

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.

## 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 Gerau haben sich besonders intensiv mit dem Kurs auseinandergesetzt und mich auf Ungenauigkeiten und Tippfehler aufmerksam gemacht. Dafür an dieser Stelle ein herzliches Dankeschön!

Zum Sommersemester 2018 wurde der Kurs auf die SPSS-Version 25 aktualisiert. Außerdem wurde das Kapitel 8.3 „Logistische Regression“ neu in den Kurs aufgenommen.

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.

Aktuelle Ergänzungen und eine Errata-Liste 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.

Hagen, im Dezember 2017

Markus Tausendpfund

## Inhaltsverzeichnis

Vorwort .....	III
Inhaltsverzeichnis .....	IV
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 .....	16
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 die 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	Deskriptive Statistik .....	92
6.2	Wie werden Lage- und Streuungsmaße berechnet? .....	93
7	Bivariate Datenanalyse .....	99
7.1	Kreuztabellen .....	99
7.2	Zusammenhangsmaße .....	104
7.2.1	Nominalskalierte Merkmale .....	105
7.2.2	Ordinalskalierte Merkmale .....	108

---

7.2.3	Metrische Merkmale.....	112
7.3	Mittelwertvergleiche.....	115
8	Multivariate Datenanalyse.....	120
8.1	Einführung.....	120
8.2	Lineare Regression.....	122
8.2.1	Das Grundmodell.....	122
8.2.2	Lineare Regression mit SPSS.....	130
8.2.3	Interpretation der Ergebnisse.....	133
8.2.4	Weitere Optionen der Regression mit SPSS.....	140
8.2.5	Anwendungsvoraussetzungen.....	146
8.2.6	Praktische Hinweise.....	148
8.3	Logistische Regression.....	150
8.3.1	Das Grundmodell.....	150
8.3.2	Logistische Regression mit SPSS.....	156
8.3.3	Interpretation der Ergebnisse.....	159
8.3.4	Praktische Hinweise.....	170
9	Grafiken mit SPSS.....	172
9.1	Einführung.....	172
9.2	Säulen- und Balkendiagramm.....	174
9.3	Kreisdiagramm.....	177
9.4	Histogramm.....	178
9.5	Boxplot.....	180
9.6	Streudiagramm.....	183
10	Weiterführende Literaturhinweise.....	185
11	Literaturverzeichnis.....	188

## 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 zu der ALLBUS 2014 .....	47
Abbildung 25: ALLBUS-Daten 2014.....	48
Abbildung 26: Anmeldung beim GESIS-Datenbestandskatalog .....	49
Abbildung 27: Registrierung für ZACAT, DBK und datorium .....	49
Abbildung 28: Registrierung bei ZACAT, DBK und datorium.....	50
Abbildung 29: Download der ALLBUS-Daten 2014 .....	51
Abbildung 30: Unicode-Codierungsmodus .....	52
Abbildung 31: Variablenansicht der ALLBUS-Daten 2014.....	52
Abbildung 32: Gewichtung über das Menü aktivieren .....	55
Abbildung 33: Exemplarische Syntax für die Analyse mit den ALLBUS-Daten 2014.....	57
Abbildung 34: Politikinteresse nach Geschlecht .....	59
Abbildung 35: Kreuztabelle des politischen Interesses nach Geschlecht.....	61
Abbildung 36: Häufigkeitstabelle des politischen Interesses .....	63
Abbildung 37: Häufigkeitstabelle des allgemeinen Schulabschlusses .....	64
Abbildung 38: Häufigkeitstabelle des sozialen Vertrauens.....	65
Abbildung 39: Häufigkeitstabelle des politischen Interesses (vor der Kodierung).....	67
Abbildung 40: Häufigkeitstabelle des politischen Interesses (nach der Kodierung).....	68
Abbildung 41: Häufigkeitstabelle des politischen Interesses (nach Formatanpassung).....	69
Abbildung 42: Häufigkeitstabelle des allgemeinen Schulabschlusses (vor der Kodierung) .....	70
Abbildung 43: Häufigkeitstabelle des allgemeinen Schulabschlusses (nach der Kodierung).....	72

