

Markus Tausendpfund

Quantitative Datenanalyse. Eine Einführung mit R

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. Wir weisen darauf hin, dass die vorgenannten Verwertungsalternativen je nach Ausgestaltung der Nutzungsbedingungen bereits durch Einstellen in Cloud-Systeme verwirklicht sein können. Die FernUniversität bedient sich im Falle der Kenntnis von Urheberrechtsverletzungen sowohl zivil- als auch strafrechtlicher Instrumente, um ihre Rechte geltend zu machen.

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

Vorwort

Die vorliegende Lerneinheit behandelt die sozialwissenschaftliche Datenanalyse, R und insbesondere die sozialwissenschaftliche Datenanalyse mit R. Dabei werden Kenntnisse vermittelt, um einfache Analysen selbstständig mit der Software R durchführen zu können.

Der Text entspricht dabei weniger einem klassischen Lehrbuch, sondern eher einem Begleitkurs für die Auseinandersetzung mit dem Programm R bzw. RStudio. Die Lerneinheit soll das Interesse an sozialwissenschaftlichen Fragestellungen wecken sowie die Möglichkeiten und Grenzen der quantitativen Datenanalyse aufzeigen.

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. Ein besonderer Dank geht an Verena Bade, Christian Cleve, Dorothee Köstlin und Simon Stocker, die sich intensiv mit der Lerneinheit beschäftigt und mich auf Ungenauigkeiten sowie Tippfehler aufmerksam gemacht haben.

Aktuelle Ergänzungen und mögliche Korrekturen zu dieser Lerneinheit 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 Videos veröffentlicht, die die Auseinandersetzung mit den Inhalten dieser Lerneinheit fördern sollen.

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

Die vorliegende Lerneinheit ist kein „Endprodukt“. Die regelmäßige Aktualisierung stellt eine Daueraufgabe dar. Deshalb freue ich mich sehr über alle Hinweise und Anregungen zur weiteren Verbesserung der Lerneinheit (E-Mail: Markus.Tausendpfund@fernuni-hagen.de).

Hagen, im Dezember 2023

Markus Tausendpfund

Inhaltsverzeichnis

Abbildungsverzeichnis	VII
Tabellenverzeichnis	VIII
1 Einführung	9
1.1 Sozialwissenschaftlicher Forschungsprozess	9
1.2 Quantitative Datenanalyse	11
1.3 Warum R?	12
1.4 Struktur der Lerneinheit	14
2 R und RStudio kennenlernen	15
2.1 Installation	15
2.1.1 R	15
2.1.2 RStudio	16
2.1.3 Pakete	20
2.1.4 Aktualisierungen	20
2.1.5 posit Cloud: eine Alternative zur lokalen R-Installation	20
2.2 Ein erster Überblick	21
2.2.1 Console	21
2.2.2 Skripte	23
2.2.3 Befehle	24
2.2.4 Objekte	25
2.2.5 Vektoren	26
2.2.6 Tabellen	28
2.2.7 Erste Analysen	29
2.2.8 Fehlende Werte	31
2.2.9 Hilfe	32
3 Arbeiten mit R	33
3.1 Pakete installieren und laden	33
3.2 Daten laden	35
3.3 Daten importieren	36
3.3.1 Excel	37
3.3.2 SPSS	37
3.3.3 CSV	38
3.4 Objekttypen	39
3.5 Datenstrukturen	43

3.6	Saubere Skripte erstellen.....	43
3.7	Projekte in R.....	45
4	Beispieldatensatz.....	46
4.1	Daten und Pakete.....	46
4.2	Beispieldatensatz kennenlernen.....	46
4.3	Pipe-Operator.....	47
4.4	Datenmanagement mit dplyr.....	48
4.4.1	Variablen auswählen.....	48
4.4.2	Variablen umbenennen.....	49
4.4.3	Variablen filtern.....	50
4.4.4	Variablen verändern.....	50
4.4.5	Einfache Berechnungen.....	51
4.4.6	Weitere Optionen.....	52
4.5	Datenaufbereitung mit sjmisc.....	52
4.5.1	Variablenwerte ändern.....	52
4.5.2	Variablen zusammenfassen.....	55
4.5.3	Werte zählen.....	57
4.5.4	Weitere Optionen.....	57
4.6	Labels konvertieren.....	58
5	Univariate Datenanalyse.....	60
5.1	Daten und Pakete.....	60
5.2	Häufigkeitstabellen.....	60
5.3	Lagemaße.....	62
5.4	Streuungsmaße.....	63
5.5	Formmaße.....	64
5.6	Kompakte Übersichten.....	66
6	Bivariate Datenanalyse.....	67
6.1	Daten und Pakete.....	67
6.2	Univariate Statistiken nach Gruppen.....	67
6.3	Kreuztabellen.....	68
6.4	Zusammenhangsmaße.....	71
6.4.1	Nominalskalierte Merkmale.....	72
6.4.2	Ordinalskalierte Merkmale.....	75
6.4.3	Metrische Merkmale.....	77

