Inhaltsverzeichnis

1 Wahrscheinlichkeitstheorie

1.1 Wahrscheinlichkeitsbegriffe

Wahrscheinlichkeit nach Laplace: $\frac{Anzah\ \ der\ günstigen\ Fälle}{Anzahl\ der\ (gleich)\ möglichen\ Fälle}$

Wahrscheinlichkeit nach Mises: P(A)= $\lim\limits_{n \to \infty} \frac{n_A}{n}$

n_A: Anzahl der Erfolge
n: Anzahl der Versuche

1.2 Bedingte Wahrscheinlichkeit

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

1.3 Satz von Bayes

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A) \cdot P(B|A)}{P(B|A) \cdot P(A) + P(B|A^C) \cdot P(A^C)} = \frac{P(B|A) \cdot P(A)}{P(B|A) \cdot P(A) + P(B|A^C) \cdot P(A^C)}$$

1.4 Stochastische Unabhängigkeit zweier Ereignisse

1. $P(B|A) = P(B)$ und $P(A|B) = P(A)$
2. $P(A \cap B) = P(A) \cdot P(B)$

1.5 Axiome der Wahrscheinlichkeitsrechnung von Kolmogoroff

$P(A) \geq 0$ für alle Ereignisse
$P(A \cup B) = P(A) + P(B)$ für disjunkte Ereignisse
$P(\Omega) = 1$

1.6 Verteilungsfunktion einer Zufallsvariable (ZV) X

$$F(x) = P(X \leq x)$$

1.7 Dichte einer diskreten Zufallsvariable (Wahrscheinlichkeitsfunktion)

$$f(x) = P(X = x)$$

1.8 Zusammenhänge von Dichten und Verteilungsfunktionen

Handelt es sich um eine stetige Zufallsgröße, so gilt allgemein:

$$F(b) = P(-\infty \leq X \leq b) = \int_{-\infty}^{b} f_X(x)\,dx$$

Es gilt:

$$P(a < X \leq b) = \sum_{a<x_i\leq b} f(x_i)$$ bei diskreter Zufallsvariable X

$$P(a \leq X \leq b) = \int_a^b f_X(x)\, dx = F(b) - F(a)$$ bei stetiger Zufallsvariable X

$$P(X = x) = 0 \quad \text{für alle } x \in \mathbb{R}\text{, falls X eine stetige ZV ist}$$
$$F'(x) = f(x)$$

1.9 Eigenschaften der Dichtefinktion

$f(x) \geq 0$

$$\int_{-\infty}^{\infty} f(x)\, dx = 1$$

1.10 Unabhängigkeit zweier Zufallsgrößen

Zwei Zufallsgrößen X und Y heißen unabhängig, falls die Ereignisse $X \leq x_0$ und $Y \leq y_0$ für beliebige Werte von x_0 und y_0 unabhängig sind. Es gilt dann:

$$P(X = x_o \text{ und } Y = y_0) = P(X = x_0) \cdot P(Y = y_0)$$

$$P(X \leq x_o \text{ und } Y \leq y_0) = P(X \leq x_0) \cdot P(Y \leq y_0)$$

1.11 Erwartungswert μ, Varianz und Standardabweichung

Erwartungswert [E(X)]

$$E(X) = \sum_{i=1}^{n} x_i P(X = x_i)$$ bei diskreter Zufallsvariable X

$$E(X) = \int_{-\infty}^{\infty} x f(x)\, dx$$ bei stetiger Zufallsvariable X

Varianz [Var(X) oder σ^2]

$$Var(X) = \sum_{i=1}^{n} \left(x_i - E(X)\right)^2 P(X = x_i)$$ bei diskreter Zufallsvariable X

$$Var(X) = \int_{-\infty}^{\infty} \left(x - E(X)\right)^2 f(x)\ dx$$ bei stetiger Zufallsvariable X

