

Reinhard Wittenberg,  
Hans Cramer | Basha Vicari

# Datenanalyse mit IBM SPSS Statistics

Eine syntaxorientierte  
Einführung

UTB 4225



### **Eine Arbeitsgemeinschaft der Verlage**

Böhlau Verlag · Wien · Köln · Weimar

Verlag Barbara Budrich · Opladen · Toronto

facultas.wuv · Wien

Wilhelm Fink · Paderborn

A. Francke Verlag · Tübingen

Haupt Verlag · Bern

Verlag Julius Klinkhardt · Bad Heilbrunn

Mohr Siebeck · Tübingen

Nomos Verlagsgesellschaft · Baden-Baden

Ernst Reinhardt Verlag · München · Basel

Ferdinand Schöningh · Paderborn

Eugen Ulmer Verlag · Stuttgart

UVK Verlagsgesellschaft · Konstanz, mit UVK/Lucius · München

Vandenhoeck & Ruprecht · Göttingen · Bristol

vdf Hochschulverlag AG an der ETH Zürich

## Die Autoren:

**Dipl.-Soz. Dr. rer. soc. Reinhard Wittenberg** ist Akad. Direktor i. R. am Lehrstuhl für Soziologie und Empirische Sozialforschung der Friedrich-Alexander-Universität Erlangen-Nürnberg.

*Kontakt:*

*Dr. Reinhard Wittenberg  
Lehrstuhl für Soziologie und Empirische Sozialforschung  
Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU)  
Findelgasse 7/9  
90402 Nürnberg  
reinhard.wittenberg@wiso.uni-erlangen.de*

**Dipl.-Ing. Hans Cramer** nimmt auch im Ruhestand regelmäßig Lehraufträge für SPSS-Kurse wahr. Von 1995 bis 2007 war er Leiter der Abteilung »Unterstützung dezentraler Systeme« mit Software-Beschaffung/ -Verteilung/ -Support am Regionalen Rechenzentrum Erlangen (RRZE) der Friedrich-Alexander-Universität Erlangen-Nürnberg.

*Kontakt:*

*Dipl.-Ing. Hans Cramer  
Zur Alten Burg 28  
91085 Weisendorf  
hans.cramer@gmx.net  
ehemals:  
Regionales Rechenzentrum Erlangen (RRZE)  
Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU)*

**Dipl.-Sozialw. (Univ.) Basha Vicari** ist seit 2010 Doktorandin im Graduiertenprogramm des Instituts für Arbeitsmarkt- und Berufsforschung in Nürnberg und dort seit 2012 auch als wissenschaftliche Mitarbeiterin im Bereich der Berufsforschung tätig.

*Kontakt:*

*Dipl.-Sozialw. Basha Vicari  
Institut für Arbeitsmarkt- und Berufsforschung (IAB)  
Forschungsgruppe Berufliche Arbeitsmärkte  
Regensburger Str. 104  
90478 Nürnberg  
basha.vicari@iab.de*

Reinhard Wittenberg,  
Hans Cramer,  
Basha Vicari

# Datenanalyse mit IBM SPSS Statistics

**Eine syntaxorientierte Einführung**

UVK Verlagsgesellschaft mbH · Konstanz  
mit UVK Lucius · München

Zusatzmaterialien zum Buch finden Sie unter  
<http://www.utb-shop.de/9783825242251>

Online-Angebote oder elektronische Ausgaben sind erhältlich unter  
[www.utb-shop.de](http://www.utb-shop.de).

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.

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 insbesondere für  
Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und  
Verarbeitung in elektronischen Systemen.

© UVK Verlagsgesellschaft mbH, Konstanz und München 2014

Einbandgestaltung: Atelier Reichert, Stuttgart  
Layout: Gesetzt mit LaTeX  
Druck und Bindung: fgb · freiburger graphische betriebe, Freiburg

UVK Verlagsgesellschaft mbH  
Schützenstr. 24 · 78462 Konstanz  
Tel. 07531-9053-0 · Fax 07531-9053-98  
[www.uvk.de](http://www.uvk.de)

UTB Nr. 4225  
ISBN 978-3-8252-4225-1

## Vorwort

Das vorliegende Buch soll an die selbständige Datenanalyse mit IBM SPSS Statistics – dem weltweit verbreitetsten Programmsystem für angewandte Statistik – heranführen. Es will darüber hinaus grundlegende, für die Datenanalyse notwendige Statistikkennntnisse vermitteln, ohne jedoch vom Leser vertiefte mathematische Vorkenntnisse zu erwarten. Beschrieben und für die exemplarischen Analyseverfahren verwendet wird die deutschsprachige Programmversion IBM SPSS Statistics 22.

Großen Wert legen wir auf die detaillierte Vorstellung und Erörterung von generell in den Erfahrungswissenschaften und in der außeruniversitären Praxis häufig benutzten Verfahren zur explorativen, deskriptiven und analytischen Datenauswertung. Die im Buchtitel genannte „Syntaxorientierung“ hat dabei den Vorteil, dass die Nutzer von vornherein und durchgängig lernen, ihre Analysen effizient, leicht reproduzierbar und schließlich sogar automatisierbar zu gestalten. Zahlreiche Übungsaufgaben und deren Lösungen intensivieren den Lernprozess.

Für die im Buch demonstrierten Auswertungen verwenden wir einen Auszug aus einem öffentlich zugänglichen Datensatz des Erhebungsprogramms „Allgemeine Bevölkerungsumfrage der Sozialwissenschaften“ (ALLBUS). Anhand des ALLBUS 2012 wird idealtypisch das Vorgehen bei der Durchführung eines Forschungsvorhabens durchgespielt, so dass die Leser nicht nur etwas über Datenanalyse im engeren Sinne erfahren, sondern auch über andere Phasen im Ablauf von Forschungsvorhaben: über Planung eines Projekts, Datenerhebung, Codieren und Bereinigen von Daten sowie über Ergebnispräsentation mittels Tabellen und Grafiken.

Informationen zu diesem Buch und zu den verfügbaren Zusatzmaterialien werden im Online-Shop des UTB-Verlags angeboten. Dort findet der Leser z. B. Fragebogen und Dateien des ALLBUS 2012-Datensatzes, die Lösungen zu den Übungsaufgaben, sonstige Ergänzungen und Aktualisierungen zum vorliegenden SPSS Statistics-Buch sowie, wenn nötig, Errata. Des Weiteren besteht dort die Möglichkeit, Fragen zum Buch zu stellen und es zu bewerten:

⇒<http://www.utb-shop.de/9783825242251>

Die Autoren, die zum Teil jahrzehntelange Erfahrung in der Durchführung von Lehrveranstaltungen zu den Methoden der empirischen Sozialforschung und

von SPSS-Kursen aufweisen, wollen die Leser dabei unterstützen, selbstständig ein Forschungsprojekt von Anfang bis Ende durchzuführen. Mit dem vorliegenden Buch nehmen wir den vielfach geäußerten Vorschlag auf, die von uns in mehreren Auflagen verfassten Handbücher „Grundlagen computerunterstützter Datenanalyse“ (Wittenberg 1998) und „Datenanalyse mit SPSS für Windows“ (Wittenberg und Cramer 2003) zu aktualisieren und in einem Band zu integrieren.

Ein herzlicher Dank geht an Sonja Rothländer im Lektorat von UVK, die uns bei der Erstellung des Buches bestens beraten hat, sowie an Daniel Bela und Rüdiger Vicari: Beide haben geholfen, einige komplizierte Hürden, die uns das Textsatzsystems  $\text{\LaTeX}$  in den Weg gestellt hatte, elegant zu überwinden. Bernhard Schrauth war, seinerzeit noch als studentische Hilfskraft, bei der Verifizierung der im Online-Zusatzmaterial zu diesem Buch vorgestellten Verfahren, Tests und Koeffizienten sowie statistischen Formeln behilflich; Dagny Cramer schließlich hat das gesamte Werk Korrektur gelesen – auch ihnen unser herzlicher Dank für die Unterstützung! Sämtliche verbliebenen Mängel gehen selbstverständlich zu Lasten der Autoren, die offen für Kritik und Vorschläge sind, wie das Arbeitsbuch für zukünftige Auflagen ergänzt und verbessert werden könnte.

