Statistika a spolehlivost v lékařství – vzorce ke zkoušce

1 Intervalové odhady parametrů normálního rozdělení $\mathrm{N}(\mu, \sigma^2)$

Odhad střední hodnoty při **známém** rozptylu σ^2: $\left\langle \bar{\boldsymbol{X}} - \frac{\sigma}{\sqrt{n}}\,\Phi^{-1}\left(1-\frac{\alpha}{2}\right), \bar{\boldsymbol{X}} + \frac{\sigma}{\sqrt{n}}\,\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\right\rangle$.

Odhad střední hodnoty při **neznámém** rozptylu: $\left\langle \bar{\boldsymbol{X}} - \frac{S_X}{\sqrt{n}}\,q_{\mathrm{t}(n-1)}(1-\frac{\alpha}{2}), \bar{\boldsymbol{X}} + \frac{S_X}{\sqrt{n}}\,q_{\mathrm{t}(n-1)}(1-\frac{\alpha}{2})\right\rangle$.

Odhad rozptylu: $\left\langle \frac{(n-1)\,S_X^2}{q_{\chi^2(n-1)}\left(1-\frac{\alpha}{2}\right)}, \frac{(n-1)\,S_X^2}{q_{\chi^2(n-1)}\left(\frac{\alpha}{2}\right)}\right\rangle$.

2 Testování hypotéz

Test střední hodnoty normálního rozdělení při **známém** rozptylu σ^2: $\frac{\bar{\boldsymbol{x}}-c}{\sigma}\sqrt{n}$ testujeme na normované normální rozdělení.
Test střední hodnoty normálního rozdělení při **neznámém** rozptylu: $\frac{\bar{\boldsymbol{x}}-c}{s_{\boldsymbol{x}}}\sqrt{n}$ testujeme na Studentovo rozdělení $\mathrm{t}(n-1)$.
Test rozptylu normálního rozdělení: $\frac{(n-1)\,s_{\boldsymbol{x}}^2}{c}$ testujeme na rozdělení $\chi^2(n-1)$.
Test rovnosti rozptylů dvou normálních rozdělení (Fisherův): $\frac{s_{\boldsymbol{x}}^2}{s_{\boldsymbol{y}}^2}$ testujeme na rozdělení $\mathrm{F}(m-1, n-1)$.
Test rovnosti středních hodnot dvou normálních rozdělení se **známým** rozptylem σ^2: $\frac{\bar{\boldsymbol{x}}_m - \bar{\boldsymbol{y}}_n}{\sigma\sqrt{\frac{1}{m}+\frac{1}{n}}}$ testujeme na normované normální rozdělení.
Test rovnosti středních hodnot dvou normálních rozdělení se (stejným) **neznámým** rozptylem: $\frac{\bar{\boldsymbol{x}}_m - \bar{\boldsymbol{y}}_n}{s\sqrt{\frac{1}{m}+\frac{1}{n}}}$, kde $s^2 = \frac{(m-1)\,s_X^2+(n-1)\,s_Y^2}{m+n-2}$, testujeme na Studentovo rozdělení $\mathrm{t}(m+n-2)$.
Test rovnosti středních hodnot dvou normálních rozdělení se **známým** rozptylem σ^2 – **párový pokus**: $\frac{\bar{\boldsymbol{x}}-\bar{\boldsymbol{y}}}{\sigma}\sqrt{\frac{n}{2}}$ testujeme na normované normální rozdělení.
Test rovnosti středních hodnot dvou normálních rozdělení se (stejným) **neznámým** rozptylem – **párový pokus**: $\frac{\bar{\boldsymbol{\delta}}}{s_\delta}\sqrt{n}$, kde $\delta_j = x_j - y_j$, testujeme na Studentovo rozdělení $\mathrm{t}(n-1)$.

