

Ústav matematiky a statistiky  
Přírodovědecká fakulta  
Masarykova univerzita

---

## Analýza prežívania

*Vybrané kapitoly z neparametrických metód pre  
cenzurované dáta, zadania príkladov a domácich úloh a  
niektoré riešenia*

Stanislav Katina

katina@math.muni.cz

10. decembra 2014

# Obsah

<b>1</b>	<b>Udalosti a cenzúrovanie</b>	<b>1</b>
1.1	Cenzúrovanie I. typu . . . . .	2
1.2	Cenzúrovanie II. typu . . . . .	2
1.3	Progresívne cenzúrovanie . . . . .	3
1.4	Ľubovoľné a náhodné cenzúrovanie . . . . .	3
1.5	Intervalové cenzúrovanie I. typu . . . . .	3
1.6	Intervalové cenzúrovanie II typu . . . . .	4
<b>2</b>	<b>Základné charakteristiky prežívania</b>	<b>5</b>
<b>3</b>	<b>Odhady základných charakteristík prežívania</b>	<b>5</b>
3.1	Empirická funkcia prežívania . . . . .	6
3.2	Odhady funkcie prežívania . . . . .	7
3.3	Odhady rizika a kumulatívneho rizika . . . . .	11
3.4	Odhady rozptylu kumulatívneho rizika a funkcie prežívania . . . . .	12
3.5	Empirické intervaly a pásy spoľahlivosti . . . . .	14
3.5.1	Intervaly spoľahlivosti . . . . .	14
3.5.2	Intervaly spoľahlivosti pre kumulatívne riziko . . . . .	15
3.5.3	Intervaly spoľahlivosti pre funkciu prežívania . . . . .	15
3.6	Odhady strednej hodnoty a mediánu prežívania . . . . .	16
3.7	Odhad strednej hodnoty a mediánu zostatkového života a ich rozptyl . . . . .	18
<b>4</b>	<b>Testy na porovnanie kriviek prežívania</b>	<b>21</b>
<b>5</b>	<b>Charakteristiky definované sčítacím procesom</b>	<b>26</b>
<b>6</b>	<b>Softvérová implementácia charakteristík prežívania</b>	<b>28</b>
<b>7</b>	<b>Štruktúra funkcií na cvičenia a do DÚ</b>	<b>30</b>

# 1 Udalosti a cenzúrovanie

**Udalosťou** nazývame *ukončenie pozorovania z dôvodu zlyhania alebo smrti pacienta* – do konca sledovaného obdobia.

Príklady udalostí:

- **overall survival** – smrť z akéhokoľvek dôvodu,
- **progression-free survival** – prvé znaky progresie choroby alebo smrť,
- **disease-free survival** – prvé znovuobjavenie sa choroby alebo smrť,
- **event-free survival** – prvé znovuobjavenie sa choroby, objavenie sa inej špecifikovanej choroby alebo smrť,
- **disease-specific survival (cause-specific survival)** – smrť ako dôsledok špecifikovanej choroby,
- **relapse-free survival (recurrence-free survival)** – prvé znaky recidívy (opakovania sa) choroby,
- **time-to-progression** – prvé znaky progresie choroby.

V mnohých prípadoch v danom pokuse musíme pozorovania ukončiť z dôvodu iného ako je zlyhanie alebo smrť pacienta. Teda do konca sledovaného obdobia dôjde k úmrtiu len niektorých subjektov (pacientov), zatiaľ čo u ostatných k úmrtiu do konca sledovaného obdobia buď nedôjde alebo sa tieto subjekty z pozorovania stratia. Cenzúrované pozorovanie obsahuje v sebe len čiastočnú informáciu o čase úmrtia.

Predpokladajme, že  $n$  subjektov s rovnakými znakmi začneme pozorovať v čase  $t = 0$ . Pre ich časy do zlyhania  $T_1, T_2, \dots, T_n$  platí, že sú nezávislé, rovnako rozdelené náhodné premenné, kde náhodná veličina  $T_i$  ( $i = 1, 2, \dots, n$ ) má distribučnú funkciu  $F(t)$ , o ktorej obyčajne predpokladáme, že je absolutne spojitá. Teoreticky by sme mali pozorovať subjekty tak dlho, pokiaľ nezískame úplný náhodný výber časov do zlyhania pre všetkých  $n$  pozorovaní. Ide o veľmi dlhý časový interval, a preto nemôžeme túto štúdiu alebo experiment z rôznych dôvodov urobiť až do konca. Musíme preto vhodným spôsobom rozhodnúť o ukončení pozorovaní, k čomu nás môžu donútiť napr. technické alebo ekonomické dôvody, a teda získame neúplné *cenzurované dáta*. V medicíne sa môžu vyskytnúť z nasledujúcich príčin:

- **ukončenie štúdie (termination of the study)**: pacient prežije časový interval experimentu,
- **konkurenčné riziko (competing risk)**: pacient zomrie z iného dôvodu, ako v dôsledku sledovanej choroby,
- **prerušenie/vysadenie liečby (drop-out)**: pacient preruší liečbu a odíde z kliniky predčasne, napr. z dôvodu zlých vedľajších účinkov liečby, pacient sa sám rozhodne nepokračovať v liečbe,
- **strata z ďalšieho sledovania (loss to follow-up)**: pacient sa rozhodne presťahovať a nemáme o ňom už žiadne informácie.

Za predpokladu, že experiment prebehne bez cenzúrovania,  $i$ -ty jedinec z populácie o rozsahu  $n$  zomrie za čas  $T_i$  od začiatku sledovania. Čas, ktorý mienime sledovať  $i$ -teho jedinca označíme  $C_i$  ( $C_i > 0$ ).

Podľa spôsobu ukončenia experimentu môžeme rozdeliť cenzúrovanie do nasledovných typov:

- cenzúrovanie I. typu,
- cenzúrovanie II. typu,
- progresívne (zrýchlené) cenzúrovanie,
- ľubovoľné a náhodné cenzúrovanie,
- intervalové cenzúrovanie I. typu,
- intervalové cenzúrovanie II. typu.

## 1.1 Cenzúrovanie I. typu

Predpokladáme, že všetkých  $n$  jedincov vstupuje do experimentu súčasne. Ide o *cenzurovanie časom*, pretože si zvolíme pevné číslo  $t_c$ , ktoré nazveme *fixovaný cenzurujúci čas*, teda jediná príčina cenzúrovania je plánované ukončenie experimentu. Takto získame informáciu o prvých  $d$  pozorovaniach  $T^{(1)} < T^{(2)} < \dots < T^{(d)}$ , kde  $T^{(d)} < t_c < T^{(d+1)}$  a o ostatných  $n - d$  nepozorovaných zlyhaniach vieme iba to, že nastali až po cenzurujúcom čase  $t_c$ . *Počet skutočne pozorovaných zlyhaní  $d$  je náhodná veličina*, ktorá môže nadobúdať hodnoty  $0, 1, \dots, n$ .

Namiesto pozorovania  $T_i$  pozorujeme  $X_1, X_2, \dots, X_n$ , kde

$$X_i = \min(T_i, t_c) = \begin{cases} T_i, T_i \leq t_c, & \text{pre necenzúrované } X_i \\ t_c, T_i > t_c, & \text{pre cenzúrované } X_i \end{cases}.$$

Pre distribučnú funkciu  $X$  platí  $\Pr(T > t_c) > 0$  v bode  $x = t_c$ , pretože  $\Pr(X = t_c) = \Pr(T > t_c) = 1 - F(t_c) > 0$ . To znamená, že  $X$  je zmiešaná náhodná premenná so spojitým a diskretným komponentom. Kumulatívna distribučná funkcia  $M(X)$  náhodnej premennej  $X$  bude teda spojitá do času  $t_c$  (v čase  $< t_c$ ) a v čase  $t_c$  skočí do jednotky (diskretný komponent). Majme binárnu premennú  $\delta$ , ktorá indikuje, či bol čas zlyhania pozorovaný alebo cenzúrovaný,

$$\delta_i = \begin{cases} 1, T_i \leq t_c, & \text{pre necenzúrované } X_i \\ 0, T_i > t_c, & \text{pre cenzúrované } X_i \end{cases}.$$

Skutočným pozorovaniam potom zodpovedá náhodný vektor  $(X_i, \delta_i)$ ,  $i = 1, 2, \dots, n$ .

Pozn.: Treba si uvedomiť, že ak  $(\delta = 0$  a  $T \leq t_c)$  implikuje, že čas zlyhania bol presne  $T = t_c$ , čo nastane s nulovou pravdepodobnosťou, ak je  $T$  spojitá premenná. Pre diskretné premenné môžeme definovať  $t_c$  rovné poslednému dostupnému času zlyhania, ktorý môže byť pozorovaný. Potom  $\Pr(\delta = 0 \wedge T \leq t_c) \neq 0$ .

## 1.2 Cenzúrovanie II. typu

Aj pri tomto type cenzúrovania vstupuje do experimentu súčasne všetkých  $n$  jedincov. Zvolíme pevné číslo  $d$ , ktoré nazveme *fixovaný počet zlyhaní*. Ukončenie potom nastáva po vopred zvolenom počte  $d$  zlyhaní, tj. ide o *cenzúrovanie zlyhaním*, kde  $d = [np] + 1, p \in (0, 1), T^{(1)} < T^{(2)} < \dots < T^{(d)}$ .

Odtiaľ vidíme, že  $d < n$  a všetkých zostávajúcich  $n - d$  pozorovaní má čas zlyhania väčší ako  $T^{(d)}$ . Dostávame tak celistvú informáciu o prvých  $d$  pozorovaniach. *Náhodnou veličinou je čas trvania experimentu  $T^{(d)}$* .

Namiesto pozorovania  $T_i$  pozorujeme  $X_1, X_2, \dots, X_n$ , kde

$$X_i = \min(T_i, T^{(d)}) = \begin{cases} T_i, T_i \leq T^{(d)}, & \text{pre necenzúrované } X_i \\ T^{(d)}, T_i > T^{(d)}, & \text{pre cenzúrované } X_i \end{cases}.$$

Skutočnému pozorovaniu potom zodpovedá náhodný vektor  $(X_i, \delta_i)$ , kde

$$\delta_i = \begin{cases} 1, T_i \leq T^{(d)}, & \text{pre necenzúrované } X_i \\ 0, T_i > T^{(d)}, & \text{pre cenzúrované } X_i \end{cases}.$$

### 1.3 Progresívne cenzúrovanie

Cieľom tohto typu cenzúrovania je čo najväčšie skrátenie času experimentu. Znovu vstupuje do experimentu súčasne všetkých  $n$  jedincov. Teraz máme dve možnosti postupu.

**Zrýchlené cenzúrovanie I. typu.** Ide o *cenzúrovanie zlyhaním*, kedy zvolíme čísla  $t_{ci}, i = 1, 2, \dots, k$ , ktoré nazveme *fixované cenzurujúce časy*. Vieme, že  $t_{c1} < t_{c2} < \dots < t_{ck}$ . V čase  $t_{ci}$  vyradíme  $m_i$  subjektov, t.j. v čase  $t_{c1}$  vyradíme  $m_1$  subjektov, v čase  $t_{c2}$  vyradíme  $m_2$  subjektov, ..., v čase  $t_{ck}$  vyradíme  $m_k$  subjektov. Po  $k$ -tom kroku máme vyradených  $m_1 + m_2 + \dots + m_k$  subjektov. *Náhodnou veličinou je počet skutočne pozorovaných zlyhaní*  $d \in \{0, 1, \dots, n\}$ .

**Zrýchlené cenzúrovanie II. typu.** Ide o *cenzúrovanie časom*, kedy zvolíme čísla  $d_i$ , ktoré nazveme *fixované počty zlyhaní*. Vyradenie teda nastáva po vopred zvolenom počte  $d_i$  zlyhaní, kde  $d_i = [np_i] + 1, p_i \in (0, 1)$ , t.j. po  $d_1$  zlyhaniach vyradíme  $m_1$  subjektov, po  $d_2$  zlyhaniach vyradíme  $m_2$  subjektov, ..., po  $d_k$  zlyhaniach vyradíme  $m_k$  subjektov. Po  $k$ -tom kroku máme vyradených  $m_1 + m_2 + \dots + m_k$  subjektov. *Náhodnou veličinou je čas trvania experimentu*.

### 1.4 Ľubovoľné a náhodné cenzúrovanie

V predchádzajúcich typoch cenzúrovania vstupuje do experimentu súčasne všetkých  $n$  jedincov a časy zlyhania sú v takom poradí, v akom nastali. Ale takýto ideálny prípad nemusí nastať.

*Čas do cenzúrovania  $i$ -teho jedinca* označme  $C_i$ , potom  $C_1, C_2, \dots, C_n$  sú nezávislé, rovnako rozdelené náhodné premenné, kde náhodná veličina  $C_i$  ( $i = 1, \dots, n$ ) má distribučnú funkciu  $G(t)$ .

Pozorujeme  $X_1, X_2, \dots, X_n$ , kde

$$X_i = \min(T_i, C_i) = \begin{cases} T_i, T_i \leq C_i, & \text{pre necenzúrované } X_i \\ C_i, T_i > C_i, & \text{pre cenzúrované } X_i \end{cases}.$$

Skutočnému pozorovaniu potom zodpovedá náhodný vektor  $(X_i, \delta_i), i = 1, 2, \dots, n$ , kde  $X_i = \min(T_i, C_i)$ ,

$$\delta_i = \begin{cases} 1, T_i \leq C_i, & \text{pre necenzúrované } X_i \\ 0, T_i > C_i, & \text{pre cenzúrované } X_i \end{cases}.$$

Ak  $C_i$  považujeme za ľubovoľné konštanty a nie za náhodné veličiny, kde napr.  $C_i = c, i = 1, 2, \dots, n$ , potom hovoríme o **ľubovoľnom cenzúrovaní**. Cenzúrovanie I. typu je jeho špeciálnym prípadom.

Ak považujeme  $C_i$  za náhodné veličiny, potom hovoríme o **náhodnom cenzúrovaní** a predpokladáme nezávislosť náhodných veličín  $C_i$  a  $T_i$ . Tento typ cenzúrovania vzniká často v medicínskych štúdiách.

Pozn.: Je to však dosť silný predpoklad. Treba si uvedomiť súvislosť s typom cenzúry drop-out (z dôvodu vedľajších účinkov liečby). Ale práve za tohoto predpokladu je funkcia vierohodnosti ľahko spočítateľná.

### 1.5 Intervalové cenzúrovanie I. typu

Majme  $n$  subjektov. Označme  $T_i, i = 1, 2, \dots, n$ , **nepozorovateľné časy zlyhania**. Skutočnému pozorovaniu potom zodpovedá náhodný vektor  $(X_i, \delta_i)$ , kde  $X_i = C_i$  sú časy cenzúr a  $\delta_i = I(T_i \leq C_i)$ , t.j.

$$\delta_i = \begin{cases} 1, T_i \leq C_i, & \text{pre necenzúrované } X_i \\ 0, T_i > C_i, & \text{pre cenzúrované } X_i \end{cases}.$$

**Príklad 1 (nádor pľúc, animálny model)** *Laboratórne myši sú injektované látkou, ktorá spôsobuje nádor. Keďže tento druh nádoru nie je smrteľný, je potrebné myš najprv zabiť, aby sme zistili, či bol nádor indukovaný, t.j. po časovom úseku náhodnej dĺžky  $C$  je myš zabitá, aby sme zistili, či sa nádor vyvinul alebo nie. Endpoint záujmu je čas  $T$  do objavenia sa nádoru.*

pred

## 1.6 Intervalové cenzúrovanie II typu

Majme opäť  $n$  subjektov. Označme  $T_i, i = 1, 2, \dots, n$ , **nepozorovateľné časy zlyhania**. Vieme len, že  $T_i$  nastalo buď vnútri nejakého náhodného časového intervalu, pred jeho ľavou hranicou alebo po jeho pravej hranici. Označme  $C_{1i}$  a  $C_{2i}$  časy dvoch vyšetrení a indikačné funkcie definujeme nasledovne  $\delta_{1i} = I(T_i \leq C_{1i})$ ,  $\delta_{2i} = I(C_{1i} < T_i \leq C_{2i})$  a  $\delta_{3i} = I(T_i > C_{2i})$ , t.j.

$$\delta_{1i} = \begin{cases} 1, T_i \leq C_{1i}, & \text{pre necenzurované } X_i \\ 0, T_i > C_{1i}, & \text{pre cenzurované } X_i \end{cases},$$

$$\delta_{2i} = \begin{cases} 1, C_{1i} < T_i \leq C_{2i}, & \text{pre necenzurované } X_i \\ 0, T_i > C_{2i}, & \text{pre cenzurované } X_i \end{cases}$$

a nakoniec  $\delta_{3i} = 0$ .

