

Zur Berechnung von Kappa (nach Cohen für 2 Rater/innen)

Formel:

$$\text{Kappa} = \frac{P_o - P_e}{1 - P_e} = \frac{\text{Anteil übereinstimmender Urteile} - \text{Anteil zufälliger Übereinstimmungen}}{1 - \text{Anteil zufälliger Übereinstimmungen}}$$

Beispiel 1:

Übereinstimmung von Rater/in 1 und Rater/in 5 hinsichtlich des Kategorisierungsaspekts „Stereotype in der Berichterstattung“; Anzahl der Kategorien ist 5 und Gesamtanzahl der Analyseeinheiten ist 21

Kreuztabelle (Rohwerte)

		Rater/in 5					Gesamt
		Kat. 1	Kat. 2	Kat. 3	Kat. 4	Kat. 5	
Rater/in 1	Kat. 1	1	0	0	0	0	1
	Kat. 2	0	3	1	0	0	4
	Kat. 3	0	0	3	0	0	3
	Kat. 4	0	0	0	1	0	1
	Kat. 5	0	0	3	0	9	12
Gesamt		1	3	7	1	9	Σ 21

→ **Diagonale** = übereinstimmende Urteile/ Zuordnungen; alle Werte außerhalb der Diagonalen sind nicht-übereinstimmende Urteile/ Zuordnungen;

Beispiel für **Übereinstimmung**: **orangene 3**: Rater/in 5 und 1 haben 3 Analyseeinheiten übereinstimmend der Kategorie 2 zugeordnet.

Beispiel für **Nicht-Übereinstimmung**: **grüne 1**: Rater/in 5 hat eine Analyseeinheit Kategorie 3 zugeordnet, die von Rater/in 1 Kategorie 2 subsumiert wurde.

→ Der **Anteil (beobachteter) übereinstimmender Urteile** (P_o für observed) lässt sich anhand der Tabelle einfach ablesen:

17 Urteile wurden übereinstimmend abgegeben (Summe der Werte in der Diagonalen ($1+3+3+1+9=17$)), 21 Urteile sind es insgesamt → $P_o = 17/21 = 0,81$

Etwas umständlicher ist das Ganze bei Caspar & Wirtz (2001) (S. 56ff.) dargestellt, die zunächst jeden einzelnen Wert an der Gesamtanzahl relativieren und dann die Summe aller Einzelanteile bilden:

Kreuztabelle (Anteilswerte)

		Rater/in 5					Gesamt
		Kat. 1	Kat. 2	Kat. 3	Kat. 4	Kat. 5	
Rater/in 1	Kat. 1	1/21=0,05	0	0	0	0	1/21=0,05
	Kat. 2	0	3/21=0,14	1/21=0,05	0	0	4/21=0,19
	Kat. 3	0	0	3/21=0,14	0	0	3/21=0,14
	Kat. 4	0	0	0	1/21=0,05	0	1/21=0,05
	Kat. 5	0	0	3/21=0,14	0	9/21=0,43	12/21=0,57
Gesamt		1/21=0,05	3/21=0,14	7/21=0,33	1/21=0,05	9/21=0,43	Σ 1,00

Berechnung des Anteils der (beobachteten) übereinstimmenden Urteile:

$$P_o = 1/21 + 3/21 + 3/21 + 1/21 + 9/21 = 0,05 + 0,14 + 0,14 + 0,05 + 0,43 = 0,81$$

→ Wir erhalten – auf beiden Wegen – eine Übereinstimmung von 0,81, d.h. in 81% der Fälle haben die Kodierer/innen das gleiche Urteil abgegeben. Dieser Wert soll nun noch um den **Anteil zufälliger Übereinstimmungen** korrigiert werden.

In die **Berechnung zufälliger Übereinstimmungen** (P_e für expected) gehen die empirischen Daten, genauer gesagt die (Zeilen- und Spalten-)Randsummen (in Anteilswerten; siehe vorherige Kreuztabelle, **blau markierte Werte**) ein. Hierzu werden die Anteilswerte der Zeilensummen mit den Anteilswerten der Spaltensummen multipliziert. Da hier nur die Werte in der Diagonalen interessieren, sind die anderen Werte nicht eingetragen worden.

		Rater/in 5				
		Kat. 1	Kat. 2	Kat. 3	Kat. 4	Kat. 5
Rater/in 1	Kat. 1	0,05x0,05
	Kat. 2	...	0,14x0,19
	Kat. 3	0,33x0,14
	Kat. 4	0,05x0,05	...
	Kat. 5	0,43x0,57

Berechnung des Anteils der zufällig übereinstimmenden Urteile:

$$P_e = (0,05 \times 0,05) + (0,14 \times 0,19) + (0,33 \times 0,14) + (0,05 \times 0,05) + (0,43 \times 0,57) = 0,0025 + 0,0264 + 0,0462 + 0,0025 + 0,2451 = 0,3227$$

→ **man beachte:** die Kategorie 5 (Restkategorie) ist mit 9 Analyseeinheiten (vgl. 1. Kreuztabelle) verhältnismäßig häufig besetzt; daher ist auch die erwartete zufällige Übereinstimmung mit 24,51% bei dieser Kategorie relativ hoch und macht ca. 75% der gesamten erwarteten zufälligen Übereinstimmung aus, die 32,27% beträgt!!!

