Studeo® Formelsammlung Induktive Statistik (aus dem Klausurtrainer Induktive Statistik)

STUDEO

Kombinatorik

n-Fakultät

$n! = n(n-1)(n-2)\cdot \ldots \cdot 2\cdot 1$

Binomischer Satz

$(a+b)^n = \sum_{k=1}^{n} \binom{n}{k} a^{n-k} b^k$

Permutation (ohne Wiederholung)

$(n) = n!$

Permutation (mit Wiederholung)

$P^W(n; g_1, \ldots, g_r) = \frac{n!}{g_1!\cdot g_2!\cdot g_3!\cdot \ldots g_r!}$

$n = g_1 + g_2 + \ldots + g_r$

Variation (ohne Wiederholung)

$V(n,k) = \frac{n!}{(n-k)!}$

Variation (mit Wiederholung)

$V^W(n,k) = n^k$

Kombination (ohne Wiederholung)

$K(n,k) = \binom{n}{k} = \frac{n!}{k!\cdot(n-k)!}$

Kombination (mit Wiederholung)

$K^W(n,k) = \binom{n+k-1}{k} = \frac{(n+k-1)!}{k!(n+k-1-k)!} = \frac{(n+k-1)!}{k!(n-1)!}$

Wahrscheinlichkeitsrechnung

Vereinigung von Ereignissen

$A = \bigcup_i A_i$

Durchschnitt von Ereignissen

$A = \bigcap_i A_i$

Komplementärereignis

$B = \overline{A} \quad P(\overline{A}) = 1 - P(A)$

Allgemeiner Additionssatz

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

Bedingte Wahrscheinlichkeit

$P(A|B) = \frac{P(A\cap B)}{P(B)}, P(B) > 0$

Unabhängige Ereignisse

$P(A|B) = P(A|\overline{B}) = P(A)$

Theorem von Bayes

$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum_i P(B|A_i)P(A_i)} \quad \forall j = 1,\ldots,n$

Eindimensionale Wahrscheinlichkeitsverteilung

Wahrscheinlichkeitsfunktion (diskrete ZV)

$f(x_i) = P(X = x_i) = p_i \quad i=1,2\ldots$

$P(a \le x \le b) = \sum_{a\le x_i \le b} P(X=x_i) = f(x_i)$

Verteilungsfunktion (diskrete ZV)

$F(x_i) = P(X \le x_i) = \sum_{k\le i} P(X = x_k)$

Wahrscheinlichkeitsdichte (stetige ZV)

$P(a < x \le b) = \int_a^b f(x)dx \quad$ für alle a,b mit $a \le b$

Verteilungsfunktion (stetige ZV)

$F(x) = P(-\infty < X \le x) = \int_{-\infty}^{x} f(t)dt$

Erwartungswert (diskrete ZV)

$E(x) = \mu = \sum_i x_i \cdot f(x_i)$

Erwartungswert (stetige ZV)

$E(x) = \mu = \int_{-\infty}^{+\infty} x\cdot f(x)dx$

$Var(x) = \sigma^2 = \sum_i (x_i - \mu)^2 \cdot f(x_i)$

$= \sum_i x_i^2 \cdot f(x_i) - \mu^2 = E(X^2) - (E(X))^2$

$Var(x) = \sigma^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)dx$

$= \int_{-\infty}^{+\infty} x^2 f(x)dx - \mu^2$

$= E[(X - E(X))^2] = E(X^2) - (E(X))^2$

Median

Der Median $x_{0,5}$ ist genau diejenige Zahl, für die gilt:

$P(X \le x_{0,5}) \ge 0{,}5 \quad$ und $\quad P(X \ge x_{0,5}) \ge 0{,}5$

Für das p-Quantil gilt:

$P(X \le x_p) \ge p \quad$ und $\quad P(X \ge x_p) \ge 1 - p$

Schiefe

$\gamma = \frac{E(x-\mu)^3}{\sigma^3} = \frac{\sum (x - E(X))^3 \cdot f(x)}{\sigma^3}$

Standardabweichung

$\sigma = \sqrt{Var(X)}$

k-faches Schwankungsintervall

$P(\mu - k\cdot\sigma \le x \le \mu + k\cdot\sigma)$

Zweidimensionale Wahrscheinlichkeitsverteilung

Zweidimensionale Wahrscheinlichkeitsfunktion

$f(x,y) = f_{XY}(x_i, y_j) = P(X = x_i, Y = y_j)$

Verteilungsfunktion (diskret)

$F_{XY}(x,y) = \sum_{x_i \le x}\sum_{y_j \le y} f_{XY}(x_i, y_j) = F(x,y)$

Verteilungsfunktion (stetig)

$F_{XY}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{XY}(x,y)dxdy = F(x,y)$ Bedingte

Verteilung (diskreter Fall)

$P(X = x_i | Y = y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)} = \frac{f_{XY}(x_i, y_j)}{f_Y(y_j)}$

$= f_X(x_i | y_j)$

Bedingte Verteilung (stetiger Fall)

$f_X(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)}$

Kovarianz

$Cov(X,Y) = E[(X - E(X))\cdot(Y - E(Y))]$

$= E(XY) - (E(X)\cdot E(Y)$

mit $E(XY) = \sum_x\sum_y x\cdot y\cdot f_{xy}(x,y)$

Korrelationskoeffizient

$\rho(X,Y) = \frac{Cov(X,Y)}{\sigma_x\sigma_y}$,es gilt: $-1 \le \rho(X,Y) \le 1$

Spezielle Verteilungmodelle

Diskrete Gleichverteilung

$f(x_i) = \begin{cases} \frac{1}{n} & i = 1\ldots n \\ 0 & \text{sonst} \end{cases}$