Potom máme nasledovné tri možnosti:

1. udalosť mohla nastať niekedy pred prvým vyšetrením  $C_{1i}$ , kde  $\delta_{1i} = 1$  a  $\delta_{2i} = \delta_{3i} = 0$ ,
2. udalosť mohla nastať niekedy medzi prvým a druhým vyšetrením, t.j. v intervale  $(C_{1i}, C_{2i})$ , kde  $\delta_{1i} = 0$ ,  $\delta_{2i} = 1$  a  $\delta_{3i} = 0$ ,
3. udalosť sa do druhého vyšetrenia nevyskytla, t.j. mohla nastať niekedy po  $C_{2i}$  (ale nevieme kedy), kde  $\delta_{1i} = 0$ ,  $\delta_{2i} = 0$  a  $\delta_{3i} = 0$ .

Nech  $X_{1i} = C_{1i}$  a  $X_{2i} = C_{2i}$ . Skutočnému pozorovaniu potom zodpovedá náhodný vektor

$$(X_{1i}, X_{2i}, \delta_{1i}, \delta_{2i}).$$

Všimnime si, že  $\delta_{3i}$  nie je potrebné použiť, pretože nemáme ďalšie vyšetrenie po  $C_{2i}$ . Keby sme mali  $C_{3i}$  alebo aj ďalšie (po ňom nasledujúce) vyšetrenia, hovorili by sme **zovšeobecnenom intervalovom cenzúrovaní**.

**Príklad 2 (nádor pľúc, pacienti)** *Pacienti navštevovali kliniku opakovane každých 4 až 6 mesiacov, kde pozorovania sú buď intervaly  $(C_{1i}, C_{2i})$  ak sa retrakcia prsníka vyskytla medzi poslednými dvoma návštevami alebo  $(C_{2i}, \infty)$ , ak sa do  $C_{2i}$  retrakcia nevyskytla.*

## 2 Základné charakteristiky prežívania

**Príklad 3 (*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  $\underbrace{1 + 1 + \dots + 1}_{t\text{-krát}} = \sum_{\xi=1}^t 1 =$*

$$\sum_{\xi=0}^{t-1} 1 = \sum_{\xi < t, \xi \in \mathbb{N}_0} 1.$$

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  $\int_0^\infty S(t) dt = \int_0^\infty (\int_0^t 1 dx) S(t) dt$ . 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 4 (stredná hodnota zostatkového života)** *Pomocou metódy per partes ukážte, že platí posledná rovnosť v nasledovnom vzťahu*

DÚ

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

**Príklad 5 (funkcia prežívania)** *Ukážte, že platí posledná rovnosť v nasledovnom vzťahu*

DÚ

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

**Príklad 6 (riziko)** *Ukážte, že platí posledná rovnosť v nasledovnom vzťahu*

DÚ

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

**Príklad 7 (hustota)** *Ukážte, že platí posledná rovnosť v nasledovnom vzťahu*

DÚ

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

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

## 3 Odhady základných charakteristík prežívania

**Príklad 8 (Akútna myelogénna leukémia, AML)** *AML pacienti boli po absolvovaní chemoterapie a zmiernení príznakov náhodne rozdelení do dvoch skupín. Prvá skupina (skupina A) dostala udržiavajúcu chemoterapiu a druhá (kontrolná; skupina B) nie (pozri tabuľku). Cieľom bolo zistiť, či udržiavajúca chemoterapia predlžuje čas remisie (opätovného zhoršenia stavu).*

cvič.

Tabuľka 1: Dáta AML

skupina	čas po kompletný relaps (v týždňoch)	$n$	udalostí	cenzúr
A	9, 13, 13+, 18, 23, 28+, 31, 34, 45+, 48, 161+	11	7	4
B	5, 5, 8, 8, 12, 16+, 23, 27, 30, 33, 43, 45	12	11	1

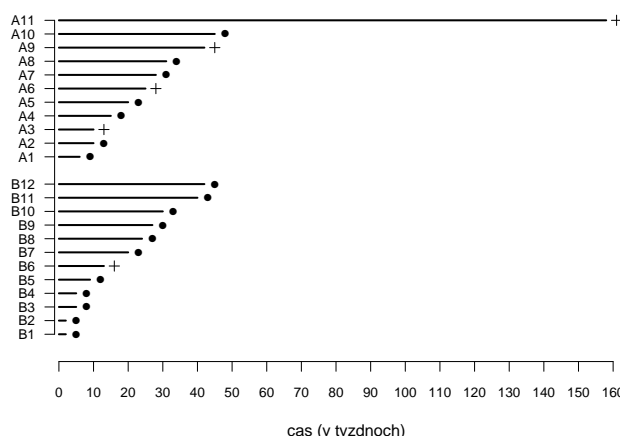
Tri náhľady na problém analýzy AML dáta:

- **problém 1:** po odstránení cenzúrovaných pozorovaní;
- **problém 2:** po ošetrení cenzúrovaných pozorovaní, ktoré zoberieme do úvahy akoby boli udalostami (zlyhaniami) a
- **problém 3:** berúc do úvahy cenzúrované pozorovania.

Tabuľka 2: Základné charakteristiky AML podľa skupín (v týždňoch)

skupina	problém 1		problém 2		problém 3	
	A	B	A	B	A	B
$\bar{x}$	25.1	21.7	38.5	21.3	52.6	22.7
$\tilde{x}$	23.0	23.0	28.0	19.5	31.0	23.0

**Príklad 9 (časy do zlyhania alebo cenzúry, AML)** Nakreslite časy do zlyhania a časy do cenzúry pre AML dáta ako horizontálne úsečky ukončené bodom  $\bullet$  (pre zlyhania) a bodom  $+$  (pre cenzúry). Na  $y$ -ovej osi označte pacientov skratkami A1–A11 a B1–B12 a na  $x$ -ovej osi zobrazte časy po desiatich týždňoch od 0 po 160.

Obr. 1: Časy do zlyhania (ozn.  $\bullet$ ) a časy do cenzúry (ozn.  $+$ ) pre AML dáta

### 3.1 Empirická funkcia prežívania

Majme diskretný prípad. Predpokladajme, že časy do zlyhania sú už zoradené, t.j. označenie  $t_{(1)} \leq t_{(2)} \leq \dots \leq t_{(I)}, I \leq n$ , preznačíme na  $t_1 \leq t_2 \leq \dots \leq t_I$ . Potom **empirickú funkciu prežívania** vypočítame nasledovne

$$\hat{S}_n(t) = \frac{\#\text{pozorovaní} > t}{n} = \frac{\#\{t_i > t\}}{n} = \frac{\sum_{i=1}^n I(t_i > t)}{n},$$

kde  $n$  je rozsah súboru. Ide teda o podiel tých pacientov, ktorí sú stále v riziku.  $S_n(t)$  je sprava spojitá schodovitá funkcia so schodmi smerom nadol v každom  $t_i$ . Ide o konzistentný odhad skutočnej funkcie prežívania  $S(t)$ . Exaktné rozdelenie  $nS_n(t)$  pre každé fixované  $t$  je  $Bi(n, p)$ , kde  $n$  je počet pozorovaní a  $p = \Pr(T > t)$ . Na základe centrálnej limitnej vety pre každé fixované  $t$  platí asymptoticky  $S_n(t) \stackrel{\mathcal{D}}{\sim} N(p, p(1-p)/n)$ .

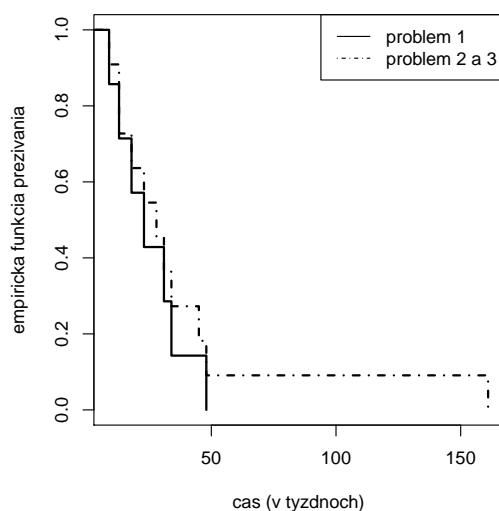


**Príklad 10 (programovanie; empirická funkcia prežívania)** Naprogramujte v  $\mathbb{R}$  algoritmus na výpočet empirickej funkcie prežívania. cvič.

**Príklad 11 (empirická funkcia prežívania, AML)** Vypočítajte empirickú funkciu prežívania vo všetkých časoch  $t$  pre skupinu  $A$  (1) pre problém 1, (2) problém 2 a (3) pre problém 3. Zobrazte tieto tri funkcie do jedného obrázka ako schodovité funkcie. Použite tri rôzne typy čiar alebo tri rôzne farby a pridajte legendu a popis osí. cvič.

**Riešenie** (čiastkové – pre problém 3; pozri tabuľku)

$t$	0	9	13	18	23	28	31	34	45	48	161
$\hat{S}_n(t)$	$\frac{11}{11}$	$\frac{10}{11}$	$\frac{8}{11}$	$\frac{7}{11}$	$\frac{6}{11}$	$\frac{5}{11}$	$\frac{4}{11}$	$\frac{3}{11}$	$\frac{2}{11}$	$\frac{1}{11}$	$\frac{0}{11}$



Obr. 2: Empirická funkcia prežívania pre AML dáta (problem 1, 2 a 3)

**Príklad 12 (programovanie; početnosti  $d_i$  a  $n_i$ )** Naprogramujte v  $\mathbb{R}$  algoritmus na výpočet  $d_i$  a  $n_i$  pre všetky časy zlyhania  $t_i$ . Výstupom bude tabuľka so stĺpcami  $t_i$ ,  $d_i$  a  $n_i$ . cvič.

### 3.2 Odhady funkcie prežívania

Klasický Kaplan-Meierov (KM) odhad funkcie prežívania je definovaný nasledovne

$$\hat{S}_{\text{KM}}(t) = \prod_{i:t_i \leq t} (1 - \hat{\lambda}(t_i)),$$

kde  $\hat{\lambda}(t_i) = \frac{d_i}{n_i}$  (ide o maximálne vierohodný odhad rizika),  $d_i$  je počet zlyhaní (úmrtí) v čase  $t_i \leq t$  a  $n_i$  počet ľudí v riziku tesne pred časom  $t$ . **KM odhad kumulatívneho rizika**  $\hat{\Lambda}_{\text{KM}}(t) = -\ln \hat{S}_{\text{KM}}(t) = -\sum_{i:t_i \leq t} \ln(1 - \hat{\lambda}(t_i))$ . Ak je najväčší čas cenzúrovaný čas  $c_{\text{max}} > t_I$ , potom podľa definície  $S(t_I) = S(c_{\text{max}})$  a keďže funkcia vierohodnosti nezávisí na  $S(t)$  pre  $t > c_{\text{max}}$ , nie je možný žiadny odhad za  $c_{\text{max}}$ .

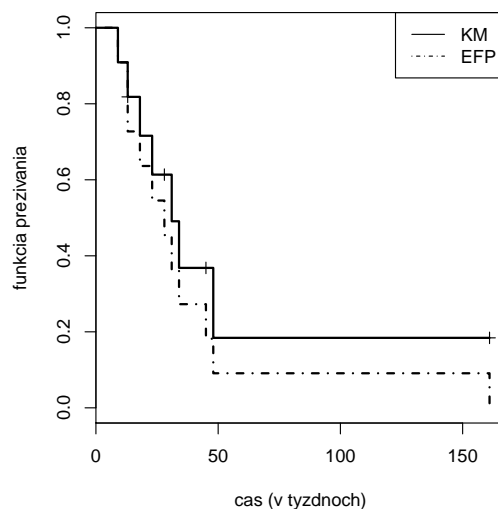
**Príklad 13 (programovanie; KM odhad funkcia prežívania)** Naprogramujte v  $\mathbb{R}$  algoritmus na výpočet KM odhadu funkcie prežívania  $\hat{S}_{\text{KM}}(t)$  vo všetkých časoch zlyhania. cvič.

**Príklad 14 (KM odhad funkcia prežívania, AML)** *Vypočítajte  $\widehat{S}_{KM}(t)$  vo všetkých časoch zlyhania a porovnajte ho s  $\widehat{S}_n(t)$  pre skupinu A. Zobrazte tieto dve funkcie do jedného obrázka ako schodovité funkcie. Použite dve rôzne typy čiar alebo dve rôzne farby a pridajte legendu a popis osí. cvič.*

**Riešenie** (čiastkové; pozri tabuľku)

$t$	0	9	13	13+	18	23	28+	31	34	45+	48	161+
$\widehat{S}_n(t)$	$\frac{11}{11}$	$\frac{10}{11}$	$\frac{8}{11}$	$\frac{8}{11}$	$\frac{7}{11}$	$\frac{6}{11}$	$\frac{5}{11}$	$\frac{4}{11}$	$\frac{3}{11}$	$\frac{2}{11}$	$\frac{1}{11}$	$\frac{0}{11}$
$\widehat{S}_{KM}(t)$	1	0.91	0.82	0.82	0.72	0.61	0.61	0.49	0.37	0.37	0.18	0.18

$$\begin{aligned} \widehat{S}_{KM}(0) &= 1 \\ \widehat{S}_{KM}(9) &= \widehat{S}_{KM}(0) \times \frac{11-1}{11} \\ \widehat{S}_{KM}(13) &= \widehat{S}_{KM}(9) \times \frac{10-1}{10} \\ \widehat{S}_{KM}(13+) &= \widehat{S}_{KM}(13) \times \frac{9-0}{9} = \widehat{S}_{KM}(13) \\ \widehat{S}_{KM}(18) &= \widehat{S}_{KM}(13) \times \frac{8-1}{8} \\ \widehat{S}_{KM}(23) &= \widehat{S}_{KM}(18) \times \frac{7-1}{7} \\ \widehat{S}_{KM}(28+) &= \widehat{S}_{KM}(23) \times \frac{6-0}{6} = \widehat{S}_{KM}(23) \\ \widehat{S}_{KM}(31) &= \widehat{S}_{KM}(23) \times \frac{5-1}{5} \\ \widehat{S}_{KM}(34) &= \widehat{S}_{KM}(31) \times \frac{4-1}{4} \\ \widehat{S}_{KM}(45+) &= \widehat{S}_{KM}(34) \times \frac{3-0}{3} = \widehat{S}_{KM}(34) \\ \widehat{S}_{KM}(48) &= \widehat{S}_{KM}(34) \times \frac{2-1}{2} \\ \widehat{S}_{KM}(161+) &= \widehat{S}_{KM}(48) \times \frac{1-0}{1} = \widehat{S}_{KM}(48) \end{aligned}$$



Obr. 3: Dva odhady funkcie prežívania –  $\widehat{S}_{KM}(t)$  a  $\widehat{S}_n$  – pre AML dáta (skupina A)

**Modifikovaný KM odhad funkcie prežívania** je definovaný nasledovne

$$\widehat{S}_{KM\text{mod}}(t) = \prod_{i:t_i \leq t} (1 - \widehat{\lambda}_i), \text{ kde } 1 - \widehat{\lambda}_i = \begin{cases} 1 - \frac{d_i}{n_i}, & \text{ak } d_i = 1 \\ \prod_{j=0}^{d_i-1} \left(1 - \frac{1}{n_i-j}\right), & \text{ak } d_i > 1 \end{cases}$$

**Breslowov (B) odhad funkcie prežívania** definujeme nasledovne

$$\widehat{S}_B(t) = \exp\left(-\widehat{\Lambda}_{NA}(t)\right) = \prod_{i:t_i \leq t} \exp\left(-\widehat{\lambda}_i\right),$$

kde  $\widehat{\Lambda}_{\text{NA}}(t) = \sum_{i:t_i \leq t} \widehat{\lambda}_i = \sum_{i:t_i \leq t} \frac{d_i}{n_i}$  je Nelson-Aalenov (NA) odhad kumulatívneho rizika.

Flemingom a Harringtonom (FH) modifikovaný B odhad funkcie prežívania definujeme nasledovne

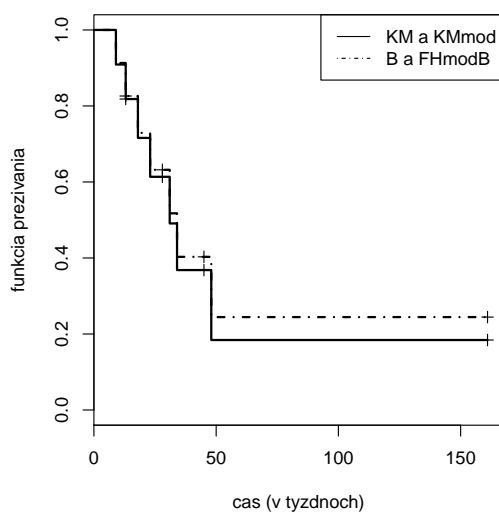
$$\widehat{S}_{\text{FHmodB}}(t) = \exp\left(-\widehat{\Lambda}_{\text{FHmodNA}}(t)\right),$$

kde  $\widehat{\Lambda}_{\text{FHmodNA}} = \sum_{i:t_i \leq t} \left(\sum_{j=0}^{d_i-1} \frac{1}{n_i-j}\right)$  je FH modifikovaný NA odhad kumulatívneho rizika.

