

**Statistik für Studierende der
Politikwissenschaften oder der Soziologie
im Nebenfach**

Vorlesungsbegleitende Materialien

Dietmar Haun

Anja Hall

Ulrich Kohler

Inhaltsverzeichnis

3.3.4. Probleme der Inhaltsanalyse	29
3.3.5. Beispiele	29
3.4.	

8.4.3. Lageregeln	64
8.4.4. Streuungswerte	64
8.5. •	

1. Wozu Sozialforschung und Statistik?

Gegenstand: Erfassung, Analyse und Interpretation empirischer Ph

2. Wissenschaftstheoretische Grundlagen

2. *Wissenschaftstheoretische Grundlagen*

2.1.2. Hypothesen, Gesetze und Theorien

Die Gesetze oder Allaussagen, die Teil der deduktiv-nomologischen{Erklärung sind, wer-

2.1. Kritischer Rationalismus

De niendum ist der ..

2. *Wissenschaftstheoretische Grundlagen*

seltener kann die Theorie angewandt werden, die Zahl der potentiellen Falsifikationen wird kleiner.

Als Beispiel dient die Aussage „Alle Schwäne können fliegen“ (i.e. „Wenn Objekt O

2. *Wissenschaftstheoretische Grundlagen*

2.2. *Wissenschaftstheoretische Kontroversen*

die objektive Begründung einer Norm nur durch die a priori Setzung einer anderen Norm erfolgen kann (*Munchhausen Trilemma*).

Neben der Debatte um die Wünschbarkeit brachte der Positivismusstreit eine Verdeutlichung des Objektivitätsbegriffs innerhalb des Begründungszusammenhangs. Von seiten des kritischen Rationalismus wurde betont, dass Wertungen auch innerhalb des Begr•

2. *Wissenschaftstheoretische Grundlagen*

Lakatos

Der bislang einflussreichste Versuch, die wissenschaftliche Praxis und die wissenschaft-

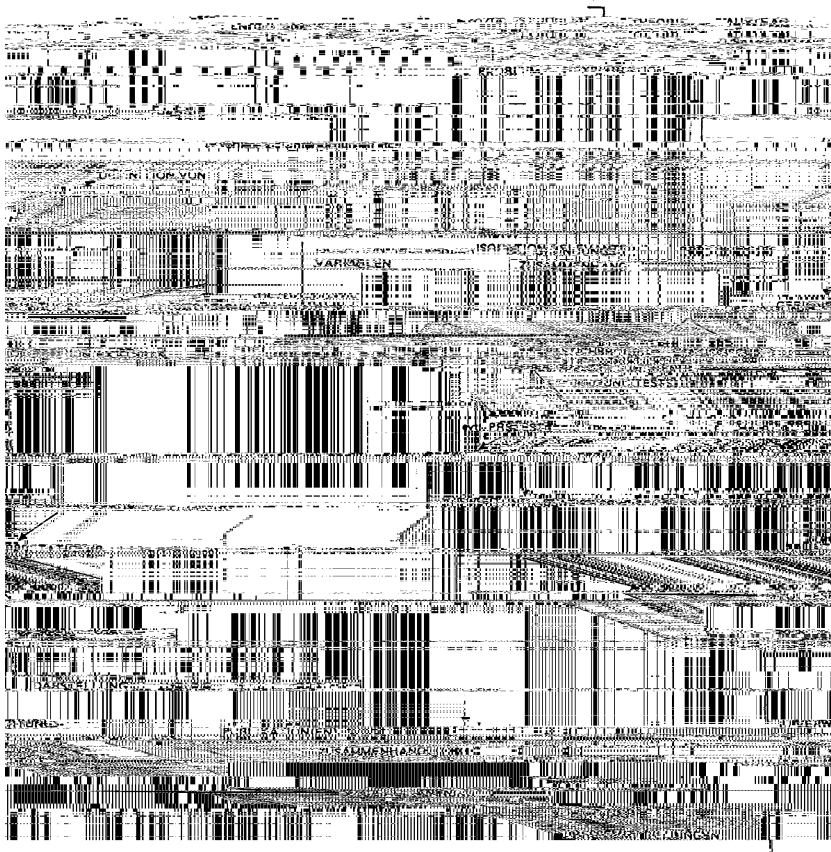


Abbildung 2.2.: Der Forschungsablauf (Friedrichs)

