

# Berechnung von Fleiss' Kappa und Krippendorff's Alpha für die erste Version des Fragebogens

## 1 Fleiss' Kappa

### 1.1 Daten aus den Fragebögen:

Frage	Ja	Nein
1	63	3
2	62	4
3	62	4
4	49	17
5	51	15
6	46	20

### 1.2 Formel der Item(Frage)-spezifischen Übereinstimmung ( $P_i$ )

Die Formel zur Berechnung von  $P_F$  lautet:

$$P_i = \frac{1}{n(n-1)} \sum_{k=1}^K n_{ik}(n_{ik} - 1) \quad (1)$$

Dabei ist  $n$  die Anzahl der Annotatoren (hier 3 [Kevin, Kubi, Mengjiao]), und  $n_{ik}$  die Anzahl der Annotatoren, die die Kategorie  $k$  gewählt haben und  $N$  für die Anzahl der Fragen.

### 1.3 Formel der durchschnittlichen beobachteten Übereinstimmung ( $\bar{P}$ )

Die durchschnittliche Übereinstimmung ergibt sich aus dem Mittelwert aller  $P_i$ -Werte:

$$\bar{P} = \frac{\sum_{i=1}^N P_i}{N} \quad (2)$$

### 1.4 Formel der erwarteten Übereinstimmung ( $P_e$ )

Die erwartete Übereinstimmung ergibt sich aus:

$$P_e = p_{ja}^2 + p_{nein}^2 \quad (3)$$

Dabei sind  $p_{ja}$  und  $p_{nein}$  die Anteile der jeweiligen Kategorien an der Gesamtzahl der Bewertungen.

## 1.5 Berechnung von Fleiss' Kappa

Die Formel zur Berechnung von Fleiss' Kappa lautet:

$$\kappa = \frac{\bar{P} - P_e}{1 - P_e} \quad (4)$$

Berechnung von:  $\bar{P}$

Hierbei wird durch 22 geteilt, da es 22 Durchläufe (22 Fragebögen) sind

$$P_1 = \frac{1}{6} \left[ \frac{63}{22} \times \left( \frac{63}{22} - 1 \right) + \frac{3}{22} \times \left( \frac{3}{22} - 1 \right) \right] = 0,87 \quad (5)$$

$$P_2 = \frac{1}{6} \left[ \frac{62}{22} \times \left( \frac{62}{22} - 1 \right) + \frac{4}{22} \times \left( \frac{4}{22} - 1 \right) \right] = 0,83 \quad (6)$$

$$P_3 = \frac{1}{6} \left[ \frac{62}{22} \times \left( \frac{62}{22} - 1 \right) + \frac{4}{22} \times \left( \frac{4}{22} - 1 \right) \right] = 0,83 \quad (7)$$

$$P_4 = \frac{1}{6} \left[ \frac{49}{22} \times \left( \frac{49}{22} - 1 \right) + \frac{17}{22} \times \left( \frac{17}{22} - 1 \right) \right] = 0,426 \quad (8)$$

$$P_5 = \frac{1}{6} \left[ \frac{51}{22} \times \left( \frac{51}{22} - 1 \right) + \frac{15}{22} \times \left( \frac{15}{22} - 1 \right) \right] = 0,473 \quad (9)$$

$$P_6 = \frac{1}{6} \left[ \frac{46}{22} \times \left( \frac{46}{22} - 1 \right) + \frac{20}{22} \times \left( \frac{20}{22} - 1 \right) \right] = 0,366 \quad (10)$$

$$\bar{P} = \frac{P_1 + P_2 + P_3 + P_4 + P_5 + P_6}{N} \quad (11)$$

Einsetzen der Werte ergibt:

$$\bar{P} = \frac{0,87 + 0,83 + 0,83 + 0,426 + 0,473 + 0,366}{6} = 0,6323 \quad (12)$$

Berechnung der durchschnittlichen beobachteten Übereinstimmung  $\bar{P}$ : Die Berechnung der Wahrscheinlichkeiten  $p_{ja}$  und  $p_{nein}$  erfolgt folgendermaßen:

$$p_{ja} = \frac{\sum \text{Ja}}{N \times n}, \quad p_{nein} = \frac{\sum \text{Nein}}{N \times n}$$

Dabei gilt:

$$p_{ja} = \frac{\sum \text{Ja}}{22 \times 3 \times 6}, \quad p_{nein} = \frac{\sum \text{Nein}}{22 \times 3 \times 6}$$

Wobei:

- $N = 6$  (Anzahl der Items (Fragen))
- $n = 3$  (Anzahl der Annotatoren pro Item)
- $22 =$  Anzahl der Durchläufe pro Item bzw. 22 Fragebögen

Berechnung der Gesamtsummen für Ja und Nein:

Die Gesamtsummen der Ja- und Nein-Antworten ergeben sich aus der Summe der jeweiligen Werte für jedes Item:

Gesamtanzahl der Ja-Antworten

$$\sum \text{Ja} = 63 + 62 + 62 + 49 + 51 + 46 = 333$$

Gesamtanzahl der Nein-Antworten

$$\sum \text{Nein} = 3 + 4 + 4 + 17 + 15 + 20 = 63$$

Gesamtanzahl aller Antworten Die Gesamtanzahl aller Antworten ergibt sich aus der Summe der Ja- und Nein-Antworten:

$$\sum \text{Gesamt} = \sum \text{Ja} + \sum \text{Nein}$$

$$p_{ja} = \frac{\sum \text{Ja}}{N \times n \times 22}, \quad p_{nein} = \frac{\sum \text{Nein}}{N \times n \times 22}$$