Abbildung 44: Häufigkeitstabelle des Vertrauens in Mitmenschen (vor der Kodierung) .....	73
Abbildung 45: Häufigkeitstabelle des Vertrauens in Mitmenschen (nach der Kodierung).....	74
Abbildung 46: Häufigkeitstabelle Musikunterricht .....	76
Abbildung 47: Häufigkeitstabelle „kein Mitglied“ .....	78
Abbildung 48: Häufigkeitstabelle von V210 (Wichtigkeit von Ruhe und Ordnung) .....	85
Abbildung 49: Häufigkeitstabelle von V211 (Wichtigkeit von Bürgereinfluss) .....	86
Abbildung 50: Häufigkeitstabelle von V212 (Wichtigkeit der Inflationsbekämpfung).....	86
Abbildung 51: Häufigkeitstabelle von V213 (Wichtigkeit von freier Meinungsäußerung).....	86
Abbildung 52: Häufigkeitstabelle des erstellten Inglehart-Index .....	89
Abbildung 53: Häufigkeitstabelle zur Lebenszufriedenheit.....	94
Abbildung 54: Deskriptive Statistik zur Lebenszufriedenheit (Standardausgabe).....	95
Abbildung 55: Deskriptive Statistik zur Lebenszufriedenheit (erweiterte Ausgabe) .....	95
Abbildung 56: Deskriptive Angaben zur Lebenszufriedenheit (erweiterte Ausgabe) .....	96
Abbildung 57: Balkendiagramm zur Lebenszufriedenheit (Angaben in Prozent).....	98
Abbildung 58: Kreuztabelle der Demokratiezufriedenheit nach Gebiet (Häufigkeiten).....	100
Abbildung 59: Kreuztabelle der Demokratiezufriedenheit nach Gebiet (Spaltenprozente) .....	102
Abbildung 60: Kreuztabelle der Demokratiezufriedenheit nach Gebiet (Zeilenprozente).....	103
Abbildung 61: Kreuztabelle von Rauchen und Geschlecht .....	105
Abbildung 62: Berechnung von Chi-Quadrat.....	107
Abbildung 63: Berechnung von Phi und Cramer's V .....	107
Abbildung 64: Kreuztabelle der Bildung und des Politikinteresses (Spaltenprozente) .....	110
Abbildung 65: Berechnung von Spearman's rho .....	110
Abbildung 66: Berechnung von Spearman's rho und Kendall's Tau-B .....	112
Abbildung 67: Deskriptive Statistiken des Einkommens und der Lebenszufriedenheit.....	113
Abbildung 68: Korrelation des Einkommens und der allgemeinen Lebenszufriedenheit.....	114
Abbildung 69: Einkommen und Lebenszufriedenheit nach Geschlecht.....	116
Abbildung 70: Aussage über die Grundgesamtheit auf Basis einer Stichprobe .....	116
Abbildung 71: Nettoeinkommen in Abhängigkeit des Geschlechts .....	118
Abbildung 72: t-Test des Einkommens in Abhängigkeit des Geschlechts.....	118
Abbildung 73: Lebenszufriedenheit in Abhängigkeit des Geschlechts .....	119
Abbildung 74: t-Test der Lebenszufriedenheit in Abhängigkeit des Geschlechts.....	119
Abbildung 75: Auswahl von regressionsanalytischen Verfahren .....	120
Abbildung 76: Streudiagramm von Alter und Einkommen .....	123
Abbildung 77: Streudiagramm von Einkommen und Alter mit Regressionsgerade.....	124
Abbildung 78: Empirischer und geschätzter Wert einer linearen Regression .....	125
Abbildung 79: Bivariate Regression (Beispieldaten) .....	127
Abbildung 80: Grafische Darstellung einer multiplen Regression.....	129
Abbildung 81: Multiple Regression (Beispieldaten).....	129
Abbildung 82: Ergebnisse einer linearen Regression .....	133
Abbildung 83: Ergebnisse der Regressionsanalyse (ANOVA).....	137
Abbildung 84: TSS, RSS und MSS .....	138
Abbildung 85: Modellzusammenfassung einer linearen Regression.....	139
Abbildung 86: Grundinformationen zur linearen Regression.....	140
Abbildung 87: Ungewichtete Häufigkeitstabelle zum allgemeinen Schulabschluss .....	141
Abbildung 88: Ergebnisse der linearen Regression mit Bildungsgruppen I .....	143



