

S. Albers / D. Klapper / U. Konradt /
A. Walter / J. Wolf (Hrsg.)

Methodik der empirischen Forschung

3. Auflage



Sönke Albers / Daniel Klapper / Udo Konradt / Achim Walter /
Joachim Wolf (Hrsg.)

Methodik der empirischen Forschung

Sönke Albers / Daniel Klapper /
Udo Konradt / Achim Walter /
Joachim Wolf (Hrsg.)

Methodik der empirischen Forschung

3., überarbeitete
und erweiterte Auflage



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.

1. Auflage erschienen 2006 im Deutschen Universitäts-Verlag Wiesbaden

2. Auflage 2007

3. Auflage 2009

Alle Rechte vorbehalten

© Springer Fachmedien Wiesbaden 2009

Ursprünglich erschienen bei Gabler | GWV Fachverlage GmbH, Wiesbaden 2009

Lektorat: Barbara Roscher | Jutta Hinrichsen

www.gabler.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.

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.

Umschlaggestaltung: KünkelLopka Medienentwicklung, Heidelberg

Gedruckt auf säurefreiem und chlorfrei gebleichtem Papier

ISBN 978-3-8349-1703-4 ISBN 978-3-322-96406-9 (eBook)
DOI 10.1007/978-3-322-96406-9

Vorwort zur 3. Auflage

Unser Konzept, Studierenden und Doktoranden mit einem Buch zu helfen, die Methodik des empirischen Forschens auf einem fortgeschrittenen Niveau zu verstehen und anzuwenden, wird offenbar gut angenommen, denn die 2. Auflage ist ebenfalls wieder ausverkauft.

Gegenüber der 2. Auflage sind die in der Zwischenzeit im Web angebotenen zusätzlichen Artikel in die 3. Auflage übernommen worden. Es handelt sich um die Beiträge zur Multidimensionalen Skalierung, zur Schätzung von Marketing-Modellen mit simulationsbasierten Verfahren, zu Nichtparametrischen Testverfahren, zu Event Studies, zur Conjoint Analyse, zur Sequenzdatenanalyse und zur Spatial Analysis. Weitere Artikel werden kontinuierlich ergänzt und im Internet kostenlos angeboten unter

<http://www.bwl.uni-kiel.de/grad-kolleg/methodenbuch/>

Wir hoffen, dass diese immer umfassender werdende Sammlung von Abhandlungen zu methodischen Fragen der empirischen Forschung immer dann herangezogen wird, wenn anspruchsvolle empirische Forschungsdesigns geplant werden und entsprechende Analysen durchzuführen sind. Wir würden uns freuen, wenn die 3. Auflage des vorliegenden Buches eine ebenso gute Aufnahme findet, wie dies bei den bisherigen Auflagen der Fall war.

Sönke Albers,
Daniel Klapper, Udo Konradt,
Achim Walter, Joachim Wolf

Vorwort zur 2. Auflage

Zu unserer großen Freude war die 1. Auflage des vorliegenden Buches schon nach wenigen Monaten ausverkauft. Offenbar besteht für Bücher ein nicht gedeckter Bedarf, die Studierenden und Doktoranden dabei helfen, die Methodik des empirischen Forschens auf einem fortgeschrittenen Niveau zu verstehen und anzuwenden. Das vorliegende ursprünglich beim Deutschen-Universitäts-Verlag verlegte Buch erscheint deshalb in der 2. Auflage als Lehrbuch beim Gabler Verlag. Mit diesem Buch wollen wir die Erfahrungen der empirischen Arbeit aus zwei Graduiertenkollegs an der Universität zu Kiel an Interessierte weitergeben.

Gegenüber der 1. Auflage sind die einzelnen Beiträge des Buches überarbeitet worden und um zwei weitere Beiträge zur Hazard-Analyse und zum Common Method bzw. Single Source Bias ergänzt worden. Daneben werden als Ergänzung in den nächsten Wochen zusätzliche Beiträge zur Realisierung angemessener Antwortquoten, Behandlung von Ausreißer-Werten, Conjoint Analysen, Soziale Netzwerkanalysen, Sequenzanalysen, Spatial Analysen, Efficient Frontier Analysen, Faktorenanalysen und zur Multidimensionale Skalierung im Internet kostenlos angeboten unter

<http://www.bwl.uni-kiel.de/grad-kolleg/methodenbuch/>

Wir hoffen, dass wir damit eine umfassende Sammlung von Abhandlungen zu methodischen Fragen der empirischen Forschung geben. Jeder Beitrag enthält einen Überblick über die jeweilige Fragestellung, ein angemessenes Vorgehen zu deren Behandlung, weiterführende Literatur und sowie Hinweise auf dafür geeignete Computer-Programmpakete. Auf diese Weise gewinnt die Leserin bzw. der Leser ein umfangreiches Know-how über anspruchsvolle Methoden des empirischen Arbeitens. Wir würden uns freuen, wenn die 2. Auflage des vorliegenden Buches und die zusätzlichen Beiträge im Internet eine ebenso gute Aufnahme finden, wie dies bei der Erstauflage der Fall war.

Sönke Albers,
Daniel Klapper, Udo Konradt,
Achim Walter, Joachim Wolf

Vorwort zur 1. Auflage

Das vorliegende Buch richtet sich an Studierende und Doktoranden, die in ihrem Studium bzw. ihrer Dissertation die Methodik des empirischen Forschens auf einem fortgeschrittenen Niveau verstehen und anwenden wollen. Gegenwärtig gibt es keine Bücher, die die Probleme der Methodik empirischen Forschens von der Forschungsstrategie über die Datensammlung bis hin zur Datenanalyse umfassend darstellen. Das vorliegende Buch möchte dazu beitragen, diese Lücke zu schließen. In den Beiträgen werden die Erfahrungen der empirischen Arbeit von zwei Graduiertenkollegs an der Universität zu Kiel ausgewertet.

Die Kapitel sind ausschließlich von Doktoranden verfasst, die selbst in empirische Projekte eingebunden sind. Sie berichten über die Erkenntnisse, die sie in ihrer Doktorandenausbildung und bei ihren Forschungsprojekten gewonnen haben. Durch eine sorgfältige Begutachtung haben die Herausgeber ihre methodische Erfahrung einfließen lassen. Das Buch ist somit für die Bedürfnisse von Anwendern geschrieben, die entweder Artikel über die Ergebnisse empirischer Forschungsarbeiten verstehen oder selbst empirische Forschungsprojekte durchführen wollen. Das Buch führt in 26 Beiträgen das Wissen zusammen, das auf der Basis eines breiten Spektrums von selbst durchgeführten Projekten und den Erkenntnissen der Literatur entstanden ist. Es besteht aus den 4 Kapiteln der Forschungsstrategien, der Datenerhebung, der Datenanalyse und der Beurteilung der Güte von empirisch gewonnenen Ergebnissen.

Das erste Kapitel enthält Diskussionen der Vor- und Nachteile grundsätzlicher Forschungsstrategien, z.B. von Fallstudien versus großzahliger sowie experimenteller versus nicht-experimenteller Forschung. Im zweiten Kapitel werden Verfahren der Datenerhebung vorgestellt. Am Anfang stehen Beiträge zum Messen und Skalieren sowie zur Stichprobenbildung. Gerade in der Erfolgsfaktorenforschung ist von Bedeutung, wie Erfolg gemessen wird, weshalb eine Gegenüberstellung der unterschiedlichen Alternativen erfolgt. Komplexe Sachverhalte misst man heute als Multifacetten-Konstrukte, de-

ren Eigenschaften ausführlich diskutiert werden. Der letzte Beitrag beschäftigt sich mit den Problemen fehlender Daten und dafür geeigneter Behandlungsverfahren.

Für die Datenauswertung existieren bereits Bücher zu multivariaten Analysemethoden. Die meisten von ihnen gehen aber nicht auf die neueren Verfahren ein, die sich aus der Berücksichtigung von Heterogenität, Endogenität und speziellen Skalenverteilungen ergeben. Nach einem klassischen Kapitel zum Hypothesentest wird auf neuere Verfahren der Clusterbildung auf der Basis nicht deterministischer, sondern stochastischer Grundlagen eingegangen. Auf die Darstellung herkömmlicher Regressionsanalysen wird verzichtet, da diese bereits Gegenstand vieler Methodenlehrbücher sind. Vielmehr werden Probleme der Multikollinearität, ordinaler und kategorialer Skalen auf Seiten der abhängigen sowie unabhängigen Variablen behandelt. Dabei werden Probleme von Mehrgleichungsmodellen ebenso angesprochen wie Schätzungen, bei denen unabhängige Variablen durch Instrumentenschätzer ersetzt worden sind. Von großer theoretischer wie praktischer Bedeutung sind ferner Moderatoren und Mediatoren, deren Berücksichtigung ausführlich behandelt wird. Heute enthalten viele Datensätze Paneldaten, weshalb hier ihre Grundprinzipien behandelt werden. Verallgemeinerungen ergeben sich bei Mehrgleichungsstrukturmodellen, deren Lösung mit dem sich gerade in der Forschung durchsetzenden Partial-Least-Squares-Verfahren dargestellt wird.

Besonders in der Marketingforschung bestehen viele Daten aus Auswahlentscheidungen, weshalb das Aufstellen und Schätzen von Choice-Modellen ausführlich erläutert wird. Dabei wird auch auf die Berücksichtigung von unbeobachteter Heterogenität durch simultane Schätzung von Segmenten und Wirkungsbeziehungen im Rahmen von Latente-Klassen-Regressionen eingegangen. Viele Organisationsprobleme wiederum zeichnen sich durch Daten auf unterschiedlichen Hierarchiestufen aus, wofür spezielle Verfahren der hierarchischen Analyse (HLM) nötig sind, deren Prinzipien erläutert werden. In der Regel liegen Daten als Querschnittsdaten vor. Sind dagegen Längsschnittanalysen durchzuführen, so stellt sich die Frage, ob bestimmte Maßnahmen langfristige Effekte auslösen, wofür geeignete Analyseverfahren vorgestellt werden.