Nürnberg, im September 2014

Reinhard Wittenberg, Hans Cramer und Basha Vicari

# Inhaltsverzeichnis

<b>Vorwort</b>	<b>v</b>
<b>1 Grundlagen computerunterstützter Datenanalyse</b>	<b>1</b>
1.1 Datenanalyse als Teil eines umfassenden Forschungsprozesses . . .	1
1.2 Programme für die computerunterstützte Datenanalyse . . . . .	4
<b>2 Der Beispieldatensatz ALLBUS 2012</b>	<b>7</b>
2.1 Das ALLBUS-Programm . . . . .	7
2.1.1 Ziele des ALLBUS-Programms . . . . .	8
2.1.2 Schwerpunktthemen des ALLBUS-Programms . . . . .	8
2.1.3 Auswahlverfahren der ALLBUS-Studien . . . . .	9
2.2 Beschreibung des Beispieldatensatzes ALLBUS 2012 . . . . .	10
2.2.1 Ausgewählte Themenbereiche und Variablen . . . . .	11
2.2.2 Ausgewählte Fragen aus dem ALLBUS 2012-Fragebogen .	13
2.2.3 Codierplan . . . . .	21
2.3 Datensicherung und Datenschutz . . . . .	28
<b>3 SPSS Statistics-Grundlagen</b>	<b>31</b>
3.1 SPSS Statistics im Überblick . . . . .	32
3.1.1 SPSS Statistics starten . . . . .	33
3.1.2 Fenster und Hauptmenüs . . . . .	34
3.1.2.1 Dateneditor . . . . .	36
3.1.2.2 Viewer . . . . .	38
3.1.2.3 Pivot-Tabellen-Editor . . . . .	40
3.1.2.4 Diagrammeditor . . . . .	42
3.1.2.5 Textausgabe-Editor . . . . .	43
3.1.2.6 Syntaxeditor . . . . .	44
3.1.2.7 Fensterverwaltung . . . . .	47
3.1.3 SPSS Statistics beenden . . . . .	48
3.1.4 Dateien . . . . .	49
3.1.5 Programmoptionen . . . . .	50
3.1.5.1 Optionen: Allgemein . . . . .	50
3.1.5.2 Optionen: Sprache . . . . .	51
3.1.5.3 Optionen: Viewer . . . . .	52



	3.1.5.4	Optionen: Ausgabe . . . . .	53
	3.1.5.5	Optionen: Dateispeicherorte . . . . .	54
3.1.6		Drucken . . . . .	55
	3.1.6.1	Viewer . . . . .	55
	3.1.6.2	Dateneditor . . . . .	56
	3.1.6.3	Syntaxeditor . . . . .	56
3.2		Eine SPSS Statistics-Beispielsitzung . . . . .	56
	3.2.1	Lesen einer SPSS Statistics-Datendatei . . . . .	57
	3.2.2	Aufruf einer Statistikprozedur . . . . .	58
	3.2.3	Die Ergebnisse im Ausgabefenster . . . . .	60
	3.2.4	Die Ergebnisse drucken und weiterverarbeiten . . . . .	62
3.3		Grundbegriffe . . . . .	63
	3.3.1	Datenmatrix . . . . .	63
	3.3.2	Rohdaten . . . . .	63
	3.3.3	Fälle, Variablen und Werte . . . . .	64
3.4		Datendateien . . . . .	64
	3.4.1	Erstmalige Datenerfassung . . . . .	65
	3.4.2	Rohdatendateien lesen . . . . .	65
		3.4.2.1 Rohdatendateien in festem Format . . . . .	65
		3.4.2.2 Rohdatendateien in freiem Format . . . . .	66
		3.4.2.3 Rohdatendateien mit Tabulatoren oder speziellen Trennzeichen . . . . .	66
	3.4.3	Datenbankdateien lesen . . . . .	67
	3.4.4	SPSS Statistics-Datendateien . . . . .	67
		3.4.4.1 SPSS Statistics-Datendateien schreiben . . . . .	67
		3.4.4.2 SPSS Statistics-Datendateien lesen . . . . .	68
	3.4.5	Datendateien schreiben . . . . .	68
3.5		Die SPSS Statistics-Befehlssprache . . . . .	69
	3.5.1	SPSS Statistics-Befehlsdateien . . . . .	69
		3.5.1.1 Neues Syntaxfenster öffnen . . . . .	70
		3.5.1.2 Befehlsdatei speichern und öffnen . . . . .	70
		3.5.1.3 SPSS Statistics-Befehle ausführen . . . . .	70
	3.5.2	Elemente und Syntax der Befehle . . . . .	71
		3.5.2.1 Unterbefehle . . . . .	72
		3.5.2.2 Variablennamen . . . . .	72
		3.5.2.3 Werte . . . . .	73
		3.5.2.4 Schlüsselwörter . . . . .	74
		3.5.2.5 Zeichenfolgen . . . . .	74
		3.5.2.6 Arithmetische Operatoren . . . . .	74
		3.5.2.7 Trennzeichen . . . . .	75

3.5.3	Variablen . . . . .	75
3.5.3.1	Variablentypen . . . . .	76
3.5.3.2	Variablengruppen . . . . .	76
3.5.3.3	Variablenlisten . . . . .	77
3.5.4	Grundlegende SPSS Statistics-Befehle . . . . .	79
3.5.5	Reihenfolge der SPSS Statistics-Befehle . . . . .	80
3.5.6	Hilfsbefehle . . . . .	80
3.5.6.1	COMMENT . . . . .	80
3.5.6.2	TITLE und SUBTITLE . . . . .	81
3.5.6.3	INSERT . . . . .	82
3.5.6.4	EXECUTE . . . . .	82
3.6	Datenbeschreibung und Datenerfassung . . . . .	83
3.6.1	Befehle zur Datenbeschreibung . . . . .	84
3.6.1.1	DATA LIST . . . . .	84
3.6.1.2	MISSING VALUES . . . . .	86
3.6.1.3	VARIABLE LEVEL . . . . .	88
3.6.1.4	VARIABLE LABELS . . . . .	89
3.6.1.5	VALUE LABELS . . . . .	90
3.6.2	Die Befehle zur Beschreibung des ALLBUS-Datensatzes . . . . .	92
3.6.2.1	DATA LIST . . . . .	92
3.6.2.2	Fehlende Werte, Messniveau und Beschriftungen . . . . .	94
3.6.3	Datenbeschreibung mit dem Dateneditor . . . . .	97
3.6.3.1	Datenbeschreibung . . . . .	98
3.6.3.2	Dateneingabe . . . . .	101
3.7	Datentransformation . . . . .	103
3.7.1	RECODE . . . . .	104
3.7.2	COMPUTE . . . . .	107
3.7.3	IF . . . . .	110
3.7.4	WEIGHT . . . . .	112
3.7.5	COUNT . . . . .	113
3.8	Datenauswahl . . . . .	114
3.8.1	SELECT IF . . . . .	114
3.8.2	SAMPLE . . . . .	115
3.9	SPSS Statistics-Datendateien . . . . .	116
3.9.1	SPSS Statistics-Datendateien schreiben (SAVE) . . . . .	117
3.9.2	SPSS Statistics-Datendateien lesen (GET) . . . . .	118
3.9.3	Variableninformationen . . . . .	119
3.9.3.1	SYSFILE INFO . . . . .	119
3.9.3.2	DISPLAY . . . . .	120
3.9.4	Daten auflisten (LIST) . . . . .	121