**Príklad 15 (programovanie; odhady funkcie prežívania)** Naprogramujte v  $\mathbb{R}$  algoritmus na výpočet  $\widehat{S}_{\text{KMmod}}(t)$ ,  $\widehat{S}_B(t)$  a  $\widehat{S}_{\text{FHmodB}}(t)$  v každom čase zlyhania. cvič.

**Príklad 16 (porovnanie; odhady funkcie prežívania, AML)** Vypočítajte nasledovné odhady funkcií prežívania –  $\widehat{S}_{\text{KM}}(t)$ ,  $\widehat{S}_{\text{KMmod}}(t)$ ,  $\widehat{S}_B(t)$  a  $\widehat{S}_{\text{FHmodB}}(t)$  pre AML dáta (skupina A). Zobrazte tieto štyri funkcie do jedného obrázka ako schodovité funkcie. Použite štyri rôzne typy čiar alebo štyri rôzne farby a pridajte legendu a popis osí. cvič.

#	t	d	n	KM	KMmod	B	FHmodB
1	9	1	11	0.9090909	0.9090909	0.9131007	0.9131007
2	13	1	10	0.8181818	0.8181818	0.8262077	0.8262077
3	18	1	8	0.7159091	0.7159091	0.7291257	0.7291257
4	23	1	7	0.6136364	0.6136364	0.6320630	0.6320630
5	31	1	5	0.4909091	0.4909091	0.5174894	0.5174894
6	34	1	4	0.3681818	0.3681818	0.4030212	0.4030212
7	48	1	2	0.1840909	0.1840909	0.2444447	0.2444447



Obr. 4: Štyri odhady funkcie prežívania –  $\widehat{S}_{\text{KM}}(t)$ ,  $\widehat{S}_{\text{KMmod}}(t)$ ,  $\widehat{S}_B(t)$  a  $\widehat{S}_{\text{FHmodB}}(t)$  – pre AML dáta (skupina A)

**Príklad 17 (odhady funkcie prežívania)** Vypočítajte odhady  $\widehat{S}_{\text{KMmod}}(t)$ ,  $\widehat{S}_B(t)$  a  $\widehat{S}_{\text{FHmodB}}(t)$  v čase 12 a 88 (pozri tabuľku). pred

t	$d_i$	$n_i$
4.5	1	70
11.5	2	68
16.0	1	65
20.7	2	55
20.8	1	53
31.0	1	47
34.5	1	45
46.0	1	34
61.0	1	25
87.5	5	15

Pre 2 zhody v čase 12 platí:

$$\widehat{S}_{KM}(12) = (69/70) (66/68) = 0.9567$$

$$\widehat{S}_{KMmod}(12) = (69/70) (67/68) (66/67) = (69/70) (66/68) = 0.9567$$

Druhý prípad predstavuje úpravu  $\widehat{S}_{KM}(t)$  pri zlome zhôd rozdelením času 11.5 na 11.48 a 11.52, z čoho je zrejmá invariantnosť  $\widehat{S}_{KM}(t)$  na zhody.

$$\widehat{S}_B(12) = \exp[-(1/70 + 2/68)] = 0.9572$$

$$\widehat{S}_{FHmodB}(12) = \exp[-(1/70 + 1/68 + 1/67)] = 0.9570$$

Pre 5 zhôd v čase 88 platí:

$$\widehat{S}_{KM}(88) = 0.5294$$

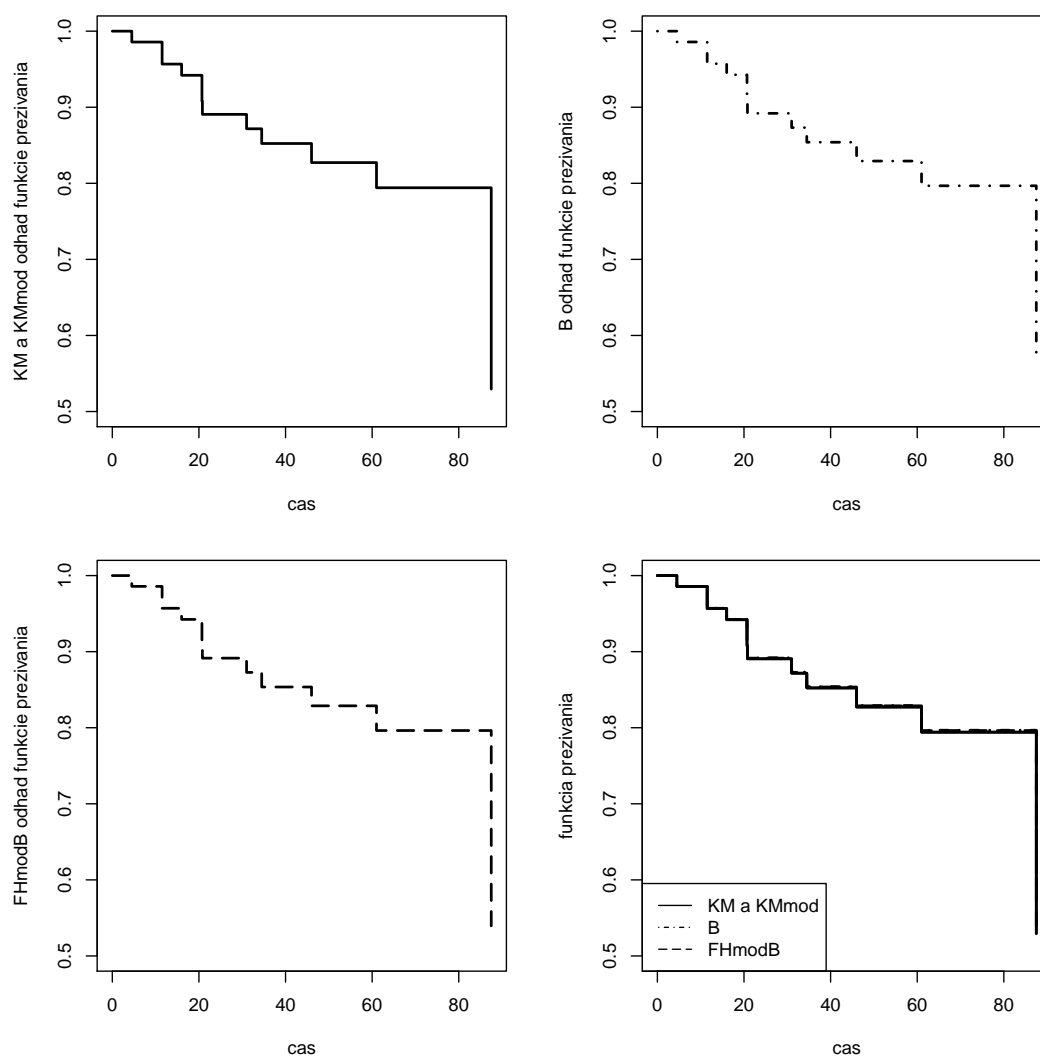
$$\widehat{S}_{FHmodB}(88) = 0.5395$$

$$\widehat{S}_B(88) = 0.5709$$

$\widehat{S}_{FHmodB}(t)$  dáva vo všeobecnosti odhad bližšie ku  $\widehat{S}_{KM}(t)$  a je menší ako  $\widehat{S}_B(t)$  pri zhodách.

**Príklad 18 (porovnanie; odhady funkcií prežívania)** *Vypočítajte nasledovné odhady funkcií prežívania –  $\widehat{S}_{KM}(t)$ ,  $\widehat{S}_{KMmod}(t)$ ,  $\widehat{S}_B(t)$  a  $\widehat{S}_{FHmodB}(t)$  pre dáta (pozri tabuľku). Zobrazte tieto štyri funkcie do jedného obrázka ako schodovité funkcie. Použite štyri rôzne typy čiar alebo štyri rôzne farby a pridajte legendu a popis osí. Použite dáta z predchádzajúceho príkladu.*

cvič.



Obr. 5: Štyri odhady funkcie prežívania –  $\widehat{S}_{KM}(t)$ ,  $\widehat{S}_{KMmod}(t)$ ,  $\widehat{S}_B(t)$  a  $\widehat{S}_{FHmodB}(t)$  – pre AML dáta (skupina A)

9	#	t	d	n	KM	KMmod	B	FHmodB
10	# [1, ]	4.5	1	70	0.9857143	0.9857143	0.9858158	0.9858158
11	# [2, ]	11.5	2	68	0.9567227	0.9567227	0.9572435	0.9570334
12	# [3, ]	16.0	1	65	0.9420039	0.9420039	0.9426294	0.9424225
13	# [4, ]	20.7	2	55	0.9077492	0.9077492	0.9089677	0.9084623
14	# [5, ]	20.8	1	53	0.8906218	0.8906218	0.8919781	0.8914822
15	# [6, ]	31.0	1	47	0.8716724	0.8716724	0.8732004	0.8727148
16	# [7, ]	34.5	1	45	0.8523019	0.8523019	0.8540099	0.8535351
17	# [8, ]	46.0	1	34	0.8272342	0.8272342	0.8292578	0.8287967
18	# [9, ]	61.0	1	25	0.7941449	0.7941449	0.7967421	0.7962991
19	# [10, ]	87.5	5	15	0.5294299	0.5294299	0.5708907	0.5395385

Rozdiely medzi odhadmi funkcií prežívania sú malé, s výnimkou veľmi malého  $n$  alebo pri veľkom počte zhôd. Odhad  $\widehat{S}_{KM}(t)$  je invariantný na zhody. Odhad  $\widehat{S}_B(t)$  môžeme vylepšiť použitím odhadu  $\widehat{S}_{FHmodB}(t)$  v prípade veľkého počtu zhôd. Ak  $e^{-x} \approx 1 - x$ , pre malé  $x$ , potom sú oba odhady  $\widehat{S}_{KM}(t)$  a  $\widehat{S}_B(t)$  podobné v prípade, že nárasty  $d\widehat{\Lambda}$  sú malé, čo znamená, že v riziku je veľa osôb. Oba odhady sú v skutočnosti asymptoticky ekvivalentné ( $n \rightarrow \infty$ ). Avšak  $e^{-x} \geq 1 - x \Rightarrow \widehat{S}_B(t) \geq \widehat{S}_{KM}(t)$  v prípade, že ide o konečné náhodné výbery (ale ak posledný príspevok je smrť, potom  $\widehat{S}_{KM}(t) = 0$ ,  $\widehat{S}_B(t) > 0$ ).

**Príklad 19 (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

### 3.3 Odhady rizika a kumulatívneho rizika

**Príklad 20 (riziko a časové jednotky)** Závislosť hodnoty rizika na jednotkách času. pred

$$\Pr(t \leq T < t + \Delta t | T \geq t) = \frac{1}{4}$$

Ak  $\Delta t = \frac{1}{3}$  dňa, potom  $\lambda(t) = \frac{1}{\frac{1}{3}} = 0.75$  na deň.

Ak  $\Delta t = \frac{1}{21}$  týždňa, potom  $\lambda(t) = \frac{1}{\frac{1}{21}} = 5.25$  na týždeň.

**Odhad rizika v časovom intervale**  $t_i \leq t < t_{i+1}$  je definovaný nasledovne

$$\tilde{\lambda}(t) = \frac{d_i}{n_i(t_{i+1} - t_i)},$$

**Príklad 21 (riziko, AML)** Vypočítajte odhad rizika  $\tilde{\lambda}(t)$  v intervale  $t_i \leq t < t_{i+1}$  v čase 26 (pre skupinu A). pred

$$\widehat{\lambda}(23) = \frac{1}{7} = 0.143$$

$$\tilde{\lambda}(26) = \widehat{\lambda}(23) = \frac{1}{7(31-23)} = 0.018$$

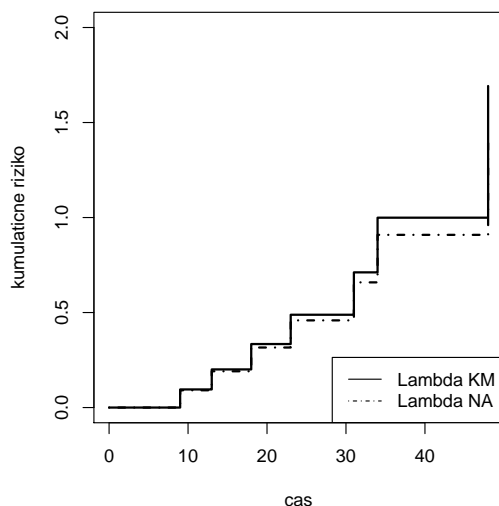
**Príklad 22 (programovanie; riziko)** Naprogramujte v  $\mathbb{R}$  algoritmus na výpočet odhadov rizika  $\widehat{\lambda}(t)$  a  $\tilde{\lambda}(t)$  pre všetky  $t_i$ . cvič.

**Príklad 23 (riziko, AML)** Vypočítajte odhad rizika  $\widehat{\lambda}(t)$  a porovnajte ho s odhadom  $\tilde{\lambda}(t)$  v intervale  $t_i \leq t < t_{i+1}$  pre všetky  $t_i$ . cvič.

20	#	cas	n.i	d.i	lambda.KM	lambda.INT
21	#1	9	11	1	0.0909	0.0227
22	#2	13	10	1	0.1000	0.0200
23	#3	18	8	1	0.1250	0.0250
24	#4	23	7	1	0.1429	0.0179
25	#5	31	5	1	0.2000	0.0667
26	#6	34	4	1	0.2500	0.0179
27	#7	48	2	1	0.5000	NA

**Príklad 24 (programovanie, kumulatívne riziko)** Naprogramujte v  $\mathbb{R}$  algoritmus na výpočet odhadov kumulatívneho rizika  $\hat{\Lambda}_{KM}$  a  $\hat{\Lambda}_{NA}$  pre všetky  $t$ . cvič.

**Príklad 25 (kumulatívne riziko, AML)** Vypočítajte  $\hat{\Lambda}_{KM}$  a  $\hat{\Lambda}_{NA}$  pre všetky  $t$  (pre skupinu A). Zobrazte tieto dve funkcie do jedného obrázka ako schodovité funkcie. Použite dve rôzne typy čiar alebo dve rôzne farby a pridajte legendu a popis osí. cvič.



Obr. 6: Dva odhady kumulatívneho rizika –  $\hat{\Lambda}_{KM}$  a  $\hat{\Lambda}_{NA}$  – pre AML dáta (skupina A)

### 3.4 Odhady rozptylu kumulatívneho rizika a funkcie prežívania

Greenwoodov (G) odhad rozptylu kumulatívneho rizika definujeme nasledovne

$$\hat{\sigma}_G^2(t) = \text{Var}_G \left[ \widehat{\Lambda}_{KM}(t) \right] = \sum_{i:t_i \leq t} \frac{d_i}{n_i(n_i - d_i)},$$

**plug-in odhad rozptylu kumulatívneho rizika** (nazýva sa aj *Kleinov odhad*; *maximálne viero-hodný odhad*)

$$\hat{\sigma}_K^2(t) = \text{Var}_K \left[ \widehat{\Lambda}(t) \right] = \sum_{i:t_i \leq t} \frac{d_i(n_i - d_i)}{n_i^3},$$

**binomický odhad rozptylu kumulatívneho rizika**

$$\hat{\sigma}_B^2(t) = \text{Var}_B \left[ \widehat{\Lambda}(t) \right] = \sum_{i:t_i \leq t} \frac{d_i(n_i - d_i)}{n_i^2(n_i - 1)},$$

**NA odhad rozptylu kumulatívneho rizika** (nazýva sa aj *Aalenov*, *Poissonov* alebo *Tsiatisov* odhad)

$$\hat{\sigma}_T^2(t) = \text{Var} \left[ \widehat{\Lambda}_{NA}(t) \right] = \sum_{i:t_i \leq t} \frac{d_i}{n_i^2},$$

**FH modifikovaný NA odhad rozptylu kumulatívneho rizika**

$$\hat{\sigma}_{FH}^2(t) = \text{Var} \left[ \widehat{\Lambda}_{FH\text{modNA}}(t) \right] = \sum_{i:t_i \leq t} \left( \sum_{j=0}^{d_i-1} \frac{1}{(n_i - j)^2} \right).$$

Odhad  $\hat{\sigma}_G^2(t)$  je o niečo väčší ako  $\hat{\sigma}_T^2(t)$  v prípade vyšších hodnôt odhadu kumulatívneho rizika. Vo všeobecnosti platí  $\hat{\sigma}_K^2(t) < \hat{\sigma}_B^2(t) \leq \hat{\sigma}_T^2(t)$ . Ak sa nevyskytujú zhody, potom  $\hat{\sigma}_B^2(t) = \hat{\sigma}_T^2(t)$ . Všetky odhady sú menšie ako  $\hat{\sigma}_G^2(t)$ . Odhad  $\hat{\sigma}_T^2(t)$  môžeme v prípade zhôd vylepšiť použitím odhadu  $\hat{\sigma}_{FH}^2(t)$ . Odhad  $\hat{\sigma}_T^2(t)$  má menšiu výchyľku ako  $\hat{\sigma}_K^2(t)$ . V praktických aplikáciách sú odhady rozptylu rôzne v závislosti od množstva zhôd v analyzovaných dátach. Ignorovaním sumy  $\sum_{i:t_i \leq t}$  dostaneme **odhady rozptylov rizika v čase  $t$** , ktoré označujeme ako  $\hat{\sigma}_G^2(t_i)$ ,  $\hat{\sigma}_K^2(t_i)$ ,  $\hat{\sigma}_B^2(t_i)$ ,  $\hat{\sigma}_T^2(t_i)$  a  $\hat{\sigma}_{FH}^2(t_i)$ .

