Ústav matematiky a statistiky
Přírodovědecká fakulta
Masarykova univerzita

Štatistická inferencia II

Zadania príkladov na cvičenia

Stanislav Katina

katina@math.muni.cz

6. mája 2015

Definícia 1 (asymptotické rozdelenie poriadkovej štatistiky) *Nech $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ sú poriadkové štatistiky náhodného výberu $X_1, X_2, \ldots, X_n$. Majme pravdepodobnosť α, kde $F(t_\alpha) = \alpha$. Asymptoticky platí, že $\sqrt{n}(\frac{j}{n} - \alpha)$ konverguje k 0. Potom je poriadková štatistika $X_{(j)}$ normálne rozdelená so strednou hodnotou $E[X_{(j)}] = t_\alpha$ a rozptylom $\sigma^2_{X_{(j)}} = \frac{\alpha(1-\alpha)}{f^2(t_\alpha)n}$. Ak $X \sim N(\mu, \sigma^2)$, potom $\sigma^2_{X_{(j)}} = \sigma^2 \frac{\pi^2}{24 \ln n}$.*

Príklad 1 (rozptyl poriadkovej štatistiky) *(a) Pomocou delta metódy odvoďte rozptyl poriadkovej štatistiky v definícii 1. (b) Pomocou definície 1 odvoďte rozptyl poriadkovej štatistiky, ak $X \sim N(\mu, \sigma^2)$.*

Príklad 2 (graf distribučnej funkcie a jej IS) *Nakreslite graf distribučnej funkcie $N(\mu, \sigma^2)$, kde $\mu = 0$ a $\sigma^2 = 1$. Do grafu dokreslite 95% pás spoľahlivosti pre $F(x)$. Jeho hranice vypočítajte pomocou simulácie pseudonáhodných čísel z $N(0, 1)$ pri $n = 50$, kde $F_n(x)$ je odhadnutá z dát. Teoretická distribučnú funkciu $\Phi(\lambda)$ naprogramujte v R alebo použite knižnicu `kolmim` a funkciu `pkolm()`; help k tejto funkcii je pristupný na http://cran.r-project.org/web/packages/kolmim/index.html.*

Príklad 3 (χ^2-test dobrej zhody) *Majme dáta `Grades` z knižnice `PASWR`, ktoré reprezentujú SAT skóre ($n = 200$) náhodne vybranej vzorky študentov z jednej univerzity v USA. Otestujte na hladine významnosti $\alpha = 0.05$, či majú dáta normálne rozdelenie. Použite intervaly $\langle \mu - 3\sigma, \mu - 2\sigma \rangle$, $(\mu - 2\sigma, \mu - \sigma \rangle$, $(\mu - \sigma, \mu \rangle$, $(\mu, \mu + \sigma \rangle$, $(\mu + \sigma, \mu + 2\sigma \rangle$ a $(\mu + 2\sigma, \mu + 3\sigma \rangle$. Nakreslite histogram použitím vyššie spomenutých intervalov a superponujte ho s očakávanými hodnotami SAT skóre v každej kategórii, keď $F_0(x) \sim N(\mu, \sigma^2)$.*

Príklad 4 (χ^2-test dobrej zhody) *Johann Gregor Mendel vo svojich pokusoch s krížením rastlín hrachu (Pisum sativum) študoval dedičnosť siedmych rôznych znakov. V každom z pokusov, pri sledovaní jedného znaku, získal po krížení dvoch čistých línií (t.j. dominantného homozygota AA s recesívnym homozygotom aa) generáciu, v ktorej mali všetky rastliny rovnaký fenotyp (t.j. heterozygoti Aa). Po ich samooplodnení (čo je prirodzený spôsob rozmnožovania hrachu) získal ďalšiu generáciu, v ktorej sa vyskytovali sledované znaky v dvoch formách, a to zakaždým v pomere veľmi blízkom 3:1. Jedným zo znakov, ktoré študoval, bola farba semien. Po krížení* 258 *hybridov získal celkove 8023 semien, z ktorých* 6022 *bolo žltých a* 2001 *zelených. Otestujte platnosť fenotypového štiepneho pomeru* $3:1$ *na hladine významnosti $\alpha = 0.05$.*

