
Sozialwissenschaftliche Datenanalyse

Johannes Kopp • Daniel Lois

Sozialwissenschaftliche Datenanalyse

Eine Einführung

2., überarbeitete und aktualisierte Auflage

 Springer VS

Prof. Dr. Johannes Kopp
Universität Trier
Deutschland

JunDoz. Dr. Daniel Lois
Universität Tübingen
Deutschland

ISBN 978-3-658-02299-0
DOI 10.1007/978-3-658-02300-3

ISBN 978-3-658-02300-3 (eBook)

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Springer VS

© Springer Fachmedien Wiesbaden 2012, 2014

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsgesetz zugelassen ist, bedarf der vorherigen Zustimmung des Verlags. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

Lektorat: Dr. Cori A. Mackrodt, Stefanie Loyal

Gedruckt auf säurefreiem und chlorfrei gebleichtem Papier

Springer VS ist eine Marke von Springer DE. Springer DE ist Teil der Fachverlagsgruppe Springer Science+Business Media
www.springer-vs.de

Vorwort zur zweiten Auflage

Es freut uns, dass wir nach kurzer Zeit die Möglichkeit erhalten, eine zweite Auflage unserer Einführung in die sozialwissenschaftliche Datenanalyse vorzustellen. Diese zweite Auflage bietet uns die Chance, einige wenige, aber eventuell verwirrende Fehler aus der ersten Auflage zu korrigieren. Wir schulden an dieser Stelle vielen Kolleginnen und Kollegen, vor allem aber unseren Studierenden in Tübingen, Chemnitz und Trier großen Dank, die uns immer wieder Hinweise auf Ungenauigkeiten und Missverständnisse gegeben haben. Besonders danken möchten wir Wolfgang Sodeur, der kritisch den gesamten Text durchgearbeitet hat und uns auf eine Vielzahl von Unklarheiten und Verbesserungsmöglichkeiten hingewiesen hat, die wir so gut wie möglich eingearbeitet haben. Zugleich haben wir die Chance einer Neuauflage genutzt, um einige neuere Entwicklungen – wie etwa im Kap. 2 die Möglichkeiten der Sekundärdatenanalyse mit Hilfe des deutschen Bildungspanel (NEPS) oder die Hinweise zu Metanalysen im Kap. 8 – zu ergänzen.

Trier und Tübingen
im Januar 2014

Johannes Kopp
Daniel Lois

Inhaltsverzeichnis

1	Einleitung	1
	Literatur	12
2	Zur Aufgabe der empirischen Sozialforschung und der quantitativen Datenanalyse im soziologischen Forschungsprozess	15
	Literatur	29
3	Zur Datenlage: eigene Datenerhebungen oder Sekundäranalysen?	31
	Literatur	42
4	Erste Schritte der Datenanalyse: Aufbereitung, Datenbeschreibung und bivariate Zusammenhänge	45
4.1	„A long and winding road“ – zu den Mühen der Datenaufbereitung	46
4.2	Zur Beschreibung von Daten: Verteilungen und Tabellen	51
4.3	Zur Beschreibung von Daten: Graphiken	53
4.4	Maßzahlen: alle für einen, einer für alle?	56
4.5	Maßzahlen der Streuung: „Gleich und Gleich gesellt sich gern“?	58
4.6	Form und Funktion: Zur Gestalt einer Verteilung	61
4.7	Zusammenhangsmaße	63
4.8	Nachbemerkung	76
	Literatur	77

