Die Formelsammlungen sind teilweise stark veraltet (Vorlesungsinhalte aus vergangenen Semestern, alte Normen...) und sollten lediglich als Hilfestellung zum Verfassen eigener Formelsammlungen dienen. Kontrolliert auf jeden Fall die Formeln, es haben sich auch Fehler eingeschlichen.

Statistik Formelsammlung

Jan Höffgen

28. Mai 2012

Diese Zusammenfassung wurde auf Basis der Vorlesung
Angewandte Statistik
im Sommersemester 2011 erstellt.

Eine Aktualisierung erfolgt nicht mehr.

Beziehungen zwischen Ereignissen

Unabhängige Ereignisse

disjunkt: $P(A \cap B) = 0,\ P(A \setminus B) = P(A \cap \bar{B}) = P(A),\ P(A \cup B) = P(A) + P(B)$

nicht disjunkt: $P(A \cap B) = P(A) \cdot P(B),\ P(A \setminus B) = P(A) - P(A \cap B),\ P(A \cup B) = P(A) + P(B) - P(A \cap B),$
$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(C \cap B) + P(A \cap B \cap C)$

Nicht unabhängige Ereignisse

Bedingte Wahrscheinlichkeiten: $P(B|A) = \frac{P(A \cap B)}{P(A)} \Leftrightarrow P(A \cap B) = P(A|B) \cdot P(B) = P(B|A) \cdot P(A)$

Kombinatorik: Anzahl möglicher Kombinationen je Fall

ohne Zurücklegen: $C^n_{k,Rf} = \frac{n!}{(n-k)!},\ C^n_k = \binom{n}{k}$ (Rf: Reihenfolge interessiert)

mit Zurücklegen: ${}^wC^n_{k,Rf} = n^k,\ {}^wC^n_k = \frac{(n+k-1)!}{(n-1)! \cdot k!}$

Diskrete Daten, stetige Daten (Stichprobe, diskrete Gesamtheit $\rightarrow$ stetige Gesamtheit)

k-tes Moment: $M_{xk} = \frac{1}{n} \sum_{i=1}^n x_i^k,\ m_{xk} = \sum_{j=1}^\infty p_j x_j^k \rightarrow m_{xk} = \int_{-\infty}^\infty x^k f(x) dx;$

Mittelwert $\bar{x}$, Erwartungswert μ: 1. Moment

k-tes Zentralmoment: $M_{cxk} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^k,\ m_{xk} = \sum_{j=1}^\infty p_j (x_j - \mu)^k \rightarrow m_{cxk} = \int_{-\infty}^\infty (x-\mu)^k f(x) dx$

Unterschreitungswahrscheinlichkeit: $P_u(x_l) = \sum_{j=1}^l \frac{n_j}{n},\ F(x_l) = \sum_{j=1}^l p_j \rightarrow F(X) = \int_{-\infty}^X f(v) dv$

Varianz : $S_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n-1} (\sum_{i=1}^n x_i^2 - n\bar{x}^2)$ (2. Zentralmoment)

Kovarianz: $S_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n-1} (\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y})$

Variationskoeffizient: $C_{vx} = \frac{s_x}{\bar{x}}$, **Median**: Quantil, das in 50% der Fälle über-/unterschritten wird

Schiefekoffizient: $\gamma = \frac{m_{cx3}}{\sigma_x^3}, \gamma > 0$: linksschief, $\gamma = 0$: symmetrisch, $\gamma < 0$: rechtsschief

Klassen (große Stichproben mit n Werten in K Klassen einteilen: im Mittel 8-10 Werte pro Klasse)

$\bar{x} = \frac{1}{n} \sum_{j=1}^K n_j \cdot \tilde{x}_j$ ($\tilde{x}_j$: Klassenmitte, n_j abs. Klassenhäufigkeit)

Diskrete Gleichverteilung: $p_j = \frac{1}{N}$, N: Anzahl der möglichen x-Werte

$E(x) = \mu = \frac{1}{N} \sum_{j=1}^N x_j,\ \sigma^2 = \frac{1}{N} \sum_{j=1}^N x_j^2 - \mu^2$

Stetige Gleichverteilung in den Grenzen a und b: $f(x) = \frac{1}{b-a}$ für $a \leq x \leq b \Rightarrow \mu = \frac{a+b}{2},\ \sigma^2 = \frac{(b-a)^2}{12}$

Binomial-Verteilung (unabhängige Versuche)

Wahrscheinlichkeit für k mal *Erfolg* in n Versuchen: $P(X = k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$

mit p= Wahrscheinlichkeit für Erfolg, $\mu = n \cdot p,\ \sigma^2 = n \cdot p \cdot (1-p)$

Poisson-Verteilung: Binomialverteilung für $p \rightarrow 0$ und $n \rightarrow \infty$: $P(X = k) = \frac{(n \cdot p)^k}{k!} e^{-n \cdot p} = \frac{\mu^k}{k!} e^{-n \cdot p}$

Lineare **Transformation stetiger Zufallsvariablen** (Wahrscheinlichkeitsdichten): $f_y(y) = f_x(x) \left| \frac{dx}{dy} \right|$

