Statistika a spolehlivost v lékařství – vzorce ke zkoušce

1 Vlastnosti podmíněných rozdělení

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

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

(Uvedeny pouze symetrické oboustranné odhady.)

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

Odhad střední hodnoty při **neznámém** rozptylu: $\bar{\boldsymbol{x}} \pm \frac{s_{\boldsymbol{x}}}{\sqrt{n}}\,q_{\mathrm{t}(n-1)}(1 - \frac{\alpha}{2})$.

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

3 Testování hypotéz

Typický tvar testu: Testovací statistiku T, která roste s parametrem ϑ (přesněji její realizaci t) porovnáme s kvantily příslušného rozdělení za předpokladu $\vartheta = c$:

H_0	H_1	zamítáme pro	dosažená významnost
$\vartheta \le c$	$\vartheta > c$	$t > q_T(1-\alpha)$	$1 - F_T(t)$
$\vartheta \ge c$	$\vartheta < c$	$t < q_T(\alpha)$	$F_T(t)$
$\vartheta = c$	$\vartheta \ne c$	$t > q_T(1-\frac{\alpha}{2})$ nebo $t < q_T(\frac{\alpha}{2})$	$2\min\left(F_T(t), 1 - F_T(t)\right)$

Test střední hodnoty normálního rozdělení při **známém** rozptylu σ^2: $\frac{\bar{\boldsymbol{x}} - c}{\sigma}\sqrt{n}$ testujeme na rozdělení $\mathrm{N}(0,1)$.

Test střední hodnoty normálního rozdělení při **neznámém** rozptylu: $\frac{\bar{\boldsymbol{x}} - c}{s_{\boldsymbol{x}}}\sqrt{n}$ testujeme na 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 hodnost 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 rozdělení $\mathrm{N}(0,1)$.

Test rovnosti středních hodnost 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 $s^2 = \frac{(m-1)\,s_{\boldsymbol{x}}^2 + (n-1)\,s_{\boldsymbol{y}}^2}{m+n-2}$, testujeme na rozdělení $\mathrm{t}(m+n-2)$.

Párový pokus – test rovnosti středních hodnost dvou normálních rozdělení se **známým** rozptylem σ^2: $\frac{\bar{\boldsymbol{x}} - \bar{\boldsymbol{y}}}{\sigma}\sqrt{\frac{n}{2}}$ testujeme na rozdělení $\mathrm{N}(0,1)$.

Párový pokus – test rovnosti středních hodnost dvou normálních rozdělení se stejným **neznámým** rozptylem: $\frac{\bar{\boldsymbol{\delta}}}{s_{\boldsymbol{\delta}}}\sqrt{n}$, kde $\delta_j = x_j - y_j$, testujeme na rozdělení $\mathrm{t}(n-1)$.

3.1 Testy korelace

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

$$z_{\boldsymbol{x},\boldsymbol{y}} = \frac{1}{2}\,\ln \frac{1 + r_{\boldsymbol{x},\boldsymbol{y}}}{1 - r_{\boldsymbol{x},\boldsymbol{y}}} = h(r_{\boldsymbol{x},\boldsymbol{y}})\,, \qquad \text{kde } h(t) = \frac{1}{2}\,\ln\frac{1+t}{1-t}\,.$$

Test nekorelovanosti dvou normálních rozdělení, $H_0\colon \varrho(X,Y) = 0$: $\dfrac{r_{\boldsymbol{x},\boldsymbol{y}}}{\sqrt{1 - r_{\boldsymbol{x},\boldsymbol{y}}^2}}\sqrt{n-2}$ testujeme na rozdělení $\mathrm{t}(n-2)$.

Test (nenulové) hodnoty korelace dvou normálních rozdělení, $H_0\colon \varrho(X,Y) = c$, kde $c \neq 0$: $(h(r_{\boldsymbol{x},\boldsymbol{y}}) - h(c))\,\sqrt{n-3} = \dfrac{z_{\boldsymbol{x},\boldsymbol{y}} - \mu_{z_{\boldsymbol{x},\boldsymbol{y}}}}{\sigma_{z_{\boldsymbol{x},\boldsymbol{y}}}}$, kde $\mu_{z_{\boldsymbol{x},\boldsymbol{y}}} = h(c) = \dfrac{1}{2}\,\ln\dfrac{1+c}{1-c}$, $\sigma_{z_{\boldsymbol{x},\boldsymbol{y}}} = \sqrt{\dfrac{1}{n-3}}$, testujeme na rozdělení $\mathrm{N}(0,1)$.

Test rovnosti korelací dvou normálních rozdělení, $H_0\colon \varrho(X,Y) = \varrho(U,V)$: $\dfrac{h(r_{\boldsymbol{x},\boldsymbol{y}}) - h(r_{\boldsymbol{u},\boldsymbol{v}})}{\sqrt{\frac{1}{n-3} + \frac{1}{m-3}}} = \dfrac{z_{\boldsymbol{x},\boldsymbol{y}} - z_{\boldsymbol{u},\boldsymbol{v}}}{\sigma_Z}$, kde $\sigma_Z = \sqrt{\dfrac{1}{n-3} + \dfrac{1}{m-3}}$, testujeme na rozdělení $\mathrm{N}(0,1)$.

