

Formelsammlung – Statistik I**1 Empirische Verteilungsfunktion**Empirische Verteilungsfunktion

$$F(x) = \begin{cases} 0 & \text{für } x < a_1 \\ \sum_{i=1}^j f_i & \text{für } a_j \leq x \text{ und } a_{j+1} > x \\ 1 & \text{für } x \geq a_k \end{cases}$$

Stetige empirische Verteilungsfunktion

$$F(x) = \begin{cases} 0 & \text{für } x \leq g_0 \\ F(g_{i-1}) + \frac{x - g_{i-1}}{d_i} * f_i & \text{für } g_{i-1} < x \leq g_i \\ 1 & \text{für } x \geq g_k \end{cases}$$

2 Statistische LagemaßeArithmetisches Mittel

a) bei unklassierten Daten

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

b) bei klassierten Daten

$$\bar{x}_g = \sum_{i=1}^k m_i * f_i$$

Median

a) bei einer ungeraden Anzahl von Werten

$$x_{med} = x_{\left(\frac{n+1}{2}\right)}$$

b) bei einer geraden Anzahl von Werten

$$x_{med} = \frac{1}{2} \left(x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)} \right)$$

Perzentilwertea) ergibt (n * p) keinen ganzzahligen Wert,
ist k die auf (n * p) folgende ganze Zahl

$$x_p = x_{(k)}$$

b) ergibt (n * p) einen ganzzahligen Wert,
entspricht k dem Ergebnis von (n * p)

$$x_p = \frac{1}{2} (x_{(k)} + x_{(k+1)})$$

Modus

$$x_{mod} = a_{x_{max}}$$

(nur interpretierbar, wenn die Häufigkeitsverteilung ein eindeutiges Maximum besitzt)

Geometrisches Mittel

$$\bar{x}_{geom} = \sqrt[n]{x_1 * \dots * x_n}$$

Harmonisches Mittel

$$\bar{x}_{har} = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$$

3 Streuungsmaße / DispersionsparameterSpannweite

$$d_s = x_{\max} - x_{\min}$$

Interquartilsabstand (IQR)

$$IQR = x_{(0,75)} - x_{(0,25)}$$

Empirische Varianz

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

Standardabweichung

$$s = +\sqrt{s^2}$$

Variationskoeffizient

$$v = \frac{s}{\bar{x}} \quad (\text{nur berechenbar, wenn das arithmetische Mittel positiv ausfällt})$$

Fünf-Werte-Zusammenfassung

$$\left[x_{\min}; x_{0,25}; x_{med}; x_{0,75}; x_{\max} \right]$$

4 VerteilungsmaßeMomentenkoeffizient der Schiefe

$$g_m = \frac{m_3}{s^3}$$

$$m_3 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3$$

$$s^3 = \left(\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \right)^3$$

Quartilskoeffizient der Schiefe

$$g_{0,25} = \frac{(x_{0,75} - x_{med}) - (x_{med} - x_{0,25})}{x_{0,75} - x_{0,25}}$$

Kurtosis / Exzeß

$$g_k = \frac{m_4}{s^4} - 3$$

$$m_4 = \frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})^4$$

$$s^4 = \left(\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \right)^4$$

5 ZusammenhangsmaßeBravais-Pearson-Korrelationskoeffizient

$$r = \frac{\sum_{i=1}^n (x_i * y_i) - n * \bar{x} * \bar{y}}{\sqrt{\sum_{i=1}^n (x_i^2) - n * \bar{x}^2} * \sqrt{\sum_{i=1}^n (y_i^2) - n * \bar{y}^2}}$$

Konkordanzkoeffizient nach Kendall

$$\tau = \frac{2 * (K - D)}{n * (n - 1)}$$

Rangkorrelationskoeffizient nach Spearman

$$rho = 1 - \frac{6 * \sum d_i^2}{(n^2 - 1) * n}$$

6 Lineare RegressionBerechnung des Regressionskoeffizienten

$$b = \frac{\sum_{i=1}^n (x_i * y_i) - n * \bar{x} * \bar{y}}{\sum_{i=1}^n (x_i^2) - n * \bar{x}^2}$$

Berechnung des konstanten Glieds

$$a = \bar{y} - b * \bar{x}$$

Bestimmtheitsmaß der Regressionsfunktion

$$R^2 = \frac{ESS}{TSS}$$

TSS = Total Sum of Squares

ESS = Explained Sum of Squares