Formelsammlung

L1-L3 - Datentabellen

Anteile

$$p_{(k)} = \frac{n_{(k)}}{n} \text{ bzw. } p_x = \frac{n_x}{n}$$

Kumulierten Anteile

$$cp_{(k)} = p\left(X \leq x_{(k)}\right) = \sum_{j=1}^{k} p_{(j)} = \frac{\sum_{j=1}^{k} n_{(j)}}{n}$$

Quantile

Rangreihe/Häufigkeitstabelle

$$i = n \cdot \alpha$$

Gruppierte Daten

$$Q_\alpha = o_{(k-1)} + \frac{\alpha - cp_{(k-1)}}{p_{(k)} / \left(o_{(k)} - o_{(k-1)}\right)} = o_{(k-1)} + \frac{\alpha - cp_{(k-1)}}{p_{(k)}} \cdot \left(o_{(k)} - o_{(k-1)}\right)$$

LE 5 – Lagemaße (ordinal)

Median (gerade Fallzahl)

$$\tilde{x} = \frac{x_{(n/2)} + x_{(n/2+1)}}{2}$$

Median (ungerade Fallzahl)

$$\tilde{x} = x_{\left(\frac{n+1}{2}\right)}$$

Median (gruppierte Daten)

$$\tilde{x} = o_{(k-1)} + \frac{0.5 - cp_{(k-1)}}{p_{(k)}} \cdot \left(o_{(k)} - o_{(k-1)}\right)$$

LE 5 – Lagemaße (metrisch)

Mittelwert (Rangreihe)

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

Mittelwert (Häufigkeitstabelle)

$$\overline{x} = \frac{1}{n}\sum_{k=1}^{K} n_k \cdot x_k$$

Mittelwert (gruppierte Daten)

$$\overline{x} = \frac{1}{n}\sum_{k=1}^{K} n_k \cdot m_k \qquad m_{(k)} = \frac{u_{(k)} + o_{(k)}}{2}$$

Geometrische Mittel

$$\overline{x}_{geom.} = \sqrt[n]{\prod_{i=1}^{n} x_i} = \exp\left(\frac{1}{n} \cdot \sum_{i=1}^{n} \ln\left(x_i\right)\right)$$

Getrimmte Mittel (gruppierte Daten)

$$u_t = u_{(1)} + \frac{\alpha_{(1)}}{p_{(1)}} \cdot \left(o_{(1)} - u_{(1)}\right)$$

$$o_t = o_{(K)} - \frac{\alpha_t}{p_{(K)}} \cdot \left(o_{(K)} - u_{(K)}\right)$$

L6 – Streuungsmaße (nominal)

Absolute Devianz

$$D_X = -2\sum_{k=1}^{K} n_k \cdot \ln\left(\frac{n_k}{n}\right) = -2\sum_{k=1}^{K} n_k \cdot \ln(p_k)$$

Relative Devianz

$$d_X = -2\sum_{k=1}^{K} p_k \cdot \ln(p_k) = \frac{D_x}{n}$$

Index Qualitativer Variation

$$IQV = \frac{K}{K-1} \cdot \left(1 - \sum_{k=1}^{K} p_k^2\right)$$

L6 – Streuungsmaße (metrisch)

Range

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

Quartilabstand

$$IQR = Q_{0.75} - Q_{0.25}$$

Mittlere Quartilabstand

$$mIQR = \frac{Q_{0.75} - Q_{0.25}}{2}$$

Durchschnittliche Abweichung

$$AD = \frac{1}{n}\sum_{i=1}^{n} \left|x_i - \overline{x}\right|$$

Variation (Häufigkeitstabelle)

$$\begin{aligned} SS_X &= \sum_{i=1}^{n} \left(x_i^2 + \overline{x}^2 - 2 \cdot x_i \cdot \overline{x}\right) \\ &= \sum_{i=1}^{n} x_i^2 + n \cdot \overline{x}^2 - 2 \cdot \overline{x} \cdot \sum_{i=1}^{n} x_i \\ &= \sum_{i=1}^{n} x_i^2 + n \cdot \overline{x}^2 - 2 \cdot \overline{x} \cdot (n \cdot \overline{x}) = \sum_{i=1}^{n} x_i^2 - n \cdot \overline{x}^2 \end{aligned}$$

