

## Literatur

- Benninghaus, H. (1990): Einführung in die sozialwissenschaftliche Datenanalyse. München: Oldenbourg
- Blalock, H.M. (1979): Social Statistics. N.Y. : McGraw-Hill
- Bortz, J. (1989): Statistik für Sozialwissenschaften. Berlin: Springer
- Clauß, G., Ebner, H. (1992)<sup>7</sup>: Statistik für Soziologen, Pädagogen, Psychologen und Mediziner. Thun: Harri Deutsch
- Clauß, G., Finze, F., Partzsch, L. (1995)<sup>2</sup>: Statistik. Für Soziologen, Pädagogen, Psychologen und Mediziner. Thun: Harri Deutsch
- Diekmann, A. (1995): Empirische Sozialforschung. Reinbek: Rowohlt
- Eckey, H.-F., Kosfeld, R., Dreger, C. (2000)<sup>2</sup>: Statistik. Grundlagen, Methoden. Beispiele. Wiesbaden: Gabler
- Ehrenberg, A. (1986): Statistik oder der Umgang mit Daten. Weinheim: VCH
- Engel, A., Möhring, M., Troitzsch, K. (1995): Sozialwissenschaftliche Datenanalyse. Mannheim: BI Wissenschaft
- Fahrmeir, L. u.a. (1999)<sup>2</sup>: Statistik. Der Weg zur Datenanalyse. Berlin: Springer
- Healey, Joseph F., Babbie, E., Halley, F. (1997): Exploring Social Issues. Using SPSS for Windows. Thousand Oaks
- Hochstädter, D. (1996)<sup>8</sup>: Statistische Methodenlehre. Thun: Harri Deutsch
- Kern, H. (1982): Empirische Sozialforschung. Ursprünge, Ansätze, Entwicklungslinien. München: Beck
- Knapp, T. (1996): Learning Statistics Through Playing Cards. London: Sage
- Krämer, W. (1992)<sup>2</sup>: Statistik verstehen. Eine Gebrauchsanweisung. Frankfurt: Campus
- Kraul, M., Troitzsch, K., Wirrer, R. (1995): Lehrerinnen und Lehrer an Gymnasien. Ms. Halle, April 1995
- Kühnel, S., Krebs, D. (2001): Statistik für Sozialwissenschaften.. Reinbek: rowohlt
- Liebetrau, A. (1983): Measures of association. London: Sage
- Lüdemann, C. (1995): Ökologisches Handeln und Schwellenwerte. Ergebnisse einer Studie zum Recycling-Verhalten. In: Zuma-Nachrichten 37:63-75
- Mohr, L.B. (1990): Understanding Significance Testing. London: Sage
- Roeser, J. (1995): Was Frauen und Männer vor dem Bildschirm erleben. Düsseldorf
- Sachs, L. (1982): Statistische Methoden. Berlin: Springer
- Schlittgen, R. (1996)<sup>6</sup>: Einführung in die Statistik. München: Oldenbourg
- Sensch, J. (1995): Statistische Modelle in der Historischen Sozialforschung I: Allgemeine Grundlagen – Deskriptivstatistik – Auswahlbibliographie. HSR - Historical Social Research Supplement Nr. 7 (1995)
- Sixtl, F. (1993): Der Mythos des Mittelwerts. Neue Methodenlehre der Statistik. München: Oldenbourg
- Thome, H. (1990): Grundkurs Statistik für Historiker Teil II: Induktive Statistik und Regressionsanalyse. HSR - Historical Social Research Supplement Nr. 3 (1990)
- Thome, H. (1989): Grundkurs Statistik für Historiker Teil I: Deskriptive Statistik. HSR - Historical Social Research Supplement Nr. 2 (1989)