2.1 Test nekorelovanosti dvou normálních rozdělení

$$r_{\boldsymbol{x},\boldsymbol{y}} = \frac{\sum\limits_{j=1}^{n}(x_j-\bar{\boldsymbol{x}})\,(y_j-\bar{\boldsymbol{y}})}{\sqrt{\left(\sum\limits_{j=1}^{n}(x_j-\bar{\boldsymbol{x}})^2\right)\left(\sum\limits_{j=1}^{n}(y_j-\bar{\boldsymbol{y}})^2\right)}} = \frac{n\sum\limits_{j=1}^{n}x_j\,y_j - \left(\sum\limits_{j=1}^{n}x_j\right)\left(\sum\limits_{j=1}^{n}y_j\right)}{\sqrt{\left(n\sum\limits_{j=1}^{n}x_j^2-\left(\sum\limits_{j=1}^{n}x_j\right)^2\right)\left(n\sum\limits_{j=1}^{n}y_j^2-\left(\sum\limits_{j=1}^{n}y_j\right)^2\right)}}$$

$$t = \frac{r_{\boldsymbol{x},\boldsymbol{y}}\sqrt{n-2}}{\sqrt{1-r_{\boldsymbol{x},\boldsymbol{y}}^2}}$$

testujeme na $\mathrm{t}(n-2)$.

3 Podmíněná rozdělení pravděpodobnosti

3.1 Rozdělení podmíněné jevem

Podmíněná pravděpodobnostní funkce $p_{Y|B}$ *diskrétní* náhodné veličiny Y za podmínky B:

$$p_{Y|B}(y) = P[Y = y|B] = \frac{P[Y = y, B]}{P[B]} \,,$$

Podmíněná distribuční funkce $F_{Y|B}$ náhodné veličiny Y za podmínky B:

$$F_{Y|B}(y) = P[Y \le y|B] = \frac{P[Y \le y, B]}{P[B]} \,,$$

podmíněná hustota $f_{Y|B}$ *spojité* náhodné veličiny

$$F_{Y|B}(y) = \int_0^y f_{Y|B}(t)\,\mathrm{d}t$$

3.2 Rozdělení podmíněné hodnotou diskrétní náhodné veličiny

$$p_{Y,X}(y,x) = P[Y = y, X = x] = P[Y = y \mid X = x] \cdot P[X = x] = p_{Y|X}(y|x) \cdot p_X(x)\,,$$

$$\begin{aligned} F_{Y,X}(y,x) &= \sum_{t:\,t\le x}\,\sum_{u:\,u\le y} p_{Y,X}(u,t) = \sum_{t:\,t\le x}\,\sum_{u:\,u\le y} p_{Y|X}(u|t) \cdot p_X(t) \\ &= \sum_{t:\,t\le x} F_{Y|X}(y|t) \cdot p_X(t)\,. \end{aligned}$$

Obecněji (pro *libovolnou* náhodnou veličinu Y)

$$F_{Y,X}(y,x) = \sum_{t:\,t\le x} F_{Y|X}(y|t) \cdot p_X(t)\,.$$

3.3 Rozdělení podmíněné hodnotou spojité náhodné veličiny

$$f_{Y,X}(y,x) = f_{Y|X}(y|x) \cdot f_X(x)\,,$$

přesněji a obecněji

$$F_{Y,X}(y,x) = \int_{-\infty}^{x} F_{Y|X}(y|t) \cdot f_X(t)\,\mathrm{d}t = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{Y|X}(u|t) \cdot f_X(t)\,\mathrm{d}u\,\mathrm{d}t\,.$$

3.4 Vlastnosti podmíněných rozdělení

$$\begin{aligned} \mathrm{E}(\mathrm{E}(Y|X)) &= \mathrm{E}Y\,, \\ \mathrm{D}Y &= \mathrm{D}(\mathrm{E}(Y|X)) + \mathrm{E}(\mathrm{D}(Y|X))\,. \end{aligned}$$