Das Buch schließt mit einem Kapitel, das sich ausführlich mit der Gütebeurteilung von statistischen Analysen beschäftigt. Zunächst werden die Kriterien, insbesondere Reliabilität, Validität und Generalisierbarkeit vorgestellt. Danach wird diskutiert, wie sich mit Bootstrap-Verfahren auch bei nicht der Normalverteilung gehorchenden Stichproben Signifikanzniveaus und die Prognosegüte von Modellen bestimmen lassen. Das Buch schließt mit einer Diskussion geeigneter Prognosegütemaße.

Jeder Beitrag ist so aufgebaut, dass ein Überblick über die jeweilige Fragestellung gegeben und weiterführende Literatur bereitgestellt wird, die den Leser in die Lage versetzen, sich intensiver mit dem Gebiet zu beschäftigen. Außerdem wird dargelegt, mit welchen Computer-Programmpaketen die jeweiligen Analysen durchgeführt werden können. Auf diese Weise gewinnt die Leserin bzw. der Leser ein umfangreiches Know-how über anspruchsvolle Methoden des empirischen Arbeitens. Wir hoffen, dass der Band eine Lücke in dem gegenwärtigen Buchangebot schließen kann, und wünschen allen Lesern viel Erfolg bei ihren empirischen Arbeiten.

Sönke Albers,
Daniel Klapper, Udo Konradt,
Achim Walter, Joachim Wolf

Inhaltsverzeichnis

Vorwort	V
Inhaltsverzeichnis	IX
Abbildungsverzeichnis	XIII
Tabellenverzeichnis	XVII

Kapitel 1: Forschungsstrategie

1. Großzahlige empirische Forschung <i>Felix Riesenhuber</i>	1
2. Experimente <i>Oliver Rack und Timo Christophersen</i>	17
3. Erkenntnisgewinnung durch Fallstudien <i>Andreas Borchardt und Stephan E. Göthlich</i>	33

Kapitel 2: Datensammlung

4. Verfahren der Datenerhebung <i>Maria Kaya</i>	49
5. Messen und Skalieren von Sachverhalten <i>Bert Greving</i>	65
6. Möglichkeiten der Stichprobenbildung <i>Maria Kaya und Alexander Himme</i>	79
7. Subjektive versus objektive Erfolgsmaße <i>Anne Bachmann</i>	89
8. Die Erfassung latenter Konstrukte mit Hilfe formativer und reflektiver Messmodelle <i>Timo Christophersen und Christian Grape</i>	103
9. Zum Umgang mit fehlenden Daten in großzahligen empirischen Erhebungen <i>Stephan E. Göthlich</i>	119
10. Common Method Variance und Single Source Bias <i>Florian Söhnchen</i>	137

Kapitel 3: Struktur entdeckende Verfahren

11. Multidimensionale Skalierung <i>Christian Rohrlack</i>	153
12. Fuzzy Clustering mit Hilfe von Mixture Models <i>Silvia Boßow-Thies und Michel Clement</i>	175

13. Sequenzdatenanalyse <i>Torsten Biemann</i>	191
---	-----

Kapitel 4: Grundlegendes zur Bestimmung von Zusammenhängen

14. Logik und Kritik des Hypothesentestens <i>Torsten Biemann</i>	205
15. Nachweis und Behandlung von Multikollinearität <i>Holger Schneider</i>	221
16. Moderatoren und Mediatoren in Regressionen <i>Dirk Müller</i>	237
17. Endogenität und Instrumentenschätzer <i>Dennis Proppe</i>	253

Kapitel 5: Grundlegende Verfahren zur Bestimmung von Zusammenhängen

18. Logistische und Ordinale Regression <i>Christian Rohrlack</i>	267
19. Conjoint-Analysen <i>Alexander Himme</i>	283
20. Discrete-Choice-Modelle <i>Jarg Temme</i>	299
21. Prinzipien der Panelanalyse <i>Arne Schröder</i>	315
22. Hazard-Raten-Modelle <i>Kerstin Reimer und Christian Barrot</i>	331
23. Mehrgleichungsmodelle: Schätzmethoden und Anwendungsperspektiven <i>Marisa Schlichthorst</i>	349
24. Analyse kausaler Wirkungszusammenhänge mit Hilfe von Partial Least Squares (PLS) <i>Silvia Boßow-Thies und Gregor Panten</i>	365

Kapitel 6: Erweiterte Verfahren zur Bestimmung von Zusammenhängen

25. Eine anwendungsbezogene Einführung in die Hierarchische Lineare Modellierung (HLM) <i>Sascha G. Walter und Oliver Rack</i>	381
---	-----

26. Simultane Schätzung von Choice-Modellen und Segmentierung <i>Markus Horenburger</i>	397
27. Spatial Analysis <i>Sina Henningsen</i>	413
28. Schätzung von Marketing-Modellen mit simulationsbasierten Verfahren <i>Dennis Proppe</i>	433
29. Die Persistenzmodellierung als Methode zur Schätzung von langfristigen Marketingwirkungen <i>Tobias Maria Günter</i>	451
30. Event-Studies <i>Björn Goerke</i>	467

Kapitel 7: Ergebnisgüte

31. Gütekriterien der Messung: Reliabilität, Validität und Generalisierbarkeit <i>Alexander Himme</i>	485
32. Nichtparametrische Testverfahren <i>Daniel Reuschenbach</i>	501
33. Bootstrapping und andere Resampling-Methoden <i>Kerstin Reimer</i>	521
34. Ausgewählte Verfahren der Holdout- und Kreuzvalidierung <i>Jan Kuhlmann</i>	537
35. Prognosegütemaße <i>Christian Barrot</i>	547
Autoren	561
Stichwortverzeichnis	569

Abbildungsverzeichnis

Abb. 1.1	Deduktive Überprüfung von Theorien	2
Abb. 1.2	Generischer Ablauf eines Forschungsprojekts	4
Abb. 1.3	Kontinuum von Daten, Informationen und Wissen	5
Abb. 1.4	Stufen der Theoriebildung, Art der Forschung und Forschungsmethode	6
Abb. 1.5	Flussdiagramm zur Auswahl univariater Testmethoden	13
Abb. 1.6	Flussdiagramm zur Auswahl multivariater Testmethoden	14
Abb. 1.7	Ablaufschema eines Forschungsprojekts und Kapitel des vorliegenden Buches	15
Abb. 3.1	Ordnungsraster zur Einordnung von Fallstudien in den Methodenkontext	35
Abb. 3.2	Datenerhebungsmethoden und ihre speziellen Formen	38
Abb. 3.3	Überblick über den Erstellungsprozess einer Fallstudie	44
Abb. 4.1	Ablauf eines empirischen Forschungsprozesses	49
Abb. 4.2	Methoden zur Gewinnung der Datenerhebung	50
Abb. 4.3	Formen von Panels	61
Abb. 5.1	Gebräuchliche Skalierungsverfahren	67
Abb. 6.1	Ablaufschema der Stichprobengewinnung (Auswahlplan)	79
Abb. 6.2	Teilerhebungsverfahren im Überblick	81
Abb. 7.1	Domänen des Unternehmenserfolgs	92
Abb. 7.2	Klassifikationsschema für Messansätze des Unternehmenserfolgs	93
Abb. 8.1	Analyse mit indirekt messbaren Konstrukten	103
Abb. 8.2	Latentes Konstrukt mit reflektiven Indikatoren	104
Abb. 8.3	Latentes Konstrukt mit formativen Indikatoren	105
Abb. 8.4	Beispiel für ein multidimensionales Konstrukt	107
Abb. 8.5	Zwei-Konstrukt-Modell zur Bestimmung der Validität einer formativen LV	113
Abb. 9.1	Fehlendmuster	122
Abb. 9.2	Übersicht über Verfahren zur Behandlung fehlender Daten	123
Abb. 9.3	Complete Case Analysis sowie Mittelwert- und Regressionsimputationen im Beispiel	126
Abb. 9.4	Multiple Imputation	129
Abb. 10.1	Systematisierung möglicher Messfehler	138
Abb. 10.2	Differenzierung verschiedener Quellen der CMV	141
Abb. 10.3	Prinzip von Harman's One-Factor-Test	142
Abb. 10.4	Berücksichtigung von Kovariaten	145
Abb. 10.5	Berücksichtigung direkt gemessener latenter Faktoren	145
Abb. 10.6	Berücksichtigung nicht-gemessener latenter Faktoren	146
Abb. 10.7	Berücksichtigung mehrerer latenter Methoden-Faktoren	147

Abb. 11.1	Beispiel einer Two-way nicht-metrischen MDS-Lösung mit ALSCAL	157
Abb. 11.2	Verschiedene Nutzenmodelle für Individuen	158
Abb. 11.3	MDS-Lösung zur Darstellung von strategischen Karten inkl. Unternehmen, Leistungskriterien sowie Geschäftsstrategien mit GENFOLD2	160
Abb. 11.4	Prototypische Partitionierungsmuster einer MDS-Lösung	163
Abb. 11.5	Anwendung unterschiedlicher Partitionierungsmuster	165
Abb. 11.6	Eindimensionale räumliche Darstellung für drei ermittelte Dimensionen	167
Abb. 11.7	Segmentierung durch VALUEMAP am Beispiel von 12 Automarken	169
Abb. 12.1	Arten der Heterogenität	175
Abb. 12.2	Methoden der deskriptiven Clusteranalyse	177
Abb. 12.3	Vorgehensweise einer Segmentierung	185
Abb. 14.1	Annahme- und Verwerfungsbereich beim einseitigen Testen	209
Abb. 14.2	Beta-Fehler und Power beim statistischen Test	210
Abb. 15.1	Regressionsmodell mit zwei unabhängigen Variablen	222
Abb. 15.2	Beispiel eines Ridge-Trace	233
Abb. 16.1	Darstellung einer Moderatorbeziehung. Unterschiedliche Niveaus und Steigungen für verschiedene Gruppen	238
Abb. 16.2	Darstellung der Moderatorbeziehung in Modellen	239
Abb. 16.3	Regression ohne (unten) und mit (oben) Interaktionsterm.	244
Abb. 16.4	Beziehungen zwischen Prädiktor-, Mediator- und Prognosevariablen	246
Abb. 16.5	Medierte Moderationsbeziehung (1) und moderierte Mediationsbeziehung (2)	250
Abb. 19.1	Präferenzmessverfahren im Überblick	284
Abb. 19.2	Beispiel für ein Choice-Set für das Produkt „Fruchtsaft“	290
Abb. 20.1	Modellansätze	303
Abb. 20.2	Entscheidungsbaum	304
Abb. 20.3	Dichtefunktion Preisparameter	312
Abb. 21.1	Unbeobachtete Heterogenität und geeignete Schätzverfahren	318
Abb. 21.2	Mögliche Ablaufschritte bei der Panelanalyse	326
Abb. 22.1	Zensierung und Trunkierung bei Ereignisdaten	332
Abb. 22.2	Systematische Übersicht über die verschiedenen Hazard-Raten-Modelle	333
Abb. 22.3	Beziehung von Dichte-, Survivor-, Verteilungsfunktion und Hazard-Rate	335
Abb. 22.4	Hazard-Raten-Verläufe der Weibull- und der Log-Logistischen-Verteilung	339
Abb. 23.1	Zusammenhang zwischen den Anforderungen eines Systems und den jeweils optimalen Schätzern	355
Abb. 24.1	Darstellung eines PLS-Modells	367

