

Kennzahlen

Arithm. Mittel: $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$

p-Quantil: $x_p = \begin{cases} x_{(\lfloor np \rfloor + 1)} & \text{falls } np \notin \mathbb{N} \\ \frac{1}{2}(x_{(np)} + x_{(np+1)}) & \text{falls } np \in \mathbb{N} \end{cases}$

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

Median: $x_{med} = x_{0.5}$

Korrigierter Kontingenzkoeffizient nach Pearson: $K_P^* = \left(\frac{\chi^2}{\chi^2 + n} \right)^{\frac{1}{2}} \cdot \left(\frac{\min(k,l)}{\min(k,l)-1} \right)^{\frac{1}{2}}$

Empirische Korrelation: $r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$

Auswahl von k aus n Objekten

	ohne Wiederholung ($k \leq n$)	mit Wiederholung
ohne Reihenf.	$\binom{n}{k}$	$\binom{n+k-1}{k}$
mit Reihenf.	$\frac{n!}{(n-k)!}$	n^k

Wahrscheinlichkeitsrechnung

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Bedingte Wahrscheinlichkeit: $P(A|B) = \frac{P(A \cap B)}{P(B)}$

Satz v. Bayes: $P(B_k|A) = \frac{P(B_k) \cdot P(A|B_k)}{\sum_{i=1}^m P(B_i)P(A|B_i)}$

Zufallsvariablen und Verteilungen

	Diskret	Stetig
Dichte	$f(x) = P(X = x)$	$f(x) \geq 0, \int_{-\infty}^{\infty} f(x)dx = 1$
Verteilungsfunktion	$F(x) = P(X \leq x) = \sum_{t \leq x, f(t) > 0} f(t)$	$F(x) = P(X \leq x) = \int_{-\infty}^x f(t)dt$
E(X)	$E(X) = \sum_{x \in \Omega} x f(x)$	$E(X) = \int_{-\infty}^{\infty} x f(x)dx$
Var(X)	$\text{Var}(X) = \sum_{x \in \Omega} (x - E(X))^2 f(x)$	$\text{Var}(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x)dx$

Name	Bezeichnung	f(x)	Träger Ω	E(X)	Var(X)
Diskrete Verteilungen					
Binomial	$X \sim \text{Bin}(n, p)$	$\binom{n}{x} p^x (1-p)^{n-x}$	$x \in \{0, 1, \dots, n\}$	np	$np(1-p)$
Hypergeometrisch	$X \sim \mathcal{H}(N, M, n)$	$\frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}}$	$0 \leq x \leq M, 0 \leq n-x \leq N-M$	$\frac{nM}{N}$	$n \frac{M}{N} (1 - \frac{M}{N}) \frac{N-n}{N-1}$
diskrete Gleich	$X \sim \text{DU}(m)$	$\frac{1}{m}$	$\{1, \dots, m\}$	$\frac{m+1}{2}$	$\frac{m^2-1}{12}$
Poisson	$X \sim \text{Poi}(\lambda)$	$\exp(-\lambda) \frac{\lambda^x}{x!}$	$x \in \mathbb{N}_0$	λ	λ
Stetige Verteilungen					
Normal	$X \sim N(\mu, \sigma^2)$	$\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	$x \in \mathbb{R}$	μ	σ^2
Standardnormal	$X \sim N(0, 1)$	$\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)$	$x \in \mathbb{R}$	0	1
stetige Gleich	$X \sim \text{Unif}(a, b)$	$\frac{1}{b-a}$	$x \in [a, b]$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponential	$X \sim \text{Exp}(\lambda)$	$\lambda \exp(-\lambda x)$	$x \geq 0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$

Konfidenzintervalle

100(1 - α)%-Konfidenzintervalle für Erwartungswert $E(X)$

$X \sim N(\mu, \sigma^2)$ oder n groß ($n \geq 30$), σ bekannt	$\left[\bar{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \quad \bar{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right]$
$X \sim N(\mu, \sigma^2)$, σ unbekannt	$\left[\bar{X} - t_{n-1, 1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}, \quad \bar{X} + t_{n-1, 1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}} \right]$
n groß ($n \geq 30$), σ unbekannt	$\left[\bar{X} - z_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}}, \quad \bar{X} + z_{1-\frac{\alpha}{2}} \frac{S}{\sqrt{n}} \right]$

100(1 - α)% Konfidenzintervall für Anteile p

$X \sim \text{Bin}(n, p)$, n groß ($n \geq 30$)	$\left[\hat{p} - z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \quad \hat{p} + z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right]$ mit $\hat{p} = \bar{X}$
--	---