## Lösungen zu den Übungsaufgaben

### Kapitel 3

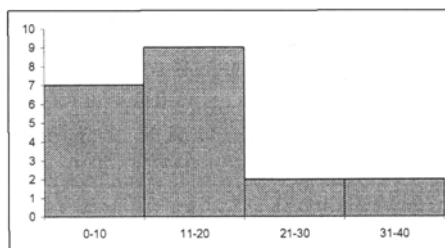
1. a) nominal, b) intervall, c) nominal (kann man nicht mit rechnen!), d) ordinal (es gibt allerdings in der Schulleistungsmessung eine intensive Diskussion über die Skalenqualität der 6 deutschen Stufen; meist werden Punktesysteme zusätzlich eingesetzt, um intervallskalierte Schulleistungen zu erhalten), e) intervall, f) rational, g) nominal, h) rational (aber diskret!).
2. a) Einkommen in DM – metrisch, quantitativ, stetig  
Gering-/Normal-/Besserverdiener - ordinal, qualitativ, diskret
- b) Anzahl Stunden – metrisch quantitativ, stetig  
allein, in der Gruppe, mit Nachhilfe – ordinal, qualitativ, diskret
- c) km/h – metrisch, quantitativ, stetig  
Zügig, normal, mit Pausen - ordinal, qualitativ, diskret
- d) Anzahl Zigaretten / Tag – metrisch, quantitativ, stetig  
kein Raucher, Gelegenheitsraucher, Raucher, Kettenraucher – ordinal, qualitativ, diskret

### Kapitel 4

1. a)

Stundenzahl	Abs. Häuf.	Rel. Häuf.
0	2	0,1
4	2	0,1
6	3	0,15
12	2	0,1
16	3	0,15
20	4	0,2
30	2	0,1
40	2	0,1
SUMME	20	1

b)



2.

Video / Bestanden?	Nein	Ja	SUMME
Nicht gesehen	6	4	10
Einmal gesehen	3	7	10
Zweimal gesehen	2	8	10
SUMME	11	19	30

*Kapitel 5*

1. a) Modus, b) Modus, Median, Mittelwert, c) Modus, Median, d) Modus, Median

2. Wenn die Datensätze vereinigt werden, lässt sich der Mittelwert der neuen Verteilung berechnen, indem der Mittelwert aus den zwei Teilgruppen gemittelt wird:

$$\bar{x} = \frac{5 \cdot 100 + 10 \cdot 100}{200} = 7,5$$

Ein neuer Median oder Modus kann nicht bestimmt werden, da man nicht weiss, wie das Merkmal in den Ausgangsdatsätzen verteilt ist.

*Kapitel 6*

1. Da hier 20 Daten vorliegen, umfasst jedes Quartil genau 5 Fälle. Deshalb ist der Wert des 1. Quartils genau der zwischen dem Wert des 5. und des 6. Fall gemittelte Wert und der Wert des 3. Quartils entsprechend der zwischen dem Wert des 15. und 16. Falls gemittelte Wert. Da Fall 5 und 6 beide den Wert 6 haben, ist auch  $Q_1 = 6$ , ebenso folgt  $Q_3 = 20$ . Dann muss der Quartilsabstand  $Q_A = (20 - 6) / 2 = 7$  sein.  $Q_A = 7$  bedeutet, dass der Fernsehkonsum der im mittleren Bereich liegenden 50% der Befragten um bis zu 14 h unterschiedlich ist.

2. Mittelwert der Ursprungsdaten = 4,0625, Standardabweichung der Ursprungsdaten = 2,1438

x	x standardisiert.
1	$(1 - 4,0625) / 2,1438 = -1,429$
2	$(2 - 4,0625) / 2,1438 = -0,962$
3	$(3 - 4,0625) / 2,1438 = -0,496$

4	$(4 - 4,0625) / 2,1438 = -0,029$
5	$(5 - 4,0625) / 2,1438 = 0,437$
6	$(6 - 4,0625) / 2,1438 = 0,904$
7	$(7 - 4,0625) / 2,1438 = 1,370$

### Kapitel 7

1. Für  $N(0,1)$  liegt der  $z_{\alpha}$ -Wert für  $\alpha = 95\%$  bei 1,645 (Tabelle 7.2), darüber liegen die eifrigsten 5%. Die Rücktransformation auf die in der Aufgabe vorgegeben  $N(10,6)$  lautet  $1,645 \cdot 6 + 10 = 19,87$ .

Die Studentin musste sich also fast 20 Stunden vorbereiten, um zu den eifrigsten 5% der Studierenden zu gehören.