---

Abbildung 89: Ergebnisse der linearen Regression mit Bildungsgruppen II.....	144
Abbildung 90: Deskriptive Statistiken der abhängigen und unabhängigen Variablen.....	145
Abbildung 91: Lineare Regressionsfunktion .....	151
Abbildung 92: Logistische Regressionsfunktion.....	153
Abbildung 93: Regressionskoeffizienten einer logistischen Regression (Beispieldaten) .....	154
Abbildung 94: Logistische Regression (Beispieldaten).....	155
Abbildung 95: Ergebnisse einer logistischen Regression .....	160
Abbildung 96: Grafische Darstellung der logistischen Regression .....	163
Abbildung 97: Bewertung der Modellgüte.....	165
Abbildung 98: Iterationsprotokoll einer logistischen Regression .....	167
Abbildung 99: Klassifizierungstabelle einer logistischen Regression (nur Konstante) .....	168
Abbildung 100: Klassifizierungstabelle einer logistischen Regression (mit uVs) .....	168
Abbildung 101: Säulendiagramm der allgemeinen Lebenszufriedenheit.....	175
Abbildung 102: Balkendiagramm der allgemeinen Lebenszufriedenheit (bearbeitet) .....	176
Abbildung 103: Kreisdiagramm für Haustierbesitz (bearbeitet).....	178
Abbildung 104: Histogramm des Alters (bearbeitet).....	179
Abbildung 105: Elemente eines Boxplots .....	180
Abbildung 106: Boxplot der allgemeinen Lebenszufriedenheit .....	181
Abbildung 107: Boxplot der allgemeinen Lebenszufriedenheit (bearbeitet).....	181
Abbildung 108: Boxplots der Demokratiezufriedenheit nach pol. Interesse (bearbeitet).....	182
Abbildung 109: Streudiagramm von Alter und Einkommen (bearbeitet).....	183
Abbildung 110: Streudiagramm von Alter und Einkommen mit Regressionsgerade.....	184

## 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: Weitere Elemente des RECODE-Befehls .....	74
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: Unteranweisungen des FREQUENCIES- und DESCRIPTIVES-Befehls .....	97
Tabelle 19: Wichtige Zusammenhangsmaße bei der bivariaten Datenanalyse (Auswahl) .....	104
Tabelle 20: Arbeitstabelle für die Berechnung von Chi-Quadrat .....	106
Tabelle 21: Interpretation von Cramer's V .....	108
Tabelle 22: Spezifikationen der STATISTICS-Unteranweisung beim CROSSTABS-Befehl .....	108
Tabelle 23: Spezifikationen der PRINT-Unteranweisung beim NONPAR CORR-Befehl .....	111
Tabelle 24: Interpretation von Spearman's rho .....	112
Tabelle 25: Interpretation des Korrelationskoeffizienten nach Pearson .....	115
Tabelle 26: Beispieldaten für Alter und Einkommen .....	122
Tabelle 27: Beispieldaten für Alter, Einkommen und Berufserfahrung .....	128
Tabelle 28: Bestimmungsfaktoren des Einkommens .....	146
Tabelle 29: Beispieldaten für Rauchen und Alter in Jahren .....	150
Tabelle 30: Angaben zur Modellgüte im Vergleich .....	166
Tabelle 31: Determinanten des Rauchens .....	169
Tabelle 32: Diagrammtypen und SPSS-Befehle .....	172

# 1 Einführung<sup>1</sup>

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 bekannte 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 Sozialwissenschaftlicher Forschungsprozess

In einem quantitativen Forschungsprojekt lassen sich idealtypisch mehrere Phasen unterscheiden (siehe Abbildung 1; ausführlicher Tausendpfund 2017). 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.

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, ESS), dann können die Phasen Forschungsdesign (6), Auswahlverfahren (7) und Datenerhebung (8) „übersprungen“ werden.