Nun haben wir alle Werte, die wir für die Berechnung von Kappa benötigen und müssen nur noch in die Formel einsetzen:

Berechnung von Kappa:

$$\text{Kappa} = \frac{P_o - P_e}{1 - P_e} = \frac{0,81 - 0,3227}{1 - 0,3227} = 0,719$$

Vergleich mit dem von **SPSS** ausgegebenen Wert:

Symmetrische Maße

	Wert	Asymptotischer Standardfehler(a)	Näherungsweise T(b)	Näherungsweise Signifikanz
Maß der Übereinstimmung Kappa	,718	,124	5,698	,000
Anzahl der gültigen Fälle	21			

a Die Null-Hypothese wird nicht angenommen.

b Unter Annahme der Null-Hypothese wird der asymptotische Standardfehler verwendet.

Kappa beträgt also 0,719 bzw. die zufallskorrigierte Übereinstimmung beträgt 71,9%!!!

Exkurs:

Veränderung von Kappa für den Fall, dass sich die Häufigkeiten in den Kategorien etwa gleich verteilt hätten (im Vergleich zum 1. Beispiel, in dem die Kat. 5 vgl.s.weise stark und alle anderen vgl.s.weise schwach besetzt waren).

		Rater/in 5					Gesamt
		Kat. 1	Kat. 2	Kat. 3	Kat. 4	Kat. 5	
Rater/in 1	Kat. 1	3	0	0	0	0	3
	Kat. 2	0	3	1	0	0	4
	Kat. 3	0	0	4	0	0	4
	Kat. 4	0	0	0	4	0	4
	Kat. 5	0	0	3	0	3	6
Gesamt		3	3	8	4	3	Σ 21

$$P_o = 17/21 = 0,81$$

		Rater/in 5					Gesamt
		Kat. 1	Kat. 2	Kat. 3	Kat. 4	Kat. 5	
Rater/in 1	Kat. 1	0,14	0	0	0	0	3/21=0,14
	Kat. 2	0	0,14	0,05	0	0	4/21=0,19
	Kat. 3	0	0	0,19	0	0	4/21=0,19
	Kat. 4	0	0	0	0,19	0	4/21=0,19
	Kat. 5	0	0	0,14	0	0,14	6/21=0,29
Gesamt		3/21=0,14	3/21=0,14	8/21=0,38	4/21=0,19	3/21=0,14	Σ 1,00

		Rater/in 5				
		Kat. 1	Kat. 2	Kat. 3	Kat. 4	Kat. 5
Rater/in 1	Kat. 1	0,14x0,14
	Kat. 2	...	0,14x0,19
	Kat. 3	0,38x0,19
	Kat. 4	0,19x0,19	...
	Kat. 5	0,14x0,29

$P_e = 0,02 + 0,03 + 0,07 + 0,04 + 0,04 = 0,2 \rightarrow$ deutlich kleiner als bei ungleicher Häufigkeitsverteilung (siehe Beispiel 1)

$$\text{Kappa} = \frac{P_o - P_e}{1 - P_e} = \frac{0,81 - 0,2}{1 - 0,2} = 0,762 \rightarrow \text{somit auch etwas höher!}$$

Beispiel 2:

Übereinstimmung von Rater/in 1 und Rater/in 3 hinsichtlich des Kategorisierungsaspekts „Stil der Berichterstattung“; Anzahl der Kategorien ist diesmal 2 und Gesamtanzahl der Analyseeinheiten ist 5

Die Berechnungen erfolgen in Analogie zu dem obigen Beispiel und werden daher nicht näher kommentiert.

Kreuztabelle

		Rater/in 3		Gesamt
		Kat. 1	Kat. 2	
Rater/in 1	Kat. 1	3	1	4
	Kat. 2	0	1	1
Gesamt		3	2	Σ 5

Berechnung des Anteils der (beobachteten) übereinstimmenden Urteile:

$$P_o = 4/5 = 0,8$$

alternativ (siehe Caspar & Wirtz, 2001):

		Rater/in 3		Gesamt
		Kat. 1	Kat. 2	
Rater/in 1	Kat. 1	3/5=0,6	1/5=0,2	4/5=0,8
	Kat. 2	0	1/5=0,2	1/5=0,2
Gesamt		3/5=0,6	2/5=0,4	Σ 1,00

$$P_o = 3/5 + 1/5 = 0,6 + 0,2 = 0,8$$

→ Die beobachtete Übereinstimmung ist also von gleicher Höhe, wie beim vorherigen Beispiel.

		Rater/in 3	
		Kat. 1	Kat. 2
Rater/in 1	Kat. 1	0,6x0,8	...
	Kat. 2	...	0,4x0,2

Berechnung des Anteils der zufällig übereinstimmenden Urteile:

$$P_e = (0,6 \times 0,8) + (0,4 \times 0,2) = 0,48 + 0,08 = 0,56$$

→ dieser Wert ist viel höher als der aus dem vorherigen Beispiel, da der Zufall sich bei „nur“ zwei Kategorien stärker auswirken kann!!!

→ daher wird auch Kappa entsprechend kleiner als im vorherigen Beispiel:

Berechnung von Kappa:

$$\text{Kappa} = \frac{P_o - P_e}{1 - P_e} = \frac{0,8 - 0,56}{1 - 0,56} = 0,54$$

Symmetrische Maße

	Wert	Asymptotischer Standardfehler(a)	Näherungsweise T(b)	Näherungsweise Signifikanz
Maß der Übereinstimmung Kappa	,545	,362	1,369	,171
Anzahl der gültigen Fälle	5			

a Die Null-Hypothese wird nicht angenommen.

b Unter Annahme der Null-Hypothese wird der asymptotische Standardfehler verwendet.

Literatur:

Bortz, J. & Lienert, G.A. (2003). *Kurzgefasste Statistik für die Klinische Forschung. Leitfaden für die verteilungsfreie Analyse kleiner Stichproben*. Heidelberg: Springer.

Wirtz, M. & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität*. Göttingen: Hogrefe.