2. a) Für einen Frauenanteil von  $p = 0,2$  ergibt sich nach Tabelle 7.5 der Binomialverteilung für Stichproben der Größe  $n = 20$  ein Wert von  $0,002 = 0,2\%$  für den Anteil der Stichproben mit genau  $k = 10$  zehn Frauen.

b) Ein Frauenanteil von 0,7 entspricht einem Männeranteil  $p = 0,3$ . Nach derselben Tabelle ergibt sich ein Wert von  $0,03 = 3,1\%$  für den Anteil von Stichproben mit genau 10 Männern, also ebenfalls 10 Frauen (bei einer Gesamtzahl von 20) in der Stichprobe.

### Kapitel 8

1.a) Die Wahrscheinlichkeit, bei einem Würfelwurf eine bestimmte Zahl zu erzielen, liegt bei  $1/6$ . Durch die Bedingung eines logischen "Oder" werden die Wahrscheinlichkeiten, beim ersten Wurf eine 5 "oder" eine 6 zu werfen, addiert; die Gesamtwahrscheinlichkeit beträgt daher  $2/6 = 1/3$ .

b) Es existieren insgesamt drei Möglichkeiten, mit zwei Würfeln eine Summe von mehr als oder gleich 11 zu erreichen: durch eine Fünf beim ersten und eine Sechs beim zweiten Wurf, durch die umgekehrte Situation sowie durch zwei Sechsen. Die Gesamtzahl aller möglichen verschiedenen Ergebnisse bei zwei hintereinander folgenden Würfeln ist 36. Da die relative Häufigkeit gleich der Wahrscheinlichkeit ist, ist die gesuchte Wahrscheinlichkeit  $3/36 = 1/12$ .

2. Die z-Transformation des Werts 133 ergibt  $(133 - 100) / 20 = 1,65$ . Bis zu diesem Wert liegen unter der Standardnormalverteilung

lung 95 % der Fläche (S.131), und genauso auch bei der Normalverteilung  $N(100,20)$ . Dementsprechend liegt die Wahrscheinlichkeit auf einen  $IQ \geq 133$  bei  $100 - 95 = 5\%$ .

### Kapitel 9

1. Es müssen zunächst die Standardfehler berechnet werden:

$$\text{Standardfehler West } s_{\text{West}} = 1,467 / \sqrt{2284} = 1,467 / 47,79 = 0,0307.$$

$$\text{Standardfehler Ost } s_{\text{Ost}} = 1,512 / \sqrt{1081} = 1,512 / 32,88 = 0,046.$$

Damit ergeben sich folgende Konfidenzintervalle:

$$\text{West: } [-1,96 \cdot 0,0307 + 3,92; 3,92 + 0,0307 \cdot 1,96] = [3,86 ; 3,98]$$

$$\text{Ost: } [-1,96 \cdot 0,046 + 3,35; 3,35 + 0,046 \cdot 1,96] = [3,26 ; 3,44].$$

Die Mittelwerte der beiden Ländergruppen unterscheiden sich signifikant zum Niveau 5% von einander auch in der Grundgesamtheit, da die Mittelwerte der einen Gruppe jeweils außerhalb des 95%- Konfidenzintervalls des Mittelwertes der anderen Gruppe liegen.

2. Hier muss nach LaPlace approximiert werden:

$$B(2102; 0,048) \approx N(0,048; \sqrt{\frac{0,048(1-0,048)}{2102}}) \approx N(0,048; 0,0047)$$

Für das 95%-Konfidenzintervall gilt dann:

$$[-1,96 \cdot 0,0047 + 0,048; 1,96 \cdot 0,0047 + 0,048] = [0,0388; 0,0572]$$

d.h. mit 95% Sicherheit liegt der Anteil der „pessimistischen“ Jugendlichen zwischen 3,88% und 5,72% (in absoluten Zahlen: zwischen 81 und 120).

### Kapitel 10

1.

$$\phi = \frac{|507 \cdot 334 - 982 \cdot 368|}{\sqrt{(507 + 368)(982 + 334)(507 + 982)(368 + 334)}} = -0,175$$

$$\chi^2_{\text{emp}} = n \cdot \Phi^2 = 2191 \cdot (-0,175)^2 = 67,4$$

Der Zusammenhang ist nicht zufällig, da  $\chi^2_{\text{emp}} = 67,4 > 3,84 = \chi^2_{1,1}$ . Aber er ist schwach, da  $\Phi$  klein ist.