Variation (Rangreihe)

$$SS_X = \sum_{i=1}^{n} (x_i - \overline{x})^2$$

L6 – Streuungsmaße (metrisch)

Varianz (Rangreihe)

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

Varianz (Häufigkeitstabelle)

$$s_X^2 = \frac{1}{n}\cdot\sum_{k=1}^{K} n_k \cdot x_k^2 - \bar{x}^2 = \frac{\sum_{k=1}^{K} n_k \cdot x_k^2}{n} - \frac{\left(\sum_{k=1}^{K} n_k \cdot x_k\right)^2}{n^2} = \frac{SS_X}{n}$$

Standardabweichung

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

Variationskoeffizient

$$V_X = \frac{s_X}{\bar{x}} = \frac{\sqrt{s_X^2}}{\bar{x}} = \frac{\sqrt{SS_X}}{\sqrt{n}\cdot\bar{x}}$$

L8 - Wahrscheinlichkeitstheorie

Additionstheorem: $Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$

Bedingte Wahrscheinlichkeit =: $Pr(A|B) = Pr(A \cap B) / Pr(B)$

z.B.:Die Wahrscheinlichkeit bei der zweiten Auswahl HH1 auszuwählen, wenn bei der 1. Auswahl HH 2 ausgewählt wurde beträgt

$Pr(C|D) = Pr(C \cap D) / Pr(D) = (1/36)/(1/6) = 1/6$

Multiplikationstheorem: $Pr(A \cap B) = Pr(A|B) \cdot Pr(B) = Pr(B|A) \cdot Pr(A)$

bei statistischer Unabhängigkeit: $Pr(A \cap B) = Pr(A) \cdot Pr(B)$

***Satz von Bayes**:* Austausch von bedingter und bedingender Wahrscheinlichkeit:

$$Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)} = \frac{Pr(B|A)\cdot Pr(A)}{Pr(B)} = \frac{Pr(B|A)\cdot Pr(A)}{Pr(B|A)\cdot Pr(A) + Pr(B|\neg A)\cdot Pr(\neg A)}$$

L9 – Kombinatorik und Variation

Mit Zurücklegen

$$mit\ Reihenfolge = {}_N V_n^{\,m} = N^n$$

$$ohne\ Reihenfolge = {}_N K_n^{\,m} = \frac{(N+n-1)!}{(N-1)! * n!}$$

Ohne Zurücklegen

$$mit\ Reihenfolge = {}_N V_n^{\,o} = \frac{N!}{(N-n)!}$$

$$ohne\ Reihenfolge = {}_N K_n^{\,o} = \frac{N!}{(N-n)! * n!}$$

L11 - Binomialverteilung

Binomialverteilung

$$b(X; n, \pi_1) = \Pr(X = x | n, \pi_1) = \binom{n}{x} \cdot \pi_1^x \cdot (1 - \pi_1)^{n-x} \quad \text{mit } x = 0, 1, 2, \ldots, n$$

Binomialkoeffizient

$$\binom{N}{x} = \frac{N!}{(N-x)! \cdot x!}$$

L11 – Hypergeometrische Verteilung

Hypergeometrische Verteilung

$$\Pr\left(X = n_1\right) = \frac{\binom{N_1}{n_1} \cdot \binom{N-N_1}{n-n_1}}{\binom{N}{n}} = \frac{\frac{N_1!}{n_1! \cdot (N_1 - n_1)!} \cdot \frac{N_0\,!}{n_0\,!(N_0 - n_o)!}}{\frac{N!}{n! \cdot (N-n)!}}$$

L11 – Pascal- und Poissonverteilung

Pascalverteilung

$$\Pr\left(X = x \middle| k, \pi_1\right) = \binom{x-1}{k-1} \cdot \pi_1^k \cdot \left(1-\pi_1\right)^{x-k} \text{ mit } x = k, k+1, k+2, \ldots$$

$$\mu_X = \frac{k}{\pi_1} \text{ und } \sigma_X^2 = k \cdot \frac{1-\pi_1}{\pi_1^2}$$ Mittelwert und Varianz der Pascalverteilung

Poissonverteilung

$$p\left(X;\lambda\right) = \Pr\left(X = x \middle| p\left(X;\lambda\right)\right) = \frac{\lambda^x}{x!} \cdot e^{-\lambda} \text{ mit x=0,1,2,...,}\infty$$

L12 - Normalverteilung

Asymptotische Annäherung an die Normalverteilung (bei zu großen Fakultäten)