3. Erhebungsverfahren

3.1. Die Befragung

3.1.1. Typen der Befragung

Die Befragung gilt als das Standardinstrument der empirischen Sozialforschung. Befragungen lassen sich entlang der Befragungssituation und entlang des Strukturierungsgrades in unterschiedliche Typen einteilen:

1. Unterscheidung nach der Befragungssituation

- Interviews bzw. mündliche *face to face* Befragung

- Schriftliche (postalische) Befragungen

3. Erhebungsverfahren

Anderung der Messmethode (variiere Fragestellungen) unterschiedliche Antworten gegeben werden. In Abbildung 3.1 sind einige Vor- und Nachteile standardisierter Befragungen gegenüber weniger standardisierten Formen aufgeführt.

Vorteile	Nachteile
Ökonomischer z.B. durch geringere Kosten der Datenerfassung und -auswertung Zuordnung zu einer Kategorie erfolgt durch den Befragten selbst	Erhebliches Vorwissen erforderlich, aufwändige Fragebogenentwicklung Antwortvorgaben evtl. unzureichend (mögl. Folge: Verzerrung, Nichtbeantwortung)

3. Erhebungsverfahren

Fragebogen: Broschüren-Form, Frontseite und letzte Seite frei, verkleinert, weißes Papier

Anordnung der Fragen: mit leichten interessanten Fragen beginnen, inhaltlich gleiche Fragen zusammen, technisch einheitlich, Komplexe aufeinander aufbauen, sensible, bzw. heikle Fragen am Ende, Demografie am Ende

Design der Fragen: übersichtlich, Art und Anordnung der Buchstaben, Vermeidung von Frageteilungen, Fragefolge von oben nach unten, visuelle Hilfen

Implementation: Anschreiben: weißes Briefpapier, Anschrift auf Brief, Datum exakt, **N**ützlichkeit der Studie, **W**ichtigkeit des Befragten, Vertraulichkeit, Erklärung

3. Es7022 r gsverfahren

3.1. Die Befragung

6. „Das Umweltproblem wird in seiner Bedeutung von vielen Umweltschützern stark übertrieben

stimme eher zu
lehne eher ab

7. „Das Auto ist gar nicht so umweltschädlich wie oft gesagt wird

stimme eher zu
lehne eher ab

3.2. *Wissenschaftliche Beobachtung*

3. Erhebungsverfahren

3.2.2. Formen der Beobachtung

Beobachtungen können nach dem Strukturierungsgrad, der Beobachterrolle, der Transparenz, der Beziehung zum Beobachtungsobjekt und dem Beobachtungskontext unter-

3. Erhebungsverfahren

3.2.4. Probleme der Beobachtung

Stichprobenprobleme: Welche Situationen und welche in ihr vorkommenden Aktivitäten und Ereignisse werden beobachtet? Fragen der Bestimmung von Grundgesamtheit, Beobachtungseinheit.

Kategorisierung: Probleme der Konstruktion eines Beobachtungsschemas und der Operationalisierung von interessierenden Variablen bzw. Beobachtungskategorien; je feiner die Kategorien, umso präziser die Beobachtung, aber umso höher die Arbeitsintensität und der Auswertungsaufwand; siehe auch das o.g. Problem selektiver Ged

3.3. Inhaltsanalyse

nach der Uhrzeit, so wird diese Zuordnung der Kategorie 7 „Erfragte Informationen“ zugeordnet, und der Beobachter notiert auf einem mit gleichmäßiger Geschwindigkeit sich fortbewegenden Papierstreifen die Zielerfolge 4-7. Je zwei Kategorien sind

