

Markus Tausendpfund

Quantitative Datenanalyse

Eine Einführung mit SPSS

kultur- und
sozialwissenschaften

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.

Inhaltsverzeichnis

Abbildungsverzeichnis	VI
Tabellenverzeichnis.....	VIII
1 Einführung	9
1.1 Sozialwissenschaftlicher Forschungsprozess.....	9
1.2 Quantitative Datenanalyse	11
1.3 SPSS, Stata oder R?	12
1.4 Aufbau des Kurses	14
2 SPSS kennenlernen	17
2.1 Vor dem Start	17
2.1.1 Lizenz.....	17
2.1.2 Installation.....	17
2.1.3 Hilfe bei Fragen oder Problemen	18
2.2 Aufruf des Programms	18
2.2.1 Daten-Fenster	19
2.2.2 Syntax-Fenster	20
2.2.3 Ausgabe-Fenster.....	21
2.2.4 Optionen	22
2.3 Grundlagen der SPSS-Syntax	23
2.3.1 Vorteile der Syntax.....	23
2.3.2 Grundregeln der Syntax	24
2.3.3 Befehlssyntaxreferenz (Command Syntax Reference).....	25
2.4 Erste Analysen	26
2.4.1 Beispiel-Datensatz.....	26
2.4.2 Öffnen des Beispiel-Datensatzes	28
2.4.3 Variablenname, Variablenlabel und Wertelabel	29
2.4.4 Skalenniveau der Variablen	32
2.4.5 Fehlende Werte	33
2.4.6 Häufigkeitstabelle	33
2.4.7 Deskriptive Statistik.....	36
2.4.8 Abspeichern der SPSS-Syntax	37

3	ALLBUS kennenlernen	38
3.1	Was ist die ALLBUS?	38
3.2	Zugang zu den ALLBUS-Daten	38
3.2.1	GESIS – Leibniz-Institut für Sozialwissenschaften	39
3.2.2	Datenarchiv für Sozialwissenschaften	39
3.2.3	GESIS-Datenbestandskatalog	39
3.2.4	Download der ALLBUS-Daten	40
3.3	Arbeiten mit ALLBUS-Daten	46
3.3.1	Öffnen des ALLBUS-Datensatzes	46
3.3.2	Designgewicht	47
3.3.3	Dokumente zum ALLBUS-Datensatz	50
4	Erste Analysen mit den ALLBUS-Daten	51
4.1	Syntax-Datei erstellen	51
4.2	Häufigkeitstabellen	52
4.3	Deskriptive Statistiken	52
4.4	Gruppenvergleiche	53
4.5	Kreuztabellen	54
4.6	Syntax-Datei abspeichern	56
5	Grundlagen der Datenmodifikation	57
5.1	Warum Datenmodifikation?	57
5.2	Werkzeuge der Datenmodifikation	61
5.2.1	RECODE	61
5.2.2	COUNT	70
5.2.3	COMPUTE	73
5.3	Praktische Hinweise der Datenmodifikation	79
6	Univariate Analyseverfahren	80
6.1	Deskriptive Statistik	80
6.2	Wie werden Lage- und Streuungsmaße berechnet?	81

7	Bivariate Analyseverfahren.....	86
7.1	Kreuztabellen.....	86
7.2	Zusammenhangsmaße.....	91
7.2.1	Nominalskalierte Merkmale.....	92
7.2.2	Ordinalskalierte Merkmale.....	95
7.2.3	Metrische Merkmale.....	99
7.3	Mittelwertvergleiche.....	101
8	Multivariate Analyseverfahren.....	107
9	Weiterführende Literaturhinweise.....	108
10	Literaturverzeichnis.....	111

