Ústav matematiky a statistiky
Přírodovědecká fakulta
Masarykova univerzita

Analýza prežívania

Zadania domácich úloh

Stanislav Katina

katina@math.muni.cz

22. decembra 2014

Inštrukcie k DÚ: Odovzdáva sa jeden pdf súbor nazvaný priezvisko-meno-text-analprez-2014.pdf (obsahuje riešenia príkladov, obrázky, R-kód napísaný v TEXu), jeden zdrojový súbor naprogramovaných funkcií priezvisko-meno–source-analprez-2014.r a jeden súbor R-kódu konkrétnych zadaní z DÚ priezvisko-meno-priklady-analprez-2014.r, ktorý používa tento zdrojový kód. Na písanie R-kódu odporúčam TEX balíček `listings` a vytvoreniu prostredia v hlavičke dokumentu ako

```
\lstset{language=R, % nastavenie jazyka R
basicstyle=\footnotesize\ttfamily, % typ pisma R-kodu
commentstyle=\ttfamily\color{farba1}, % farba  komentara k funkciam
numberstyle=\color{farba2}\footnotesize, % farba a velkost cislovania
numbers=left, % cislovanie vlavo
stepnumber=1, % cislovanie po krokoch jedna
frame=leftline, % vytvorenie lavej hranicnej ciary
breaklines=true} % zalomenie riadkov
```

a potom v texte medzi begin a end.
DÚ je potrebné odovzdať 7 dní pred termínom skúšky, na ktorý sa prihlásite.

Tabuľka 1: Dáta AML

skupina	čas po kompletný relaps (v týždňoch)
A	9, 13, 13+, 18, 23, 28+, 31, 34, 45+, 48, 161+
B	5, 5, 8, 8, 12, 16+, 23, 27, 30, 33, 43, 45

Tabuľka 2: Dáta Koncentrácie

koncentrácie	čas do objavenia sa nádoru (v týždňoch)
2.0	41+, 41+, 47, 47+, 47+, 58, 58, 58, 100+, 117
1.5	43+, 44+, 45+, 67, 68+, 136, 136, 150, 150, 150
0	73+, 74+, 75+, 76, 76, 76+, 99, 166, 246+,

Príklad 1 (k-ty moment času prežívania) *Nech nezáporná náhodná veličina T je charakterizovaná funkciou prežívania $S(T)$. Nech je k-ty moment, $E[T^k]$, konečný, $E[T^k] < \infty, k \in \mathbb{N}$.*
a) Ukážte, že platí $E[T] = \sum_{t\in\mathbb{N}_0} \Pr(T \geq t) = \sum_{t\in\mathbb{N}_0} 1 - F(T) = \sum_{t\in\mathbb{N}_0} S(T)$. Použite pri tom definíciu strednej hodnoty $E[T] = \sum_{t\in\mathbb{N}_0} t\Pr(t)$ a pomocné tvrdenie $\underbrace{1+1+\ldots+1}_{t\text{-krát}} = \sum_{\xi=1}^{t} 1 = \sum_{\xi=0}^{t-1} 1 = \sum_{\xi<t,\xi\in\mathbb{N}_0} 1$.
b) Ukážte, že platí $E[T] = \int_0^\infty S(t)\,dt$. Použite pri tom definíciu strednej hodnoty $E[T] = \int_0^\infty tf(t)dt$, aplikujte vlastnosti súm z (a) ako aj $\int_0^\infty S(t)\,dt = \int_0^\infty (\int_0^t 1dx) S(t)\,dt$. Výpočet Vám uľahčí metóda per-partes.
c) Pomocou metódy per-partes ukážte, že $E[T^k] = k\int_0^\infty t^{k-1}S(t)dt$. DÚ

Príklad 2 (odvodenie charakteristík prežívania) *(a)* ***stredná hodnota zostatkového života*** *– pomocou metódy per partes ukážte, že platí posledná rovnosť v nasledovnom vzťahu*

$$mrl(t) = E[T-t|T \geq t] = \frac{\int_t^\infty (u-t)f(u)du}{S(t)} = \frac{\int_t^\infty S(u)du}{S(t)}.$$

(b) ***funkcia prežívania*** *– ukážte, že platí posledná rovnosť v nasledovnom vzťahu*

$$S(t) = \int_t^\infty xf(x)dx = \exp\left(\int_0^t \lambda(x)dx\right) = \exp\left(-\Lambda(t)\right) = \frac{mrl(0)}{mrl(t)} \exp\left(-\int_0^t \frac{du}{mrl(u)}\right).$$

(c) ***riziko*** *– ukážte, že platí posledná rovnosť v nasledovnom vzťahu*

$$\lambda(t) = -\frac{\partial}{\partial t}\ln S(t) = \frac{f(t)}{S(t)} = \left(\frac{\partial}{\partial t}mrl(t) + 1\right)\frac{1}{mrl(t)}.$$

(d) ***hustota*** *– ukážte, že platí posledná rovnosť v nasledovnom vzťahu*

$$f(t) = -\frac{\partial}{\partial t}S(t) = \lambda(t)S(t) = \left(\frac{\partial}{\partial t}mrl(t) + 1\right)\frac{mrl(0)}{(mrl(t))^2}\exp\left(-\int_0^t \frac{du}{mrl(u)}\right).$$