Príklad 5 (χ^2-test dobrej zhody; početnosti úmrtí) *Otestujte zhodu početností X Pruských armádnych jednotiek, v ktorých nastalo n úmrtí zapríčinených kopnutím koňom za rok (pozri príklad zo SI1) s Poissonovým rozdelením s parametrom λ, t.j. $X \sim Poiss(\lambda)$ na hladine významnosti $\alpha = 0.05$.* DÚ

Príklad 6 (χ^2-test dobrej zhody; početnosti chlapcov) *Otestujte zhodu početností rodín X s n chlapcami (pozri príklad zo SI1) s binomickým rozdelením s parametrami N a π, t.j. $X \sim Bin(N, \pi)$ na hladine významnosti $\alpha = 0.05$.* DÚ

Príklad 7 (χ^2-test dobrej zhody; úrazy robotníkov) *Otestujte zhodu početností robotníkov X s n úrazmi v továrni (pozri príklad zo SI1)* DÚ
(a) s Poissonovým rozdelením s parametrom λ, t.j. $X \sim Poiss(\lambda)$ a
(b) s negatívne binomickým rozdelním s parametrami α a π, t.j. $Negbinom(\alpha, \pi)$ na hladine významnosti $\alpha = 0.05$.

Príklad 8 (χ^2-test dobrej zhody; fetálna aktivita) *Nech X predstavuje početnosti päťsekundových intervalov (z 240) v posledných 2/3 ťarchavosti zaznamenaných ultrazvukom, v ktorých sa plod ovce n-krát pohol (pozri tabuľku). Vypočítajte očakávané početnosti za predpokladu, že $X \sim Poiss(\lambda)$. Otestujte zhodu pozorovaných a teoretických početností na hladine významnosti $\alpha = 0.05$.* DÚ

Tabuľka 1: Pozorované početnosti m_n päťsekundových intervalov v posledných 2/3 ťarchavosti zaznamenaných ultrazvukom, v ktorých sa plod ovce n-krát pohol

n	0	1	2	3	4	5	6	7
pozorované m_n	182	41	12	2	2	0	0	1

Príklad 9 (Kolmogorov-Smirnovov test dobrej zhody) *Majme výšky $n = 12$ náhodne vybraných 10-ročných dievčat $\mathbf{x} = (131, 132, 135, 141, 141, 141, 141, 142, 143, 146, 146, 151)^T$. Otestujte na hladine významnosti $\alpha = 0.05$, či majú dáta normálne rozdelenie, kde $F_0(x) \sim N(\mu, \sigma^2)$.*

Príklad 10 (nezávislosť μ a σ^2; pravdepodobnosť pokrytia) *Nech $X \sim N(\mu, \sigma^2)$, kde $\mu = 20$ a $\sigma^2 = 100$. Vypočítajte Pearsonov korelačný koeficient $r_{\overline{X},S}$ pomocou simulačnej štúdie. Nakreslite sivou farbou rozptylový graf $(\overline{x}_m, s_m)$, kde $m = 1, 2, \ldots, M$, kde $M = 100000$. Dokreslite do grafu čiernou farbou také body, pre ktoré platí $t_{W,m} = \left|\frac{\overline{x}_m - \mu}{s_m}\sqrt{n}\right| < t_{n-1}(\alpha/2)$, ako aj hranice, ktoré definujú také body $(\overline{x}_m, s_m)$, pre ktoré $t_{W,m} = t_{n-1}(\alpha/2)$. Vypočítajte pravdepodobnosť pokrytia 95% DIS pre μ ako podiel $\sum_m I\left(t_{W,m} < t_{n-1}(\alpha/2)\right)/M$. Zvoľte (a) $n = 5$, (b) $n = 50$ a (c) $n = 100$.*

