



# Korrelationen und Skalentransformation

Marie Schmidt und Laura Dietl

Seminar: Evaluation und Forschungsstrategien

08.11.2018



# Die Skala

„...ein Messinstrument, mit dem man empirischen Gegenständen Zahlenwerte zuordnet, die der Stärke bestimmter Eigenschaften dieser Gegenstände entsprechen.“

Definition „Skala“ nach DORSCH, Lexikon der Psychologie

# Skalenniveaus

Skalenniveau	Nominalskala	Ordinalskala	Intervallskala	Ratioskala	Absolutskala
empirische Relation	Unterscheidbarkeit/ Gleichheit	+ Rangordnung	+ Linearität/ konstante Abstände	+ natürlicher Nullpunkt	+ natürliche Einheiten
Term	$a \neq b$ , $a = b$	$a < b$ , $a > b$	$a < b < c < d$ und damit $b - a = d - c$	$a/b = c/d$	$a = a$ Wert an sich
Beispiel	Geschlecht	höchster Schulabschluss	Jahr der Geburt	Einnahmen im Monat (in €)	Anzahl der Geschwister

zunehmender Informationsgehalt →

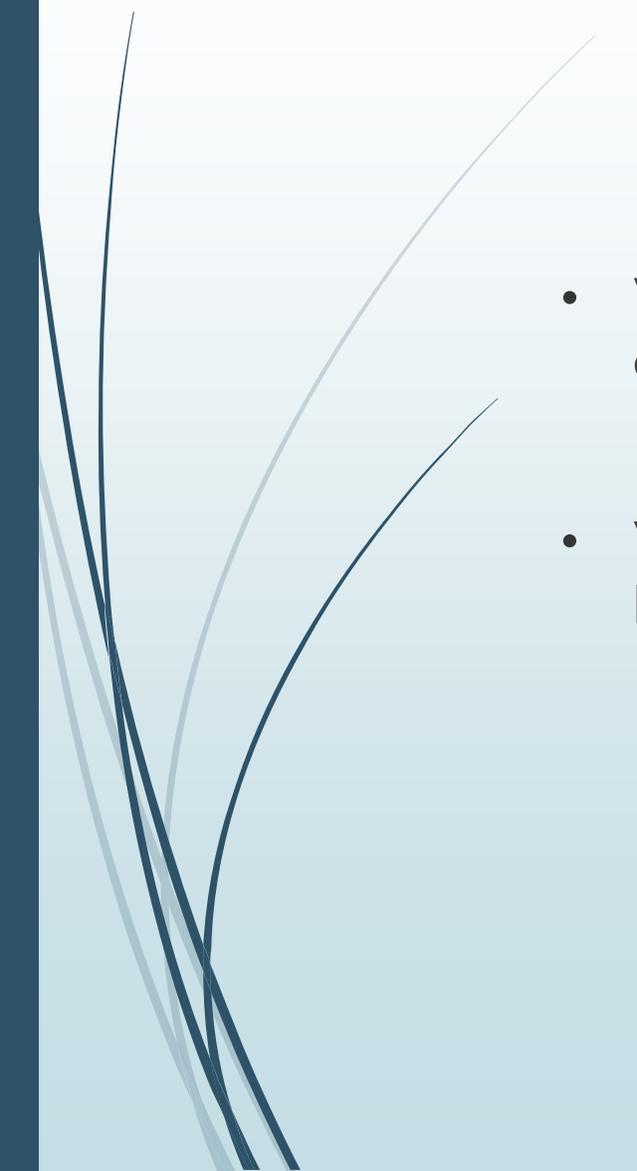


# Skalentransformation – z-Standardisierung

- Hintergrund:
  - Umwandlung, durch die Stichprobenwerte zueinander in Beziehung gesetzt werden
  - Werte können zu ihrer Lage in der Verteilung verglichen und beurteilt werden



# Anwendungsbereiche



- Vergleich der Werte zweier Personen unterschiedlicher Gruppen oder Stichproben
- Vergleich von Ausprägungen verschiedener Merkmale von einer Person



# z-Standardisierung

Ziele:

- Relative Lage von Werten in einer Verteilung darstellen
- Umwandlung von einer Skala in eine andere mithilfe von Erwartungswert und Standardabweichung

→ Nach der Transformation werden sie in Vielfachen der Standardabweichung gemessen

