

Markus Tausendpfund

Quantitative Datenanalyse. Eine Einführung mit SPSS

Fakultät für
**Kultur- und
Sozialwissen-
schaften**

Das Werk ist urheberrechtlich geschützt. Die dadurch begründeten Rechte, insbesondere das Recht der Vervielfältigung und Verbreitung sowie der Übersetzung und des Nachdrucks, bleiben, auch bei nur auszugsweiser Verwertung, vorbehalten. Kein Teil des Werkes darf in irgendeiner Form (Druck, Fotokopie, Mikrofilm oder ein anderes Verfahren) ohne schriftliche Genehmigung der FernUniversität reproduziert oder unter Verwendung elektronischer Systeme verarbeitet, vervielfältigt oder verbreitet werden.

Der Inhalt dieses Studienbriefs wird gedruckt auf Recyclingpapier (80 g/m², weiß), hergestellt aus 100 % Altpapier.

Vorwort

Der vorliegende Kurs behandelt die sozialwissenschaftliche Datenanalyse, SPSS und insbesondere die sozialwissenschaftliche Datenanalyse **mit** SPSS. Dabei werden Kenntnisse vermittelt, um einfache Analysen selbstständig durchführen zu können. Dazu gehören der Download von Sekundärdatensätzen, die sachgemäße Kodierung von Variablen (Datenmodifikation), die uni- und bivariate Datenanalyse sowie die Anwendung multivariater Analyseverfahren.

Der Text entspricht dabei weniger einem klassischen Lehrbuch, sondern eher einem Begleitkurs bei der Auseinandersetzung mit dem Statistikprogramm SPSS. In diesem Sinne soll der Kurs das Interesse an sozialwissenschaftlichen Fragestellungen wecken sowie die Möglichkeiten und Grenzen der quantitativen Datenanalyse aufzeigen. Für die Datenanalyse wird der ALLBUS-Datensatz 2014 verwendet. Dadurch beziehen sich die Beispiele auf klassische sozialwissenschaftliche Fragestellungen (z.B. Einkommensunterschiede zwischen Männern und Frauen).

Tuuli-Marja Kleiner, Sebastian Kuhn, Reinhold Melcher und Toni Schell haben einzelne Kapitel gelesen und kommentiert. Christian Cleve und Jasmin Genau haben sich besonders intensiv mit dem Kurs auseinandergesetzt und mich auf Ungenauigkeiten und Tippfehler aufmerksam gemacht. Lehrmaterial, das in erster Linie zum Selbststudium angelegt ist, profitiert insbesondere durch Rückmeldungen der Leserinnen und Leser. Deshalb möchte ich mich herzlich bei allen Personen bedanken, die mich auf Fehler und Verbesserungsmöglichkeiten hingewiesen haben.

Aktuelle Ergänzungen und eine mögliche Fehlerliste zu diesem Kurs finden Sie in der Moodle-Lernumgebung des Moduls M1 „Quantitative Methoden der Sozialwissenschaften“ im BA-Studiengang „Politikwissenschaft, Verwaltungswissenschaft und Soziologie“. Dort werden auch Übungsaufgaben und Videotutorials veröffentlicht, die die Auseinandersetzung mit den Inhalten des Kurses fördern sollen. Die aktuelle PDF-Version des Kurses ist über den Virtuellen Studienplatz erhältlich.

Der vorliegende Kurs ist kein „Endprodukt“. Die regelmäßige Aktualisierung stellt eine Daueraufgabe dar. Deshalb freue ich mich sehr über alle Hinweise und Anregungen zur weiteren Verbesserung des Kurses. Sie erreichen mich unter der E-Mail-Adresse Markus.Tausendpfund@fernuni-hagen.de