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

1.12 Rechenregeln von i.i.d. Zufallsgrößen

Ist $E(X_i) = \mu$ und $Var(X_i) = \sigma^2$, so gilt:

$$E(X1 + X2 + \cdots + Xn) = E(X1) + E(X2) + \ldots + E(Xn) \qquad = n\mu$$
$$Var(X1 + X2 + \cdots + Xn) = Var(X1) + Var(X2) + \ldots + Var(Xn) \qquad = n\sigma^2$$
$$E(\frac{1}{n}(X1 + X2 + \cdots + Xn)) = E(\bar{X}) = \frac{1}{n} E(X1 + X2 + \cdots + Xn) = \frac{1}{n} n\mu \qquad = \mu$$
$$Var(\frac{1}{n}(X1 + X2 + \cdots + Xn)) = Var(\bar{X}) = (\frac{1}{n})^2 \cdot Var(X1 + X2 + \cdots + Xn) = \frac{1}{n^2} \cdot n\sigma 2 \qquad = \frac{\sigma^2}{n}$$

1.13 Standardisierung von Zufallsvariablen

Lineare Transformation der Normalverteilung in Standardnormalverteilung.
Wenn $X \sim N(\mu; \sigma^2)$, dann gilt für die Zufallsgröße Z mit:

$$Z = \frac{X - \mu}{\sigma} \sim N(0; 1)$$

1.14 Verteilungen

Verteilung			E(X)	Var(X)
Bernoulli-Verteilung X~Be(p)	$P(X = 1) = p$ $P(X = 0) = 1 - p$		p	$p \cdot (1 - p)$
Binomial-verteilung X~B(n;p)	$f(x) = \binom{n}{x} \cdot p^x \cdot q^{n-x}$		np	$np(1 - p)$
Stetige Gleichverteilung X~G([a ; b])	$f(x) = \begin{cases} \frac{1}{b-a} & \text{für } a \leq x \leq b \\ 0 & \text{sonst} \end{cases}$	$F(x) = \begin{cases} 0 & \text{für } x \leq a \\ \frac{1}{b-a} \cdot (x - a) & \text{für } a \leq x \leq b \\ 1 & \text{für } x \geq b \end{cases}$	$\frac{a + b}{2}$	$\frac{(b - a)^2}{12}$
Normal-verteilung X~N(μ;σ²)	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-0,5(\frac{x-\mu}{\sigma})^2}$	$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \int_a^b e^{-0,5(\frac{x-\mu}{\sigma})^2} dx$	μ	σ^2
χ^2-Verteilung Y~χ^2(n)		$Y = \sum_{i=1}^{n} X_i^2$	$E(Y) = n$	$Var(Y) = 2n$

1.15 Grenzwertsätze

Gesetz der großen Zahl

$\lim_{n\to\infty} P(|\bar{X} - \mu| \leq \varepsilon) = 1 \quad \forall\varepsilon > 0$; $\lim_{n\to\infty} P(|R - p| \leq \varepsilon) = 1 \quad \forall\varepsilon > 0$

Zentaler Grenzwertsatz

$n \to \infty$: $\bar{X} \overset{as}{\sim} N(\mu; \frac{\sigma^2}{n})$; $n \to \infty$: $R \overset{as}{\sim} N(p; \frac{p(1-p)}{n})$

Standardisierung

Wenn X N(μ;σ²)-verteilt ist, dann ist Z=(X-μ)/σ N(0,1)-verteilt.

Wenn $\bar{X} \overset{as}{\sim} N(\mu; \frac{\sigma^2}{n})$, dann gilt: $\frac{\bar{X} - \mu}{\sigma} \cdot \sqrt{n} \overset{as}{\sim} N(0;1)$

bzw.