Abb. 24.2	Vorgehen des PLS-Algorithmus	369
Abb. 25.1	Spektrum von Variablenzusammenhängen bei hierarchisch strukturierten Daten	381
Abb. 25.2	Vier mögliche Beziehungen von Regressions-koeffizienten und -steigungen	384
Abb. 25.3	Ergebnisdarstellung Submodell 1 mit HLM 6.0	394
Abb. 25.4	Ergebnisdarstellung Submodell 3 mit HLM 6.0	395
Abb. 26.1	Dichtefunktionen zweier Segmente	398
Abb. 26.2	Mischung der Dichtefunktionen zweier Segmente	400
Abb. 29.1	Beispiel für Effekte einer TV-Werbemaßnahme	452
Abb. 29.2	Die zwei Fälle der Gesamtwirkung im Rahmen der Persistenzmodellierung	453
Abb. 29.3	Die Persistenzmodellierung als Multi-Step-Verfahren	454
Abb. 29.4	Typische Verläufe von Zeitreihen	455
Abb. 29.5	Grafische Darstellungen von IR-Funktionen	459
Abb. 29.6	Die Marken- und Preis-Zeitreihen der beiden Marken	462
Abb. 30.1	Confounding Event	471
Abb. 30.2	Kursverlauf ohne bewertungsrelevanten Event	473
Abb. 30.3	Kursverlauf mit bewertungsrelevantem Event	473
Abb. 30.4	Umfang von Schätz- und Ereignisfenster	475
Abb. 30.5	Tägliche abnormale Renditen	480
Abb. 30.6	Kumulierte tägliche abnormale Renditen, eigene Erstellung	481
Abb. 32.1	Nichtparametrische Testverfahren im Überblick	501
Abb. 32.2	Veranschaulichung des Kolmogorov-Smirnov-Tests	505
Abb. 32.3	Vorgehensweise des Kolmogorov-Smirnov-Tests	512
Abb. 33.1	Bootstrap-Verteilung	530
Abb. 34.1	Das Problem des overfitting	538

Tabellenverzeichnis

Tab. 2.1	Anwendungsbeispiel für ein sog. Lateinisches Quadrat	22
Tab. 2.2	Unifaktorieller Vierzufallsversuchsgruppenplan mit Vor- und Nachhermessung	23
Tab. 2.3	Zweifaktorieller Zufallsversuchsgruppenplan (2x2)	24
Tab. 2.4	Zweistichprobenversuchsplan mit Wiederholungsmessung	25
Tab. 2.5	Vollständiges und unvollständiges Ausbalancieren	30
Tab. 4.1	Vor- und Nachteile von Befragungsmethoden	54
Tab. 4.2	Vergleich Labor- und Feldexperiment	59
Tab. 6.1	Verteilung von „Geschlecht“ und „Alter“ in der Grundgesamtheit	82
Tab. 6.2	Quotenanweisung	82
Tab. 8.1	Entscheidungshilfen zur Frage, ob ein reflektives oder ein formatives Messmodell erforderlich ist	110
Tab. 9.1	Literaturquellen mit Evaluation verschiedener Missing-Data-Techniken	130
Tab. 10.1	Zusammensetzung der erklärten Varianz	139
Tab. 11.1	Ausgangsdaten Unähnlichkeiten zwischen 7 verschiedenen Automarken	157
Tab. 11.2	STRESS-Werte für Zufallszahlen (nicht-metrische Two-way-Modelle)	161
Tab. 11.3	Korrekturfaktor k^* für STRESS und S-STRESS Werte von Zufallszahlen	162
Tab. 11.4	STRESS-Werte und S-STRESS-Werte für strukturierte Daten (nicht-metrische Three-way-Modelle)	162
Tab. 11.5	Ausgangsdaten (Interkorrelationen) zum Testen der strukturellen Hypothese	164
Tab. 11.6	Beispielhafte Datenmatrix für „pick any/n“-Daten	166
Tab. 12.1	Zuordnungsmatrizen verschiedener Clustermethoden	178
Tab. 12.2	Informationskriterienbasierte Maße	183
Tab. 12.3	Statistische Entscheidungskriterien für die Segmentlösung	187
Tab. 12.4	Zugehörigkeitswahrscheinlichkeiten ausgewählter Cases zu den Segmenten	187
Tab. 12.5	Mittelwertunterschiede in den Segmenten	189
Tab. 13.1	Fiktive Beispiele für die Entwicklung der Organisations- struktur	191
Tab. 13.2	Beispielsequenzen für berufliche Karriereverläufe	193
Tab. 13.3	Beispiel für eine Transition Cost Matrix	195
Tab. 13.4	Beispiel für eine Distanzmatrix	196
Tab. 13.5	Beispiele für Karrieresequenzen von Topmanagern	197
Tab. 13.6	Überblick über sechs ermittelte Karrieremuster	199
Tab. 14.1	Mögliche (Fehl-)Entscheidungen beim Hypothesentesten	207

Tab. 14.2	Regressionsdaten aus drei Replikationsstudien (n=100)	214
Tab. 14.3	Drei Arten von Power in einer Regression mit 5 Prädiktoren	215
Tab. 14.4	Überblick über einige wichtige Effektgrößen	217
Tab. 15.1	Beispiel einer Korrelationsmatrix	224
Tab. 15.2	Multipler Korrelationskoeffizient, Toleranz und Variance-Inflation-Factor	226
Tab. 15.3	Varianzzerlegungsverhältnisse für das Digital-TV-Beispiel	228
Tab. 16.1	Berechnungsprozedur für Interaktionen in Abhängigkeit des Skalenniveaus	241
Tab. 16.2	SPSS-Output für die Änderung von R und R ²	243
Tab. 16.3	SPSS-Output: Regressionskoeffizienten	243
Tab. 16.4	Test auf Mediation	247
Tab. 16.5	Mediationsbeziehung mit Koeffizienten für statistische Tests und Testkriterien	249
Tab. 17.1	Kriterien bei der Auswahl von Instrumentvariablen	262
Tab. 18.1	Dummy-Kodierung der nominalen und ordinalen Variablen	269
Tab. 18.2	Klassifikationsmatrix binär-logistische Regression	271
Tab. 18.3	Gütemaße der logistischen Regression und deren Wertebereiche	272
Tab. 18.4	Koeffizienten, odds ratios und Konfidenzintervalle für „Mitarbeiterzahl“	273
Tab. 18.5	Freiheitsgrade und Wald-Statistik binär-logistische Regression	274
Tab. 18.6	Likelihood-Quotienten-Test binär-logistische Regression	274
Tab. 18.7	Dummy-Kodierung der nominalen und ordinalen Variablen	275
Tab. 18.8	Logit-Funktionen der multinomial logistischen Regression	276
Tab. 18.9	Klassifikationsmatrix multinomial logistische Regression	276
Tab. 18.10	Wald-Statistik multinomial logistische Regression	278
Tab. 18.11	Likelihood-Quotienten-Test	278
Tab. 18.12	Dummy-Kodierung der nominalen und ordinalen Variablen	279
Tab. 18.13	Logit-Funktion der ordinalen Regression	280
Tab. 18.14	Klassifikationsmatrix für ordinale Regression	281
Tab. 19.1	Effekt-Codierung der Eigenschaftsausprägungen	292
Tab. 19.2	Geschätzte Nutzenparameter für das Beispiel „Fruchtsaft“	293
Tab. 19.3	Spannweiten und Bedeutungsgewichte der Eigenschaften	294
Tab. 20.1	Beispiel Daten Konsumentenpanel	301
Tab. 20.2	Schätzergebnisse MNL und MMNL-Modell	311
Tab. 20.3	Darstellung verschiedener Softwareprogramme zur Discrete-Choice-Schätzung	313
Tab. 21.1	Absätze im Beispiel-Panel	316
Tab. 21.2	Schätzergebnisse der Standardmodelle aus Abschnitt 2	327
Tab. 22.1	Beispieldaten für Amtszeiten von Vorständen	334
Tab. 22.2	Wichtige Verteilungen für parametrische Hazard-Raten-Modelle	338
Tab. 22.3	Hazard-Funktionen und korrespondierende unbestimmte Integrale	341

