

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 2018 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. 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 mögliche Korrekturen 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 Mai 2020

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.....	36
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	Download der ALLBUS-Daten 2018.....	43
3.3	Arbeiten mit ALLBUS-Daten	49
3.3.1	Öffnen des ALLBUS-Datensatzes	49
3.3.2	Designgewicht.....	50
3.3.3	Variablennamen im ALLBUS	53
3.3.4	Fehlende Werte im ALLBUS.....	53
3.3.5	Dokumente zum ALLBUS-Datensatz	54
4	Erste Analysen mit den ALLBUS-Daten.....	55
4.1	Syntax-Datei erstellen	55
4.2	Häufigkeitstabellen	56
4.3	Deskriptive Statistiken.....	57
4.4	Gruppenvergleiche	57
4.5	Kreuztabellen	59
4.6	Syntax-Datei abspeichern.....	60
5	Grundlagen der Datenmodifikation	61
5.1	Warum Datenmodifikation?.....	61
5.2	Werkzeuge der Datenmodifikation.....	65
5.2.1	RECODE	65
5.2.2	COUNT	74
5.2.3	COMPUTE	77
5.2.4	IF	83
5.3	Praktische Hinweise der Datenmodifikation	89
6	Univariate Datenanalyse.....	91
6.1	Lagemaße	91
6.2	Streuungsmaße	94
6.3	Formmaße.....	96
6.4	Zusammenfassung.....	99
7	Bivariate Datenanalyse	103
7.1	Kreuztabellen	103
7.2	Zusammenhangsmaße.....	108
7.2.1	Nominalskalierte Merkmale.....	109

7.2.2	Ordinalskalierte Merkmale	113
7.2.3	Metrische Merkmale	117
7.3	Mittelwertvergleiche	119
8	Multivariate Datenanalyse	124
8.1	Einführung	124
8.2	Lineare Regression	126
8.2.1	Das Grundmodell	126
8.2.2	Lineare Regression mit SPSS	134
8.2.3	Interpretation der Ergebnisse	137
8.2.4	Weitere Optionen der Regression mit SPSS	144
8.2.5	Anwendungsvoraussetzungen	151
8.2.6	Praktische Hinweise	153
8.3	Logistische Regression	154
8.3.1	Das Grundmodell	154
8.3.2	Logistische Regression mit SPSS	160
8.3.3	Interpretation der Ergebnisse	164
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	184
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 – Suchergebnis	45
Abbildung 22: GESIS – ALLBUS-Übersichtsseite	45
Abbildung 23: GESIS – Übersicht des ALLBUS 2018	46
Abbildung 24: GESIS – Hinweis „Login nötig“	46
Abbildung 25: GESIS – Login	47
Abbildung 26: GESIS – Registrierung	48
Abbildung 27: GESIS – Download der ALLBUS-Daten 2018 (ZA5270)	49
Abbildung 28: Variablenansicht der ALLBUS-Daten 2018	50
Abbildung 29: Gewichtung über das Menü aktivieren	53
Abbildung 30: Exemplarische Syntax für die Analyse mit den ALLBUS-Daten 2018	56
Abbildung 31: Politikinteresse nach Geschlecht	58
Abbildung 32: Kreuztabelle des politischen Interesses nach Geschlecht	60
Abbildung 33: Häufigkeitstabelle des politischen Interesses	62
Abbildung 34: Häufigkeitstabelle des allgemeinen Schulabschlusses	63
Abbildung 35: Häufigkeitstabelle des sozialen Vertrauens	64
Abbildung 36: Häufigkeitstabelle des politischen Interesses (vor der Kodierung)	66
Abbildung 37: Häufigkeitstabelle des politischen Interesses (nach der Kodierung)	67
Abbildung 38: Häufigkeitstabelle des politischen Interesses (nach Formatanpassung)	68
Abbildung 39: Häufigkeitstabelle des allgemeinen Schulabschlusses (vor der Kodierung)	69
Abbildung 40: Häufigkeitstabelle des allgemeinen Schulabschlusses (nach der Kodierung)	71
Abbildung 41: Häufigkeitstabelle des Vertrauens in Mitmenschen (vor der Kodierung)	72
Abbildung 42: Häufigkeitstabelle des Vertrauens in Mitmenschen (nach der Kodierung)	73
Abbildung 43: Häufigkeitstabelle Mitglied in Gewerkschaft und Partei	75