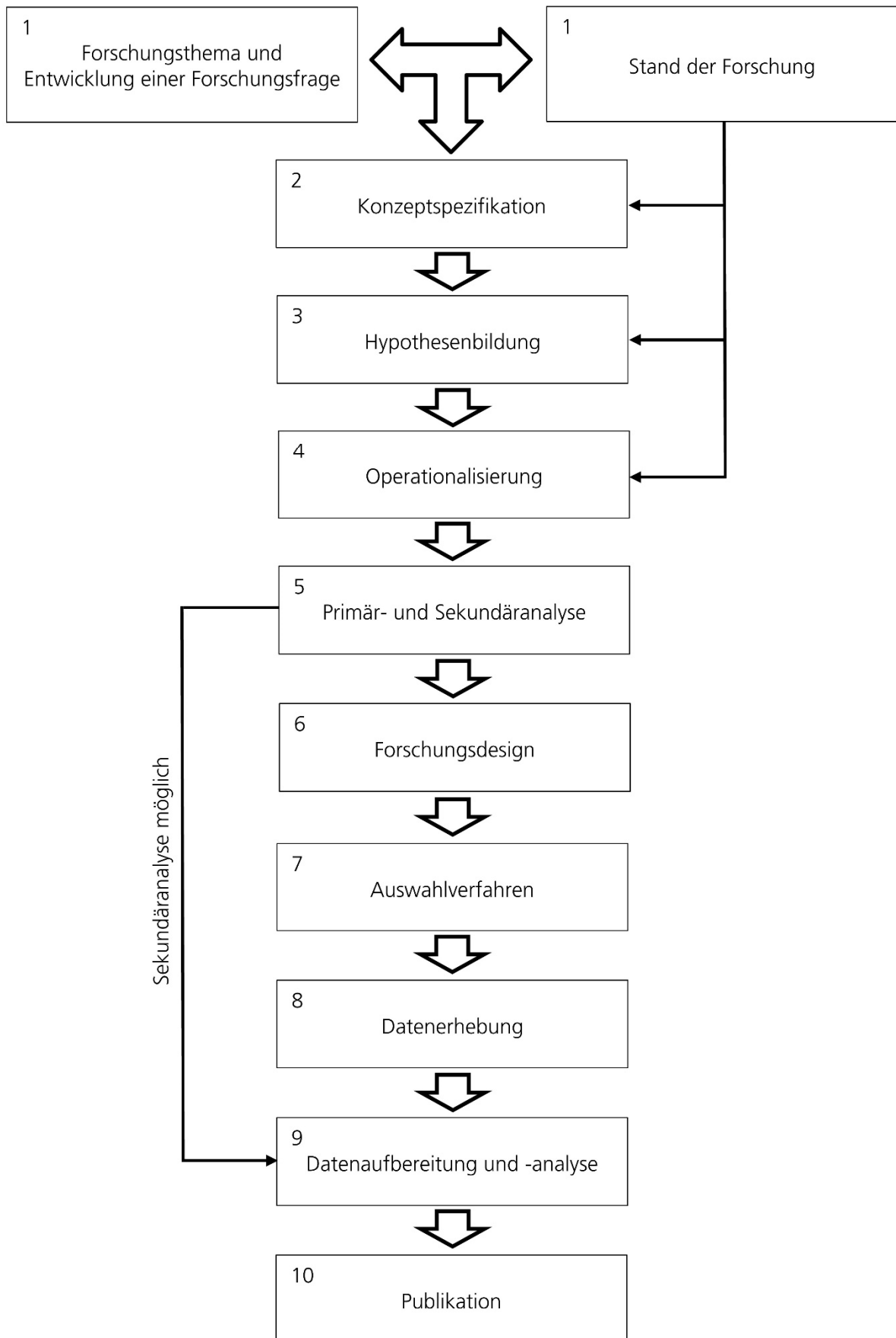
Die Datenaufbereitung und -analyse stellt eine Phase in einem sozialwissenschaftlichen Forschungsprojekts dar (9). In dieser Phase werden die theoretisch formulierten Hypothesen empirisch geprüft. Mittlerweile existieren zahlreiche Verfahren der Datenanalyse (für einen Überblick siehe z.B. Wolf und Best 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.

---

<sup>1</sup> 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



## 1.2 Quantitative Datenanalyse

Sozialwissenschaftlerinnen wollen soziale Sachverhalte beschreiben und erklären. Fördert ein höheres Einkommen die Lebenszufriedenheit? Warum beteiligen sich Menschen mit höherer Bildung eher an Wahlen als Personen mit geringer Bildung? Welche Faktoren beeinflussen die Scheidungswahrscheinlichkeit?

Ein Blick in die gängigen sozialwissenschaftlichen Fachzeitschriften wie die Kölner Zeitschrift für Soziologie und Sozialpsychologie (KZfSS), die Zeitschrift für Soziologie (ZfS), die Politische Vierteljahresschrift (PVS) oder auch die Zeitschrift für Vergleichende Politikwissenschaft (ZfVP) zeigt, dass in Fachaufsätzen häufig quantitative Analyseverfahren genutzt werden, um entsprechende Forschungsfragen zu bearbeiten. Ohne grundlegende Methodenkompetenz können die empirischen Ergebnisse allerdings weder interpretiert noch kritisiert werden. Wer sich also gehaltvoll mit einer (quantitativen) empirischen Studie auseinandersetzen möchte, der muss die Möglichkeiten und Grenzen quantitativer Analyseverfahren kennen.