Hagen, im Juni 2019

Markus Tausendpfund

Inhaltsverzeichnis

Abbildungsverzeichnis	VII
Tabellenverzeichnis	X
1 Einführung	11
1.1 Sozialwissenschaftlicher Forschungsprozess	11
1.2 Quantitative Datenanalyse	13
1.3 SPSS, Stata oder R?	14
1.4 Aufbau des Kurses.....	17
2 SPSS kennenlernen	19
2.1 Vor dem Start.....	19
2.1.1 Lizenz	19
2.1.2 Installation	19
2.1.3 Hilfe bei Fragen oder Problemen	20
2.2 Aufruf des Programms.....	20
2.2.1 Daten-Fenster.....	21
2.2.2 Syntax-Fenster	22
2.2.3 Ausgabe-Fenster.....	23
2.2.4 Optionen	24
2.3 Grundlagen der SPSS-Syntax.....	25
2.3.1 Vorteile der Syntax	25
2.3.2 Grundregeln der Syntax.....	26
2.3.3 Befehlssyntaxreferenz (Command Syntax Reference).....	27
2.4 Erste Analysen	28
2.4.1 Beispiel-Datensatz	28
2.4.2 Öffnen des Beispiel-Datensatzes.....	30
2.4.3 Variablenname, Variablenlabel und Wertelabel.....	31
2.4.4 Skalenniveau der Variablen	34
2.4.5 Fehlende Werte	35
2.4.6 Häufigkeitstabelle.....	35
2.4.7 Deskriptive Statistik	38
2.4.8 Öffnen und Speichern der SPSS-Syntax.....	39
2.4.9 Layout der SPSS-Tabellen	40
3 ALLBUS kennenlernen.....	42
3.1 Was ist der ALLBUS?.....	42

3.2	Zugang zu den ALLBUS-Daten	42
3.2.1	GESIS – Leibniz-Institut für Sozialwissenschaften	43
3.2.2	Datenarchiv für Sozialwissenschaften	43
3.2.3	GESIS-Datenbestandskatalog	43
3.2.4	Download der ALLBUS-Daten.....	44
3.3	Arbeiten mit ALLBUS-Daten	51
3.3.1	Öffnen des ALLBUS-Datensatzes	51
3.3.2	Designgewicht.....	52
3.3.3	Dokumente zum ALLBUS-Datensatz	55
4	Erste Analysen mit den ALLBUS-Daten.....	56
4.1	Syntax-Datei erstellen	56
4.2	Häufigkeitstabellen	57
4.3	Deskriptive Statistiken.....	58
4.4	Gruppenvergleiche	58
4.5	Kreuztabellen	60
4.6	Syntax-Datei abspeichern.....	61
5	Grundlagen der Datenmodifikation	62
5.1	Warum Datenmodifikation?	62
5.2	Werkzeuge der Datenmodifikation.....	66
5.2.1	RECODE	66
5.2.2	COUNT	75
5.2.3	COMPUTE	78
5.2.4	IF	84
5.3	Praktische Hinweise der Datenmodifikation	90
6	Univariate Datenanalyse.....	92
6.1	Lagemaße	92
6.2	Streuungsmaße	95
6.3	Formmaße.....	97
6.4	Zusammenfassung.....	100
7	Bivariate Datenanalyse	104
7.1	Kreuztabellen	104
7.2	Zusammenhangsmaße.....	109
7.2.1	Nominalskalierte Merkmale.....	110
7.2.2	Ordinalskalierte Merkmale	114

7.2.3	Metrische Merkmale	118
7.3	Mittelwertvergleiche	120
8	Multivariate Datenanalyse	125
8.1	Einführung	125
8.2	Lineare Regression	127
8.2.1	Das Grundmodell	127
8.2.2	Lineare Regression mit SPSS	135
8.2.3	Interpretation der Ergebnisse	138
8.2.4	Weitere Optionen der Regression mit SPSS	145
8.2.5	Anwendungsvoraussetzungen	152
8.2.6	Praktische Hinweise	153
8.3	Logistische Regression	155
8.3.1	Das Grundmodell	155
8.3.2	Logistische Regression mit SPSS	161
8.3.3	Interpretation der Ergebnisse	165
8.3.4	Praktische Hinweise	175
9	Grafiken mit SPSS	177
9.1	Einführung	177
9.2	Säulen- und Balkendiagramm	179
9.3	Kreisdiagramm	182
9.4	Histogramm	183
9.5	Boxplot	185
9.6	Streudiagramm	188
10	Weiterführende Literaturhinweise	190
11	Literaturverzeichnis	193

