BM21 Business Analytics - Formuleblad

Betrouwbaarheidsinterval (confidence interval)

Steekproefstandaardafwijking (sample standard deviation)

$$\text{CI} = \bar{x} \pm t^{\star} \cdot \frac{s}{\sqrt{n}}$$

Bekende populatiestandaardafwijking of grote n (known population standard deviation or large n)

$$\text{CI} = \bar{x} \pm z^{\star} \cdot \frac{\sigma}{\sqrt{n}}$$

Samenhang (association)

Covariantie (covariance)

$$\text{Cov}(X, Y) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})$$

Correlatie (Pearson's correlation)

$$r_{XY} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)}\,\sqrt{\text{Var}(Y)}}$$

Toetsingsgrootheden (test statistics)

t-toetsen: één steekproef (one-sample)

$$t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

t-toetsen: twee steekproeven (two-sample)

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

t-toetsen: gepaard (paired)

$$t = \frac{\bar{d}}{s_d / \sqrt{n}}$$

z-toetsen: één steekproef (one-sample)

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

z-toetsen: twee steekproeven (two-sample)

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma^2_1}{n_1} + \frac{\sigma^2_2}{n_2}}}$$

z-toetsen: proportie (proportion)

$$z = \frac{p - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Chikwadraattoets (chi-square test)

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

O = waargenomen (observed), E = verwachte frequentie (expected frequency)

Cronbachs alfa (Cronbach's alpha)

Cronbachs alfa (Cronbach's alpha)

$$\rho_Y = \frac{\text{Var}(K \cdot T)}{\text{Var}(Y)} = \cdots = \frac{K}{K-1}\left(1 - \frac{\sum_{k=1}^{K} \text{Var}(X_k)}{\text{Var}\left(\sum_{k=1}^{K} X_k\right)}\right) = \alpha$$

Lineaire regressie (linear regression)

Determinatiecoëfficiënt (R-squared)

$$R^2 = \frac{\text{SSE}}{\text{SST}} = 1 - \frac{\text{SSR}}{\text{SST}}$$

Gecorrigeerde R² (adjusted R²)

$$R^2_{\text{adj}} = 1 - (1 - R^2) \cdot \frac{n-1}{n-k-1}$$

Toets op één coëfficiënt (coefficient test)

$$t = \frac{b_j - \beta_{j,0}}{\text{SE}(b_j)}$$

SST = totale variatie (total sum of squares); SSR = som van gekwadrateerde residuen (sum of squared residuals); SSE = SST − SSR = verklaarde variatie (explained sum of squares); n = aantal waarnemingen; k = aantal voorspellers.

Logistische regressie (logistic regression)

Log-likelihood

$$\ell(\beta) = \sum_{i=1}^{n} \left[y_i \log(p_i) + (1 - y_i) \log(1 - p_i) \right]$$

Logitmodel (logit model)

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$$

Pseudo-R² (McFadden)

$$R^2_{\text{McF}} = 1 - \frac{\ln L_{\text{model}}}{\ln L_{\text{null}}}$$

Akaike-informatiecriterium (AIC)

$$\text{AIC} = 2k - 2\ln(\hat{L})$$