Príklad 11 (nezávislosť μ a σ^2; pravdepodobnosť pokrytia) *Nech $X \sim [pN(\mu, \sigma_1^2) + (1-p)N(\mu, \sigma_2^2)]$, kde $p = 0.9$, $\mu = 20$, $\sigma_1^2 = 100$ a $\sigma_2^2 = 400$. Vypočítajte Pearsonov korelačný koeficient $r_{\overline{X},S}$ pomocou simulačnej štúdie. Nakreslite sivou farbou rozptylový graf $(\overline{x}_m, s_m)$, kde $m = 1, 2, \ldots, M$, kde $M = 100000$. Dokreslite do grafu čiernou farbou také body, pre ktoré platí $t_{W,m} = \left|\frac{\overline{x}_m - \mu}{s_m}\sqrt{n}\right| < t_{n-1}(\alpha/2)$, ako aj hranice, ktoré definujú také body $(\overline{x}_m, s_m)$, pre ktoré $t_{W,m} = t_{n-1}(\alpha/2)$. Vypočítajte pravdepodobnosť pokrytia 95% DIS pre μ ako podiel $\sum_m I\left(t_{W,m} < t_{n-1}(\alpha/2)\right)/M$. Zvoľte (a) $n = 5$, (b) $n = 50$ a (c) $n = 100$.*

Príklad 12 (necentrálne t-rozdelenie) *Nakreslite distribučnú funkciu necentrálneho t-rozdelenia $t_{n-1,\lambda}$, kde $\delta = \mu - \mu_0$ a $\lambda = \delta/(\sigma/\sqrt{n})$. Použite $\mu_0 = 0$, $\delta = 1$, $\sigma = 1.4$ a $n = 26$. Vypočítajte pravdepodobnosť nad kvantilom $x_{0.975}$ pod krivkou hustoty tohoto rozdelenia.*

Príklad 13 (necentrálne t-rozdelenie) *Nakreslite hustoty jedného centrálneho a štyroch necentrálnych t-rozdelení $t_{n-1,\lambda}$ ($\delta = \mu - \mu_0$ a $\lambda = \delta/(\sigma/\sqrt{n})$) do jedneho obrázka tak, aby boli odlíšiteľné farbou alebo typom čiary. Použite $\mu_0 = 0$, $\delta = 0, 0.5, 0.8, 1$ a 1.2, $\sigma = 1.4$ a $n = 26$.*

Nech $X \sim N(\mu, \sigma^2)$, kde σ^2je naznáma. Testujme $H_{01} : \mu = \mu_0$ vs. $H_{11} : \mu \neq \mu_0$. Potom

$$U_W = \frac{n}{n-1}T_W^2,\ U_{LR} = n \ln\left(1 + \frac{U_W}{n}\right) \text{ a } U_S = \frac{nU_W}{n + U_W}.$$

Príklad 14 (pravdepodobnosť empirickej CHPD) *Nech* $X \sim N(\mu, \sigma^2)$, *kde* $\mu = 0$ *a* $\sigma^2 = 2.5^2$. *Testujte* $H_0 : \mu = 0$ *oproti* $H_{11} : \mu \neq 0$ *na* $\alpha = 0.05$, σ^2 *je neznáme. Použite* R *na simuláciu empirickej* $\Pr(CHPD)$, *kde počet simulácií je* $M = 10000$ *a rozsah náhodného výberu je* $n = 5, 10, 20, 30, 50, 100, 500, 1000$. *Vypočítajte testovacie štatistiky (1)* $u_{W,m} = \frac{n}{n-1} t^2_{W,m}$, *(2)* $u_{LR,m} = n \ln\left(1 + \frac{u_{W,m}}{n}\right)$ *a (3)* $u_{S,m} = \frac{n}{n+u_{W,m}} u_{W,m}$, *kde* $m = 1, 2, \ldots, M$. *Vypočítajte relatívnu početnosť* p *zamietnutých* H_{01} *na hladine významnosti* $\alpha = 0.05$ *medzi* M *testami, kde* $p = \Pr(CHPD) = \frac{\sum_{m=1}^{M} I(H_0 \text{ zamietame})}{M}$. *Porovnajte výsledky vzhľadom na rýchlosť konvergencie s rastúcim* n.