Wenn $\bar{X} \sim N(\mu; \frac{\sigma^2}{n})$, dann gilt: $\frac{\bar{X} - \mu}{\sigma} \cdot \sqrt{n} \sim N(0;1)$

2 Datendarstellung

2.1 Lagemaßzahlen

Modus

x_{mod}: Ausprägung mit der größten Häufigkeit

Median
Seien $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$ der Größe nach geordnete Messwerte, so liegt der Median „in der Mitte" dieser Zahlen. Genauer gilt:

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

Arithmetisches Mittel

$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ für einzelne Werte
$\bar{x} = \frac{1}{n}\left(\sum_{j=1}^{k} n_j \cdot x_j\right)$ für k unterschiedliche Ausprägungen
$\bar{x} = \frac{1}{n}\left(\sum_{j=1}^{k} n_j \cdot m_j\right)$ für gruppierte Werte

(Bei Klassenbildung/Gruppenbildung: m_j als Klassenmitte der Klasse j bzw. evtl. der sinnvoll geschlossenen Klasse j)

Geometrisches Mittel

$$\bar{x}_G = \sqrt[n]{\prod_{i=1}^{n} x_i}$$

2.2 Besondere Lagemaßzahlen: Quantile

Gegeben seien die n Daten x_1, x_2,..., x_n in aufsteigender Ordnung: $x^{(1)}, x^{(2)}, \ldots, x^{(n)}$

$$\tilde{x}_p := \begin{cases} x^{(i)}, falls\ np\ nicht\ ganzzahlig\ mit\ der\ kleinsten\ ganzen\ Zahl\ i > np \\ \frac{1}{2}\left(x^{(np)} + x^{(np+1)}\right), falls\ np\ ganzzahlig \end{cases}$$

2.3 Streuungsmaßzahlen

Spannweite (Range)

$$R = x_{max} - x_{min} = x_{(n)} - x_{(1)}$$

Quartilsabstand

$$d_Q = \tilde{x}_{0,75} - \tilde{x}_{0,25}$$

Mittlere Abweichung

$$DAA_a = \frac{1}{n}\sum_{i=1}^{n} |x_i - a|$$

Empirische Varianz S^2 und Standardabweichung S

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

$$S = \sqrt{S^2}$$

Variationskoeffizient (nur für positive Merkmale sinnvoll anwendbar)

$$v = \frac{s}{\bar{x}}$$

3 Mehrdimensionale Merkmale

3.1 Zusammenhangsmaße

Kovarianz S_{XY}

$$s_{XY} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

Korrelationskoeffizient nach Pearson (Voraussetzung: X und Y sind metrische Merkmale)

$$r_{XY} = \frac{s_{XY}}{s_X s_y} = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2}} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

X und Y heißen …
… positiv korreliert, wenn $r_{XY} > 0$
… unkorreliert, wenn $r_{XY} = 0$
… negativ korreliert, wenn $r_{XY} < 0$

Liegen alle Punkte auf einer Geraden, dann gilt $|r_{XY}| = 1$

Rangkorrelationskoeffizient nach Spearman

$$r_{XY}^{S} = r_{UV} = \frac{s_{UV}}{\sqrt{s_U^2 \cdot s_V^2}} = \frac{s_{UV}}{s_U \cdot s_V}$$

mit $U = R(X)$ und $V = R(Y)$

3.2 Regression

Annahme: $y_i = \alpha + \beta x_i + \varepsilon_i$ mit α,β Modellparameter
ε Störterm, Zufallsvariable

Geschätzte $\boldsymbol{y_i}$: $\hat{y}_i = a + bx_i$
Residuen: $e_i = y_i - \hat{y}_i$

Kleinst-Quadrate-Bedingung

$$Q(a;b) = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - a - bx_i)^2 \rightarrow min$$

Kleinst-Quadrate-Schätzer (KQ-Methode)

$a = \bar{y} - b\bar{x}$ und $b = \frac{s_{XY}}{s_X^2}$

Streuungszerlegung

$S_Y^2 = S_{\hat{Y}}^2 + S_e^2$ $\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$

Bestimmtheitsmaß (Anteil der durch das lineare Modell erklärten Varianz) der Regression von Y bzgl. X

$$B = r^2 = \frac{S_{\hat{Y}}^2}{S_Y^2} = r_{XY}^2$$

Geschätzte Standardabweichung

$$s_e = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n} e_i^2}$$

Partielle Regression
Gegeben seien die Merkmale X, Y und Z mit den unten stehenden Regressionsgleichungen. $\mathcal{E}$ und F sind dabei die Residuen der Regressionen von X auf Z bzw. von Y auf Z.