Deshalb werden im Modul M1 „Quantitative Methoden der Sozialwissenschaften“ im BA-Studiengang „Politikwissenschaft, Verwaltungswissenschaft und Soziologie“ wichtige Verfahren der deskriptiven Statistik (z.B. Mittelwert), Zusammenhangsmaße (z.B. Pearson's  $r$ ) sowie die Grundlagen der linearen Regression vorgestellt (ausführlicher Mittag 2018). Übungsaufgaben und Schritt-für-Schritt-Anleitungen in der Moodle-Lernumgebung sollen das Verständnis für die einzelnen Analyseverfahren fördern. Dabei wird beispielsweise der Mittelwert und die Standardabweichung einzelner Merkmale (z.B. Alter) oder der Zusammenhang von zwei Merkmalen (z.B. Bildung und Einkommen) berechnet.

### Analyseverfahren kennenlernen

Mit der Zahl der Untersuchungsobjekte (z.B. Personen) steigt der rechnerische Aufwand allerdings erheblich an. Deshalb basieren solche Übungsaufgaben in der Regel auf kleinen – häufig auch fiktiven – Datensätzen. So wird beispielsweise das durchschnittliche Alter oder auch der Zusammenhang zwischen Bildung und Einkommen bei zehn Befragten ermittelt. Im Mittelpunkt steht das Kennenlernen und Verstehen quantitativer Analyseverfahren, weniger die Auseinandersetzung mit „echten“ Daten.

In der Forschungspraxis sind quantitativ arbeitende Sozialwissenschaftler mit deutlich größeren Fallzahlen konfrontiert. Die Allgemeine Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) enthält beispielsweise Informationen von knapp 3500 Befragten, der European Social Survey (ESS) sogar Angaben für mehrere zehntausend Personen. Es ist allerdings weder sinnvoll noch fördert es vermutlich das Verständnis, wenn das durchschnittliche Alter oder die Beziehung zwischen Bildung und Einkommen auf Basis von mehreren tausend Untersuchungsobjekten „händisch“ berechnet wird. Für diese Aufgaben nutzt die Sozialwissenschaftlerin ein Statistikprogramm. Programme wie SPSS, Stata oder R sollen dem Anwender diese Routinetätigkeiten abnehmen.

### Anwendung in der Forschungspraxis

An dieser Stelle setzt der vorliegende Kurs an. Der Kurs „Quantitative Datenanalyse“ will einerseits die Vorgehensweise bei der quantitativen Datenanalyse verdeutlichen und andererseits praktische Kenntnisse in der Anwendung eines Statistikprogramms vermitteln. Der Kurs behandelt dabei ty-

pische Herausforderungen der Datenaufbereitung und Datenanalyse im Rahmen eines quantitativen Forschungsprojekts. Für eine möglichst einfache und verständliche Präsentation werden andere Phasen des Forschungsprojekts (z.B. Konzeptspezifikation) an dieser Stelle zurückgestellt.

Die Kenntnis wichtiger quantitativer Analyseverfahren und die Anwendung in der Forschungspraxis sollen in erster Linie die gehaltvolle Auseinandersetzung mit empirischen Studien fördern. Wer selbst einmal mit einem Statistikprogramm gearbeitet hat, der kann meist besser die Herausforderungen in der Forschungspraxis nachvollziehen und entsprechende Entscheidungen bei einem veröffentlichten Aufsatz in einer Fachzeitschrift bewerten (z.B. Umgang mit fehlenden Werten). Deshalb dient die Arbeit mit einem Statistikprogramm auch unmittelbar der Lesefähigkeit für die Auseinandersetzung mit empirischen Studien.

### Qualifikation auf dem Arbeitsmarkt

(Erste) Erfahrungen mit einem Statistikprogramm sind aber nicht nur wertvoll, um quantitative Forschungsergebnisse besser nachvollziehen zu können, sondern stellen auch eine wichtige Qualifikation auf dem Arbeitsmarkt dar. Dies unterstreicht eine inhaltsanalytische Auswertung der Stellenanzeigen in der Wochenzeitung „Die Zeit“ (Schreiber 2016). Bei fast jeder zweiten Ausschreibung werden Methodenkenntnisse erwartet.

*„Dabei überwiegt zurzeit die Nachfrage nach quantitativen Methodenkenntnissen speziell für Befragungen, Evaluationen und statistische Auswertungen, beispielsweise mit SPSS.“ (Schreiber 2016, S. 29)*

Auch frühere Veröffentlichungen (z.B. Engel 2002; Schnapp et al. 2004) haben bereits auf die Praxisrelevanz der Methodenausbildung hingewiesen. Methodenkenntnisse bzw. die praktische Kompetenz mit einem Statistikprogramm haben unmittelbare Berufsrelevanz.