2. Die Behauptung kann durch die vorliegenden Daten nicht gestützt werden, da hier voreilig der Schluss auf Frauen und Männer im Allgemeinen gezogen wird. Zu diesem Schluss müsste ein Zusammenhangsmaß der Vierfeldertafel (Frauen/Männer, viel/wenig) berechnet und auf Signifikanz geprüft werden. Bei der geringen Anzahl Daten und der geringen Differenz von 20% wäre eher keine Signifikanz zu erwarten. Zudem könnte das Maß darüber Auskunft geben, wie stark der Zusammenhang wäre.

### Kapitel 11

1. a) Summe der konkordanten Paare  $N_c = 3(7+5) + 35 = 36 + 15 = 51$ , Summe der diskordanten Paare  $N_d = 1(1+7) + 3 \cdot 1 = 8 + 3 = 11$ ,  $\text{Gamma} = (51-11)/(51+11) = 0,64$ . Gamma weist damit einen stärkeren (mehr als mittelstarken) Zusammenhang auf. Ein Wert von ca.  $2/3$  bedeutet, dass man bei ca.  $2/3$  aller nicht gebundenen Paare aus der Differenz bei dem einen Merkmal auf dieselbe Differenz bei dem anderen Merkmal schließen kann.

b) Nach der Formel gilt für die Varianz der Differenz  $S = N_c - N_d$ :  $s_S^2 = 1/18 \cdot 20 \cdot (20-1) \cdot (2 \cdot 20 - 5) = 738,9$ , und für die Standardabweichung  $s_S = \sqrt{738,9} = 27,18$ . Damit ist der standardisierte Wert von  $S$  (der Testwert)  $(N_c - N_d) / s_S = (51-11)/27,18 = 40/27,18 = 1,47$ . Dieser Wert ist kleiner als 1,96, der kritischen Grenze, innerhalb derer 95% aller zufällig zustande gekommenen Differenzen  $S$  (standardisiert) liegen. Damit kann man auf der Grundlage dieser 20 Daten nicht behaupten, dass in der Grundgesamtheit überhaupt ein Zusammenhang besteht. Der in a) festgestellte starke Zusammenhang ist zufällig. Das gilt um so stärker für ein höheres Sicherheitsniveau, z.B. 99%.

### Kapitel 12

1. Variable qm: Mittelwert: 81,25,  
 Varianz =  $8187,5 / 7 = 1169,643$   
 Variable Miene: Mittelwert: 980,00  
 Varianz =  $1265400 / 7 = 180771,4$

$$\text{Standardabweichung } qm = \sqrt{1169,34} = 34,2$$

$$\text{Standardabweichung } \text{Miete} = \sqrt{180771,4} = 425,17$$

Zur Berechnung der Kovarianz sind die Tabelle 12.2 entsprechenden Berechnungen durchzuführen. Dann ergibt sich

$$\text{Kovarianz zwischen } qm \text{ und } \text{Miete} = 91200 / 7 = 13028,6.$$

Damit ist die

$$\begin{aligned} \text{Korrelation } r \text{ zwischen } qm \text{ und } \text{Miete} &= 13028,6 / (34,2 \cdot 425,17) \\ &= 0,896. \end{aligned}$$

Um die Korrelation zu verallgemeinern, muss die Tabelle für die Signifikanz von  $r$  (Tabelle 12.3) herangezogen werden. Die Freiheitsgrade betragen hier  $n - 2 = 6$ . Die Tabelle sagt aus, dass für 5 Freiheitsgrade durch Zufall für ein Fehler-Niveau von 5%  $r$  schon Werte bis 0,75 annehmen kann. Da die berechnete Korrelation darüber liegt, kann der festgestellte Zusammenhang „je größer die Wohnung, desto höher die Miete“ verallgemeinert werden.