Abbildungsverzeichnis

Abbildung 1: Phasen eines quantitativen Forschungsprojekts.....	10
Abbildung 2: SPSS-Begrüßungsdialogfeld.....	18
Abbildung 3: SPSS-Datenansicht.....	19
Abbildung 4: SPSS-Variablenansicht.....	20
Abbildung 5: SPSS-Syntax-Fenster.....	21
Abbildung 6: SPSS-Ausgabe-Fenster.....	22
Abbildung 7: SPSS-Optionen (Dateispeicherorte).....	23
Abbildung 8: Befehlssyntaxreferenz (Command Syntax Reference).....	25
Abbildung 9: Hilfefunktion bei der SPSS-Syntax.....	26
Abbildung 10: Beispieldatensatz in der SPSS-Datenansicht.....	28
Abbildung 11: Beispieldatensatz in der SPSS-Variablenansicht.....	29
Abbildung 12: Ändern der Variablenlabels.....	31
Abbildung 13: Häufigkeitstabelle des Schulabschlusses.....	34
Abbildung 14: Häufigkeitstabelle des Schulabschlusses mit Variablen- und Wertelabels.....	35
Abbildung 15: SPSS-Optionen (Ausgabe).....	36
Abbildung 16: Deskriptive Angaben des Einkommens.....	37
Abbildung 17: GESIS – Leibniz-Institut für Sozialwissenschaften.....	40
Abbildung 18: GESIS-Datenbestandskatalog.....	41
Abbildung 19: Suche nach ALLBUS 2014.....	42
Abbildung 20: Übersicht zu der ALLBUS 2014.....	42
Abbildung 21: ALLBUS-Daten 2014.....	43
Abbildung 22: Anmeldung erforderlich.....	44
Abbildung 23: Registrierung für ZACAT, DBK und datorium.....	44
Abbildung 24: Registrierung.....	45
Abbildung 25: Download der ALLBUS-Daten 2014.....	46
Abbildung 26: Variablenansicht der ALLBUS-Daten 2014.....	47
Abbildung 27: Gewichtung über das Menü aktivieren.....	50
Abbildung 28: Exemplarische Syntax für die Analyse mit den ALLBUS-Daten 2014.....	52
Abbildung 29: Politikinteresse nach Geschlecht.....	54
Abbildung 30: Kreuztabelle des politischen Interesses nach Geschlecht.....	55
Abbildung 31: Häufigkeitstabelle des politischen Interesses.....	58
Abbildung 32: Häufigkeitstabelle des allgemeinen Schulabschlusses.....	59
Abbildung 33: Häufigkeitstabelle des sozialen Vertrauens.....	60
Abbildung 34: Häufigkeitstabelle des politischen Interesses (vor der Kodierung).....	62
Abbildung 35: Häufigkeitstabelle des politischen Interesses (nach der Kodierung).....	63
Abbildung 36: Häufigkeitstabelle des politischen Interesses (nach Formatanpassung).....	64
Abbildung 37: Häufigkeitstabelle des allgemeinen Schulabschlusses (vor der Kodierung).....	65
Abbildung 38: Häufigkeitstabelle des allgemeinen Schulabschlusses (nach der Kodierung).....	67
Abbildung 39: Häufigkeitstabelle des Vertrauens in Mitmenschen (vor der Kodierung).....	68
Abbildung 40: Häufigkeitstabelle des Vertrauens in Mitmenschen (nach der Kodierung).....	69
Abbildung 41: Häufigkeitstabelle Musikunterricht.....	71
Abbildung 42: Häufigkeitstabelle „kein Mitglied“.....	73
Abbildung 43: Häufigkeitstabelle zur Lebenszufriedenheit.....	82