Binomialverteilung

$$\Pr\left(X = x \middle| b(X;n,\pi_1)\right) = \binom{n}{x}\cdot \pi_1^x \cdot (1-\pi_1)^{n-x} \approx \Phi\left(\frac{x+0.5-n\cdot\pi_1}{\sqrt{n\cdot\pi_1\cdot(1-\pi_1)}}\right) - \Phi\left(\frac{x-0.5-n\cdot\pi_1}{\sqrt{n\cdot\pi_1\cdot(1-\pi_1)}}\right)$$

Hypergeometrischen Verteilung

$$\rightarrow \Pr(x) = \frac{\binom{N_1}{x}\cdot\binom{N-N_1}{n-x}}{\binom{N}{n}} \approx \Phi\left(\frac{x+0.5-n\cdot\frac{N_1}{N}}{\sqrt{n\cdot\frac{N_1}{N}\cdot\frac{N-N_1}{N}\cdot\frac{N-n}{N-1}}}\right) - \Phi\left(\frac{x-0.5-n\cdot\frac{N_1}{N}}{\sqrt{n\cdot\frac{N_1}{N}\cdot\frac{N-N_1}{N}\cdot\frac{N-n}{N-1}}}\right)$$

L13 – Schätzen und Testen

Umwandlung: Stichprobenvarianz <-> Geschätzte Populationsvarianz

$$\hat{\sigma}_X^2 = \frac{n}{n-1}\cdot s_X^2 = \frac{SS_X}{n-1} = \frac{1}{n-1}\cdot\sum_{i=1}^{n}(x_i-\bar{x})^2 \qquad \hat{\sigma}_X = \sqrt{\hat{\sigma}_X^2} = \sqrt{\frac{1}{n-1}\cdot\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

$$\text{mit } \mu\left(\hat{\sigma}_X^2\right) = \sigma_X^2 \qquad \text{aber: } \mu\left(\hat{\sigma}_X\right) \neq \sigma_X$$

Standardfehler (Mittelwert)

$$\hat{\sigma}_{\bar{X}} = \begin{cases} \dfrac{\hat{\sigma}_X}{\sqrt{n}} = \sqrt{\dfrac{SS_X}{n\cdot(n-1)}} & \text{mit Zurücklegen} \\ \dfrac{\hat{\sigma}_X}{\sqrt{n}}\cdot\sqrt{\dfrac{N-n}{N-1}} = \sqrt{\dfrac{SS_X}{n\cdot(n-1)}\cdot\dfrac{N-n}{N-1}} & \text{ohne Zurücklegen} \end{cases}$$

Standardfehler (Anteile)

$$\sigma(p_1) = \begin{cases} \sqrt{\dfrac{\pi_1\cdot(1-\pi_1)}{n}} \\ \sqrt{\dfrac{\pi_1\cdot(1-\pi_1)}{n}\cdot\dfrac{N-n}{N-1}} \end{cases} \text{bzw. } \hat{\sigma}(p_1) = \begin{cases} \sqrt{\dfrac{p_1\cdot(1-p_1)}{n}} & \text{mit Zurücklegen} \\ \sqrt{\dfrac{p_1\cdot(1-p_1)}{n}\cdot\dfrac{N-n}{N-1}} & \text{ohne Zurücklegen} \end{cases}$$

L13 + L14 – Schätzen und Testen

Teststatistiken

$$Z = \frac{X - \bar{x}}{s_X} \qquad Z = \frac{\bar{X} - \mu_X}{\hat{\sigma}_x / \sqrt{n}} \qquad T = \frac{\bar{X} - \mu_X}{\hat{\sigma}_x / \sqrt{n}} \qquad Z = \frac{p_1 - \pi_1}{\sqrt{\frac{\pi_1 \cdot (1 - \pi_1)}{n}}}$$

Konfidenzintervall (Anteil)

$$\text{c.i.}(\pi_1) = p_1 \pm \sqrt{\frac{p_1 \cdot (1 - p_1)}{n}} \cdot z_{\alpha/2} = p_1 \pm \sqrt{\frac{p_1 \cdot (1 - p_1)}{n}} \cdot z_{1-\alpha/2}$$

