universität innsbruck

Institut für Geographie

Everything hangs together: Zusammenhangsmaße

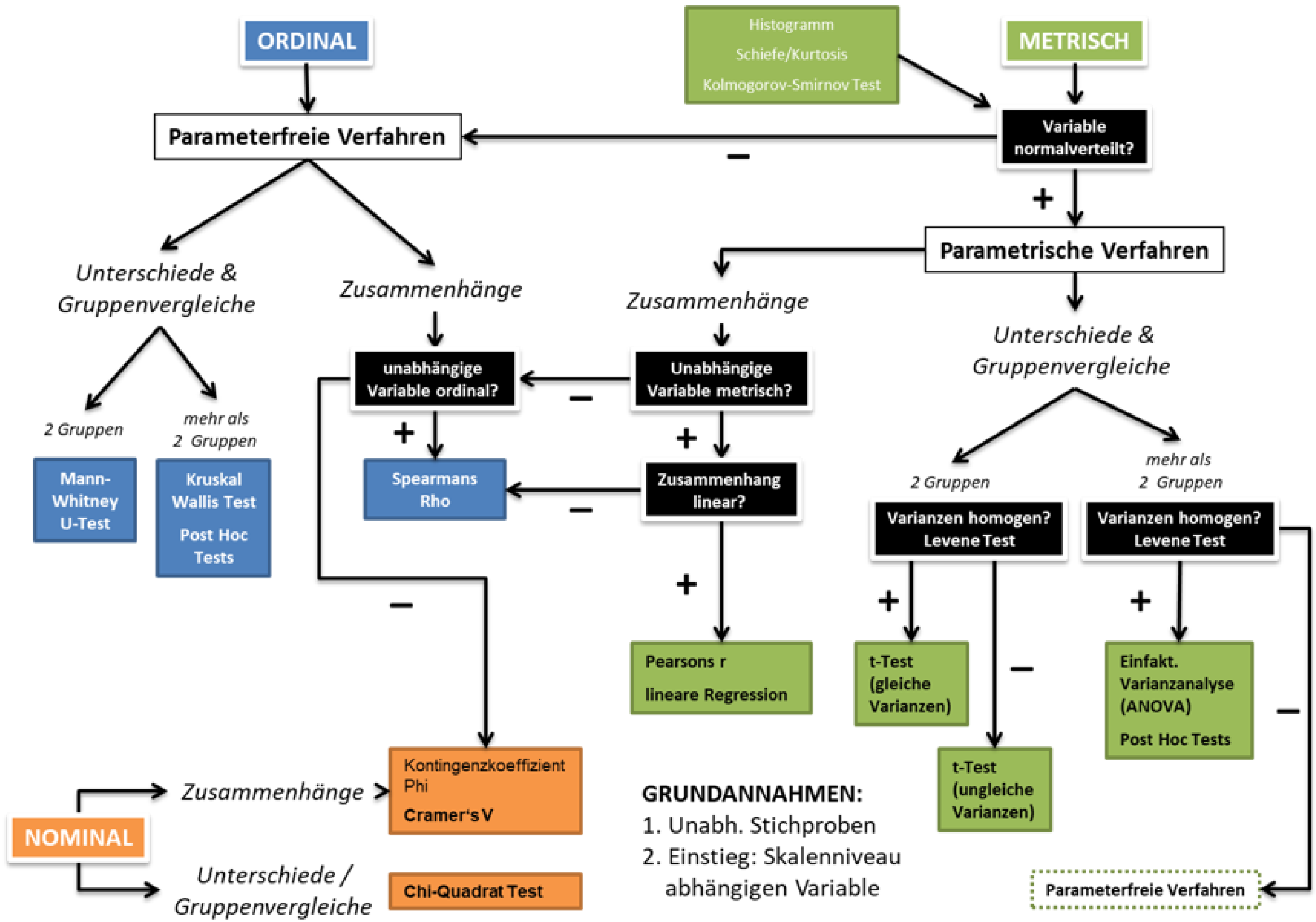
716408 | Sozialwiss. Methoden – How 2 do Things with Numbers

KMH

SS 22 (updated: 2022-06-05)

Everything hangs together?

Ein Überblick auf ausgewählte Maße



(Eigene Überarbeitung 2016 von Hager, 2011)

Exkurs: Normalverteit?

Warum ist das wichtig?