**Príklad 26 (rozptyl kumulatívneho rizika, AML)** Vypočítajte odhady rozptylu kumulatívneho rizika  $\hat{\sigma}_G^2(t)$ ,  $\hat{\sigma}_K^2(t)$ ,  $\hat{\sigma}_B^2(t)$ ,  $\hat{\sigma}_T^2(t)$  a  $\hat{\sigma}_{FH}^2(t)$  v čase 26 pre AML dáta (skupina A).

pred

$$\hat{\sigma}_G^2(26) = \text{Var}_G[\widehat{\Lambda}_{KM}(26)] = \frac{1}{11(11-1)} + \frac{1}{10(10-1)} + \frac{1}{8(8-1)} + \frac{1}{7(7-1)} = 0.0619$$

$$\hat{\sigma}_K^2(26) = \text{Var}_K[\widehat{\Lambda}(26)] = \frac{1(11-1)}{11^3} + \frac{1(10-1)}{10^3} + \frac{1(8-1)}{8^3} + \frac{1(7-1)}{7^3} = 0.0478$$

$$\hat{\sigma}_B^2(26) = \text{Var}_B[\widehat{\Lambda}(26)] = \frac{1(11-1)}{11^2(11-1)} + \frac{1(10-1)}{10^2(10-1)} + \frac{1(8-1)}{8^2(8-1)} + \frac{1(7-1)}{7^2(7-1)} = 0.0543$$

$$\hat{\sigma}_T^2(26) = \text{Var}[\widehat{\Lambda}_{NA}(26)] = \frac{1}{11^2} + \frac{1}{10^2} + \frac{1}{8^2} + \frac{1}{7^2} = 0.0543$$

**Príklad 27 (programovanie, rozptyl kumulatívneho rizika)** Naprogramujte v  $\mathbb{R}$  algoritmus na výpočet odhadov rozptylu kumulatívneho rizika  $\hat{\sigma}_G^2(t)$ ,  $\hat{\sigma}_K^2(t)$ ,  $\hat{\sigma}_B^2(t)$ ,  $\hat{\sigma}_T^2(t)$  a  $\hat{\sigma}_{FH}^2(t)$ .

cvič.

**Príklad 28 (rozptyl kumulatívneho rizika, AML)** Vypočítajte odhady rozptylu kumulatívneho rizika  $\hat{\sigma}_G^2(t)$ ,  $\hat{\sigma}_K^2(t)$ ,  $\hat{\sigma}_B^2(t)$ ,  $\hat{\sigma}_T^2(t)$  a  $\hat{\sigma}_{FH}^2(t)$  pre AML dáta (skupina A).

cvič.

#	cas	ni	di	rozptyl.G	rozptyl.K	rozptyl.B	rozptyl.T	rozptyl.FH	
28	#1	9	11	1	0.0091	0.0075	0.0083	0.0083	0.0083
29	#2	13	10	1	0.0202	0.0165	0.0183	0.0183	0.0183
30	#3	18	8	1	0.0381	0.0302	0.0339	0.0339	0.0339
31	#4	23	7	1	0.0619	0.0477	0.0543	0.0543	0.0543
32	#5	31	5	1	0.1119	0.0797	0.0943	0.0943	0.0943
33	#6	34	4	1	0.1952	0.1266	0.1568	0.1568	0.1568
34	#7	48	2	1	0.6952	0.2516	0.4068	0.4068	0.4068

**Greenwoodov odhad rozptylu KM odhadu funkcie prežívania** je definovaný ako

$$\text{Var}_G[\widehat{S}_{KM}(t)] = [\widehat{S}_{KM}(t)]^2 \text{Var}[\ln \widehat{S}_{KM}(t)] = [\widehat{S}_{KM}(t)]^2 \hat{\sigma}_G^2(t).$$

**Aalen-Johansenov odhad rozptylu KM odhadu funkcie prežívania** píšeme ako

$$\text{Var}_{AJ}[\widehat{S}_{KM}(t)] = [\widehat{S}_{KM}(t)]^2 \text{Var}[\widehat{\Lambda}_{NA}(t)] = [\widehat{S}_{KM}(t)]^2 \hat{\sigma}_T^2(t).$$

Oba odhady rozptylu funkcie prežívania odhadnú rozptyl menší ako v skutočnosti (*underestimation*) pre malé a stredne veľké výbery. V priemere je Greenwoodov odhad bližšie k skutočnému rozptylu. Podobne ako pre KM odhad môžeme zapísať aj rozptyl B odhadu funkcie prežívania (Klein 2003)

$$\text{Var}_G[\widehat{S}_B(t)] = [\widehat{S}_B(t)]^2 \text{Var}[\widehat{\Lambda}_{KM}(t)] = [\widehat{S}_B(t)]^2 \hat{\sigma}_G^2(t).$$

a

$$\text{Var}_{NA}[\widehat{S}_B(t)] = [\widehat{S}_B(t)]^2 \text{Var}[\widehat{\Lambda}_{NA}(t)] = [\widehat{S}_B(t)]^2 \hat{\sigma}_T^2(t).$$

Pokiaľ máme v dátach zhody, použijeme nasledovné rozptyly FHmodB odhadu funkcie prežívania

$$\text{Var}_G[\widehat{S}_{FHmodB}(t)] = [\widehat{S}_{FHmodB}(t)]^2 \text{Var}[\widehat{\Lambda}_{KM}(t)] = [\widehat{S}_{FHmodB}(t)]^2 \hat{\sigma}_G^2(t).$$

a

$$\text{Var}_{\text{NA}} \left[ \widehat{S}_{\text{FHmodB}}(t) \right] = \left[ \widehat{S}_{\text{FHmodB}}(t) \right]^2 \text{Var} \left[ \widehat{\Lambda}_{\text{NA}}(t) \right] = \left[ \widehat{S}_{\text{FHmodB}}(t) \right]^2 \widehat{\sigma}_{\text{FH}}^2(t).$$

Vo všeobecnosti potom **odhad rozptylu funkcie prežívania** definujeme ako

$$\text{Var} \left[ \widehat{S}(t) \right] = \left[ \widehat{S}(t) \right]^2 \text{Var} \left[ \ln \widehat{S}(t) \right].$$

**Príklad 29 (rozptyl funkcie prežívania, AML)** *Vypočítajte odhad rozptylu KM a B odhadu funkcie prežívania v čase 13 pre skupinu A pomocou Greenwoodovej formule.*

pred

$$\text{Var}_{\text{G}} \left[ \widehat{S}_{\text{KM}}(13) \right] = 0.82^2 \left( \frac{1}{11(11-1)} + \frac{1}{10(10-1)} \right) = 0.0136$$

$$\text{Var}_{\text{NA}} \left[ \widehat{S}_{\text{B}}(13) \right] = 0.83^2 \left( \frac{1}{11^2} + \frac{1}{10^2} \right) = 0.0125$$

**Príklad 30 (programovanie, rozptyl funkcie prežívania)** *Naprogramujte v  $\mathbb{R}$  algoritmus na výpočet odhadov rozptylu KM a B odhadu funkcie prežívania.*

cvič.

**Príklad 31 (rozptyl funkcie prežívania, AML)** *Vypočítajte odhad rozptylu KM a B odhadu funkcie prežívania vo všetkých časoch pre skupinu A pomocou Greenwoodovej formule.*

cvič.

```

36 rozptyl.funkcie.prezivania(survobjA.KM)
37 # kontrola -- G1 je rozptyl vypočítaný funkciou z knižnice survival
38 #   cas ni di rozptyl.G1 rozptyl.G2
39 #1   9 11 1   0.0075   0.0075
40 #2  13 10 1   0.0135   0.0135
41 #3  18  8 1   0.0195   0.0195
42 #4  23  7 1   0.0233   0.0233
43 #5  31  5 1   0.0270   0.0270
44 #6  34  4 1   0.0265   0.0265
45 #7  48  2 1   0.0236   0.0236
46 B.rozptyl.funkcie.prezivania(survobjA.B)
47 # kontrola -- B1 je rozptyl vypočítaný funkciou z knižnice survival
48 #   cas ni di rozptyl.B1 rozptyl.B2
49 #1   9 11 1   0.0069   0.0069
50 #2  13 10 1   0.0125   0.0125
51 #3  18  8 1   0.0180   0.0180
52 #4  23  7 1   0.0217   0.0217
53 #5  31  5 1   0.0253   0.0253
54 #6  34  4 1   0.0255   0.0255
55 #7  48  2 1   0.0243   0.0243

```

## 3.5 Empirické intervaly a pásy spoľahlivosti

### 3.5.1 Intervaly spoľahlivosti

Pri výpočte hraníc empirických intervalov spoľahlivosti (ISs) sa používajú nasledovné princípy

1. **Waldov princíp, skóre princíp a vierohodnostný princíp** (všetky tri vychádzajúce z funkcie vierohodnosti),
2. **princíp transformácie funkcie vierohodnosti v  $S(t)$  pomocou nejakej funkcie  $g(S(t))$**  aplikovaný na Waldov princíp (hranice IS vypočítané pomocou  $g(S(t))$ ) sa spätne transformujú do škály  $S(t)$ ), kde podľa  $g(S(t))$  rozoznávame nasledovné škály
  - $g(S(t)) = S(t)$  – škála funkcie prežívania



- $g(S(t)) = \ln S(t)$  – škála logaritmu funkcie prežívania (log škála funkcie prežívania; škála kumulatívneho rizika) – spätná transformácia pre funkciu prežívania  $\exp(\ln S(t))$ ,
- $g(S(t)) = \ln(-\ln S(t))$  – log-log škála funkcie prežívania (log škála kumulatívneho rizika; škála logaritmu kumulatívneho rizika) –  
spätná transformácia pre funkciu prežívania  $\exp(\exp(-\ln S(t)))$ ,  
spätná transformácia pre kumulatívne riziko  $\exp(-\ln S(t))$ ,
- $g(S(t)) = \arcsin\left((S(t))^{1/2}\right)$  – arcus sínus odmocninová škála funkcie prežívania –  
spätná transformácia pre funkciu prežívania  $\left(\sin\left(\arcsin\left((S(t))^{1/2}\right)\right)\right)^2$
- $g(S(t)) = \arcsin\left(\exp\left(-\frac{\ln S(t)}{2}\right)\right)$  – arcus sínus odmocninová škála kumulatívneho rizika –  
spätná transformácia pre kumulatívne riziko  $-2 \ln \sin\left(\left(\arcsin\left(\exp\left(-\frac{\ln S(t)}{2}\right)\right)\right)\right)$

Pre extrémne malé alebo extrémne veľké hodnoty  $t$  sa môže stať, že empirický interval spoľahlivosti pre  $S(t)$  má hranice mimo intervalu  $\langle 0, 1 \rangle$ . Ak budeme predpokladať asymptoticky normálne rozdelenie nejakej transformácie  $S(t)$ , ktorej hodnoty nie sú ohraničené, môžeme sa tomuto problému vyhnúť.

3. princíp úpravy hraníc IS pomocou efektívneho rozsahu súboru v čase  $t$  ( $n^*(t)$  alebo  $n^{**}(t)$ ) z dôvodu výskytu cenzúr v dátach (resp. rozdielu medzi rozptylom  $S(t)$  v čase  $t$  vypočítaným pre cenzúrované dáta a rozptylom za predpokladu, že cenzúry v dátach nie sú)
4. princíp korekcie hraníc IS z dôvodu zlých štatistických vlastností (konzervatívny alebo liberálny IS, t.j. predpokladáme, že nominálny koeficient spoľahlivosti  $1 - \alpha$  je iný ako teoretický)

### 3.5.2 Intervaly spoľahlivosti pre kumulatívne riziko

Pri tvorbe IS pre  $\Lambda(t)$  máme tri možnosti.

1.  $\widehat{\Lambda}_{\text{KM}}(t)$  a  $\widehat{\sigma}_{\text{G}}^2(t)$ ,
2.  $\widehat{\Lambda}_{\text{NA}}(t)$  a  $\widehat{\sigma}_{\text{T}}^2(t)$  (namiesto  $\widehat{\sigma}_{\text{T}}^2(t)$  môžeme použiť aj  $\widehat{\sigma}_{\text{B}}^2(t)$  alebo  $\widehat{\sigma}_{\text{K}}^2(t)$ ),
3.  $\widehat{\Lambda}_{\text{FHmodNA}}(t)$  a  $\widehat{\sigma}_{\text{FH}}^2(t)$ ,

### 3.5.3 Intervaly spoľahlivosti pre funkciu prežívania

Pri tvorbe IS pre  $S(t)$  máme päť možností podľa použitých odhadov funkcií prežívania a ich rozptylov

1.  $\widehat{S}_{\text{KM}}(t)$  a  $\widehat{Var}_{\text{G}}\left[\widehat{S}_{\text{KM}}(t)\right]$ ,
2.  $\widehat{S}_{\text{B}}(t)$  a  $\widehat{Var}_{\text{NA}}\left[\widehat{S}_{\text{B}}(t)\right]$ ,
3.  $\widehat{S}_{\text{B}}(t)$  a  $\widehat{Var}_{\text{AJ}}\left[\widehat{S}_{\text{B}}(t)\right]$ ,
4.  $\widehat{S}_{\text{FHmodB}}(t)$  a  $\widehat{Var}_{\text{NA}}\left[\widehat{S}_{\text{FHmodB}}(t)\right]$  a
5.  $\widehat{S}_{\text{FHmodB}}(t)$  a  $\widehat{Var}_{\text{G}}\left[\widehat{S}_{\text{FHmodB}}(t)\right]$ .

**Príklad 32 (intervaly spoľahlivosti pre kumulatívne riziko)** Naprogramujte v  $\mathbb{R}$   $100 \times (1 - \alpha)\%$  intervaly spoľahlivosti pre kumulatívne riziko

(a) v škále kumulatívneho rizika,

(b) v log škále kumulatívneho rizika.

Pri výpočte použite  $\hat{\Lambda}_{KM}(t)$  a  $\hat{\Lambda}_{FHmodNA}(t)$  a ich rozptyly.

cvič.

**Príklad 33 (intervaly spoľahlivosti pre kumulatívne riziko)** Vypočítajte v  $\mathbb{R}$  95% intervaly spoľahlivosti pre kumulatívne riziko v oboch škálach pre AML dáta (skupina A). Zobrazte do okna  $2 \times 1$  odhady kumulatívneho rizika a IS pre kumulatívne riziko v tvare písmena „I“ v každom čase zlyhania v oboch škálach. Použite (a)  $\hat{\Lambda}_{KM}(t)$  a (b)  $\hat{\Lambda}_{FHmodNA}(t)$  a ich rozptyly.

cvič.

**Príklad 34 (intervaly spoľahlivosti pre funkciu prežívania)** Naprogramujte v  $\mathbb{R}$   $100 \times (1 - \alpha)\%$  intervaly spoľahlivosti pre funkciu prežívania

(a) v škále funkcie prežívania,

(b) v log škále funkcie prežívania a

(c) v log-log škále funkcie prežívania.

Pri výpočte použite  $\hat{S}_{KM}(t)$  a  $\hat{S}_{FHmodB}(t)$  a ich rozptyly.

cvič.

**Príklad 35 (intervaly spoľahlivosti pre funkciu prežívania)** Vypočítajte v  $\mathbb{R}$  95% intervaly spoľahlivosti pre funkciu prežívania vo všetkých troch škálach pre AML dáta (skupina A). Zobrazte do okna  $3 \times 1$  odhady funkcie prežívania a IS pre funkciu prežívania v tvare písmena „I“ v každom čase zlyhania vo všetkých troch škálach. Použite (a)  $\hat{S}_{KM}(t)$  a (b)  $\hat{S}_{FHmodB}(t)$  a ich rozptyly.

cvič.

### 3.6 Odhady strednej hodnoty a mediánu prežívania

Pre odhad strednej hodnoty času zlyhania  $E[T]$  platí

$$\hat{\mu} = \int_0^{t_{\max}} \hat{S}(t) dt,$$

kde  $\hat{S}(t)$  je KM odhad funkcie prežívania a  $t_{\max}$  je maximum pozorovaných časov. Tomuto odhadu hovoríme aj *urezaný*.