Príklad 15 (pravdepodobnosť empirickej CHPD t***-testu)*** *Nech* $X \sim N(\mu, \sigma^2)$, *kde* $\mu = 500$ *a* $\sigma^2 = 100$. *Testujte* $H_0 : \mu = 500$ *oproti* $H_1 : \mu > 500$, *ak* $\alpha = 0.05$, σ *je neznáme. Použite* R *na simuláciu empirickej* $\Pr(CHPD)$, *kde počet simulácií je* $M = 10000$ *a rozsah náhodného výberu je* $n = 20$ *pre jednovýberový Studentov t-test o strednej hodnote* μ. *Použite funkciu* `t.test(x, alternative = "greater", mu = mu0)` *a pre každú testovaciu štatistiku* $t_m, m = 1, 2, \ldots, M$ *vypočítajte p-hodnotu a jej štandardnú chybu za platnosti* H_0. *Ide o zistenie relatívnej početnosti p zamietnutých* H_0 *na hladine významnosti* $\alpha = 0.05$ *medzi* M *testami, kde* $p = \Pr(CHPD) = \frac{\sum_{m=1}^{M} I(H_0 \text{ zamietame})}{M}$.

Príklad 16 (pravdepodobnosť teoretickej CHPD pri danom n***)*** *Nech* $X \sim N(\mu, \sigma^2)$, *kde* σ^2*je naznáma. Majme* $H_{01} : \mu = \mu_0$ *vs.* $H_{11} : \mu \neq \mu_0$. *Nakreslite tri vyššie uvedené pravdepodobnosti ako funkcie n, t.j. (1)* $\alpha_W(n)$, *(2)* $\alpha_{LR}(n)$ *a (3)* $\alpha_S(n)$, *kde* $n \in (1, 1000)$. *Označte v grafe také n, pre ktoré* $\alpha(n)$ *prvýkrát prekroči hranicu* 0.052 *pre* U_W *a* U_{LR} *zhora a hranicu* 0.048 *pre* U_S *zdola. Porovnajte výsledky vzhľadom na rýchlosť konvergencie s rastúcim n. Vypočítajte podiely* n_W/n_S *a* n_{LR}/n_S *a okomentujte.*

Príklad 17 *Nech* $X \sim N(\mu, \sigma^2)$, *kde odhady* $\overline{x} = 4$ *a* $s^2 = 2.89^2$. *Rozsah náhodného výberu* $n = 25$.

(a) *Vypočítajte silu* $1 - \beta$ *pre* $\mu_0 = 2.5$ *a* $\mu_1 = 4$ *(*μ_1 *predstavuje hodnotu* μ *za platnosti* H_1*) za predpokladu, že* $\sigma = 2.5$.

(b) *Použite* R *na simuláciu hustoty rozdelenia* $t_{n-1,\lambda}$ *testovacích štatistík* $t^{(m)}_{W,\lambda} = \frac{\overline{x}_m - \mu_0}{s_m}\sqrt{n}$ *(necentrálne t-rozdelenie s* $n - 1$ *stupňami voľnosti a parametrom necentrality* λ*), kde* $n = 25$, $\lambda = 3$, $m = 1, 2, \ldots, M$, *pri* $M = 20000$ *opakovaniach. Na základe tohoto rozdelenia výpočítajte silu testu pre* $\mu_0 = 2.5$ *a* $\mu_1 = 4$ *(pozri obrázok 45). (1)* $X \sim N(4, 2.5^2)$ *a (2)* $X \sim [pN(4, 2.5^2) + (1-p)N(4, 4.5^2)]$, *kde* $p = 0.9$.