4 Regrese

4.1 Odhad náhodné veličiny: lineární model dimenze 1

Úloha 3: Odhadujeme spojitou náhodnou veličinu Y, předpokládáme

$$Y = \vartheta_0 + \vartheta_1\,X + \mathcal{E}\,,$$

Vstupy: realizace náhodného výběru ze sdruženého rozdělení n. vektoru (Y, X)

$$(\boldsymbol{y}, \boldsymbol{x}) = ((y_1, x_1), \ldots, (y_n, x_n))\,.$$

Předpokládáme $y_j = \vartheta_0 + \vartheta_1\,x_j + e_j$, kde $(e_1, \ldots, e_n)$ je realizace náhodného výběru z rozdělení $\mathrm{N}(0, \sigma^2)$.

$$\hat\vartheta_0 = \bar y - \hat\vartheta_1\,\bar x\,.$$

$$\hat\vartheta_1 = \frac{\sum_j x_j\,y_j - n\,\bar x\,\bar y}{\sum_j x_j^2 - n\,\bar x^2} = \frac{\sum_j (x_j - \bar x)\,(y_j - \bar y)}{\sum_j (x_j - \bar x)^2}$$

4.2 Regresní přímka 1

Je tvořena body (x, y), které splňují rovnici

$$y = \hat{\vartheta}_0 + \hat{\vartheta}_1\, x\,.$$

$$y - \bar{y} = \hat{\vartheta}_1\, (x - \bar{x})\,.$$

4.3 Regresní přímka 2

$$\hat{\vartheta}_1^* = \frac{\sum_j\, (x_j - \bar{x})\, (y_j - \bar{y})}{\sum_j\, (y_j - \bar{y})^2}\,.$$

$$\begin{aligned} x &= \hat{\vartheta}_0^* + \hat{\vartheta}_1^*\, y\,, \\ x - \bar{x} &= \hat{\vartheta}_1^*\, (y - \bar{y})\,. \end{aligned}$$

4.4 Interpretace regresních koeficientů

$$\begin{aligned} \hat{\sigma}_{\boldsymbol{x}}^2 &:= \frac{1}{n} \sum_j (x_j - \bar{x})^2 = \frac{n-1}{n}\, s_{\boldsymbol{x}}^2 = \mathrm{D}\,\mathrm{Emp}(\boldsymbol{x})\,, \\ \hat{\sigma}_{\boldsymbol{y}}^2 &:= \frac{1}{n} \sum_j (y_j - \bar{y})^2 = \frac{n-1}{n}\, s_{\boldsymbol{y}}^2 = \mathrm{D}\,\mathrm{Emp}(\boldsymbol{y})\,, \\ c_{\boldsymbol{x},\boldsymbol{y}} &:= \frac{1}{n} \sum_j (x_j - \bar{x})\,(y_j - \bar{y}) = \mathrm{cov}(\mathrm{Emp}(\boldsymbol{x},\boldsymbol{y}))\,. \end{aligned}$$

Odhad korelace = realizace výběrového koeficientu korelace =

$$r_{\boldsymbol{x},\boldsymbol{y}} := \frac{\sum_j (x_j - \bar{x})\,(y_j - \bar{y})}{\sqrt{\sum_j (x_j - \bar{x})^2\, \sum_j (y_j - \bar{y})^2}} = \frac{c_{\boldsymbol{x},\boldsymbol{y}}}{\hat{\sigma}_{\boldsymbol{x}}\, \hat{\sigma}_{\boldsymbol{y}}} = \varrho(\mathrm{Emp}(\boldsymbol{x},\boldsymbol{y}))\,.$$

Lze psát

$$\hat{\vartheta}_1 = \frac{c_{\boldsymbol{x},\boldsymbol{y}}}{\hat{\sigma}_{\boldsymbol{x}}^2}\,.$$