Tab. 22.4	Datenaufbereitung für Hazard-Raten-Modelle mit TVC	342
Tab. 23.1	Identifikationsstatus und Identifikationskriterium	353
Tab. 24.1	PLS und LISREL im Vergleich	372
Tab. 25.1	Übersicht über Metriken der Ebene-1-Prädiktoren	388
Tab. 25.2	Übersicht über Schätzer bei HLM	390
Tab. 25.3	Überblick über Testverfahren bei HLM	391
Tab. 26.1	Bestrafungsfaktoren von Informationskriterien	403
Tab. 26.2	Bestimmung der Segmentanzahl	409
Tab. 26.3	Beste Optimallösungen für vier Segmente	410
Tab. 27.1	Ergebnisse der OLS-Schätzung (3)	415
Tab. 27.2	Ergebnisse des Mixed regressive - spatial autoregressive model (7)	420
Tab. 27.3	Ergebnisse des Spatial autoregressive error model (9)	422
Tab. 27.4	Ergebnisse des Spatial lag model with spatially regressive error term (11)	423
Tab. 28.1	Realisationen von 6 Münzwürfen	434
Tab. 28.2	Wahrscheinlichkeiten für verschiedene Werte von p	434
Tab. 29.1	Geschätzte Parameter des VAR-Modells	463
Tab. 30.1	Renditendatensatz von 1998 bis 2007	479
Tab. 30.2	Eventinformationen und Ergebnisse der Event-Study-Rechnung	479
Tab. 30.3	Kursverlauf ohne bewertungsrelevanten Event, eigene Erstellung	479
Tab. 30.4	Kursverlauf ohne bewertungsrelevanten Event	481
Tab. 31.1	Reliabilitätsformen im Rahmen der Kausalanalyse	490
Tab. 31.2	Beispiel für eine MTMM-Matrix	494
Tab. 31.3	Übersicht der Schwellenwerte	496
Tab. 31.4	Ergebnis der Varianzanalyse und Darstellung der geschätzten Varianzkomponenten	498
Tab. 32.1	Fallkonstellationen und deren Rahmenbedingungen	502
Tab. 32.2	Erwartete Ergebnisse des χ^2 -Tests für verschiedene Stichproben	504
Tab. 32.3	Erwartete Ergebnisse des Vorzeichen-Test	507
Tab. 32.4	Erwartete Ergebnisse des Wald-Wolfowitz Runs-Tests	509
Tab. 32.5	Erwartete Ergebnisse des Moses-Tests	511
Tab. 32.6	Erwartete Ergebnisse des Kruskal-Wallis H-Test	514
Tab. 32.7	Mittelwerte und Verteilungsbasis der generierten Datensätze	515
Tab. 32.8	Ergebnisse zur Anwendbarkeit des χ^2 -Tests und Kolmogorov-Smirnov-Tests	516
Tab. 32.9	Ergebnisse der Testverfahren basierend auf unabhängigen Stichproben	518
Tab. 33.1	Monatsumsätze der Kunden pro Gruppe	523
Tab. 33.2	Kundendaten	526

Tab. 33.3	Ergebnis der Regressionsanalyse für N=25 Kunden mit $R^2=0,689$	526
Tab. 33.4	Pseudowerte für die erste Replikation	526
Tab. 33.5	Jackknife-Koeffizienten und Überprüfung der t-Werte	527
Tab. 34.1	Beispieldatensatz mit Umsatz, Werbeaufwand und Weihnachts-Dummy	540
Tab. 34.2	Ergebnis der Regression von Umsatz auf Werbeaufwand	541
Tab. 34.3	Validierung der ersten Modellvariante anhand der 8 Beobachtungen des Validierungsdatensatzes	541
Tab. 34.4	Ergebnis der Regression von Umsatz auf Werbe- aufwand und Weihnachts-Dummy	542
Tab. 34.5	Ergebnis der Regression von Umsatz auf Werbeaufwand und Weihnachts-Dummy anhand von Schätz- und Validierungsdatensatz	542
Tab. 35.1	„Kollege Zufall unerreicht“	547
Tab. 35.2	Verbreitung von Prognosegütemaßen in Wissenschaft und Praxis	556
Tab. 35.3	Beurteilung von Prognosegütemaßen	557

Großzahlige empirische Forschung

Felix Riesenhuber

1 Zusammenfassung

Mit dem folgenden Beitrag soll ein Überblick über Ablauf und Methoden großzahliger empirischer Forschung gegeben werden. Ausgehend vom kritischen Rationalismus als Erkenntnisgrundlage konzentriert sich der Artikel auf empirische betriebswirtschaftliche Forschung als angewandte Sozialwissenschaft. Grundzüge qualitativer und quantitativer Forschungsmethoden werden erläutert und entsprechend ihrem Beitrag zur Theoriebildung geordnet. Zuletzt geht der Beitrag auf potentielle Fehler quantitativer Forschung ein und steckt damit den Bezugsrahmen der folgenden Buchbeiträge ab.

2 Der kritische Rationalismus als Grundlage empirischer Forschung

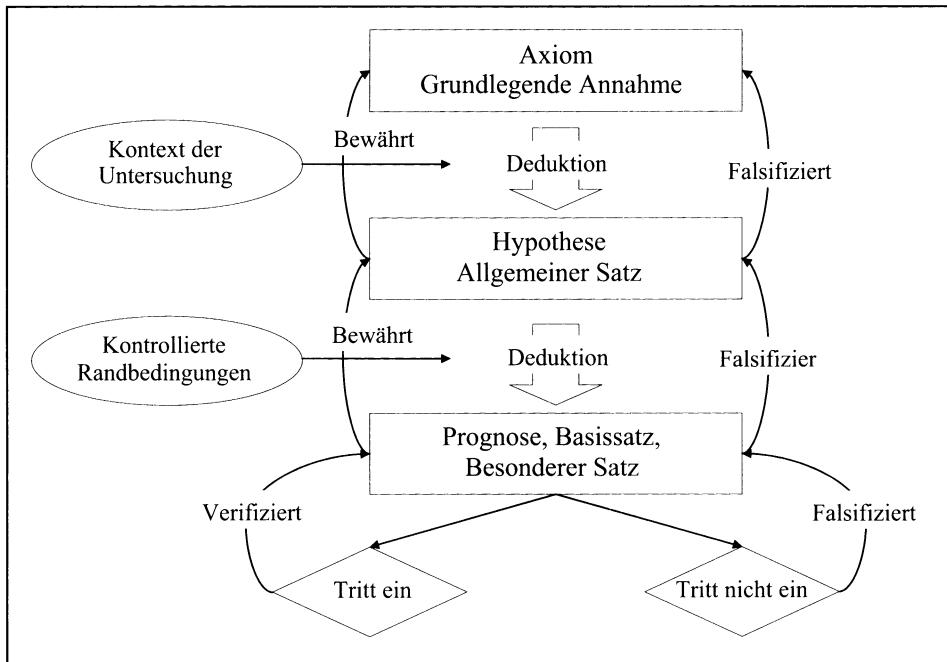
Der Beitrag vertritt ein theoriegeleitetes, deduktives Forschungsverständnis auf Basis des von Popper definierten kritischen Rationalismus. Der vorliegende Abschnitt skizziert Grundgedanken dieser Denkweise, um hierauf aufbauend den konkreten Ablauf eines Forschungsprojekts und die dabei anwendbaren Methoden erläutern zu können (für eine generelle Einführung in die Wissenschaftstheorie siehe Chalmers 2001).

Nach Popper ist Wissenschaft ausschließlich wegen ihres Beitrages zur Lösung des Problems, „die Welt zu verstehen“, interessant (Popper 1966, S. XVI). Theorien sind dabei „das Netz, das wir auswerfen, um „die Welt“ einzufangen, - sie zu rationalisieren, zu erklären, zu beherrschen“ (Popper 1966, S. 31). In seinen Abhandlungen über die „Logik der Forschung“ beschreibt Popper eine Methodenlehre, die dem systematischen Aufbau von Erkenntnis über die Wirklichkeit dient. Leitend ist dabei nicht die Frage nach der einzelnen Tatsache, sondern nach dem, „was gilt“, d.h. „ob und wie ein Satz begründet werden kann, ob er nachprüfbar ist und ob er von gewissen anderen Sätzen logisch abhängt oder mit ihnen im Widerspruch steht usw.“ (Popper 1966, S. 6). Ein Satz ist dabei eine Aussage über die Wirklichkeit. Er kann mit anderen Sätzen in Beziehung stehen und mit ihnen ein theoretisches System bilden. Unterschieden wird zwischen „besonderen Sätzen“, die für einen bestimmten Fall bzw. unter bestimmten Randbedingungen gelten, „allgemeinen Sätzen“, die als Hypothesen oder Naturgesetze generell gelten, und Axiomen, die als grundlegende Annahmen in Bezug auf die Wirklichkeit das theoretische System begründen. Eine Theorie ist damit als System von Axiomen und ihren Folgerungen zu verstehen.

Die Geltung einer Theorie wird deduktiv überprüft. Zunächst werden die Folgerungen der Theorie untereinander verglichen und damit das übergeordnete System auf Widerspruchslosigkeit untersucht. Weiter untersucht man die logische Form des Systems, um festzustellen, ob es den Anforderungen einer empirisch wissenschaftlichen Theorie genügt. Eine Theorie ist empirisch (im Gegensatz zu metaphysisch), wenn aus ihr prüf-

bare Prognosen zu Vorgängen in der Wirklichkeit abgeleitet werden können. Der empirische Gehalt einer Theorie steigt in dem Maße, in dem ihre Folgerungen konkreter, eindeutiger und damit leichter überprüfbar sind, bzw. in dem Maße, in dem die Theorie logisch mögliche Vorgänge verbietet. Der dritte Schritt überprüft, ob die Theorie im Vergleich zu bestehenden Theorien mehr erklärt und damit einem wissenschaftlichen Fortschritt gleichkommt. Im letzten Schritt folgt der empirische Test der Theorie. Hier können sich die theoretischen Behauptungen im „wissenschaftlichen Experiment oder in der technisch-praktischen Anwendung“ praktisch behaupten (Popper 1966, S. 7 ff.).