Príklad 18 (empirická silofunkcia t***-testu)*** *Nech (a) X pochádza z normálneho rozdelenia,* $X \sim N(\mu_1, 100^2)$, *a (b) X pochádza zo zmesi dvoch normálnych rozdelení,* $X \sim [pN(\mu_1, 100^2) + (1-p)N(\mu_1, 200^2)]$, *kde* $p = 0.9$. *Rozsah náhodného výberu* $n = 20$. *Použite* R *na simuláciu empirickej silofunkcie pre jednovýberový Studentov t-test. Testujeme* $H_0 : \mu = 500$ *oproti* $H_1 : \mu \neq 500$, *kde* $\mu_1 = 450, 460, \ldots, 640, 650$ *(ide o obojstrannú alternatívu). Použite funkciu* `t.test(x,mu=500)`, *na výpočet každej testovacej štatistiky* $t_m, m = 1, 2, \ldots, M$, *kde* $M = 10000$, *vypočítajte p-hodnotu korešpondujúcu* t_m *a porovnajte ju s hladinou významnosti* $\alpha = 0.05$. *Tak získate empirickú silofunkciu* $1 - \widehat{\beta}(\mu_1)$ *pri danej alternatíve. Do grafu zakreslite* $1 - \widehat{\beta}(\mu_1)$ *pri danej alternatíve ako aj ich štandardné chyby* $SE[\widehat{1 - \widehat{\beta}(\mu_1)}] = \sqrt{\frac{(1-\widehat{\beta}(\mu_1))\widehat{\beta}(\mu_1)}{M}}$ *v podobe chybovej úsečky* $1 - \widehat{\beta}(\mu_1) \pm SE[\widehat{1 - \widehat{\beta}(\mu_1)}]$. *Do grafu vkreslite aj teoretickú silofunkciu* $1 - \beta(\mu_1), \mu_1 \in \langle 450, 650 \rangle$ *(použite funkciu* `power.t.test()`*).*

***Príklad 19 (MC odhad koeficientu spoľahlivosti** $1-\alpha$**)** Vypočítajte v R MC odhad koeficientu spoľahlivosti (pravdepodobnosti pokrytia) pre pravostranný (horný) 95% JIS pre σ^2 pri $M = 1000$ a $n = 20$. Tento JIS je ekvivalentný s testom H_{02} oproti H_{12}. (a) Nech $X \sim N(0,4)$, (b) $X \sim \chi^2(2)$ a (c) $X \sim [pN(0,4) + (1-p)N(0,9)]$, kde $p = 0.9$.*

***Príklad 20 (minimálny rozsah súboru)** Vypočítajte v R minimálny rozsah náhodného výberu pre test $H_{03} : \sigma^2 \geq \sigma_0^2$ oproti $H_{13} : \sigma^2 < \sigma_0^2$ pri $\alpha = 0.05$ a $1-\beta = 0.8$, ak podiel $\frac{\sigma_0^2}{\sigma^2}$ je rovný (a) 1.1, (b) 1.5 a (c) 5.*

***Príklad 21 (minimálny rozsah** n**)** Vypočítajte minimálny rozsah n pre $\rho = 0.1, 0.2, \ldots, 0.9$, $\rho_0 = 0$ pri $\alpha = 0.05$, $1-\beta = 0.8$ a obojstrannej alternatíve H_{11}.*

***Príklad 22 (minimálny rozsah** n**)** Vypočítajte minimálny rozsah n pre $\rho = 0.1, 0.2, \ldots, 0.9$, ρ_0 vždy o 0.1 menšie ako ρ, pri $\alpha = 0.05$, $\beta = 0.8$ a obojstrannej alternatíve H_{11}.*

***Príklad 23 (konvergencia** ρ **a** ξ **k normálnemu rozdeleniu)** Urobte v R simuláciu pseudonáhodných čísel z $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, kde $\mu_1 = 0, \mu_2 = 0$, $\sigma_1 = 1, \sigma_2 = 1$ (pozri príklad zo SI1), kde $n = 5, 10, 20, 50$ a 100, $M = 10000$. Použite (a) $\rho = 0$, (b) $\rho = 0.50$ a (c) $\rho = 0.9$. Pre každé $m = 1, 2, \ldots, M$, vypočítajte Pearsonov korelačný koeficient r_m a Fisherovu Z-premennú $z_{R,m}$. Zobrazte histogramy simulovaných r_m a $z_{R,m}$ a superponujte ich teoretickými hustotami prislúchajúcich normálnych rozdelení.*