V diskretnom prípade odhad strednej hodnoty vypočítame nasledovne

$$\hat{\mu} = \sum_{i=0}^{\max(t_{\max}, c_{\max})-1} (t_{i+1} - t_i) \hat{S}(t_i),$$

kde  $t_0 = 0$  a  $I \leq n$  je počet rôznych zlyhaní a  $t_I = t_{\max}$ . Ak  $t_{\max} < c_{\max}$ , kde  $c_{\max}$  je maximálnym časom do cenzúry, potom odhad strednej hodnoty počítame na intervale  $\langle 0, c_{\max} \rangle$ . Pre KM odhad funkcie prežívania je odhad strednej hodnoty prežívania nedefinovateľný, ak je posledné pozorovanie cenzúra. Je možné definíciu modifikovať tak, že odhad je nulový za posledným pozorovaním. Takýto odhad strednej hodnoty je síce vychýlený smerom k nule, ale doteraz nebola objavená lepšia alternatíva. Podľa použitých odhadov funkcií prežívania rozlišujeme nasledovné odhady strednej hodnoty  $\hat{\mu}_{KM} = \hat{\mu}_{KMmod}$ ,  $\hat{\mu}_B$  a  $\hat{\mu}_{FHmodB}$ .

Potom pre **odhad rozptylu strednej hodnoty času prežívania** platí

$$\widehat{Var}[\hat{\mu}] = \int_0^{t_{\max}} \left[ \int_t^{t_{\max}} \hat{S}(u) du \right]^2 \hat{\sigma}^2(t) dt$$

a v diskretnom prípade

$$\widehat{Var}[\hat{\mu}] = \sum_{i:t_i \leq \max(t_{\max}, c_{\max})-1} \left[ \sum_{j:t_j \leq t_i \leq \max(t_{\max}, c_{\max})-1} (t_{j+1} - t_j) \hat{S}(t_j) \right]^2 \hat{\sigma}^2(t_i).$$

Za  $\widehat{S}(t)$  dosadíme  $\widehat{S}_{KM}(t)$ ,  $\widehat{S}_B(t)$  alebo  $\widehat{S}_{FHmodB}(t)$ . Podobne ako pri rozptyle funkcie prežívania dostaneme päť nasledovných rozptylov

$$Var_G[\widehat{\mu}_{KM}] = Var_G[\widehat{\mu}_{KMmod}], Var_{NA}[\widehat{\mu}_B], Var_{AJ}[\widehat{\mu}_{KM}], Var_{NA}[\widehat{\mu}_{FHmodB}] \text{ a } Var_G[\widehat{\mu}_{FHmodB}].$$

**Príklad 36 (priemerný čas prežívania, AML)** Vypočítajte priemerný čas prežívania  $\widehat{\mu}$ , jeho rozptyl  $Var[\widehat{\mu}]$  a  $100 \times (1 - \alpha)\%$  intervalov spoľahlivosti pre  $\mu$  pre AML dáta (skupina A). pred

$$\widehat{\mu} = 52.6 \text{ týždňa}, Var[\widehat{\mu}] = 19.8^2, 95\%IS = (13.792, 91.408) \text{ týždňa}$$

**Príklad 37 (programovanie, priemerný čas prežívania)** Naprogramujte v  $\mathbb{R}$  algoritmus na výpočet obsahu pod krivkou prežívania. cvič.

**Príklad 38 (priemerný čas prežívania, AML)** Vypočítajte priemerný čas prežívania pre AML dáta (skupina A). Pri výpočte použite  $\widehat{S}_{KM}(t)$  a  $\widehat{S}_B(t)$ . Okomentujte výsledky. Nakreslite krivku odhadu funkcie prežívania a pomocou funkcie `polygon()` vyfarbite obsah pod ňou.

```
56 # vysledok 60.30259
57 # kontrola
58 print(survobjA.B, rmean="individual") # 60.3
```

Keďže  $\widehat{S}_{KM}(t) \leq \widehat{S}_B(t)$ , potom  $\widehat{\mu}_{KM} \leq \widehat{\mu}_{KM}$ .

**Príklad 39 (programovanie, rozptyl priemerného času prežívania)** Naprogramujte v  $\mathbb{R}$  algoritmus na výpočet rozptylu priemerného času prežívania. cvič.

**Príklad 40 (rozptyl priemerného času prežívania, AML)** Vypočítajte rozptyl a smerodajnú odchýlku priemerného času prežívania pre AML dáta (skupina A). Pri výpočte použite  $\widehat{S}_{KM}(t)$  a  $\widehat{S}_B(t)$ . Okomentujte výsledky.

```
59 rozptyl.mu(survobjA.KM, error="G") # 393.1735
60 rozptyl.mu(survobjA.B, error="T") # 414.8907
61 rozptyl.mu(survobjA.B, error="G") # 655.1461 (ako prednastavene v R)
62 sqrt(rozptyl.mu(survobjA.KM, error="G")) # 19.8286
63 sqrt(rozptyl.mu(survobjA.B, error="T")) # 20.36887
64 sqrt(rozptyl.mu(survobjA.B, error="G")) # 25.59582
```

**Medián času prežívania** je 50-ty percentil  $t_{0.5}$ . Medián funkcie prežívania je potom  $S(t_{0.5}) = 0.5$ . Výberový medián je definovaný ako prvý čas, v ktorom  $\widehat{S}(\widehat{t}_{0.5}) = 0.5$ . Teda  $(\tilde{\mu} = \widehat{t}_{0.5} \leq \widehat{S}^{-1}(0.5))$ . Niekedy je potrebné použiť lineárnu interpoláciu pre  $\widehat{t}_{0.5}$  v podobe

$$\tilde{\mu}_{int} = t_i + (t_{i+1} - t_i) \frac{\widehat{S}(t_i) - 0.5}{\widehat{S}(t_i) - \widehat{S}(t_{i+1})} = t_i + \frac{\widehat{S}(t_i) - 0.5}{\tilde{f}(t_i)}, t_i \leq t_{0.5} \leq t_{i+1},$$

kde  $\tilde{f}(t_i) = \frac{\widehat{S}(t_i) - \widehat{S}(t_{i+1})}{t_{i+1} - t_i}$ , čo je pravdepodobnosť zlyhania v intervale  $(t_{i+1} - t_i)$  škálovaná jeho dĺžkou.

Horná a dolná hranica intervalu spoľahlivosti (IS) pre medián je definovaná na základe IS pre  $S(t)$  v danom čase, t.j. horná hranica IS pre medián je prvý čas, v ktorom je horná hranica IS pre  $S(t)$  väčšia alebo rovná 0.5 (rovnaký vzťah platí pre dolnú hranicu IS pre medián). To korešponduje s narysovaním horizontálnej úsečky na grafe krivky prežívania, tj. prenutím tejto úsečky s krivkou prežívania, dolnou a hornou hranicou IS pre  $S(t)$  v danom čase.

Nech  $t_p$  je  $p$ -ty kvantil rozdelenia  $T$  ( $100 \times p$ -ty percentil), potom

$$\widehat{S}(\widehat{t}_p) = \Pr(T \geq \widehat{t}_p) = 1 - p, \widehat{t}_p \leq \widehat{S}^{-1}(1 - p).$$

Keďže KM krivka prežívania je schodovitá funkcia, inverzia  $\widehat{S}^{-1}(t_p)$  nie je jednoznačne definovaná. Odhad kvantilu bude potom

$$\widehat{t}_p = \min\{t_i : \widehat{S}(t_i) \leq 1 - p\}.$$

Aplikovaním *delta metódy* na  $\widehat{Var}_G[\widehat{S}(t_p)]$  dostaneme

$$\widehat{Var}[\widehat{t}_p] = \frac{\widehat{Var}_G[\widehat{S}(t_p)]}{[\widehat{f}(t_p)]^2}, \quad \widehat{f}(t_p) = \frac{\widehat{S}(\widehat{u}_p) - \widehat{S}(\widehat{l}_p)}{\widehat{l}_p - \widehat{u}_p},$$

kde  $\widehat{u}_p = \max\{t_i : \widehat{S}(t_i) \geq 1 - p + \epsilon\}$  a  $\widehat{l}_p = \min\{t_i : \widehat{S}(t_i) \leq 1 - p - \epsilon\}$  pre  $i = 1, 2, \dots, I \leq n$ , kde  $u$  korešponduje s hornou hranicou,  $l$  s dolnou,  $I$  je počet rozdielnych časov zlyhania,  $\epsilon$  je veľmi malé číslo. Najčastejšie  $\epsilon = 0.05$  akceptovateľné, ale musí byť veľké, ak  $|\widehat{l}_p - \widehat{u}_p| \approx 0$ . Za  $\widehat{S}(t)$  dosadíme  $\widehat{S}_{\text{KM}}(t)$ ,  $\widehat{S}_{\text{B}}(t)$  alebo  $\widehat{S}_{\text{FHmodB}}(t)$ . Potom dostaneme  $\widehat{t}_p^{(\text{KM})} = \widehat{t}_p^{(\text{KMmod})}$ ,  $\widehat{t}_p^{(\text{B})}$  a  $\widehat{t}_p^{(\text{FHmodB})}$  a  $\widehat{Var}[\widehat{t}_p^{(\text{KM})}] = \widehat{Var}[\widehat{t}_p^{(\text{KMmod})}]$ ,  $\widehat{Var}[\widehat{t}_p^{(\text{B})}]$  a  $\widehat{Var}[\widehat{t}_p^{(\text{FHmodB})}]$ .

**Príklad 41 (medián, AML)** Vypočítajte medián času prežívania  $\widetilde{\mu}$ , jeho rozptyl  $\widehat{Var}[\widetilde{\mu}]$  a  $100 \times (1 - \alpha)\%$  intervalov spoľahlivosti pre  $t_{0.5}$  pre AML dáta (skupina A). pred

$$\widetilde{\mu} = \widehat{t}_{0.5}, \quad \widehat{Var}[\widetilde{\mu}] = \frac{\widehat{Var}_G[\widehat{S}(t_{0.5})]}{[\widehat{f}(t_{0.5})]^2}$$

$$\widehat{t}_{0.5} = 31 \text{ týždňov}, \quad \widehat{u}_{0.5} = \max\{t_i : \widehat{S}(t_i) \geq 0.55\} = 23 \text{ a } \widehat{l}_{0.5} = \min\{t_i : \widehat{S}(t_i) \leq 0.45\} = 34$$

$$\widehat{f}(31) = \frac{\widehat{S}(\widehat{u}_{0.5}) - \widehat{S}(\widehat{l}_{0.5})}{\widehat{l}_{0.5} - \widehat{u}_{0.5}} = \frac{\widehat{S}(23) - \widehat{S}(34)}{34 - 23} = \frac{0.6136364 - 0.3681818}{11} = 0.022$$

$$\widehat{Var}[31] = \left(\frac{0.16419327}{0.02231405}\right)^2 = 54.144, \quad \sqrt{\widehat{Var}[31]} = 7.358, \quad 95\% \text{ IS} = (16.578, 45.422) \text{ týždňa}$$

**Príklad 42 (programovanie, kvantily)** Naprogramujte v  $\mathbb{R}$  funkciu na výpočet odhadu kvantilov času prežívania  $\widehat{t}_p$ ,  $\widehat{Var}[\widehat{t}_p]$  a  $100 \times (1 - \alpha)\%$  intervalov spoľahlivosti pre  $t_p$ . DÚ

**Príklad 43 (medián, AML)** Vypočítajte medián  $t_{0.5}$  a jeho  $100 \times (1 - \alpha)\%$  intervalov spoľahlivosti pre AML dáta (skupina A).

```
65 | kvantily.km(survobjA.KM,p = 0.5, eps = 0.05, conf.coef = 0.975)
66 | #   kvp sd(kvp)   DHIS   HHIS
67 | # [1,]  31  7.3583 16.578 45.422
68 | kvantily.km(survobjA.B,p = 0.5, eps = 0.05, conf.coef = 0.975)
69 | #   kvp sd(kvp)   DHIS   HHIS
70 | # [1,]  34  8.5516 17.2392 50.7608
```

### 3.7 Odhad strednej hodnoty a mediánu zostatkového života a ich rozptyl

Odhad strednej hodnoty zostatkového života v čase  $t$  nazývame **priemerný zostatkový život** v čase  $t$  a vypočítame ho v spojitom prípade ako

$$\widehat{\text{mrl}}(t) = \frac{\int_t^{t_{\max}} \widehat{S}(u) du}{\widehat{S}(t)}$$

a v diskretnom prípade ako

$$\widehat{\text{mrl}}(t) = \frac{1}{\widehat{S}(t)} \left( (t_{i+1} - t) \widehat{S}(t_i) + \sum_{j: t_{i+1} \leq t_j \leq \max(t_{\max}, c_{\max}) - 1} (t_{j+1} - t_j) \widehat{S}(t_j) \right),$$

kde  $t_i \leq t < t_{i+1}$ .

Za  $\widehat{S}(\cdot)$  dosadíme  $\widehat{S}_{\text{KM}}(\cdot)$ ,  $\widehat{S}_{\text{B}}(\cdot)$  alebo  $\widehat{S}_{\text{FHmodB}}(\cdot)$ . Rozptyl priemerného zostatkového života odhadneme nasledovne

$$\widehat{\text{Var}}[\widehat{\text{mrl}}(t)] = \frac{1}{\widehat{S}^2(t)} \left( \int_t^{t_{\max}} \left[ \int_u^{t_{\max}} \widehat{S}(x) dx \right]^2 \widehat{\sigma}_{\text{G}}^2(u) du + \left[ \int_t^{t_{\max}} \widehat{S}(u) du \right]^2 \int_0^t \widehat{\sigma}_{\text{G}}^2(u) du \right),$$

pre spojité prípad, kde  $u \in \langle t, t_{\max} \rangle$ , a ako

$$\begin{aligned} \widehat{\text{Var}}[\widehat{\text{mrl}}(t)] = & \frac{1}{\widehat{S}^2(t)} \left( \sum_{i:t_i \leq t_i \leq \max(t_{\max}, c_{\max})-1} \left[ \sum_{j:t_i \leq t_j \leq \max(t_{\max}, c_{\max})-1} (t_{j+1} - t_j) \widehat{S}(t_j) \right]^2 \widehat{\sigma}_{\text{G}}^2(t_i) \right. \\ & \left. + \left[ (t_{i+1} - t) \widehat{S}(t_i) + \sum_{j:t_{i+1} \leq t_j \leq \max(t_{\max}, c_{\max})-1} (t_{j+1} - t_j) \widehat{S}(t_j) \right]^2 \sum_{i:t_i \leq t} \widehat{\sigma}_{\text{G}}^2(t_i) \right) \end{aligned}$$

pre diskretný prípad, kde  $t_i \leq t < t_{i+1}$ . Za  $\widehat{S}(t)$  dosadíme  $\widehat{S}_{\text{KM}}(t)$ ,  $\widehat{S}_{\text{B}}(t)$  alebo  $\widehat{S}_{\text{FHmodB}}(t)$ . Potom dostaneme  $\widehat{\text{mrl}}_{\text{KM}}(t)$ ,  $\widehat{\text{mrl}}_{\text{B}}(t)$  a  $\widehat{\text{mrl}}_{\text{FHmodB}}(t)$  päť nasledovných rozptylov

$$\begin{aligned} \widehat{\text{Var}}_{\text{G}}[\widehat{\text{mrl}}_{\text{KM}}] &= \widehat{\text{Var}}_{\text{G}}[\widehat{\text{mrl}}_{\text{KMmod}}], \quad \widehat{\text{Var}}_{\text{NA}}[\widehat{\text{mrl}}_{\text{B}}], \quad \widehat{\text{Var}}_{\text{AJ}}[\widehat{\text{mrl}}_{\text{KM}}], \\ &\widehat{\text{Var}}_{\text{NA}}[\widehat{\text{mrl}}_{\text{FHmodB}}] \text{ a } \widehat{\text{Var}}_{\text{G}}[\widehat{\text{mrl}}_{\text{FHmodB}}]. \end{aligned}$$

**Medián zostatkového života** (median residual life, median remaining lifetime) definujeme ako

$$\widetilde{\text{mrl}}(t) = (t_i - t) + (t_{i+1} - t_i) \frac{\widehat{S}(t_i) - 0.5\widehat{S}(t)}{\widehat{S}(t_i) - \widehat{S}(t_{i+1})} = (t_i - t) + \frac{\widehat{S}(t_i) - 0.5\widehat{S}(t)}{\widetilde{f}(t_i)}$$

kde  $t_i < t < t_{i+1}$ ,  $\widehat{S}(t_i) \geq 0.5\widehat{S}(t)$  a  $\widehat{S}(t_{i+1}) < 0.5\widehat{S}(t)$ . To znamená, že ak sme mali pravdepodobnosť prežitia  $\widehat{S}(t)$  v čase  $t$ , potom budeme mať nažive polovicu z tohoto množstva, t.j.  $0.5\widehat{S}(t)$ , v čase  $t + \widetilde{\text{mrl}}(t)$ . Rozptyl odhadneme nasledovne

$$\widehat{\text{Var}}[\widetilde{\text{mrl}}(t)] \approx \frac{1}{4n_i} \left( \frac{\widehat{S}(t_i)}{\widetilde{f}(t_i)} \right)^2.$$

Za  $\widehat{S}(t)$  dosadíme  $\widehat{S}_{\text{KM}}(t)$ ,  $\widehat{S}_{\text{B}}(t)$  alebo  $\widehat{S}_{\text{FHmodB}}(t)$ . Potom dostaneme  $\widetilde{\text{mrl}}_{\text{KM}}(t)$ ,  $\widetilde{\text{mrl}}_{\text{B}}(t)$  a  $\widetilde{\text{mrl}}_{\text{FHmodB}}(t)$  päť nasledovných rozptylov

$$\begin{aligned} \widehat{\text{Var}}_{\text{G}}[\widetilde{\text{mrl}}_{\text{KM}}] &= \widehat{\text{Var}}_{\text{G}}[\widetilde{\text{mrl}}_{\text{KMmod}}], \quad \widehat{\text{Var}}_{\text{NA}}[\widetilde{\text{mrl}}_{\text{B}}], \quad \widehat{\text{Var}}_{\text{AJ}}[\widetilde{\text{mrl}}_{\text{KM}}], \\ &\widehat{\text{Var}}_{\text{NA}}[\widetilde{\text{mrl}}_{\text{FHmodB}}] \text{ a } \widehat{\text{Var}}_{\text{G}}[\widetilde{\text{mrl}}_{\text{FHmodB}}]. \end{aligned}$$


**Príklad 44 (programovanie, priemerný zostatkový život)** Naprogramujte v  $\mathbb{R}$  funkciu na výpočet odhadu strednej hodnoty zostatkového života v čase  $t$ . cvič.