Abbildung 1.1: Deduktive Überprüfung von Theorien



Für den empirischen Test werden aus den Axiomen der Theorie und dem Untersuchungskontext allgemeine Sätze (Hypothesen) abgeleitet. Aus den Hypothesen werden zusammen mit definierten Randbedingungen besondere Sätze bzw. Basissätze deduziert. Diese Basissätze sind konkrete überprüfbare Aussagen bzw. Prognosen der Ergebnisse eines Experiments. Falls diese Prüfung positiv ausfällt, gilt die Folgerung als verifiziert und das System als vorerst bewährt. Falls sich die prognostizierten Vorgänge nicht beobachten lassen, gelten sowohl die Hypothese als auch das System, aus dem sie abgeleitet wurde, als falsifiziert. Von der Falsifikation der Prognose, des „besonderen Satzes“, wird damit auf die Falsifikation ihrer zugrunde liegenden Hypothese und von dieser auf die Falsifikation der Axiome der Theorie geschlossen. Da die Axiome einer Theorie voneinander unabhängig sind, trifft die Falsifikation unter Umständen nur einzelne Hypothesen bzw. Theorieteile, nicht aber die ganze Theorie (Popper 1966, S. 44 f.). Eine

Theorie gilt damit als falsifiziert, falls eine widersprechende Hypothese von niedrigerer Allgemeinheitsstufe aufgestellt wird und sich im Experiment empirisch bewährt.

Die Überprüfung der Theorie muss grundsätzlich von jedem Sachverständigen nachvollzogen und mit gleichem Ergebnis wiederholbar sein. Weiter kann durch seine Überprüfung ein Satz nie begründet werden. Die Begründung liegt in seiner logischen Ableitung aus den Axiomen bzw. Hypothesen der Theorie und den Randbedingungen seiner Überprüfung. Ebenso handelt es sich bei den Axiomen um Definitionen, die für sich keine Erkenntnisse liefern. Erst durch die Deduktion von Hypothesen und prüfbaren Sätzen werden Einsichten gewonnen.

Eine Theorie steht in ihrem Erklärungsgehalt immer im Wettstreit mit anderen Theorien. Es setzt sich dabei diejenige Theorie durch, die besser prüfbar ist und dabei bereits bewährte Theorien zumindest annäherungsweise enthält. Bei der Überprüfung der neueren, allgemeineren Theorie interessieren vor allem die Hypothesen, die von den bestehenden, bewährten Theorien nicht abgeleitet werden konnten. Da sich eine Theorie nie als „wahr“ herausstellt bzw. Wissenschaft nie Wahrheit erreichen kann und es damit keine „Stufen der Wahrscheinlichkeit“ (Popper 1966, S. 223) von Hypothesen bzw. Theorien gibt, ist der Wettbewerb der Theorien endlos, und im Ergebnis stehen immer erklärungskräftigere, allgemeinere Theorien. Von entscheidender Bedeutung ist, dass die Wissenschaft nie von der Beobachtung zur Theorie (induktives Prinzip) fortschreiten kann und nicht die Anzahl der Experimente, die die betreffende Theorie prüfen, ausschlaggebend ist für ihre empirische Geltung, sondern die Strenge der Prüfung (Popper 1966, S. 5).

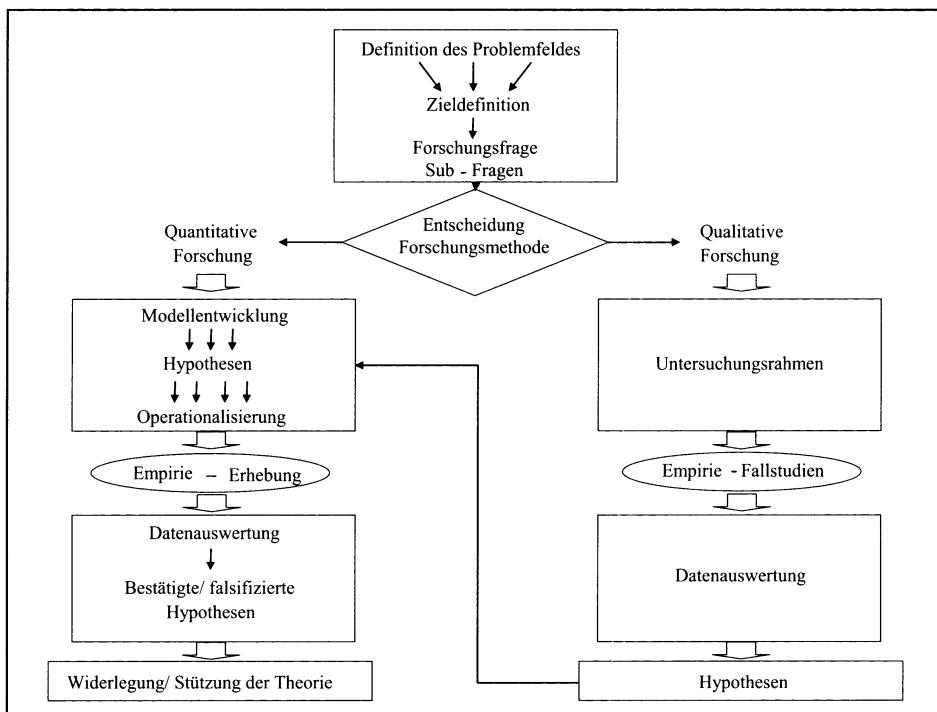
Die weiteren Ausführungen fokussieren vor diesem Hintergrund inhaltlich auf betriebswirtschaftliche Forschung als angewandte Sozialwissenschaft. Als solche bedient sie sich grundlegender Theorien aus beispielsweise den Sozialwissenschaften und der Psychologie, um „disziplinspezifische Sachverhalte zu erklären und ggf. vorauszusagen“ (Schanz 1988, S. 15) sowie letzten Endes eigene Theorien der Organisation, des Managements und der Unternehmensführung zu entwickeln.

Methodisch fokussiert der Beitrag auf empirische Forschung als Mittel der Erkenntnisgewinnung. Sachlich-analytische Forschung, deren Erkenntnisse von reinen Plausibilitätsüberlegungen mit eventueller exemplarischer empirischer Unterstützung herrühren, soll ausgeklammert werden (Grochla 1979, S. 72 ff.). Ebenso wird das formal-analytische Vorgehen zur Erkenntnisgewinnung ausgeklammert, das Problemstrukturen in vereinfachenden abstrakten, oft mathematischen Modellen beschreibt und damit logische Herangehensweisen zur Lösung des Problems entwickeln will (Grochla 1979, S. 85 ff.).

3 Ablauf eines Forschungsprojekts

Abbildung 1.2 stellt den idealtypischen Ablauf eines empirischen Forschungsprojekts dar, an dem sich dieser Abschnitt orientiert.

Abbildung 1.2: Generischer Ablauf eines Forschungsprojekts

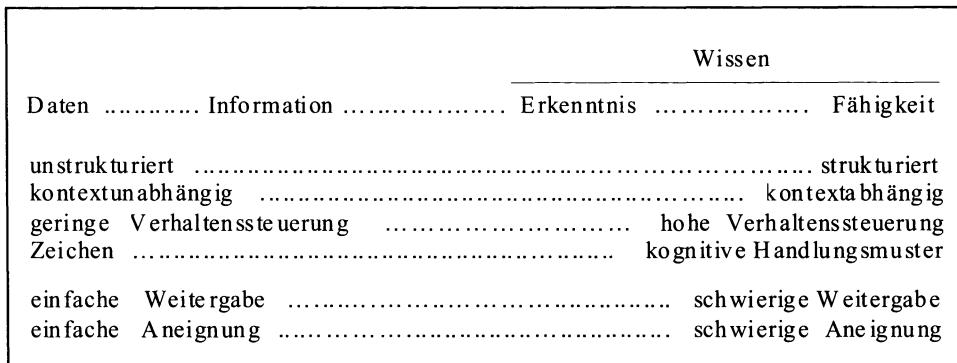


3.1 Definition des Problemfeldes, Zieldefinition und Forschungsfrage

Zu Projektbeginn steht die Definition des Problemfeldes, das das Forschungsprojekt bearbeiten soll. Hierbei ist es sinnvoll, sich das generelle Ziel von Forschung, die Erkenntnisgewinnung bzw. die Generierung von Wissen, zu vergegenwärtigen.

Wissen entsteht aus der Verknüpfung von Informationen. Informationen entstehen wiederum aus der Verknüpfung von Daten, die ihrerseits aus einer geregelten Kombination von Zeichen entstehen. Der Übergang zwischen Daten, Information und Wissen ist fließend. Ihn kennzeichnet das ordnende Prinzip, das Elemente der einen Stufe durch Verknüpfung und Verdichtung in ein Element der höher liegenden Stufe überführt. Während zwischen Zeichen und Daten eine feste Syntax diese Ordnungsvorschrift bildet, ist es bei Daten der Bezug zu einem bestimmten Problem, d.h. Kontext, der diese in Informationen überführt. Bei Wissen ist es das einzelne Subjekt, das Informationen miteinander verknüpft und in Beziehung setzt. Wissen kann weiter nach Erkenntnis bzw. Kenntnis und Fähigkeit differenziert werden. Während Kenntnis Faktenwissen entspricht, d.h. Wissen, dass etwas der Fall ist („knowing that“), definiert Fähigkeit Wissen, wie eine bestimmte Aufgabe zu erledigen ist („knowing how“; Ryle 1958, S. 27 f.; Polanyi 1985, S. 16 und 85).

Abbildung 1.3: Kontinuum von Daten, Informationen und Wissen



Forschung kann nur Wissen in Form von Erkenntnissen hervorbringen. Aufgrund der Personengebundenheit von Wissen ist der Wert dieser Erkenntnisse bzw. die Tatsache, ob es sich bei den Forschungsergebnissen überhaupt um nützliche Erkenntnisse handelt, kaum objektiv bestimmbar, sondern abhängig von dem Vorwissen des Rezipienten und dem Kontext, in dem er sich bewegt.

Im ersten Schritt muss daher geklärt werden, *was* untersucht werden soll und *warum* das *für wen* interessant ist sowie *welche Ergebnisse* man beabsichtigt. Das Ziel des Forschungsprojekts lässt sich gewöhnlich in ein *praktisches* und ein *theoretisches* Ziel aufteilen (van der Velde und Anderson 2004, S. 14). Das praktische Ziel bedeutet einen Beitrag zur Lösung eines bestimmten praktischen Problems, das theoretische Ziel bezieht sich auf den Beitrag des Projekts zum wissenschaftlichen Fortschritt. Vor dem Hintergrund des kritischen Rationalismus kann ein sinnvolles Forschungsprojekt nicht ausschließlich praktische Ziele oder ausschließlich theoretische Ziele verfolgen. Vielmehr bedingen sich durch die theoriegeleitete Erklärung schwach verstandener Realphänomene und deren empirischer Überprüfung beide Zielsetzungen. Aus dem Forschungsziel leitet sich die generelle Forschungsfrage ab, die sich in ihren Facetten in Sub-Fragen zerlegt.

