Analýza prežívania

Zlyhávanie a cenzúrovanie, funkcia vierohodnosti, základné charakteristiky a ich odhady, intervaly a pásy spoľahlivosti

Stanislav Katina[1]

[1]Ústav matematiky a statistiky
Přírodovědecká fakulta
Masarykova univerzita

ZS 2013

INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Poďakovanie

Tento učební text vznikl za přispění Evropského sociálního fondu a státního rozpočtu ČR prostřednictvím Operačního programu Vzdělávání pro konkurenceschopnost v rámci projektu Univerzitní výuka matematiky v měnícím se světě

(CZ.1.07/2.2.00/15.0203)

Aké otázky v analýze prežívania riešime?

Príklady z praxe

- Odhadujeme a interpretujeme funkciu prežívania a riziko
- Porovnávame funkcie prežívania a riziká
- Modelujeme vzťah medzi vysvetľujúcimi premennými a časom prežívania

Otázky v analýze prežívania v aplikáciách

Príklady z praxe

Prežívanie pacientov po infarkte myokardu (IM) v rámci sekundárnej prevencie závažných kardiovaskulárnych problémov u pacientov s polymorfizmom glykoproteínu IV (GP VI 13254C/T) v membráne krvných doštičiek. [**Thrombosis Research** 125, 2: 61–4, 2009]

105 pacientov sledovaných v priemere 19($\pm$10.8) mesiacov

Otázky v analýze prežívania v aplikáciách
Príklady z praxe

Zlyhania: *smrť*, *ďalší IM*, *ďalšia selektívna koronarografia* (SKG: percutaneous coronary intervention (*PCI*, coronary angioplasty), coronary artery bypass graft (*CABG*)), *ďalša cievna mozgová príhoda* (CMP; stroke), *ďalšia hospitalizácia* (re-intervencia); sledované kombinácie: smrť/IM/re-intervencia a smrť/IM/re-intervencia/CMP [**MACE; Major Adverse Cardiac Events**, hlavné nepriaznivé srdcové udalosti]

Adjustujúce (rizikové) premenné:

1. *pohlavie* (žena=0, muž=1)
2. *hypertenzia* (nie=0, áno=1)
3. *hyperlipidémia* (nie=0, áno=1)
4. *fajčenie* (nefajčiar=0, fajčiar a bývalý fajčiar=1)
5. *diabetes* (nie=0, áno=1)
6. *srdcové zlyhanie* (**NYHA**; **New York Heart Association**; Classes: I = 0; II, III, IV = 1)

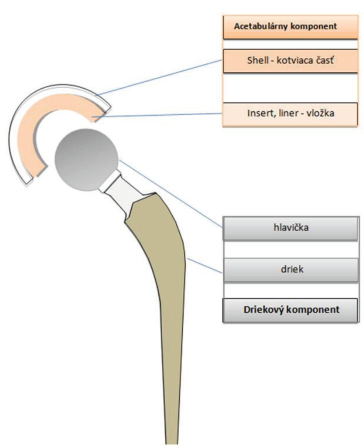
Otázky v analýze prežívania v aplikáciách
Príklady z praxe

Analýza prežívania implantátov bedra a kolena na Slovensku v rokoch 2003−2011. [**Acta Chir. Orthop. Traum. Čech.** 80: 1−85, 2013]

49 668 operácií (primárnych operácií a revízií) zo všetkých slovenských ortopedických kliník za roky 2003−2011:

- 38 485 THA (Total Hip Arthroplasty)
- 11 183 TKA (Total Knee Arthroplasty)

Otázky v analýze prežívania v aplikáciách
Príklady z praxe



Otázky v analýze prežívania v aplikáciách
Príklady z praxe

Zlyhania: zlyhanie komponentu implantátu

Adjustujúce (rizikové, prognostické) premenné:

1. *typ komponentu* (acetabulárny=0, femorálny=1)
2. *fixácia komponentu* (necementovaný=0, cementovaný=0)
3. *pohlavie* (žena=0, muž=1)
4. *cementovacia technika* (necementovaný=0, generácia cementu I = 1, generácia cementu II = 2, generácia cementu III = 3)
5. *diagnóza pri primárnej operácii* (primárna coxartróza = 1, dysplastická coxartróza = 2, poúrazová coxartróza = 3, aseptická nekróza hlavy = 4, M.Perthes = 5, reumatoidná artritída = 6, zlomenina krčku = 7)
6. *dôvod revízie* (spolu 18 dôvodov)
7. *revidované časti* (spolu 19 častí) a pod.