Abbildungsverzeichnis

Abbildung 1: Phasen eines quantitativen Forschungsprojekts	12
Abbildung 2: SPSS-Begrüßungsdialogfeld	20
Abbildung 3: SPSS-Datenansicht	21
Abbildung 4: SPSS-Variablenansicht	22
Abbildung 5: SPSS-Syntax-Fenster	23
Abbildung 6: SPSS-Ausgabe-Fenster	24
Abbildung 7: SPSS-Optionen (Dateispeicherorte)	25
Abbildung 8: Befehlssyntaxreferenz (Command Syntax Reference)	27
Abbildung 9: Hilfefunktion bei der SPSS-Syntax	28
Abbildung 10: Beispieldatensatz in der SPSS-Datenansicht	30
Abbildung 11: Beispieldatensatz in der SPSS-Variablenansicht	31
Abbildung 12: Ändern der Variablenlabels	33
Abbildung 13: Häufigkeitstabelle des Schulabschlusses	36
Abbildung 14: Häufigkeitstabelle des Schulabschlusses mit Variablen- und Wertelabels	37
Abbildung 15: SPSS-Optionen (Ausgabe)	38
Abbildung 16: Deskriptive Angaben des Einkommens	39
Abbildung 17: Tabellenlayout Systemstandard	40
Abbildung 18: Tabellenlayout Klassikstandard	40
Abbildung 19: Tabellenlayout festlegen	41
Abbildung 20: GESIS – Leibniz-Institut für Sozialwissenschaften	44
Abbildung 21: GESIS – „Unser Angebot“	45
Abbildung 22: GESIS – Datenservice für Sekundäranalysen	45
Abbildung 23: GESIS-Datenbestandskatalog	46
Abbildung 24: Übersicht zur ALLBUS 2014	47
Abbildung 25: ALLBUS-Daten 2014	48
Abbildung 26: Datenbestandskatalog Download	49
Abbildung 27: Anmelden für den Datenbestandskatalog	49
Abbildung 28: Neues Konto anlegen für DBK/datorium	50
Abbildung 29: Download der ALLBUS-Daten 2014	51
Abbildung 30: Variablenansicht der ALLBUS-Daten 2014	52
Abbildung 31: Gewichtung über das Menü aktivieren	55
Abbildung 32: Exemplarische Syntax für die Analyse mit den ALLBUS-Daten 2014	57
Abbildung 33: Politikinteresse nach Geschlecht	59
Abbildung 34: Kreuztabelle des politischen Interesses nach Geschlecht	61
Abbildung 35: Häufigkeitstabelle des politischen Interesses	63
Abbildung 36: Häufigkeitstabelle des allgemeinen Schulabschlusses	64
Abbildung 37: Häufigkeitstabelle des sozialen Vertrauens	65
Abbildung 38: Häufigkeitstabelle des politischen Interesses (vor der Kodierung)	67
Abbildung 39: Häufigkeitstabelle des politischen Interesses (nach der Kodierung)	68
Abbildung 40: Häufigkeitstabelle des politischen Interesses (nach Formatanpassung)	69
Abbildung 41: Häufigkeitstabelle des allgemeinen Schulabschlusses (vor der Kodierung)	70
Abbildung 42: Häufigkeitstabelle des allgemeinen Schulabschlusses (nach der Kodierung)	72
Abbildung 43: Häufigkeitstabelle des Vertrauens in Mitmenschen (vor der Kodierung)	73