**Príklad 45 (priemerný zostatkový život, AML)** Vypočítajte priemerný zostatkový život v čase  $t = 0$  a  $t = 30$  pre AML dáta (skupina A). Nakreslite krivku odhadu funkcie prežívania a pomocou funkcie `polygon()` vyfarbite obsah pod ňou za bodom  $t$ . cvič.


```
71 | mrl(survobjA.KM, tcko = 0)$mrl # 52.64545 52.64545 52.64545
72 | mrl(survobjA.KM, tcko = 30)$mrl # 45.70000 52.73077 52.57692
```

**Príklad 46 (programovanie, rozptyl priemerného zostatkového života)** Naprogramujte v  funkciu na výpočet odhadu rozptylu priemerného zostatkového života v čase  $t$ . DÚ

**Príklad 47 (rozptyl priemerného zostatkového života, AML)** Vypočítajte odhadu rozptylu priemerného zostatkového života v čase  $t = 0$  a  $t = 30$  pre AML dáta (skupina A). DÚ

**Príklad 48 (programovanie, medián zostatkového života)** Naprogramujte v  funkciu na výpočet odhadu mediánu zostatkového života v čase  $t$ . DÚ

**Príklad 49 (medián zostatkového života, AML)** Vypočítajte medián zostatkového života v čase  $t = 0$  a  $t = 30$  pre AML dáta (skupina A). DÚ

**Príklad 50 (programovanie, rozptyl mediánu zostatkového života)** Naprogramujte v  funkciu na výpočet odhadu rozptylu mediánu zostatkového života v čase  $t$ . DÚ

**Príklad 51 (rozptyl mediánu zostatkového života, AML)** Vypočítajte odhadu rozptylu mediánu zostatkového života v čase  $t = 0$  a  $t = 30$  pre AML dáta (skupina A). DÚ

## 4 Testy na porovnanie kriviek prežívania

**Príklad 52 (nádor pľúc)** *Nech  $t_{ij}, i = 1, \dots, n_j, j = 1, 2$  sú časy do zlyhania (úmrtia) od diagnostiky nádoru pľúc v mesiacoch, kde  $j = 1$  predstavuje I. typ terapie a  $j = 2$  zasa II. typ terapie (pozri tabuľku). Otestujte  $H_0 : S_1(t) = S_2(t)$  oproti alternatíve  $H_1 : S_1(t) \neq S_2(t)$  pomocou  $S_W$  a  $S_{MW}$  nasledovne (1)  $S_W = W_Y$  a  $S_W = W_X$ , (2)  $S_{MW} = U_Y$  a  $S_{MW} = U_X$ . Vždy presne naformulujte  $H_1$ . Okomentujte výsledky.*

cvič.

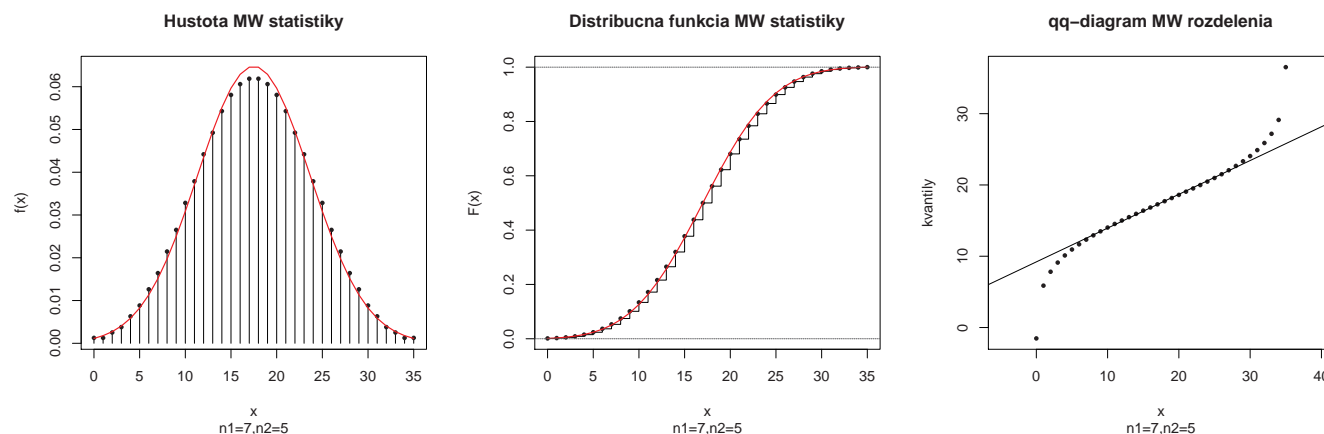
$t_{ij}$	52	240	19	53	15	43	340	133	111	231	378	49
skup	1	2	2	1	1	2	2	1	1	2	1	1

Určenie poradí

$t_{(i)}$	15	19	43	49	52	53	111	133	231	240	340	378
skup	1	2	2	1	1	1	1	1	2	2	2	1
$r_i^{(1)}$	1	-	-	4	5	6	7	8	-	-	-	12
$r_i^{(2)}$	-	2	3	-	-	-	-	-	9	10	11	-

**Príklad 53 (asymptotická normalita  $S_{MW}$ )** *Pre  $n_1 = 7$  a  $n_2 = 5$  porovnajzte v  $\mathbb{R}$  asymptotické rozdelenie  $S_{MW}$  s jej exaktným rozdelením. Na výpočet asymptotickej hustoty použite funkciu `dnorm()` a na výpočet asymptotickej distribučnej funkcie použite funkcie `dnorm()` a `cumsum()`. Na výpočet exaktnej hustoty použite funkciu `dwilcox()` a na výpočet exaktnej distribučnej funkcie použite funkciu `pwilcox()`. 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  $\min(S_{MW})$  po teoreticky možné  $\max(S_{MW})$  a na y-ovej osi teoretické kvantily  $y$  vypočítané pomocou funkcie `qnorm()`; qq-priamka bude prechádzať bodmi  $(\tilde{x}_{0.25}, \tilde{x}_{0.75})$  a  $(\tilde{y}_{0.25}, \tilde{y}_{0.75})$ ).*

cvič.



Obr. 7: Rozdelenie Mann-Whitneyho štatistiky  $S_{MW}$  ( $n_1 = 7, n_2 = 5$ )

**Príklad 54 (asymptotická normalita  $S_{MW}$ )** *Porovnajzte v  $\mathbb{R}$  asymptotické rozdelenie  $S_{MW}$  s jej exaktným rozdelením pre (1)  $n_1 = 5$  a  $n_2 = 50$ , (2)  $n_1 = 50$  a  $n_2 = 50$ , (3)  $n_1 = 50$  a  $n_2 = 100$  a (4)  $n_1 = 100$  a  $n_2 = 100$ .*

cvič.

**Príklad 55 (nádor pľúc pokrač.)** *Nech  $t_{ij}, i = 1, \dots, n_j, j = 1, 2$  sú časy do zlyhania (úmrtia) od diagnostiky nádoru pľúc v mesiacoch, kde  $j = 1$  predstavuje I. typ terapie a  $j = 2$  zasa II. typ terapie (pozri tabuľku). Otestujte v  $\mathbb{R}$   $H_0 : Var[S_1(t)] = Var[S_2(t)]$  oproti alternatíve  $H_1 : Var[S_1(t)] \neq Var[S_2(t)]$  pomocou  $S_{ST}$  a  $S_{ST}^{alt}$ . Okomentujte výsledky.*

cvič.

**Riešenie (aj v  $\mathbb{R}$ ):**

Siegel-Tukey test

$t_{ij}$	52	240	19	53	15	43	340	133	111	231	378	49
skup	1	2	2	1	1	2	2	1	1	2	1	1
$R_i^{(1)}$	9	-	-	11	1	-	-	10	12	-	2	7
$R_i^{(2)}$	-	6	3	-	-	5	4	-	-	8	-	-

**Príklad 56** Naprogramujte v  $\mathbb{R}$  test  $H_0 : Var[S_1(t)] = Var[S_2(t)]$  oproti alternatíve  $H_1 : Var[S_1(t)] \neq Var[S_2(t)]$  pomocou  $S_{ST}$  a  $S_{ST}^{alt}$  použitím algoritmu 1. Okomentujte výsledky. cvič.

**Príklad 57 (dvojčatá; Wilcoxonov znamienkový test)** Pri skúmaní agresívnosti dvojčat psychológovia zaznamenali nasledovné skóre (pozri tabuľku). Otestujte  $H_0$  na  $\alpha = 0.05$ , že prvorodené dvojča je agresívnejšie. Použite Wilcoxonov znamienkový test, t.j. štatistiky (1)  $S_W$ , (2)  $S_W^*$  a (3)  $S_W$  s korekciou rozptylu na zhody. Porovnajte výsledky. pred.

$i$	1	2	3	4	5	6	7	8	9	10	11
$X_i$ dvojča 1	85	70	76	68	90	72	77	90	70	71	87
$Y_i$ dvojča 2	87	76	75	64	95	72	65	89	65	80	80
$Z_i = X_i - Y_i$	-2	-6	1	4	-5	0	12	1	5	-9	7
$ Z_i $	2	6	1	4	5	-	12	1	5	9	7
$R_i$	3	7	1.5	4	5.5	-	10	1.5	5.5	9	8

**Príklad 58 (asymptotická normalita  $S_{KW}$ )** Pre  $n_1 = n_2 = n_3 = 5$  porovnajte v  $\mathbb{R}$  asymptotické rozdelenie  $S_{KW}$  s jej exaktným rozdelením. 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  $(\tilde{x}_{0.25}, \tilde{x}_{0.75})$  a  $(\tilde{y}_{0.25}, \tilde{y}_{0.75})$  alebo alternatívne bude qq-priamku reprezentovať os prvého a tretieho kvadrantu). DÚ

**Príklad 59 (asymptotická normalita  $S_{KW}$ )** Porovnajte v  $\mathbb{R}$  asymptotické rozdelenie  $S_{KW}$  s jej exaktným rozdelením pre (1)  $n_1 = n_2 = 5$  a  $n_3 = 50$ , (2)  $n_1 = n_2 = n_3 = 50$ , (3)  $n_1 = n_2 = 50$  a  $n_3 = 100$  a (4)  $n_1 = n_2 = n_3 = 100$ . DÚ

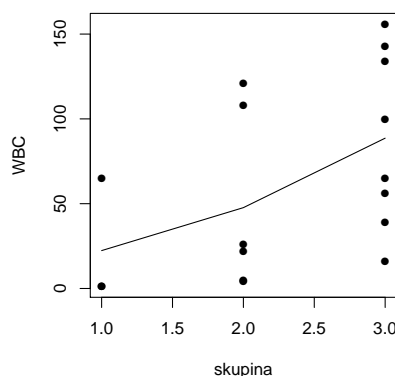
**Príklad 60 (WBC pokrač.)** Majme pacientov s akútnou myeloidnou leukémiou a v rámci nich skupinu AG-pozitívnych (výskyt určitých špecifických indikátorov choroby v kostnej dreni). Pre chorobu je charakteristické, že s počtom bielych krviniek (white blood cells counts, WBC) vzrastá závažnosť choroby. Nech  $t_i, i = 1, 2, \dots, 17$  sú časy do zlyhania v týždňoch (pozri tab.). (a) Otestujte  $H_0 : S_1(t) = S_2(t) = S_3(t)$ , alternatíva a)  $H_1 : S_i(t) \neq S_j(t), i < j; i, j = 1, 2, 3$ , b)  $S_1(t) \stackrel{st}{>} S_2(t) \stackrel{st}{>} S_3(t)$ , c)  $S_1(t) \stackrel{st}{<} S_2(t) \stackrel{st}{<} S_3(t)$ . Použite (1)  $S_{KW}$ , (2)  $S_J$ , (3)  $S_C$  a (4)  $S_L$ . (b) Vypočítajte Kendalov korelačný koeficient  $\tau$  medzi časom do zlyhania a príslušnosťou do skupiny 1, 2, a 3 pomocou  $S_J$ . (c) Porovnajte  $S_C$  a  $S_N$ . (d) Otestujte odklon od trendu. (e) Graficky znázornite príslušnosť do skupiny voči času do zlyhania spolu s krivkou spájajúcou priemerný čas do zlyhania v každej skupine. Okomentujte adekvátnosť lineárneho trendu. pred.

**Riešenie:**



Tabuľka: WBC	$t_i$	Rozdeľme AG-pozitívnych pacientov do troch skupín nasledovne																																																																																	
2300	65	<ul style="list-style-type: none"> <li>• Skupina 1: <math>WBC \geq 100000</math>, <math>n_1 = 3</math>, (1, 1, 65)</li> <li>• Skupina 2: <math>WBC \in \langle 10000, 100000 \rangle</math>, <math>n_2 = 6</math>, (108, 121, 4, 26, 22, 5),</li> <li>• Skupina 3: <math>WBC &lt; 10000</math>, <math>n_3 = 8</math>, (65, 156, 100, 134, 16, 39, 143, 56)</li> </ul>																																																																																	
750	156																																																																																		
4300	100																																																																																		
2600	134																																																																																		
6000	16;																																																																																		
10500	108																																																																																		
10000	121																																																																																		
17000	4																																																																																		
5400	39																																																																																		
7000	143																																																																																		
9400	56	<table border="1"> <thead> <tr> <th>Tabuľka:</th> <th colspan="8"></th> </tr> </thead> <tbody> <tr> <td>sk.1</td> <td>1</td> <td>1</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> </tr> <tr> <td>sk.2</td> <td>-</td> <td>-</td> <td>4</td> <td>5</td> <td>-</td> <td>22</td> <td>26</td> <td>-</td> </tr> <tr> <td>sk.3</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>16</td> <td>-</td> <td>-</td> <td>39</td> </tr> <tr> <td>poradie</td> <td>1.5</td> <td>1.5</td> <td>3</td> <td>4</td> <td>5</td> <td>6</td> <td>7</td> <td>8</td> </tr> <tr> <td>sk.1</td> <td>-</td> <td>65</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> <td>-</td> </tr> <tr> <td>sk.2</td> <td>-</td> <td>-</td> <td>-</td> <td>108</td> <td>121</td> <td>-</td> <td>-</td> <td>-</td> </tr> <tr> <td>sk.3</td> <td>56</td> <td>65</td> <td>100</td> <td>-</td> <td>-</td> <td>134</td> <td>143</td> <td>156</td> </tr> <tr> <td>poradie</td> <td>9</td> <td>10.5</td> <td>12</td> <td>13</td> <td>14</td> <td>15</td> <td>16</td> <td>17</td> </tr> </tbody> </table>	Tabuľka:									sk.1	1	1	-	-	-	-	-	-	sk.2	-	-	4	5	-	22	26	-	sk.3	-	-	-	-	16	-	-	39	poradie	1.5	1.5	3	4	5	6	7	8	sk.1	-	65	-	-	-	-	-	-	sk.2	-	-	-	108	121	-	-	-	sk.3	56	65	100	-	-	134	143	156	poradie	9	10.5	12	13	14	15	16	17
Tabuľka:																																																																																			
sk.1	1		1	-	-	-	-	-	-																																																																										
sk.2	-		-	4	5	-	22	26	-																																																																										
sk.3	-		-	-	-	16	-	-	39																																																																										
poradie	1.5		1.5	3	4	5	6	7	8																																																																										
sk.1	-		65	-	-	-	-	-	-																																																																										
sk.2	-		-	-	108	121	-	-	-																																																																										
sk.3	56		65	100	-	-	134	143	156																																																																										
poradie	9		10.5	12	13	14	15	16	17																																																																										
32000	26																																																																																		
35000	22																																																																																		
100000	1																																																																																		
100000	1																																																																																		
52000	5																																																																																		
100000	65																																																																																		