Stetige Gleichverteilung

$f(x) = \begin{cases} \frac{1}{b-a} & a \le x \le b \\ 0 & \text{sonst} \end{cases} \quad F(x) = \begin{cases} 0 & x<a \\ \frac{x-a}{b-a} & a \le x<b \\ 1 & x \ge b \end{cases}$

Binomialverteilung

$f_B(x;n,p) = B(n,p) =$

$\begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0,1,\ldots,n \\ 0 & \text{sonst} \end{cases}$

Poissonverteilung

$f_{PO}(x;\lambda) = \frac{\lambda^x}{x!} e^{-\lambda} \quad$ für $x=0,1,2\ldots; \lambda > 0$

Geometrische Verteilung

$f(x;p) = p(1-p)^{x-1} \quad$ für $x=1,2\ldots$

Hypergeometrische Verteilung

$f_H(x;N,M,n) = H(N,M,n) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$

Exponentialverteilung

$EX(\lambda) = f_{EX}(x;\lambda) = \begin{cases} \lambda\cdot e^{-\lambda x} & x \ge 0, \lambda > 0 \\ 0 & x < 0 \end{cases}$

Normalverteilung (Wahrscheinlichkeitsdichte)

$f_{NV}(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\cdot e^{\frac{-(x-\mu)^2}{2\sigma^2}} = N(\mu,\sigma)$

Einfache statistische Schätzverfahren

Likelihood-Funktion

$L(\vartheta) = L(\vartheta | x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f(x_i|\vartheta)$

Methode der kleinsten Quadrate

$Q(\hat{\vartheta}) = \sum_i [X_i - E(X_i)]^2 = \sum_i [X_i - g(\hat{\vartheta})]^2$

Mittlere quadratische Abweichung

$E[(\hat{\theta} - \vartheta)^2] = E[\{\hat{\theta} - E(\hat{\theta})\}^2] + \{E(\hat{\theta}) - \vartheta\}^2$

Zentraler Grenzwertsatz

Sind X_1, X_2, ..X_n unabhängige Zufallsvariablen mit identischer Verteilung, also auch gleichem Erwartungswert $E(X_i) = \mu$ und $Var(X_i) = \sigma^2 > 0$, dann ist die standardisierte Summe Z_n dieser Zufallsvariablen:

$Z_n = \frac{\sum_{i=1}^{n} X_i - n\cdot\mu}{\sqrt{n\cdot\sigma^2}} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{x-\mu}{\sigma}$

Näherungsweise standartnormalverteilt.

Für $n \to \infty$ konvergiert die Verteilung von Z_n gegen die Standardnormalverteilung $N(0,1)$ (praktisch gilt diese Näherung ab n>30)

Stichprobenvertei-lung des Stich-probenmittelwertes

Stichprobenfunktion Stichprobenwert

$\overline{X} = \frac{1}{n}\sum_i X_i \qquad \overline{x} = \frac{1}{n}\sum_i x_i$

Testtheorie

Hypothesen eines zweiseitigen Tests

Nullhypothese Alternativhypothese

$H_0: \upsilon = \upsilon_0 \qquad H_1: \upsilon \ne \upsilon_0$

Hypothesen eines einseitigen Tests

Nullhypothese Alternativhypothese

rechtsseitiger Test:

$H_0: \upsilon \le \upsilon_0 \qquad H_1: \upsilon > \upsilon_0$

linksseitiger Test:

$H_0: \upsilon \ge \upsilon_0 \qquad H_1: \upsilon < \upsilon_0$

Gütefunktion

$G(\upsilon) = P(H_1 | H_0: \upsilon) \le \alpha$

rechtsseitiger Test

$G(\mu) = 1 - P(z \le z_{1-\alpha} - \frac{\mu - \mu_0}{\sigma}\sqrt{n})$

linksseitiger Test:

$G(\mu) = P(z \le -z_{1-\alpha} - \frac{\mu - \mu_0}{\sigma}\sqrt{n})]$

Anpassungstest

$T = \sum_{j=1}^{m}\frac{(h_{0_j} - h_{e_j})^2}{h_{e_j}}$

ist unter H_0 χ^2-verteit mit

df = m - 1 - k (für $h_{e_{ij}}$ ≥5)

m = Anzahl Klassen/ Ausprägungen

k = Anzahl zu schätzender Parameter

χ^2 - Unabhängigkeitstest

ist unter H_0 χ^2 -verteilt mit

$T = \sum_{i=1}^{l}\sum_{j=1}^{m}\frac{(h_{0_{ij}} - h_{e_{ij}})^2}{h_{e_{ij}}}$

df=(l-1)(m-1),

falls $h_{e_{ij}}$ ≥5 gilt für alle i,j

Test auf Mittelwertsunterschiede (zwei unabhängige Stichproben)

Differenzentest, wenn σ_1 und σ_2 bekannt

$T = \frac{\overline{x}_1 - \overline{x}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$

$\left[\overline{x}_1 - \overline{x}_2 - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}};\ \overline{x}_1 - \overline{x}_2 + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right]$

$T = \frac{\overline{x}_1 - \overline{x}_2 - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$

σ_1 und σ_2 identisch, aber unbekannt

ist unter H_0 t-verteilt mit

df=n_1+n_2-2

s_p-Stichprobenvarianz

$s_p = \frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}$

Konfidenzintervall:

$\left[F_{\frac{\alpha}{2};df1,df2}\cdot\frac{s_1^2}{s_2^2};\ F_{1-\frac{\alpha}{2};df1,df2}\cdot\frac{s_1^2}{s_2^2}\right]$

Fortsetzung im Klausurtrainer Statistik