4 Regrese

4.1 Lineární model dimenze 1

$$Y = \vartheta_0 + \vartheta_1\, X + \mathcal{E}\,, \qquad y_j = \vartheta_0 + \vartheta_1\, x_j + e_j\,.$$

Regresní přímka 1:

$$\begin{aligned} y &= \hat\vartheta_0 + \hat\vartheta_1\, x\,, & y - \bar y &= \hat\vartheta_1\,(x - \bar x)\,,\\ \hat\vartheta_0 &= \bar y - \hat\vartheta_1\,\bar x\,, & \hat\vartheta_1 &= \tfrac{\sum_j x_j\, y_j - n\,\bar x\,\bar y}{\sum_j x_j^2 - n\,\bar x^2} = \tfrac{\sum_j (x_j - \bar x)\,(y_j - \bar y)}{\sum_j (x_j - \bar x)^2}\,. \end{aligned}$$

Odhad hodnot nezávisle proměnné v jednotlivých realizacích a chyby (**rezidua**): $\hat y_j = \hat\vartheta_0 + \hat\vartheta_1\, x_j\,,\quad \hat e_j = y_j - \hat y_j\,.$
Regresní přímka 2: $x = \hat\vartheta_0^* + \hat\vartheta_1^*\, y\,,\qquad x - \bar x = \hat\vartheta_1^*\,(y - \bar y)\,.$

4.2 Interpretace regresních koeficientů

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

$$r_{\boldsymbol{x},\boldsymbol{y}} = \frac{c_{\boldsymbol{x},\boldsymbol{y}}}{\hat\sigma_{\boldsymbol{x}}\,\hat\sigma_{\boldsymbol{y}}}, \qquad \hat\vartheta_1 = \frac{c_{\boldsymbol{x},\boldsymbol{y}}}{\hat\sigma_{\boldsymbol{x}}^2}\,, \qquad \hat\vartheta_1\,\hat\vartheta_1^* = r_{\boldsymbol{x},\boldsymbol{y}}^2\,.$$

Regresní přímka 1: $y - \bar{\boldsymbol{y}} = \dfrac{c_{\boldsymbol{x},\boldsymbol{y}}}{\hat\sigma_{\boldsymbol{x}}^2}\,(x - \bar{\boldsymbol{x}})\,,\qquad \dfrac{y - \bar{\boldsymbol{y}}}{\hat\sigma_{\boldsymbol{y}}} = r_{\boldsymbol{x},\boldsymbol{y}}\,\dfrac{x - \bar{\boldsymbol{x}}}{\hat\sigma_{\boldsymbol{x}}}\,.$

4.3 Složky rozptylu regresního odhadu

Rozptyl modelu $= \hat{\sigma}^2_{\hat{\boldsymbol{y}}} = \dfrac{1}{n} \sum_j (\hat{y}_j - \bar{\boldsymbol{y}})^2 .$

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

Celkový rozptyl $= \hat{\sigma}^2_{\boldsymbol{y}} = \dfrac{1}{n} \sum_j (y_j - \bar{\boldsymbol{y}})^2 = \hat{\sigma}^2_{\hat{\boldsymbol{y}}} + \hat{\sigma}^2_{\hat{\boldsymbol{e}}} .$ $\qquad r^2_{\boldsymbol{x},\boldsymbol{y}} = \dfrac{\hat{\sigma}^2_{\hat{\boldsymbol{y}}}}{\hat{\sigma}^2_{\boldsymbol{y}}} = 1 - \dfrac{\hat{\sigma}^2_{\hat{\boldsymbol{e}}}}{\hat{\sigma}^2_{\boldsymbol{y}}} .$

Odhady rozptylu σ^2 *původního* rozdělení:

- max. věrohodný: $\hat{\sigma}^2_{\hat{\boldsymbol{e}}} = \dfrac{1}{n} \sum_j \hat{e}_j^2 = \mathrm{D}\,\mathrm{Emp}(\hat{\boldsymbol{e}}) ,$
- nestranný: $\hat{\sigma}^2 = \dfrac{1}{n-2} \sum_j \hat{e}_j^2 .$