$X = \hat{X} + \mathcal{E} = a + bZ + \mathcal{E}$ $Y = \hat{Y} + \mathrm{F} = c + dZ + \mathrm{F}$

Partieller Korrelationskoeffizient (Korrelation von X und Y bei gegebenem Z)

$$r_{XY/Z} = \frac{r_{XY} - r_{XZ} \cdot r_{YZ}}{\sqrt{1 - r_{XZ}^2} \cdot \sqrt{1 - r_{YZ}^2}}$$

3.3 Kontingenz

Häufigkeiten und Häufigkeitsverteilungen

n: Gesamtzahl der Untersuchungseinheiten

n_{ij}: Häufigkeit der Merkmalskombination (A_i , B_j)
$n_{i\cdot}$: Häufigkeit von Merkmal A_i
$n_{\cdot j}$: Häufigkeit von Merkmal B_j

$f_{ij} = \frac{n_{ij}}{n}$: Relative Häufigkeit der Merkmalskombination (A_i , B_j)

$f_{i\cdot} = \frac{n_{i\cdot}}{n}$: Relative Häufigkeit von Merkmal A_i

$f_{\cdot j} = \frac{n_{\cdot j}}{n}$: Relative Häufigkeit von Merkmal B_j

Bedingte relative Häufigkeiten

$f\left(B_j / A_i\right) := \frac{n_{ij}}{n_{i\cdot}}$ $\qquad$ $f\left(A_i / B_j\right) := \frac{n_{ij}}{n_{\cdot j}}$

Bedingte Verteilungen

Bedingte Verteilung von $Y/X = B_j$
Bedingte Verteilung von $X/Y = A_i$

Unabhängigkeit von Merkmalen

$$\frac{n_{ij}}{n_{i\bullet}} = \frac{n_{\bullet j}}{n} \leftrightarrow \frac{n_{ij}}{n_{\bullet j}} = \frac{n_{i\bullet}}{n} \leftrightarrow n_{ij} = \frac{n_{i\bullet} \cdot n_{\bullet j}}{n} \leftrightarrow \frac{n_{ij}}{n} = \frac{n_{i\bullet}}{n} \cdot \frac{n_{\bullet j}}{n} \qquad \forall\ i \ und\ j$$

3.4 Assoziationsmaße

Quadratische Kontingenz χ^2

$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(n_{ij}-e_{ij})^2}{e_{ij}}$ Wertebereich: $0 \leq \chi^2 \leq n \cdot \min(r-1, s-1)$

Mittlere quadratische Kontingenz ϕ^2

$\phi^2 = \frac{\chi^2}{n}$ Wertebereich: $0 \leq \chi^2 \leq \min(r-1, s-1)$

ϕ-Koeffizient

$\phi = \sqrt{\frac{\chi^2}{n}}$ Wertebereich: $0 \leq \phi \leq \min(\sqrt{r-1}, \sqrt{s-1})$

Kontingenzmaß nach Tschuprov

$$T = \sqrt{\frac{\chi^2}{n \cdot \sqrt{(r-1)(s-1)}}}$$

Kontingenzmaß nach Cramér

$V = \sqrt{\frac{\chi^2}{n \cdot \min(r-1, s-1)}}$ Wertebereich: $0 \leq V \leq 1$

Kontingenzkoeffizient von Pearson C

$C = \sqrt{\frac{\chi^2}{\chi^2 + n}}$ Wertebereich: $0 \leq C \leq \sqrt{\frac{\min(r,s)-2}{\min(r,s)-1}}$

Normierter Kontingenzkoeffizient C_{Corr}

$C_{Corr} = \sqrt{\frac{\min(r,s)-1}{\min(r,s)-2} \cdot \frac{\chi^2}{\chi^2+n}}$ Wertebereich: $0 \leq C_{Corr} \leq 1$

4 Inferenzstatistik

4.1 Schätzfunktionen

(a) für den Mittelwert μ:

$\bar{X}$ als Schätzer für den Mittelwert

$$E(\bar{X}) = E\left[\frac{1}{n}(X_1 + X_2 + \cdots + X_n)\right] = \mu$$

$$Var(\bar{X}) = Var\left[\frac{1}{n}(X_1 + X_2 + \cdots + X_n)\right] = \frac{\sigma^2}{n}$$

$$STE(\bar{X}) = \sqrt{Var(\bar{X})}$$

(b) für den Anteilswert p:

R als Schätzer für den Anteilswert.

