

Volker Müller-Benedict

Grundkurs Statistik
in den Sozialwissenschaften

Volker Müller-Benedict

Grundkurs Statistik in den Sozialwissenschaften

*Eine leicht verständliche,
anwendungsorientierte Einführung
in das sozialwissenschaftlich
notwendige statistische Wissen*

2., überarbeitete Auflage

Westdeutscher Verlag

Bibliografische Information Der Deutschen Bibliothek
Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen
Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über
<<http://dnb.ddb.de>> abrufbar.

1. Auflage März 2001
2., überarbeitete Auflage April 2003

Alle Rechte vorbehalten
© Westdeutscher Verlag GmbH, Wiesbaden 2003

Lektorat: Frank Engelhardt

Der Westdeutsche Verlag ist ein Unternehmen der
Fachverlagsgruppe BertelsmannSpringer.
www.westdeutscher-verlag.de



Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung des Verlags unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, 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.

Umschlaggestaltung: Horst Dieter Bürkle, Darmstadt

Gedruckt auf säurefreiem und chlorfrei gebleichtem Papier

ISBN 978-3-531-33635-0

ISBN 978-3-663-05700-0 (eBook)

DOI 10.1007/978-3-663-05700-0

Inhaltsverzeichnis

| | |
|---|-----|
| Kapitel 1: Was ist Statistik? | 17 |
| 1.1 Die Wichtigkeit von Statistik..... | 17 |
| 1.2 Was ist Statistik nicht?..... | 18 |
| 1.3 Welche Funktion hat Statistik?..... | 19 |
| 1.4 Einteilung der Statistik..... | 21 |
| Kapitel 2: Sozialwissenschaftliche Datensammlungen | 26 |
| Kapitel 3: Skalenniveau | 33 |
| 3.1 Daten, Merkmale, Variable..... | 33 |
| 3.2 Skalenniveaus..... | 35 |
| Kapitel 4: Häufigkeiten | 41 |
| 4.1 Urlisten und Rohdaten..... | 41 |
| 4.2 Häufigkeitsverteilungen..... | 42 |
| 4.3 Kumulierte Häufigkeiten..... | 47 |
| 4.4 Grafische Darstellung der Häufigkeitsverteilung eines Merkmals..... | 49 |
| 4.5 Die Häufigkeitsverteilung zweier Merkmale..... | 56 |
| 4.6 Grafische Darstellung von bivariaten Verteilungen..... | 59 |
| Kapitel 5: Lageparameter | 64 |
| 5.1 Der Modus..... | 65 |
| 5.2 Der Median..... | 67 |
| 5.3 Der Mittelwert..... | 71 |
| 5.4 Verteilungsformen..... | 78 |
| 5.5 Transformationen..... | 88 |
| 5.6 Wachstumsraten..... | 90 |
| Kapitel 6: Streuungen | 94 |
| 6.1 Die Spannweite oder Variationsbreite..... | 94 |
| 6.2 Zentile, Quartile, Quartilsabstand..... | 95 |
| 6.3 Varianz, Standardabweichung, Variationskoeffizient..... | 102 |
| 6.4 Momente..... | 108 |
| 6.5 Ein Streuungswert für bivariable Verteilungen: die Kovarianz..... | 109 |
| 6.6 Ein Maß für die Heterogenität nominalskaliert er Variable: Simpsons D..... | 113 |
| Kapitel 7: Die Normalverteilung und andere theoretische Verteilungen | 116 |
| 7.1 Stichprobenverteilungen..... | 119 |