4.4 Rozdělení odhadů a testy hypotéz o nich

Odhady rozptylů regresních koeficientů:

$$\hat{\sigma}^2_{\hat{\vartheta}_0} = \frac{\hat{\sigma}^2}{n^2\,\hat{\sigma}^2_{\boldsymbol{x}}} \sum_j x_j^2 = \frac{\hat{\sigma}^2_{\hat{\boldsymbol{e}}}}{n\,(n-2)\,\hat{\sigma}^2_{\boldsymbol{x}}} \sum_j x_j^2 , \qquad \hat{\sigma}^2_{\hat{\vartheta}_1} = \frac{\hat{\sigma}^2}{n\,\hat{\sigma}^2_{\boldsymbol{x}}} = \frac{\hat{\sigma}^2_{\hat{\boldsymbol{e}}}}{(n-2)\,\hat{\sigma}^2_{\boldsymbol{x}}} .$$

Test abolutního členu, $H_0\colon \vartheta_0 = c$: $\dfrac{\hat{\vartheta}_0 - c}{\frac{\hat{\sigma}_{\hat{\boldsymbol{e}}}}{\hat{\sigma}_{\boldsymbol{x}}} \sqrt{\frac{1}{n} \sum_j x_j^2}} \sqrt{n-2}$ testujeme na rozdělení $\mathrm{t}(n-2)$.

Test směrnice, $H_0\colon \vartheta_1 = c$: $\dfrac{\hat{\vartheta}_1 - c}{\frac{\hat{\sigma}_{\hat{\boldsymbol{e}}}}{\hat{\sigma}_{\boldsymbol{x}}}} \sqrt{n-2}$ testujeme na rozdělení $\mathrm{t}(n-2)$.

Test chyby regrese pro dané x, $H_0\colon \vartheta_0 + \vartheta_1\, x = c$: $\dfrac{\hat{\vartheta}_0 + \hat{\vartheta}_1\, x - c}{\hat{\sigma}_{\hat{\boldsymbol{e}}} \sqrt{n + 1 + \frac{n\,(x-\bar{\boldsymbol{x}})^2}{\hat{\sigma}^2_{\boldsymbol{x}}}}} \sqrt{n-2}$ testujeme na rozdělení $\mathrm{t}(n-2)$.

4.5 Lineární model dimenze k

$$Y = \sum_{i=1}^{k} \vartheta_i\, X_i + \mathcal{E} , \qquad y_j = \sum_{i=1}^{k} \vartheta_i\, x_{ji} + e_j , \qquad \boldsymbol{y} = \boldsymbol{X}\,\boldsymbol{\vartheta} + \boldsymbol{e} .$$

Soustava normálních rovnic a její řešení: $\boldsymbol{X^T}\,\boldsymbol{X}\,\hat{\boldsymbol{\vartheta}} = \boldsymbol{X^T}\,\boldsymbol{y} ,$ $\qquad \hat{\boldsymbol{\vartheta}} = \left(\boldsymbol{X^T}\,\boldsymbol{X}\right)^{-1} \boldsymbol{X^T}\,\boldsymbol{y} .$

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

Reziduální součet čtverců: $R_{SS} = \sum_{j=1}^{n} \hat{e}_j^2 .$

$\dfrac{R_{SS}}{\sigma^2}$ pochází z rozdělení $\chi^2(n-k)$.

Odhady rozptylu σ^2 původního rozdělení:

- maximálně věrohodný: $\hat{\sigma}^2_{\hat{\boldsymbol{e}}} = \dfrac{1}{n}\, R_{SS} = \dfrac{1}{n} \sum_{j=1}^{n} \hat{e}_j^2 ,$
- nestranný: $\hat{\sigma}^2 = \dfrac{1}{n-k}\, R_{SS} = \dfrac{1}{n-k} \sum_{j=1}^{n} \hat{e}_j^2 .$

4.6 Intervalové odhady regresních koeficientů

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

Intervalový odhad regresních koeficientů při **známém** rozptylu σ^2: $\hat{\boldsymbol{\vartheta}}_i \pm \Phi^{-1}\left(1 - \dfrac{\alpha}{2}\right)\, \sigma\, \sqrt{c_{ii}} .$

Intervalový odhad regresních koeficientů při **neznámém** rozptylu:

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

5 χ^2 testy

Test dobré shody pozorovaného rozdělení s předpokládaným diskrétním rozdělením: $\displaystyle\sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i}$ testujeme na rozdělení $\chi^2(k-1)$.

Test dobré shody 2 diskrétních rozdělení: $\displaystyle\left(\frac{1}{m} + \frac{1}{n}\right) \sum_{i=1}^{k} \frac{(m_i - mp_i)^2}{mp_i}$, kde $p_i = \dfrac{m_i + n_i}{m + n}$, testujeme na rozdělení $\chi^2(k-1)$.

Test nezávislosti 2 diskrétních rozdělení: $\displaystyle\sum_{i=1}^{k} \sum_{j=1}^{m} \frac{(n_{ij} - np_i q_j)^2}{np_i q_j}$, kde $p_i = \dfrac{1}{n} \displaystyle\sum_{j=1}^{m} n_{ij}$ a $q_j = \dfrac{1}{n} \displaystyle\sum_{i=1}^{k} n_{ij}$, testujeme na rozdělení $\chi^2((k-1)(m-1))$.