

Download der aktuellen Version unter <https://johannes-titz.github.io/formelsammlung/main.pdf>

Diese Formelsammlung wurde für Studierende der Psychologie entwickelt, ist aber vermutlich auch für Studierende anderer Sozialwissenschaften hilfreich. Die Formelsammlung basiert auf einer alten Sammlung der Professur für Forschungsmethoden und Evaluation in der Psychologie an der TU Chemnitz, die weder Autoren noch Lizenz erhielt. Für eine sinnvolle Nutzung und Weiterentwicklung ist eine Lizenz notwendig, weshalb wir das vorliegende Dokument unter CC BY-SA 4.0 veröffentlichen. Die Formelsammlung orientiert sich am Lehrbuch von Sedlmeier und Renkewitz (2018). Die Autoren der Formelsammlung sind: Johannes Titz, Nils Heimhuber, Feline Baumgärtel. Weitere Mitwirkende: Vivien Lungwitz, Annika Sternkopf. Für wertvolle Hinweise bedanken wir uns bei: Lea Haase.

Einige Hinweise vorab: Lateinische Buchstaben ( $\bar{x}, s^2, s$ ) werden für die Stichprobe verwendet, griechische Buchstaben ( $\mu, \sigma^2, \sigma$ ) für die Population. Ein Dach über einer Statistik ( $\hat{\sigma}$ ) steht für eine Schätzung des Parameters. Abkürzungen:

MWU: Mittelwertsunterschied US: Unabhängige Stichproben AS: Abhängige Stichproben

## **Lagemaße**

### **Modalwert**

Der Modalwert ist der häufigste Wert aller Messwerte. Es kann mehrere Modal-Werte geben.

### **Mittelwert**

Auch bezeichnet als arithmetisches Mittel, für den Mittelwert der Population wird der Buchstabe  $\mu$  verwendet.

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

### **Median**

Der Median ist der Wert in der Mitte aller in einer Rangreihe geordneten Messwerte.

$$\text{Tiefe}_{\text{Med}} = \frac{n+1}{2}$$

### **Quartile**

Die Werte müssen in eine Rangreihe gebracht werden. Für das untere Quartil wird die Tiefe von unten gezählt, fürs obere Quartil von oben. Quartile werden manchmal auch als die Quantile  $Q_{25}, Q_{75}$  bezeichnet.

$$\text{Tiefe}_{\text{Quartil}} = \frac{\lfloor \text{Tiefe}_{\text{Med}} \rfloor + 1}{2}$$

### **Zäune bei Boxplots**

Die Zäune werden benötigt um die Whiskers zu bestimmen. Die Whiskers sind durch tatsächlich vorkommende Werte repräsentiert. Von den Zäunen aus geht man in Richtung Box, bis man den ersten vorkommenden Wert findet.

$$\begin{aligned} \text{Zaun}_{\text{unten}} &= Q_{25} - 1,5 \cdot \text{IQA} \\ \text{Zaun}_{\text{oben}} &= Q_{75} + 1,5 \cdot \text{IQA} \end{aligned}$$

## **Streuungsmaße**

### **Standardabweichung**

$$s = \sqrt{s^2}$$

### **Varianz**

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

### **Schätzung für Populationsvarianz**

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{n}{n-1} \cdot s^2$$

Hinweis: In der mathematischen Statistik wird die *Stichprobenvarianz* üblicherweise mit dem Nenner  $n-1$  definiert, da dieser Schätzer erwartungstreue für die Populationsvarianz ist. In den Sozialwissenschaften wird hingegen oft die Variante mit  $n$  als *Varianz der Stichprobe* bezeichnet. Hat man dagegen die gesamte Population, so wird auch die *Populationsvarianz* mit  $n$  im Nenner berechnet.

### **Spannweite (Range)**

$$R = x_{\max} - x_{\min}$$

### **Mittlerer absoluter Abstand**

$$\text{MAD} = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|$$

### **Interquartilsabstand**

$$\text{IQA} = Q_{75} - Q_{25}$$

## Transformation

### **z-Standardisierung**

$$z_i = \frac{x_i - \bar{x}}{s}$$

$$\bar{z} = 0$$

$$s_z = 1$$

### **Zentrierung**

$$c_i = x_i - \bar{x}$$

### **IQ-Standardisierung**

$$IQ_i = z_i \cdot 15 + 100$$

## Standardfehler für Stichprobenverteilungen

### **Binomialverteilung**

$$\sigma_{\text{Anzahl}} = \sqrt{np(1-p)}$$

$$\sigma_{\text{Anteil}} = \sqrt{\frac{p(1-p)}{n}}$$

### **Mittelwerte**

$$\hat{\sigma}_{\bar{x}} = \frac{\hat{\sigma}}{\sqrt{n}} = \frac{s}{\sqrt{n-1}}$$