Rovnice regresní přímky pro závislost Y na X:

$$\begin{aligned} y - \bar{y} &= \hat{\vartheta}_1\, (x - \bar{x})\,, \\ y - \bar{y} &= \frac{c_{\boldsymbol{x},\boldsymbol{y}}}{\hat{\sigma}_{\boldsymbol{x}}^2}\, (x - \bar{x})\,, \\ \frac{y - \bar{y}}{\hat{\sigma}_{\boldsymbol{y}}} &= r_{\boldsymbol{x},\boldsymbol{y}}\, \frac{x - \bar{x}}{\hat{\sigma}_{\boldsymbol{x}}}\,, \end{aligned}$$

rovnice regresní přímky pro závislost X na Y (to je **jiná** přímka!):

$$\begin{aligned} x - \bar{x} &= \hat{\vartheta}_1^*\, (y - \bar{y})\,, \\ x - \bar{x} &= \frac{c_{\boldsymbol{x},\boldsymbol{y}}}{\hat{\sigma}_{\boldsymbol{y}}^2}\, (y - \bar{y})\,, \\ \frac{x - \bar{x}}{\hat{\sigma}_{\boldsymbol{x}}} &= r_{\boldsymbol{x},\boldsymbol{y}}\, \frac{y - \bar{y}}{\hat{\sigma}_{\boldsymbol{y}}}\,. \end{aligned}$$

Směrnice obou regresních přímek mají součin $\hat{\vartheta}_1\, \hat{\vartheta}_1^* = \dfrac{c_{\boldsymbol{x},\boldsymbol{y}}^2}{\hat{\sigma}_{\boldsymbol{x}}^2\, \hat{\sigma}_{\boldsymbol{y}}^2} = r_{\boldsymbol{x},\boldsymbol{y}}^2$
a geometrický průměr absolutních hodnot $\sqrt{\hat{\vartheta}_1\, \hat{\vartheta}_1^*} = |r_{\boldsymbol{x},\boldsymbol{y}}|$.

4.5 Chyba lineární regrese

Odhadli jsme lineární regresní funkci $\hat{g}$,

$$\hat{g}(x) := \hat{\vartheta}_0 + \hat{\vartheta}_1\, x\,,$$

pomocí ní hodnoty nezávisle proměnné v jednotlivých realizacích

$$\hat{y}_j := \hat{g}(x_j) = \hat{\vartheta}_0 + \hat{\vartheta}_1\, x_j\,,$$
$$\hat{\boldsymbol{y}} := (\hat{y}_1, \ldots, \hat{y}_n)$$

a chyby (**rezidua**)

$$\hat{e}_j := y_j - \hat{y}_j = y_j - \hat{\vartheta}_0 - \hat{\vartheta}_1\, x_j = y_j - \bar{y} - \hat{\vartheta}_1\, (x_j - \bar{x})\,,$$
$$\hat{\boldsymbol{e}} := (\hat{e}_1, \ldots, \hat{e}_n)\,.$$

Věta: $\frac{1}{n} \sum_j \hat{y}_j = \bar{y}\,.$

Důkaz: $\frac{1}{n} \sum_j \hat{y}_j = \frac{1}{n} \sum_j \left(\hat{\vartheta}_0 + \hat{\vartheta}_1\, x_j\right) = \hat{\vartheta}_0 + \hat{\vartheta}_1\, \bar{x} = \bar{y}\,.$

4.6 Složky rozptylu regresního odhadu

Celkový rozptyl (angl. *total variation*; někdy se nedělí n) $:= \hat{\sigma}^2_{\boldsymbol{y}} = \frac{1}{n} \sum_j (y_j - \bar{y})^2\,.$

Rozptyl modelu (angl. *explained variation*) $:= \hat{\sigma}^2_{\hat{\boldsymbol{y}}} = \frac{1}{n} \sum_j (\hat{y}_j - \bar{y})^2\,.$

Reziduální rozptyl (angl. *unexplained variation*) :=

$$\hat{\sigma}^2_{\hat{\boldsymbol{e}}} = \frac{1}{n} \sum_j \hat{e}_j^2 = \frac{1}{n} \sum_j (y_j - \hat{y}_j)^2\,.$$