Es gilt P(X_i = 1) = p und P(X_i = 0) = 1-p mit X_1, X_2, …, X_n i.i.d.

$$E(X) = p$$

$$Var(X) = p(1-p)$$

$$R = \frac{1}{n}(X_1 + X_2 + \cdots + X_n) = \bar{X}$$

für $np(1-p) > 9$ ist $R \underset{\sim}{as} N(p; \frac{p(1-p)}{n})$

4.2 Konfidenzintervalle

Standardisierung und $\frac{1+\gamma}{2}$-Quantil der Standardnormalverteilung

Konfidenzintervall für den Mittelwert μ bei Normalverteilung und bekannter σ^2

Annahme: X ist normalverteilt und es gilt $\bar{X} \sim N\left(\mu; \frac{\sigma^2}{n}\right)$

$$Z = \frac{\bar{X} - \mu}{\sigma} \cdot \sqrt{n} \sim N(0;1)$$

$$P(z_{\frac{1-\gamma}{2}} \leq Z \leq z_{\frac{1+\gamma}{2}}) = P(-z_{\frac{1+\gamma}{2}} \leq Z \leq z_{\frac{1+\gamma}{2}}) = \gamma$$

Es ergibt sich: $P\left(\bar{X} - z_{\frac{1+\gamma}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\frac{1+\gamma}{2}} \frac{\sigma}{\sqrt{n}}\right) = \gamma$

Konfidenzintervall für μ zum Sicherheitsniveau γ: $[\bar{X} - z_{\frac{1+\gamma}{2}} \frac{\sigma}{\sqrt{n}} ; \bar{X} + z_{\frac{1+\gamma}{2}} \frac{\sigma}{\sqrt{n}}]$

mit $z_{\frac{1+\gamma}{2}}$ ist das $\frac{1+\gamma}{2}$-Quantil der Standardnormalverteilung

Konfidenzintervall für den Mittelwert μ bei Normalverteilung und unbekannter Varianz σ^2 sowie kleiner Stichprobe

Annahme: X ist normalverteilt

Dann ist die Zufallsgröße $T = \frac{\bar{X}-\mu}{S} \cdot \sqrt{n}$ t-verteilt mit n-1 Freiheitsgraden

Konfidenzintervall für μ zum Sicherheitsniveau γ lautet: $[\bar{X} - t_{\left(n-1;\frac{1+\gamma}{2}\right)} \frac{S}{\sqrt{n}} ; \bar{X} + t_{\left(n-1;\frac{1+\gamma}{2}\right)} \frac{S}{\sqrt{n}}]$

mit $t_{(n-1;\frac{1+\gamma}{2})}$ als $\frac{1+\gamma}{2}$-Quantil der t-Verteilung mit n-1 Freiheitsgraden

Konfidenzintervall für den Mittelwert μ bei Normalverteilung und unbekannter Varianz σ^2 sowie großer Stichprobe

Annahme: X ist normalverteilt und n groß (Faustregel: $n \geq 30$)

Dann ist die Zufallsgröße $Z = \frac{\bar{X}-\mu}{S} \cdot \sqrt{n}$ approximativ standardnormalverteilt.