Zentraler Grenzwertsatz der Statistik:

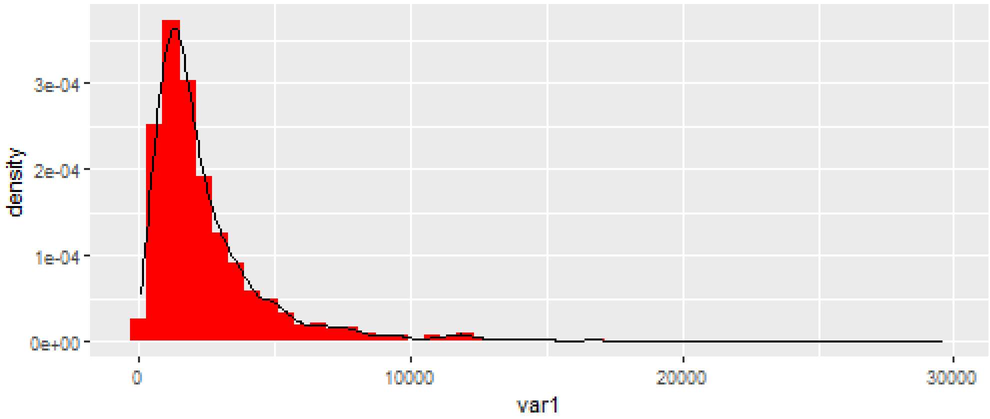
Verteilungen (= Messungen) die unter einer großen Anzahl unabhängiger Einflüsse (= Abweichungen) ermittelt werden:
👉 **normalverteilt**

BSP: 500 Zufallszahlen

Überprüfen der Normalverteilung 1/3

1. Visuell: Histogramm & Normalverteilungskurve

```
ggplot(beispieldaten, aes(x = var1, y = ..density..)) +
  geom_histogram(bins = 50, fill = "red") +
  geom_density()
```



Bei Normalverteilung:

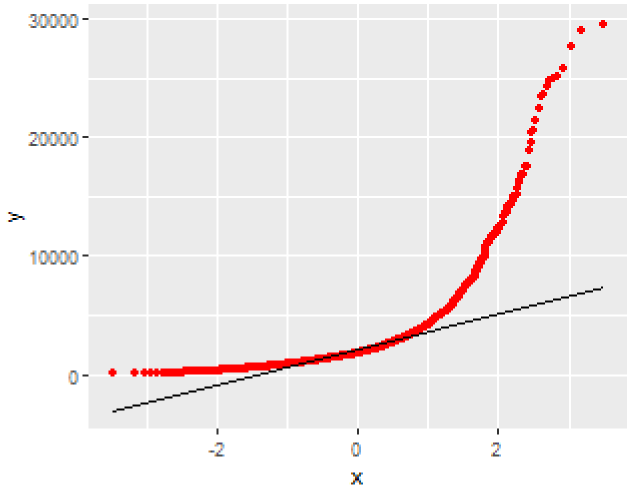
- symmetrische Verteilung um Mittelwert
- Dichtekurve = Glockenkurve

👉 hier nicht gegeben → **vermutlich keine Normalverteilung**

Überprüfen der Normalverteilung 2/3

2. Visuell: Q(uantil)-Q(uantil) Plots

```
ggplot(beispieldaten, aes(sample = var1)) +
  geom_qq(color="red") +
  geom_qq_line()
```



Bei Normalverteilung:

- beobachtete Werte liegen auf der Geraden (= ideale Normalverteilung)

👉 hier nicht gegeben → **vermutlich keine Normalverteilung**

Überprüfen der Normalverteilung 3/3

3. Numerisch: [Der Shapiro-Wilk-Test]

```
shapiro.test(beispieldaten$var1)
```

```
##
##      Shapiro-Wilk normality test
##
## data:  beispieldaten$var1
## W = 0.62872, p-value < 2.2e-16
```

Interpretation:

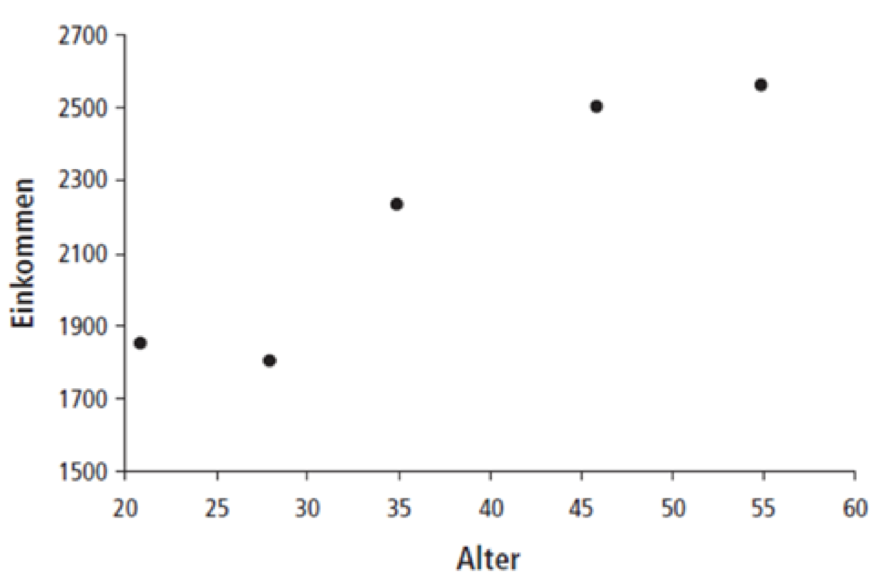
- **H0:** Normalverteilung ist gegeben
- **H1:** Normalverteilung ist nicht gegeben