### Empirische Arbeiten verfassen

Schließlich stehen heute zahlreiche hochwertige Datensätze zur Verfügung, die sich hervorragend für sekundäranalytische Auswertungen eignen. Diese können (und sollen) auch im Rahmen des Studiums genutzt werden, um beispielsweise eine eigene empirische Haus- oder Abschlussarbeit zur Wahlbeteiligung oder sozialen Ungleichheit zu verfassen. Neben der Kenntnis wichtiger quantitativer Analyseverfahren sind dafür natürlich auch praktische Fähigkeiten mit einem Statistikprogramm erforderlich.

## 1.3 SPSS, Stata oder R?

Für die quantitative Datenanalyse stehen heute mehrere leistungsfähige Statistikprogramme zur Verfügung. Die bekanntesten Programme sind SPSS, Stata und R.<sup>2</sup> Gelegentlich sind im wissenschaftlichen Alltag emotionale Debatten überzeugter Anhänger der einzelnen Statistikprogramme zu beobachten, die die Vorteile (nicht aber die Nachteile) des „eigenen“ Programms hervorheben und alternative Statistikprogramme abwerten. Von der deskriptiven Statistik (z.B. Median, Mittelwert) über einfache Zusammenhangsmaße (z.B. Cramer's V, Spearman's rho, Pearson's r) bis hin

<sup>2</sup> Weitere Programme sind beispielsweise SAS, GRAFSTAT, gretl oder auch STATISTICA, die aber in der sozialwissenschaftlichen Forschungspraxis eine untergeordnete Rolle spielen. Für einfache Berechnungen sind auch Tabellenkalkulationsprogramme wie Excel ausreichend.

zu fortgeschrittenen Analyseverfahren (z.B. lineare und logistische Regression) sowie Signifikanztests decken die drei Programme die grundlegenden Verfahren der quantitativen Datenanalyse ab. Im Folgenden werden die drei Programme kurz vorgestellt; eine Übersicht einzelner Aspekte der drei Statistikprogramme bietet Tabelle 1.

Das Statistikprogramm **Statistical Package for the Social Sciences (SPSS)** wurde 1968 an der University of Stanford entwickelt (Bühl 2016). Mit der ersten Programmversion entstand die Firma SPSS Inc., die 2009 von IBM übernommen wurde. Von 2009 bis 2010 wurde das Statistikprogramm unter dem Namen PASW (Predictive Analysis SoftWare) vermarktet, aktuell wird das Programm unter der Bezeichnung IBM SPSS Statistics vertrieben. SPSS ist ein modular aufgebautes Programmpaket. Das Basismodul ermöglicht das grundlegende Datenmanagement und deckt die grundlegenden Verfahren der quantitativen Datenanalyse ab. Für fortgeschrittene Analyseverfahren stehen weitere (kostenpflichtige) Module zur Verfügung. SPSS kann über das Menü oder die Befehlssprache bedient werden; das Programm steht auch in einer deutschen Sprachversion zur Verfügung. Bei SPSS handelt es sich um ein kommerziell vertriebenes Statistikprogramm; die Kosten für SPSS Statistics Base liegen – je nach Lizenztyp – zwischen 3000 und 7500 Euro. Für Hochschulen und Studierende gibt es erhebliche Rabatte. Studierende der FernUniversität in Hagen können eine SPSS-Semesterlizenz für 7,50 Euro erwerben (siehe Kapitel 2).

### SPSS im Überblick

Tabelle 1: SPSS, Stata und R im Überblick

	<b>SPSS</b>	<b>Stata</b>	<b>R</b>
Menüsteuerung	Ja	Ja	Nein
Befehlssprache	Ja	Ja	Ja
Deutsche Sprachversion	Ja	Nein	Ja
Kosten	Ja	Ja	Nein
Homepage	<a href="http://e.feu.de/spss">http://e.feu.de/spss</a>	<a href="http://www.stata.com">www.stata.com</a>	<a href="http://www.r-project.org">www.r-project.org</a>
Einführende Literatur	Wittenberg et al. (2014)	Kohler und Kreuter (2017)	Hatzinger et al. (2011)