$\bar{R}_1 = 4.50, \bar{R}_2 = 7.8\bar{3}3, \bar{R}_3 = 11.5625$   
 $S_{KW} = \frac{12}{n(n+1)} \sum_{j=1}^k \frac{R_j^2}{n_j} - 3(n+1) = \frac{12}{17 \times 18} [3 \times 4.5^2 + 6 \times 7.8\bar{3}3^2 + 8 \times 11.5625^2] - 3 \times 18 = 4.762662,$   
 p-hodnota= 0.0924  
 $S_L = \sum_{j=1}^3 n_j (L_j - M_j) \bar{R}_j = 3 \times (0 - 14) \times 4.5 + 6 \times (3 - 8) \times 7.8\bar{3}3 + 8 \times (9 - 0) \times 11.5625 = 408.5$   
 $Var [S_L] = \frac{n(n+1)}{12} \sum_{j=1}^k n_j (L_j - M_j)^2 = \frac{17(17+1)}{12} [3 \times (0 - 14)^2 + 6 \times (3 - 8)^2 + 8 \times (9 - 0)^2] = 187.9973^2$   
 $Z_L = 408.5 / 187.9973 = 2.172903, \text{ p-hodnota} = 0.0298$   
 $S_{KW} - (Z_L)^2 = 4.762662 - 2.172903^2 = 0.0412, \text{ p-hodnota} = 0.8392$



Obr. 8: Lineárny trend pre WBC dáta

**Príklad 61 (nádor pľúc pokrač.)** *Nech  $t_{ij}, i = 1, \dots, n_j, j = 1, 2$  sú časy do zlyhania (úmrtia) od diagnostiky nádoru pľúc v mesiacoch, kde  $j = 1$  predstavuje I. typ terapie a  $j = 2$  zasa II. typ terapie (pozri tabuľku). Otestujte  $H_0 : S_1(t) = S_2(t)$ , alternatíva  $H_1 : S_1(t) \neq S_2(t)$ . Použite (1)  $S_{KW}$ , (2)  $S_J$ , (3)  $S_C$  a (4)  $S_L$ . Vždy presne naformulujte  $H_1$ .*

**Príklad 62 (pokrač. WBC)** *Majme pacientov s akútnou myeloidnou leukémiou a v rámci nich skupinu AG-pozitívnych (výskyt určitých špecifických indikátorov choroby v kostnej dreni). Pre chorobu je charakteristické, že s počtom bielych krviniek (white blood cells counts, WBC) vzrastná*

závažnosť choroby. Nech  $t_i, i = 1, 2, \dots, 17$  sú časy do zlyhania v týždňoch prislúchajúce zoradeným WBC (pozri tab.). Vypočítajte Kendallov korelačný koeficient  $\hat{\tau}$ . Otestujte nezávislosť medzi počtom bielych krviniek a časmi do zlyhania na hladine významnosti  $\alpha = 0.05$ .

DÚ

WBC	$t_i$	$c_{ij}$	$d_{ij}$
750	156	0	16
2300	65	5	9
2600	134	1	13
4300	100	3	10
5400	39	5	7
6000	16	7	4
7000	143	0	10
9400	56	3	6
10000	121	0	8
10500	108	0	7
17000	4	4	2
32000	26	1	4
35000	22	1	3
52000	5	1	2
100000	1	1	0
100000	1	1	0
100000	65	0	0

Riešenie:

$$c_{ij} = \#(\uparrow t_i, \uparrow WBC_i) \text{ pod } i, c = \sum c_{ij},$$

$$d_{ij} = \#(\downarrow t_i, \uparrow WBC_i) \text{ pod } i, d = \sum d_{ij},$$

$$c = 33, d = 101, \frac{n(n-1)}{2} = \frac{17 \times 16}{2},$$

$$\hat{\tau} = \frac{c-d}{\frac{n(n-1)}{2}} = -0.5,$$

$$\widehat{Var}[\hat{\tau}] = \frac{2(2 \times 17 + 5)}{9 \times 17(17-1)} = 0.032,$$

$$z_{\hat{\tau}} = -0.5 / \sqrt{0.032} = -2.801 \text{ a p-hodnota} = 0.005.$$

**Príklad 63 (WBC pokrač.)** Vypočítajte Spearmanov korelačný koeficient  $r_S$ . Otestujte nezávislosť medzi počtom bielych krviniek a časmi do zlyhania pomocou  $Z_S$  na hladine významnosti  $\alpha = 0.05$ .

**Príklad 64 (testy pre cenzurované dáta)** Majme dáta z klinickej štúdie zhrnuté v nasledovnej tabuľke (pozri tabuľku). (a) Vytvorte kontingenčné tabuľky v každom čase zlyhania  $t_i, i = 1, 2, \dots, 7$  použitím celkového počtu subjektov v riziku  $n_i$  v čase  $t_i$ , celkového počtu zlyhaní  $d_i$  v čase  $t_i$ , celkového počtu subjektov prvej skupiny v riziku  $n_{1i}$  v čase  $t_i$  a celkového počtu zlyhaní  $d_{1i}$  subjektov prvej skupiny v čase  $t_i$ . (b) Vypočítajte stredné hodnoty  $E_0[d_{1i}]$ , rozdiely empirických a očakávaných početností  $d_{1i} - E_0[d_{1i}]$ , ako aj rozptyly  $Var_0[d_{1i}]$ . (c) Otestujte  $H_0 : \lambda_1(t) = \lambda_2(t)$  oproti  $H_1 : \lambda_1(t) = \theta \lambda_2(t)$  pomocou testovacích štatistík  $Q_{GW}$ ,  $Q_{CM}$ ,  $Q_{TW}$  a  $Q_{PP}$ . (d) Nakreslite Kaplan-Meierove odhady funkcie prežívania pre obe skupiny do jedného obrázka. (e) Vypočítajte (1)  $\hat{\theta}_P$ ,  $Var[\hat{\theta}_{MH}]$  a 95%IS pre  $\theta_P$ , (2)  $\hat{\theta}_{MH}$  a 95%IS pre  $\theta_{MH}$  a (3)  $\hat{\theta}_{MH}^*$ .

cvič.

DÚ

Riešenie:

$t_i$	$n_i$	$d_i$	$n_{1i}$	$d_{1i}$	$E_0[d_{1i}]$	$d_{1i} - E_0[d_{1i}]$	$Var_0[d_{1i}]$
3	10	1	5	1	0.50	0.50	0.2500
5	9	1	4	1	0.44	0.56	0.2469
7	8	1	3	1	0.38	0.62	0.2344
12	6	1	1	0	0.17	-0.17	0.1389
18	5	1	1	1	0.20	0.80	0.1600
19	4	1	0	0	0.00	0	0
20	3	1	0	0	0.00	0	0
suma				4	1.69	2.31	1.0302

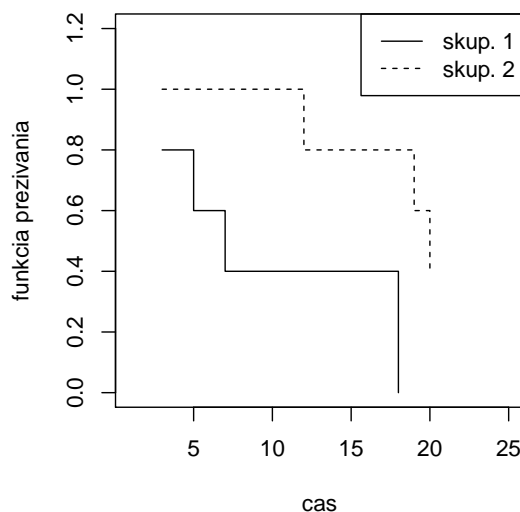
$Q = 2.31^2/1.0302 = 5.179674$ , p-hodnota=0.02285261.

$z_Q = \sqrt{2.31^2/1.0302} = 2.275890$ , p-hodnota= $2 \times 0.01142630 = 0.02285259$

KT v každom čase  $t_i$ :

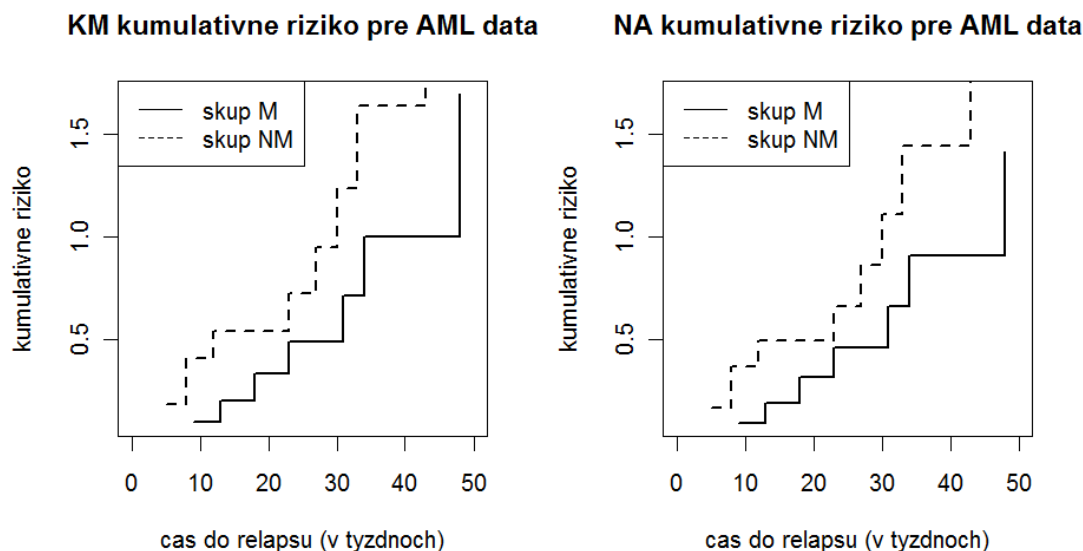
	$d_1$	$a_1$	sp	$d_2$	$a_2$	sp	$d_3$	$a_3$	sp	$d_4$	$a_4$	sp
skup. 1	1	4	5	1	3	4	1	2	3	0	1	1
skup. 2	0	5	5	0	5	5	0	5	5	1	4	5
sp	1	9	10	1	8	9	1	7	8	1	5	6
	$d_5$	$a_5$	sp	$d_6$	$a_6$	sp	$d_7$	$a_7$	sp			
skup. 1	1	0	1	0	0	0	0	0	0			
skup. 2	0	4	4	1	3	4	1	2	3			
sp	1	4	5	1	3	4	1	2	3			

testovacie kritérium	$Q$	p-hodnota
GW	4.695652	0.03023902
CM	5.197242	0.02262275
TW	4.970637	0.02578116
PP	4.732935	0.02959035

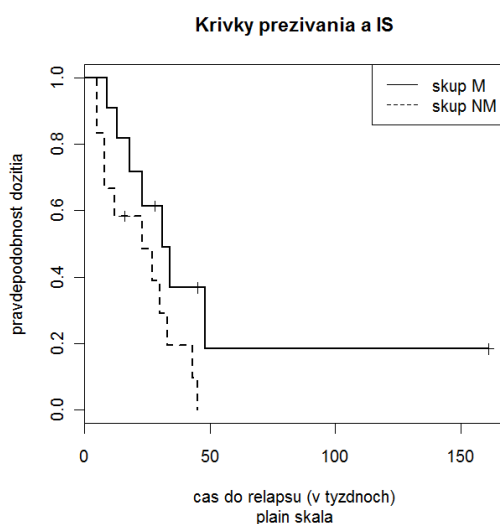


Obr. 9: Kaplan Meierove odhady funkcie prežívania

**Príklad 65 (AML, pokrač.)** Nakreslite (a) kumulatívne riziko  $\hat{\Lambda}_{KM,j}(t)$ ,  $j = 1, 2$  pre obe skupiny, (b) kumulatívne riziko  $\hat{\Lambda}_{NA,j}(t)$ ,  $j = 1, 2$ , (c) Kaplan-Meierove krivky prežívania  $\hat{S}_{KM,j}(t)$ ,  $j = 1, 2$ . (d) Otestujte  $H_0 : \lambda_1(t) = \lambda_2(t)$  oproti  $H_1 : \lambda_1(t) = \theta\lambda_2(t)$  pomocou testovacích štatistík  $Q_{MH}$  a  $Q_{PP}$ . Použite funkcie `survdiff()` s argumentami `rho=0` ( $Q_{MH}$ ) a `rho=1` ( $Q_{PP}$ ). (e) Otestujte  $H_0 : \lambda_1(t) = \lambda_2(t)$  oproti  $H_1 : \lambda_1(t) = \theta\lambda_2(t)$  pomocou testovacích štatistík  $Q_{GW}$ ,  $Q_{CM}$ ,  $Q_{TW}$  a  $Q_{PP}$ . (f) Vypočítajte (1)  $\hat{\theta}_P$ ,  $Var[\hat{\theta}_{MH}]$  a 95%IS pre  $\theta_P$ , (2)  $\hat{\theta}_{MH}$  a 95%IS pre  $\theta_{MH}$  a (3)  $\hat{\theta}_{MH}^*$ .



Obr. 10: Odhady kumulatívneho rizika pre AML dáta



Obr. 11: Kaplan Meierove odhady funkcie prežívania pre AML dáta

## 5 Charakteristiky definované sčítacím procesom

Vo formuláciách sčítacím procesom  $(X_i, \delta_i)$  nahradíme  $(N_i(t), Y_i(t))$ , kde  $N_i(t)$  je počet pozorovaných udalostí v intervale  $\langle 0, t \rangle$  v jednotke  $i$ ,

$$Y_i(t) = \begin{cases} 1 & \text{jednotka } i \text{ je v riziku v čase } t \text{ (pozorujem ju)} \\ 0 & \text{inak} \end{cases}$$

Táto formulácia obsahuje dáta s pravým typom cenzúr ako špeciálny prípad, teda  $Y_i(t) = I(\{T_i \geq t\})$  a  $N_i(t) = I(\{T_i \leq t, \delta = 1\})$ . Všimnime si, že  $N(t)$  je zprava spojitá a  $Y(t)$  zľava spojitá.  $Y(t)$  je príkladom predikovatelného procesu, ktorého hodnoty v čase  $t$  sú známe nekonečne krátko pred  $t$ , v čase  $t^-$ , ak nie skôr. Sčítací proces je stochastický proces začínajúci v čase 0, ktorého trajektória je zprava spojitá funkcia so skokmi veľkosti 1. Pre  $N(t)$  potom platí  $\{N(t) : t \geq 0\}$ ,  $N(0) = 0$  [7].

Odhad kumulatívneho rizika je definovaný na základe agregovaného procesu  $\bar{Y}(t) = \sum_i Y_i(t)$ ,  $\bar{N}(t) = \sum_i N_i(t)$ ,  $d\bar{N}(t) = \Delta\bar{N}(t) = \bar{N}(t) - \bar{N}(t^-)$ , kde  $\bar{N}(t)$  je suma udalostí do času  $t$  vrátane,

$\bar{Y}(t)$  je počet jednotiek v riziku v čase  $t$  (formálne ide o počet jednotiek v riziku v časovom intervale  $(t - \epsilon, t)$  pre malé  $\epsilon$ ).

**Príklad 66** Majme náhodný vektor  $(X_i, \delta_i)$ , definovaný nasledovne (pre nejakú fiktívnu  $i$ -tu štatistickú jednotku, t.j. subjekt)

1)  $(X_i, \delta_i) = (3, 0)$ , t.j. v čase  $X_i = 3$  je cenzúra,  $N_i(t) = N_i(3) = 0$ ,  $Y_i(3) = Y_i(3) = 1$   
 $\rightarrow (N_i(3), Y_i(3)) = (0, 1)$ ,

2)  $(X_i, \delta_i) = (4, 1)$ , t.j. v čase  $X_i = 4$  je udalosť (zlyhanie),  $N_i(4) = 1$ ,  $Y_i(4) = 1$ , t.j.  
 $(N_i(4), Y_i(4)) = (1, 1)$ ,

3) Ak máme viac udalostí:  $(N_i(0.5), Y_i(0.5)) = (1, 1)$ ,  $(N_i(2), Y_i(2)) = (2, 1)$ .