3.3. *Inhaltsanalyse*

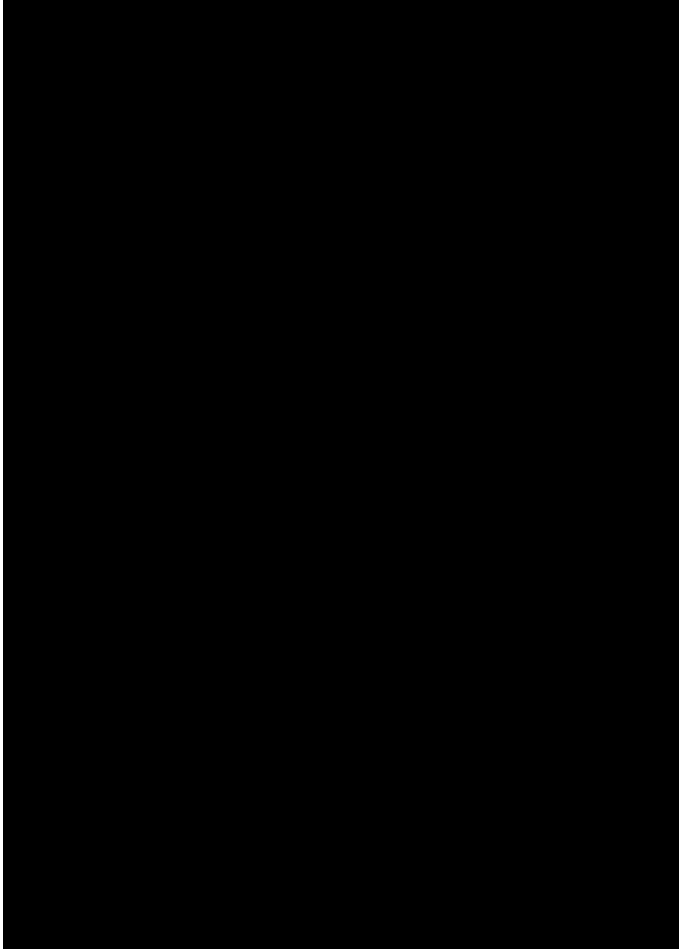
des Inhalts und der durch ihn repräsentierten sozialen Bedingungen aus ihm selbst

3. Erhebungsverfahren

3.4. •

4. *Forschungsdesign*

4. *Forschungsdesign*



4.4. Übungsaufgaben

3

5. Operationalisierung, Messung und Skalierung

5.1. Operationalisierung

Die Operationalisierung ist die Angabe einer Anweisung, wie *Objekten mit Eigenschaften*,

5. Operationalisierung, Messung und Skalierung

5.2.1. Skalen und Repräsentativitätstheorem

Eine

5. Operationalisierung, Messung und Skalierung

Nominalskalen sind Klassifikation der Objekte nach der *Relation der Gleichheit und Verschiedenheit*. Bei Nominalskalen sind alle *ein- eindeutigen* Transformatio-

5. Operationalisierung, Messung und Skalierung

Ein wahrer Wert kann als Mittelwert einer großen Zahl unabhängiger Messungen angesehen werden, wenn:

$$(E) = 0$$

$$TE = 0$$

$$E_1$$

5.2. Messen

Splithalf{Method: Das Instrument wird in zwei Hälften geteilt. Die Korrelation der entstehenden Instrumente wird als Maß für die Reliabilität verwendet. Da die Korrelation der Instrumente mit der Zahl der verwendeten Items steigt, wird sie entsprechend korrigiert:

Cronbachs{Alpha

5. *Operationalisierung, Messung und Skalierung*

Konstruktvalidit

Spezielle Skalen

Thurstone{Skala

1. Sammlung einer gro ßen Zahl von Items mit extremen und neutralen Aussagen
2. Beurteilung der Items durch „Beurteiler“
3. Auswahl der Items, die von den Beurteilern ähnlich eingeschätzt wurden

