Formelsammlung zur beschreibende Statistik

Übersicht: Merkmalsarten und Merkmalsskalen

Merkmal

Art	qualitativ		quantitativ
Skala	Nominalskala	Ordinalskala	metrische Skala
Eigenschaft	Ausprägungen stehen gleichberechtigt nebeneinander	Nur (natürliche) Rangfolgen können angegeben werden	Rangfolge und Differenzen können gebildet werden.
Beispiele	Nationalität Beruf Automarke	Noten Dienstgrade Kleidergröße	Euro - Beträge Temperatur Körpergröße
Erhebung durch	Feststellen	Vergleichen	Messen, Zählen

Arithmetisches Mittel einer Datenreihe

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\left(x_1 + x_2 + \ldots + x_n\right)$$

n: Anzahl der Beobachtungswerte x_i x_i: i - ter Beobachtungswert $i \in \mathbb{N}$

Berechnung des arithmetischen Mittels aus einer Häufigkeitstabelle

Fall I: Absolute Häufigkeit n_i

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{j} x_i \cdot n_i = \frac{1}{n}\left(x_1 \cdot n_1 + x_2 \cdot n_2 + \ldots. + x_j \cdot n_j\right) \qquad n = \sum_{i=1}^{j} n_i = n_1 + n_2 + \ldots + n_j$$

Fall II: Relative Häufigkeit $h_i = \frac{n_i}{n}$

$$\bar{x} = \sum_{i=1}^{j} x_i \cdot h_i = \left(x_1 \cdot h_1 + x_2 \cdot h_2 + \ldots. + x_j \cdot h_j\right)$$

n_i: absolute Häufigkeit der Merkmalsausprägung x_i n: Summe der absoluten Häufigkeiten

h_i: relative Häufigkeit der Merkmalsausprägung x_i j: Anzahl der Merkmalsausprägungen x_i

Berechnung des arithmetischen Mittels bei klassierten Daten

Fall I: Absolute Häufigkeit n_i

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} m_i \cdot n_i = \frac{1}{n}\left(m_1 \cdot n_1 + m_2 \cdot n_2 + \ldots + m_k \cdot n_k\right) \qquad n = \sum_{i=1}^{k} n_i = n_1 + n_2 + \ldots + n_k$$

Fall II: Relative Häufigkeit $h_i = \frac{n_i}{n}$

$$\bar{x} = \sum_{i=1}^{k} m_i \cdot h_i = m_1 \cdot h_1 + m_2 \cdot h_2 + \ldots + m_k \cdot h_k$$

n_i : absolute Häufigkeit der i - ten Klasse n : Summe der absoluten Häufigkeiten

h_i : relative Häufigkeit der i - ten Klasse k : Anzahl der Klassen

m_i : Klassenmitte der i -ten Klasse

Allgemeine Rechenvorschrift zur Berechnung des Medians

Ist n ist die Anzahl der Beobachtungswerte x_i , dann gilt:

n ungerade $\Rightarrow x_{Med} = x_{\frac{n+1}{2}}$ n gerade $\Rightarrow x_{Med} = \frac{1}{2}\left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}\right)$

Die Spannweite

Spannweite = größter Beobachtungswert - kleinster Beobachtungswert

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

Quartilsabstand

Der mittlere 50% - Bereich aller Beobachtungswerte heißt Quartilsabstand.

Berechnung: $Q_A = Q_3 - Q_1$

Varianz einer Datenreihe

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 = \frac{\left(x_1 - \bar{x}\right)^2 + \left(x_2 - \bar{x}\right)^2 + \ldots + \left(x_n - \bar{x}\right)^2}{n}$$

n: Anzahl der Beobachtungswerte, x_i : i - ter Beobachtungswert, $\bar{x}$: Mittelwert

Standardabweichung

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

Berechnung der Varianz aus einer Häufigkeitstabelle

Fall I: Absolute Häufigkeit n_i

$$n = \sum_{i=1}^{j} n_i = n_1 + n_2 + \ldots + n_j$$

$$s^2 = \frac{1}{n}\sum_{i=1}^{j}\left(x_i - \overline{x}\right)^2 \cdot n_i = \frac{\left(x_1 - \overline{x}\right)^2 \cdot n_1 + \left(x_2 - \overline{x}\right)^2 \cdot n_2 + \ldots + \left(x_j - \overline{x}\right)^2 \cdot n_j}{n}$$

Fall II: Relative Häufigkeit $h_i = \frac{n_i}{n}$

$$s^2 = \sum_{i=1}^{j}\left(x_i - \overline{x}\right)^2 \cdot h_i = \left(x_1 - \overline{x}\right)^2 \cdot h_1 + \left(x_2 - \overline{x}\right)^2 \cdot h_2 + \ldots + \left(x_j - \overline{x}\right)^2 \cdot h_j$$

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

n_i : absolute Häufigkeit der Merkmalsausprägung x_i — n : Summe der absoluten Häufigkeiten

h_i : relative Häufigkeit der Merkmalsausprägung x_i — j : Anzahl der Merkmalsausprägungen x_i

Berechnung der Varianz aus einer klassierten Häufigkeitstabelle

Fall I: Absolute Häufigkeit n_i

$$n = \sum_{i=1}^{k} n_i = n_1 + n_2 + \ldots + n_k$$

$$s^2 = \frac{1}{n}\sum_{i=1}^{k}\left(m_i - \overline{x}\right)^2 \cdot n_i = \frac{\left(m_1 - \overline{x}\right)^2 \cdot n_1 + \left(m_2 - \overline{x}\right)^2 \cdot n_2 + \ldots + \left(m_k - \overline{x}\right)^2 \cdot n_k}{n}$$

Fall II: Relative Häufigkeit $h_i = \frac{n_i}{n}$

$$s^2 = \sum_{i=1}^{k}\left(m_i - \overline{x}\right)^2 \cdot h_i = \left(m_1 - \overline{x}\right)^2 \cdot h_1 + \left(m_2 - \overline{x}\right)^2 \cdot h_2 + \ldots + \left(m_k - \overline{x}\right)^2 \cdot h_k$$

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

n_i : absolute Häufigkeit der i - ten Klasse — n : Summe der absoluten Häufigkeiten

h_i : relative Häufigkeit der i - ten Klasse — k : Anzahl der Klassen

m_i : Klassenmitte der i -ten Klasse