

Formelsammlung Statistik

Lageparameter und Streumaße:

Teil 1: Diskrete Merkmalsdarstellung (Einzelwerte)

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

Gewogenes arithmetisches Mittel:

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

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

Quantile: $\bar{x}_p = \begin{cases} x_{[n \cdot p]+1} & \text{für } (n \cdot p) \text{ nicht ganzzahlig} \\ \frac{1}{2} \left(x_{n \cdot p} + x_{n \cdot p+1} \right) & \text{für } (n \cdot p) \text{ ganzzahlig} \end{cases}$

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.

Teil 2: Klassierte Merkmalsausprägungen

Arithmetisches Mittel bei klassierten Merkmalsausprägungen:

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

Modus / Modalwert:

⇒ **Modale Klasse:** Klasse mit max. Häufigkeitsdichte

⇒ **Modus / Modalwert:** Wert der Klassenmitte der modalen Klasse

Median und Quartile:

Median (bei intervallskaliertem Merkmal): $\bar{X}_{0,5} \in [a;b]$, $\bar{X}_{0,5} = a + \frac{\Delta_i \cdot (0,5 - F(a))}{p_i}$

unteres Quartil (bei intervallskaliertem Merkmal): $\bar{X}_{0,25} \in [a;b]$, $\bar{X}_{0,25} = a + \frac{\Delta_i \cdot (0,25 - F(a))}{p_i}$

oberes Quartil (bei intervallskaliertem Merkmal): $\bar{X}_{0,75} \in [a;b]$, $\bar{X}_{0,75} = a + \frac{\Delta_i \cdot (0,75 - F(a))}{p_i}$

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: $0 \leq K \leq 1/2$.

Wegen $K_{\max} = \frac{1}{2} - \frac{1}{2n}$ lautet der normierte Gini-Koeffizient:

$$\text{norm. Gini} = K \cdot \frac{2n}{n-1}$$

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

Preisindizes

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

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

Lineare Regression und Korrelation

$$\text{Ansatz: } y = b_0 + b_1 x \quad \mu_x = \frac{1}{n} \sum_{i=1}^n x_i \quad \mu_y = \frac{1}{n} \sum_{i=1}^n y_i$$

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

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

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

Stetige Wahrscheinlichkeitsverteilung - (Standard)Normalverteilung:

Wahrscheinlichkeitsdichte / Gauß-Glockenfunktion:

$$\varphi_{\mu;\sigma}(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2} \xrightarrow[\text{und } \sigma=1]{\text{mit } \mu=0} \varphi_{0;1}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

$$\Rightarrow P(X \leq k) = \Phi_{\mu;\sigma}\left(\frac{k-\mu}{\sigma}\right) \quad \text{und} \quad z = \frac{k-\mu}{\sigma}$$

$$\Rightarrow \Phi(-z) = 1 - \Phi(z) \quad \text{und} \quad \Phi(z) - \Phi(-z) = 2\Phi(z) - 1$$

Gaußsche Summenfunktion

z	0	1	2	3	4	5	6	7	8	9
0,0	0,500	0,504	0,508	0,512	0,516	0,520	0,524	0,528	0,532	0,536
0,1	0,540	0,544	0,548	0,552	0,556	0,560	0,564	0,567	0,571	0,575
0,2	0,579	0,583	0,587	0,591	0,595	0,599	0,603	0,606	0,610	0,614
0,3	0,618	0,622	0,626	0,629	0,633	0,637	0,641	0,644	0,648	0,652
0,4	0,655	0,659	0,663	0,666	0,670	0,674	0,677	0,681	0,684	0,688
0,5	0,691	0,695	0,698	0,702	0,705	0,709	0,712	0,716	0,719	0,722
1,0	0,841	0,844	0,846	0,848	0,851	0,853	0,855	0,858	0,860	0,862
1,1	0,864	0,867	0,869	0,871	0,873	0,875	0,877	0,879	0,881	0,883
1,2	0,885	0,887	0,889	0,891	0,893	0,894	0,896	0,898	0,900	0,901
1,3	0,903	0,905	0,907	0,908	0,910	0,911	0,913	0,915	0,916	0,918
1,9	0,971	0,972	0,973	0,973	0,974	0,974	0,975	0,976	0,976	0,977
2,0	0,977	0,978	0,978	0,979	0,979	0,980	0,980	0,981	0,981	0,982
2,1	0,982	0,983	0,983	0,983	0,984	0,984	0,985	0,985	0,985	0,986
2,2	0,986	0,986	0,987	0,987	0,987	0,988	0,988	0,988	0,989	0,989