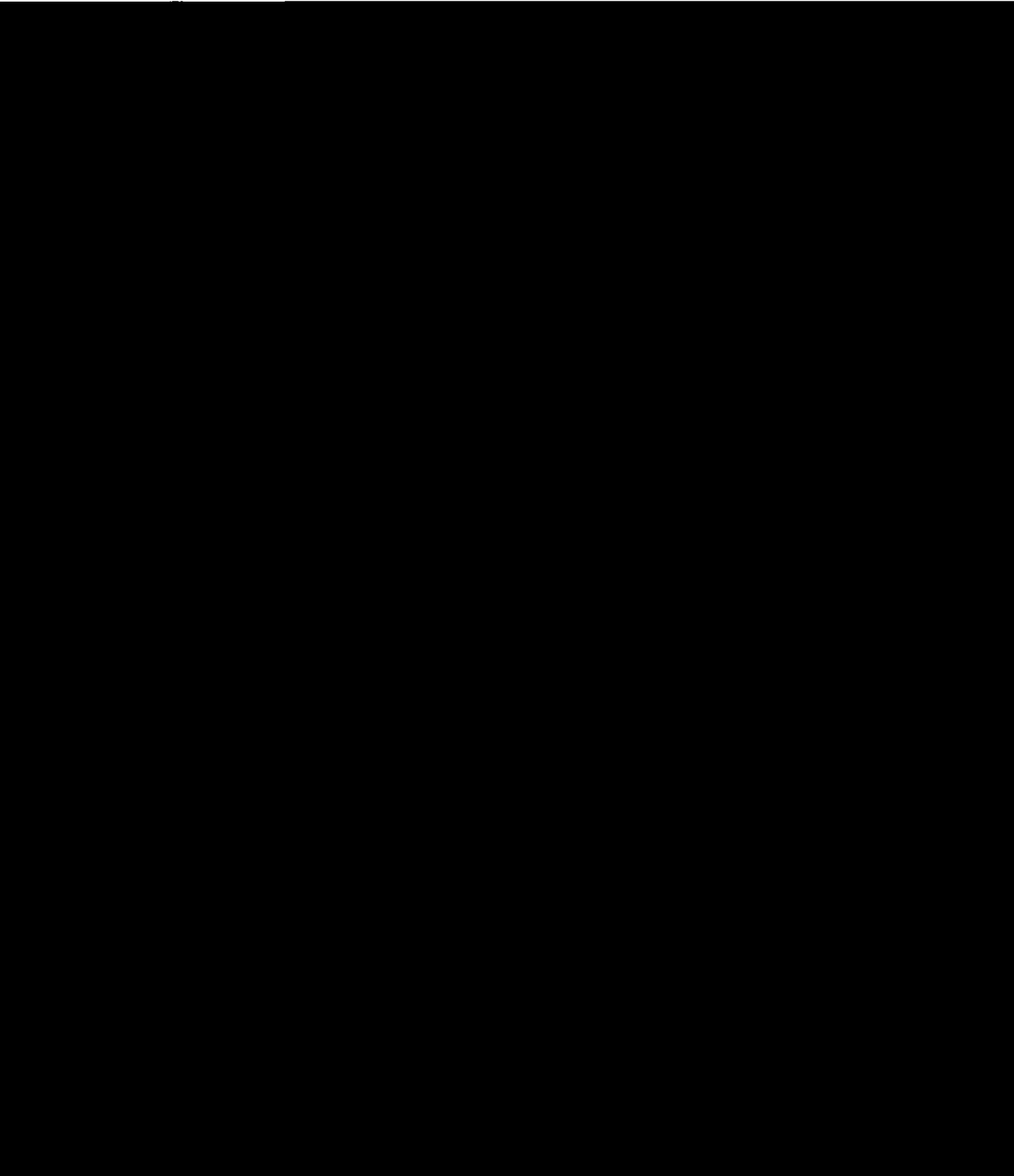
Es sollten ca. 20 - 30 Items ausgewählt werden, und es sollten sowohl extreme als auch neutrale Items enthalten sein.

Likert{Skala

1. Sammlung einer gro ßen Zahl von Items mit monotoner Itemcharakteristik
2. Abfragen der Items in einer Stichprobe
3. Berechnung von Skalenwerten für jeden Befragten (additiver Index)
4. Item{Analyse und Aussonderung ungeeigneter Items (Items, die von Personen

5. *Operationalisierung, Messung und Skalierung*

d) Blutgruppe von Personen?



7. Datenaufbereitung

7. Datenaufbereitung

7.3. Auswahlbibliografie

Eine Einführung in die Techniken der Datenaufbereitung findet sich bei (

7.3. Auswahlbibliografie

Variable	Variablenname	Codes	Spalte
lfn	Befragtennummer		1-4
v1	Geschlecht		5
	weiblich	1	
	m•		

8. Beschreibung von Verteilungen

8. Beschreibung von Verteilungen

wenn Sie ab und zu eine andere Partei wählen. Wie ist das bei Ihnen? Neigen auch Sie einer bestimmten Partei zu? Wenn ja: Welcher?