7	Multivariate Datenanalyse.....	80
7.1	Einführung.....	80
7.2	Lineare Regression.....	82
7.2.1	Das Grundmodell.....	82
7.2.2	Daten und Pakete.....	90
7.2.3	Lineare Regression mit R.....	90
7.2.4	Interpretation der Ergebnisse.....	93
7.2.5	Weitere Möglichkeiten.....	96
7.2.6	Anwendungsvoraussetzungen.....	102
7.2.7	Praktische Hinweise.....	103
7.3	Logistische Regression.....	105
7.3.1	Das Grundmodell.....	105
7.3.2	Daten und Pakete.....	111
7.3.3	Logistische Regression mit R.....	112
7.3.4	Interpretation der Ergebnisse.....	113
7.3.5	Weitere Möglichkeiten.....	116
7.3.6	Praktische Hinweise.....	121
8	Inferenzstatistik.....	123
8.1	Daten und Pakete.....	123
8.2	Konfidenzintervalle.....	124
8.3	Mittelwertvergleiche (t-Test).....	125
9	Grafiken.....	131
9.1	Einführung.....	131
9.2	Ausgewählte Diagramme.....	133
9.2.1	Säulen- und Balkendiagramm.....	133
9.2.2	Kreisdiagramm.....	135
9.2.3	Histogramm.....	136
9.2.4	Boxplot.....	137
9.2.5	Streudiagramm.....	139
9.2.6	Liniendiagramm.....	141
9.3	Weitere Pakete.....	142
	Pakete im Überblick.....	143
	Beispieldatensatz im Überblick.....	145
	Literaturverzeichnis.....	149

Abbildungsverzeichnis

Abbildung 1: Phasen eines quantitativen Forschungsprojekts	10
Abbildung 2: The Comprehensive R Archive Network (CRAN)	15
Abbildung 3: Startbildschirm von R	16
Abbildung 4: RStudio mit drei Fenstern	17
Abbildung 5: RStudio mit vier Fenstern	18
Abbildung 6: Global Options bei RStudio	19
Abbildung 7: posit Cloud	21
Abbildung 8: Streudiagramm des Einkommens in Abhängigkeit des Alters	31
Abbildung 9: Installierte Pakete	34
Abbildung 10: „Import Dataset“-Funktion in RStudio	36
Abbildung 11: Excel-Datensatz importieren	37
Abbildung 12: SPSS-Datensatz importieren	38
Abbildung 13: Normalverteilung	65
Abbildung 14: Auswahl von regressionsanalytischen Verfahren	80
Abbildung 15: Streudiagramm von Alter und Einkommen	83
Abbildung 16: Streudiagramm von Einkommen und Alter mit Regressionsgerade	84
Abbildung 17: Empirischer und geschätzter Wert einer linearen Regression	85
Abbildung 18: Grafische Darstellung einer multiplen Regression	89
Abbildung 19: Lineare Regressionsfunktion	107
Abbildung 20: Logistische Regressionsfunktion	108
Abbildung 21: Logistische Regression (Beispieldaten)	111
Abbildung 22: Grafische Darstellung der logistischen Regression I	120
Abbildung 23: Grafische Darstellung der logistischen Regression II	121
Abbildung 24: Aussage über die Grundgesamtheit auf Basis einer Zufallsstichprobe	127
Abbildung 25: Säulen- und Balkendiagramm	134
Abbildung 26: Kreisdiagramm	135
Abbildung 27: Histogramm	136
Abbildung 28: Elemente eines Boxplots	137
Abbildung 29: Boxplots	138
Abbildung 30: Plotsymbole bei R	140
Abbildung 31: Streudiagramm	140
Abbildung 32: Liniendiagramm	141