Abbildung 44: Häufigkeitstabelle des Vertrauens in Mitmenschen (nach der Kodierung).....	74
Abbildung 45: Häufigkeitstabelle Musikunterricht	76
Abbildung 46: Häufigkeitstabelle „kein Mitglied“	78
Abbildung 47: Häufigkeitstabelle von V210 (Wichtigkeit von Ruhe und Ordnung).....	85
Abbildung 48: Häufigkeitstabelle von V211 (Wichtigkeit von Bürgereinfluss).....	86
Abbildung 49: Häufigkeitstabelle von V212 (Wichtigkeit der Inflationsbekämpfung).....	86
Abbildung 50: Häufigkeitstabelle von V213 (Wichtigkeit von freier Meinungsäußerung).....	86
Abbildung 51: Häufigkeitstabelle des erstellten Inglehart-Index.....	89
Abbildung 52: Häufigkeitstabelle zur Lebenszufriedenheit	93
Abbildung 53: Lagemaße mit dem FREQUENCIES-Befehl	94
Abbildung 54: DESCRIPTIVES-Befehl (Standardausgabe)	95
Abbildung 55: Lagemaße mit dem DESCRIPTIVES-Befehl	95
Abbildung 56: Streuungsmaße mit dem FREQUENCIES-Befehl	96
Abbildung 57: Streuungsmaße mit dem DESCRIPTIVES-Befehl	97
Abbildung 58: Normalverteilung	98
Abbildung 59: Formmaße mit dem FREQUENCIES-Befehl.....	99
Abbildung 60: Formmaße mit dem DESCRIPTIVES-Befehl.....	100
Abbildung 61: Lage-, Streuungs- und Formmaße mit dem FREQUENCIES-Befehl	100
Abbildung 62: Lage-, Streuungs- und Formmaße mit dem DESCRIPTIVES-Befehl	101
Abbildung 63: Balkendiagramm zur Lebenszufriedenheit (Angaben in Prozent).....	103
Abbildung 64: Kreuztabelle der Demokratiezufriedenheit nach Gebiet (Häufigkeiten).....	105
Abbildung 65: Kreuztabelle der Demokratiezufriedenheit nach Gebiet (Spaltenprozente)	107
Abbildung 66: Kreuztabelle der Demokratiezufriedenheit nach Gebiet (Zeilenprozente).....	108
Abbildung 67: Kreuztabelle von Rauchen und Geschlecht	110
Abbildung 68: Berechnung von Chi-Quadrat.....	112
Abbildung 69: Berechnung von Phi und Cramer's V	112
Abbildung 70: Kreuztabelle der Bildung und des Politikinteresses (Spaltenprozente)	115
Abbildung 71: Berechnung von Spearman's Rho	116
Abbildung 72: Berechnung von Spearman's Rho und Kendall's Tau-B	117
Abbildung 73: Deskriptive Statistiken des Einkommens und der Lebenszufriedenheit.....	119
Abbildung 74: Korrelation des Einkommens und der allgemeinen Lebenszufriedenheit.....	119
Abbildung 75: Einkommen und Lebenszufriedenheit nach Geschlecht	121
Abbildung 76: Aussage über die Grundgesamtheit auf Basis einer Stichprobe	121
Abbildung 77: Nettoeinkommen in Abhängigkeit des Geschlechts.....	123
Abbildung 78: t-Test des Einkommens in Abhängigkeit des Geschlechts	123
Abbildung 79: Lebenszufriedenheit in Abhängigkeit des Geschlechts.....	124
Abbildung 80: t-Test der Lebenszufriedenheit in Abhängigkeit des Geschlechts.....	124
Abbildung 81: Auswahl von regressionsanalytischen Verfahren	125
Abbildung 82: Streudiagramm von Alter und Einkommen	128
Abbildung 83: Streudiagramm von Einkommen und Alter mit Regressionsgerade.....	129
Abbildung 84: Empirischer und geschätzter Wert einer linearen Regression.....	130
Abbildung 85: Bivariate Regression (Beispieldaten)	132
Abbildung 86: Grafische Darstellung einer multiplen Regression	134
Abbildung 87: Multiple Regression (Beispieldaten).....	134
Abbildung 88: Ergebnisse einer linearen Regression	138