Abbildung 44: Deskriptive Statistik zur Lebenszufriedenheit (Standardausgabe)	82
Abbildung 45: Deskriptive Statistik zur Lebenszufriedenheit (erweiterte Ausgabe).....	83
Abbildung 46: Deskriptive Angaben zur Lebenszufriedenheit (erweiterte Ausgabe).....	84
Abbildung 47: Balkendiagramm zur Lebenszufriedenheit (Angaben in Prozent)	85
Abbildung 48: Kreuztabelle der Demokratiezufriedenheit nach Gebiet (absolute Häufigkeiten) 87	
Abbildung 49: Kreuztabelle der Demokratiezufriedenheit nach Gebiet (Spaltenprozent).....	89
Abbildung 50: Kreuztabelle der Demokratiezufriedenheit nach Gebiet (Zeilenprozent).....	90
Abbildung 51: Kreuztabelle von Rauchen und Geschlecht	92
Abbildung 52: Berechnung von Chi-Quadrat	94
Abbildung 53: Berechnung von Phi und Cramer's V	94
Abbildung 54: Kreuztabelle der Bildung und des Politikinteresses (Spaltenprozent)	97
Abbildung 55: Berechnung von Spearman's roh	97
Abbildung 56: Berechnung von Spearman's roh und Kendall's Tau-b.....	99
Abbildung 57: Deskriptive Statistiken des Einkommens und der Lebenszufriedenheit.....	100
Abbildung 58: Korrelation des Einkommens und der allgemeinen Lebenszufriedenheit.....	101
Abbildung 59: Einkommen und Lebenszufriedenheit nach Geschlecht	102
Abbildung 60: Aussage über die Grundgesamtheit auf Basis einer Stichprobe.....	103
Abbildung 61: Nettoeinkommen in Abhängigkeit des Geschlechts.....	104
Abbildung 62: t-Test des Einkommens in Abhängigkeit des Geschlechts	105
Abbildung 63: Lebenszufriedenheit in Abhängigkeit des Geschlechts.....	105
Abbildung 64: t-Test der Lebenszufriedenheit in Abhängigkeit des Geschlechts.....	106

Tabellenverzeichnis

Tabelle 1: SPSS, Stata und R im Überblick	13
Tabelle 2: Beispieldaten	27
Tabelle 3: Beschreibung der Beispieldaten	27
Tabelle 4: Traditionelle Skalenniveaus und SPSS-Skalenniveaus im Vergleich	32
Tabelle 5: Datengrundlage für das Designgewicht in der ALLBUS 2014	48
Tabelle 6: Vergleich Mikrozensus 2013 und gewichtete ALLBUS-Daten	49
Tabelle 7: Datenmodifikation des politischen Interesses	61
Tabelle 8: Datenmodifikation des allgemeinen Schulabschlusses	66
Tabelle 9: Weitere Elemente des RECODE-Befehls	70
Tabelle 10: Mitgliedschaft in Vereinen (Anwendung des COUNT-Befehls)	72
Tabelle 11: Einfache Berechnungen mit dem COMPUTE-Befehl	74
Tabelle 12: Berechnung des Alters bei der ersten Berufstätigkeit	74
Tabelle 13: Möglichkeiten der Indexbildung mit SPSS	76
Tabelle 14: Funktionen mit COMPUTE	77
Tabelle 15: Unteranweisungen des FREQUENCIES- und DESCRIPTIVES-Befehls	84
Tabelle 16: Wichtige Zusammenhangsmaße bei der bivariaten Datenanalyse (Auswahl)	91
Tabelle 17: Arbeitstabelle für die Berechnung von Chi-Quadrat	93
Tabelle 18: Interpretation von Cramer's V	95
Tabelle 19: Spezifikationen der STATISTICS-Unteranweisung beim CROSSTABS-Befehl	95
Tabelle 20: Spezifikationen der PRINT-Unteranweisung beim NONPAR CORR-Befehl	98
Tabelle 21: Interpretation des Korrelationskoeffizienten nach Pearson	101

1 Einführung¹

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 2016). Nach der Entscheidung für ein Forschungsthema und der Entwicklung einer Forschungsfrage (1) müssen zunächst die zentralen Konzepte des Forschungsprojekts 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