| | |
|--|------------|
| 7.2 Die Normalverteilung | 124 |
| 7.3 Flächen unter der Normalverteilung | 129 |
| 7.4 Die χ^2 -Verteilung..... | 136 |
| 7.5 Die Binomialverteilung | 140 |
| 7.6 Die Übergänge in die Normalverteilung von χ^2 und B(n,p) | 144 |
| Kapitel 8: Wahrscheinlichkeiten | 150 |
| 8.1 Definition der Wahrscheinlichkeit..... | 151 |
| 8.2 Eigenschaften der Wahrscheinlichkeit..... | 155 |
| 8.3 Entsprechungen zwischen empirischen Verteilungen und Wahrscheinlichkeitsverteilungen..... | 159 |
| Kapitel 9: Konfidenzintervalle | 164 |
| 9.1 Umformulierung empirischer Daten als Ergebnisse von Zufallsexperimenten | 164 |
| 9.2 Konfidenzintervalle | 167 |
| 9.3 Konfidenzintervalle und Test auf Null | 175 |
| 9.4 Die Berechnung von Konfidenzintervallen für Anteilswerte | 177 |
| 9.5 Berechnung der nötigen Sample-Größe..... | 179 |
| 9.6 Exkurs: Schätzung des Konfidenzintervall für eine Varianz | 181 |
| Kapitel 10: Nominalskalierte Zusammenhangsmaße..... | 185 |
| 10.1 Die Berechnung von Zusammenhangsmaßen..... | 185 |
| 10.2 Die Prozentsatzdifferenz..... | 189 |
| 10.3 Das Maß χ^2_{emp} | 192 |
| 10.4 Auf χ^2_{emp} – basierende Maße | 198 |
| 10.5 Maße mit PRE-Interpretation..... | 201 |
| Kapitel 11: Ordinalskalierte Zusammenhangsmaße..... | 210 |
| 11.1 Wie sollten Maße für ordinalskalierte Merkmale aussehen? | 210 |
| 11.2 Relationen von Paaren und ihre Summen | 214 |
| 11.3 Die Berechnung der Maße | 219 |
| 11.4 Konfidenzintervalle und Signifikanz | 222 |
| 11.5 Vergleich und Kritik der Maße..... | 227 |
| 11.6 PRE-Interpretation der Maße..... | 232 |
| 11.7 Monotonie und Nichtlinearität..... | 235 |
| Kapitel 12: Korrelation und Regression | 241 |
| 12.1 Die Beispiel-Daten..... | 241 |

| | |
|---|------------|
| 12.2 Die Korrelation r | 245 |
| 12.3 Die Regressionsgerade..... | 252 |
| 12.4 Der Determinationskoeffizient | 256 |
| 12.5 Die Beziehung zwischen Korrelation r , Regressionskoeffizient b und R^2 | 262 |
| 12.6 Rang-Korrelation | 263 |
| 12.7 Übersicht über alle Zusammenhangsmaße | 267 |
| 12.8 Korrelation ist nicht Kausalität | 271 |
| Literatur | 274 |
| Lösungen zu den Übungsaufgaben..... | 275 |
| Glossar | 281 |

Abbildungsverzeichnis

| | |
|--|----|
| Abbildung 4.1: Balkendiagramm „Konfessionszugehörigkeit“ | 51 |
| Abbildung 4.2: Histogramm der Altersverteilung in 10-Jahres-Messwertklassen | 52 |
| Abbildung 4.3: Altersverteilung in 10-Jahres-Klassen als Polygonzug | 52 |
| Abbildung 4.4: Altersverteilung absolut..... | 53 |
| Abbildung 4.5: Altersverteilung mit nur 3 Messklassen..... | 54 |
| Abbildung 4.6: „Dichtefunktion“ der Altersverteilung und relative Häufigkeit einer Messwertklasse | 54 |
| Abbildung 4.7: absolute kumulierte Verteilungsfunktion des Alters | 55 |
| Abbildung 4.8: Gruppiertes Balkendiagramm | 60 |
| Abbildung 4.9: Drei – dimensionales Säulendiagramm | 60 |
| Abbildung 4.10: Punkte- bzw. Streudiagramm einer bivariaten Verteilung: Fernsehanschlusssdichte und Kinobesucher pro Jahr 1962 - 1986 | 61 |
| Abbildung 5.1: kumulierte Verteilung der Schulabschlüsse der unter 50-Jährigen | 69 |
| Abbildung 5.2: Verschiedene Daten mit Mittelwert 5.4 | 78 |
| Abbildung 5.3: Verteilungsformen | 79 |
| Abbildung 5.4: Lehrerinnen-Anteil an Gymnasien in Rheinland-Pfalz (bis zur Aufhebung der getrennt-geschlechtlichen Schulen (ca. 1970) bimodal) | 80 |
| Abbildung 5.5: Residuen (Mittelwert = 0-Linie)..... | 81 |
| Abbildung 5.6: Kinderzahl | 82 |
| Abbildung 5.7: Wichtigkeit des Ziels, „das Leben zu genießen“, für Jugendliche | 83 |
| Abbildung 5.8: U-förmige Verteilung des Umweltschutzverhaltens | 84 |
| Abbildung 5.9: Einkommensverteilung in einer Firma als Piktogramm | 87 |
| Abbildung 5.10: Dauer der Arbeitslosigkeit, absolut und logarithmiert. | 89 |
| Abbildung 6.1: kumulierte Altersverteilung mit C_{10} und C_{90} | 95 |