3.9.5	SPSS Statistics-Datendateien manipulieren . . . . .	122
3.9.5.1	SPSS Statistics-Datendateien sortieren . . . . .	122
3.9.5.2	SPSS Statistics-Datendateien aufteilen . . . . .	122
<b>4</b>	<b>Theoretische und methodische Vorarbeiten der Datenanalyse</b>	<b>125</b>
4.1	Erkenntnisinteresse . . . . .	125
4.2	Literaturrecherchen . . . . .	127
4.3	Theoretischer Analyserahmen . . . . .	128
4.4	Hypothesenbildung und -prüfung . . . . .	131
4.5	Begriffe, Indikatoren und Operationalisierung . . . . .	134
<b>5</b>	<b>Univariate Analyseschritte</b>	<b>139</b>
5.1	Univariate explorative Datenanalyse . . . . .	140
5.1.1	Datenprüfung und Datenbereinigung . . . . .	140
5.1.1.1	Statistikprozeduren DESCRIPTIVES und FRE- QUENCIES . . . . .	142
5.1.2	Skalenniveau und Verteilungsform der Variablen . . . . .	143
5.1.2.1	Statistikprozedur EXAMINE . . . . .	153
5.1.2.2	Statistikprozedur NPAR TESTS . . . . .	161
5.1.3	Univariate „Ausreißer“-Analyse . . . . .	162
5.1.4	Univariate „item nonresponse“-Analyse . . . . .	167
5.2	Univariate deskriptive und konfirmative Datenanalyse . . . . .	170
5.3	Übungsaufgaben . . . . .	186
5.3.1	Normalverteilung . . . . .	186
5.3.2	Nettoeinkommen nach Geschlechtszugehörigkeit . . . . .	186
5.3.3	Berechnung eines „objektiven Schichtindex“ . . . . .	187
<b>6</b>	<b>Bivariate Analyseschritte</b>	<b>189</b>
6.1	Bivariate explorative Datenanalyse . . . . .	189
6.2	Bivariate deskriptive und konfirmative Datenanalyse . . . . .	194
6.2.1	Voraussetzungen bei der Anwendung statistischer Signi- fikanztests . . . . .	195
6.2.2	Variablenzusammenhänge . . . . .	198
6.2.2.1	Kontingenztabellen und die Statistikprozedur CROSSTABS . . . . .	200
6.2.2.2	Korrelationen und Regressionen sowie die Sta- tistikprozeduren REGRESSION, CORRELA- TIONS und NONPAR CORR . . . . .	217
6.2.3	Variablenunterschiede . . . . .	229
6.2.3.1	Mittelwerte und die Statistikprozeduren T-TEST und NPAR TESTS M-W . . . . .	231

6.2.3.2	Streuungen und die Statistikprozeduren ONE-WAY und NPAR TESTS K-W . . . . .	237
6.3	Übungsaufgaben . . . . .	247
6.3.1	Kreuztabellenanalyse des Zusammenhangs zwischen ausgewählten Anomievariablen . . . . .	247
6.3.2	Kreuztabellenanalyse des Zusammenhangs zwischen Schulabschluss und Schichtzugehörigkeit . . . . .	248
6.3.3	Lineare Regression des Haushaltsnettoeinkommens auf das Lebensalter . . . . .	248
6.3.4	Korrelationsanalyse antisemitischer Einstellungen . . . . .	248
6.3.5	Rangkorrelationsanalyse antisemitischer Einstellungen . . . . .	249
6.3.6	Unterschiede im Haushaltsnettoeinkommen in Abhängigkeit vom Lebensalter: T-TEST . . . . .	249
6.3.7	Unterschiede im Haushaltsnettoeinkommen in Abhängigkeit von der Haushaltsgröße: NPAR TESTS K-W . . . . .	249
<b>7</b>	<b>Multivariate Analyseschritte</b>	<b>251</b>
7.1	Multivariate explorative Datenanalyse . . . . .	251
7.1.1	Multivariate Kasten- und Streudiagramme . . . . .	251
7.1.2	Reliabilität, Validität und Homogenität von Indizes und Skalen . . . . .	255
7.1.2.1	Statistikprozedur RELIABILITY . . . . .	257
7.1.2.2	Statistikprozedur FACTOR . . . . .	261
7.2	Multivariate deskriptive und konfirmative Datenanalyse . . . . .	277
7.2.1	Drei-Variablen-Kontingenztabellen und Partialkorrelation . . . . .	277
7.2.2	Multiple lineare Regressionsanalyse . . . . .	282
7.2.3	Logistische Regression und die Statistikprozedur LOGISTIC REGRESSION . . . . .	290
7.3	Übungsaufgaben . . . . .	306
7.3.1	Reliabilitätsanalyse der Anomievariablen . . . . .	306
7.3.2	Faktorenanalyse der Anomievariablen . . . . .	306
7.3.3	Bildung eines additiven Anomie-Indexes . . . . .	306
7.3.4	Multiple OLS-Regression des persönlichen Nettoeinkommens auf verschiedene Prädiktoren . . . . .	307
7.3.5	Logistische Regression von Zukunftsskepsis auf verschiedene Determinanten . . . . .	307
<b>8</b>	<b>Präsentationsgrafiken</b>	<b>309</b>
8.1	Balkendiagramme . . . . .	311
8.2	Kreisdiagramme . . . . .	320

---

8.3	Streudiagramme . . . . .	322
8.4	Editieren von Diagrammen . . . . .	330
8.4.1	Balkendiagramme . . . . .	331
8.4.2	Kreisdiagramme . . . . .	334
8.4.3	Streudiagramme . . . . .	339
8.5	Übungsaufgaben . . . . .	341
8.5.1	Balkendiagramm . . . . .	341
8.5.2	Balkendiagramm bearbeiten . . . . .	341
8.5.3	Kreisdiagramm . . . . .	341
8.5.4	Kreisdiagramm bearbeiten . . . . .	342
8.5.5	Streudiagramm . . . . .	342
8.5.6	Streudiagramm bearbeiten . . . . .	342
<b>9</b>	<b>Anhang</b>	<b>343</b>
9.1	Download des Zusatzmaterials . . . . .	343
9.2	Arbeiten mit der SPSS Statistics-Datendatei . . . . .	344
9.3	Erstellen der SPSS Statistics-Datendatei . . . . .	345
	<b>Literaturverzeichnis</b>	<b>347</b>
	<b>Sachverzeichnis</b>	<b>351</b>

# Abbildungsverzeichnis

4.1	Pfadmodell zur Erklärung der gesellschaftlichen Rolle der Frau . . .	130
5.1	Normalverteilung . . . . .	151
5.2	Unimodale und bimodale Verteilung . . . . .	152
5.3	Schmalgipflige und breitgipflige Verteilung . . . . .	152
5.4	Linkssteile und rechtssteile Verteilung . . . . .	152
5.5	Abfallende und u-förmige Verteilung . . . . .	153
6.1	Linearer und kurvenlinearer bivariater Zusammenhang . . . . .	194
6.2	Tests zur Überprüfung von Zusammenhangshypothesen . . . . .	199
6.3	Formen bivariater Verteilungen bei gleichen Stichprobengrößen und statistischen Maßzahlen . . . . .	219
6.4	Tests zur Überprüfung von Unterschiedshypothesen . . . . .	232
7.1	Unrotierte Faktorladungen . . . . .	269
7.2	Rotierte Faktorladungen . . . . .	270

## Tabellenverzeichnis

2.1	Codierplan für die ausgewählten ALLBUS 2012-Variablen . . . . .	22
2.2	Rohdatenmatrix – die ersten 30 Fälle . . . . .	27
5.1	Skalierungsarten . . . . .	148
5.2	Häufigkeitstabelle des Lebensalters der Befragten . . . . .	172
5.3	Maßzahlen zur Beschreibung univariater Häufigkeitsverteilungen und vorausgesetztes Skalenniveau . . . . .	175
5.4	Sprachregelungen für die Signifikanz von Hypothesentests . . . . .	180
6.1	Signifikanztests . . . . .	196
6.2	Formen statistischer Zusammenhangshypothesen . . . . .	200
6.3	Schicht-Selbsteinstufung in Abhängigkeit vom allgemeinen Schul- abschluss – <i>Kontingenztafel</i> . . . . .	202
6.4	Schicht-Selbsteinstufung in Abhängigkeit vom allgemeinen Schul- abschluss - <i>Indifferenztafel</i> . . . . .	205
6.5	Arbeitstabelle zur Berechnung des $\chi^2$ - Wertes . . . . .	206
6.6	Sprachregelungen für die Stärke von Zusammenhängen . . . . .	210
7.1	Cronbach's Alpha . . . . .	256