Hypothesentests

Nullhypothese	Alternativhypothese	Teststatistik	Ablehnungsbereich
(approximativer) Gaußtest ($X \sim N(\mu, \sigma^2)$ oder $n \geq 30, \sigma$ bekannt)			
$\mu = \mu_0$	$\mu \neq \mu_0$	$Z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$	$ z > z_{1-\frac{\alpha}{2}}$
$\mu \geq \mu_0$	$\mu < \mu_0$		$z < -z_{1-\alpha}$
$\mu \leq \mu_0$	$\mu > \mu_0$		$z > z_{1-\alpha}$
t-Test auf Lage ($X \sim N(\mu, \sigma^2), \sigma$ unbekannt)			
$\mu = \mu_0$	$\mu \neq \mu_0$	$T = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}}$	$ t > t_{n-1, 1-\frac{\alpha}{2}}$
$\mu \geq \mu_0$	$\mu < \mu_0$		$t < -t_{n-1, 1-\alpha}$
$\mu \leq \mu_0$	$\mu > \mu_0$		$t > t_{n-1, 1-\alpha}$
Approximativer Gaußtest ($n \geq 30, \sigma$ unbekannt, $E(X) = \mu$)			
$\mu = \mu_0$	$\mu \neq \mu_0$	$Z = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}}$	$ z > z_{1-\frac{\alpha}{2}}$
$\mu \geq \mu_0$	$\mu < \mu_0$		$z < -z_{1-\alpha}$
$\mu \leq \mu_0$	$\mu > \mu_0$		$z > z_{1-\alpha}$
t-Test auf Lagedifferenz ($X \sim N(\mu_X, \sigma_X^2), Y \sim N(\mu_Y, \sigma_Y^2), \sigma_X, \sigma_Y$ unbekannt)			
$\mu_X - \mu_Y = \delta_0$	$\mu_X - \mu_Y \neq \delta_0$	$T = \frac{\bar{X} - \bar{Y} - \delta_0}{\sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}}}$	$ t > t_{k, 1-\frac{\alpha}{2}}$
$\mu_X - \mu_Y \geq \delta_0$	$\mu_X - \mu_Y < \delta_0$		$t < -t_{k, 1-\alpha}$
$\mu_X - \mu_Y \leq \delta_0$	$\mu_x - \mu_Y > \delta_0$		$t > t_{k, 1-\alpha}$
mit $k = \left\lceil \left(\frac{S_X^2}{n} + \frac{S_Y^2}{m} \right)^2 \middle/ \left(\frac{1}{n-1} \left(\frac{S_X^2}{n} \right)^2 + \frac{1}{m-1} \left(\frac{S_Y^2}{m} \right)^2 \right) \right\rceil$			
Approximativer Binomialtest ($X \sim \text{Bin}(1, p), n \geq 30$)			
$p = p_0$	$p \neq p_0$	$Z = \frac{\bar{X} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$	$ z > z_{1-\frac{\alpha}{2}}$
$p \geq p_0$	$p < p_0$		$z < -z_{1-\alpha}$
$p \leq p_0$	$p > p_0$		$z > z_{1-\alpha}$
Chi-Quadrat-Unabhängigkeitstest (X und Y kategorial, $\tilde{h}_{ij} > 5$)			
Merkmale X und Y sind	Merkmale X und Y sind	$\chi^2 = \sum_{i=1}^k \sum_{j=1}^l \frac{(h_{ij} - \tilde{h}_{ij})^2}{\tilde{h}_{ij}}$	$\chi^2 > \chi^2_{1-\alpha, (k-1)(l-1)}$
stochastisch unabhängig	stochastisch abhängig	$\tilde{h}_{ij} = \frac{h_{i\bullet} \cdot h_{\bullet j}}{n}$	

Regression

Regressionsmodell: $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$

$\epsilon_i \sim N(0, \sigma^2) \quad i = 1, \dots, n$

Schätzer: $\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2}$

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$

$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-2} \quad \hat{\sigma}_{\hat{\beta}_1} = \sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} \quad \hat{\sigma}_{\hat{\beta}_0} = \sqrt{\hat{\sigma}^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)}$

Angepasste Regressionsgrade: $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$

$(\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i)$

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

Residuen: $e_i = y_i - \hat{y}_i$

Nullhypothese	Alternativhypothese	Teststatistik	Ablehnungsbereich
$\beta_1 = \beta_{1,0}$	$\beta_1 \neq \beta_{1,0}$	$T = \frac{\hat{\beta}_1 - \beta_{1,0}}{\hat{\sigma}_{\hat{\beta}_1}}$	$ t > t_{n-2, 1-\alpha/2}$
$\beta_1 \geq \beta_{1,0}$	$\beta_1 < \beta_{1,0}$		$t < -t_{n-2, 1-\alpha}$
$\beta_1 \leq \beta_{1,0}$	$\beta_1 > \beta_{1,0}$		$t > t_{n-2, 1-\alpha}$

100(1 - α)-%**Konfidenzintervall für Prognose:** $\hat{Y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$

$$\left[\hat{Y}_0 - t_{n-2, 1-\alpha/2} \sqrt{\hat{\sigma}^2 \left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)}, \hat{Y}_0 + t_{n-2, 1-\alpha/2} \sqrt{\hat{\sigma}^2 \left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)} \right]$$