Formelsammlung Statistik

Lageparameter und Streumaße:

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

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:

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

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

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

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

Gini-Koeffizient:

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

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

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

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

Preisindizes

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

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

Lineare Regression und Korrelation

Ansatz: $y = b_0 + b_1 x$

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

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

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

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:

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

Stochastische Unabhängigkeit:

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

Bedingte Wahrscheinlichkeit:

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

Totale Wahrscheinlichkeit:

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

Satz von Bayes:

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

Dichtefunktionen:

Eine Funktion f(x) heißt Wahrscheinlichkeitsdichte (Dichtefunktion) genau dann, wenn gilt:

$$(1) \quad f(x) \geq 0 \quad \forall x\in\Re \qquad und \qquad (2) \quad \int_{-\infty}^{\infty} f(x)\,dx = 1$$

Ist X eine stetige Zufallsgröße mit der Dichtefunktion f(x), so heißt die reelle Zahl

(i) $\mu(X) = \int_{-\infty}^{\infty} x\cdot f(x)\,dx$ der Erwartungswert der Zufallsgröße X.

(ii) $V(X) = \int_{-\infty}^{\infty} (x-\mu)^2\cdot f(x)\,dx$ die Varianz der Zufallsgröße X.

Diskrete Wahrscheinlichkeitsverteilungen

Binomialverteilung: $B(X=k) = \binom{n}{k}p^k\cdot(1-p)^{n-k} \quad mit \quad k\leq n$

Poissonverteilung: $P(X=k) = \frac{\mu^k}{k!}e^{-\mu} \quad mit \quad \mu = n\cdot p$

Stetige Wahrscheinlichkeitsverteilung - (Standard)Normalverteilung:

Wahrscheinlichkeitsdichte / Gauß-Glockenfunktion:

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

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

=> $\Phi(-x) = 1-\Phi(x) \quad und \quad \Phi(x)-\Phi(-x) = 2\Phi(x)-1$