

Formelsammlung Statistik

Lageparameter und Streumaße:

Einfaches arithmetisches Mittel: $\mu = \frac{1}{n} \sum_{i=1}^n x_i \quad \left[\text{oder} \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \right]$

Gewogenes arithmetisches Mittel:

$$\mu = \frac{\sum_{i=1}^k (x_i \cdot n_i)}{\sum_{i=1}^k n_i} = \frac{1}{n} \sum_{i=1}^k (x_i \cdot n_i) = \sum_{i=1}^k \left(x_i \cdot \frac{n_i}{n} \right)$$

Arithmetisches Mittel bei klassierten Merkmalsausprägungen:

$$\mu = \frac{1}{n} \sum_{i=1}^k [(x_i)_m \cdot n_i] = \sum_{i=1}^k \left[(x_i)_m \cdot \frac{n_i}{n} \right] \quad \text{mit } (x_i)_m \text{ als Klassenmitte der Klasse } i$$

Median (Zentralwert): $\overline{x_M} = \begin{cases} x_{\frac{n+1}{2}} & \text{für } n \text{ ungerade} \\ \frac{1}{2} \left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1} \right) & \text{für } n \text{ gerade} \end{cases}$

Quantile: $\overline{x_p} = \begin{cases} x_{[n \cdot p]+1} & \text{für } (n \cdot p) \text{ nicht ganzzahlig} \\ \frac{1}{2} (x_{n \cdot p} + x_{n \cdot p+1}) & \text{für } (n \cdot p) \text{ ganzzahlig} \end{cases}$

Varianz: $\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 = \frac{1}{n} \sum_{i=1}^n (x_i)^2 - \mu^2$

Varianz bei Häufigkeitsverteilung: $\sigma^2 = \frac{1}{n} \sum_{i=1}^k (x_i - \mu)^2 \cdot n_i$

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

Modus: *Häufigster Wert einer Verteilung.*

Gini-Koeffizient:

Berechnung der Flächen mit relativen kumulierten Häufigkeiten:

$$Gini = \frac{\text{Konzentrationsfläche } K}{\text{maximale Konzentrationsfläche } K_{\max}}$$

Anmerkung: Als „Konzentrationsfläche“ K bezeichnet man die Fläche, zwischen der Gleichverteilungsdiagonalen und der LORENZ-Kurve: $0 \leq K \leq 1/2$.

Der Wertebereich des Gini-Koeffizienten liegt zwischen 0 (= Gleichverteilung) und 1 (= vollständige Konzentration auf einen Merkmalsträger).

Preisindizes

nach Laspeyres: $L_p = \frac{\sum p_{1i} \cdot q_{0i}}{\sum p_{0i} \cdot q_{0i}} = \sum \frac{p_{1i}}{p_{0i}} \cdot \frac{p_{0i} \cdot q_{0i}}{\sum p_{0i} \cdot q_{0i}} \quad \text{mit } i \in N$

nach Paasche: $P_p = \frac{\sum p_{1i} \cdot q_{1i}}{\sum p_{0i} \cdot q_{1i}} = \sum \frac{p_{1i}}{p_{0i}} \cdot \frac{p_{0i} \cdot q_{1i}}{\sum p_{0i} \cdot q_{1i}} \quad \text{mit } i \in N$

Lineare Regression und Korrelation

Ansatz: $y = b_0 + b_1 x$

$$\mu_x = \frac{1}{n} \sum_{i=1}^n x_i \quad \mu_y = \frac{1}{n} \sum_{i=1}^n y_i$$

$$b_1 = \frac{\sigma_{xy}}{\sigma_x^2} = \frac{\sum_{i=1}^n (x_i y_i) - n \cdot \mu_x \cdot \mu_y}{\sum_{i=1}^n (x_i^2) - n \cdot \mu_x^2} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i y_i) - \mu_x \cdot \mu_y}{\frac{1}{n} \sum_{i=1}^n (x_i^2) - \mu_x^2}$$

$$b_0 = \mu_y - b_1 \cdot \mu_x$$

Korrelation nach Brevais-Pearson:

$$r = \frac{\sum_{i=1}^n (x_i - \mu_x) \cdot (y_i - \mu_y)}{\sqrt{\sum_{i=1}^n (x_i - \mu_x)^2} \cdot \sqrt{\sum_{i=1}^n (y_i - \mu_y)^2}} = \frac{\sum_{i=1}^n (x_i y_i) - n \cdot \mu_x \cdot \mu_y}{\sqrt{\sum_{i=1}^n (x_i^2) - n \cdot \mu_x^2} \cdot \sqrt{\sum_{i=1}^n (y_i^2) - n \cdot \mu_y^2}}$$

Rangkorrelation nach Spearman

$$r_s = 1 - \frac{6 \sum_{i=1}^n (d_i)^2}{n(n^2 - 1)}$$

Stochastik

Erwartungswert einer diskreten Zufallsvariablen:

$$\mu = E(X) = \sum_{i=1}^n x_i \cdot P(X = x_i) \quad \text{mit} \quad X = \{x_1, x_2, \dots, x_n\}$$

Stochastische Unabhängigkeit:

$$A, B \text{ sind stochastisch unabhängig} \Leftrightarrow P(A \cap B) = P(A) \cdot P(B)$$

Bedingte Wahrscheinlichkeit:

$$P_B(A) = \frac{P(A \cap B)}{P(B)} \quad \text{W'keit für A unter der Bedingung B (B ist bereits eingetreten)}$$

Totale Wahrscheinlichkeit:

$$P(B) = \sum_{i=1}^k [P_{A_i}(B) \cdot P(A_i)]$$

Satz von Bayes:

$$P_B(A_j) = \frac{P_{A_j}(B) \cdot P(A_j)}{\sum_{i=1}^k [P_{A_i}(B) \cdot P(A_i)]}$$