2. Der Regressionskoeffizient  $b$  aus der Regressionsgerade ‘Miete =  $a + b \cdot qm$ ’ gibt an, um wie viel die Miete „durchschnittlich“ pro  $qm$  steigt. Aus der Definition der Regressionsgeraden ergibt sich, dass  $b = \text{cov}_{xy} / s_x^2$ . Aus der Beziehung zwischen  $R^2$  und  $r$  ergibt sich, dass  $r = \text{cov}_{xy} / (s_x \cdot s_y)$ . Also kann man berechnen  $b = r \cdot s_y / s_x$ . In diesem Fall also

$$b = 0,896 \cdot 425,17 / 34,2 = 11,14 \text{ DM.}$$

## Glossar

- abhängige Variable 187
- absolute Häufigkeit 43
- Abstand 70
- Abstandsquadrat 77
- Achsenabschnitt 254
- Additionssatz des Mittelwerts 75
- ALLBUS 27
- Anteilswerte 177
- arithmetisches Mittel 71
- Ausreißer 100
- Balkendiagramm 50
- bedingte Verteilung 58
- Berufsprestige 112
- bimodal 79
- Binomialverteilung 140
- Bivariate Verteilung 57
- Boxplot 97
- $\text{Chi}^2$  – Verteilung 137
- Chi-Quadrat-empirisch 195
- concordant 216
- Cramer's V 201
- deskriptive Statistik 21
- Dichotomisierung 230
- Dichtefunktion 55, 124
- discordant 216
- Drittvariablen 237
- Durchschnittliche Abweichung 102
- Ereignisraum 153
- erklärte Abstandsquadrate 257
- erwartete Häufigkeit 194
- Erwartungswert 161
- Extremwerte 100
- fehlende Werte 49
- Fehlerreduktion PRE 233
- flach 83
- Freiheitsgrad 137
- Freiheitsgrade 197
- Gamma 219
- Gauß 124
- gebunden 216
- geometrisches Mittel 92
- Gerade 254
- Gewichtung 74
- Gleichverteilung 83
- Goodman und Kruskals Gamma 219
- Grundgesamtheit 120
- Häufigkeits-Verteilung 43
- Histogramm 51
- Index 75
- Induktion 20
- Inferenz-Statistik 21
- inverser Zusammenhang 214
- Kategorisierung 231
- Kausalität 271
- Kendalls Tau 220
- Kennwert 64
- Konfidenzintervall 169
- Kontingenzkoeffizient C 200
- Kontingenztafel 193
- Korrelation 249
- Kovarianz 111
- Kreuztabelle 57
- kumulierte Verteilung 47
- Kurtosis 108
- L – förmige Verteilung 84
- Lambda 203



- Laplace, Satz von 146
- lineare Transformation 88
- linearer Zusammenhang
  - 236
- linkssteil 81
- logarithmische
  - Transformation 89
- Median 67
- Messwertklassen 45
- Mittelwert 71
- Modus 65
- monotone Funktion 88
- monotoner Zusammenhang
  - 236
- nichtlinearer
  - Zusammenhang 236
- Nichtlinearität 235
- Normalverteilung 125
- Parameter 64
- Phi – Koeffizient 199
- Polygonzug 51
- positiver Zusammenhang
  - 214
- PRE-Konzept 233
- Prozentsatzdifferenz 190
- prozentuale Häufigkeit 46
- Punktwolke 111
- Quadrant 110
- Quartil 97
- Quartilsabstand 97
- Querschnittsanalyse 28
- Quetelet 118
- Randverteilung 57
- Rangkorrelationskoeffizient
  - 264
- Rangkorrelationsmaße 210
- rechtssteil 81
- Regressionsgerade 255
- Regressionskoeffizient 255
- Relationen von Paaren 216
- relative Häufigkeit 46
- repräsentativ 166
- Residuum 81
- robust 70
- Rohdaten 41
- Sample-Größe 179
- SAQ 103
- Säulendiagramm 59
- Scheinkausalität 271
- Schiefe 108
- schwacher Zusammenhang
  - 198
- Schwankungsintervall 134
- signifikant 175
- Simspons D 114
- Somer's d 221
- Sozialstruktur 63
- Spannweite 94
- Standardabweichung 105
- Standard-Demografie 31
- Standardfehler 126
- standardisierte Fragen 31
- Standardisierung 107
- Standardnormalverteilung
  - 128
- starker Zusammenhang 198
- statistische Unabhängigkeit
  - 188
- statistischer Zusammenhang
  - 188
- Steigung 254
- Stichprobe 120
- Stichprobenverteilung 122
- stochastisch unabhängig
  - 156
- Streuungsdiagramm 61