| | |
|---|------|
| Abbildung 6.2: Histogramm der Altersverteilung (1 „Stäbchen“ = Häufigkeit eines Jahrgangs)..... | 96 |
| Abbildung 6.3: „Boxplot“ mit x_{\min} , Q_1 , Z , Q_3 , x_{\max} | 97 |
| Abbildung 6.4: SPSS - Boxplot der Einkommensverteilung für Männer und Frauen Erläuterung mit Berufsangabe für Ausreißer- und Extremwerte (Stichprobe von 8% aus ALLBUS 94). | 101 |
| Abbildung 6.5: Nettoeinkommen von Männern und Frauen | 102 |
| Abbildung 6.6: Einkommensverteilung Männer mit Mittelwert und Standardabweichung..... | 106 |
| Abbildung 6.7: Streudiagramm mit Quadranten..... | 110 |
| Abbildung 6.8: Streudiagramm Einkommen - Berufsprestige.... | 113 |
| Abbildung 7.1: Quetelets Ermittlung der Größe von Rekruten . | 118 |
| Abbildung 7.2: Verteilung von Mittelwerten von 90 Stichproben, für Jungen (StdAbw = 0,18) und Mädchen (StdAbw = 0,21) | 121 |
| Abbildung 7.3: Häufigkeitsverteilung und theoretische Verteilung eines Parameters | 123 |
| Abbildung 7.4: Flächenstücke unter einer theoretischen Verteilung..... | 124 |
| Abbildung 7.5: Grafik und Funktion der Normalverteilung auf dem 10-Mark-Schein..... | 125 |
| Abbildung 7.6: Verschiedene Gestalten der Normalverteilung . | 127 |
| Abbildung 7.7: Normalverteilung mit Flächenangaben der $\pm 1, 2$ oder 3σ - Bereiche | 129 |
| Abbildung 7.9: Verschiedene Gestalten der χ^2 -Verteilung | 137 |
| Abbildung 7.10: Verschiedene Gestalten der Binomialverteilung | 0141 |
| Abbildung 7.11: Approximation der Binomialverteilung durch die Normalverteilung | 147 |
| Abbildung 8.1: Anteil von „Zahl“ beim Münzwurf | 152 |
| Abbildung 8.2: Zerlegung einer Menge, $E =$ Ereignisraum, $A_i =$ Elementar-Ereignisse | 153 |
| Abbildung 8.3: Das Galton-Brett | 156 |
| Abbildung 9.1: Wahrscheinlichkeitsverteilung bei 2- und 3-maligem Würfeln | 165 |

| | |
|--|-----|
| Abbildung 9.2: Verschiedene mögliche Konfidenzintervalle um einen wahren Parameter | 169 |
| Abbildung 9.3: Links-Rechts-Selbsteinschätzungsskala aus dem ALLBUS..... | 171 |
| Abbildung 9.4: Test und Konfidenzintervall: nur unten sind δ und μ signifikant unterschiedlich | 175 |
| Abbildung 11.1: Mögliche Lagen von Paaren in Bezug auf ordinale Ordnungsrelationen | 217 |
| Abbildung 12.1: Streudiagramm Ausbildungsdauer vs. durchschnittliche subjektive Schichteinstufung..... | 245 |
| Abbildung 12.2: Streudiagramm der standardisierten Variablen mit dem Rechteck zum Fall i | 247 |
| Abbildung 12.3: Punktwolken im Streudiagramm für verschiedene Werte von r | 250 |
| Abbildung 12.4a: Mittelwert als Referenz für die Residuen | 253 |
| Abbildung 12.4b: Regressionsgerade als Referenz für die Residuen | 254 |
| Abbildung 12.5: Residuenzerlegung | 257 |
| Abbildung 12.6: Streudiagramm und Regressionsgerade von Ausbildungsjahren und subjektiver Schichteinstufung..... | 261 |