# 1 Grundlagen computerunterstützter Datenanalyse

Egal, ob man nur die tägliche Zeitungslektüre kritisch beurteilen oder sich der Prüfung und Weiterentwicklung von sozialwissenschaftlichen Theorien widmen möchte: Empirischen Daten über gesellschaftliche Zusammenhänge und Entwicklungen begegnet man in allen Themenbereichen der Gesellschaft. Um die Qualität der dort präsentierten Ergebnisse sicher beurteilen, aber auch um eigene fundierte Ergebnisse produzieren zu können, sind Kenntnisse der Methoden der empirischen Sozialforschung unumgänglich. Dieses Buch will die Leser mit den Grundlagen der *quantitativen* Datenanalysemethoden soweit vertraut machen, dass sie zur selbständigen Untersuchung einfacher Forschungsfragen oder zur kritischen Beurteilung veröffentlichter Texte in der Lage sind.

## 1.1 Datenanalyse als Teil eines umfassenden Forschungsprozesses

Die statistische Auswertung von Daten, kurz Datenanalyse, ist in einer empirischen Untersuchung nur eine von vier interdependenten Phasen. In Anlehnung an von Alemann (1991) und Diekmann (2007: 192f.) bezeichnen wir diese vier sachlich und zeitlich aufeinander aufbauenden und miteinander eng verzahnten Phasen als Definitions-, Erhebungs-, Analyse- und Verwertungsphase des Forschungsprozesses.<sup>1</sup>

1. In der **Definitionsphase** geht es um die Klärung der Fragestellung und Zielsetzung der geplanten Untersuchung, die Erarbeitung des jeweiligen theoretischen Bezugsrahmens und die Aufstellung der zu überprüfenden Hypothesen sowie um die Operationalisierung der verwendeten Begriffe. Ferner folgen die Bestimmung von Grundgesamtheit sowie Erhebungs- und Untersuchungseinheiten (z. B. Einzelpersonen, Haushalte, Betriebe) und schließlich die Aufstellung eines Forschungsplans.
2. In der **Erhebungsphase** ist zunächst das Datenerhebungsinstrument (z. B. der Fragebogen für Umfragen oder der Leitfaden für Gruppendiskussionen) zu entwickeln und gegebenenfalls ein Plan zur Auswahl der jeweiligen

---

<sup>1</sup> Der an mehr Informationen über diese Arbeitsschritte interessierte Leser sei auf die ausführliche Darstellung bei von Alemann (1991) hingewiesen.



Erhebungseinheiten (z. B. einfache Zufallsauswahl, geschichtete Klumpenziehung) aufzustellen. Danach muss das Datenerhebungsinstrument einem Pretest unterzogen und entsprechend den Ergebnissen nachgebessert werden. Vorbereitung und Durchführung der Hauptuntersuchung – die sog. Feldphase – bilden den Abschluss der Erhebungsphase.

3. In der **Analysephase** geht es um die Codierung der erhobenen Daten und ihre Übertragung auf digitale Datenträger. Danach folgen Datenprüfung und -korrektur sowie die Einrichtung von sog. Dateien (engl.: „files“) und ihre Sicherung gegen Verlust, Zerstörung oder Missbrauch durch Unbefugte. Erst nach Beendigung dieser eher technischen Arbeiten kann mit der eigentlichen explorativen, deskriptiven und konfirmativen Auswertung in Form von uni-, bi- und multivariater Datenanalyse begonnen werden. Die Datenanalyse sollte dabei tunlichst einem detaillierten Auswertungsplan folgen, der in Abhängigkeit von Fragestellung, theoretischem Bezugsrahmen und eingesetzten Messmethoden aufgestellt wird.
4. In der **Verwertungsphase** als letztem Abschnitt einer empirischen Untersuchung geht es v. a. darum, die mittels der Datenanalyse erzielten Ergebnisse im Hinblick auf die explizite Fragestellung und deren theoretische Fundierung zu interpretieren – und zu Papier zu bringen. Ob sich dies im Schreiben von Forschungsberichten und Gutachten erschöpft oder auch in Publikationen mündet und ob daraus zudem spezifische Handlungsempfehlungen abgeleitet werden können, hängt insbesondere mit der jeweiligen Art der Forschung zusammen: „Freie“ Forschung steht hier gegen Auftragsforschung, „angewandte“ gegen Grundlagenforschung, Abschlussarbeit gegen Artikel in wissenschaftlicher Zeitschrift, um nur einiges zu nennen, was den Stellenwert von Forschungsergebnissen beeinflusst.

Die Darstellung des Ablaufs von Forschungsprozessen ist keineswegs normativ zu verstehen; sie soll der Verortung der **Datenanalysephase** innerhalb des umfassenden Untersuchungsprozesses dienen.<sup>2</sup> Selbstverständlich bedingt die Vielfalt von Forschungsgegenständen, Erkenntnisinteressen und -zielen in den verschiedenen Erfahrungswissenschaften auch eine Vielfalt der Forschungsdesigns.<sup>3</sup> Forschungslogik und -organisation einer jeden empirischen Untersuchung werfen jeweils spezifische Probleme auf und verlangen nach Schwerpunktsetzungen auf und innerhalb der Untersuchungsphasen. Sachliche und zeitliche Abweichungen vom oben skizzierten „Modell“ des phasenhaften Forschungsablaufes sind infolgedessen eher Regel als Ausnahme.

<sup>2</sup> Eine grafische Darstellung der aufeinander folgenden Phasen des Forschungsprozesses ist im ⇒*Online-Zusatzmaterial* unter <http://www.utb-shop.de/9783825242251> zu finden.

<sup>3</sup> „Erfahrungswissenschaften“ sind empirische Wissenschaften, die ihre theoretischen Aussagen an der beobachteten Wirklichkeit dahingehend überprüfen, ob sie „wahr“ oder „falsch“, nicht aber, ob sie „wünschenswert“ oder „verwünschenswert“ sind.

Im Fall einer Sekundäranalyse bspw., d. h. der erneuten Analyse des Datenmaterials einer bereits durchgeführten Untersuchung,<sup>4</sup> die wir exemplarisch im vorliegenden Handbuch vorführen wollen, werden die Arbeitsschritte der Erhebungsphase gerade nicht durchgeführt. Im Weglassen dieser Phase liegt ihr besonderer Wert: Mittels Sekundäranalysen können eine Vielzahl interessanter Fragestellungen ohne die ansonsten erforderliche zeit- und kostenaufwendige eigene Feldarbeit angegangen werden.<sup>5</sup>

Bei aller Variation von Forschungsdesigns gilt allerdings ausnahmslos, dass das Endprodukt einer empirischen Untersuchung – das Forschungsergebnis – danach zu bewerten ist, inwieweit jeder durchgeführte Arbeitsschritt für sich und in der Verzahnung mit den anderen Arbeitsschritten als gelungen bezeichnet werden kann. Abgesehen davon, dass selbstverständlich Forschungsprojekte durchgeführt werden, deren Problemstellung in sozialer und/oder wissenschaftlicher Hinsicht bedeutungslos, trivial oder unlösbar ist, werden bereits in der Definitionsphase entscheidende Weichen für das Gelingen des Forschungsunternehmens insgesamt gestellt: Neben der Ausarbeitung des theoretischen Bezugsrahmens ist hier insbesondere die Operationalisierung der verwendeten Begriffe, also die Angabe von Anweisungen, wie die durch einen Begriff bezeichneten Sachverhalte erfasst werden können und sollen,<sup>6</sup> von herausragender Bedeutung für die Entwicklung des Erhebungsinstrumentes und den Verlauf der Erhebungsphase. Die Konstruktion des Erhebungsinstrumentes und eines angemessenen Auswahlplans sowie die Realisierung der Untersuchung im sog. „Feld“ wiederum bestimmen die Rahmenbedingungen, Möglichkeiten und Grenzen der Analysephase.