Konfidenzintervall für μ zum Sicherheitsniveau γ lautet: $[\bar{X} - z_{\frac{1+\gamma}{2}} \frac{S}{\sqrt{n}} ; \bar{X} + z_{\frac{1+\gamma}{2}} \frac{S}{\sqrt{n}}]$

mit $z_{\frac{1+\gamma}{2}}$ ist das $\frac{1+\gamma}{2}$-Quantil der Standardnormalverteilung

Approximatives Konfidenzintervall für den Anteilswert p bei hinreichend großem Stichprobenumfang (Faustregel: $n \geq 30$)

$$[R - z_{\frac{1+\gamma}{2}} \cdot \sqrt{\frac{R \cdot (1-R)}{n}} \ ; \ R + z_{\frac{1+\gamma}{2}} \cdot \sqrt{\frac{R \cdot (1-R)}{n}}]$$

mit $z_{\frac{1+\gamma}{2}}$ ist das $\frac{1+\gamma}{2}$-Quantil der Standardnormalverteilung

Approximatives Konfidenzintervall für die Erwartungswertdifferenz ($\mu_X - \mu_Y$) bei i.i.d. Stichproben, unbekannten Varianzen und jeweils hinreichend großen Stichprobenumfängen (Faustregel: $m \geq 30$ und $n \geq 30$)

Es gilt: $Var(\bar{X} - \bar{Y}) = Var(\bar{X}) + Var(\bar{Y}) = \frac{\sigma_X^2}{m} + \frac{\sigma_Y^2}{n}$

Schätzung der Varianz: $\widehat{Var(\bar{X} - \bar{Y})} = \frac{S_X^2}{m} + \frac{S_Y^2}{n}$

$$[(\bar{X} - \bar{Y}) - z_{\frac{1+\gamma}{2}} \cdot S_d \ ; \ (\bar{X} - \bar{Y}) + z_{\frac{1+\gamma}{2}} \cdot S_d]$$

mit $S_d = \sqrt{\frac{S_X^2}{m} + \frac{S_Y^2}{n}}$

mit $z_{\frac{1+\gamma}{2}}$ ist das $\frac{1+\gamma}{2}$-Quantil der Standardnormalverteilung

Konfidenzintervall für die Erwartungswertdifferenz ($\mu_X - \mu_Y$) bei i.i.d. Stichproben, unbekannten, aber gleichen Varianzen ($\sigma_X^2 = \sigma_Y^2 = \sigma^2$) und Normalverteilungsannahme

Es gilt: $Var(\bar{X} - \bar{Y}) = Var(\bar{X}) + Var(\bar{Y}) = \frac{\sigma^2}{m} + \frac{\sigma^2}{n} = \sigma^2 \cdot (\frac{1}{m} + \frac{1}{n})$

Erwartungstreuer Punktschätzer für σ^2: $S_{X,Y}^2 = \frac{1}{m+n-2} \cdot \left[\sum_{i=1}^{m}(X_i - \bar{X})^2 + \sum_{j=1}^{n}(Y_i - \bar{Y})^2\right]$

Dann gilt für $U = \frac{(\bar{X}-\bar{Y})-(\mu_X-\mu_Y)}{S_{X,Y} \cdot \sqrt{\frac{1}{m}+\frac{1}{n}}} \sim$ t(m+n-2)

Das Konfidenzintervall für $\mu_X - \mu_Y$ zum Sicherheitsniveau γ lautet damit:

$$\left[(\bar{X} - \bar{Y}) - t_{\left(m+n-2;\frac{1+\gamma}{2}\right)} \cdot S_{X,Y} \cdot \sqrt{\frac{1}{m} + \frac{1}{n}} \ ; \ (\bar{X} - \bar{Y}) + t_{\left(m+n-2;\frac{1+\gamma}{2}\right)} \cdot S_{X,Y} \cdot \sqrt{\frac{1}{m} + \frac{1}{n}}\right]$$

mit $t_{\left(m+n-2;\frac{1+\gamma}{2}\right)}$ ist das $\frac{1+\gamma}{2}$-Quantil der t-Verteilung mit m+n-2 Freiheitsgraden

