## Lineární regresní model

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + e_i, \quad i = 1, 2, \ldots n$$

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \mathbf{e} = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}, \mathbf{X} = \begin{pmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nk} \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}$$

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

Normální rovnice

$$\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{Y}$$

Odhady metodou nejmenších čtverců

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}'\mathbf{Y}.$$

Reziduální součet čtverců

$$S_e = (\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}) = \mathbf{Y}'\mathbf{Y} - \widehat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y}$$

Reziduální rozptyl

$$s_e^2 = \frac{S_e}{n-k} = \frac{1}{n-k}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Směrodatné chyby odhadů

$$s(\widehat{\beta}_j) = \sqrt{s_e^2 v_{jj}},$$

kde $v_{11}, v_{22}, \ldots, v_{kk}$ jsou prvky na hlavní diagonále matice $(\mathbf{X}'\mathbf{X})^{-1}$

Oboustranný interval spolehlivosti pro odhad parametru $\beta_j$ při riziku odhadu $\alpha$

$$\widehat{\beta}_j - t_{1-\alpha/2}(n-k) \cdot s(\widehat{\beta}_j) < \beta_j < \widehat{\beta}_j + t_{1-\alpha/2}(n-k) \cdot s(\widehat{\beta}_j)$$

Testy významnosti parametrů $\beta_j$, $j = 1, \ldots, k$

$H$: $\beta_j = 0 \to A$: $\beta_j \neq 0$.

$$t = \frac{\widehat{\beta}_j}{s(\widehat{\beta}_j)}$$

$W_\alpha$: $|t| \geq t_{1-\alpha/2}(n-k)$

Interval spolehlivosti pro regresní funkci v bodě $\mathbf{x}_0$

$$\widehat{y}(\mathbf{x}_0) - t_{1-\alpha/2}(n-k) \cdot s(\widehat{y}(\mathbf{x}_0)) < y(\mathbf{x}_0) < \widehat{y}(\mathbf{x}_0) + t_{1-\alpha/2}(n-k) \cdot s(\widehat{y}(\mathbf{x}_0)),$$

kde $s(\widehat{y}(\mathbf{x}_0)) = s_e\sqrt{\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0}$

Interval spolehlivosti pro předpověď v bodě $\mathbf{x}_0$

$$\widehat{y}(\mathbf{x}_0) - t_{1-\alpha/2}(n-k) \cdot s_0 < Y_0 < \widehat{y}(\mathbf{x}_0) + t_{1-\alpha/2}(n-k) \cdot s_0,$$

kde $s_0 = s_e\sqrt{1 + \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0}$

Celkový součet čtverců

$$S_Y = \mathbf{Y}'\mathbf{Y} - n\overline{y}^2 = \sum_{i=1}^{n}(y_i - \overline{y})^2 = n \cdot s_n^2(y), \text{ kde } s_n^2(y) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2$$

Reziduální součet čtverců

$$S_e = \mathbf{Y}'\mathbf{Y} - \widehat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y} = \sum_{i=1}^{n}(y_i - \widehat{y}_i)^2 = (n-k)\cdot s_e^2, \text{ kde } s_e^2 = \frac{1}{n-k}\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2$$

Teoretický součet čtverců

$$S_T = \widehat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y} - n\overline{y}^2 = \sum_{i=1}^{n}(\widehat{y}_i - \overline{y})^2 = n\cdot s_n^2(\hat{y}), \text{ kde } s_n^2(\widehat{y}) = \frac{1}{n}\sum_{i=1}^{n}(\widehat{y}_i - \overline{y})^2$$

Index determinace

$$R^2 = \frac{S_T(y)}{S_c(y)} = \frac{\sum_{i=1}^{n}(\widehat{y}_i - \overline{y})^2}{\sum_{i=1}^{n}(y_i - \overline{y_i})^2}$$

Index determinace (adjusted)

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

**Test významnosti regresního modelu (s konstantou)**

$H$: $\beta_2 = \beta_3 = \cdots = \beta_k = 0 \rightarrow A$: $\beta_j \neq 0$ pro alespoň jedno $j = 2, 3, \ldots, k$.

$$F = \frac{S_T(y)}{k-1} : \frac{S_e(y)}{n-k}$$

$W_\alpha$: $F \geq F_{1-\alpha}(k-1, n-k)$

**Test obecné lineární hypotézy**

$\mathbf{A}$ je reálnou matice typu $m \times k$ a hodnosti $m \leq k$, $\mathbf{a}$ je vhodný $m-$rozměrný reálný vektor

$H$: $\mathbf{A}\boldsymbol{\beta} = \mathbf{a} \rightarrow A$: $\mathbf{A}\boldsymbol{\beta} \neq \mathbf{a}$.

$$F = \frac{1}{ms_e^2}(\mathbf{A}\widehat{\boldsymbol{\beta}} - \mathbf{a})'[\mathbf{A}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}']^{-1}(\mathbf{A}\widehat{\boldsymbol{\beta}} - \mathbf{a})$$

$W_\alpha$: $F \geq F_{1-\alpha}(m, n-k)$