Tabellenverzeichnis

Tabelle 1: SPSS, Stata und R im Überblick.....	13
Tabelle 2: Mathematische Funktionen in R	22
Tabelle 3: Logische Abfragen in R	22
Tabelle 4: Installieren und Laden von Paketen	34
Tabelle 5: Beispieldatensatz (peanuts_r)	35
Tabelle 6: Ausgewählte Funktionen zum Testen und Konvertieren von Objekten.....	41
Tabelle 7: Objekttypen in den Datensätzen	42
Tabelle 8: Ausgewählte Funktionen von dplyr	48
Tabelle 9: Auswahl von Filter-Möglichkeiten.....	50
Tabelle 10: Ausgewählte Möglichkeiten des rec-Arguments.....	54
Tabelle 11: Interpretation von Schiefe und Wölbung	66
Tabelle 12: Kreuztabelle zwischen Wahlbeteiligung und Bildung.....	70
Tabelle 13: Zusammenhangsmaße bei der bivariaten Datenanalyse	72
Tabelle 14: Arbeitstabelle für die Berechnung von Chi-Quadrat	73
Tabelle 15: Interpretation von Cramer's V	75
Tabelle 16: Interpretation von Spearman's Rho	77
Tabelle 17: Interpretation des Korrelationskoeffizienten nach Pearson.....	79
Tabelle 18: Beispieldaten für Alter und Einkommen.....	82
Tabelle 19: Bivariate Regression (Beispieldaten)	87
Tabelle 20: Beispieldaten für Alter, Einkommen und Berufserfahrung.....	88
Tabelle 21: Multiple Regression (Beispieldaten).....	89
Tabelle 22: Bestimmungsfaktoren der Lebenszufriedenheit I.....	91
Tabelle 23: Bestimmungsfaktoren der Lebenszufriedenheit II	92
Tabelle 24: Bestimmungsfaktoren der Lebenszufriedenheit III.....	93
Tabelle 25: Bestimmungsfaktoren der Lebenszufriedenheit IV	97
Tabelle 26: Bestimmungsfaktoren der Lebenszufriedenheit V (standardisierte Koeffizienten)....	98
Tabelle 27: Bestimmungsfaktoren der Lebenszufriedenheit V (mit Faktoren).....	100
Tabelle 28: Bestimmungsfaktoren der Lebenszufriedenheit VI (Regressionstabelle mit sjPlot)..	101
Tabelle 29: Informationen zu einer linearen Regression.....	102
Tabelle 30: Beispieldaten für Rauchen und Alter in Jahren	106
Tabelle 31: Bestimmungsfaktoren des Tabakkonsums	110
Tabelle 32: Bestimmungsfaktoren der Wahlbeteiligung	113
Tabelle 33: Bedeutung der Asteriske	115
Tabelle 34: Logistische Regressionskoeffizienten und Odds Ratio im Vergleich.....	117
Tabelle 35: t-Test der Lebenszufriedenheit in Abhängigkeit des Geschlechts.....	128
Tabelle 36: t-Test der Demokratiezufriedenheit in Abhängigkeit des politischen Interesses....	129
Tabelle 37: Argumente beim t-Test.....	130
Tabelle 38: Argumente bei Grafikbefehlen	131
Tabelle 39: Pakete zur Erstellung von Grafiken	142
Tabelle 40: Pakete im Überblick	143
Tabelle 41: Beispieldatensatz im Überblick	145

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 dieser Lerneinheit und verweist auf ergänzende Materialien in der Moodle-Lernumgebung.

Vorschau



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 2018).

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.