Konfidenzintervall (Mittelwert)

$$\text{c.i.}(\mu_X) = \bar{x} \pm \hat{\sigma}(\bar{x}) \cdot z_{1-\alpha/2} = \bar{x} \pm \sqrt{\frac{SS_X}{n \cdot (n-1)}} \cdot z_{1-\alpha/2} = \bar{x} \pm \frac{\hat{\sigma}_X}{\sqrt{n}} \cdot z_{1-\alpha/2} = \bar{x} \pm \frac{s_X}{\sqrt{n-1}} \cdot z_{1-\alpha/2}$$

L15 - Kreuztabellen

Anteile

$$p_{i(j)} = \frac{n_{ij}}{n_{\bullet j}} \text{ bzw. } p_{i(j)}\% = 100 \cdot p_{i(j)}$$

Prozentsatzdifferenz

$$d_{YX}\% = 100 \cdot \left(p_{1(1)} - p_{1(2)}\right) = 100 \cdot \left(\frac{n_{11}}{n_{\bullet 1}} - \frac{n_{12}}{n_{\bullet 2}}\right) = 100 \cdot \left(\frac{a}{a+c} - \frac{b}{b+d}\right)$$

Konfidenzintervall

$$\text{c.i.}(\delta_{YX}\%) = d_{YX}\% \pm z_{1-\alpha/2} \cdot \hat{\sigma}(d_{YX}\%) = d_{YX}\% \pm z_{1-\alpha/2} \cdot 100 \cdot \sqrt{\frac{a \cdot c}{(a+c)^3} + \frac{b \cdot d}{(b+d)^3}}$$

L15 - Kreuztabellen

Standardfehler

$$\hat{\sigma}\left(d_{YX}\%\right) = 100 \cdot \sqrt{\frac{p_{1(1)} \cdot p_{2(1)}}{n_{\bullet 1}} + \frac{p_{1(2)} \cdot p_{2(2)}}{n_{\bullet 2}}} = 100 \cdot \sqrt{\frac{\frac{a}{a+c} \cdot \frac{c}{a+c}}{a+c} + \frac{\frac{b}{b+d} \cdot \frac{d}{b+d}}{b+d}}$$

$$= 100 \cdot \sqrt{\frac{a \cdot c}{(a+c)^3} + \frac{b \cdot d}{(b+d)^3}}$$

Teststatistik

$$Z = \frac{d_{YX}\% - d\%}{\hat{\sigma}\left(d_{YX}\%\right)} = \frac{100 \cdot \left(\frac{a}{a+c} - \frac{b}{b+d}\right) - d\%}{100 \cdot \sqrt{\frac{a \cdot c}{(a+c)^3} \cdot \frac{b \cdot d}{(b+d)^3}}} = \frac{\left(\frac{a}{a+c} - \frac{b}{b+d}\right) - \frac{d\%}{100}}{\sqrt{\frac{a \cdot c}{(a+c)^3} \cdot \frac{b \cdot d}{(b+d)^3}}}$$

L15 - Kreuztabellen

Standardfehler (wenn gegen 0 geprüft wird)

$$\hat{\sigma}\left(Z \middle| \delta_{YX}\% = 0\right) = \sqrt{\frac{p_{1\bullet} \cdot p_{2\bullet}}{n_{\bullet 1}} + \frac{p_{1\bullet} \cdot p_{2\bullet}}{n_{\bullet 2}}} = \sqrt{p_{1\bullet} \cdot p_{2\bullet} \cdot \left(\frac{1}{n_{\bullet 1}} + \frac{1}{n_{\bullet 2}}\right)}$$

$$= \sqrt{\frac{(a+b)}{n} \cdot \frac{(c+d)}{n} \cdot \left(\frac{1}{a+c} + \frac{1}{b+d}\right)}$$

Teststatistik

$$Z = \frac{p_{1(1)} - p_{2(1)}}{\sqrt{p_{1\bullet} \cdot p_{2\bullet} \cdot \left(\frac{1}{n_{\bullet 1}} + \frac{1}{n_{\bullet 2}}\right)}} = \frac{\frac{a}{a+c} - \frac{b}{b+d}}{\sqrt{\frac{(a+b) \cdot (c+d)}{n^2} \cdot \left(\frac{1}{a+c} + \frac{1}{b+d}\right)}}$$

L16 + 17 – (A)Symmetrische Zusammenhänge (4-Felder-Tabelle)