### **MWU für AS**

$$\hat{\sigma}_{\bar{d}} = \frac{\hat{\sigma}_d}{\sqrt{n}}$$

### **MWU für US**

wenn  $n_A = n_B$ :

$$\hat{\sigma}_{\bar{x}_A - \bar{x}_B} = \sqrt{\hat{\sigma}_{\bar{x}_A}^2 + \hat{\sigma}_{\bar{x}_B}^2}$$

wenn  $n_A \neq n_B$

$$\hat{\sigma}_{\bar{x}_A - \bar{x}_B} = \sqrt{\frac{(n_A - 1) \cdot \hat{\sigma}_A^2 + (n_B - 1) \cdot \hat{\sigma}_B^2}{(n_A - 1) + (n_B - 1)} \left( \frac{1}{n_A} + \frac{1}{n_B} \right)}$$

Bei Ungleichheit der Varianzen in der Population

$$\hat{\sigma}_{\bar{x}_A - \bar{x}_B} = \sqrt{\frac{\hat{\sigma}_A^2}{n_A} + \frac{\hat{\sigma}_B^2}{n_B}}$$

## Zusammenhangsmaße

### **Kovarianz**

$$\text{cov}(x, y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

	nein	ja
nein	a	b
ja	c	d

### **Korrelation**

Pearson-Korrelationskoeffizient  
(Produkt-Moment-Korrelation)

$$r = \frac{\text{cov}(x, y)}{s_x s_y} = \frac{1}{n} \sum_{i=1}^n z_{x_i} \cdot z_{y_i}$$

### **Partialkorrelation**

$$r_{xy.z} = \frac{r_{xy} - r_{xz} \cdot r_{yz}}{\sqrt{1 - r_{xz}^2} \cdot \sqrt{1 - r_{yz}^2}}$$

### **Phi-Koeffizient (für dichotome Variablen)**

$$\Phi = \frac{a \cdot d - b \cdot c}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

### **Interrater-Reliabilität (Kappa)**

$$\kappa = \frac{p_o - p_e}{1 - p_e}$$

$$p_e = \sum_{i=1}^k p_{i \cdot} p_{\cdot i}$$

$p_o$ : beobachtete Übereinstimmung

$p_e$ : erwartete Übereinstimmung unter Zufall

$p_{i \cdot}$ : die Randwahrscheinlichkeit der Zeile  $i$

$p_{\cdot i}$ : die Randwahrscheinlichkeit der Spalte  $i$

$k$ : die Anzahl der Kategorien

### Regressionsgleichung

$$\begin{aligned}\hat{y}_i &= a + b \cdot x_i \\ b &= \frac{\text{cov}(x,y)}{s_x^2} = r \frac{s_y}{s_x} \\ a &= \bar{y} - b\bar{x} \\ y_i &= \hat{y}_i + e_i\end{aligned}$$

### Determinationskoeffizient

$$R^2 = \frac{s_{\hat{y}}^2}{s_y^2}$$

### Standardschätzfehler

$$\begin{aligned}s_e &= \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} = s_y \cdot \sqrt{1 - R^2} \\ \hat{\sigma}_e &= \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - 2}}\end{aligned}$$

## Multiple Regression

In den folgenden Formeln wird zur Veranschaulichung der Fall mit zwei Prädiktoren dargestellt. Für mehr als zwei Prädiktoren lassen sich die Regressionsgleichung, die Beziehung zwischen  $b$  und  $\beta$ , die Berechnung von  $b_0$  sowie die Formeln für den Standardschätzfehler analog erweitern. Die Berechnung der  $\beta$ -Gewichte und von  $R^2$  erfordert jedoch eine Matrixschreibweise, die nicht überall unterrichtet wird.

### Regressionsgleichung

für Originalwerte:

$$\hat{y}_i = b_0 + b_1 \cdot x_{1i} + b_2 \cdot x_{2i}$$

wobei:

$$\begin{aligned}b_1 &= \beta_1 \cdot \frac{s_y}{s_{x_1}} \\ b_2 &= \beta_2 \cdot \frac{s_y}{s_{x_2}} \\ b_0 &= \bar{y} - b_1 \cdot \bar{x}_1 - b_2 \cdot \bar{x}_2\end{aligned}$$

für z-standardisierte Werte:

$$\hat{z}_{y_i} = \beta_1 \cdot z_{x_{1i}} + \beta_2 \cdot z_{x_{2i}}$$

wobei:

$$\begin{aligned}\beta_1 &= \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1 x_2}}{1 - r_{x_1 x_2}^2} \\ \beta_2 &= \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1 x_2}}{1 - r_{x_1 x_2}^2}\end{aligned}$$