Approximatives Konfidenzintervall für die Anteilswertdifferenz bei zwei voneinander stochastisch unabhängigen i.i.d.-Stichproben für $(p_X - p_Y)$ und jeweils hinreichend großen Stichprobenumfängen (Faustregel: $m \geq 30$ und $n \geq 30$) für das Sicherheitsniveau γ

$$[(R_X - R_Y) - z_{\frac{1+\gamma}{2}} \cdot S_d\ ;\ (R_X - R_Y) + z_{\frac{1+\gamma}{2}} \cdot S_d]$$

mit $S_d = \sqrt{\frac{R_X \cdot (1-R_X)}{m} + \frac{R_Y \cdot (1-R_Y)}{n}}$

und mit $z_{\frac{1+\gamma}{2}}$ ist das $\frac{1+\gamma}{2}$-Quantil der Standardnormalverteilung

Konfidenzintervall für die Erwartungswertdifferenz $(\mu_W = \mu_X - \mu_Y)$ bei hinreichend großem Stichprobenumfang (Faustregel: $n \geq 30$) und bei verbunden Stichproben für das Sicherheitsniveau γ

$$[\bar{W} - z_{\frac{1+\gamma}{2}} \cdot \frac{S_W}{\sqrt{n}}\ ;\ \bar{W} + z_{\frac{1+\gamma}{2}} \cdot \frac{S_W}{\sqrt{n}}]$$

mit $W_i = X_i - Y_i$ und $S_W^2 = \frac{1}{n-1} \sum_{i=1}^{n} (W_i - \bar{W})^2$

und mit $z_{\frac{1+\gamma}{2}}$ ist das $\frac{1+\gamma}{2}$-Quantil der Standardnormalverteilung

4.3 Statistische Testverfahren

Einseitiger approximativer Test auf den Anteilswert p zum Signifikanzniveau α

Als Faustregel muss gelten: $n \cdot p_0 \cdot (1 - p_0) > 9$

$H_0\colon\ p \leq p_0$
$H_1\colon\ p > p_0$

Ablehnungsbereich: $\left[p_0 + z_{1-\alpha} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}\ ; 1\right]$

$H_0\colon\ p \geq p_0$
$H_1\colon\ p < p_0$

Ablehnungsbereich: $\left[0\ ;\ p_0 - z_{1-\alpha} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}\right]$

mit $z_{1-\alpha}$ ist das $1 - \alpha$-Quantil der Standardnormalverteilung

Zweiseitiger approximativer Test auf den Anteilswert p zum Signifikanzniveau α

Als Faustregel muss gelten: $n \cdot p_0 \cdot (1 - p_0) > 9$

$H_0\colon\ p = p_0$
$H_1\colon\ p \neq p_0$

Annahme- oder Nichtablehnungsbereich: $\left[p_0 - z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}\ ;\ p_0 + z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{p_0(1-p_0)}{n}}\right]$

mit $z_{1-\frac{\alpha}{2}}$ ist das $1 - \frac{\alpha}{2}$-Quantil der Standardnormalverteilung

(Zweiseitiger) Test auf Erwartungswert (t-Test / Gauß-Test) zum Signifikanzniveau α

H_o: $\mu = \mu_0$
H_1: $\mu \neq \mu_0$
Testgröße $T = \frac{\bar{X}-\mu_0}{S} \cdot \sqrt{n}$
H_o wird abgelehnt, falls $|T| > t(n-1; 1-\frac{\alpha}{2})$
mit $t(n-1; 1-\frac{\alpha}{2})$ ist das $1-\frac{\alpha}{2}$-Quantil der t-Verteilung mit n-1 Freiheitsgraden
oder alternativ mit Testgröße $\bar{X}$
H_o wird beibehalten, falls $\bar{X} \in \left[\mu_0 - t\left(n-1; 1-\frac{\alpha}{2}\right) \cdot \frac{s}{\sqrt{n}} \; ; \; \mu_0 + t\left(n-1; 1-\frac{\alpha}{2}\right) \cdot \frac{s}{\sqrt{n}}\right]$

Zu beachten:
Für $n \leq 30$ ist der Test nur gültig, wenn $X \sim N(\mu; \sigma^2)$
Für $n > 30$ (und Verteilung, die nicht zu Ausreißern neigt) darf das $1-\frac{\alpha}{2}$-Quantil der t-Verteilung mit n-1 Freiheitsgraden durch $z_{1-\frac{\alpha}{2}}$, also das $1-\frac{\alpha}{2}$-Quantil der Standardnormalverteilung, ersetzt werden