5	Faktorenanalyse und Skalierung	79
5.1	Begriffsdefinitionen: Was ist unter einer Skala zu verstehen und worin besteht der Unterschied zu einem Index?	80
5.2	Das Grundprinzip der explorativen Faktorenanalyse	84
5.3	Anwendungsempfehlungen zur Faktorenanalyse	94
5.4	Wie wird die Reliabilität einer Likert-Skala beurteilt?	96
5.5	Anwendungsempfehlungen zur Reliabilitätsanalyse	101
5.6	Skalenbildung	102
5.7	Wie wird die Validität einer Likert-Skala beurteilt?	103
5.8	Nachbemerkung	104
	Literatur	105
6	Lineare Regressionen	107
6.1	Grundlogik der bivariaten Regression	108
6.2	Bivariate Regression: ein Beispiel aus der Praxis	113
6.3	“You’ll never walk alone” – multivariate Regression	115
6.4	Dummy-Variablen	119
6.5	Gleiche Ergebnisse – verschiedene Darstellungen	127
6.6	Nachbemerkung: eine kleine to-do-Liste	129
	Literatur	130
7	Zur Logik der Datenanalyse: Welche Auswertungsstrategie passt am besten zu meiner Fragestellung?	133
7.1	Das Brutto-Netto-Modell	134
7.2	Die Mediationsanalyse	139
7.3	Die Moderationsanalyse	150
7.4	Nachbemerkung	157
	Literatur	158
8	Logistische Regressionen	161
8.1	Zwei Grundkonzepte der logistischen Regression: Chance und Wahrscheinlichkeit	164
8.2	Wie interpretiert man den Output der logistischen Regression?	166
8.3	Wahrscheinlichkeiten, Chancen, logarithmierte Chancen: Richtlinien für die Ergebnisinterpretation	170
8.4	Ein Beispiel: Welche Merkmale beeinflussen die Wahrscheinlichkeit, konfessionslos zu sein?	176
8.5	Fallstricke der logistischen Regression	181
8.6	Nachbemerkung: weiterführende Literaturhinweise	185
	Literatur	185

9	Ein Ausblick auf fortgeschrittene statistische Analyseverfahren	187
9.1	Ereignisdatenanalyse	188
9.2	Hierarchisch strukturierte Daten: Mehrebenenanalysen	194
9.3	Kausalanalysen mit Paneldaten	199
9.4	Kovarianzbasierte Pfad- und Strukturgleichungsmodelle	204
9.5	Metaanalysen	208
9.6	Nachbemerkung	212
	Literatur	212
	Sachverzeichnis	215

Abbildungsverzeichnis

Abb. 2.1	Anteil der Arbeiter und Angestellten an der Kickerleserschaft (in Prozent)	21
Abb. 2.2	Anteil empirischer Studien in der deutschsprachigen soziologischen Fachliteratur (in Prozent)	25
Abb. 3.1	Anteil der Arbeiten auf Grund von Sekundärdaten an allen empirischen Studien in der deutschsprachigen soziologischen Fachliteratur (in Prozent)	33
Abb. 4.1	Beispiel zur Datenaufbereitung	47
Abb. 4.2	Beispiel zur Datenaufbereitung	49
Abb. 4.3	Balkendiagramm (barchart) der Konfessionszugehörigkeit	54
Abb. 4.4	Histogramme des Religiositätsindex mit unterschiedlicher Klassenbreite	55
Abb. 4.5	Boxplots der Religiosität in Ost- und Westdeutschland	61
Abb. 4.6	Scatterplot (Streudiagramm) Religiosität und Kirchengangshäufigkeit	75
Abb. 5.1	Das Grundmodell der explorativen Faktorenanalyse	84
Abb. 5.2	Extraktion eines Faktorvektors aus zwei Variablenvektoren	87
Abb. 5.3	Faktorextraktion in einem 7-Variablen-Beispiel	89
Abb. 5.4	Rotation der Faktorenlösung	90
Abb. 6.1	Zusammenhang zwischen Religiosität und Kirchengangshäufigkeit	109
Abb. 6.2	Überprüfung der Homoskedastizitätsannahme	125
Abb. 6.3	Boxplot der standardisierten Residuen in Abhängigkeit der Vorhersagewerte	126
Abb. 7.1	Beispiele für Mediation	140
Abb. 7.2	Beispiele für Suppression	141
Abb. 7.3	Partielle Mediation des Einflusses einer ausländischen Herkunft auf die Religiosität durch das Alter	145

