am und be)	496
14.17	Arbeitsmoral und Berufsethik als zwei korre- lierte latente Dimensionen unter Berücksichti- gung der Ordinalität	497
14.18	a) geschätzte latente Klassenwahrschein- lichkeiten und b) geschätzte latente Klas- senhäufigkeiten (Beispiel bezüglich der Ein- stellung zur Arbeit unter Berücksichtigung der Ordinalität)	498
14.19	Arbeitsmoral und Berufsethik als zwei un-kor- relierte latente Dimensionen	501
14.20	Latente-Klassen-Analyse der Gruppe 'Junge' (Beispiel bezüglich der Einstellung zur Arbeit)	503
14.21	Latente-Klassen-Analyse der Gruppe 'Alte' (Beispiel bezüglich der Einstellung zur Arbeit)	503

14.22	Kreuztabelle der latenten Variablen und der Gruppenvariablen (Beispiel bezüglich der Einstellung zur Arbeit)	505
14.23	Simultane Latente-Klassen-Analyse in zwei Gruppen (Beispiel bezüglich der Einstellung zur Arbeit)	507
14.24	Überprüfung, ob latente Klassen in den beiden Gruppen gleichbedeutend sind (Beispiel bezüglich der Einstellung zur Arbeit)	509
14.25	Überprüfung, ob latente Klassen in den beiden Gruppen gleichbedeutend <i>und</i> gleich gross sind (Beispiel bezüglich der Einstellung zur Arbeit)	510
14.26	Teststatistiken für die schrittweise Überprüfung der einzelnen Hypothesen (Beispiel bezüglich der Einstellung zur Arbeit)	512
B.1	Verschiedene Fehlermöglichkeiten	564

1. Überblick

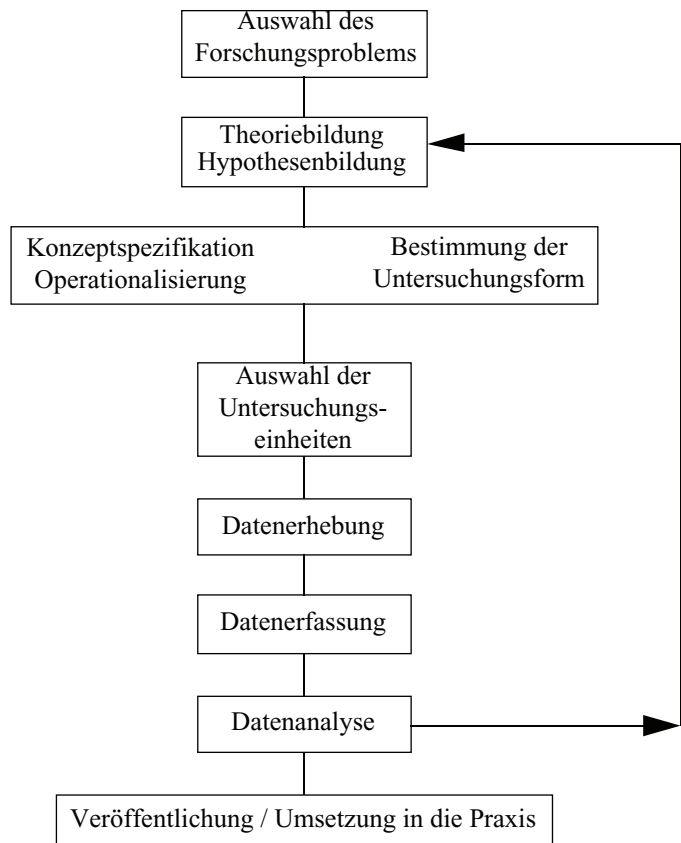
1.1 Einleitung

Ziel dieses Kapitels ist es, einen Überblick über die üblichsten statistischen Methoden zu geben, und zwar so, dass die weniger statistisch gebildeten Forschenden anhand von seiner oder ihrer Problemstellung und Datenlage entscheiden kann, welche Methode am geeignetsten ist. Dies sollte eben einen ersten selektiven und zielgerichteten Zugang zur Statistik ermöglichen. Um unnötige Komplexität zu vermeiden, werden weniger übliche statistische Verfahren in der in diesem Kapitel vorgelegten Übersicht nicht berücksichtigt. Allenfalls wird darauf hingewiesen. Bevor wir uns diesem Überblick zuwenden können, müssen wir uns aber erst einige Begriffe zurechtlegen (Abschnitte 1.2 bis 1.4), anhand derer sich die behandelten Methoden typisieren lassen (Abschnitt 1.5).

1.2 Grundbegriffe

Die statistische Analyse von Daten steht nie nur für sich selbst, sondern ist in einen Forschungsprozess eingebunden. Die Gründe für die Verwendung einer bestimmten Auswertungsmethode haben meistens bereits in den der Auswertungsphase vorangehenden Forschungsschritten oder früheren Forschungsarbeiten ihren Ursprung. In Figur 1.1 haben wir einen möglichen Ablauf des empirischen Forschungsprozesses schematisch dargestellt. So werden schon in der Phase der Konzeptspezifikation und der Operationalisierung entscheidende methodische Weichen gestellt. Schauen wir also die Phase der Konzeptspezifikation und Operationalisierung etwas genauer an.

Zumindest zu Beginn eines Forschungsprozesses sind die Theorien häufig noch nicht explizit und eindeutig formuliert und die Begriffe nicht deutlich von anderen Begriffen abgegrenzt. Theorien bestehen am Anfang meistens nur aus vagen Vorstellungen und Ideen. Wenn wir eine Theorie aufstellen möchten, müssen wir also erst genauer wissen, worüber wir reden. Wir müssen die Bedeutung der von uns verwendeten Begriffe festlegen; wir müssen unsere Begriffe definieren.



In Anlehnung an Schnell, Hill und Esser (1989, S. 110)

Abbildung 1.1: Ein mögliches Ablaufschema des empirischen Forschungsprozesses