Tabellenverzeichnis

| | |
|---|-----|
| Tabelle 2.1: Einige sozialwissenschaftliche Trenddatenbestände | 28 |
| Tabelle 2.2: Standard-Demografie (ALLBUS)..... | 31 |
| Tabelle 2.3: Ausschnitt aus dem Eurobarometer | 32 |
| Tabelle 3.1: Begriffe zur Beschreibung von Forschungsobjekten | 33 |
| Tabelle 3.2: Beispiel einer Rohdaten-Matrix | 34 |
| Tabelle 3.3: Aufbau einer Matrix | 34 |
| Tabelle 3.4: Datenmatrix des ALLBUS 1994 in SPSS (Ausschnitt) | 35 |
| Tabelle 3.5: Die unterschiedlichen Skalentypen..... | 38 |
| Tabelle 4.1: Beispiele für Rohdaten, ihre Fälle, Merkmale und Ausprägungen..... | 41 |
| Tabelle 4.2: Häufigkeitstabelle in SPSS | 49 |
| Tabelle 4.3: Weiter unterteilte Kreuztabelle | 61 |
| Tabelle 5.1: Religionszugehörigkeit in ländlichen Gebieten und Großstädten..... | 65 |
| Tabelle 5.2: Altersverteilung in zwei verschiedenen Messklassen | 66 |
| Tabelle 5.3: Allgemeiner Schulabschluss mit Alter ≥ 50 | 67 |
| Tabelle 5.4 Allgemeiner Schulabschluss mit Alter < 50 | 68 |
| Tabelle 5.5: Mittelwerte auf einer 7er-Skala | 72 |
| Tabelle 5.6: SPSS-Ausgabe von Lageparametern | 73 |
| Tabelle 5.7: Beschreibende Merkmale von Häufigkeitsverteilungen..... | 85 |
| Tabelle 5.8: Beziehungen zwischen Lageparametern und Verteilungsform..... | 86 |
| Tabelle 5.9 : Abiturientenzahlen in der Bundesrepublik (mit 1980 = 100) | 91 |
| Tabelle 6.1: Schulabschluss Männer und Frauen (ALLBUS 1994) | 98 |
| Tabelle 6.2: Quartile und Median der Altersverteilung | 100 |
| Tabelle 6.3: Lage- und Streuungsparameter der Einkommensverteilung (ohne Einkommen = 0)..... | 105 |
| Tabelle 6.4: Ethnische Verteilung in den USA..... | 115 |
| Tabelle 7.1: „Wichtigkeit von Rücksicht nehmen“ für Mädchen und Jungen..... | 119 |

| | |
|--|-----|
| Tabelle 7.2: Ausschnitt aus einer Tabellierung der Standardnormalverteilung | 131 |
| Tabelle 7.3: Vertrauen in das Gesundheitswesen | 135 |
| Tabelle 7.4: Ausschnitt aus der Tabelle der χ^2 -Verteilung | 139 |
| Tabelle 7.5: Ausschnitt aus der Tabelle der Binomialverteilung | 142 |
| Tabelle 7.6: Theoretische und Stichproben-Verteilung von 3 wichtigen Parametern | 144 |
| Tabelle 8.1: Gegenüberstellung empirischer und theoretischer Begriffe..... | 162 |
| Tabelle 9.1: Links-Rechts-Selbsteinstufung 1994 | 171 |
| Tabelle 9.2: Links-Recht-Selbsteinstufung 1996..... | 176 |
| Tabelle 10.1: Parameter und Fragestellung der zugehörigen statistischen Analyse..... | 186 |
| Fernsehgebrauch und Geschlecht | 188 |
| Tabelle 10.2: Fünf Verfahren zur Feststellung von statistischem Zusammenhang..... | 189 |
| Tabelle 10.3: Notation einer Vierfeldertafel..... | 190 |
| Tabelle 10.4: Fernsehgebrauch und Geschlecht in Prozent | 191 |
| Tabelle 10.5: Eine Kontingenztafel (Kreuztafel), sowohl mit f = absolute, als auch mit p = relative Häufigkeiten möglich... | 193 |
| Tabelle 10.6: Eigenschaften von Zusammenhangsmaßen | 198 |
| Tabelle 10.7: SPSS-Ausgabe nominalskalierter Zusammenhangsmaße | 206 |
| Tabelle 10.8: Übersicht über die Zusammenhangs-Maße von nominalskalierten Variablen..... | 208 |
| Tabelle 11.1: Mögliche Relationen von Paaren von Untersuchungsfällen | 216 |
| Tabelle 11.2a: Zusammenhangsmaße für die Beziehung zwischen Schulabschluss und politischem Interesse (mit SPSS) | 225 |
| Tabelle 11.2b: wie oben, nur nach Dichotomisierung | 231 |
| Tabelle 11.3: Zusammenhang zwischen Alter und Zustimmung zum „Hausfrauen-Dasein“ bei Frauen | 238 |
| Tabelle 11.4: Zusammenhang zwischen Alter und Zustimmung zum „Hausfrauen-Dasein“ bei Frauen mit mindestens Fachhochschulreife..... | 239 |