Asymmetrisches Zusammenhangsmaß Prozentsatzdifferenz

$$d_{YX}\% = 100 \cdot \left(\frac{a}{a+c} - \frac{b}{b+d} \right) \; ; \; \hat{\sigma}_{d_{YX}\%} = 100 \cdot \sqrt{\frac{a \cdot c}{(a+c)^3} + \frac{b \cdot d}{(b+d)^3}}$$

Symmetrische Zusammenhangsmaße Chi², Phi², Phi, Yules Q

$$\chi^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \qquad \Phi^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{\left(p_{ij} - \frac{e_{ij}}{n} \right)^2}{\frac{e_{ij}}{n}} = \chi^2 / n \qquad Q = \frac{a \cdot d - b \cdot c}{a \cdot d + b \cdot c}$$

$$\Phi^2 = \frac{(a \cdot d - b \cdot c)^2}{(a+b) \cdot (c+d) \cdot (a+c) \cdot (b+d)} \text{ bzw. } \chi^2 = n \cdot \frac{(a \cdot d - b \cdot c)^2}{(a+b) \cdot (c+d) \cdot (a+c) \cdot (b+d)}$$

Teststatistik Z

$$Z = \Phi \cdot \sqrt{n}$$

L17 – Zusammenhangsmaße Mehrfeldertabelle (nominal)

Symmetrische Zusammenhangsmaße

Cramers V

$$V = \sqrt{\frac{\chi^2}{\chi^2_{max}}} = \sqrt{\frac{\chi^2}{n \cdot \min(I-1, J-1)}}$$

Lambda

$$\lambda_{YX} = 1 - \frac{E_1}{E_0} = 1 - \frac{\sum_{j=1}^{J} \left(n_{\bullet j} - \max_i (n_{ij}) \right) \text{ in Spalte } j}{n_{\bullet\bullet} - \max_i (n_{i\bullet})}$$

Asymmetrisches Zusammenhangsmaß (Relative Devianzreduktion)

$$\rightarrow R'_{YX} = 1 - \frac{D_{YX}}{D_Y} = 1 - \frac{-2 \sum_{j=1}^{J} \sum_{i=1}^{I} n_{ij} \cdot \ln \left(\frac{n_{ij}}{n_{\bullet j}} \right)}{-2 \sum_{i=1}^{I} n_{i\bullet} \cdot \ln \left(\frac{n_{i\bullet}}{n_{..}} \right)}$$

Teststatistik

$$L^2 = D_Y - D_{YX} = 2 \cdot \sum_{i=1}^{I} \sum_{j=1}^{J} n_{ij} \cdot \ln \left(\frac{n_{ij}}{e_{ij}} \right)$$

L18 – Symmetrische Zusammenhangsmaße (metrisch)

Kovariation und Kovarianz

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

$$s_{XY} = \frac{1}{n}\cdot\sum_{i=1}^{n}(x_i - \bar{x})\cdot(y_i - \bar{y}) = \frac{1}{n}\cdot\sum_{i=1}^{n} x_i \cdot y_i - \bar{x}\cdot\bar{y} = \frac{1}{n}\cdot\sum_{i=1}^{n} x_i \cdot y_i - \frac{\left(\sum_{i=1}^{n} x_i\right)\cdot\left(\sum_{i=1}^{n} y_i\right)}{n^2}$$

Umwandlung: Stichprobenkovarianz <-> geschätzte Populationskovarianz

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

L18 – Symmetrische Zusammenhangsmaße (metrisch)

Korrelation

$$r_{XY} = \frac{SP_{XY}}{\sqrt{SS_X \cdot SS_Y}} = \frac{s_{XY}}{\sqrt{s_X^2 \cdot s_Y^2}} = \frac{s_{XY}}{s_X \cdot s_Y} = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i \cdot y_i - \bar{x}\cdot\bar{y}}{\sqrt{\left(\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2\right)\cdot\left(\frac{1}{n}\sum_{i=1}^{n} y_i^2 - \bar{y}^2\right)}}$$

$$= \frac{\sum_{i=1}^{n}(x_i - \bar{x})\cdot(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot \sum_{i=1}^{n}(y_i - \bar{y})^2}} = \frac{\sum_{i=1}^{n} x_i \cdot y_i - n\cdot\bar{x}\cdot\bar{y}}{\sqrt{\left(\sum_{i=1}^{n} x_i^2 - n\cdot\bar{x}^2\right)\cdot\left(\sum_{i=1}^{n} y_i^2 - n\cdot\bar{y}^2\right)}}$$