***Príklad 24 (minimálny rozsah** N**)** Vypočítajte minimálny rozsah n pre $p = 0.1, 0.2, \ldots, 0.9$, $p_0 = 0$ pri $\alpha = 0.05$, $\beta = 0.8$ a obojstrannej alternatíve H_{11}. Skontrolujte, či je splnená Haldova podmienka. Ak nie je, doplňte minimálne N, ktoré túto podmienku splňa.*

***Príklad 25 (minimálny rozsah** N**)** Vypočítajte minimálny rozsah N pre $p = 0.1, 0.2, \ldots, 0.9$, p_0 vždy o 0.1 menšie ako p, pri $\alpha = 0.05$, $\beta = 0.8$ a obojstrannej alternatíve H_{11}. Skontrolujte, či je splnená Haldova podmienka. Ak nie je, doplňte minimálne N, ktoré túto podmienku splňa.*

***Príklad 26 (pravdepodobnost pokrytia)** Nech $X \sim Bin(N,p)$, kde $N = 30$ a $p = 0.8$ a pravdepodobnosť úspechu $\widehat{p} = \frac{24}{30} = 0.8$, kde $x = 24$ a $N = 30$. Waldov 95% empirický DIS pre p je rovný $(d,h) = (0.657, 0.943)$. Vypočítajte pravdepodobnosť pokrytia tohoto intervalu. Pozn.: pravdepodobnosť pokrytia Waldovho 95% DIS pre p vypočítame nasledovne*

$$\Pr(pokrytie) = \sum_j \Pr(X = Np_j : p \in Waldov\ 95\%\ DIS\ pre\ p_j),$$

kde $p_j \in \mathcal{M}_J = \left\{\frac{1}{30}, \frac{2}{30}, \ldots, 1-\frac{1}{30}\right\}$, t.j. ide o súčet takých funkčných hodnôt pravdepodobnostnej funkcie v bodoch Np_j, kde $p \in$ Waldovmu 95% DIS pre p_j. Výsledky usporiadajte do tabuľky, ktorej stĺpce budú x_j, p_j, d_j (dolná hranica Waldovho 95% DIS pre p_j), h_j (horná hranica Waldovho 95% DIS pre p_j), $\Pr(pokrytie)$ a pokrytie (indikácia toho, či p patrí alebo nepatrí Waldovmu 95% DIS pre p_j).

***Príklad 27 (pravdepodobnost pokrytia)** Nech $X_i \sim Bin(N, p_i)$. Vypočítajte pravdepodobnosti pokrytia (a) Waldovho 95% DIS a (b) skóre 95% DIS pre každé p_i, kde p_i patria množine $\mathcal{M}_I = \left\langle \frac{1}{N}, 1-\frac{1}{N} \right\rangle$, sú ekvidištantne vzdialené medzi $\frac{1}{N}$ a $1-\frac{1}{N}$ a ich počet $M = 5000$. Nakreslite obrázok, kde na osi x budú p_i a na osi y pravdepodobnosť pokrytia $\Pr_i(pokrytie)$. Zvoľte (a) $N = 30$, (b) $N = 100$ a (c) $N = 1000$. Pozn.: pravdepodobnosti pokrytia Waldovho 95% DIS pre p_i vypočítame nasledovne*

$$\Pr_i(\textit{pokrytie}) = \sum_j \Pr(X = Np_j : p_i \in \textit{Waldov } 95\% \textit{ DIS pre } p_j),$$

kde $p_j \in \mathcal{M}_J = \left\{\frac{1}{N}, \frac{2}{N}, \ldots, 1 - \frac{1}{N}\right\}$*, t.j. ide o súčet takých funkčných hodnôt pravdepodobnostnej funkcie v bodoch* Np_j*, kde* $p_i \in$ *Waldovmu* 95% *DIS pre* p_j.