Věta: $\hat{\sigma}^2_{\boldsymbol{y}} = \hat{\sigma}^2_{\hat{\boldsymbol{y}}} + \hat{\sigma}^2_{\hat{\boldsymbol{e}}}\,.$

Důkaz:

Věta: $r^2_{\boldsymbol{x},\boldsymbol{y}} = \frac{\hat{\sigma}^2_{\hat{\boldsymbol{y}}}}{\hat{\sigma}^2_{\boldsymbol{y}}} = 1 - \frac{\hat{\sigma}^2_{\hat{\boldsymbol{e}}}}{\hat{\sigma}^2_{\boldsymbol{y}}}\,.$

4.7 Odhad rozptylu

Max. věrohodný odhad rozptylu *původního* rozdělení:

$$\hat{\sigma}^2 = \frac{1}{n} \sum_j \hat{e}_j^2 = \mathrm{D}\,\mathrm{Emp}(\hat{\boldsymbol{e}})\,.$$

nestranný odhad:

$$\frac{1}{n-2} \sum_j \hat{e}_j^2 = \frac{1}{n-2} \sum_j \left(y_j - \hat{\vartheta}_0 - \hat{\vartheta}_1\, x_j\right)^2\,.$$

4.8 Rozdělení odhadů a testy hypotéz o nich

Odhady regresních koeficientů jako funkce náhodného výběru (nikoli jeho realizace) jsou náhodné veličiny,

$$\hat{\Theta}_1 = \frac{C_{\boldsymbol{X},\boldsymbol{Y}}}{\hat{\sigma}^2_{\boldsymbol{X}}}\,,$$
$$\hat{\Theta}_0 = \bar{Y} - \hat{\Theta}_1\, \bar{X}\,,$$

kde

$$\hat{\sigma}^2_{\boldsymbol{X}} := \frac{1}{n}\sum_j (X_j - \bar{X})^2 = \frac{n-1}{n}\, S^2_{\boldsymbol{X}},$$

$$C_{\boldsymbol{X},\boldsymbol{Y}} := \frac{1}{n}\sum_j (X_j - \bar{X})\,(Y_j - \bar{Y})$$

podobně odhad rozptylu,

$$\hat{\sigma}^2 := \frac{n}{n-2}\,\hat{\sigma}^2_{\mathcal{E}} = \frac{1}{n-2}\sum_j \mathcal{E}_j^2 = \frac{1}{n-2}\sum_j \left(Y_j - \hat{\Theta}_0 - \hat{\Theta}_1\, X_j\right)^2 .$$

Věta: Odhady $\hat{\Theta}_0, \hat{\Theta}_1, \hat{\sigma}^2$ skutečných parametrů $\vartheta_0, \vartheta_1, \sigma^2$ jsou nestranné, konzistentní, asymptoticky normální.
Odhady rozptylů regresních koeficientů

$$\hat{\sigma}^2_{\hat{\Theta}_0} = \frac{\hat{\sigma}^2}{n^2\,\hat{\sigma}^2_{\boldsymbol{X}}}\sum_j X_j^2 = \frac{\hat{\sigma}^2_{\mathcal{E}}}{n\,(n-2)\,\hat{\sigma}^2_{\boldsymbol{X}}}\sum_j X_j^2\,,$$

$$\hat{\sigma}^2_{\hat{\Theta}_1} = \frac{\hat{\sigma}^2}{n\,\hat{\sigma}^2_{\boldsymbol{X}}} = \frac{\hat{\sigma}^2_{\mathcal{E}}}{(n-2)\,\hat{\sigma}^2_{\boldsymbol{X}}}\,.$$