Abbildung 89: Ergebnisse der Regressionsanalyse (ANOVA)	142
Abbildung 90: TSS, RSS und MSS.....	143
Abbildung 91: Modellzusammenfassung einer linearen Regression	144
Abbildung 92: Grundinformationen zur linearen Regression	145
Abbildung 93: Ungewichtete Häufigkeitstabelle zum allgemeinen Schulabschluss	146
Abbildung 94: Ergebnisse der linearen Regression mit Bildungsgruppen I.....	148
Abbildung 95: Ergebnisse der linearen Regression mit Bildungsgruppen II	149
Abbildung 96: Deskriptive Statistiken der abhängigen und unabhängigen Variablen.....	150
Abbildung 97: Lineare Regressionsfunktion	156
Abbildung 98: Logistische Regressionsfunktion	158
Abbildung 99: Regressionskoeffizienten einer logistischen Regression (Beispieldaten)	160
Abbildung 100: Logistische Regression (Beispieldaten).....	161
Abbildung 101: Ergebnisse einer logistischen Regression	166
Abbildung 102: Grafische Darstellung der logistischen Regression	169
Abbildung 103: Bewertung der Modellgüte	170
Abbildung 104: Iterationsprotokoll einer logistischen Regression.....	172
Abbildung 105: Klassifizierungstabelle einer logistischen Regression (nur Konstante).....	174
Abbildung 106: Klassifizierungstabelle einer logistischen Regression (mit uVs).....	174
Abbildung 107: Säulendiagramm der allgemeinen Lebenszufriedenheit	180
Abbildung 108: Balkendiagramm der allgemeinen Lebenszufriedenheit (bearbeitet)	181
Abbildung 109: Kreisdiagramm für Haustierbesitz (bearbeitet).....	183
Abbildung 110: Histogramm des Alters (bearbeitet).....	184
Abbildung 111: Elemente eines Boxplots.....	185
Abbildung 112: Boxplot der allgemeinen Lebenszufriedenheit	186
Abbildung 113: Boxplot der allgemeinen Lebenszufriedenheit (bearbeitet)	186
Abbildung 114: Boxplots der Demokratiezufriedenheit nach pol. Interesse (bearbeitet).....	187
Abbildung 115: Streudiagramm von Alter und Einkommen (bearbeitet)	188
Abbildung 116: Streudiagramm von Alter und Einkommen mit Regressionsgerade.....	189

Tabellenverzeichnis

Tabelle 1: SPSS, Stata und R im Überblick.....	15
Tabelle 2: Beispieldaten.....	29
Tabelle 3: Beschreibung der Beispieldaten.....	29
Tabelle 4: Traditionelle Skalenniveaus und SPSS-Skalenniveaus im Vergleich.....	34
Tabelle 5: Datengrundlage für das Designgewicht in der ALLBUS 2014.....	53
Tabelle 6: Vergleich Mikrozensus 2013 und gewichtete ALLBUS-Daten.....	54
Tabelle 7: Datenmodifikation des politischen Interesses.....	66
Tabelle 8: Datenmodifikation des allgemeinen Schulabschlusses.....	71
Tabelle 9: Elemente des RECODE-Befehls.....	75
Tabelle 10: Mitgliedschaft in Vereinen (Anwendung des COUNT-Befehls).....	77
Tabelle 11: Einfache Berechnungen mit dem COMPUTE-Befehl.....	79
Tabelle 12: Berechnung des Alters bei der ersten Berufstätigkeit.....	79
Tabelle 13: Möglichkeiten der Indexbildung mit SPSS.....	81
Tabelle 14: Funktionen mit COMPUTE.....	82
Tabelle 15: Konstruktion des Inglehart-Index.....	87
Tabelle 16: Logische Operatoren des IF-Befehls.....	89
Tabelle 17: Relationale Operatoren des IF-Befehls.....	90
Tabelle 18: Interpretation von Schiefe und Wölbung.....	99
Tabelle 19: Unteranweisungen des FREQUENCIES- und des DESCRIPTIVES-Befehls.....	102
Tabelle 20: Wichtige Zusammenhangsmaße bei der bivariaten Datenanalyse (Auswahl).....	109
Tabelle 21: Arbeitstabelle für die Berechnung von Chi-Quadrat.....	111
Tabelle 22: Interpretation von Cramer's V.....	113
Tabelle 23: Spezifikationen der STATISTICS-Unteranweisung beim CROSSTABS-Befehl.....	113
Tabelle 24: Spezifikationen der PRINT-Unteranweisung beim NONPAR CORR-Befehl.....	116
Tabelle 25: Interpretation von Spearman's Rho.....	117
Tabelle 26: Interpretation des Korrelationskoeffizienten nach Pearson.....	120
Tabelle 27: Beispieldaten für Alter und Einkommen.....	127
Tabelle 28: Beispieldaten für Alter, Einkommen und Berufserfahrung.....	133
Tabelle 29: Bestimmungsfaktoren des Einkommens.....	151
Tabelle 30: Beispieldaten für Rauchen und Alter in Jahren.....	155
Tabelle 31: Angaben zur Modellgüte im Vergleich.....	172
Tabelle 32: Determinanten des Rauchens.....	175
Tabelle 33: Diagrammtypen und SPSS-Befehle.....	177