3.2 Auswahl der Forschungsmethode und Art der Ergebnisse

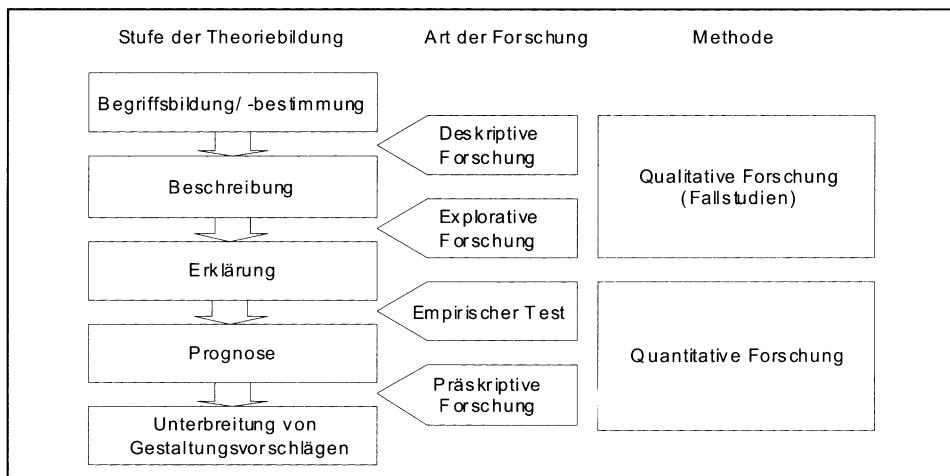
Der Erkenntnisstand in einem gewählten Forschungsgebiet bestimmt die Art der Fragestellung. Die Fragestellung bestimmt ihrerseits die Natur der eigenen Forschung und damit die anzuwendende Forschungsmethode. Eine Forschungsfrage ist nur dann sinnvoll, wenn sie methodisch sinnvoll bearbeitet werden kann (Kutschker, Bärle und Schmid 1997, S. 5). Ebenso setzen verwertbare Ergebnisse eine der Forschungsfrage angemessene Methode voraus. Die Entwicklung einer empirisch gehaltvollen, d.h. testbaren Theorie erfolgt deshalb in Stufen (folgend: Wolf 2005, S. 7 ff.; Grochla 1978, S. 68 ff.).

In der ersten Stufe werden die zugrunde liegenden *Begriffe* des Untersuchungsfeldes definiert bzw. ein einheitliches Verständnis der zu verwendenden Konzepte hergestellt. Die Definition der Begriffe ist notwendig, um in der nächsten Stufe die *Beschreibung*

der Ausprägungen des zu untersuchenden Realphänomens vornehmen zu können. Eine Ausprägung ist zu verstehen als eine bestimmte Variablenkonfiguration eines Untersuchungsobjekts; die Variablen sind die in der ersten Stufe definierten Begriffe, mit denen das Objekt beschrieben werden kann. Darauf aufbauend untersucht die folgende, dritte Stufe der Theoriebildung die Ursachen für das Zustandekommen der Ausprägungen. Sie liefert damit eine vergangenheitsbezogene Erklärung des Realphänomens. Auf Basis der Erklärungen der vorhergehenden Stufe werden in der vierten Stufe Prognosen über zukünftige Entwicklungen abgeleitet und getestet. Die vierte Theoriebildungsstufe entspricht damit dem von Popper generell beschriebenen Vorgehen beim Theorietest. Die fünfte Theoriebildungsstufe leitet aus den Prognosen bzw. bestätigten Variablenzusammenhängen der vierten Stufe praktische Verhaltensempfehlungen ab (Abbildung 1.1).

Forschungstätigkeit überführt eine Theorie des gewählten Feldes von einer Stufe in die nächste. Wie in Abbildung 1.4 dargestellt, kann damit unterschieden werden zwischen (1) deskriptiver Forschung, (2) explorativer Forschung, (3) empirischem Test und (4) der präskriptiven, beratenden Forschung. Entsprechend der Natur der Aufgabe unterscheiden sich die zu verwendenden Werkzeuge. Grundsätzlich lassen sich großzahlige, quantitative empirische Forschung und kleinzahlige, qualitative empirische Forschung unterscheiden.

Abbildung 1.4: Stufen der Theoriebildung, Art der Forschung und Forschungsmethode



3.2.1 Qualitative Forschung

Qualitative Forschung erfasst die Variabilität der Merkmale des Untersuchungsobjekts durch verbale Beschreibung am Beispiel sorgfältig ausgewählter Einzelfälle. Qualitative Forschung in Form von Fallstudien wird in Gebieten mit geringem Kenntnisstand mit dem Ziel eingesetzt, ein Tiefenverständnis des komplexen Realphänomens zu gewinnen. Auf Grundlage der vergangenheitsbezogenen Erklärung bzw. dieses Tiefenverständnisses formuliert der Forscher als Ergebnis Hypothesen, die über die untersuchten Fälle hinaus Gültigkeit haben sollen. Fallstudien können dabei vor dem Hintergrund theoretisch-

scher Vorüberlegungen (Yin 1984) oder theorieilos als Grundlage der Theorieentwicklung (Glaser und Strauss 1967) durchgeführt werden. Bei den Ergebnissen von Fallstudien handelt es sich nicht um vollwertige Theorien oder um einen Theoretetest, sondern um Hypothesen. Fallstudienergebnisse sind damit nur dann wertvoll, wenn (1) ihr Bezug zu existierenden Theorien dargestellt werden kann und (2) sie empirisch testbare Aussagen liefern können, die über die existierenden Theorien hinausgehen (siehe den Beitrag von Göthlich und Borchardt zu Fallstudien in diesem Buch).

3.2.2 *Quantitative Forschung*

Im Gegensatz zur verbalen Beschreibung qualitativer Forschung erfasst quantitative Forschung die Variabilität eines Merkmals über die definierte Zuordnung von Zahlenwerten. Die Menge aller Merkmalsmessungen einer Untersuchung wird als Daten bezeichnet. Quantitative Daten lassen sich erheblich einfacher verarbeiten als qualitative Daten. Aus diesem Grund kann quantitative Forschung mit wesentlich größeren Stichproben arbeiten als qualitative Forschung. Ein grundlegender Unterschied der Ergebnisse kleinzahliger qualitativer und großzahliger quantitativer Forschung liegt damit in der Generalisierbarkeit bzw. Repräsentativität der Ergebnisse für die Grundgesamtheit, aus der die Stichprobe entnommen wurde. Quantitative Forschung eignet sich damit zum Hypothesentest und zur Ableitung von Handlungsempfehlungen (Bamberg und Baur 1998, S. 6 f.).

Entsprechend der Zieldefinition des Forschungsprojekts bzw. des Vorgehens der angewandten Sozialwissenschaften werden daher eine oder mehrere generelle Theorien zur Klärung des interessierenden Sachverhalts herangezogen. Dieses theoretische Analyseaster erlaubt die Zerlegung des Forschungsproblems in Einzelkomponenten, die als Modell miteinander in Beziehung stehen. Zur Prüfung des Modells werden hieraus einzelne Hypothesen – die Forschungshypothesen – abgeleitet und diese entsprechend dem Forschungsdesign und der Stichprobe sowie der Ergebnisse qualitativer Studien zu überprüfbaren Einzelaussagen – den operationalen Hypothesen – konkretisiert, d.h. operationalisiert. Entsprechend der Datengrundlage werden aus den operationalen Hypothesen statistische Hypothesen formuliert und mittels geeigneter Tests überprüft (Bortz und Döring 1995, S. 461).

Zur Überprüfung der Hypothesen können entweder Primärdaten erhoben, Sekundärdaten herangezogen oder eine Kombination von beiden genutzt werden. Entsprechend der Datengrundlage bzw. der Skalierung der Daten kommen unterschiedliche Auswertungsverfahren zum Zuge. Das Ergebnis dieser Verfahren sind bestätigte und nicht bestätigte statistische Hypothesen. Von den statistischen Hypothesen wird zurückgeschlossen auf die operationalen Hypothesen, von diesen auf die Forschungshypothesen und damit auf die Geltung des entwickelten Modells und der zugrunde liegenden Theorie (vgl. Abbildung 1.1). Damit ergeben sich aus dem Forschungsprojekt die empirische Bestätigung (oder Widerlegung) einer oder mehrerer Theorien sowie ein Beitrag zum Verständnis eines bisher nicht oder nur schwach verstandenen Realphänomens.

Das Funktionieren dieses Prozesses ist an mehrere Voraussetzungen gebunden.

- ⇒ *Theoriegeleitetes Vorgehen*: Es muss vor der Erhebung bzw. der Datenanalyse definiert sein, was anhand der Daten untersucht werden soll und welche Ergebnisse man erwartet. In der Regel sind hierfür neben dem Studium der wissenschaftlichen Literatur qualitative Vorstudien notwendig. Die Auswertung eines Datensatzes mittels statistischer Methoden ohne zugrunde liegende Theorie ist aufgrund des Problems der Scheinkorrelation mit erheblichen Problemen behaftet (siehe hierzu den Beitrag von Biemann zu Hypothesentests in diesem Buch).
- ⇒ *Klare Definition der Grundgesamtheit*: Die Generalisierbarkeit der Ergebnisse kann sich nur auf die Grundgesamtheit beziehen, aus der die Stichprobe gezogen wurde. Falls das Untersuchungsobjekt nicht oder unklar definiert wurde bzw. die Stichprobe nicht dem Untersuchungsobjekt entspricht, sind die Ergebnisse kaum über den speziellen Datensatz hinaus generalisierbar.
- ⇒ *Korrekte Bildung der Stichprobe*: Die Stichprobe muss bezüglich der Verteilung der zu untersuchenden Merkmale die Grundgesamtheit widerspiegeln.
- ⇒ *Korrekte Operationalisierung* (valide Messung): Die operative Messung der Merkmalsausprägungen muss dem entsprechen, was in der Hypothese miteinander in Beziehung gesetzt wurde.
- ⇒ *Verzerrungsfreie Durchführung der Erhebung* (Erhebungsdesign): Systematisch verzerrende Einflüsse müssen in der Datenerhebung kontrolliert werden.
- ⇒ *Korrekte Verarbeitung der Daten* (methodische Angemessenheit): Die Methode der Datenverarbeitung muss dem zugrunde liegenden Messniveau bzw. Skalenniveau entsprechen.