Abb. 7.4	Suppression des Einflusses einer ausländischen Herkunft auf die Religiosität durch die subjektive SchichtEinstufung	146
Abb. 7.5	Schematische Darstellung einer Moderation	151
Abb. 7.6	Streudiagramm zum geschlechtsspezifischen Zusammenhang zwischen Bildungsniveau und Religiosität	153
Abb. 8.1	Streudiagramm für eine dichotome abhängige Variable y und eine metrische unabhängige Variable x inklusive Regressionsgerade	162
Abb. 8.2	Lineares Wahrscheinlichkeitsmodell und logistisches Regressionsmodell im Vergleich	163
Abb. 8.3	Auswirkungen positiver und negativer b-Koeffizienten auf den Verlauf der Logit-Funktion	171
Abb. 9.1	Übergangsraten zur ersten Ehe bei wenig und starkreligiösen Frauen	190
Abb. 9.2	Überlebensfunktionen zum Übergang in die Ehe bei wenig und stark religiösen Frauen	192
Abb. 9.3	Beispiele für hierarchisch strukturierte Daten	195
Abb. 9.4	Streudiagramm zum Zusammenhang zwischen Religiosität der Eltern und Religiosität des Befragten	196
Abb. 9.5	Streudiagramme und Regressionsgerade zum Zusammenhang zwischen derReligiosität der Eltern und individueller Religiosität in zwei Ländern	197
Abb. 9.6	Veränderung der Kirchgangshäufigkeit von fünf fiktiven Personen über vier Panelwellen	200
Abb. 9.7	Kovarianzbasiertes Pfadmodell zu vermittelten Einflüssendes Bildungsniveaus der Frau auf ihre Religiosität	205
Abb. 9.8	Das Prinzip statistischer Modellierung	206

Tabellenverzeichnis

Tab. 2.1	Kompetenzen, Zertifikate oder Vorurteile: zur Erklärung des unterschiedlichen Einflusses von in- und ausländischen Bildungsabschlüssen	20
Tab. 2.2	Sozialstrukturelle Bestimmungsgründe der regelmäßigen Kicker-Lektüre	23
Tab. 4.1	Konfessionszugehörigkeit in Deutschland (Spaltenprozent)	52
Tab. 4.2	Religiosität in Deutschland (Spaltenprozent)	53
Tab. 4.3	Kirchgangshäufigkeit in Deutschland (Spaltenprozent sowie kumulierte Spaltenprozent)	58
Tab. 4.4	Konfessionszugehörigkeit in Ost- und Westdeutschland (Spaltenprozent)	59
Tab. 4.5	Messniveau und Darstellung bivariater Zusammenhänge	64
Tab. 4.6	Konfessionszugehörigkeit in Ost- und Westdeutschland (Spaltenprozent bei den Gesamtangaben, Zeilenprozent hinsichtlich der Ost-West-Verteilung)	65
Tab. 4.7	Indifferenztabelle der Konfessionszugehörigkeit in Ost- und Westdeutschland (absolute erwartete Werte bzw. Fälle)	66
Tab. 4.8	Konfessionszugehörigkeit Deutschland (Absolutzahlen und Spaltenprozent)	68
Tab. 4.9	Konfessionszugehörigkeit in West- und Ostdeutschland (Anzahl beziehungsweise Spaltenprozent)	69
Tab. 4.10	Zusammenhang zwischen Einstellung zu Gott und Naturgesetzen (Spaltenprozent)	70
Tab. 4.11	Messniveau und Zusammenhangsmaße	77
Tab. 5.1	Beispiel für einen Index aus zwei Merkmalsdimensionen	81
Tab. 5.2	Zwölf Fragen zur religiösen Weltanschauung aus dem ALLBUS 2002	83
Tab. 5.3	Korrelationsmatrix mit spiegelbildlichen Vektorwinkeln	88