1 Einführung

Die quantitative Datenanalyse ist die Phase im sozialwissenschaftlichen Forschungsprozess, in der die theoretisch entwickelten Hypothesen empirisch geprüft werden. In diesem Kapitel werden die einzelnen Phasen des Forschungsprozesses knapp skizziert, die Bedeutung der Methodenkompetenz für die Auseinandersetzung mit empirischen Studien dargelegt und Statistikprogramme vorgestellt. Die Einführung schließt mit einem Ausblick auf die weiteren Kapitel dieses Kurses und verweist auf ergänzende Materialien in der Moodle-Lernumgebung.¹

1.1 Sozialwissenschaftlicher Forschungsprozess

In einem quantitativen Forschungsprojekt lassen sich idealtypisch mehrere Phasen unterscheiden (siehe Abbildung 1). Nach der Entscheidung für ein Forschungsthema und der Entwicklung einer Forschungsfrage (1) müssen zunächst die zentralen Konzepte der Forschungsfrage identifiziert und theoretisch geklärt werden (2). Auf dieser Grundlage können Hypothesen formuliert (3) und Operationalisierungen der Konzepte (4) entwickelt werden (ausführlicher Tausendpfund 2018a).

Diese Phasen eines Forschungsprojekts erfolgen in intensiver Auseinandersetzung mit der existierenden Fachliteratur. Nur wer den Forschungsstand zu seinem Forschungsthema kennt, kann eine gehaltvolle Forschungsfrage entwickeln. Die Auseinandersetzung mit der Fachliteratur ist aber auch für die Konzeptspezifikation und die Entwicklung von Hypothesen erforderlich. Schließlich ist auch bei der „Übersetzung“ theoretischer Konzepte in empirische Indikatoren ein Überblick existierender Operationalisierungen notwendig.