Príklad 28 (pravdepodobnosť pokrytia, ak $\sigma_1^2 = \sigma_2^2$ ***sú neznáme)*** *Nech* $X_j \sim N(\mu_j, \sigma^2)$*, kde* $j = 1, 2$*,* $\mu_1 = 20$*,* $\mu_2 = 35$ *a* $\sigma^2 = 100$*. Pomocou simulačnej štúdie (*$M = 100000$*) vypočítajte pravdepodobnosť pokrytia* 95% *DIS pre* $\mu_1 - \mu_2$ *ako podiel* $\sum_{m=1}^{M} I\left(t_{W,m} < t_{df}(\alpha/2)\right)/M$*, kde* $t_{W,m}$ *sú testovacie štatistiky klasického dvojvýberového t-testu. Zvoľte (a)* $n = 5$*, (b)* $n = 50$ *a (c)* $n = 100$*.*

Príklad 29 (pravdepodobnosť pokrytia, ak $\sigma_1^2 \neq \sigma_2^2$ ***sú neznáme)*** *Nech* $X_j \sim N(\mu_j, \sigma_j^2)$*, kde* $j = 1, 2$*,* $\mu_1 = 20$*,* $\mu_2 = 35$*,* $\sigma_1^2 = 100$ *a* $\sigma_2^2 = 150$*. Pomocou simulačnej štúdie (*$M = 100000$*) vypočítajte pravdepodobnosť pokrytia* 95% *DIS pre* $\mu_1 - \mu_2$ *ako podiel* $\sum_{m=1}^{M} I\left(t_{W,m} < t_{df}(\alpha/2)\right)/M$*, kde* $t_{W,m}$ *sú (1) testovacie štatistiky klasického dvojvýberového t-testu a (2) testovacie štatistiky Welchovho dvojvýberového t-testu. Zvoľte (a)* $n = 5$*, (b)* $n = 50$ *a (c)* $n = 100$*.*

Príklad 30 (pravdepodobnosť pokrytia, ak $\sigma_1^2 = \sigma_2^2$ ***sú neznáme)*** *Nech* $X_j \sim [pN(\mu_j, \sigma^2) + (1-p)N(\mu_j, \sigma_a^2)]$*, kde* $p = 0.9$*,* $j = 1, 2$*,* $\mu_1 = 20$*,* $\mu_2 = 35$*,* $\sigma^2 = 100$ *a* $\sigma_a^2 = 400$*. Pomocou simulačnej štúdie (*$M = 100000$*) vypočítajte pravdepodobnosť pokrytia* 95% *DIS pre* $\mu_1 - \mu_2$ *ako podiel* $\sum_{m=1}^{M} I\left(t_{W,m} < t_{df}(\alpha/2)\right)/M$*, kde* $t_{W,m}$ *sú testovacie štatistiky klasického dvojvýberového t-testu. Zvoľte (a)* $n = 5$*, (b)* $n = 50$ *a (c)* $n = 100$*.*

Príklad 31 (pravdepodobnosť pokrytia, ak $\sigma_1^2 \neq \sigma_2^2$ ***sú neznáme)*** *Nech* $X_j \sim [pN(\mu_j, \sigma_j^2) + (1-p)N(\mu_j, \sigma_{ja}^2)]$*, kde* $p = 0.9$*,* $j = 1, 2$*,* $\mu_1 = 20$*,* $\mu_2 = 35$*,* $\sigma_1^2 = 100$*,* $\sigma_2^2 = 150$*,* $\sigma_{1a}^2 = 400$ *a* $\sigma_{2a}^2 = 450$*. Pomocou simulačnej štúdie (*$M = 100000$*) vypočítajte pravdepodobnosť pokrytia* 95% *DIS pre* $\mu_1 - \mu_2$ *ako podiel* $\sum_{m=1}^{M} I\left(t_{W,m} < t_{df}(\alpha/2)\right)/M$*, kde* $t_{W,m}$ *sú (1) testovacie štatistiky klasického dvojvýberového t-testu a (2) testovacie štatistiky Welchovho dvojvýberového t-testu. Zvoľte (a)* $n = 5$*, (b)* $n = 50$ *a (c)* $n = 100$*.*

