1. Beschreibung von eindimensionalen Häufigkeitsverteilungen durch Maßzahlen

1.1. Häufigkeiten

Merkmal x_{ij} ... Merkmalsausprägungen $j=1(1)n_i,\ i=1(1)k$
n ... Beobachtungen
k ... Gruppen

absolute Häufigkeiten n_i $\sum_{i=1}^{k} n_i = n$

relative Häufigkeiten $h_i = \frac{n_i}{n}$ $\sum_{i=1}^{k} h_i = 1$

Häufigkeitsdichte $f_h(x_i) = \frac{h_i}{x_{i,o} - x_{i,u}}$ oder $f_n(x_i) = \frac{n_i}{x_{i,o} - x_{i,u}}$

empirische Häufigkeitsverteilung $h_i = f(x_i)$

empirische Verteilungsfunktion ergibt sich durch Kumulierung der relativen Häufigkeiten über alle Merkmalsausprägungen mit $x_j \leq x_i$: $H_i = F(x_i) = \sum_{x_j \leq x_i} h_j$

Gesamtmerkmalsgröße

absolute Häufigkeiten $m_i = x_i \cdot n_i$ $\sum_{i=1}^{k} m_i = m$

relative Häufigkeiten $l_i = \frac{m_i}{m}$ $\sum_{i=1}^{k} l_i = 1$

empirische Häufigkeitsverteilung $l_i = f(x_i)$

empirische Verteilungsfunktion $L_i = F(x_i) = \sum_{x_j \leq x_i} l_j$

1.2. Mittelwerte

Dichtemittel (Modalwert)

- nominalskalierte Merkmale $x_D = A_j$, wenn $n(A_j) = \underset{i}{\mathrm{Max}}[n(A_i)]$

- ungruppierte Merkmale (ordinal- und kardinalskaliert) $x_D = x_j$, wenn $n(x_j) = \underset{i}{\mathrm{Max}}\,[n(x_i)]$

- gruppierte Merkmale (kardinalskaliert) $x_D = x_j$ (Gruppenmitte) wenn $f_n(x_j) = \underset{i}{\mathrm{Max}}[\, f_n(x_j)]$

Zentralwert (Median)

- ungruppiertes Material x

$$x_{0,5} = x_{\left(\frac{n+1}{2}\right)} \quad \text{für n ungerade}$$

$$x_{0,5} = \frac{1}{2}\left(x_{\left(\frac{n}{2}\right)} + x_{\left(\left(\frac{n}{2}\right)+1\right)} \right) \quad \text{für n gerade}$$

- gruppiertes Material

$$x_{0,5} = x_{i,u} + \frac{0{,}5 - H_{i-1}}{H_i - H_{i-1}} \cdot (x_{i,o} - x_{i,u})$$

Arithmetisches Mittel

$i=1(1)k$; k ... Anzahl der Gruppen
$j=1(1)n_i$ n_i ... Anzahl der Elemente in einer Gruppe

- ungruppiertes Material

$$\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j\,; \qquad \text{mit i=k=1 (da Daten ungruppiert)}$$

- gruppiertes Material

$$\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij} \qquad \text{(Gruppenmittelwert)}$$

Berechnungen des gewogenen arithmetischen Mittels:

- bei Verwendung der Mittelwerte der Gruppen $\bar{x}_i$:

$$\bar{x} = \frac{\sum_{i=1}^{k} \bar{x}_i \cdot n_i}{\sum_{i=1}^{k} n_i} = \sum_{i=1}^{k} \bar{x}_i \cdot h_i$$

- bei Verwendung der Gruppenmitten x_i^*:

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i^* \cdot n_i}{\sum_{i=1}^{k} n_i} \quad \text{mit} \quad x_i^* = \frac{x_{i,u} + x_{i,o}}{2}$$

Harmonisches Mittel

- ungewichtet $\qquad x_h = \dfrac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$

- gewichtet $\qquad x_h = \dfrac{\sum_{i=1}^{k} n_i}{\sum_{i=1}^{k} \frac{1}{x_i} \cdot n_i} = \dfrac{1}{\sum_{i=1}^{k} \frac{1}{x_i} \cdot h_i}$

Geometrisches Mittel

- ungewichtet: $$x_g = \sqrt[n]{\prod_{i=1}^{n} x_i}$$

- gewichtet: $$x_g = \sqrt[n]{\prod_{i=1}^{k} x_i^{n_i}} \; ; \qquad \text{mit} \quad \sum_{i=1}^{k} n_i = n$$

Sonderfall: Wachstumsrate

$$p = \sqrt[n]{\prod_{t=1}^{n} \frac{y_t}{y_{t-1}}} = \sqrt[n]{\frac{y_n}{y_0}}$$; wobei y_t ... Wert einer Zeitreihe t=0(1)n

Chronologisches Mittel (= spezieller Mittelwert für Zeitpunktreihen)

$$y_c = \frac{\frac{y_1 + y_2}{2} + \frac{y_2 + y_3}{2} + \ldots + \frac{y_{n-1} + y_n}{2}}{n-1} = \frac{\frac{1}{2}\left(y_1 + y_n\right) + \sum_{i=2}^{n-1} y_i}{n-1}$$

1.3. Streuungsmaße

Schwankungsbereich (Variationsbreite, Spannweite)

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

Mittlerer Quartilsabstand

absolut: $$QA_m = \frac{\left(x_{0,75} - x_{0,5}\right) + \left(x_{0,5} - x_{0,25}\right)}{2}$$

relativ: $$QA_m^{'} = \frac{QA_m}{x_{0,5}}$$

Durchschnittliche absolute lineare Abweichung (lineares Streuungsmaß)

- ungewichtet: $$\bar{d} = \frac{\sum_{j=1}^{n_i} |x_j - \bar{x}|}{n_i} \qquad \text{mit } i = k = 1$$

- gewichtet: $$\bar{d} = \frac{\sum_{i=1}^{k} |x_i - \bar{x}| \cdot n_i}{\sum_{i=1}^{k} n_i} = \sum_{i=1}^{k} |x_i - \bar{x}| \cdot h_i$$

Vermerk: statt x_i entweder $\bar{x}_i$ oder x_i^*

Durchschnittliche relative lineare Abweichung

$$\bar{d}' = \frac{\bar{d}}{\bar{x}}$$

Varianz und Standardabweichung

Varianz

- ungewichtet: $$s_x^2 = \frac{\sum_{j=1}^{n} (x_j - \bar{x})^2}{n} = \frac{\sum_{j=1}^{n} x_j^2}{n} - \bar{x}^2$$

- gewichtet: $$s_x^2 = \frac{\sum_{i=1}^{k} (x_i - \bar{x})^2 \cdot n_i}{\sum_{i=1}^{k} n_i} = \frac{\sum_{i=1}^{k} x_i^2 \cdot n_i}{\sum_{i=1}^{k} n_i} - \bar{x}^2$$

Vermerk: statt x_i entweder $\bar{x}_i$ oder x_i^*

Standardabweichung $$s_x = \sqrt{s_x^2}$$

Variationskoeffizient $$v_x = \frac{s_x}{\bar{x}}$$