Ausprägung	j	h_j	f_j	100
------------	-----	-------	-------	-----

Lesebeispiel: 92.88 Prozent der Befragten identifizieren sich zumindest mäßig mit ihrer Partei.

Bei Variablen mit vielen Ausprägungen ist eine Häufigkeitstabelle nicht mehr übersichtlich. Handelt es sich um eine Variable mit mindestens ordinalem Skalenniveau, so kann die Variable jedoch in k Intervalle $[c_{j-1}; c_j)$ mit $j = 1; \dots; k$ unterteilt werden. In einer *gruppierten Häufigkeitstabelle* können die einzelnen Intervalle wie die Ausprägungen einer Variable mit wenig Ausprägungen dargestellt werden.

Beispiel In Tabelle 8.3 wurde das Einkommen in 5 gleichgroße Intervalle unterteilt. Das erste Intervall ist $[0;$

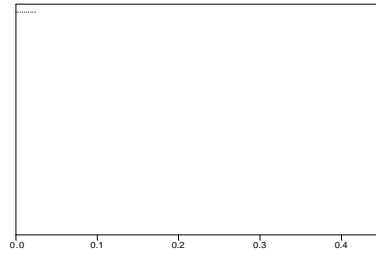


Abbildung 8.3.: Beispiel eines Dot-Charts

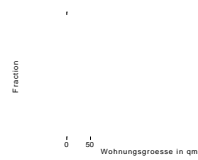


Abbildung 8.4.: Beispiel für Histogramme mit varrierendem Ursprung

8. *Beschreibung von Verteilungen*

davon erh•

8. Beschreibung von Verteilungen

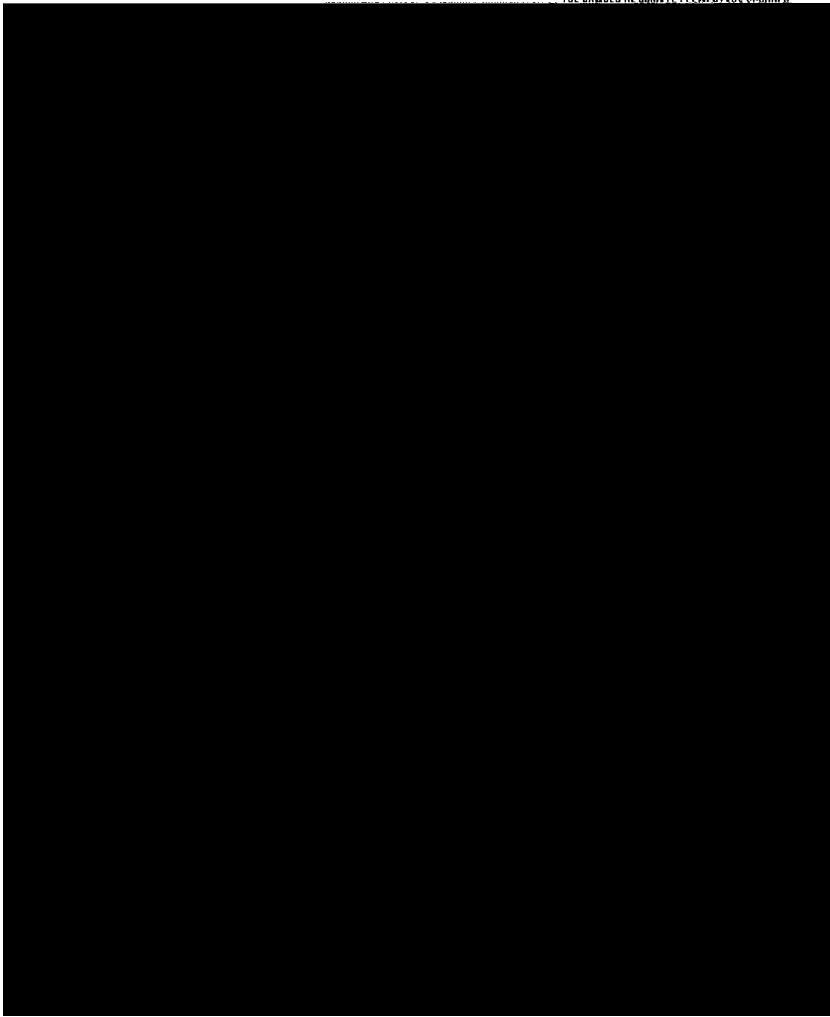


Abbildung 8.8.: Gestaltung von Grafen (Wainer

Aus einer gruppierten H

8. Beschreibung von Verteilungen

Aus der Häufigkeitsverteilung

$$\begin{aligned}x &= \sum_{j=1}^k a_j f_j \\ &= \frac{1}{n} \sum_{j=1}^k a_j h_j\end{aligned}\tag{8.7}$$

Aus einer gruppierten Häufigkeitsverteilung

$$\begin{aligned}x_{\text{grupp}} &= \sum_{j=1}^k f_j m_j \\ &= \frac{1}{n} \sum_{j=1}^k h_j m_j\end{aligned}\tag{8.8}$$

Der Interquartilsabstand

9. Kreuztabellenanalyse

9.1. Bivariate Kreuztabellenanalyse

9.1.1. Formale Struktur einer Kreuztabelle (Kontingenztafel)

Die Kontingenztafel gibt die gemeinsame Verteilung der Merkmale X und Y in absoluten H.

9. Kreuztabellenanalyse