Príklad 32 (sila a silofunkcia testu rozdielu stredných hodnôt) *Použite R na simuláciu hustoty rozdelenia testovacej štatistiky* $T_{df,\lambda}$ *dvojvýberového testu rozdielu stredných hodnôt* $\mu_1 - \mu_2$ *za platnosti dvojstrannej alternatívy* H_{11} *pri* $M = 20000$ *opakovaniach. Túto hustotu v podobe histogramu v relatívnej škále zakreslite do obrázka a superponujte ju s teoretickou hustotou. Vypočítajte silu za platnosti alternatívy* $H_{11} : \mu_1 - \mu_2 = 2$*, kde* $n_1 = n_2 = 25$*. Použite (a) klasický dvojvýberový t-test a (b) Welchov dvojvýberový t-test. (1)* $X_1 \sim N(4, 2.5^2)$*,* $X_2 \sim N(2, 2.5^2)$ *a (2)* $X_1 \sim [pN(4, 2.5^2) + (1-p)N(4, 4.5^2)]$*,* $X_2(2, 4.5^2)$*, kde* $p = 0.9$*.*

Príklad 33 (maximálne vierohodné odhady; nádor prsníka) *Majme početnosti subjektov* X_1*, ktoré majú rozšírené metastázy nádoru prsníka, kde* $X_1 \sim Bin(N_1, p_1)$ *a početnosti subjektov* X_2*, ktoré majú lokalizované metastázy nádoru prsníka, kde* $X_2 \sim Bin(N_2, p_2)$*. (a) Aplikujte funkciu vierohodnosti* $L(\boldsymbol{\theta}|\mathbf{x}_1, \mathbf{x}_2)$*, kde* $\boldsymbol{\theta} = (\theta, \eta)^T$*, logaritmus pomeru šancí* $\theta = \ln \frac{p_1/(1-p_1)}{p_2/(1-p_2)}$ *a rušivý parameter* $\eta = \ln \frac{p_2}{1-p_2}$ *na dáta v tabuľke a vypočítajte* $\widehat{\theta}$*. (b) Nakreslite funkciu vierohodnosti ako aj profilovú funkciu vierohodnosti a DIS. Zopakujte pre* $n_1 = 6$ *a* $n_2 = 0$*. (c) Vypočítajte vierohodnostný* 95% *DIS pre* θ *pomocou metodiky* 15% *cut-off štandardizovanej profilovej funkcie vierohodnosti. DIS dokreslite do jedného obrázka k profilovej funkcii vierohosnoti v jej* 15% *cut-off.*

Tabuľka 2: Početnosti subjektov s rozšírenými a lokalizovanými metastázami

metastázy	rozšírené	lokalizované	spolu
áno	5	1	6
nie	10	9	19
spolu	15	10	25

Príklad 34 (sila ANOVA F-testu a minimálny rozsah) *Majme štyri populácie, ktorých stredné hodnoty sú $\mu_1 = 390$, $\mu_2 = 405$, $\mu_3 = 415$ a $\mu_4 = 410$. Predpokladajme, že $Y_{ji} \sim N(\mu_j, \sigma_e^2)$.*
(a) Vypočítajte silu $1 - \beta$ ANOVA F-testu rovnosti stredných hodnôt za predpokladu, že $K = 6$, $\sigma_e^2 = 20^2$ a $\alpha = 0.05$.
(b) Vypočítajte silu $1 - \beta$ ANOVA F-testu rovnosti stredných hodnôt za predpokladu, že $K = 6$, $\sigma_e^2 = 10^2$ a $\alpha = 0.05$.
(c) Použite R na simuláciu hustoty necentrálneho $F_{df_A, df_e, \lambda^2}$ testovacích štatistík $F_{W,\lambda}^{(m)}$ (necentrálne F rozdelenie s df_A a df_e stupňami voľnosti a parametrom necentrality λ^2), kde $\alpha = 0.05$, $K = 6$, $\sigma_e^2 = 20^2$, $\lambda^2 = \sum_{j=1}^{J} (\mu_j - \widehat{\mu})^2 / (\sigma_e^2/K)$, $m = 1, 2, \ldots, M$, pri $M = 1000$ opakovaniach.