### Gütemaße

Multipler Determinationskoeffizient

$$R^2 = \beta_1 \cdot r_{yx_1} + \beta_2 \cdot r_{yx_2}$$

Multipler Korrelationskoeffizient

$$R = \sqrt{R^2}$$

Standardschätzfehler für Originalwerte

$$\begin{aligned}s_e &= \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} = s_y \sqrt{1 - R^2} \\ \hat{\sigma}_e &= \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - k - 1}}\end{aligned}$$

$k$  ist die Anzahl der Prädiktoren

Standardschätzfehler für z-standardisierte Werte

$$s_e = \sqrt{1 - R^2}$$

## Konfidenzintervalle

### Binomialverteilung (Approximation)

Die folgende Approximation über die Normalverteilung ist nur nutzbar, wenn  $\sigma_{\text{Anzahl}}^2 = n \cdot p \cdot (1 - p) > 9$

### Mittelwert

$$\bar{x} \pm \hat{\sigma}_{\bar{x}} \cdot t_{\text{df}, \text{Konfidenz}}$$

### MWU, US

für Anteile

$$\hat{p} \pm \sigma_{\text{Anteil}} \cdot z_{\text{Konfidenz}}$$

$$(\bar{x}_A - \bar{x}_B) \pm \hat{\sigma}_{\bar{x}_A - \bar{x}_B} \cdot t_{\text{df}, \text{Konfidenz}}$$

für Anzahlen/Objekte

### MWU, AS

$$\hat{p} \cdot n \pm \sigma_{\text{Anzahl}} \cdot z_{\text{Konfidenz}}$$

$$\bar{d} \pm \hat{\sigma}_{\bar{d}} \cdot t_{\text{df}, \text{Konfidenz}}$$

## t-Test

### Mittelwert gegen Konstante

wobei c (constant) vorgegeben

$$t = \frac{\bar{x} - c}{\hat{\sigma}_{\bar{x}}}$$

$$\text{df} = n - 1$$

### MWU, US

### MWU, AS

$$t = \frac{\bar{x}_A - \bar{x}_B}{\hat{\sigma}_{\bar{x}_A - \bar{x}_B}}$$

$$\text{df} = (n_A - 1) + (n_B - 1)$$

$$t_{\text{AS}} = \frac{\bar{d}}{\hat{\sigma}_{\bar{d}}}$$

$$\text{df}_{\text{AS}} = n - 1$$

Hinweis: Berechnung der df gilt nur bei Gleichheit der Populationsvarianzen.

## ANOVA US

### F-Wert

### Quadratsummen

$$F(\text{df}_{\text{zw}}, \text{df}_{\text{inn}}) = \frac{\hat{\sigma}_{\text{zw}}^2}{\hat{\sigma}_{\text{inn}}^2}$$

### Schätzer der Populationsvarianzen

$$\hat{\sigma}_{\text{zw}}^2 = \frac{\text{QS}_{\text{zw}}}{\text{df}_{\text{zw}}}$$

$$\hat{\sigma}_{\text{inn}}^2 = \frac{\text{QS}_{\text{inn}}}{\text{df}_{\text{inn}}}$$

Hinweis: diese werden auch oft als Mean Squares (MS) bezeichnet.

### Quadratsummenzerlegung

$$\text{QS}_{\text{ges}} = \text{QS}_{\text{zw}} + \text{QS}_{\text{inn}}$$

$$\text{QS}_{\text{zw}} = \sum_{j=1}^k n_j \cdot (\bar{x}_j - \bar{x})^2$$

$$\text{QS}_{\text{inn}} = \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2$$

### Freiheitsgrade

$$\text{df}_{\text{zw}} = k - 1$$

$$\text{df}_{\text{inn}} = \sum_{j=1}^k (n_j - 1) = N - k$$

### Notation

$k$ : Anzahl der Bedingungen

$n_j$ : Anzahl der Messwerte pro Bedingung  $j$

$N$ : Gesamtzahl der Messwerte

$\bar{x}$ : Gesamtmittelwert

Hinweis: Gilt für balanciertes Design.