(Einseitiger) Test auf Erwartungswert (t-Test / Gauß-Test) zum Signifikanzniveau α

H_o: $\mu \leq \mu_0$
H_1: $\mu > \mu_0$
Testgröße $T = \frac{\bar{X}-\mu_0}{S} \cdot \sqrt{n}$
H_o wird abgelehnt, falls $T > t(n-1; 1-\alpha)$
mit $t(n-1; 1-\alpha)$ ist das $1-\alpha$-Quantil der t-Verteilung mit n-1 Freiheitsgraden
oder alternativ mit Testgröße $\bar{X}$
H_o wird abgelehnt, falls $\bar{X} \in [\mu_0 + t(n-1; 1-\alpha) \cdot \frac{s}{\sqrt{n}}; \infty)$

H_o: $\mu \geq \mu_0$
H_1: $\mu < \mu_0$
Testgröße $T = \frac{\bar{X}-\mu_0}{S} \cdot \sqrt{n}$
H_o wird abgelehnt, falls $T <$ -$t(n-1; 1-\alpha)$
mit $t(n-1; 1-\alpha)$ ist das $1-\alpha$-Quantil der t-Verteilung mit n-1 Freiheitsgraden
oder alternativ mit Testgröße $\bar{X}$
H_o wird abgelehnt, falls $\bar{X} \in (-\infty \; ; \; \mu_0 - t(n-1; 1-\alpha) \cdot \frac{s}{\sqrt{n}}]$

Zu beachten:
Für $n \leq 30$ ist der Test nur gültig, wenn $X \sim N(\mu; \sigma^2)$
Für $n > 30$ (und Verteilung, die nicht zu Ausreißern neigt) darf das $1-\alpha$-Quantil der t-Verteilung mit n-1 Freiheitsgraden durch $z_{1-\alpha}$, also das $1-\alpha$-Quantil der Standardnormalverteilung, ersetzt werden

Zweiseitiger doppelter t-Test auf Erwartungswertdifferenz zum Signifikanzniveau α

Voraussetzung: zwei unabhängige Stichproben, normalverteilte Zufallsgrößen X und Y, Varianzen gleich (aber unbekannt)

H_o: μ_x-$\mu_y = d_0$ $\quad \left(d_0 = 0 \; entspricht \; \mu_x = \mu_y\right)$
H_1: μ_x-$\mu_y \neq d_0$ $\quad \left(d_0 \neq 0 \; entspricht \; \mu_x \neq \mu_y\right)$

Annahme- bzw. Nichtablehnungsbereich:

$[d_o - t_{\left(m+n-2\,;\,1-\frac{\alpha}{2}\right)} \cdot S_d \,;\; d_o + t_{\left(m+n-2\,;\,1-\frac{\alpha}{2}\right)} \cdot S_d]$
mit $S_d = S_{X,Y} \cdot \sqrt{\frac{1}{m} + \frac{1}{n}}$ und $S_{X,Y} = \frac{1}{m+n-2} [\sum_{i=1}^{m}(x_i - \bar{x})^2 + \sum_{j=1}^{n}\left(y_j - \bar{y}\right)^2]$

χ^2-Unabhängigkeitstest zum Signifikanzniveau α

Dieser Test prüft die Hypothese, ob zwei kategoriale Merkmale stochastisch unabhängig sind oder nicht.

H_0: $p_{ij} = p_{i\cdot} \cdot p_{\cdot j}$
H_1: $p_{ij} \neq p_{i\cdot} \cdot p_{\cdot j}$

Teststatistik: χ^2

$$\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \qquad \text{mit} \quad e_{ij} = \frac{n_{i\cdot} \cdot n_{\cdot j}}{n}$$

Faustregel: alle $n_{ij} \geq 5$

H_0 wird abgelehnt, falls χ^2 größer dem (1-α)-Quantil der χ^2-Verteilung mit (k-1) · (l-1) Freiheitsgraden.