Die in der Stadt College Station im US-Bundesstaat Texas ansässige Firma StataCorp ist Herausgeber des Statistikprogramms Stata, das erstmals 1985 veröffentlicht wurde. Stata ist ein umfangreiches statistisches Softwarepaket für den Einsatz in Forschung und Lehre. Stata kann – wie SPSS – über das Menü oder die Befehlssprache bedient werden; eine deutsche Sprachversion steht allerdings nicht zur Verfügung. Durch sogenannte (kostenfreie) Ado-Dateien wird der Leistungsumfang von Stata erweitert. Ados sind kleine Programme, die selbst oder von anderen Personen erstellt werden können. Diese Programme stehen über das Internet zur Verfügung und können unkompliziert in Stata integriert werden. Mit Stata/IC, Stata/SE und Stata/MP gibt es drei Programmversionen von Stata, die sich insbesondere

### Stata im Überblick

in der Zahl der maximal zu verarbeitenden Variablen und der Anzahl der Untersuchungseinheiten unterscheiden. So kann beispielsweise Stata/IC maximal 2048 Variablen verarbeiten; bei Stata/SE können bis zu 32.767 Variablen untersucht werden und bei Stata/MP sind es 120.000 Variablen (alle Angaben beziehen sich auf die Stata-Version 15). Stata ist ein kostenpflichtiges Statistikprogramm, die Kosten liegen je nach Nutzertyp (Educational oder Business) und Programmversion zwischen 400 und 3200 Euro. Beim Kauf mehrerer Lizenzen werden Rabatte eingeräumt; auch Studentenlizenzen sind deutlich günstiger. Deutscher Vertriebspartner für Stata ist die Firma Ditrach & Partner Consulting.

### R im Überblick

R ist ein Open-Source-Programm und daher auch frei (kostenlos) verfügbar. Auf der R-Homepage finden sich Informationen zur aktuellen Programmversion und den Downloadmöglichkeiten. R steht mit einer englischen Sprachversion zur Verfügung und wird weitgehend über eine Befehlssprache gesteuert. Allerdings existieren mittlerweile zahlreiche Ergänzungsprogramme, die den Einstieg in R erleichtern (z.B. RStudio). Diese Pakete werden von (anderen) R-Nutzern programmiert und der Öffentlichkeit zur Verfügung gestellt. Durch diese Programmpakete (Packages) kann der (zunächst begrenzte) Leistungsumfang von R weit über die Standardverfahren hinaus erweitert werden. Die kostenfreie Nutzung und die flexible Erweiterung haben R eine ständig wachsende Anzahl an Nutzern beschert. Der Verein R Foundation, der das Urheberrecht von R besitzt und verwaltet, ist das „Sprachrohr“ des Open-Source-Programms.

### Warum SPSS?

Für die ersten Schritte der quantitativen Datenanalyse sind SPSS, Stata und R gleichermaßen geeignet. Bei der Konzeption und Entwicklung dieses Kurses sprachen mehrere Gründe dafür, sich am Statistikprogramm SPSS zu orientieren bzw. die Einführung an SPSS auszurichten. Erstens wird SPSS an der FernUniversität in mehreren Studiengängen und Fakultäten eingesetzt (z.B. Psychologie und Wirtschaftswissenschaften), so dass zum einen für Studierende kostengünstige SPSS-Lizenzen zur Verfügung stehen und zum anderen die Studiengänge mit Blick auf das Statistikprogramm kompatibel sind. Zweitens werden an den Regional- und Studienzentren der FernUniversität – häufig in Kooperation mit den örtlichen Volkshochschulen – auch SPSS-Kurse angeboten. Drittens stehen über die Universitätsbibliothek Hagen mehrere SPSS-Bücher zum Download zur Verfügung (z.B. Akremi et al. 2011; Fromm 2012; Wittenberg et al. 2014; Schendera 2014), die die Ausführungen dieses Kurses ergänzen und vertiefen. Viertens hat SPSS auch außerhalb des Hochschulbereichs eine sehr große Verbreitung gefunden. Nach Bühl (2016, S. 39) ist SPSS „das weltweit verbreitetste Anwendungssystem zur statistischen Datenanalyse“.

Wichtiger als (endlose) Debatten über die Vor- und Nachteile einzelner Statistikprogramme ist die eigene Arbeit mit einem Statistikprogramm. Wer erste praktische Kenntnisse mit einem Programm gesammelt hat, kann dieses Wissen meist relativ einfach auf die anderen Programme übertragen. Mit anderen Worten: Wer weiß, wie SPSS funktioniert, der wird sich relativ schnell in Stata oder R einarbeiten können (und umgekehrt).



## 1.4 Aufbau des Kurses

Der Kurs „Quantitative Datenanalyse“ besteht aus elf Kapiteln. Im folgenden zweiten Kapitel wird das Statistikprogramm SPSS (Version 25) vorgestellt. Für die ersten Schritte mit SPSS wird ein Beispieldatensatz genutzt, der in der Moodle-Lernumgebung des Moduls M1 „Quantitative Methoden der Sozialwissenschaften“ zur Verfügung steht.