Setzt man die Werte ein:

$$p_{ja} = \frac{333}{396} = 0,8409, \quad p_{nein} = \frac{63}{396} = 0,1591$$

Nun werden die Wahrscheinlichkeiten quadriert und addiert:

$$P_e = \left(\frac{333}{396}\right)^2 + \left(\frac{63}{396}\right)^2$$

Das ergibt:

$$P_e = 0.7071 + 0.0253$$

Somit:

$$P_e = 0.7324$$

Einsetzen der Werte ergibt:

$$\bar{P} = 0.6323, \quad P_e = 0.7324 \quad (13)$$

$$\kappa = \frac{0.6323 - 0.7324}{1 - 0.7324} = -0.3741 \quad (14)$$

Dieser negative Wert deutet darauf hin, dass wir Annotatoren schlechter als zufällig übereinstimmen. Mögliche Ursachen könnten sein:

- Eine ungleiche Verteilung der Antworten, da es sehr viele Ja-Antworten und wenige Nein-Antworten gab.
- Systematische Unterschiede im Bewertungsverhalten der Annotatoren aufgrund von Interpretationsschwierigkeiten bei den letzten drei Fragen im Fragebogen.

Aus diesem Grund haben wir die Berechnung zusätzlich mit **Krippendorff's Alpha** durchgeführt, da die Formel weniger sensibel für unausgeglichene Kategorien ist. In unserem Fall sind mehr "Ja" -Antworten als "Nein" -Antworten sehr wahrscheinlich und auch so gewollt.

## 2 Krippendorff's Alpha

2.1 Berechnung der beobachteten Diskrepanz  $D_o$

2.2 Berechnung der erwarteten Diskrepanz  $D_e$

2.3 Berechnung von Krippendorff's Alpha

—

### 2.1 Berechnung der beobachteten Diskrepanz $D_o$

Die beobachtete Diskrepanz  $D_o$  misst die tatsächliche Unstimmigkeit zwischen den Bewertungen der Annotatoren. Die Formel lautet:

$$D_o = \frac{1}{N} \sum_{i=1}^N \sum_{j,k} (x_{ij} - x_{ik})^2$$

-  $N$ : Gesamtanzahl der Bewertungen -  $x_{ij}, x_{ik}$ : Die Bewertungen der einzelnen Annotatoren für jedes Item

Da es sich um binäre Daten (Ja = 1, Nein = 0) handelt, gilt:

$$(1 - 0)^2 = 1, \quad (1 - 1)^2 = 0, \quad (0 - 0)^2 = 0$$

Das bedeutet, die beobachtete Diskrepanz entspricht der Anzahl der ungleichen Paare geteilt durch die Gesamtzahl der Paare.

—

#### Berechnung für jedes Item:

Die Anzahl der Paare pro Item ergibt sich aus:

$$\text{Paare pro Item} = \frac{66 \times (66 - 1)}{2} = 2145$$

Für jedes Item gilt:

$$\begin{aligned} D_{o1} &= \frac{63 \times 3}{2145} = 0.0881 \\ D_{o2} &= \frac{62 \times 4}{2145} = 0.1156 \\ D_{o3} &= \frac{62 \times 4}{2145} = 0.1156 \\ D_{o4} &= \frac{49 \times 17}{2145} = 0.3882 \\ D_{o5} &= \frac{51 \times 15}{2145} = 0.3567 \\ D_{o6} &= \frac{46 \times 20}{2145} = 0.4283 \end{aligned}$$

—

Durchschnittliche beobachtete Diskrepanz:

Nun wird der Durchschnitt aller Werte berechnet:

$$D_o = \frac{0.0881 + 0.1156 + 0.1156 + 0.3882 + 0.3567 + 0.4283}{6}$$

$$D_o = \frac{1.4925}{6} = 0.2489$$

—

## 2.2 Berechnung der erwarteten Diskrepanz $D_e$

Die erwartete Diskrepanz berücksichtigt die Wahrscheinlichkeit, dass zwei zufällig gewählte Annotatoren unterschiedliche Bewertungen abgeben. Die Formel lautet:

$$D_e = \sum_c p_c(1 - p_c)$$

-  $p_c$ : Anteil der jeweiligen Kategorie (Ja oder Nein)

Die Gesamtanzahl aller Antworten beträgt:

$$6 \times 66 = 396$$

Gesamtanzahl der Ja- und Nein-Antworten:

$$\sum \text{Ja} = 333 \quad \sum \text{Nein} = 63$$

Anteile:

$$p_{ja} = \frac{333}{396} = 0.8409, \quad p_{nein} = \frac{63}{396} = 0.1591$$

Berechnung:

$$D_e = 0.8409 \times (1 - 0.8409) + 0.1591 \times (1 - 0.1591)$$

$$D_e = 0.8409 \times 0.1591 + 0.1591 \times 0.8409$$

$$D_e = 0.1339 + 0.1339 = 0.2676$$

—

## 2.3 Berechnung von Krippendorff's Alpha

Nun setzen wir die Werte in die Hauptformel ein:

$$\alpha = 1 - \frac{D_o}{D_e}$$

$$\alpha = 1 - \frac{0.2489}{0.2676}$$

$$\alpha = 1 - 0.9302 = 0.0698$$

Der Alpha Wert ist sehr niedrig mit 0,0698.

### 3 Zusammenfassung der Ergebnisse

- Fleiss' Kappa:  $-0.3741$  (Geringe Übereinstimmung, schlechter als zufällig)
- Krippendorff's Alpha:  $0.0698$  (Sehr geringe Übereinstimmung, knapp über Zufall)