| | |
|---|-----|
| Tabelle 12.1: Subjektive Schichteinstufung und Anzahl der schulischen Ausbildungsjahre | 242 |
| Tabelle 12.2: Ausbildungsdauer und durchschnittliche subjektive Schichteinstufung und daraus berechnete Werte | 244 |
| Tabelle 12.3: „Zufallshöchstwerte“ von r | 251 |
| Tabelle 12.4: Korrelationsausgabe in SPSS..... | 252 |
| Tabelle 12.5: Berechnung der nicht erklärten SAQ..... | 259 |
| Tabelle 12.6: Rangplatzdifferenzen..... | 264 |
| Tabelle 12.7: Rangreihen von zwei verschiedenen Gutachtern.. | 265 |
| Tabelle 12.8: Übersicht über die Zusammenhangsmaße..... | 268 |

Vorwort

Dieses Buch ist aus mehrjährigen Vorlesungen für Studierende der sozialwissenschaftlichen Fakultät entstanden. Seine Konzeption und Schwerpunkte wollen den Problemen und Anforderungen dieser Gruppe von Studierenden Rechnung tragen. Zu den Problemen und Anforderungen gehören insbesondere:

- wenig mathematische Vorkenntnisse und wenig Motivation, sich auf strenge mathematische Formulierungen und Beweise einzulassen,
- Statistik nicht als Methodenlehre anzusehen, sondern als eine Möglichkeit, sozialwissenschaftliche Daten zu interpretieren und zu verstehen,
- die auch für Sozialwissenschaftler unabdingbare Qualifikation, Daten mit einem EDV-Programm auswerten zu können oder zumindest den von einem solchen Programm produzierten Output, z.B. in wissenschaftlichen Publikationen, lesen zu können,
- die mittlerweile ebenso für Sozialwissenschaftler unabdingbare Qualifikation, Ergebnisse der Umfrage- und Marktforschung substantiell auch statistisch kritisieren zu können.

Aus Punkt 4 ist der Umfang der hier präsentierten Methoden abgeleitet: da in vielen sozialwissenschaftlichen Studiengängen nur Statistik I (und nicht auch noch multivariate Statistik = Statistik II) verlangt wird, sind hier auch Wahrscheinlichkeits- und Schätztheorie insoweit aufgenommen, dass Konfidenzintervalle und damit die Frage der Repräsentativität und Signifikanz von deskriptiven statistischen Werten behandelt werden können. Von Punkt 2. und 3. zusammen stammt die Entscheidung, mit einem ALLBUS-Datensatz als Grundlage der Beispiele und mit dem mittlerweile auch in der Wirtschaft weit verbreiteten Statistik-Programm SPSS als Grundlage für EDV-Outputs zu arbeiten.

Punkt 1 führt dazu, dass in diesem Buch auf Beweise weitgehend verzichtet wird und „höhere“ mathematische Symbole in Exkursen erklärt werden, und dass der Ausdruck umfangreicher statistischer Tabellen unterbleibt. Um Statistik zu verstehen, muss man diese Tabellen einmal kennen gelernt haben, den praktischen Umgang damit nehmen uns heute die EDV-Programme ab.

Ich hoffe, dass mit diesen Entscheidungen die Attraktivität dieses Buches für Studierende der Sozialwissenschaften erhöht werden kann, ohne dass die inhaltliche statistische Substanz zu sehr verloren geht.

Bedanken möchte ich mich vor allem und ganz herzlich bei meiner Kollegin Elisabeth Klaus, mit der zusammen das Projekt der Statistik-Vorlesung begonnen hat und die eine ganze Reihe der Grafiken und Ideen in diesem Buch beigesteuert hat. Weiter möchte ich mich bei meinen studentischen Tutoren Mirco Arnhold, Christian Klees, Sascha Münnich, Carina Siefken, Ravena Pennig und Matthias Glebe bedanken, die den Text korrigiert haben, und ebenso bei den Teilnehmerinnen und Teilnehmern meiner Vorlesung, die mich immer wieder auf neue Fehler aufmerksam machen.