### Quadratsummenzerlegung

#### F-Wert

$$QS_{\text{ges}} = QS_{\text{UV}} + QS_{\text{Person}} + QS_{\text{res}}$$

$$F = \frac{\hat{\sigma}_{\text{UV}}^2}{\hat{\sigma}_{\text{res}}^2}$$

#### Freiheitsgrade

#### Schätzer der Populationsvarianzen

$$\hat{\sigma}_{\text{UV}}^2 = \frac{QS_{\text{UV}}}{df_{\text{UV}}}$$

$$\hat{\sigma}_{\text{res}}^2 = \frac{QS_{\text{res}}}{df_{\text{res}}}$$

$$df_{\text{gesamt}} = N - 1$$

$$df_{\text{UV}} = k - 1$$

$$df_{\text{Person}} = n - 1$$

$$df_{\text{res}} = (k - 1)(n - 1)$$

Hinweis: diese werden auch oft als Mean Squares (MS) bezeichnet.

#### Quadratsummen

$$QS_{\text{ges}} = \sum_{j=1}^k \sum_{i=1}^n (x_{ji} - \bar{x})^2$$

$$QS_{\text{UV}} = \sum_{j=1}^k n \cdot (\bar{x}_j - \bar{x})^2$$

$$QS_{\text{Person}} = \sum_{i=1}^n k \cdot (\bar{x}_i - \bar{x})^2$$

$$QS_{\text{res}} = QS_{\text{ges}} - QS_{\text{UV}} - QS_{\text{Person}}$$

#### Notation

$k$ : Anzahl der Bedingungen

$n$ : Anzahl der Personen

$N$ : Gesamtzahl der Messwerte

$\bar{x}$ : Gesamtmittelwert

Hinweis: gilt nur bei balancierten Designs (gleiche Stichprobengröße)

### F-Werte

$$\begin{aligned} F_A &= \frac{\hat{\sigma}_A^2}{\hat{\sigma}_{\text{inn}}^2} \\ F_B &= \frac{\hat{\sigma}_B^2}{\hat{\sigma}_{\text{inn}}^2} \\ F_{A \times B} &= \frac{\hat{\sigma}_{A \times B}^2}{\hat{\sigma}_{\text{inn}}^2} \end{aligned}$$

### Populationsvarianzen

$$\begin{aligned} \hat{\sigma}_A^2 &= \frac{QS_A}{df_A} \\ \hat{\sigma}_B^2 &= \frac{QS_B}{df_B} \\ \hat{\sigma}_{A \times B}^2 &= \frac{QS_{A \times B}}{df_{A \times B}} \\ \hat{\sigma}_{\text{inn}}^2 &= \frac{QS_{\text{inn}}}{df_{\text{inn}}} \end{aligned}$$

### Freiheitsgrade

$$\begin{aligned} df_{\text{inn}} &= \sum_j^k \sum_l^m (n_{jl} - 1) = N - km \\ df_A &= k - 1 \\ df_B &= m - 1 \\ df_{A \times B} &= (k - 1)(m - 1) \end{aligned}$$

### Quadratsummen

$$QS_{\text{ges}} = QS_A + QS_B + QS_{A \times B} + QS_{\text{inn}}$$

$$QS_{\text{ges}} = \sum_j^k \sum_l^m \sum_i^n (x_{ijl} - \bar{x})^2$$

$$QS_{\text{inn}} = \sum_j^k \sum_l^m \sum_i^n (x_{ijl} - \bar{x}_{jl})^2$$

$$QS_A = n \cdot m \sum_j^k (\bar{x}_{j \cdot} - \bar{\bar{x}})^2$$

$$QS_B = n \cdot k \sum_l^m (\bar{x}_{\cdot l} - \bar{\bar{x}})^2$$

$$QS_{A \times B} = n \sum_{j=1}^k \sum_{l=1}^m (\bar{x}_{jl} - \bar{x}_{j \cdot} - \bar{x}_{\cdot l} + \bar{\bar{x}})^2$$

$$QS_{A \times B} = QS_{\text{ges}} - QS_A - QS_B - QS_{\text{inn}}$$

### Notation

- N: Gesamtzahl der Messwerte der Untersuchung
- n: Anzahl der Messwerte pro Bedingung
- k: Anzahl der Stufen des Faktors A
- m: Anzahl der Stufen des Faktors B
- $\bar{x}$ : Gesamtmittelwert
- $\bar{x}_{j \cdot}$ : Mittelwert für Faktor A, Bedingung j
- $\bar{x}_{\cdot l}$ : Mittelwert für Faktor B, Bedingung l

## Kontrastanalyse US

Voraussetzung:  $\sum_i^k \lambda_i = 0$

$$F_{\text{Kontrast}} = \frac{\hat{\sigma}_{\text{Kontast}}^2}{\hat{\sigma}_{\text{inn}}^2}$$

oder

$$F = t^2$$

$$|t_{\text{Kontrast}}| = \sqrt{F_{\text{Kontrast}}}$$

### Quadratsummen

$$QS_{\text{Kontrast}} = \frac{\left( \sum_{i=1}^k \lambda_i \bar{x}_i \right)^2}{\sum_{i=1}^k \frac{\lambda_i^2}{n_i}}$$