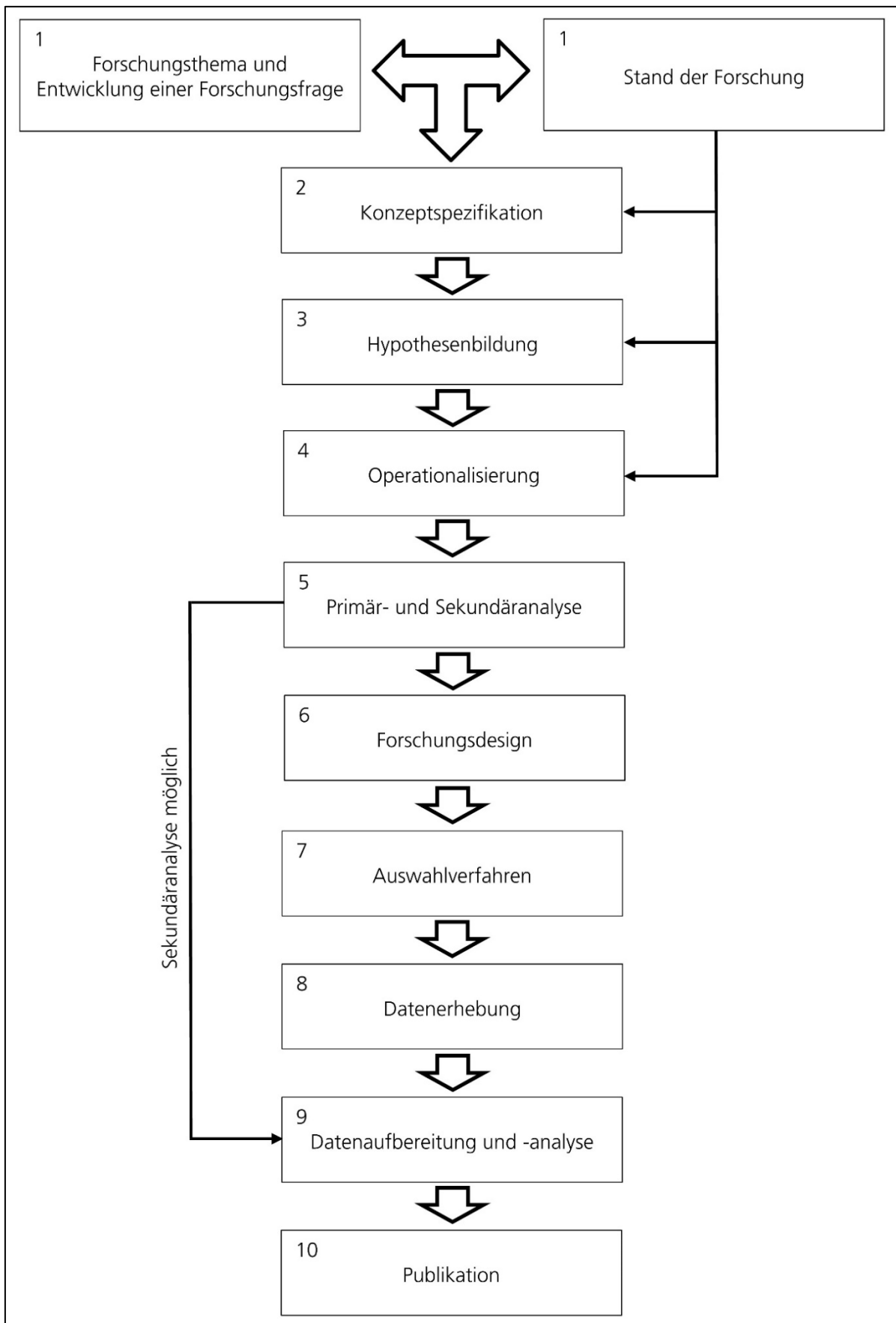
Bedeutung des Forschungsstands

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, 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; Backhaus et al. 2021); 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.

Abbildung 1: Phasen eines quantitativen Forschungsprojekts



Quelle: Eigene Darstellung

1.2 Quantitative Datenanalyse

Sozialwissenschaftlerinnen wollen soziale Sachverhalte beschreiben und erklären (King et al. 1994, S. 7-8). Fördert ein höheres Einkommen die Lebenszufriedenheit? Warum beteiligen sich Menschen mit höherer Bildung eher an Wahlen als Personen mit geringer Bildung? Welche Faktoren beeinflussen den Tabakkonsum?