Odhady $\hat{\Theta}_0, \hat{\Theta}_1$ **nejsou nezávislé**.
$\dfrac{\hat{\Theta}_0 - \vartheta_0}{\frac{\hat{\sigma}_{\mathcal{E}}}{\hat{\sigma}_{\boldsymbol{X}}}\sqrt{\frac{1}{n}\sum_j X_j^2}}\sqrt{n-2}$ má rozdělení $\mathrm{t}(n-2)\,,$
$\dfrac{\hat{\Theta}_1 - \vartheta_1}{\frac{\hat{\sigma}_{\mathcal{E}}}{\hat{\sigma}_{\boldsymbol{X}}}}\sqrt{n-2}$ má rozdělení $\mathrm{t}(n-2)\,,$
Pro dané x má $\dfrac{Y - \hat{\Theta}_0 - \hat{\Theta}_1\, x}{\hat{\sigma}_{\mathcal{E}}\sqrt{n+1+\frac{n\,(x-\bar{X})^2}{\hat{\sigma}^2_{\boldsymbol{X}}}}}\sqrt{n-2}$ rozdělení $\mathrm{t}(n-2)\,.$

4.9 Testy korelace

Test nekorelovanosti (nulovosti korelace): Za předpokladu $\varrho(X,Y) = 0$ má

$$\frac{R_{\boldsymbol{X},\boldsymbol{Y}}}{\sqrt{1 - R^2_{\boldsymbol{X},\boldsymbol{Y}}}}\sqrt{n-2}$$

rozdělení $\mathrm{t}(n-2)\,,$
Test (nenulové) hodnoty korelace: Za předpokladu $\varrho(X,Y) = c \neq 0$ má

$$Z_{\boldsymbol{X},\boldsymbol{Y}} := \frac{1}{2}\ln\frac{1 + R_{\boldsymbol{X},\boldsymbol{Y}}}{1 - R_{\boldsymbol{X},\boldsymbol{Y}}}$$

rozdělení přibližně $\mathrm{N}\left(\mu_{Z_{\boldsymbol{X},\boldsymbol{Y}}}, \sigma^2_{Z_{\boldsymbol{X},\boldsymbol{Y}}}\right)$, kde

$$\mu_{Z_{\boldsymbol{X},\boldsymbol{Y}}} := \frac{1}{2}\ln\frac{1+c}{1-c}\,,$$

$$\sigma^2_{Z_{\boldsymbol{X},\boldsymbol{Y}}} := \frac{1}{n-3}\,,$$

$$\sigma_{Z_{\boldsymbol{X},\boldsymbol{Y}}} = \sqrt{\frac{1}{n-3}}\,,$$

tedy

$$\frac{Z_{\boldsymbol{X},\boldsymbol{Y}} - \mu_{Z_{\boldsymbol{X},\boldsymbol{Y}}}}{\sigma_{Z_{\boldsymbol{X},\boldsymbol{Y}}}}$$

má rozdělení přibližně $\mathrm{N}\,(0,1)$.

Test rovnosti dvou korelací: Odhadneme $R_{\boldsymbol{X},\boldsymbol{Y}}$, resp. $R_{\boldsymbol{U},\boldsymbol{V}}$, z nezávislých výběrů rozsahu n, resp. m,

$$Z_{\boldsymbol{X},\boldsymbol{Y}} := \frac{1}{2} \ln \frac{1+R_{\boldsymbol{X},\boldsymbol{Y}}}{1-R_{\boldsymbol{X},\boldsymbol{Y}}},$$
$$Z_{\boldsymbol{U},\boldsymbol{V}} := \frac{1}{2} \ln \frac{1+R_{\boldsymbol{U},\boldsymbol{V}}}{1-R_{\boldsymbol{U},\boldsymbol{V}}};$$