### Freiheitsgrade

### Geschätzte Populationsvarianz

$$\begin{aligned} \hat{\sigma}_{\text{Kontast}}^2 &= \frac{QS_{\text{Kontast}}}{df_{\text{Kontast}}} \\ \hat{\sigma}_{\text{inn}}^2 &= \frac{QS_{\text{inn}}}{df_{\text{inn}}} \end{aligned}$$

Siehe auch ANOVA US.

$$df_{\text{Kontast}} = 1$$

$$df_{\text{inn}} = N - k$$

### Notation

- N: Gesamtzahl der Messwerte
- i: Bedingung
- k: Anzahl der Bedingungen (Gruppen)
- $n_i$ : Anzahl der Messwerte pro Bedingung i

Hinweis: Gilt für den Fall ohne Subgruppen.

Voraussetzung:  $\sum_j^k \lambda_j = 0$

$$t_{\text{Kontrast}} = \frac{\bar{L}}{\sqrt{\frac{\hat{\sigma}_L^2}{n}}}$$

$$L_i = \sum_{j=1}^k (x_{ij} \cdot \lambda_j)$$

$$\hat{\sigma}_L^2 = \frac{1}{n-1} \sum_{i=1}^n (L_i - \bar{L})^2$$

$$df_{\text{Kontrast}} = n - 1$$

### Notation

$\bar{L}$ : Mittelwert L-Werte

$\hat{\sigma}_L^2$ : geschätzte Populationsvarianz der L-Werte

$i$ : Objekt/Person

$n$ : Anzahl der Objekte/Personen, für die mehrere Messungen durchgeführt wurden

$j$ : Bedingung

$k$ : Anzahl Bedingungen

## Nonparametrische Verfahren

### $\chi^2$ -Anpassungstests

für eine Variable

$$\chi^2 = \sum_i^k \frac{(f_{b,i} - f_{e,i})^2}{f_{e,i}}$$

$$f_{e,i} = N \cdot P_i$$

$$df = k - 1$$

für zwei Variablen

$$\chi^2 = \sum_i^k \sum_j^m \frac{(f_{b,ij} - f_{e,ij})^2}{f_{e,ij}}$$

$$f_{e,ij} = N \cdot P_{ij}$$

$$df = k \cdot m - 1$$

### $\chi^2$ -Unabhängigkeitstest

$$\chi^2 = \sum_i^k \sum_j^m \frac{(f_{b,ij} - f_{e,ij})^2}{f_{e,ij}}$$

$$f_{e,ij} = \frac{Z_i \cdot S_j}{N}$$

$$df = (k-1) \cdot (m-1)$$

Bei 2 dichotomen Variablen:  $\chi^2 = \phi^2 \cdot N$

### U-Test nach Mann und Whitney (Wilcoxon Rangsummen-Test)

1. Messwerte insgesamt in Rangreihe bringen und jedem Messwert einen Rangplatz zuweisen

2.  $T_1$ : Summe der Rangplätze der Gruppe 1  $T_2$ : Summe der Rangplätze der Gruppe 2

$$3. U = n_1 \cdot n_2 + \frac{n_1 \cdot (n_1+1)}{2} - T_1$$

$$U' = n_1 \cdot n_2 + \frac{n_2 \cdot (n_2+1)}{2} - T_2 = n_1 \cdot n_2 - U$$

4. Prüfgröße für die bei uns verwendete Tabelle ist der kleinere der beiden U-Werte

### Wilcoxon-Test für AS (Vorzeichenrangtest)

1. Differenzen der Messwertpaare bilden

2. den Beträgen der Differenzen Rangplatz zuweisen (beim kleinsten Betrag mit Rangplatz 1 beginnen)

3.  $T_-$ : Summe der Rangplätze der negativen Differenzen  $T_+$ : Summe der Rangplätze der positiven Differenzen

4. Prüfgröße für die Tabelle ist der kleinere der beiden T-Werte

### Notation

$f_{b,i}$ : beobachtete Häufigkeit

$f_{e,i}$ : erwartete Häufigkeit

$P_i$ : in der Merkmalsausprägung erwarteter Anteil

$N$ : Anzahl der Untersuchungsteilnehmer

$k, m$ : Anzahl der Merkmalsausprägungen beider Merkmale

$Z_i$ : Zeilenhäufigkeit

$S_j$ : Spaltenhäufigkeit

$\phi$ : Phi-Koeffizient

## Faktorenanalyse

Bei  $z$ -standardisierten Variablen, Hauptkomponenten-Analyse und orthogonaler Extraktion gilt:

1. Faktorladung:  $a_{m,f} = r_{m,f}$
2. Eigenwert:  $\lambda_f = \sum_{m=1}^M a_{m,f}^2$
3. Kommunalität:  $h_m^2 = \sum_{f=1}^F a_{m,f}^2$
4. durch Faktor aufgeklärter Varianzanteil:  $\frac{\lambda_f}{M}$

## Notation

- $m$ : Variable
- $M$ : Anzahl der Variablen
- $f$ : Faktor
- $F$ : Anzahl der ausgewählten Faktoren
- $i$ : Person/Objekt
- $r$ : Korrelation

## Clusteranalyse

### Nominalskalierte Variablen

		Fall x	
		+	-
Fall y	+	a	c
	-	b	d

Tanimoto-Koeffizient

$$T = \frac{a}{a + b + c}$$

M-Koeffizient

$$M = \frac{a + d}{a + b + c + d}$$

### Intervallskalierte Variablen

- $a, b$ : Fälle
- $J$ : Anzahl der Variablen (Dimensionen)
- $r$ : Minkowski-Konstante

Minkowski-Metrik

$$d_{a,b} = \left[ \sum_{j=1}^J |x_{aj} - x_{bj}|^r \right]^{\frac{1}{r}}$$

Euklidische Distanz

$$d_{a,b} = \sqrt{\sum_{j=1}^J |X_{aj} - X_{bj}|^2}$$

Manhattan-Distanz / City-Block-Metrik

$$d_{a,b} = \sum_{j=1}^J |X_{aj} - X_{bj}|$$

### Konventionen nach Cohen

”These values are necessarily somewhat arbitrary, but were chosen so as to seem reasonable. The reader can render his own judgment as to their reasonableness.” (Cohen, 1962)

	d/g	r/w/φ	η²
klein	±0.2	±0.1	0.01
mittel	±0.5	±0.3	0.06
groß	±0.8	±0.5	0.14

### t-Test: Einstichprobenfall

$$g = \frac{t}{\sqrt{n}}$$

$$d = \frac{t}{\sqrt{df}}$$

### t-Test: US

$$g = t_{US} \cdot \sqrt{\frac{n_A + n_B}{n_A \cdot n_B}}$$

$$d = t_{US} \cdot \frac{n_A + n_B}{\sqrt{df} \cdot \sqrt{n_A \cdot n_B}}$$

$$r = \sqrt{\frac{t_{US}^2}{t_{US}^2 + df}}$$

Letzteres gilt auch für  $F$ , wenn  $df_{zw}$  1 sind, da in diesem Fall  $F = t^2$ .

Vereinfachung, wenn  $n_A = n_B$ :

$$d = \frac{2 \cdot t_{US}}{\sqrt{df}}$$

$$g = \frac{2 \cdot t_{US}}{\sqrt{n}}$$

### t-Test: AS

$$g = \frac{t_{AS}}{\sqrt{n}}$$

$$d = \frac{t_{AS}}{\sqrt{df}}$$

### Kontrastanalyse US

aus F-Test

$$r_{effectsize} = \sqrt{\frac{F_{Kontast}}{F_{zw} \cdot df_{zw} + df_{inn}}}$$

$$r_{contrast} = \sqrt{\frac{F_{Kontast}}{F_{Kontast} + df_{inn}}} = \sqrt{\frac{t_{Kontast}^2}{t_{Kontast}^2 + df}}$$

### Kontrastanalyse AS

aus t-Test

$$g = \frac{t}{\sqrt{n}}$$

### χ²-Tests

$$w = \sqrt{\sum_{i=1}^k \frac{(P_{b,i} - P_{e,i})^2}{P_{e,i}}}$$

aus Ergebnis χ²-Test:

$$w = \sqrt{\frac{\chi^2}{N}}$$

aus 2 dichotomen Variablen

$$w = \sqrt{\frac{\chi^2}{N}} = \phi$$

### ANOVA

US

$$\eta^2 = \frac{F \cdot df_{zw}}{F \cdot df_{zw} + df_{inn}}$$

$$\eta^2 = \frac{QS_{zw}}{QS_{ges}}$$

AS

$$\eta^2 = \frac{QS_{UV}}{QS_{ges}}$$

$$\eta_p^2 = \frac{QS_{UV}}{QS_{UV} + QS_{res}}$$

mehr faktoriell

$$\eta^2 = \frac{QS_{Effekt}}{QS_{ges}}$$

$$\eta_p^2 = \frac{QS_{Effekt}}{QS_{Effekt} + QS_{inn}}$$

## Effektgrößen aus Rohwerten

$$d = \frac{\bar{x} - c}{s}$$

$$d = \frac{\bar{x}_A - \bar{x}_B}{s_{AB}}$$

$$g = \frac{\bar{x} - c}{\hat{\sigma}}$$

$$g = \frac{\bar{x}_A - \bar{x}_B}{\hat{\sigma}_{AB}}$$

$$s_{AB} = \sqrt{\frac{s_A^2 n_A + s_B^2 n_B}{n_A + n_B}}$$

$$\hat{s}_{AB} = \sqrt{\frac{(n_A - 1) \cdot \hat{\sigma}_A^2 + (n_B - 1) \cdot \hat{\sigma}_B^2}{n_A + n_B - 2}}$$

$$r_{\text{effectsize}} = \frac{\frac{1}{n} \sum_i^n (x_i - \bar{x}) \cdot (\lambda_i - \bar{\lambda})}{s_x \cdot s_\lambda}$$

$$r_{\text{contrast}} = \frac{QS_{\text{Kontast}}}{QS_{\text{Kontast}} + QS_{\text{Nicht-Kontast}}}$$

$$r_{\text{alerting}}^2 = \frac{QS_{\text{Kontast}}}{QS_{\text{zw}}}$$

$$g = \frac{\bar{L}}{\hat{\sigma}_L}$$

## Odds Ratio (OR)

$$OR = \frac{\frac{a}{b}}{\frac{c}{d}} = \frac{a \cdot d}{b \cdot c}$$

	Risikofaktor	ohne Risikofaktor
Krankheit	a	b
Keine Krankheit	c	d

## Effektgrößen aus Effektgrößen

### Abstandsmaße aus Abstandsmaßen

$$d = g \cdot \sqrt{\frac{n}{df}}$$

$$g = d \cdot \sqrt{\frac{df}{n}}$$

### Korrelationen aus Abstandsmaßen

$$r = \frac{d}{\sqrt{d^2 + \frac{1}{p \cdot q}}}$$

$$r = \sqrt{\frac{g^2(n_A \cdot n_B)}{g^2(n_A \cdot n_B) + (n_A + n_B)df}}$$

### Abstandsmaße aus Korrelationen

$$d = \frac{r}{\sqrt{1 - r^2}} \cdot \sqrt{\frac{1}{p \cdot q}}$$

$$g = \frac{r}{\sqrt{1 - r^2}} \cdot \sqrt{\frac{(n_A + n_B) \cdot df}{n_A \cdot n_B}}$$

wobei

$$p = \frac{n_A}{n_A + n_B}$$

und

$$q = \frac{n_B}{n_A + n_B}$$

## Meta-Analyse

Mittlere, an der Stichprobe gewichtete, Effektgröße:  $\bar{r} = \frac{\sum_{i=1}^n (N_i r_i)}{\sum_{i=1}^n N_i}$

### Bedingte Wahrscheinlichkeit

Die bedingte Wahrscheinlichkeit von  $A$  unter der Bedingung  $B$  wird notiert als  $P(A|B)$  und berechnet sich durch:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Hierbei ist  $P(A \cap B)$  die Wahrscheinlichkeit, dass sowohl  $A$  als auch  $B$  eintreten.

### Produktregel

Die Produktregel beschreibt die Wahrscheinlichkeit des gemeinsamen Eintretens von zwei Ereignissen  $A$  und  $B$ :

$$P(A \cap B) = P(A|B) \cdot P(B)$$

Alternativ kann die Produktregel auch in umgekehrter Reihenfolge geschrieben werden:

$$P(A \cap B) = P(B|A) \cdot P(A)$$

### Summenregel

Die Wahrscheinlichkeit, dass mindestens eines der Ereignisse  $A$  oder  $B$  eintritt:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

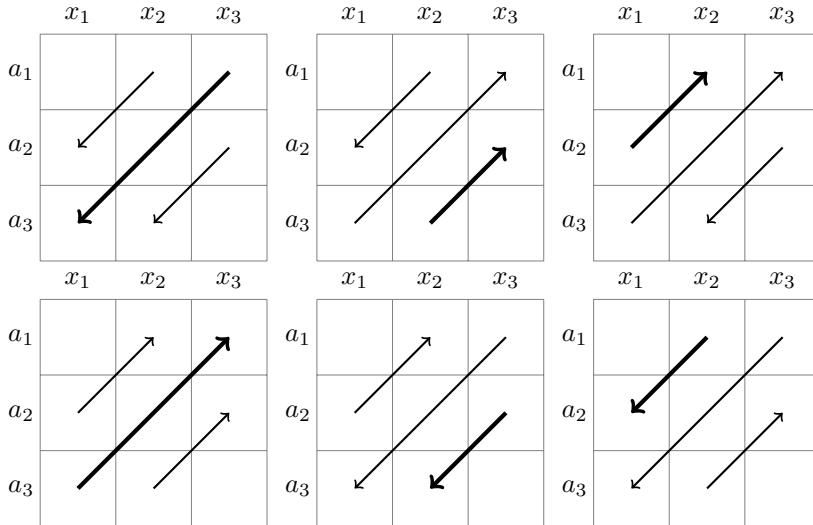
Dies ist ein Spezialfall des allgemeinen Ein-/Ausschlussprinzips.