Táto rovnosť vyplýva zo vzťahu $f(t) = \lambda(t)S(t)$ a z výsledku predchádzajúcich dvoch príkladov (b) a (c). DÚ

Príklad 3 (porovnanie odhadov funkcie prežívania) *Použitím Taylorovho rozvoja (prvého rádu) mínus logaritmu $\widehat{\Lambda}_{NA}(t)$ ukážte, že $\widehat{\Lambda}_{KM}(t)$ je prvým členom tohoto rozvoja.* DÚ

Príklad 4 (programovanie a aplikácia, rozptyl priemerného zostatkového života) *(a) Naprogramujte v R funkciu na výpočet odhadu rozptylu priemerného zostatkového života v čase t. (b) Vypočítajte odhadu rozptylu priemerného zostatkového života v čase $t = 0$ a $t = 30$ pre AML dáta (skupina A).* DÚ

Príklad 5 (programovanie a aplikácia, medián zostatkového života) *(a) Naprogramujte v R funkciu na výpočet odhadu mediánu zostatkového života v čase t. (b) Vypočítajte medián zostatkového života v čase $t = 0$ a $t = 30$ pre AML dáta (skupina A).* DÚ

Príklad 6 (programovanie a aplikácia, rozptyl mediánu zostatkového života) *(a) Naprogramujte v R funkciu na výpočet odhadu rozptylu mediánu zostatkového života v čase t. (b) Vypočítajte odhadu rozptylu mediánu zostatkového života v čase $t = 0$ a $t = 30$ pre AML dáta (skupina A).* DÚ

Príklad 7 (programovanie a aplikácia, IS a pásy spoľahlivosti pre $S(t)$) *(a) Naprogramujte v R algoritmus na výpočet Nairových $100 \times (1-\alpha)\%$ pásov spoľahlivosti pre funkciu prežívania v arcus-sínusovej odmocninovej škále funkcie prežívania $S(t)$, kde $t \in \langle t_0, t_{\max}\rangle$. Výsledok porovnajte s IS pre $S(t)$ v škále $S(t)$. Na obrázku zobrazte IS ako vertikálne úsečky v tvare písmena* `I` *v časoch zlyhania a pásy spoľahlivosti pomocou funkcie* `polygon()`. DÚ

Príklad 8 (asymptotická normalita S_{KW}) *Porovnajte v R asymptotické rozdelenie S_{KW} s jej exaktným rozdelením pre (a) $n_1 = n_2 = n_3 = 5$, (b) $n_1 = n_2 = 5$ a $n_3 = 50$, (c) $n_1 = n_2 = n_3 = 50$, (d) $n_1 = n_2 = 50$ a $n_3 = 100$ a (e) $n_1 = n_2 = n_3 = 100$. Na výpočet asymptotickej hustoty použite funkciu* `dchisq()` *a na výpočet asymptotickej distribučnej funkcie použite funkcie* `dchisq()` *a* `cumsum()`*. Na výpočet exaktnej hustoty použite funkciu* `dwilcox()` *a na výpočet exaktnej distribučnej funkcie použite funkciu* `dKruskalWallis()` *v knižnici* `SuppDists`*. Teoretické a exaktné rozdelenie superponujte v podobe (1) hustoty, (2) distribučnej funkcie a (3) qq-diagramu s qq-priamkou (na x-ovej osi bude sekvencia x od teoreticky možného $\min(S_{KW})$ po teoreticky možné $\max(S_{KW})$ a na y-ovej osi teoretické kvantily y vypočítané pomocou funkcie* `qchisq()`*; qq-priamka bude prechádzať bodmi $(\widetilde{x}_{0.25}, \widetilde{x}_{0.75})$ a $(\widetilde{y}_{0.25}, \widetilde{y}_{0.75})$ alebo alternatívne bude qq-priamku reprezentovať os prvého a tretieho kvadrantu).* DÚ

Príklad 9 (koncentrácie; relatívne riziko, trend, odklon od trendu) *Majme 3 rôzne koncentrácie látky ($konc_1 = 2.0$, $konc_2 = 1.5$ a $konc_3 = 0$) a hľadáme jej účinok na pacientov, u ktorých sme sledovali objavenie sa nádoru. Zaujíma nás testovanie (a) $H_0 : \lambda_1(t) = \lambda_2(t) = \lambda_3(t)$ oproti $H_1 : \exists$ aspoň jedno $i < j, \lambda_i(t) = \theta\lambda_j(t)$ pomocou testovacích štatistík Q_{GB}, Q_{CM}, Q_{TW} a Q_{PP}; (b) $H_0 : \lambda_1(t) = \lambda_2(t) = \lambda_3(t)$ oproti $\lambda_1(t) = w_1^*\lambda_3(t), \lambda_2(t) = w_2^*\lambda_3(t)$, kde $w_j^* = konc_j$, $j = 1, 2$ pomocou testovacej v statistiky Q_{trend}. Otestujte ak odklon od trendu pomocou Q_{resid}. (c) Vypočítajte relatívne riziko (1) $\widehat{\theta}_P$, $Var[\widehat{\theta}_P]$ a 95%IS pre θ_P, (2) $\widehat{\theta}_{MH}$ a 95%IS pre θ_{MH} a (3) $\widehat{\theta}^*_{MH}$ (ide o relatívne riziká $konc_1 : konc_2$ a $konc_1 : konc_3$).* DÚ