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 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. arithmetisches Mittel), bivariate Zusammenhangsmaße (z.B. Pearson's r) sowie die Grundlagen der linearen und logistischen Regression behandelt. Videos, Tests und Übungsaufgaben in der Moodle-Lernumgebung sollen das Verständnis für die einzelnen Analyseverfahren fördern. Dabei werden beispielsweise das arithmetische Mittel 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 stehen 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 die vorliegende Lerneinheit an. Die Lerneinheit will einerseits die Vorgehensweise bei der quantitativen Datenanalyse verdeutlichen und andererseits praktische Kenntnisse in der Anwendung eines Statistikprogramms vermitteln. Die Lerneinheit behandelt dabei typische 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 aber 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“, bei fast jeder zweiten Ausschreibung werden Methodenkenntnisse erwartet (Schreiber 2016). 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 (Kohler 2016).

Heute stehen 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 Warum R?

Für die Datenanalyse stehen heute mehrere leistungsfähige Statistikprogramme zur Verfügung. Die bekanntesten Programme sind sicherlich SPSS, Stata und R. 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 zu fortgeschrittenen Analyseverfahren (z.B. lineare und logistische Regression) sowie Signifikanztests decken alle drei Programme die grundlegenden Verfahren der quantitativen Datenanalyse ab. Eine Übersicht der drei Programme bietet Tabelle 1.

Diese Lerneinheit bietet eine Einführung in die Programme R und RStudio. R ist ein Open-Source-Programm und frei (kostenlos) verfügbar. Auf der R-Homepage unter <https://www.r-project.org> finden sich Informationen zur aktuellen Programmversion und den Downloadmöglichkeiten. R wird weitgehend über eine Befehlssprache gesteuert. Allerdings existieren mittlerweile zahlreiche Ergänzungsprogramme, die den Einstieg in R erleichtern (z.B. RStudio). Durch weitere Programmpakete (Packages) kann der 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.

Tabelle 1: SPSS, Stata und R im Überblick

	SPSS	Stata	R
Menüsteuerung	Ja	Ja	Nein
Befehlssprache	Ja	Ja	Ja
Deutsche Sprachversion	Ja	Nein	Ja
Kosten	Ja	Ja	Nein
Homepage	https://e.feu.de/spss	www.stata.com	www.r-project.org
Einführende Literatur	Tausendpfund (2022)	Kohler und Kreuter (2017)	Luhmann (2020)

Quelle: Eigene Zusammenstellung

Seit einigen Jahren erleben die Sozialwissenschaften ein neues Zeitalter: Vielfalt und Umfang sozialwissenschaftlicher Daten nehmen rapide zu, unterschiedliche Datenbestände werden systematisch verknüpft und immer leistungsfähigere Hardware erlaubt die Analyse immer größerer Datenbestände. Diese Datenbestände sowie neuere Analysetechniken erfordern allerdings neue Kompetenzen, die im Rahmen der Methodenausbildung vermittelt werden müssen. Dabei ist auch die Softwareausbildung in den Blick zu nehmen, die für die Arbeit mit den alten und neuen Datenbeständen erforderlich ist (Munzert 2018; Döring und Hocks 2021). Dabei sprechen mehrere Gründe für R.

Erstens ist R ein Open-Source-Programm und steht für mehrere Plattformen (Windows, Mac und Linux) kostenfrei zur Verfügung. Über frei verfügbare Erweiterungen (Packages) kann der Funktionsumfang von R beträchtlich erweitert werden. Mit Blick auf Aufbereitung, Visualisierung, Analyse von Daten und Ankopplung an Datenbanksysteme fungiert R damit als Programmierumgebung, die für unterschiedlichste Aufgaben genutzt werden kann.

Zweitens ist R methoden-agnostisch und kann sowohl in der quantitativen und qualitativen Sozialforschung eingesetzt werden. Es existieren Erweiterungspakete sowohl für die quantitative (z.B. Regression, Faktorenanalyse) als auch für die qualitative Sozialforschung (z.B. Qualitative Comparative Analysis).

Drittens gilt R als zukunftssicher. Als Open-Source-Programm wird R ständig weiterentwickelt. R überwindet zudem die Ein-Datensatzlogik und verfügt über Schnittstellen zu webbasierten Datensätzen. R kann daher auch genutzt werden, um unstrukturierte oder strukturierte Daten zu sammeln und damit weiterzuarbeiten.