# Berechnung – die z-Transformation

$$z = \frac{x - \mu_x}{\sigma_x}$$

z = transformierter Stichprobenwert, auch z-Wert

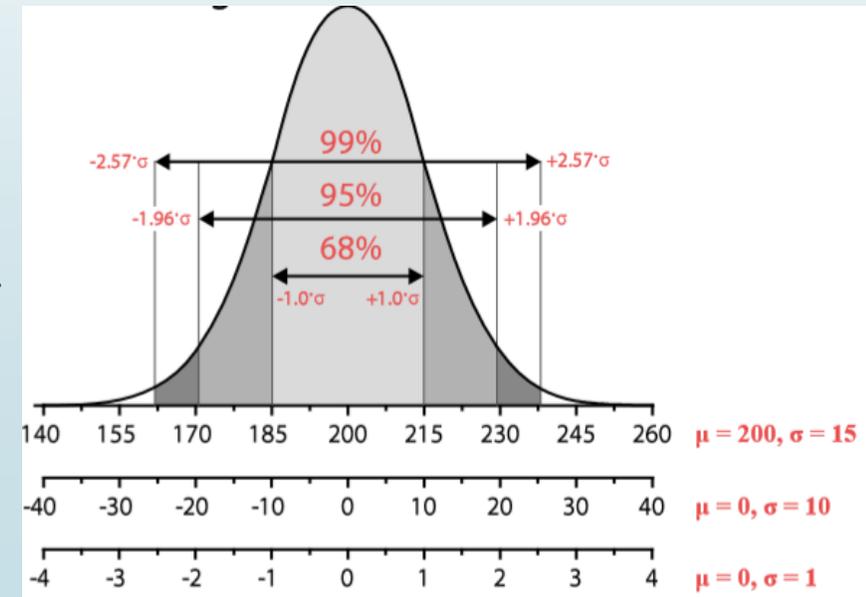
x = Stichprobenwert

$\mu$  = Erwartungswert

$\sigma$  = Standardabweichung

# z-Standardisierung und Normalverteilung

- Interpretation der z-Werte meist nur bei Normalverteilung sinnvoll
- Für normalverteilte z-Werte gilt:
  - $\mu = 0$
  - $\sigma = 1$
- Achtung! Durch eine z-Standardisierung sind die Werte **nicht** automatisch normalverteilt  
→ Die Verteilung der Werte bleibt gleich



# Skalentransformation

- Bildung von Werten definierten Erwartungswerts und Standardabweichung

1. z-Standardisierung

2.  $x_{neu} = (z * sneu) + \mu_{neu}$



# Wie verändern sich die Werte?

- Erwartungswert
  - Standardabweichung
  - Korrelation
- 

A dark grey arrow points to the right from the left edge of the slide. Several thin, light blue lines curve downwards from the arrow area towards the bottom left corner of the slide.

# Linearer Zusammenhang zwischen zwei Variablen:

## Kovarianz

- Problem: nicht invariant gegenüber erlaubter Transformationen
- Bei Addition einer Konstanten: kein Problem
- Bei Multiplikation: schwer zu interpretieren

# Produkt-Moment-Korrelation (Korrelationskoeffizient nach Pearson)

$$r_{xy} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{s_{xy}}{s_x \cdot s_y}$$

- Für intervallskalierte Daten
- Zähler: die Kovarianz
- Nenner: Standardabweichungen von X und Y
- Invariant gegenüber linearer Transformationen
- Immer zwischen -1 und 1
- Positiv: gleichsinniger Zusammenhang
- Negativ: gegensinniger Zusammenhang
- 0: kein Zusammenhang zwischen zwei Variablen