Dies gilt selbstverständlich ebenso für die Sekundäranalyse. Will der „Sekundäranalytiker“ nicht u. U. ausschließlich Forschungsartefakte produzieren, ist er gehalten, das zur Verfügung stehende Datenmaterial im Hinblick auf seinen Entstehungszusammenhang so genau wie möglich zu recherchieren. Ist dies nicht oder nur unzureichend möglich, weil z. B. der Fragebogen oder das Codebuch nicht verfügbar sind, sollte er besser auf die Sekundäranalyse verzichten.

Das in der Verwertungsphase zu erbringende Resultat einer Forschungstätigkeit ist schließlich danach zu beurteilen, inwieweit die vorgestellten Ergebnisse

---

4 Vgl. zu Sekundäranalysen Hyman (1972), Klingemann und Mochmann (1975), Hakim (1982), Kiecolt und Nathan (1985), Dale et al. (1988), Meulemann (2002) und Lewis-Beck et al. (2004).

5 Der Arbeitsbereich „Datenservice“ des *GESIS* – Leibniz-Institut für Sozialforschung (<http://www.gesis.org/unser-angebot/daten-analysieren/datenservice/>) hat mittlerweile die Daten von mehreren Tausend Untersuchungen aus verschiedenen Bereichen der Sozialwissenschaften für Sekundäranalysen aufbereitet. Es stellt diese Studien Organisationen und Personen, Wissenschaftlern und Studierenden für Forschungszwecke zur Verfügung. Das Datenmaterial selbst wird online und auf DVD-ROM zusammen mit Codebooks bereit gestellt.

6 Einführend zu Problemen der Operationalisierung vgl. z. B. Bortz und Döring (2005: Kap. 2.3.5), Diekmann (2007: Kap. IV und V), Kromrey (2006: Kap. 3 und 4) oder Schnell et al. (2013: Kap. 4).

mit Art und Durchführung der vorangehenden Untersuchungsphasen und Arbeitsschritte vereinbar sind.<sup>7</sup>

Wenn wir uns in dieser Einführung nun nahezu ausschließlich mit der Phase der Datenanalyse beschäftigen – und dabei die „inhaltlich-substanziellen“ Gesichtspunkte des Umgangs mit Datenanalyzesystemen weitgehend zugunsten der gründlichen Behandlung der eher „mathematisch-statistisch-technischen“ Probleme vernachlässigen –, sollten die Anwender die vielfältigen Implikationen der Definitions- und Erhebungsphase für die Analysephase auf keinen Fall außer Acht lassen, wenn sie Auswertungen an Daten vornehmen, die sie selbst nicht erhoben haben.

Über den von uns zur (Sekundär-)Analyse von Daten zusammengestellten Beispieldatensatz namens **ALLBUS 2012** – eine gekürzte Fassung des im zweijährigen Rhythmus erhobenen ALLBUS („**ALL**gemeine **B**evölkerungs**U**mfrage der **S**ozialwissenschaften“) aus dem Jahr 2012 – versuchen wir in Kapitel 2 so zu informieren, dass ein Einblick in seinen Entstehungszusammenhang gewonnen werden kann. Erst danach widmen wir uns im Kapitel 3 der Vorstellung der Grundlagen des Statistikpakets IBM SPSS Statistics 22 inklusive seiner Befehlssyntax, gefolgt von einem kurzen Überblick zu theoretischen und methodischen Vorarbeiten der Datenanalyse im Kapitel 4, bevor dann, mit Kenntnissen über die zu untersuchenden Daten und über die Befehlssprache des Statistikpakets ausgestattet, mit der Beschreibung der eigentlichen Datenanalyse begonnen wird. Die in den Kapiteln 5, 6 und 7 vorgestellte Datenanalyse befasst sich mit der Beschreibung der explorativen sowie der deskriptiven und konfirmativen Datenanalyse-schritte. Abschließend wird im Kapitel 8 noch die Erstellung von Präsentationsgrafiken beschrieben.

Zuvor wollen wir aber im folgenden Unterkapitel einen kleinen Überblick über Programme für die computerunterstützte Datenanalyse geben.

## 1.2 Programme für die computerunterstützte Datenanalyse

Früher mussten Wissenschaftler oder Praktiker für die computerunterstützte Datenanalyse entweder selbst über Programmierkenntnisse verfügen oder sich der Hilfe von Programmierern bedienen, um ein für die Lösung ihres statistischen Problems geeignetes Einzelprogramm zu schreiben oder schreiben zu lassen. Programme sind dabei eine Sequenz aus Instruktionen, mit Hilfe derer die Verarbeitung von Daten im Computer gesteuert wird.

---

<sup>7</sup> Kriterien zur Bewertung eines soziologischen Forschungsberichtes inklusive des ihm unterliegenden Forschungsprozesses listet Friedrichs (1990: 394–399) in Anlehnung an ein Bewertungsschema der American Sociological Association auf.

**Datenanalyzesysteme** stellen eine spezialisierte Alternative für die umständliche Handhabung vieler Einzelprogramme dar. Sie können als ein **Paket** beschrieben werden, in dem eine Vielzahl von (Einzelzweck-)Programmen zusammengeschnürt ist. Sie zeichnen sich jedoch nicht nur durch das umfassende Angebot an statistischen Prozeduren aus, sondern darüber hinaus durch weit reichende Möglichkeiten zur Datenselektion und -transformation sowie zur Ergebnisdarstellung in Form von Tabellen und Grafiken. Datenanalyzesysteme verlangen vom Anwender prinzipiell keine Kenntnisse von Programmiersprachen; vielmehr können auch Anfänger nach Erlernen einer einfachen, (meist englischen) Befehlsprache selbstständig Auswertungen vornehmen. Neben IBM SPSS Statistics sind Microsoft Excel, Stata, SAS und R die in den Erfahrungswissenschaften am häufigsten verwendeten Analyzesysteme für allgemeine Datenauswertungen.

Die **Vorteile** der Programmsysteme haben seit etwa 1970 zu einem enormen Anstieg der computerunterstützten Datenanalyse in den Erfahrungswissenschaften geführt. Die Verwendung von (Einzelzweck-)Programmen ist dagegen stark zurückgedrängt worden – sie bleibt aber selbstverständlich dort notwendig und sinnvoll, wo neue Verfahren der Datenanalyse entwickelt werden und wo diese Lücken aufweisen oder für bestimmte Fragestellungen (noch) nicht genügend ausgefüllt sind.