### Wahrscheinlichkeitsrevision (Bayes)

$$P(A|B) = \frac{P(A) \cdot P(B|A)}{P(A) \cdot P(B|A) + P(\neg A) \cdot P(B|\neg A)}$$

**Conjoint-Measurement-Theory**

Messaxiomie fürs Intervallskalenniveau sind teilweise prüfbar über ein Conjoint-Measurement-Experiment. Hierbei geht man davon aus, dass  $P = f(A, X)$  wobei  $f$  eine non-interaktive Funktion ist. Geprüft werden muss die einfache Aufhebung (single cancellation) und die doppelte Aufhebung (double cancellation). Die einfache Aufhebung besagt, dass alle Zeilen und Spalten der Conjoint-Measurement-Matrix die gleiche Ordnung haben müssen. Die Doppelaufhebung restriktiert die möglichen Ordnungen der Matrix. Wenn alle Spalten und Zeilen bereits geordnet sind, lässt sich über folgende Visualisierung die Doppelaufhebung prüfen. Hierbei entsprechen die Pfeile, dem  $\leq$ -Zeichen. Dünne Pfeile zwischen den Zellen repräsentieren die Antezedenz-Ordnungsbeziehungen, und dicke Pfeile repräsentieren die Konsequenz-Ordnungsbeziehung.



Funktion	Verteilung	Parameter
dnorm	Normalverteilung	x, mean, sd
pnorm	Normalverteilung	q, mean, sd, lower.tail, log.p
qnorm	Normalverteilung	p, mean, sd, lower.tail, log.p
rnorm	Normalverteilung	n, mean, sd
dbinom	Binomialverteilung	x, size, prob
pbinom	Binomialverteilung	q, size, prob, lower.tail, log.p
qbinom	Binomialverteilung	p, size, prob, lower.tail, log.p
rbinom	Binomialverteilung	n, size, prob
dt	t-Verteilung	x, df, ncp
pt	t-Verteilung	q, df, ncp, lower.tail, log.p
qt	t-Verteilung	p, df, ncp, lower.tail, log.p
rt	t-Verteilung	n, df, ncp
df	F-Verteilung	x, df1, df2, ncp
pf	F-Verteilung	q, df1, df2, ncp, lower.tail, log.p
qf	F-Verteilung	p, df1, df2, ncp, lower.tail, log.p
rf	F-Verteilung	n, df1, df2, ncp

#### Präfix Erklärung

- d Dichtefunktion: Berechnet die Wahrscheinlichkeitsdichte (z. B.  $f(x)$ )
- p Verteilungsfunktion: Berechnet die kumulierte Wahrscheinlichkeit (z. B.  $P(X \leq x)$ )
- q Quantilfunktion: Berechnet den Wert  $x$ , sodass  $P(X \leq x) = p$
- r Zufallszahlengenerator: Generiert Zufallswerte aus der Verteilung

#### Parameter Erklärung

- x Werte, an denen die Dichte berechnet wird
- q Quantile
- p Wahrscheinlichkeiten
- n Anzahl der zu generierenden Zufallswerte
- mean Mittelwert der Verteilung
- sd Standardabweichung
- size Anzahl der Versuche
- prob Erfolgswahrscheinlichkeit
- df Freiheitsgrade
- df1 Freiheitsgrade des Zählers (F-Verteilung)
- df2 Freiheitsgrade des Nenners (F-Verteilung)
- ncp Nichtzentralitätsparameter (optional)
- lower.tail Logisch; wenn TRUE (Standard), werden Wahrscheinlichkeiten  $P[X \leq x]$  berechnet
- log.p Logisch; wenn TRUE, werden Wahrscheinlichkeiten als Log-Werte ausgegeben

#### Literatur

Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *The Journal of Abnormal and Social Psychology*, 65(3), 146. <https://doi.org/10.1037/h0045186>

Sedlmeier, P., & Renkewitz, F. (2018). Forschungsmethoden und Statistik für Psychologen und Sozialwissenschaftler: Pearson