Teststatistik

$$T = r_{XY}\cdot\sqrt{\frac{n-2}{1-r_{XY}^2}}$$

L19 – Asymmetrische Zusammenhangsmaße (metrisch)

Regressionsgleichung (lineare Regression)

$\hat{Y} = a + b \cdot X$ $\qquad$ $Y = \hat{Y} + E = a + b \cdot X + E$

Steigung und Regressionskonstante

$$b = \frac{SP_{YX}}{SS_X} = \frac{\sum_{i=1}^{n}(y_i - \bar{y}) \cdot (x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sum_{i=1}^{n} y_i \cdot x_i - \frac{\sum_{i=1}^{n} y_i \cdot \sum_{i=1}^{n} x_i}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}} = \frac{\sum_{i=1}^{n} y_i \cdot x_i - n \cdot \bar{y} \cdot \bar{x}}{\sum_{i=1}^{n} x_i^2 - n \cdot \bar{x}^2} = \frac{s_{YX}}{s_X^2} = \frac{\hat{\sigma}_{YX}}{\hat{\sigma}_X^2}$$

$$a = \bar{y} - b \cdot \bar{x}$$

L19 – Asymmetrische Zusammenhangsmaße (metrisch)

Standardfehler für b und a

$$\hat{\sigma}(b_{Y.X}) = \sqrt{\frac{1}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \cdot \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2}} = \sqrt{\frac{1}{s_X^2} \cdot \frac{s_Y^2 - b_{Y.X}^2 \cdot s_X^2}{n-2}}$$

$$\hat{\sigma}(a_{Y.X}) = \sqrt{\frac{\frac{1}{n} \cdot \sum_{i=1}^{n} x_i^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \cdot \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2}} = \sqrt{\frac{s_X^2 + \bar{x}^2}{s_X^2} \cdot \frac{s_Y^2 - b_{Y.X}^2 \cdot s_X^2}{n-2}}$$

Teststatistik (für b)

$$T = \frac{b_{Y.X} - \beta}{\hat{\sigma}(b_{Y.X})}$$

L19 – Asymmetrische Zusammenhangsmaße (metrisch)

Determinationskoeffizient

$$R^2 = 1 - \frac{s_E^2}{s_Y^2} = 1 - \frac{SS_E}{SS_Y} = \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad = \quad \frac{s_{\hat{Y}}^2}{s_Y^2} = \frac{SS_{\hat{Y}}}{SS_Y} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

$$R^2 = \frac{SS_{\hat{Y}}}{SS_Y} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = \frac{\sum_{i=1}^{n}(a + b \cdot x_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = \frac{\sum_{i=1}^{n}((\bar{y} - b \cdot \bar{x}) + b \cdot x_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = \frac{\sum_{i=1}^{n}(b \cdot (x_i - \bar{x}))^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

$$= b^2 \cdot \frac{SS_X}{SS_Y} = b^2 \cdot \frac{s_X^2}{s_Y^2} = \frac{(s_{YX})^2}{(s_X^2)^2} \cdot \frac{s_X^2}{s_Y^2} = \frac{(s_{YX})^2}{s_x^2 \cdot s_Y^2} = \left(\frac{s_{YX}}{s_X \cdot s_Y}\right)^2 = (r_{YX})^2$$

L20 – Zusammenhangsmaße Mehrfeldertabelle (ordinal)

Asymmetrisches Zusammenhangsmaß Somer‘s d

$$d_{YX} = \frac{C - D}{C + D + T_Y} = \frac{a \cdot d - b \cdot c}{a \cdot d + b \cdot c + a \cdot b + c \cdot d} \qquad d_{XY} = \frac{C - D}{C + D + T_X} = \frac{a \cdot d - b \cdot c}{a \cdot d + b \cdot c + a \cdot c + b \cdot d}$$

$$= 100 \cdot d_{YX} \% = \left(\frac{a}{a + c} - \frac{b}{b + d}\right) \qquad = 100 \cdot d_{XY} \% = \left(\frac{a}{a + b} - \frac{c}{c + d}\right)$$

Symmetrische Zusammenhangsmaße

Gamma ; Tau-b

$$\gamma = \frac{C - D}{C + D} \; ; \; \tau_b = \sqrt{d_{YX} \cdot d_{XY}}$$