Nicht verschwiegen werden sollen an dieser Stelle aber auch einige **Nachteile**, die mit der Verwendung von Datenanalyzesystemen einhergehen (können):

1. Der erste bezieht sich auf die Unvollständigkeit der in den Paketen zusammengeschnürten Einzelprogramme. Diese betrifft sowohl statistische Analyseprogramme insgesamt als auch die Darstellung von Ergebnissen. Fairness halber sei allerdings betont, dass die wohl am häufigsten in den Erfahrungswissenschaften verwendeten statistischen Prozeduren in allen oben genannten Datenanalyzesystemen angeboten werden.
2. Den zweiten Nachteil von Datenanalyzesystemen haben bereits Böhning und Wilke (1980: 293) hervorgehoben. Er bezieht sich auf den normativen Effekt, den die Logik eines Programmsystems auf die Durchführung von Analysen ausüben kann: „Die massive Beeinflussung des Analysegangs durch die Logik des verwendeten Paketes führt den unbedarften Benutzer dazu, die Lösungen seines Paketes mit den Lösungen der empirischen Sozialforschung gleichzusetzen. Er läßt sich in der Formulierung von Fragestellungen, in der Wahl der Verfahren und des Analysestils von Vorgaben des gewählten Pakets leiten.“ Diese Gefahr für die Fortentwicklung empirischer Forschung im Allgemeinen wird umso größer, je marktbeherrschender ein Programmsystem wird. Umso wichtiger ist es sowohl für die empirische Forschung insgesamt, als auch für den einzelnen Wissenschaft-

- ler, wenn möglichst viele alternative Analysepakete auf dem Markt sind – und wenn er gleichzeitig auch ausreichende Kenntnisse über diese hat.
3. Der dritte Nachteil ist u. U. noch schwerwiegender als der zweite: Der Aufwand, eigens Programme für die Datenauswertung und die jeweilige umfangreiche Einleseroutine zu schreiben, führte vor der Verfügbarkeit handlicher Datenanalyseysteme dazu, dass Wissenschaftler und Praktiker sich einen jeden (zusätzlichen) Schritt der Datenanalyse sehr genau vergegenwärtigten und überlegten, ob er zur Überprüfung der einmal getroffenen theoretischen Annahmen sinnvoll und notwendig sowie mit der Struktur der erhobenen Daten überhaupt kompatibel ist. Heutzutage hingegen wird bei der Benutzung von Datenanalyseystemen – die ja so einfach ist, so viele statistische Prozeduren verfügbar macht und alle möglichen Koeffizienten umgehend zu berechnen erlaubt – oftmals allzu leichtfertig vorgegangen: Weder wird bedacht, welche Voraussetzungen die erhobenen Daten für die Durchführung bestimmter statistischer Analyseverfahren aufweisen müssen, noch bleibt die Theoriegeleitetheit der Auswertung unangestastet.<sup>8</sup> Ein derartiges Vorgehen führt jedoch zwangsläufig zur Produktion von Forschungsartefakten. Anfänger wie Fortgeschrittene sollten daher der immensen Verführungskraft von Datenanalyseystemen widerstehen – ein theoriegeleiteter, noch vor der Analysephase ausgearbeiteter und detaillierter Auswertungsplan verschafft die nötige innere Kraft dazu.

---

<sup>8</sup> Vgl. zu derartigen Problemen bspw. Allerbeck (1994: 180ff.; 1977).

## 2 Der Beispieldatensatz ALLBUS 2012

Die Einführung in die praktische Anwendung von IBM SPSS Statistics und die Sprache, Logik, statistischen Prozeduren etc., die diesem Statistikpaket eigen ist, werden nachfolgend an Hand eines einzigen Datensatzes beschrieben, erläutert und diskutiert, nämlich anhand eines von uns zusammengestellten Auszugs aus dem Datensatz ALLBUS („ALLgemeine BevölkerungsUmfrage der Sozialwissenschaften“) aus dem Jahr 2012.<sup>1</sup>

Die ALLBUS 2012-Daten des Beispieldatensatzes werden als ⇒ *Online-Zusatzmaterial* unter <http://www.utb-shop.de/9783825242251> angeboten. Download und Installation sind im Anhang beschrieben. Der Datensatz ALLBUS 2012 wird in den nachfolgenden Abschnitten vorgestellt. Deren Lektüre ist empfehlenswert, enthalten sie doch Hintergrundinformationen, deren Kenntnis zum inhaltlichen und methodischen Verständnis der demonstrierten Vorgehens- und Verfahrensweisen im Rahmen einer Sekundäranalyse erforderlich ist.

### 2.1 Das ALLBUS-Programm

Hinter dem Kürzel ALLBUS<sup>2</sup> verbirgt sich eine multithematische Umfrageserie zu Einstellungen, Verhaltensweisen und Sozialstruktur der Bevölkerung in der Bundesrepublik Deutschland. Organisiert und durchgeführt wird die Umfrageserie vom GESIS – Leibniz-Institut für Sozialwissenschaften, wobei die Datenerhebung selbst durch externe Sozialforschungsinstitute geschieht, zuletzt durch TNS Infratest München. Das ALLBUS-Programm umfasst kontinuierlich wiederholte Bevölkerungsumfragen mit einem teils konstanten, teils variablen Fragenprogramm, das bedeutsame Forschungsbereiche der empirischen Sozialforschung abdecken soll, um vorwiegend für Forschung und Lehre Daten zur Verfügung zu stellen. Bisher sind seit 1980 im zweijährigen Turnus – inklusive einer

---

1 Bei diesem Beispieldatensatz ALLBUS 2012 handelt es sich um eine stark verkürzte und geringfügig modifizierte Fassung der Originalversion, die bei GESIS abgerufen werden kann (vgl. Fußnote 5 auf Seite 3).

2 Informationen zum ALLBUS-Programm findet man bei den Umfragedaten des GESIS – Leibniz-Institut für Sozialwissenschaften unter <http://www.gesis.org/allbus>, bzw. doi:10.4232/1.11753. Eine ausgezeichnete Einführung in das Arbeiten mit ALLBUS-Daten überhaupt geben Porst (2000) sowie Koch und Wasmer (2004). Sämtliche bisher mit ALLBUS-Daten erarbeiteten Publikationen sind online unter <http://www.gesis.org/allbus/recherche/allbus-bibliographie> dokumentiert.

zusätzlichen „Baseline Study“ nach der Vereinigung beider deutscher Teilstaaten im Jahr 1991 – bis heute 18 Erhebungen durchgeführt worden.

### 2.1.1 Ziele des ALLBUS-Programms

Das ALLBUS-Programm dient vor allem vier Zielen:

1. dem wissenschaftlichen Ziel der **Untersuchung des sozialen Wandels**,
2. dem politischen Ziel der beschreibenden **Sozialberichterstattung**,
3. der systematischen **international vergleichenden Gesellschaftsanalyse**,
4. dem für die Sozialwissenschaften wichtigen Ziel der **Datengenerierung** für Studenten und Forscher, die keinen unmittelbaren Zugang zu Primärdaten haben.

Diese Zielsetzungen bestimmen gemeinsam mit der Entscheidung für eine mündliche, z. T. computerunterstützte Umfrage als Datenerhebungsmethode die Entwicklung des Fragenprogramms, an dessen Zustandekommen eine Vielzahl von Sozialwissenschaftlern beteiligt waren und sind.<sup>3</sup>

### 2.1.2 Schwerpunktthemen des ALLBUS-Programms

Jeder ALLBUS weist bestimmte Schwerpunktthemen auf. Neben der ausführlichen GESIS-Standarddemografie,<sup>4</sup> die bei jedem ALLBUS einen großen Teil der Variablen umfasst, sind bei den einzelnen ALLBUS-Umfragen zusätzlich verschiedene Schwerpunktthemen bearbeitet worden.

Jedem Fragenkomplex zu diesen Schwerpunktthemen sind in der Regel mehrere Fragen zugeordnet, die sich möglichst in früheren Umfragen methodisch und inhaltlich bewährt haben, die zum Teil international vergleichbar sind und die mit den anderen Fragen und Fragestellungen des Forschungsprogramms theoretisch sinnvoll in Zusammenhang gebracht werden können (siehe Übersicht der Schwerpunktthemen seit 1980 auf der nächsten Seite).

---

<sup>3</sup> Vgl. zum Folgenden auch Wasmer et al. (2007).

<sup>4</sup> Mittels der GESIS-Standarddemografie werden wichtige demografische und sozialstrukturelle Hintergrundmerkmale der Befragten und ihrer Familienangehörigen wie z. B. Lebensalter, Geschlechtszugehörigkeit, Schul- und Berufsausbildung, Berufstätigkeit und berufliche Situation erfasst. Die „Demografischen Standards“ gehen auf einen Arbeitskreis aus Vertretern des Statistischen Bundesamtes, der Arbeitsgemeinschaft Sozialwissenschaftlicher Institute e. V. (ASI) und des Arbeitskreises Deutscher Markt- und Sozialforschungsinstitute e. V. (ADM) zurück, die sie etwa alle vier Jahre überarbeiten und aktualisieren: <http://www.gesis.org/unser-angebot/studien-planen/demographische-und-regionale-standards>. Vgl. zusätzlich Pappi (1979) und Wasmer et al. (2007).