Abbildung 44: Häufigkeitstabelle „kein Mitglied“	77
Abbildung 45: Häufigkeitstabelle von va01 (Wichtigkeit von Ruhe und Ordnung).....	84
Abbildung 46: Häufigkeitstabelle von va02 (Wichtigkeit von Bürgereinfluss).....	85
Abbildung 47: Häufigkeitstabelle von va03 (Wichtigkeit der Inflationsbekämpfung)	85
Abbildung 48: Häufigkeitstabelle von va04 (Wichtigkeit von freier Meinungsäußerung).....	85
Abbildung 49: Häufigkeitstabelle des erstellten Inglehart-Index.....	88
Abbildung 50: Häufigkeitstabelle zur Lebenszufriedenheit	92
Abbildung 51: Lagemaße mit dem FREQUENCIES-Befehl	93
Abbildung 52: DESCRIPTIVES-Befehl (Standardausgabe)	93
Abbildung 53: Lagemaße mit dem DESCRIPTIVES-Befehl	94
Abbildung 54: Streuungsmaße mit dem FREQUENCIES-Befehl	95
Abbildung 55: Streuungsmaße mit dem DESCRIPTIVES-Befehl	96
Abbildung 56: Normalverteilung	97
Abbildung 57: Formmaße mit dem FREQUENCIES-Befehl.....	98
Abbildung 58: Formmaße mit dem DESCRIPTIVES-Befehl.....	98
Abbildung 59: Lage-, Streuungs- und Formmaße mit dem FREQUENCIES-Befehl	99
Abbildung 60: Lage-, Streuungs- und Formmaße mit dem DESCRIPTIVES-Befehl	100
Abbildung 61: Säulendiagramm zur Lebenszufriedenheit (Angaben in Prozent).....	102
Abbildung 62: Kreuztabelle der Demokratiezufriedenheit nach Gebiet (Häufigkeiten).....	104
Abbildung 63: Kreuztabelle der Demokratiezufriedenheit nach Gebiet (Spaltenprozente)	106
Abbildung 64: Kreuztabelle der Demokratiezufriedenheit nach Gebiet (Zeilenprozente).....	107
Abbildung 65: Kreuztabelle von Rauchen und Geschlecht	109
Abbildung 66: Berechnung von Chi-Quadrat.....	111
Abbildung 67: Berechnung von Phi und Cramer's V	111
Abbildung 68: Kreuztabelle der Bildung und des Politikinteresses (Spaltenprozente).....	114
Abbildung 69: Berechnung von Spearman's Rho	115
Abbildung 70: Berechnung von Spearman's Rho und Kendall's Tau-B	116
Abbildung 71: Deskriptive Statistiken des Einkommens und der Lebenszufriedenheit.....	118
Abbildung 72: Korrelation des Einkommens und der Lebenszufriedenheit	118
Abbildung 73: Einkommen und Lebenszufriedenheit nach Geschlecht	120
Abbildung 74: Aussage über die Grundgesamtheit auf Basis einer Zufallsstichprobe.....	120
Abbildung 75: Einkommen in Abhängigkeit des Geschlechts	122
Abbildung 76: t-Test des Einkommens in Abhängigkeit des Geschlechts	122
Abbildung 77: Lebenszufriedenheit in Abhängigkeit des Geschlechts.....	123
Abbildung 78: t-Test der Lebenszufriedenheit in Abhängigkeit des Geschlechts.....	123
Abbildung 79: Auswahl von regressionsanalytischen Verfahren	124
Abbildung 80: Streudiagramm von Alter und Einkommen	127
Abbildung 81: Streudiagramm von Einkommen und Alter mit Regressionsgerade.....	128
Abbildung 82: Empirischer und geschätzter Wert einer linearen Regression.....	129
Abbildung 83: Bivariate Regression (Beispieldaten)	131
Abbildung 84: Grafische Darstellung einer multiplen Regression	133
Abbildung 85: Multiple Regression (Beispieldaten).....	133
Abbildung 86: Ergebnisse einer linearen Regression	137
Abbildung 87: Ergebnisse der Regressionsanalyse (ANOVA).....	141
Abbildung 88: TSS, RSS und MSS	142