$$a_1 \left| \begin{array}{ccc} b_1 & \dots & b_m \\ \hline f_{11} & \dots & f_{1m} \end{array} \right.$$

9.2. Drittvariablen-Kontrolle

Partei praef	Umwel tbewusstsein, 3er Teil ung			
erenz	niedrig	mittel	hoch	Total
-----+-----+-----+-----+-----				

9. Kreuztabellenanalyse

als Erklärung der Beziehung zwischen X und Y ausgeschlossen werden können. Die Kausalanalyse setzt strenggenommen somit ein korrekt spezifiziertes und vollständiges theoretisches Erklärungsmodell voraus, das alle relevanten Variablen zur Erklärung eindeutig

9.2.3. Beispiel 2: Erfolgsaussichten von ABM

Tabelle 9.9 zeigt die Anzahl von ehemaligen Teilnehmern an einer Arbeitsbeschäftigungs-

9.3. Übungsaufgaben

1. Sie haben die Vermutung, dass sich ältere Personen stärkere Sorgen •

10. Zusammenhangsanalyse in Kontingenztafeln

10.1. Prozentsatzdiagramm

$$a_1 \begin{array}{|c|c|} \hline b_1 & b_2 \\ \hline h_{11} & h_{12} \\ \hline \end{array} h$$

10. Zusammenhangsanalyse in Kontingenztabellen

sex	survi ved		Total
	nei n	j a	
frau	126 26.81	344 73.19	470 100.00
mann	1364 78.80	367 21.20	1731 100.00
Total	1490 67.70	711 32.30	2201 100.00

Tabelle 10.1.: Überlebende der Titanic-Katastrophe nach Geschlecht

10.2. Kreuzproduktverh

10.3. χ^2 -Quadrat basierte Maßzahlen

Der Chi-Quadrat-Koeffizient nach Pearson misst den Abstand einer empirischen $(m \times k)$ -Kontingenztafel zu einer $(m \times k)$

10. Zusammenhangsanalyse in Kontingenztabellen

Da X^2_{Pears} [0; 1

10. Zusammenhangsanalyse in Kontingenztabellen

Beispiel: $c \times f$

11. Zusammenhangsanalyse von metrischen Variablen

11.1. Graphische Darstellung

Zusammenhänge zwischen metrischen Variablen können graphisch in sog. Scatterplots bzw.