1980	Politik, Freundschaftsbeziehungen
1982	Religion und Weltanschauung
1984	Soziale Ungleichheit und Wohlfahrtsstaat
1986	Bildung und Kulturfertigkeiten
1988	Einstellungen zum politischen System und politische Partizipation
1990	Sanktionen und abweichendes Verhalten; Politik, Freundschaftsbeziehungen (Replikation ALLBUS 1980)
1991	„DFG-Baseline-Study“: Zusätzliche Erhebung nach der Vereinigung beider deutscher Staaten – Replikation kleinerer Schwerpunkte der bisherigen Umfragen (Familie, Beruf, Ungleichheit, Politik, Freundschaft, Probleme der Vereinigung)
1992	Religion und Weltanschauung (Replikation ALLBUS 1982)
1994	Soziale Ungleichheit und Wohlfahrtsstaat (Replikation ALLBUS 1984)
1996	Einstellungen gegenüber ethnischen Gruppen in Deutschland
1998	Politische Partizipation und Einstellungen zum politischen System
2000	Replikation von Fragen aus dem gesamten bisherigen ALLBUS-Programm
2002	Religion und Weltanschauung (Replikation ALLBUS 1982 und 1992), Werte
2004	Soziale Ungleichheit (Replikation ALLBUS 1984 und 1994), Gesundheit und Digital Divide
2006	Einstellungen gegenüber ethnischen Gruppen in Deutschland (Replikation ALLBUS 1996)
2008	Politische Partizipation und politische Kultur (Replikation ALLBUS 1988 und 1998), Sozialkapital
2010	Replikation aus dem gesamten Frageprogramm, Freundschaftsbeziehungen
2012	Religion und Weltanschauung (Replikation ALLBUS 1982, 1992 und 2002)

### 2.1.3 Auswahlverfahren der ALLBUS-Studien

Als Auswahlverfahren kamen bei den ALLBUS-Untersuchungen bis einschließlich 1992 sowie nochmals 1998 dreistufige Wahrscheinlichkeitsauswahlen zum Einsatz, wobei auf der ersten Stufe Stimmbezirke/Sample-Points, auf der zweiten Stufe Haushalte und auf der dritten Stufe jeweils eine Befragungsperson pro Haushalt zufällig ausgewählt wurden (ADM-Stichprobendesign).<sup>5</sup> Bis 1990 wurden als Grundgesamtheit „alle während des Befragungszeitraumes in Privathaushalten der Bundesrepublik Deutschlands oder Westberlins lebenden deutschen Staatsangehörigen im Alter von mindestens 18 Jahren“ definiert, ab 1991 bilden „alle erwachsenen Personen (Deutsche und deutsch sprechende Ausländer, die in der Bundesrepublik Deutschland (Ost und West) in Privathaushalten wohnen“ die Grundgesamtheit. Zwischen 1994-1996 und seit 2000 werden zweistufige, geschichtete, disproportionale Zufallsstichproben gezogen, wobei auf der

5 Vgl. hierzu ausführlich Wasmer et al. (2007).



ersten Stufe Gemeinden/Sample-Points und auf der zweiten Stufe Personen aus den Einwohnermelderegistern gezogen werden. Im ostdeutschen Erhebungsgebiet wurden – um differenzierte Analysen für einzelne Bevölkerungsumfragen durchführen zu können – seit 1996 ungefähr doppelt so viele Personen befragt wie ihrem Anteil an der Grundgesamtheit entspricht. Diese Überrepräsentation der ostdeutschen Befragten ist bei Auswertungen für Gesamtdeutschland auf Personenebene durch einen entsprechenden Gewichtungsfaktor wieder aufzuheben.<sup>6</sup>

Der Ausschöpfungsgrad – d. h. der Anteil realisierter Interviews an den ursprünglich für die Befragung gezogenen Adressen – betrug 2012 sowohl für Ost- als auch für Westdeutschland 37,6% (vgl. Terwey und Baltzer 2013). Dabei sank der Ausschöpfungsgrad in den letzten Jahren kontinuierlich ab: Bei der ersten Erhebung im Jahr 1980 lag die Quote der realisierten Interviews noch bei 69,5%, 1992 bei 51,9% in West- und 54,7% in Ostdeutschland und ab dem Jahr 2002 unterschritten beide Teile Deutschlands die 50%-Marke.

## 2.2 Beschreibung des Beispieldatensatzes ALLBUS 2012

Im Allgemeinen ist es bei empirischen Untersuchungen – seien es Befragungen, Beobachtungen, Dokumentenanalysen oder Experimente – üblich, Daten mittels der folgenden Arbeitsschritte zunächst zu erheben, dann zu codieren, technisch zu erfassen und schließlich zu digitalisieren:

1. Ein Erhebungsinstrument wird entwickelt, in dem man die wie auch immer erhobenen empirischen Informationen festhält.
2. Ein sogenannter Codierplan wird erstellt, mit dem exakt festgelegt wird, welche dieser erhobenen Informationen *wie* – d. h. mit welchem Schlüssel (engl.: „code“) bzw. mit welchem (alpha-)numerischen Äquivalent –, und *wo* – d. h. an welcher Stelle der Eingabezeile („record“) eines Editors – registriert werden soll. Häufig sind die Anweisungen des Codierplans bereits auf dem Erhebungsinstrument – Fragebogen, Beobachtungsschema, Kategorienschema einer Inhaltsanalyse usw. – mit abgedruckt. So auch beim Fragebogen zum ALLBUS 2012.
3. Nach der Datenerhebung sind nun die im Erhebungsinstrument festgehaltenen Informationen gemäß den Anweisungen des Codierplans auf digitale Datenträger zu übertragen. Erst dann haben wir es mit „Daten“ zu tun, dem Rohmaterial also, das wir computerunterstützt analysieren wollen.
4. Sind die Informationen aus dem Erhebungsinstrument auf einen Datenträger übertragen, kann mit Hilfe des Computers eine Auflistung der auf ihm

---

<sup>6</sup> Der Gewichtungsfaktor beträgt im ALLBUS 2012 für in Ostdeutschland Befragte  $W_O = 0,556878638$ , für in Westdeutschland Befragte  $W_W = 1,210849096$ . Des Weiteren ist eine Ost-West-Gewichtung ausschließlich für die deutsche Bevölkerung möglich.

festgehaltenen Informationen erstellt werden. Wir erhalten eine rechteckige Datenmatrix, bei der in den Spalten die verschiedenen Variablen und in den Zeilen die verschiedenen Fälle (engl.: „cases“) angeordnet sind.

Im hier vorliegenden speziellen Fall stellt sich das Problem der Datenerfassung und -übertragung nicht, da wir es mit der ALLBUS-Studie mit bereits erhobenen und auf Datenträger übertragenen Daten zu tun haben; wir haben mit dem Beispieldatensatz ALLBUS 2012 sogar schon eine beschriebene Datenmatrix der uns interessierenden Daten vorliegen, die wir nun einer **Sekundäranalyse** unterziehen wollen. Was dafür zu tun bleibt, ist bereits oben angeführt worden: Der Beispieldatensatz und das IBM SPSS Statistics-Programm mit den Syntaxbefehlen zur Datenbeschreibung müssen, wie im Anhang beschrieben, aus dem [⇒ Online-Zusatzmaterial](http://www.utb-shop.de/9783825242251) unter <http://www.utb-shop.de/9783825242251> heruntergeladen und installiert werden.