Abbildung 89: Modellzusammenfassung einer linearen Regression	143
Abbildung 90: Grundinformationen zur linearen Regression	144
Abbildung 91: Ungewichtete Häufigkeitstabelle zum allgemeinen Schulabschluss	145
Abbildung 92: Ergebnisse der linearen Regression mit Bildungsgruppen I	147
Abbildung 93: Ergebnisse der linearen Regression mit Bildungsgruppen II	148
Abbildung 94: Deskriptive Statistiken der abhängigen und unabhängigen Variablen.....	150
Abbildung 95: Lineare Regressionsfunktion	155
Abbildung 96: Logistische Regressionsfunktion	157
Abbildung 97: Regressionskoeffizienten einer logistischen Regression (Beispieldaten)	159
Abbildung 98: Logistische Regression (Beispieldaten).....	160
Abbildung 99: Ergebnisse einer logistischen Regression	165
Abbildung 100: Grafische Darstellung der logistischen Regression	168
Abbildung 101: Bewertung der Modellgüte	170
Abbildung 102: Iterationsprotokoll einer logistischen Regression.....	172
Abbildung 103: Klassifizierungstabelle einer logistischen Regression (nur Konstante).....	173
Abbildung 104: Klassifizierungstabelle einer logistischen Regression (mit uVs).....	173
Abbildung 105: Säulendiagramm der Lebenszufriedenheit.....	180
Abbildung 106: Balkendiagramm der Lebenszufriedenheit (bearbeitet)	181
Abbildung 107: Kreisdiagramm (bearbeitet)	183
Abbildung 108: Histogramm des Alters (bearbeitet).....	184
Abbildung 109: Elemente eines Boxplots.....	185
Abbildung 110: Boxplot der Lebenszufriedenheit	186
Abbildung 111: Boxplot der Lebenszufriedenheit (bearbeitet)	186
Abbildung 112: Boxplots der Demokratiezufriedenheit nach pol. Interesse (bearbeitet).....	187
Abbildung 113: Streudiagramm von Alter und Einkommen (bearbeitet)	188
Abbildung 114: 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 im ALLBUS 2018.....	51
Tabelle 6: Vergleich Mikrozensus 2017 und gewichtete ALLBUS-Daten 2018.....	52
Tabelle 7: Datenmodifikation des politischen Interesses.....	65
Tabelle 8: Datenmodifikation des allgemeinen Schulabschlusses.....	70
Tabelle 9: Elemente des RECODE-Befehls.....	74
Tabelle 10: Mitgliedschaft in Vereinen (Anwendung des COUNT-Befehls).....	76
Tabelle 11: Einfache Berechnungen mit dem COMPUTE-Befehl.....	78
Tabelle 12: Berechnung des Alters am Ende der Erwerbstätigkeit.....	78
Tabelle 13: Möglichkeiten der Indexbildung mit SPSS.....	80
Tabelle 14: Funktionen mit COMPUTE.....	81
Tabelle 15: Konstruktion des Inglehart-Index.....	86
Tabelle 16: Logische Operatoren des IF-Befehls.....	88
Tabelle 17: Relationale Operatoren des IF-Befehls.....	89
Tabelle 18: Interpretation von Schiefe und Wölbung.....	97
Tabelle 19: Unteranweisungen des FREQUENCIES- und des DESCRIPTIVES-Befehls.....	100
Tabelle 20: Wichtige Zusammenhangsmaße bei der bivariaten Datenanalyse (Auswahl).....	108
Tabelle 21: Arbeitstabelle für die Berechnung von Chi-Quadrat.....	110
Tabelle 22: Interpretation von Cramer's V.....	112
Tabelle 23: Spezifikationen der STATISTICS-Unteranweisung beim CROSSTABS-Befehl.....	112
Tabelle 24: Spezifikationen der PRINT-Unteranweisung beim NONPAR CORR-Befehl.....	115
Tabelle 25: Interpretation von Spearman's Rho.....	116
Tabelle 26: Interpretation des Korrelationskoeffizienten nach Pearson.....	119
Tabelle 27: Beispieldaten für Alter und Einkommen.....	126
Tabelle 28: Beispieldaten für Alter, Einkommen und Berufserfahrung.....	132
Tabelle 29: Bestimmungsfaktoren des Einkommens.....	151
Tabelle 30: Beispieldaten für Rauchen und Alter in Jahren.....	155
Tabelle 31: Angaben zur Modellgüte im Vergleich.....	171
Tabelle 32: Determinanten der Parteiidentifikation.....	174
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.