Gauß- oder Normalverteilung

Dichtefunktion: $f_x(x) = \frac{1}{\sqrt{2\pi}\cdot\sigma} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, Verteilungsfunktion: $F_x(x) = \frac{1}{\sqrt{2\pi}\cdot\sigma} \int_{-\infty}^{x} e^{-\frac{(v-\mu)^2}{2\sigma^2}} dv$
Transformation der Normalverteilung $x = \mu_x + z\sigma_x \rightarrow$ standartisierte Normalvert. N(0,1) mit $\mu = 0,\ \sigma = 1$:
$\varphi_z(z) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{z^2}{2}},\ \Phi_z(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{u^2}{2}} du\ \rightarrow$ Werte aus Tabelle abzulesen
Logarithmische Transformation der Normalverteilung: $y = \ln x \Rightarrow \frac{dy}{dx} = \frac{1}{x}$
$\Rightarrow f_x(x) = \frac{f_y(y)}{x} = \frac{1}{x\sqrt{2\pi}\cdot\sigma_y} \cdot e^{-\frac{(\ln x-\mu_y)^2}{2\sigma_y^2}}$
3-parametrige LOG-Normalverteilung: aus $y = \ln x$ wird $y = \ln(x - x_0)$

Gammaverteilung: $f(x) = \frac{\lambda}{\Gamma(r)}[\lambda \cdot (x - x_0)]^{r-1} \cdot e^{-\lambda(x-x_0)}$
mit $\Gamma(r) = \int_0^{\infty} x^{r-1}e^{-x}dx \rightarrow \Gamma(n+1) = n!, \Gamma(r+1) = r\Gamma(r)$
3-parametrige Verteilung: $r = \frac{4}{\gamma^2}, \lambda = \frac{2}{\sigma_x\cdot\gamma}, x_0 = \mu_x - 2\frac{\sigma_x}{\gamma}$ (aus Momenten)

Weibull-Verteilung: $f(x) = s \cdot \lambda \cdot (x - x_0)^{s-1} \cdot e^{-\lambda(x-x_0)^s},\ F(x) = 1 - e^{-\lambda(x-x_0)^s}$

Schließende Statistik: Entspricht ein aus einer Stichprobe gegebener Wert mit einer Koinzidenz von $\gamma = 1-\alpha$ (Irrtumswahrscheinlichkeit) der Grundgesamtheit? $\rightarrow$ Umrechnung: Varianz des Mittelwertes $s_{\bar{x}} = \frac{s_x}{\sqrt{n}}$
kleine Stichproben (n<30): **Tschebyscheff'sche Ungleichung**: $P((\bar{x}-h\cdot s_{\bar{x}}) \leq \mu_x \leq (\bar{x}+h\cdot s_{\bar{x}})) \geq 1-\frac{1}{h^2} = \gamma$
große Stichproben: **asymptotische Normalverteilung**: $P((\bar{x} - z \cdot s_{\bar{x}}) \leq \mu_x \leq (\bar{x} + z \cdot s_{\bar{x}})) \approx D(z) = \gamma$
Berechnung mit $\ln y$ anstatt x: wie gehabt, am Ende Umrechnung: $C_{u/o} = e^{C_{u/o,log}}$
Allgemeines Vorgehen: Gegeben: μ^*, Stichprobe mit $n, \bar{x}, s_x, /\sigma_x$, Signifikanzniveau $\rightarrow$ Einseitige/ zweiseitige Fragestellung: Abweichung des Stichprobenmittels nach oben (und/oder) unten signifikant? $\rightarrow$ Hypothese: $H_0 : \mu = \mu^*$ (Mittelwert aus Stichprobe gleicht Grundgesamtheit), $H_1 : \mu \lesseqgtr \mu^*$ (Mittelwert aus Stichprobe größer/kleiner Grundgesamtheit) $\rightarrow$ Für die Verteilung von $\bar{x}$ (Teststatistik) kann nach zentralem Grenzwertsatz eine Normalverteilung angenommen werden $\rightarrow$ Standartabweichung von $\bar{x}$ $\rightarrow$ Kritischer Wert ist das Quantil C_u/C_o der NV der Zufallsvariable $\bar{x}$, das eine Unterschreitungswahrscheinlichkeit von α hat $\rightarrow$ Quantile ausrechnen $\rightarrow$ Entscheidung: $\bar{x}$ auf der richtigen Seite von C?

Lineare Regression: geg: n Wertepaare, ges: linearer Zusammenhang $y = a + bx$
$a = \bar{y} - b\bar{x},\ b = \frac{S_{xy}}{S_x^2}$, Maß für die Abhängigkeit: Korrelationskoeffizient: $r = \frac{S_{xy}}{S_xS_y} = \sqrt{b_xb_y}$
($r = \pm 1$: deterministischer Zusammenhang, r=0: stochastisch unabhängig)
Lin. Reg mit **transformierten Variablen**: $y = a + \frac{b}{x} \rightarrow x^* = \frac{1}{x} \rightarrow y = a + bx^*$,
$y = ab^x \rightarrow y^* = \ln y = \ln a + \ln b \cdot x \rightarrow y^* = a^* + b^*x,\ y = ax^b \rightarrow y^* = \ln y = \ln a + b\ln x \rightarrow y^* = a^* + bx^*$