### 2.2.1 Ausgewählte Themenbereiche und Variablen

Unser aus dem Originaldatensatz des ALLBUS 2012 entwickelter Beispieldatensatz sollte mehreren Ansprüchen gleichzeitig genügen:

- Er sollte verschiedene Fragenkomplexe aus dem ALLBUS 2012 repräsentieren, um zu gewährleisten, dass möglichst viele Benutzer ein inhaltlich interessantes Problem angesprochen finden und dieses u. U. auch über die beispielhaften Analyseprozeduren hinaus selbstständig bearbeiten können.
- Er sollte weiterhin so gestaltet sein, dass möglichst viele grundlegende statistische Verfahren explorativer, deskriptiver und konfirmativer Statistik behandelt werden können, ohne dass Skalenniveau und Verteilungsform der ausgewählten Variablen schwerwiegende Verstöße gegen die mathematisch-statistischen Voraussetzungen der Analyseprozeduren in den Beispielen erfordern.
- Er sollte schließlich von der Anzahl der Fälle her so geschnitten sein, dass er es ermöglicht, einen „repräsentativen Schluss“ von der Stichprobe auf die „wahre Verteilung“ in der Grundgesamtheit ziehen zu können.

Aus diesen Anforderungen resultiert ein Datensatz, der in der Variablenzahl im Vergleich zur originären ALLBUS-Studie 2012 zwar stark reduziert ist (82 Variablen aus 752), dafür aber die Antworten aller 3.480 befragten Personen ausschöpft; Repräsentationsschlüsse sind daher für die involvierten Variablen – zumindest von der Stichprobenziehung her – ohne jegliche Einschränkung möglich.

Aus der inhaltlichen Vielfalt der ALLBUS-Studie 2012 sind **sieben Themenbereiche** in unseren Beispieldatensatz übernommen worden: Anomie, Lebenszufriedenheit, Rollenbilder in der Familie, Einstellung zu in Deutschland lebenden Ausländern, Antisemitismus, Selbstverortung in der Gesellschaft sowie politisches Interesse und Wahlbeteiligung. Dazu kommen eine Reihe sogenannter

**soziodemografischer Angaben** aus der GESIS-Standarddemografie sowie **Angaben zum Interview**. Folgende Übersicht gibt über die ausgewählten Themenbereiche näheren Aufschluss.

1. **Anomie**
2. **Lebenszufriedenheit**
3. **Rollenbilder in der Familie**  
Einstellungen zur Rolle der Frau in der Familie  
Einstellungen zur Rolle der Eltern in der Familie
4. **Einstellung zu in Deutschland lebenden Ausländern**  
Diskriminierung  
Kontakt zu Ausländern
5. **Antisemitismus**
6. **Selbstverortung in der Gesellschaft**  
Subjektive SchichtEinstufung  
Links-Rechts-Selbsteinstufung
7. **Politisches Interesse und Wahlbeteiligung**  
Politisches Interesse  
Wahlabsicht  
Parteipräferenz
8. **Soziodemografische und sozialstrukturelle Variablen**  
Geschlechtszugehörigkeit  
Lebensalter  
Schulabschluss  
Staatsangehörigkeit und Staatsbürgerschaft  
Familienstand  
Lebensalter des/r Ehepartners/in  
Lebenspartner und Lebensalter des/r Lebenspartners/in  
Berufstätigkeit, berufliche Stellung und Berufsprestige  
Arbeitsstunden  
Persönliches und Haushaltsnettoeinkommen  
Haushaltsgröße  
Konfession und Kirchengangshäufigkeit  
Wohnungstyp und -gebäude  
Gemeindegröße, Erhebungsgebiet, Bundesland
9. **Daten zum Interview**  
Interviewdauer  
Geschlecht und Alter des/r Interviewers/in  
Schulabschluss des/r Interviewers/in
10. **Personenbezogene Ost-West-Gewichtung**

### 2.2.2 Ausgewählte Fragen aus dem ALLBUS 2012-Fragebogen

Die im Folgenden abgedruckten Auszüge aus dem ALLBUS 2012-Fragebogen entsprechen in der Formulierung der Fragen, Antworten und Listen,<sup>7</sup> nicht jedoch im Layout und der Anzahl der Variablen dem Original. Für einen direkten Vergleich haben wir die Nummerierung der Fragen aus dem Originalfragebogen übernommen. Wurden aus mehreren Fragen des Originalfragebogens neue Fragen zusammengesetzt, wird darauf durch Fußnoten verwiesen. Die Antworten der 3.480 Befragten auf die in ihm enthaltenen Fragen stellen die Basis der Rohdatendatei „allbus2012.dat“ dar. Insgesamt sind aus Platzgründen allerdings nur einige Fragetypen zur Veranschaulichung ausgewählt worden.

---

Spaltennummer in der Rohdatenmatrix:  
Identifikationsnummer des Befragten: Spalten 1-4

#### Anomie

F005. Ich werde Ihnen nun einige Aussagen vorlesen. Sagen Sie mir bitte zu jeder einzelnen dieser Aussagen, ob Sie persönlich derselben Meinung sind oder ob Sie anderer Meinung sind.

⇒ INT.: Vorgaben bitte vorlesen!

	Bin der- selben Meinung	Bin anderer Meinung	wn	kA
A) Egal, was manche Leute sagen: Die Situation der einfachen Leute wird nicht besser, sondern schlechter.	1	2	8	9
B) So wie die Zukunft aussieht, kann man es kaum noch verantworten, Kinder auf die Welt zu bringen.	1	2	8	9
C) Die meisten Politiker interessieren sich in Wirklichkeit gar nicht für die Probleme der einfachen Leute.	1	2	8	9
D) Die meisten Leute kümmern sich in Wirklichkeit gar nicht darum, was mit ihren Mitmenschen geschieht.	1	2	8	9

Spalten 5-8

---

<sup>7</sup> Der Originalfragebogen inklusive der entsprechenden Listen ist im ⇒ *Online-Zusatzmaterial* unter <http://www.utb-shop.de/9783825242251> auf der Webseite des Verlags zu finden.

### Lebenszufriedenheit

S103. Und jetzt noch eine allgemeine Frage. Wie zufrieden sind Sie gegenwärtig – alles in allem – mit Ihrem Leben?

Bitte sagen Sie es mir anhand dieser Liste.

⇒ INT.: Liste S103 vorlegen!

10	Ganz und gar zufrieden
9	
8	
7	
6	
5	
4	
3	
2	
1	
0	Ganz und gar unzufrieden
<hr/>	
99	kA

Spalten 9-10

### Rollenbilder in der Familie

F020A. **(Split 1)** Über die Aufgaben der Frau in der Familie und bei der Kindererziehung gibt es verschiedene Meinungen.

Bitte sagen Sie mir nun zu jeder Aussage auf dieser Liste, ob Sie ihr –

⇒ INT.: Liste 20A vorlegen!

- ▶ voll und ganz zustimmen,
- ▶ eher zustimmen,
- ▶ eher nicht zustimmen oder
- ▶ überhaupt nicht zustimmen.

	Stimme voll zu	Stimme eher zu	Stimme eher nicht zu	Stimme gar nicht zu	wn	kA
A) Eine berufstätige Mutter kann ein genauso herzliches und vertrauensvolles Verhältnis zu ihren Kindern finden wie eine Mutter, die nicht berufstätig ist.	1	2	3	4	8	9
B) Für eine Frau ist es wichtiger, ihrem Mann bei seiner Karriere zu helfen, als selbst Karriere zu machen.	1	2	3	4	8	9
C) Ein Kleinkind wird sicherlich darunter leiden, wenn seine Mutter berufstätig ist.	1	2	3	4	8	9
D) Es ist für alle Beteiligten viel besser, wenn der Mann voll im Berufsleben steht und die Frau zu Hause bleibt und sich um den Haushalt und die Kinder kümmert.	1	2	3	4	8	9
E) Es ist für ein Kind sogar gut, wenn seine Mutter berufstätig ist und sich nicht nur auf den Haushalt konzentriert.	1	2	3	4	8	9
F) Eine verheiratete Frau sollte auf eine Berufstätigkeit verzichten, wenn es nur eine begrenzte Anzahl von Arbeitsplätzen gibt und wenn ihr Mann in der Lage ist, für den Unterhalt der Familie zu sorgen.	1	2	3	4	8	9

Spalten 11-16