**Klasický Kaplan-Meierov (KM) odhad funkcie prežívania** je definovaný nasledovne

$$\hat{S}_{KM}(t) = \prod_{i:t_i \leq t} [1 - \Delta \hat{\Lambda}_{KM}(t_i)],$$

kde  $\Delta \hat{\Lambda}_{KM}(t_i) = \frac{\Delta \bar{N}(t_i)}{\bar{Y}(t_i)}$  je odhad rizika a **KM odhad kumulatívneho rizika**  $\hat{\Lambda}_{KM}(t) = -\ln \hat{S}_{KM}(t) = -\sum_{i:t_i \leq t} \ln(1 - \Delta \hat{\Lambda}_{KM}(t_i))$ .

**Breslowov (B) odhad funkcie prežívania** definujeme nasledovne

$$\hat{S}_B(t) = \exp(-\hat{\Lambda}_{NA}(t)) = \prod_{i:t_i \leq t} e^{-\Delta \hat{\Lambda}_{NA}(t_i)}$$

kde  $\hat{\Lambda}_{NA}(t) = \sum_{i:t_i \leq t} \hat{\lambda}_i = \sum_{i:t_i \leq t} \frac{d_i}{n_i}$  je **Nelson-Aalenov (NA) odhad kumulatívneho rizika**.

**Greenwoodov (G) odhad rozptylu kumulatívneho rizika** definujeme nasledovne

$$\hat{\sigma}_G^2(t) = \text{Var}_G[\hat{\Lambda}_{KM}(t)] = \sum_{i:t_i \leq t} \frac{\Delta \bar{N}(s)}{\bar{Y}(s) [\bar{Y}(s) - \Delta \bar{N}(s)]},$$

**plug-in odhad rozptylu kumulatívneho rizika** (nazýva sa aj Kleinov odhad; maximálne viero-  
hodný odhad)

$$\hat{\sigma}_P^2(t) = \text{Var}_P[\hat{\Lambda}(t)] = \sum_{i:t_i \leq t} \frac{\Delta \bar{N}(s) (\bar{Y}(s) - \Delta \bar{N}(s))}{\bar{Y}^3(s)},$$

**binomický odhad rozptylu kumulatívneho rizika**

$$\hat{\sigma}_B^2(t) = \text{Var}_B[\hat{\Lambda}(t)] = \sum_{i:t_i \leq t} \frac{\Delta \bar{N}(s) (\bar{Y}(s) - \Delta \bar{N}(s))}{\bar{Y}^2(s) (\bar{Y}(s) - 1)},$$

**NA odhad rozptylu kumulatívneho rizika** (nazýva sa aj Aalenov, Poissonov alebo Tsiatisov odhad)

$$\hat{\sigma}_T^2(t) = \text{Var}_T[\hat{\Lambda}_{NA}(t)] = \sum_{i:t_i \leq t} \frac{\Delta \bar{N}(s)}{\bar{Y}^2(s)}$$

## 6 Softvérová implementácia charakteristík prežívania

☞ knižnica *library(survival)* ponúka nasledovné možnosti. Označme

```
surv.obj <- survfit(Surv(cas,status)~1, type="...", error="...", conf.type
                    ="...").
```

Potom

1. **odhady funkcie prežívania** vypočítame ako

(a)  $\widehat{S}_{KM}(t)$ : `type="kaplan-meier"` (prednastavené);

(b)  $\widehat{S}_B(t)$ : `type="fleming-harrington"`;

(c)  $\widehat{S}_{FHmodB}(t)$ : `type="fh2"`;

2. **odhady rozptylu** ako

(a)  $Var_G[\widehat{S}_{KM}(t)]$ : `error="greenwood"` (prednastavené);

(b)  $Var_G[\widehat{S}_B(t)]$ : `error="tsiatis"`;

3. **typy intervalov spoľahlivosti (IS)** ako

(a) žiadny: `conf.type="none"`;

(b) škála funkcie prežívania: `conf.type="plain"`;

(c) škála kumulatívneho rizika: `conf.type="log"` (prednastavené);

(d) škála logaritmu kumulatívneho rizika: `conf.type="log-log"`.

Ďalšími argumentami sú

4. koeficient spoľahlivosti `conf.int=0.95` (prednastavené);

5. úprava spodnej hranice IS, kde argument

(a) `conf.lower="usual"` (nemodifikovaná dolná hranica IS);

(b) `conf.lower="peto"` (používa Petov efektívny rozsah súboru  $n^{**}(t)$ );

(c) `conf.lower="modified"` (používa Dorey-Korn modifikáciu).

**Priemerný vek prežívania a jeho smerodajná odchýlka** ako aj **medián a jeho smerodajná odchýlka** sa vypočítajú ako

1. `print(surv.obj, print.rmean=TRUE)` alebo

2. `print(surv.obj, rmean="individual")`.

Na rozlíšenie **typu cenzúrovania** je dôležitý počet argumentov funkcie `Surv()`. Ak sú dva, t.j. `Surv(cas,status)`, ide o **pravý typ cenzúrovania**. Ak sú tri, t.j. `Surv(cas,cas1,status)`, potom ide o **intervalové cenzúrovanie**. Pomocným argumentom je `type="..."`, kde rozlišujeme

1. `type="right"` (pravý typ), `type="interval"` (intervalový typ cenzúrovania I. typu, kde interval  $(-\infty, t_i)$  označujeme  $(NA, t_i)$ );

2. `type="interval2"` (intervalový typ cenzúrovania II. typu; kde interval je typu  $(t_{1i}, t_{2i})$ ) alebo

3. interval  $(t_i, \infty)$ , ktorý označujeme  $(t_i, NA)$ .

Dolnou hranicou intervalu môže byť aj 0 a hornou hranicou  $t_{\max}$ .

Výstupmi objektu *surv.obj* a *summary(surv.obj)* sú nasledovné

1. rozsah – **n**,
2. počet jedincov v riziku v jednotlivých časoch – **n.risk**,
3. počet udalostí (zlyhaní) v jednotlivých časoch – **n.event**,
4. počet cenzúr v jednotlivých časoch – **n.censor**,
5. odhad funkcie prežívania v jednotlivých časoch – **surv**,
6. odhad odmocniny z rozptylu funkcie prežívania v jednotlivých časoch – **std.err**,
7. dolná hranica IS pre funkciu prežívania v jednotlivých časoch – **lower**,
8. horná hranica IS pre funkciu prežívania v jednotlivých časoch – **upper**,
9. časy, v ktorých nastalo
  - zlyhanie alebo cenzúra – **time** (pre *surv.obj*).
  - zlyhanie – **time** (pre *summary(surv.obj)*).

## 7 Štruktúra funkcií na cvičenia a do DÚ

Štruktúra funkcie pre odhad kumulatívneho rizika, jeho rozptylu, intervaly spoľahlivosti a pás spoľahlivosti pre kumulatívne riziko a graf, tj. funkcie, ktorá bude „všetko“ počítat' a zobrazovať v jednotlivých časoch zlyhania  $t_0, t_1, \dots, \max(t_{\max}, c_{\max})$ , je nasledovná (číselný výstup je vo forme tabuľky *TAB.kum.riz.IS*, pozri nižšie)

```
kumulativne.riziko.all(cas, status, type="KM", error="G",
                      conf.coef=0.95, conf.type="plain",
                      plot=FALSE, plot.conf.int=FALSE, plot.conf.band=FALSE),
```

kde sú možné nasledovné

### 1. kombinácie odhadov

- $type="KM"$  pre  $\widehat{\Lambda}_{KM}(t)$  a  $error="G"$   $\widehat{\sigma}_G^2(t)$ ; prednastavené;
- $type="NA"$  pre  $\widehat{\Lambda}_{NA}(t)$  a  $error="T"$   $\widehat{\sigma}_T^2(t)$ ; namiesto  $\widehat{\sigma}_T^2(t)$  môžeme použiť aj  $\widehat{\sigma}_B^2(t)$  alebo  $\widehat{\sigma}_K^2(t)$ , potom  $error="B"$ , resp.  $error="K"$ ;
- $type="FHmodNA"$  pre  $\widehat{\Lambda}_{FHmodNA}(t)$  a  $error="FH"$   $\widehat{\sigma}_{FH}^2(t)$ ;

### 2. typy IS

- škála kumulatívneho rizika:  $conf.type="plain"$  (prednastavené);
- škála logaritmu kumulatívneho rizika:  $conf.type="log"$ ;
- škála arcus-sínusová odmocninová:  $conf.type="arc-sin"$ .

### 3. voľby kreslenia obrázkov (číselné výpočty sú vždy výstupom funkcie)

- $plot=FALSE, plot.conf.int=FALSE, plot.conf.band=FALSE$  – nekreslíme žiaden obrázok;
- $plot=TRUE, plot.conf.int=FALSE, plot.conf.band=FALSE$  – kreslíme obrázok len s odhadom kumulatívneho rizika;
- $plot=TRUE, plot.conf.int=TRUE, plot.conf.band=TRUE$  – kreslíme obrázok s odhadom kumulatívneho rizika, a intervalmi a pásmi spoľahlivosti pre kumulatívne riziko.

Štruktúra funkcie pre funkciu prežívania, jej rozptylu, intervaly spoľahlivosti a pás spoľahlivosti pre funkciu prežívania a graf, tj. funkcie, ktorá bude „všetko“ počítat' a zobrazovať v jednotlivých časoch zlyhania  $t_0, t_1, \dots, \max(t_{\max}, c_{\max})$ , je nasledovná (číselný výstup je vo forme tabuľky *TAB.fcia.prez.IS*, pozri nižšie)

```
funkcia.prezivania.all(cas, status, type="KM", error="G",
                      conf.coef=0.95, conf.type="plain",
                      plot=FALSE, plot.conf.int=FALSE, plot.conf.band=FALSE),
```

kde sú možné nasledovné

### 1. kombinácie odhadov

- $type="KM"$  pre  $\widehat{S}_{KM}(t)$  a  $error="GKM"$   $Var_G \left[ \widehat{S}_{KM}(t) \right]$ ,
- $type="B"$  pre  $\widehat{S}_B(t)$  a  $error="NAB"$   $Var_{NA} \left[ \widehat{S}_B(t) \right]$ ,



- (c)  $type="B"$  pre  $\widehat{S}_B(t)$  a  $error="AJB"$   $Var_{AJ}[\widehat{S}_B(t)]$ ,
- (d)  $type="FHmodB"$  pre  $\widehat{S}_{FHmodB}(t)$  a  $error="NAFH"$   $Var_{NA}[\widehat{S}_{FHmodB}(t)]$  a
- (e)  $type="FHmodB"$  pre  $\widehat{S}_{FHmodB}(t)$  a  $error="GFH"$   $Var_G[\widehat{S}_{FHmodB}(t)]$ .

## 2. typy IS

- (a) škála funkcie prežívania:  $conf.type="plain"$  (prednastavené);
- (b) škála logaritmu funkcie prežívania:  $conf.type="log"$ ;
- (c) škála logaritmu kumulatívneho rizika:  $conf.type="log-log"$ ;
- (d) škála arcus-sínusová odmocninová:  $conf.type="arc-sin"$ .

## 3. voľby kreslenia obrázkov (číselné výpočty sú vždy výstupom funkcie)

- (a)  $plot=FALSE$ ,  $plot.conf.int=FALSE$ ,  $plot.conf.band=FALSE$  – nekreslíme žiaden obrázok;
- (b)  $plot=TRUE$ ,  $plot.conf.int=FALSE$ ,  $plot.conf.band=FALSE$  – kreslíme obrázok len s odhadom kumulatívneho rizika;
- (c)  $plot=TRUE$ ,  $plot.conf.int=TRUE$ ,  $plot.conf.band=TRUE$  – kreslíme obrázok s odhadom kumulatívneho rizika, intervalmi a pásmi spoľahlivosti pre kumulatívne riziko.

Vyššie uvedené funkcie budú načítavať vo svojom tele ďalšie funkcie

1.  $pocetnosti(time, status)$  s výstupom v podobe tabuľky obsahujúcej časy do zlyhania  $t_i$  (ozn.  $ti$ ), počty jedincov v riziku  $n_i$  tesne pred časom  $t_i$  (ozn.  $ni$ ) a počty zlyhaní  $d_i$  v čase  $t_i$  (ozn.  $di$ );
2.  $kumulativne.riziko(ti, ni, di, type="KM", error="G")$ , ktorej výstupom je tabuľka  $TAB.kum.riz$  obsahujúca stĺpce  $ti$ ,  $di$ ,  $ni$ ,  $Lambda$  a  $Var[Lambda]$  (nastavenia argumentov  $type$  a  $error$  kopírujú nastavenia funkcie  $kumulativne.riziko.all()$ );
3.  $kumulativne.riziko.IS(TAB.kum.riz, conf.type="plain")$ , ktorej výstupom je tabuľka  $TAB.kum.riz.IS$  obsahujúca stĺpce  $ti$ ,  $di$ ,  $ni$ ,  $Lambda$ ,  $Var[Lambda]$ ,  $DH.IS$  a  $HH.IS$  (nastavenie argumentov  $conf.type$  a  $conf.coef$  kopírujú nastavenie funkcie  $kumulativne.riziko.all()$ );
4.  $funkcia.prezivania(ti, ni, di, type="KM", error="G")$ , ktorej výstupom je tabuľka  $TAB.fcia.prez$  obsahujúca stĺpce  $ti$ ,  $di$ ,  $ni$ ,  $S$  a  $Var[S]$  (nastavenia argumentov  $type$  a  $error$  kopírujú nastavenia funkcie  $funkcia.prezivania.all()$ );
5.  $funkcia.prezivania.IS(TAB.fcia.prez, conf.coef=0.95, conf.type="plain")$ , ktorej výstupom je tabuľka  $TAB.fcia.prez.IS$  obsahujúca stĺpce  $ti$ ,  $di$ ,  $ni$ ,  $S$ ,  $Var[S]$ ,  $DH.IS$  a  $HH.IS$  (nastavenia argumentu  $conf.type$  a  $conf.coef$  kopírujú nastavenie funkcie  $funkcia.prezivania.all()$ ).

Ďalšou funkciou bude funkcia

$zabl.char.prezivania(TAB.fcia.prez, conf.coef=0.95, t, epsilon=0.05,$   
 $plot.mu=FALSE, plot.mrl=FALSE),$

ktorá bude načítavať vo svojom tele ďalšie funkcie (každá z nich bude mať výstup v podobe vektora obsahujúceho `parameter`, `Var[parameter]`, `DH.IS` a `HH.IS` a celkový výstup je tabuľka so štyrmi riadkami (ozn. ako `priemer`, `median`, `priemerny.zostatkovy.zivot` a `median.zostatkoveho.zivota`) a štyrmi stĺpcami; vstupom je okrem iného samotný odhad funkcie prežívania, preto nie je nutné mať ako vstup typ odhadu a jeho rozptyl)

1. `priemerny.cas.prezivania(TAB.fcia.prez,conf.coef=0.95,plot.mu=FALSE)`, ktorej výstupom bude vektor obsahujúci `parameter`  $\hat{\mu}$ , `Var[parameter]`  $\widehat{Var}[\hat{\mu}]$ , `DH.IS` dolná a `HH.IS` horná hranica Waldovho empirického  $100 \times (1 - \alpha)\%$  IS pre strednú hodnotu času prežívania (argument `plot.mu` znamená, že funkcia nakreslí obsah pod krivkou prežívania pomocou funkcie `polygon()`);
2. `median.casu.prezivania(TAB.fcia.prez,conf.coef=0.95,plot.mrl=FALSE)`, ktorej výstupom bude vektor obsahujúci `parameter`  $\tilde{\mu}$ , `Var[parameter]`  $\widehat{Var}[\tilde{\mu}]$ , `DH.IS` dolná a `HH.IS` horná hranica Waldovho empirického  $100 \times (1 - \alpha)\%$  IS pre medián času prežívania (argument `plot.mrl` znamená, že funkcia nakreslí za bodom  $t$  obsah pod krivkou prežívania pomocou funkcie `polygon()`);
3. `priemerny.zostatkovy.zivot(TAB.fcia.prez,t,conf.coef=0.95)`, ktorej vstupom je okrem iného čas  $t$  a výstupom vektor obsahujúci `parameter`  $\widehat{mrl}(t)$ , `Var[parameter]`  $\widehat{Var}[\widehat{mrl}(t)]$ , `DH.IS` dolná a `HH.IS` horná hranica Waldovho empirického  $100 \times (1 - \alpha)\%$  IS pre strednú hodnotu zostatkového života;
4. `median.zostatkoveho.zivota(TAB.fcia.prez,t,conf.coef=0.95)`, ktorej vstupom je okrem iného čas  $t$  a výstupom vektor obsahujúci `parameter`  $\widetilde{mrl}(t)$ , `Var[parameter]`  $\widehat{Var}[\widetilde{mrl}(t)]$ , `DH.IS` dolná a `HH.IS` horná hranica Waldovho empirického  $100 \times (1 - \alpha)\%$  IS pre medián zostatkového života.