Im dritten Kapitel wird mit der Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS) ein zentraler Datensatz der Sozialwissenschaften in Deutschland vorgestellt. Dieser Datensatz eignet sich ausgezeichnet, um einerseits die Struktur eines quantitativen Datensatzes kennenzulernen und andererseits erste Erfahrungen in der quantitativen Datenanalyse zu sammeln. Erste Analysen mit dem ALLBUS-Datensatz werden im vierten Kapitel durchgeführt.

Bei der quantitativen Datenanalyse müssen immer wieder Variablen verändert und/oder erstellt werden. Deshalb werden im fünften Kapitel die Notwendigkeit der Datenmodifikation erläutert und einfache Techniken der Datenmodifikation vorgestellt. Diese Techniken werden exemplarisch an ALLBUS-Variablen demonstriert.

Im sechsten Kapitel werden univariate Analyseverfahren und im siebten Kapitel bivariate Analyseverfahren vorgestellt. Bei univariaten Analyseverfahren interessiert sich der Forscher für ein Merkmal (z.B. Einkommen). Mit bivariaten Analyseverfahren werden mögliche Zusammenhänge von zwei Merkmalen untersucht (z.B. Bildung und Einkommen). Bei der Prüfung von Hypothesen werden häufig bivariate Analyseverfahren verwendet.

Das achte Kapitel behandelt multivariate Analyseverfahren. Dabei werden Analyseverfahren vorgestellt, die den Einfluss mehrerer unabhängiger Variablen (z.B. Bildung und Geschlecht) auf eine abhängige Variable (z.B. Einkommen) schätzen können. Im Mittelpunkt steht die lineare Regression, das wichtigste und bekannteste Verfahren in den Sozialwissenschaften. Außerdem wird in Kapitel 8 auch die logistische Regression vorgestellt.

Im neunten Kapitel werden typische Grafiken vorgestellt, die mit SPSS erstellt werden können. Kapitel 10 bietet eine Übersicht weiterer Literatur zu SPSS.

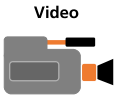
In diesem Kurs werden die Ergebnisse der Berechnungen jeweils anhand der SPSS-Abbildungen und SPSS-Tabellen erläutert und kommentiert.<sup>3</sup> In dieser Hinsicht dienen die Ausführungen als „Leseanleitung“ für die SPSS-Programmausgaben und sollen den geschriebenen Text mit der SPSS-Ausgabe verbinden. Bei Veröffentlichungen in Fachzeitschriften werden keine Abbildungen und Tabellen aus SPSS verwendet. Dies hat (mindestens) zwei Gründe: Erstens sind die üblichen SPSS-Tabellen nicht sonderlich ästhetisch. Zweitens sind die Angaben in SPSS-Abbildungen häufig auch redundant. Deshalb werden für die Präsentation empirischer Ergebnisse in Fachaufsätzen Tabellen und/oder Abbildungen gesondert erstellt. Auch in

**Verwendung  
des SPSS-Outputs**

---

<sup>3</sup> Die SPSS-Abbildungen basieren auf der Programmversion IBM SPSS Statistics 25. Als Vorlage für die Abbildungen wird das Format „Klassikstandard“ verwendet. Dieses kann über die Optionen (Reiter „Pivot-Tabellen“) ausgewählt werden.

Haus- und Abschlussarbeiten sollten SPSS-Tabellen nicht 1:1 übernommen werden. Eine erste Orientierung zur Gestaltung von Tabellen und Abbildungen bieten die Veröffentlichungen in einschlägigen Fachzeitschriften; detaillierte Informationen bieten die Ausführungen von Bauer (2010, 2015) und Plümper (2012, S. 168-174). Umfassende Darstellungen hat Miller (2005, 2015) vorgelegt.



Die Arbeit mit einem Statistikprogramm muss geübt werden. Deshalb werden ergänzend zu diesem Kurs in der Moodle-Lernumgebung Aufgabenblätter bereitgestellt, um typische Techniken der Datenmodifikation und Datenanalyse zu trainieren. Die Auseinandersetzung mit den Aufgaben soll die Vertrautheit mit dem Programm unterstützen und die Herausbildung (erster) eigenständiger Forschungsfähigkeit fördern. In der Moodle-Lernumgebung werden auch Video-Tutorials zur Verfügung gestellt, die die Abläufe bestimmter Analyseschritte darstellen.