Der folgende Abschnitt geht kurz auf die angesprochenen Punkte ein. Für eine vertiefende Auseinandersetzung wird auf die entsprechenden Kapitel dieses Buches verwiesen.

4 Grundbegriffe großzahliger empirischer Forschung

Dieser Abschnitt gibt einen knappen Überblick über Grundbegriffe großzahliger empirischer Forschung. Es werden der Begriff der Hypothese definiert und unterschiedliche Abstraktionsniveaus von Hypothesen vorgestellt. Hierauf aufbauend wird kurz auf unterschiedliche Mess- bzw. Skalenniveaus eingegangen, um abschließend einen Überblick über Methoden großzahliger empirischer Forschung zu geben.

4.1 Hypothesen

Je nach Konkretisierungsgrad kann zwischen Forschungshypothesen, operationalen Hypothesen und statistischen Hypothesen unterschieden werden.

Forschungshypothesen sind „aus Voruntersuchungen, eigenen Beobachtungen, Überlegungen und wissenschaftlichen Theorien“ abgeleitete „Vermutungen bezüglich eines in Frage stehenden Untersuchungsgegenstandes“ (Bortz und Döring 1995, S. 461). Forschungshypothesen beziehen sich auf Zusammenhänge in der zu untersuchenden Grundgesamtheit, über die in der Untersuchung durch Stichprobenziehung Erkenntnisse

gewonnen werden sollen. Sie legt damit den Geltungsbereich ihrer Aussage (die Grundgesamtheit) und die zu ziehende Stichprobe fest. Es können drei Arten von Hypothesen unterschieden werden:

- ⇒ *Zusammenhangshypothesen*, die Aussagen über einen positiven oder negativen Zusammenhang zwischen zwei Merkmalen treffen.
- ⇒ *Unterschiedshypothesen*, die Aussagen zu Unterschieden zwischen zwei oder mehreren Populationen treffen.
- ⇒ *Veränderungshypothesen*, die Aussagen zur Veränderung eines Merkmals im Zeitverlauf treffen.

Für die durchzuführende Untersuchung wird aus der Forschungshypothese eine operationale Hypothese abgeleitet, d.h. die Forschungshypothese wird für die konkrete Untersuchung operationalisiert. Die operationale Hypothese bezieht sich auf Ergebnisse der Untersuchung und ist damit Produkt der Forschungshypothese und der Randbedingungen der Untersuchung (vgl. Abbildung 1.1, folgend in Anlehnung an Bortz und Döring 1995, S. 460 ff.).

Die *operationale Hypothese* definiert, wie die Forschungshypothese in der Untersuchung auf operationaler Ebene geprüft werden soll. Im nächsten Schritt wird für die operationale Hypothese festgelegt, wie mit ihr statistisch verfahren werden soll, d.h. ab wann die Hypothese als bestätigt bzw. widerlegt gelten soll. Da die empirische Geltung einer Hypothese in der Betriebswirtschaft aufgrund der Komplexität des menschlichen Verhaltens nur „im Prinzip“ gezeigt werden kann, wird Falsifizierbarkeit durch die willkürliche Einführung statistischer Prüfkriterien, insbesondere der statistischen Signifikanz, erzeugt. Zu diesem Zweck wird die operationale Hypothese in statistische Hypothesen übersetzt, die durch einen Signifikanztest überprüft werden. Der Signifikanztest überprüft formal zwei komplementäre, sich einander ausschließende Hypothesen: die Nullhypothese (H_0) und die Alternativhypothese (H_1). Das Signifikanzniveau gibt dabei die Wahrscheinlichkeit des Fehlers 1. Art an, mit dem die Nullhypothese irrtümlich verworfen bzw. abgelehnt wird. Die Nullhypothese wird nicht abgelehnt, wenn das Beobachtungsmaterial nicht im signifikanten Widerspruch zu H_0 steht. Das bedeutet jedoch nicht, dass H_0 zutrifft, sondern nur, „dass die Beobachtungsdaten nicht zu einer Ablehnung von H_0 ausreichen“ (Bamberg und Bauer 1998, S. 182). Der Signifikanztest gibt damit an, ob die Befunde einer Stichprobe zu dem gegebenen Signifikanzniveau auf die Grundgesamtheit übertragen werden können. Für eine ausführliche Auseinandersetzung mit dem Thema Hypothesen und Hypothesentests wird auf den Beitrag von Biemann in diesem Buch verwiesen.

Die Eignung bestimmter Signifikanztests und statistischer Analyseverfahren ist abhängig von dem zugrunde liegenden Datenmaterial, das sich aus der Art der Messung des interessierenden Merkmals ergibt. Auf unterschiedliche Skalenniveaus geht der folgende Abschnitt ein.

4.2 Messung und Skalierung

Durch Messung wird ein qualitatives Merkmal quantifiziert, d.h. in einen Zahlenwert übersetzt. Die „Übersetzungsvorschrift“ bildet dabei die Skala. Merkmale unterscheiden

sich in ihrer Messbarkeit, d.h. in ihrer Beschreibbarkeit durch Zahlen. Je nachdem, inwieweit sich die Ausprägungen eines Merkmals in Zahlen ausdrücken lassen, können unterschiedliche Skalenniveaus unterschieden werden (Backhaus, Erichson, Plinke und Weiber 2000, S. XVIII ff.):

- ⇒ Nominalskala
- ⇒ Ordinalskala
- ⇒ Intervallskala
- ⇒ Ratioskala.

Die Art der Skala definiert damit den Informationsgehalt der Daten und die Anwendbarkeit von Rechenoperationen und damit statistischen Analyseverfahren.

Durch die *Nominalskala* werden qualitative Merkmale willkürlich durch Zahlen ersetzt (z.B. rot = 1, grün = 2), um sie mit Computern verarbeitbar zu machen. Sie stellt die primitivste Art des Messens dar. Aufgrund der zufälligen Zuordnung von Zahlenwerten zu Merkmalsklassen sind arithmetische Rechenoperationen auf dieser Skala nicht zulässig (sobald die Zahl verändert wird, verliert sie ihre Information). Durch Zählen lassen sich lediglich die Häufigkeiten einzelner Ausprägungen ermitteln.

Durch eine *Ordinalskala* können Rangordnungen von Untersuchungsobjekten erstellt werden. Sie erlaubt den paarweisen Vergleich von Objekten („besser als“, „schlechter als“). Von einer Ordinalskala kann jedoch nicht abgelesen werden, wie groß der Abstand zwischen zwei Objekten ist, es dürfen daher ebenfalls keine arithmetischen Operationen mit ordinal-skalierten Merkmalen durchgeführt werden. Neben Häufigkeiten sind Median und Quantile aussagekräftige statistische Maße. Bei den oft verwendeten Ratingskalen (z.B. Bewertung eines Objekts auf der Skala von 1 bis 7) handelt es sich um Ordinalskalen. Um jedoch statistische Auswertungsmethoden nutzen zu können, nimmt man oft gleich große Skalenabschnitte einer Intervallskala an.

Intervallskalen sind in gleich große Skalenabschnitte eingeteilt, verfügen jedoch über keinen natürlichen Nullpunkt (z.B. Celsiusskala). Die Differenzen zwischen Skalenabschnitten beinhalten daher Information, und die arithmetischen Operationen Addition und Subtraktion sind zulässig. Neben den bereits genannten liefern Mittelwert und Standardabweichung gehaltvolle Aussagen, nicht aber die Summe.

Über die Eigenschaften der Intervallskala hinaus verfügt die *Ratio- (oder Verhältnis-) Skala* über einen natürlichen Nullpunkt (Merkmal hat die Ausprägung „0“ bzw. „nicht vorhanden“, z.B. Größe oder Preis). Sie stellt damit das höchste Messniveau dar. Durch die Fixierung des Nullpunktes sind sowohl Differenzen als auch Quotient oder Produkt zulässige arithmetische Operationen. Neben bereits genannten statistischen Maßen sind die Anwendung des geometrischen Mittels und des Variationskoeffizienten erlaubt.

Mit der Thematik des Messens befassen sich drei weitere Artikel in diesem Buch. Während Greving eine vertiefende Einführung in Messung und Skalierung von Sachverhalten gibt, diskutiert der Beitrag von Bachmann die Problematik der Erfolgsmessung über subjektive und objektive Erfolgsmaße. Abschließend behandelt der Beitrag von Christophersen und Grape die Erfassung latenter Konstrukte über formative und reflektive Messmodelle.

Vor dem Hintergrund der theoretischen Hypothesen und der Messung der interessierenden Merkmale kann der folgende Abschnitt eine kurze Einführung in Grundlagen der Stichprobenziehung geben.

4.3 Stichprobenbildung

In der Regel ist es nicht möglich, das Forschungsproblem an allen betroffenen Objekten zu untersuchen. Aus diesem Grund greift man auf die Ziehung einer Stichprobe zurück, die für die interessierenden Eigenschaften der Grundgesamtheit repräsentativ ist. Hierauf aufbauend ermitteln Verfahren der induktiven Statistik, ob Beobachtungen in der Stichprobe auf die theoretisch definierte Grundgesamtheit übertragen werden können.

