

# Formelsammlung Statistik

03/2025

**Relative Häufigkeiten:**  $f_i = h_i/n$

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

**Geometrisches Mittel:** (mit dem relativen Wachstum  $\{q_i\}$ )

$$\bar{q}_G = \sqrt[n]{(1 + q_1) \cdot (1 + q_2) \cdot \dots \cdot (1 + q_n)} - 1$$

**Harmonisches Mittel:**

$$\bar{x}_H = \frac{1}{\frac{1}{n} \left( \frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \dots + \frac{1}{x_n} \right)}$$

**Median bei einer Stichprobe einzelner Merkmalsausprägungen:**

$$\bar{x}_Z = \begin{cases} x_i & \text{mit } i = \frac{n+1}{2} \quad \text{falls } n \text{ ungerade} \\ \frac{x_i + x_{i+1}}{2} & \text{mit } i = \frac{n}{2} \quad \text{falls } n \text{ gerade} \end{cases}$$

**Median (klassierte Daten):** Median in Klasse  $k \Rightarrow$

$$\bar{x}_Z = x_k^u + (x_k^o - x_k^u) \frac{F(\bar{x}_Z) - F(x_k^u)}{F(x_k^o) - F(x_k^u)} = x_k^u + (x_k^o - x_k^u) \frac{0,5 - F(x_k^u)}{F(x_k^o) - F(x_k^u)}$$

**$p$ -Quantil bei einer Stichprobe einzelner Merkmalsausprägungen:**

$$\bar{x}_p = \begin{cases} x_i & \text{mit } i = [n \cdot p] + 1 \quad \text{falls } n \cdot p \text{ nicht ganzzahlig} \\ \frac{x_i + x_{i+1}}{2} & \text{mit } i = n \cdot p \quad \text{falls } n \cdot p \text{ ganzzahlig} \end{cases}$$

**p-Quantil (klassierte Daten):** p-Quantil in Klasse k  $\Rightarrow$

$$\bar{x}_p = x_k^u + (x_k^o - x_k^u) \frac{p - F(x_k^u)}{F(x_k^o) - F(x_k^u)}$$

**Spannweite:**

$$s_{max} = \max\{x_i\} - \min\{x_i\}$$

**Empirische Stichprobenvarianz, Standardabweichung, Variationskoeffizient:**

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

**Mittlere absolute Abweichung vom (beliebig wählbaren) Lageparameter  $\bar{x}$ :**

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

**Gini-Koeffizient:**

$$G = 1 - \sum_{i=1}^n f_i(Q_i + Q_{i-1})$$

**Kovarianz:**

$$s_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$
$$\Rightarrow s_{xx} (= s^2) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x}) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

**Lineare Regression, Pearson-Korrelationskoeffizient:**  $y = a \cdot x + b$  (linearer Zusammenhang)

$$a = \frac{s_{xy}}{s_{xx}}, \quad b = \bar{y} - a\bar{x}, \quad r_{xy} = \frac{s_{xy}}{\sqrt{s_{xx}} \cdot \sqrt{s_{yy}}}$$

**Bestimmtheitsmaß:**

$$R^2 := \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^n (e_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

mit der unerklärten Variation der Residuen  $e_i$

$$\sum_{i=1}^n (e_i)^2$$