- subjektive Schichteinstufung 241
- Summe der Abstandsquadrate 103
- symmetrisch 80
- Tau 220
- tau - Goodman und Kruskal 204
- Test 175
- theoretische Verteilung 122
- Transformation 88
- Trendanalyse 28
- t-Verteilung 174
- U-förmige Verteilung 83
- unabhängige Variable 187
- unimodal 79
- univariate Statistik 24
- Ursache 187
- Variabilität 20
- Varianz 103
- Variationskoeffizient 108
- Verallgemeinerung 21
- Verteilungsfunktion 48
- Vierfeldertafel 190
- Wachstumsrate 90
- Wahlprognose 117
- Wahrscheinlichkeit 153
- Wirkung 187
- $z_\alpha$  - Wert 132
- ZA - Zentralarchiv für empirische Sozialforschung 26
- Zelle 57
- Zentil 95
- Zentraler Grenzwertsatz 126, 164
- Zentralwert 67
- Zerlegung 74
- z-Funktion 128
- z-Transformation 128
- Zufallsexperiment 153
- Zufallsvariable 160
- Zusammenhangsmaße 185

Klaus Feldmann

**Soziologie kompakt**

Eine Einführung

2., durchges. Aufl. 2001. 369 S. mit 59 Abb. Br. € 18,90

ISBN 3-531-23188-X

Dieses Buch wendet sich an Studienanfänger und Studierende mit dem Nebenfach Soziologie. Es bietet eine umfassende und leicht verständliche Einführung in die Grundlagentheorien und zentralen Bereiche der Soziologie. Die ausgewählten Daten, Beispiele und empirischen Untersuchungen beziehen sich primär auf den deutschen Sprachraum. Zusätzlich wurden Vergleiche mit anderen europäischen Staaten, den USA und auch außer-europäischen Ländern einbezogen.

Christian Stegbauer

**Reziprozität**

Einführung in soziale Formen der Gegenseitigkeit

2002. 182 S. Br. € 19,90

ISBN 3-531-13851-0

Warum senden sich die Menschen gegenseitig Weihnachtskarten? Weshalb kommt es in Wohngemeinschaften zu Konflikten, falls der Eindruck entsteht, nicht jeder beteiligt sich an den Hausarbeiten ungefähr gleich viel wie man selbst? Solche Themen behandelt das einführende Buch. Es wird argumentiert, dass die Ursachen von Gegenseitigkeit nicht mit individuellen Zweck-Mittel-Kalkülen erklärbar sind. Die Art und Weise, wie Gegenseitigkeit ausgestaltet wird, ist immer abhängig von der Beziehung, in der die Austauschpartner zueinander stehen.

Heiner Meulemann

**Soziologie von Anfang an**

Eine Einführung in Themen, Ergebnisse und Literatur

2001. 428 S. mit 27 Abb. und 22 Tab. Br. € 34,50

ISBN 3-531-13742-5

Diese Einführung will „Soziologie von Anfang an“ darstellen. Die Soziologie befasst sich mit der Gesellschaft als einem Produkt sozialen Handelns. Sie beginnt mit dem Begriff des sozialen Handelns und zielt auf die Gesellschaft. Sie versucht, auf dem Weg vom sozialen Handeln zur Gesellschaft die Themen oder Grundbegriffe der Soziologie systematisch darzustellen: soziale Ordnung, soziale Differenzierung, soziale Integration, Sozialstruktur, soziale Ungleichheit, soziale Mobilität und sozialer Wandel.

**www.westdeutscherverlag.de**

Erhältlich im Buchhandel oder beim Verlag,  
Änderungen vorbehalten. Stand: Oktober 2002.

Abraham-Lincoln-Str.46  
65189 WiesbadenTel. 06 11. 78 78 - 285  
Fax. 06 11. 78 78 - 400

  
West-  
deutscher  
Verlag