Bedeutung des Forschungsstands

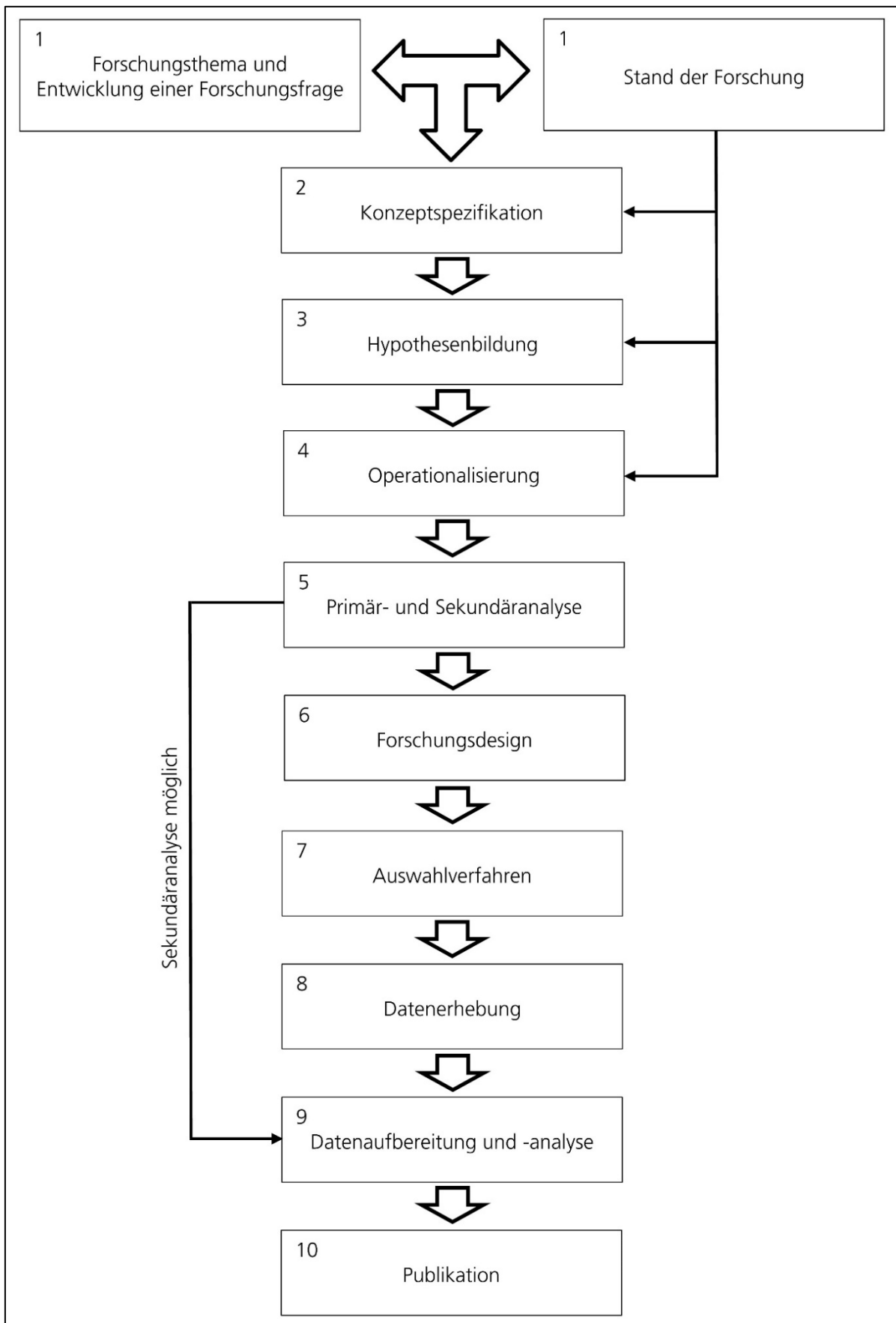
Bei einer Primäranalyse werden neue Daten erhoben, um die Forschungsfrage zu untersuchen. Bei einer Sekundäranalyse werden existierende Daten genutzt, um die Forschungsfrage zu bearbeiten (5). Falls für die Bearbeitung einer Forschungsfrage bereits geeignetes Datenmaterial existiert (z.B. ALLBUS), dann können die Phasen Forschungsdesign (6), Auswahlverfahren (7) und Datenerhebung (8) „übersprungen“ werden.

Die Datenaufbereitung und -analyse stellt eine Phase in einem sozialwissenschaftlichen Forschungsprojekt dar (9). In dieser Phase werden die theoretisch formulierten Hypothesen empirisch geprüft. Mittlerweile existieren zahlreiche Verfahren der Datenanalyse (für einen Überblick siehe z.B. Wolf und Best 2010a); alle Verfahren setzen jedoch eine vorherige intensive Auseinandersetzung mit dem jeweiligen Forschungsstand voraus. Mit anderen Worten: Die Datenanalyse kann die vorherige Auseinandersetzung mit dem Forschungsstand nicht ersetzen.

In Publikationen (10) werden die Forschungsergebnisse der Öffentlichkeit zugänglich gemacht.

¹ Ausschließlich aus Gründen der besseren Lesbarkeit wird in diesem Kurs nicht durchgängig eine geschlechterneutrale Sprache verwendet. Männliche, weibliche und genderneutrale Formen wechseln sich in diesem Kurs zufallsverteilt ab. Mit den Bezeichnungen sind jeweils alle Geschlechter gemeint.

Abbildung 1: Phasen eines quantitativen Forschungsprojekts



Quelle: Eigene Darstellung