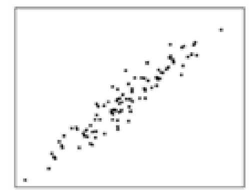
👉 $p = 2.2e\text{-}16 < \alpha \ (= 0{,}05)$
👉 H0 ablehnen → **keine Normalverteilung gegeben**

Back 2 the story: Metrische Zusammenhänge messen?

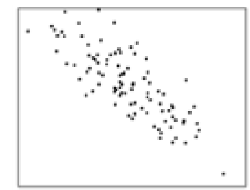
Zusammenhänge graphisch erschließen

- **Scatterplot (aka "Streudigramm"):**
  grafische Darstellung von Wertpaaren zweier metrischer Merkmale



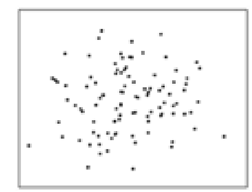


Positiver Zusammenhang



negativer Zusammenhang

Person	Alter	Einkommen [€]
A	21	1.850
B	46	2.500
C	55	2.560
D	35	2.230
E	28	1.800



kein Zusammenhang

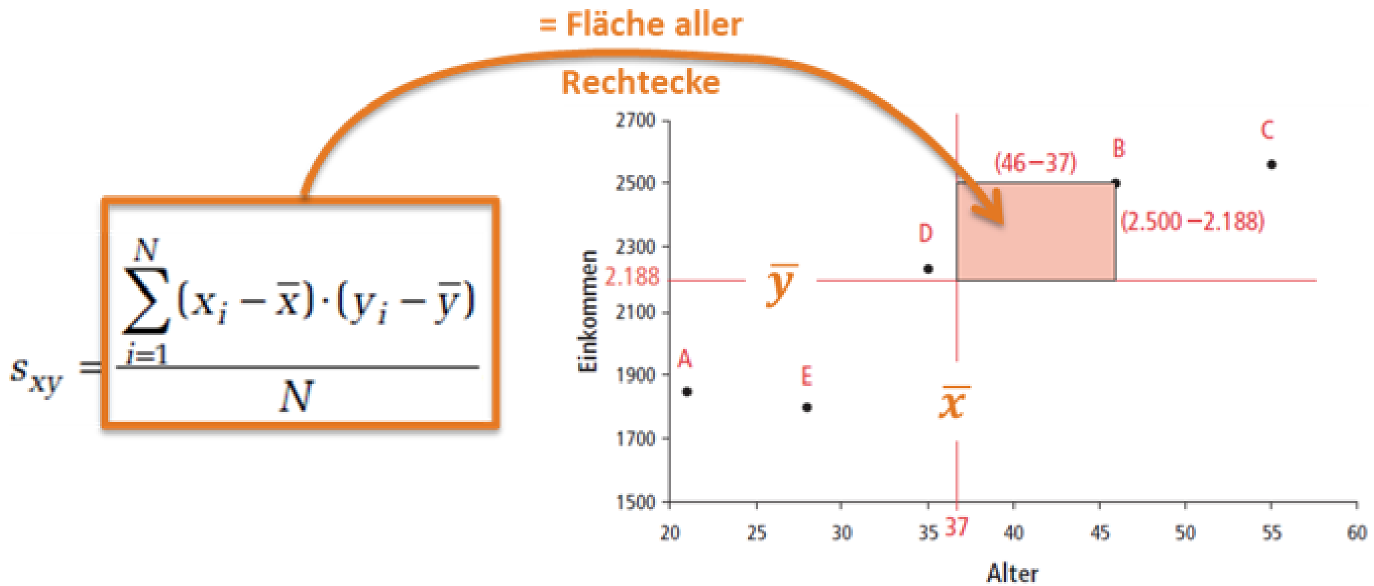
(Eigene Überarbeitung 2018 von: Quatember 2007:667)

👉 nach **Verteilungsmustern** Ausschau halten

Zusammenhänge numerisch messen

Ermittlung der Korrelation r:

- Berechnen der Kovarianz S_{xy}



(Eigene Überarbeitung 2019 von: Quatember 2007:69)

- Kovarianz ohne obere & untere Beschränkung
  → **Normierung auf Wertbereich -1 bis +1** notwendig