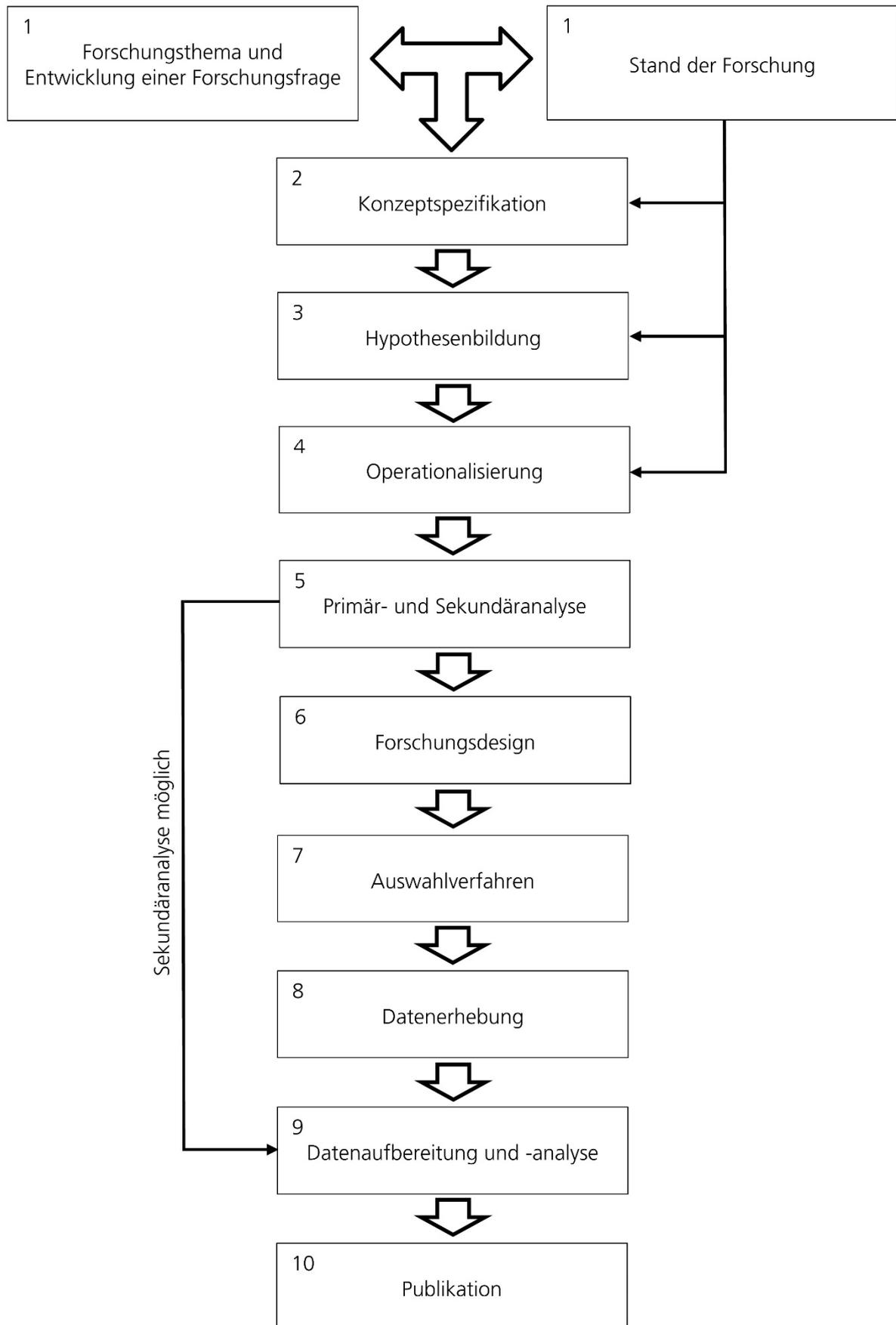
Bei einer Primäranalyse werden eigens Daten erhoben, um die Forschungsfrage zu untersuchen. Bei einer Sekundäranalyse werden existierende Daten genutzt, um die Forschungsfrage zu bearbeiten (5). Falls für die Bearbeitung einer Forschungsfrage bereits geeignetes Datenmaterial existiert (z.B. ALLBUS, ESS), 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 u. Best 2010); alle Verfahren setzen 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 Buch 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 2014). 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 steigt der rechnerische Aufwand allerdings erheblich an. Deshalb basieren solche Übungsaufgaben in der Regel auf kleinen 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 praktische 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 sicherlich SPSS, Stata und R.² 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, Pear-

² 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.

son's r) bis hin 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 2014). 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 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 12,50 Euro erwerben (siehe Kapitel 2).