Tab. 5.4	Unrotierte Komponentenmatrix der Hauptkomponentenanalyse	91
Tab. 5.5	Komponentenmatrix der Hauptkomponentenanalyse nach der Varimax-Rotation	92
Tab. 5.6	Kommunalitäten in der Hauptkomponentenanalyse	93
Tab. 5.7	Eigenwerte in der Hauptkomponentenanalyse	94
Tab. 5.8	Reliabilitätsanalyse für die Skala „religiöse Sinngebung“	98
Tab. 5.9	Reliabilitätsanalyse für die Skala „naturalistisches Weltbild“	100
Tab. 5.10	Revidierte Version der Skala „naturalistisches Weltbild“	100
Tab. 5.11	Deskriptive Statistiken	102
Tab. 5.12	Zusammenhänge zwischen den zu testenden Messinstrumenten und zwei Außenkriterien (β -Koeffizienten linearer Regressionen)	104
Tab. 6.1	Religiosität und Kirchgangshäufigkeit	110
Tab. 6.2	Fehler 1. Art bei der Vorhersage der Kirchgangshäufigkeit	110
Tab. 6.3	Fehler 2. Art bei der Vorhersage der Kirchgangshäufigkeit	112
Tab. 6.4	Modellgüte für den Zusammenhang zwischen Religiosität und Kirchgangshäufigkeit (Anova-Block)	114
Tab. 6.5	Bestimmung der Regressionsgeraden	114
Tab. 6.6	Bestimmung der Regressionsebene	116
Tab. 6.7	Bestimmung der Koeffizienten	116
Tab. 6.8	Religiosität, Weltbild und Kirchgangshäufigkeit	118
Tab. 6.9	Prognostizierte Werte und Residuen	119
Tab. 6.10	Determinanten der Kirchgangshäufigkeit	121
Tab. 6.11	Determinanten der Kirchgangshäufigkeit (β -Koeffizienten)	122
Tab. 6.12	Determinanten der Kirchgangshäufigkeit	129
Tab. 7.1	Brutto-Netto-Modell zur Überprüfung des Zusammenhangs zwischen soziodemografischen Merkmalen und Religiosität (lineare Regression, b-Koeffizienten)	136
Tab. 7.2	Deskriptive Statistiken zu sozialstrukturellen Unterschieden zwischen Frauen mit und ohne ausländische Herkunft	142
Tab. 7.3	Hierarchisches lineares Regressionsmodell zur Erklärung des Effektes einer ausländischen Herkunft auf die Religiosität der Frau (b-Koeffizienten)	144
Tab. 7.4	Hierarchisches lineares Regressionsmodell zur Erklärung des Effektes einer ausländischen Herkunft auf die Religiosität der Frau (Fortsetzung)	147

Tab. 7.5	Einfluss der Bildungsjahre auf die Religiosität und der ausländischen Herkunft auf die Bildung (lineare Regression) ...	149
Tab. 7.6	Brutto-Effekte verschiedener soziodemografischer Merkmale auf die Religiosität in Abhängigkeit vom Geschlecht (lineare Regression, b-Koeffizienten)	152
Tab. 7.7	Moderation von Effekten soziodemografischer Variabler auf die Religiosität durch das Geschlecht (lineare Regression, b-Koeffizienten)	155
Tab. 8.1	Zusammenhang von Konfessionslosigkeit und Wohnort (beobachtete Häufigkeiten mit Zeilenprozentwerten in Klammern)	164
Tab. 8.2	Wahrscheinlichkeiten, Odds und Logits im Vergleich	167
Tab. 8.3	Output einer bivariaten logistischen Regression mit der abhängigen Variablen Konfessionslosigkeit (0 = nein, 1 = ja)	168
Tab. 8.4	Auswirkungen positiver und negativer Regressionskoeffizienten auf die Wahrscheinlichkeit für $y=1$...	170
Tab. 8.5	Beispiel zur Berechnung des Log-Likelihood-Funktionswertes	173
Tab. 8.6	Logistisches Regressionsmodell zur abhängigen Variablen Konfessionslosigkeit (0 = nein, 1 = ja; Logit-Koeffizienten inklusive Standardfehlern, z-Werten und Odds-Ratios)	177
Tab. 8.7	Logistische Regression zu Determinanten der Konfessionslosigkeit (0 = nein, 1 = ja; Odds-Ratios)	180
Tab. 8.8	Logistische Regression zu Determinanten der Konfessionslosigkeit (b-Koeffizienten mit z-Werten)	181
Tab. 8.9	Output einer logistischen Regression mit der abhängigen Variablen Konfessionslosigkeit (0 = nein, 1 = ja) und Interaktionseffekt	183
Tab. 8.10	Output des STATA-Befehls „inteff“ zum Interaktionseffekt „Bildungsjahre \times Frau“ aus Tab. 8.9	184
Tab. 9.1	Beispielfragestellungen für Ereignisdatenanalysen	188
Tab. 9.2	Veränderungen der jährlichen Kirchgangshäufigkeit im Zuge verschiedener biografischer Übergänge. (Fixed-Effects-Regressionenmodelle, b-Koeffizienten, t-Werte in Klammern)	203