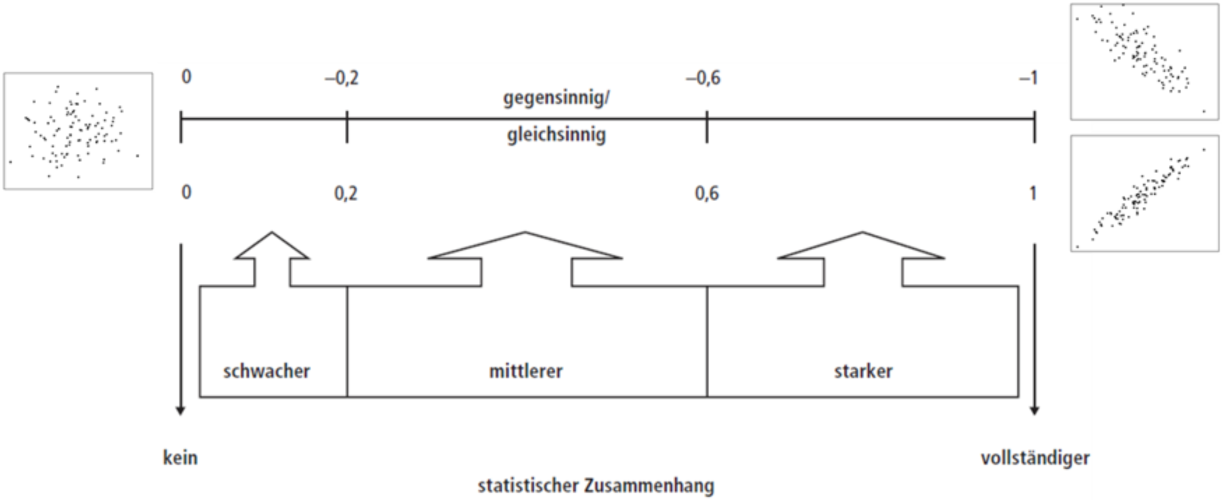
Zusammenhänge numerisch messen

Korrelationskoeffizient r [-1; +1]

$$r = \frac{S_{xy}}{S_x * S_y}$$

S_x ... Standardabweichung Variable x
S_y ... Standardabweichung Variable y

👉 Interpretation:



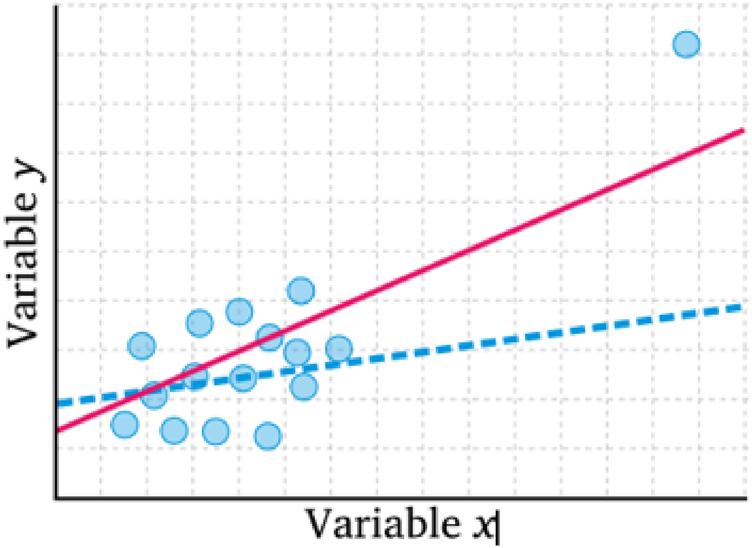
(Eigene Überarbeitung 2019 von: Quatember 2007:71)

… und wenn meine Variablen NICHT normalverteilt sind?

Spearmans ρ für (ordinale) Zusammenhänge

- **Pearsons Problem:** linearer Zusammenhang → Ausreißer

- **Umgang damit:**
  - Prüfen der Normalverteilung der verwendeten Variablen
  - Pearsons r & Spearmans ρ ermitteln → „deutliche" Unterschiede = Ausreißer verzerren Pearsons r
    - Ausreißer ausschließen



(Matheguru, o.J.)

Spearmans ρ für (ordinale) Zusammenhänge

- Misst **beliebige monotone Zusammenhänge:**
  - Wertbereich: wie Pearsons r (-1 bis +1)
  - Grundlage: Differenz in den Rängen eines Wertpaares
  - Praktisch: Pearsons r auf Basis von Rängen
- **Problem:** Bindungen („Ties“)
  - Daumenregel: unter 20% aller Wertpaare
    → Ränge über Mittelwerte

$$r_s = 1 - \frac{6 \sum_i d_i^2}{n \cdot (n^2 - 1)}$$

$$d_i = rg_x i - rg_y i$$

n ... Anzahl der Wertpaare
rg_{xi} bzw. rg_{yi} ... Rang des i-ten Ausprägung in Variable x bzw. y