Stichprobenbildung ist nicht sinnvoll bei sehr kleinen oder heterogenen Grundgesamtheiten. Stichprobenbildung ist außerdem unnötig, wenn es sich in Bezug auf das Untersuchungsmerkmal um eine vollkommen homogene Grundgesamtheit handelt (hier genügt die Untersuchung eines Objekts). Für die Repräsentativität einer Stichprobe gibt es keine definierten Kriterien. Sie ist damit mehr „theoretische Zielvorgabe als ein Attribut konkreter Untersuchungen“ (Bortz und Döring 1995, S. 372). Da Methoden der induktiven Statistik unabhängig von der Qualität der Stichprobe funktionieren, sind Ergebnisse der statistischen Auswertung immer vor dem theoretischen Hintergrund der Untersuchung und vor den Eigenschaften der konkreten Stichprobe zu diskutieren. Grundsätzlich lassen sich probabilistische und nicht-probabilistische Stichproben unterscheiden.

Probabilistische Stichproben gehen davon aus, dass jedes Untersuchungsobjekt die gleiche Wahrscheinlichkeit hat, gezogen zu werden, und deshalb die Stichprobe in ihrer Zusammensetzung der Grundgesamtheit entspricht. Je „zufälliger“ die Stichprobe zusammengesetzt ist, desto besser eignet sie sich für die Untersuchung. Unterschieden werden kann zwischen der einfachen Zufallsstichprobe und Stichproben, die das Zufallsprinzip mehrfach anwenden. Dies kann geschehen, indem die Grundgesamtheit aufgeteilt und dann zufällig gezogen wird (geschichtete Stichprobe), per Zufall nicht einzelne Objekte, sondern natürliche Gruppen von Objekten gezogen werden (Klumpenstichprobe) oder das Zufallsprinzip auf andere Weise erweitert wird (mehrstufige Stichprobe) (Bortz und Döring 1995, S. 451 ff.).

Bei nicht-probabilistischen Stichproben sind die Auswahlwahrscheinlichkeiten nicht bekannt oder unkontrollierbar. Hierzu zählen die Ad-hoc-Stichprobe, die theoretische Stichprobe und die Quotenstichprobe. Die Ad-hoc-Stichprobe setzt sich aus Objekten zusammen, die bequem erreichbar sind. Sie ist von sehr begrenztem theoretischem Aussagewert. Die theoretische Stichprobe wählt die Objekte nach zuvor überlegten Kriterien einzeln aus. Sie ist vor allem in der qualitativen Forschung von Bedeutung. Die Quotenstichprobe versucht durch eine passende Auswahl der Objekte, die Zusammensetzung der Stichprobe der der Grundgesamtheit anzugeleichen (Bortz und Döring 1995, S. 451 ff.).

Probabilistische Stichproben sind von wesentlich höherer Aussagekraft als nicht-probabilistische Stichproben. Ihre Zufälligkeitsvoraussetzung wird in der Praxis angestrebt, kann jedoch kaum streng eingehalten werden. Der Beitrag von Kaya in diesem Buch geht vertiefend auf das Thema Stichprobenbildung ein.

4.4 Datenerhebung

Zur Beschreibung der Stichprobe können Primär- und Sekundärdaten herangezogen werden (folgend: Churchill 1991, S. 247 ff.).

Sekundärdaten sind Daten, die nicht speziell für den Zweck der vorliegenden Untersuchung erhoben wurden. Es kann sich dabei um interne und externe Sekundärdaten handeln. Interne Sekundärdaten sind Daten, die intern in der zu untersuchenden Organisation gesammelt wurden, beispielsweise Daten aus dem internen Rechnungswesen, dem Vertrieb oder der Personalabteilung. Externe Sekundärdaten lassen sich in öffentliche und kommerzielle Sekundärdaten unterteilen. Öffentliche Sekundärdaten sind jedem frei zugänglich, beispielsweise amtliche Statistiken, Zeitungen oder Geschäftsberichte. Kommerzielle Sekundärdaten werden von Unternehmen gegen Gebühren angeboten. Hierunter fallen beispielsweise Scannerdaten oder Paneldaten. Der Vorteil von Sekundärdaten liegt in ihrer kostengünstigen und zeitsparenden Beschaffung. Nachteile ergeben sich aus ihrem oft fehlenden direkten Bezug zum Forschungsproblem bzw. Untersuchungsobjekt (problem of fit). Weiter kann ihre Verlässlichkeit bzw. Genauigkeit in Bezug auf korrekte Erhebung und Auswertung oft nicht überprüft werden (problem of accuracy). Da Sekundärdaten oft nicht alle Variablen beinhalten, die die interessierende Variable nach theoretischen Überlegungen beeinflussen, ist das Problem nicht kontrollierter systematischer Einflüsse hoch. Sekundärdaten müssen deshalb in der Regel durch Primärdaten ergänzt werden. Außerdem sollte immer mit den sekundären Rohdaten gearbeitet werden.

Primärdaten werden extra zur Untersuchung des Forschungsproblems durch den For- scher erhoben (folgend: Churchill 1991, S. 305 f.). Typische Primärdaten sind sozioökonomische oder demographische Daten (Alter, Geschlecht, Beruf etc.), psychologische Eigenschaften (Persönlichkeit, Verhalten), Einstellungen und Meinungen, Motivation, Verhalten, Kenntnis von/Wissen bzgl. eines bestimmten Objekts oder Phänomens und Absichten bzgl. zukünftigen Verhaltens. Primärdaten lassen sich durch Beobachtung und Befragung gewinnen.

Bei Befragung werden Untersuchungsteilnehmer direkt zu dem interessierenden Sachverhalt interviewt. Befragungen unterscheiden sich nach Standardisierungsgrad, Methode und Geheimhaltung des Befragungsziels. In Bezug auf Standardisierung lassen sich der standardisierte Fragebogen mit geschlossenen (festen) Antworten an einem Pol und das unstrukturierte Interview mit offenen Fragen am anderen Pol unterscheiden. Typische Methoden der Befragung sind das persönliche Interview, das Telefoninterview, die Verschickung eines Fragebogens und das computergestützte Interview. Das Ziel der Befragung kann dem Befragten dabei entweder mitgeteilt oder vor ihm geheim gehalten werden.

Die Beobachtung beinhaltet keine Befragung der Studienteilnehmer, sondern eine Untersuchung der interessierenden Situation auf bestimmte Tatsachen. Beobachtungen können in kontrollierter Umgebung (Laborexperiment) oder in natürlicher Umgebung durchgeführt werden. Je nach Verständnis des zu beobachtenden Phänomens kann weiter zwischen strukturierter und unstrukturierter Beobachtung unterschieden werden. Im Gegensatz zu letzterer weiß der Beobachter bei strukturierten Beobachtungen, wonach er Ausschau hält und wie er seine Beobachtungen zu klassifizieren und zu bewerten hat. Weiter können Beobachtungen dahingehend unterschieden werden, ob das Beobach-

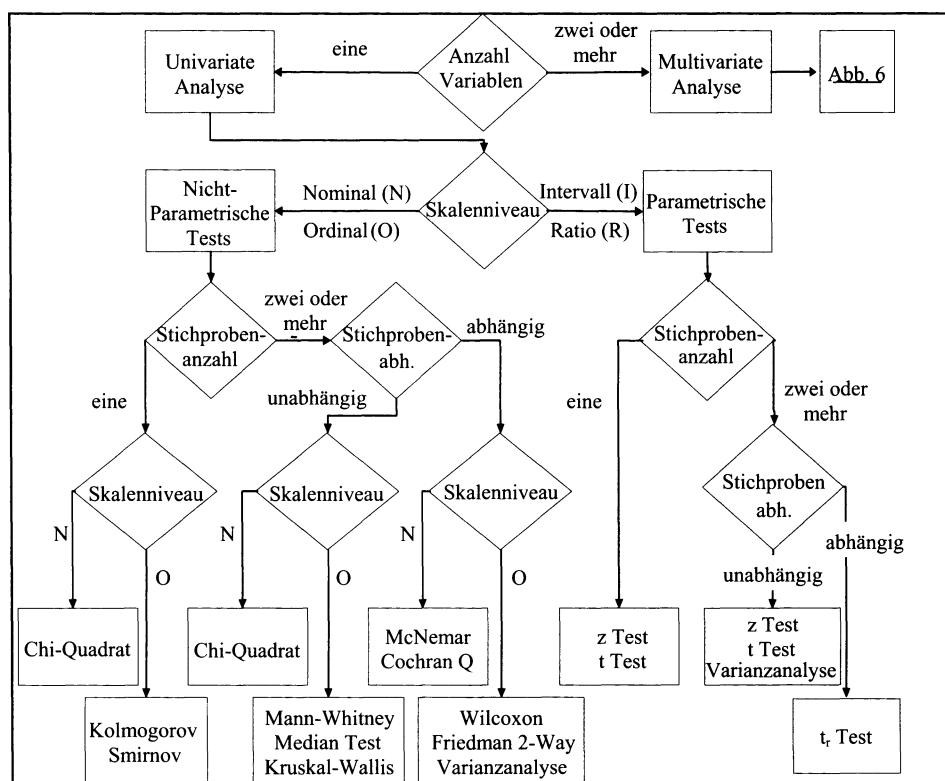
tungsobjekt von seiner Teilnahme an der Untersuchung weiß oder nicht.

Während Sekundärdaten mit Fit- und Genauigkeitsproblemen behaftet sind, ergeben sich je nach Erhebungsdesign bei Primärdaten Probleme durch verzerrnde Einflüsse. Diese können beispielsweise in der missverständlichen Formulierung von Fragen, in der Interaktion mit den Befragungsteilnehmern oder in der Interpretation von Beobachtungen durch den Wissenschaftler liegen (für weitere Ausführungen zum Thema siehe den Beitrag von Kaya zu Verfahren der Datenerhebung in diesem Buch).

4.5 Auswertungsmethoden

Je nach Art der Forschungshypothese und der zugrunde liegenden Daten der Untersuchung eignen sich unterschiedliche Verfahren zum Test der statistischen Hypothese. Die angeführten Tests untersuchen, ob das Ergebnis der Stichprobe zu einem zuvor definierten Signifikanzniveau auf die Grundgesamtheit übertragbar ist. Die Auswahl erfolgt entlang der Arme der Flussdiagramme in Abbildung 1.5 und Abbildung 1.6.

Abbildung 1.5: Flussdiagramm zur Auswahl univariater Testmethoden



Quelle: Churchill 1991, S. 755.