11. Zusammenhangsanalyse von metrischen Variablen

durch:

$$r = r_{XY} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{s_{XY}}{s_X s_Y} \quad (11.1)$$

wobei s

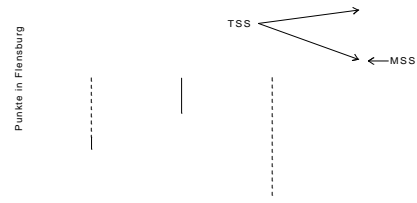
11. *Zusammenhangsanalyse von metrischen Variablen*

Zur Berechnung des Korrelationskoeffizienten empfiehlt es sich, die Daten in einer Arbeitstabelle zu übertragen. Diese sollte Spalten für

11.3. Determinationskoeffizient und lineare Regression

$$\sum (x_i - \bar{x})(y_i - \bar{y}) = \sum (x_i - \bar{x})$$

11. Zusammenhangsanalyse von metrischen Variablen



12. •

8. In einer Doppelblindstudie über den therapeutischen Nutzen von Vitamin C f•

12. *Übungsaufgaben*

3. Berechnen Sie den Korrelationskoeffizienten r
4. Wie hoch ist der Anteil linear erklärter Varianz?
5. Welches Einkommen lässt sich für eine Person voraussagen, die insgesamt 12 Bildungsjahre hat?
6. Wieviel Einkommenszuwachs bringt eine dreijährige Lehre?

11. Untersuchen Sie mit Hilfe des Verfahrens der linearen Regression, inwiefern die

13. Ergebnisse der Übungsaufgaben

L

13. Ergebnisse der Übungsaufgaben

Arbeitstabelle

Wohnungsgröße
[cm²]

Lösung zu Aufgabe 6

$$x_{\text{grupp}} = \frac{1}{100} \cdot 832 = 8$$

13. Ergebnisse der Übungsaufgaben

Nebenrechnung Quartilsabstand

$$x_{.25, \text{grupp}} = 600,5 + \frac{400 (0,25 - ,105)}{,161} = 961$$

$$x_{.75, \text{grupp}} = 1800,5 + \frac{400 (0,75 - ,636)}{,156} = 2093$$

Arbeitstabelle Standardabweichung

m_j	h_j	m_j	x_{grupp}	$(m_j - x_{\text{grupp}})^2$	$h_j (m_j - x_{\text{grupp}})^2$
-------	-------	-------	--------------------	------------------------------	----------------------------------

nicht unproblematisch. Insbesondere bei sehr breiten Intervallen kann die Abweichung gegenüber dem Midpoint recht ausgeprägt und das Gesamtergebnis entsprechend verfälscht sein.

Lösung zu Aufgabe 8

1.

$$d\% = 100(.78 - .88) = -10$$

d.h. Die Einnahme von Vitamin C verringert die Wahrscheinlichkeit, sich zu erkälten, um 10 Prozentpunkte.

Erkältung nach Einnahme von Vitamin C

	Erkältung j/n		
	nein	ja	
Placebo	.78	.22	1.0

Lösung zu Aufgabe 11

$$b = 9624$$

13. Ergebnisse der Übungsaufgaben

Literaturverzeichnis

Bandilla, Wolfgang Hauptmanns, Peter (1998). *Internetbasierte Umfragen als Datenerhebungstechnik für die empirische Sozialforschung*. ZUMA Nachrichten, 43:36{53.

LITERATURVERZEICHNIS

Esser, Hartmut, K. Kl enovits und H. Zehnpfennig (1977). *Wissenschaftstheorie*.

LITERATURVERZEICHNIS

- Popper, Karl R. (1994a). *Ausgangspunkte. Meine intellektuelle Entwicklung*. 2. Auflage. Hoffmann und Campe, Hamburg.
- Popper, Karl R. (1994b). *Logik der Forschung*. 10. Auflage. J.C.B. Mohr, Tübingen.
- Schmid, Josef (1995). *Expertenbefragung und Informationsgespräch in der Parteien-*

LITERATURVERZEICHNIS

A. Formelsammlung

A.5. Zusammenhangsanalyse metrischer Daten

Kreuzproduktverhältnis (Odds{ratio)

$$\text{Odds: } (1;2jX = a_i) = \frac{h_{i1}}{h_{i2}} \quad (\text{A.15})$$

Odds{ratio: (1;2jX

A. Formelsammlung

Korrelationskoeffizient

auch Produkt-Moment-Korrelationskoeffizient oder Bravais-Pearson-Korrelationskoeffizient genannt

$$r = \frac{\text{Cov}(x, y)}{\sigma_x \sigma_y}$$