Za předpokladu $\varrho(X,Y) = \varrho(U,V)$ má $Z := Z_{\boldsymbol{X},\boldsymbol{Y}} - Z_{\boldsymbol{U},\boldsymbol{V}}$ rozdělení přibližně $\mathrm{N}\left(\mu_Z, \sigma_Z^2\right)$, kde

$$\mu_Z := \mu_{Z_{\boldsymbol{X},\boldsymbol{Y}}} - \mu_{Z_{\boldsymbol{U},\boldsymbol{V}}} = \frac{1}{2} \ln \frac{1+R_{\boldsymbol{X},\boldsymbol{Y}}}{1-R_{\boldsymbol{X},\boldsymbol{Y}}} - \frac{1}{2} \ln \frac{1+R_{\boldsymbol{U},\boldsymbol{V}}}{1-R_{\boldsymbol{U},\boldsymbol{V}}},$$
$$\sigma_Z^2 := \sigma^2_{Z_{\boldsymbol{X},\boldsymbol{Y}}} + \sigma^2_{Z_{\boldsymbol{U},\boldsymbol{V}}} = \frac{1}{n-3} + \frac{1}{m-3},$$
$$\sigma_Z = \sqrt{\frac{1}{n-3} + \frac{1}{m-3}},$$

tedy

$$\frac{Z-\mu_Z}{\sigma_Z}$$

má rozdělení přibližně $\mathrm{N}(0,1)$.

4.10 Odhad náhodné veličiny: lineární model dimenze k

Úloha 4: Odhadujeme spojitou náhodnou veličinu Y pomocí k **vysvětlujících** náhodných veličin $X_1, \ldots, X_k$ na základě realizace náhodného výběru

$$((y_1, x_{11}, \ldots, x_{1k}), \ldots, (y_n, x_{n1}, \ldots, x_{nk})),$$

kde $n > k$, obvykle $n \gg k$. Předpokládáme lineární model

$$Y = \sum_{i=1}^{k} \vartheta_i X_i + \mathcal{E},$$

kde $\boldsymbol{\vartheta} = (\vartheta_1, \ldots, \vartheta_k)^T \in \mathbb{R}^k$ je (sloupcový) vektor neznámých parametů (**regresních koeficientů**),
$\mathcal{E}$ je náhodná veličina s rozdělením $\mathrm{N}(0, \sigma^2)$,
σ^2 je (konstantní, známý nebo neznámý) rozptyl,
Pro realizace dostáváme

$$y_j = \sum_{i=1}^{k} \vartheta_i x_{ji} + e_j,$$
$$\boldsymbol{y} = \boldsymbol{X}\,\boldsymbol{\vartheta} + \boldsymbol{e},$$

kde $\boldsymbol{y} = (y_1, \ldots, y_n)^T, \boldsymbol{e} = (e_1, \ldots, e_n)^T \in \mathbb{R}^n$ jsou (sloupcové) vektory realizací *(nezávislých)* náhodných veličin,

$$\boldsymbol{X} = \begin{pmatrix} x_{11} & x_{12} & \ldots & x_{1k} \\ x_{21} & x_{22} & \ldots & x_{2k} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \ldots & x_{nk} \end{pmatrix}.$$

reziduální součet čtverců (angl. **residual sum of squares**, **RSS**)

$$R_{SS} = \sum_{j=1}^{n} \hat{e}_j^2 = \sum_{j=1}^{n} \left(y_j - \sum_{i=1}^{k} \hat{\vartheta}_i x_{ji} \right)^2 = \|\boldsymbol{e}\|^2 = \left\| \boldsymbol{y} - \boldsymbol{X}\,\hat{\boldsymbol{\vartheta}} \right\|^2.$$

soustava normálních rovnic

$$\boldsymbol{X}^{\boldsymbol{T}} (\boldsymbol{y} - \boldsymbol{X}\,\hat{\boldsymbol{\vartheta}}) = \boldsymbol{0},$$

$$\boldsymbol{X^T}\,\boldsymbol{X}\,\hat{\boldsymbol{\vartheta}} = \boldsymbol{X^T}\,\boldsymbol{y}\,.$$