# Faustregeln nach Cohen

$r < \pm 0.10$  → keine Korrelation

$r < \pm 0.30$  → kleine Korrelation

$r < \pm 0.50$  → mittlere Korrelation

$r \geq \pm 0.50$  → hohe Korrelation



**X: natürlich dichotom  
nominalskaliert**

**Y: intervallskaliert**

**Punktbiseriale Korrelation**

**X: künstlich dichotom  
nominalskaliert**

**Y: intervallskaliert**

**Biseriale Korrelation**

Bei beiden gleiche Interpretation wie  
bei Korrelation nach Pearson

# X und Y: künstlich dichotomisiert und eigentlich normalverteilt

## Tetrachorische Korrelation

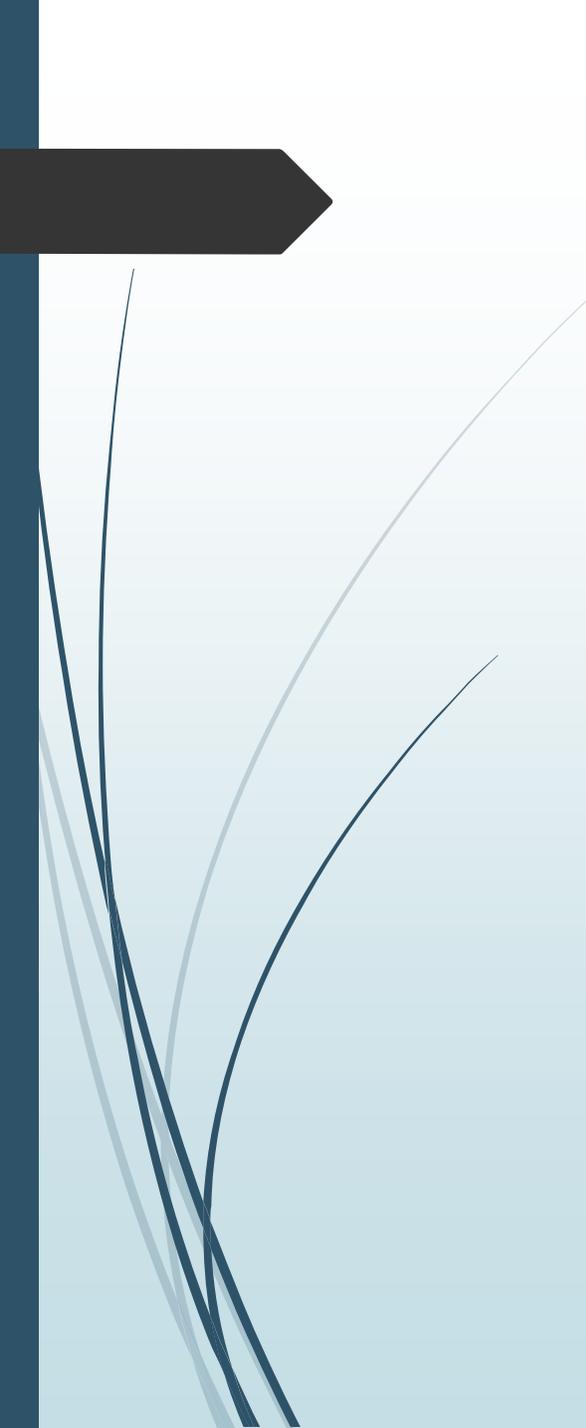
	$x_1$	$x_2$	
$y_1$	$n_{11}$	$n_{12}$	$n_{1\bullet}$
$y_2$	$n_{21}$	$n_{22}$	$n_{2\bullet}$
	$n_{\bullet 1}$	$n_{\bullet 2}$	$n_{\bullet\bullet}$

- 2X2 Kontingenztabelle
- Überschätzung der wahren Korrelation, wenn die Randverteilungen stark asymmetrisch sind oder ein  $n_{xy} < 5$
- Excel:  $\text{COS}(\text{PI}() / (1 + \text{WURZEL}((n_{11} * n_{22}) / (n_{12} * n_{21}))))$

A dark blue arrow points to the right at the top left. Below it, several thin, curved lines in shades of blue and grey sweep across the left side of the slide.

# Spearman's $r_s$

- X und Y sind ordinalskaliert
  - die Ausprägungen sind unterscheidbar
  - die Ausprägungen haben eine natürliche Ordnung
  - Abstände sind nicht interpretierbar
- Zunächst werden die Ränge gebildet
  - Excel: RANG.MITTELW
- Danach Produkt-Moment-Korrelation der Ränge
- Außerdem Vorteil: robust gegenüber Ausreißern



**Korrelation  
bedeutet nicht  
Kausalität**



## X und Y: natürlich dichotom

### Phi-Koeffizient

- Für absolute und relative Häufigkeiten
- Problem: bei schiefen Randverteilungen kann die Grenze  $\pm 1$  nicht erreicht werden
- Daher muss man dann an der max. möglichen Korrelation normieren

# Chi<sup>2</sup>

- für kxm Kontingenztabellen
- nur für absolute Häufigkeiten
- Vergleich der Tabelle mit einer Tabelle, die bei perfekter Unabhängigkeit entstanden wäre

$$\frac{(beob - erw)^2}{erw}$$

- bei perfekter Unabhängigkeit 0, ansonsten >0
- Kann beliebig große Werte annehmen → schwer zu interpretieren

# Cramérs V

- für kxm Kontingenztabellen
- nur für absolute Häufigkeiten
- Interpretierbar
- Zwischen 0 und 1

$$V = \sqrt{\frac{\chi^2}{n_{..} \cdot (\min(k, m) - 1)}}$$

mit k = Anzahl der Zeilen; m = Anzahl der Spalten

# Quellen

- ▶ Bortz, J. & Schuster, C. (2010). Statistik für Human- und Sozialwissenschaftler (7. Aufl.). Heidelberg: Springer. S 35 - 36
- ▶ K. Völkl & C. Korb (2017). Deskriptive Statistik, Elemente der Politik. S. 15
- ▶ [https://iversity.org/de/my/courses/einfuehrung-in-die-wahrscheinlichkeitsrechnung--17/lesson\\_units/22553](https://iversity.org/de/my/courses/einfuehrung-in-die-wahrscheinlichkeitsrechnung--17/lesson_units/22553) (25.10.18 19:40 Uhr)
- ▶ [https://iversity.org/de/my/courses/primer-deskriptive-statistik/lesson\\_units](https://iversity.org/de/my/courses/primer-deskriptive-statistik/lesson_units) (29.10.18 13:54)
- ▶ <https://www.spektrum.de/lexikon/psychologie/z-transformation/17279> (25.10.18 19:00 Uhr)
- ▶ [http://www.statistics4u.info/fundstat\\_germ/ee\\_ztransform.html](http://www.statistics4u.info/fundstat_germ/ee_ztransform.html) (25.10.18 19:20 Uhr)