SPSS im Überblick

Tabelle 1: SPSS, Stata und R im Überblick

	SPSS	Stata	R
Menüsteuerung	Ja	Ja	Nein
Befehlssprache	Ja	Ja	Ja
Deutsche Sprachversion	Ja	Nein	Nein
Kosten	Ja	Ja	Nein
Homepage	http://e.feu.de/spss	www.stata.com	www.r-project.org
Einführende Literatur	Wittenberg et al. (2014)	Kohler u. Kreuter (2012)	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

Stata im Überblick

Small, Stata/IC, Stata/SE und Stata/MP gibt es vier Programmversionen von Stata, die sich insbesondere in der Zahl der maximal zu verarbeitenden Variablen und der Anzahl der Untersuchungseinheiten unterscheiden. So kann beispielsweise Stata Small maximal 99 Variablen und rund 1200 Untersuchungseinheiten verarbeiten; bei Stata/SE können bis zu 32.767 Variablen untersucht werden und die Anzahl der Untersuchungseinheiten ist nur durch den Arbeitsspeicher begrenzt. 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 erhebliche Rabatte eingeräumt; auch Studentenlizenzen sind deutlich günstiger. Deutscher Vertriebspartner für Stata ist die Firma Dittrich & 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, 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), die die Ausführungen dieses Kurses ergänzen und vertiefen. Viertens hat SPSS auch außerhalb des Hochschulbereichs eine sehr große Verbreitung gefunden. Nach Bühl (2014, 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 aktuell aus zehn Kapiteln (Stand: Sommersemester 2016). Im folgenden zweiten Kapitel wird das Statistikprogramm SPSS 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.

Der vorliegende (gedruckte) Kurs stellt noch kein fertiges Produkt dar. Weitere Kapitel werden unter anderem auch multivariate Analyseverfahren (z.B. lineare Regression) in den Blick nehmen (Kapitel 8). Der Kapitelentwurf wird in der Moodle-Lernumgebung des Moduls M1 „Quantitative Methoden der Sozialwissenschaften“ zur Verfügung gestellt. Die aktuelle pdf-Version des Kurses ist über den Virtuellen Studienplatz erhältlich. Kapitel 9 bietet eine Übersicht weiterer Literatur zu SPSS.

**Weitere Kapitel
in Moodle**

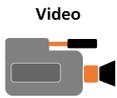
In diesem Kurs werden die Ergebnisse der Berechnungen jeweils an Hand der SPSS-Abbildungen und Tabellen erläutert und kommentiert.³ In dieser Perspektive dienen die Ausführungen als „Leseanleitung“ für die SPSS-Programmausgaben und sollen den geschriebenen Text mit der SPSS-Ausgabe verbinden. Bei Veröffentlichungen in Fachzeitschriften werden keine SPSS-Abbildungen und Tabellen 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 (z.B. t-Test). Deshalb werden für die Präsentation empirischer Ergebnisse in Fachaufsätzen Tabellen und/oder Abbildungen gesondert erstellt. Auch in 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). Eine umfassende Darstellung hat Miller (2005) vorgelegt.

**Verwendung des
SPSS-Outputs**

Die Arbeit mit einem Statistikprogramm muss geübt werden. Deshalb werden ergänzend zu diesem Kurs in der Moodle-Lernumgebung Aufgabenblätter bereitgestellt, um typische Techniken

³ Die SPSS-Abbildungen basieren auf der Programmversion IBM SPSS Statistics 23.

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 anschaulich darstellen. Mit den Video-Tutorials soll auch möglichen Ängsten bei der Arbeit mit einem Statistikprogramm entgegengewirkt werden.