řešení je

$$\hat{\boldsymbol{\vartheta}} = \left(\boldsymbol{X^T}\,\boldsymbol{X}\right)^{-1}\,\boldsymbol{X^T}\,\boldsymbol{y}\,.$$

Kovarianční matice vektoru odhadů $\hat{\Theta}$ je

$$\boldsymbol{\Sigma}_{\hat{\Theta}} = \sigma^2\,\left(\boldsymbol{X^T}\,\boldsymbol{X}\right)^{-1}\,.$$

Věta: Hodnota regresní funkce v bodě $\boldsymbol{x} = (x_1, \ldots, x_k)^T \in \mathbb{R}^k$,

$$\boldsymbol{x^T}\,\hat{\boldsymbol{\vartheta}} = \sum_{i=1}^{k} \hat{\vartheta}_i\,x_i\,,$$

je nejlepší nestranný odhad vysvětlované náhodné veličiny Y v bodě $\boldsymbol{x}$.

4.11 Intervalové odhady regresních koeficientů při známém rozptylu

Věta: (Symetrický) $(1-\alpha)$-konfidenční interval pro odhad parametru $\boldsymbol{\vartheta}_i$ je

$$\hat{\boldsymbol{\vartheta}}_i \pm \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\,\sqrt{\left(\boldsymbol{\Sigma}_{\hat{\Theta}}\right)_{ii}} = \hat{\boldsymbol{\vartheta}}_i \pm \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\,\sigma\,\sqrt{c_{ii}}\,,$$

kde $\left(\boldsymbol{\Sigma}_{\hat{\Theta}}\right)_{ii}$, resp. c_{ii}, je i-tý prvek na diagonále matice $\boldsymbol{\Sigma}_{\hat{\Theta}}$, resp. $\boldsymbol{C} := \left(\boldsymbol{X^T}\,\boldsymbol{X}\right)^{-1}$.

4.12 Odhady rozptylu σ^2 původního rozdělení

Maximálně věrohodný:

$$\hat{\sigma}^2 = \frac{1}{n}\,R_{SS} = \frac{1}{n}\sum_{j=1}^{n}\hat{e}_j^2 = \frac{1}{n}\sum_{j=1}^{n}\left(y_j - \sum_{i=1}^{k}\hat{\vartheta}_i\,x_{ji}\right)^2,$$

nestranný:

$$\hat{\sigma}_{\mathcal{E}}^2 = \frac{1}{n-k}\;R_{SS} = \frac{1}{n-k}\sum_{j=1}^{n}\hat{e}_j^2 = \frac{1}{n-k}\sum_{j=1}^{n}\left(y_j - \sum_{i=1}^{k}\hat{\vartheta}_i\,x_{ji}\right)^2.$$

$\dfrac{R_{SS}}{\sigma^2}$ pochází z rozdělení $\chi^2(n-k)$.
Odhady R_{SS} a $\hat{\boldsymbol{\vartheta}}$ jsou nezávislé. (Nikoli však jednotlivé složky vektoru $\hat{\boldsymbol{\vartheta}}$ mezi sebou!)

4.13 Intervalové odhady regresních koeficientů při neznámém rozptylu

Věta: (Symetrický) $(1-\alpha)$-konfidenční interval pro odhad parametru $\boldsymbol{\vartheta}_i$ je

$$\hat{\boldsymbol{\vartheta}}_i \pm q_{\mathrm{t}(n-k)}\left(1-\frac{\alpha}{2}\right)\,\hat{\sigma}_{\mathcal{E}}\,\sqrt{c_{ii}} = \hat{\boldsymbol{\vartheta}}_i \pm q_{\mathrm{t}(n-k)}\left(1-\frac{\alpha}{2}\right)\cdot\sqrt{\frac{R_{SS}\,c_{ii}}{n-k}}\,,$$

kde c_{ii} je i-tý prvek na diagonále matice $\boldsymbol{C} = \left(\boldsymbol{X^T}\,\boldsymbol{X}\right)^{-1}$.