Prežívanie pacientov s chronickou myeloidnou leukémiou (CML). [**Neoplasma**, 92, 5: 381–7, 2005]

589 pacientov s CML, z ktorých 78 absolvovalo transplantáciu krvotvorných kmeňových buniek kostnej drene (allogeneic transplantation; *transplantácia od HLA-identického súrodenca alebo nepríbuzného darcu*; HLA znamená human leukocyte antigen) a zároveň majú odobrané vzorky periférnej krvi a kostnej drene pred a po transplantácii na Katedre genetiky Národného onkologického ústavu v Bratislave v rokoch 1990 až 2002

Zlyhania: úmrtie pacienta

Adjustujúce (prognostické, rizikové) premenné:

1. *vek* pacienta v čase transplantácie (skupina 1: $<$20 rokov, skupina 2: [20,40), skupina 3: $\geq$40)
2. *fáza CML* (spolu dve fázy; prvá chronická fáza = 1, ďalšie chronické fázy = 2)
3. *pohlavie darcu a príjemcu* (m–m, m–ž, ž–m, ž–ž)
4. *čas od diagnózy po transplantáciu* ($<$ 1 rok, $\geq$ 1 rok)

Example

Akútna myelogénna leukémia (acute myelogenous leukemia, **AML**). Po absolvovaní chemoterapie a zmiernení príznakov, boli pacienti náhodne rozdelení do dvoch skupín. Prvá skupina (skupina A) dostala udržujúcu chemoterapiu a druhá (kontrolná; skupina B) nie. Cieľom bolo zistiť, či udržujúca chemoterapia predlžuje čas do remisie (opätovného zhoršenia stavu).

Sk	čas po kompletnú remisiu (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

(číslo = čas do zlyhania, číslo a plus ($+$) = čas do cenzúry)

Klasický prístup vs. analýza prežívania
Tri náhľady na problém analýzy AML dát

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

Klasický prístup vs. analýza prežívania
Príklady

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

(čísla sú v týždňoch)

Klasický prístup vs. analýza prežívania
Príklady

Example

Cystická fibróza (CF) je autozomálna genetická choroba spôsobená mutáciou génu pre CFTR (cystic fibrosis transmembrane conductance regulator). Postihuje prevažne pľúca, ale aj pankreas, pečeň a črevo. V celoslovenskej databáze pacientov CF rozlišujeme pacientov s **jasnou klinickou formou** (**typická forma**, 259 živých, 112 zomrelých) a pacientov s **atypickou formu** (188 živých). Spolu teda 559 pacientov, 447 živých a 112 zomrelých. Aký je priemerný vek (prežívania) a medián (prežívania) v rokoch?

Klasický prístup vs. analýza prežívania
Príklady

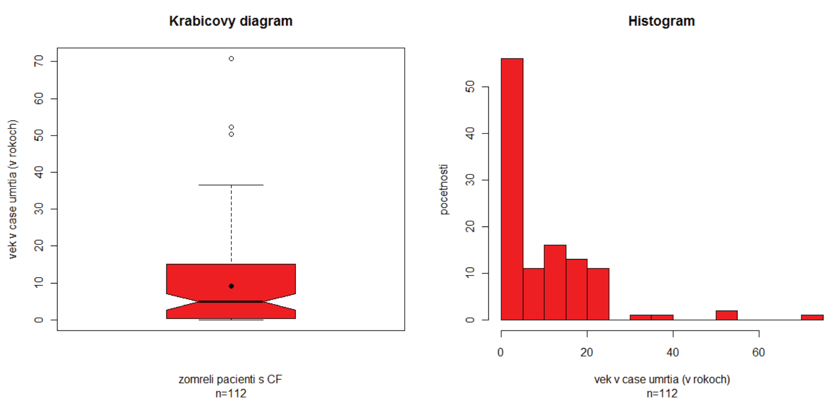
skupina/počty	typická forma CF	atypická forma CF	spolu
živí	259	188	447
zomrelí	112	0	112
spolu	371	188	559

	problém 1	**problém 3**
	typická CF	typická CF
$\overline{x}$	9.22	45.05
$\widetilde{x}$	4.90	52.26

(čísla sú v rokoch)

- Rozdiel medzi priemerným vekom prežívania pacientov s typickou formou CF a priemerným vekom zomrelých je 35.83 roka (podobne pre medián je tento rozdiel 47.36 roka)
- 95% IS pre strednú hodnotu je $(7.02, 11.41)$ a pre medián $(2.72, 7.08)$

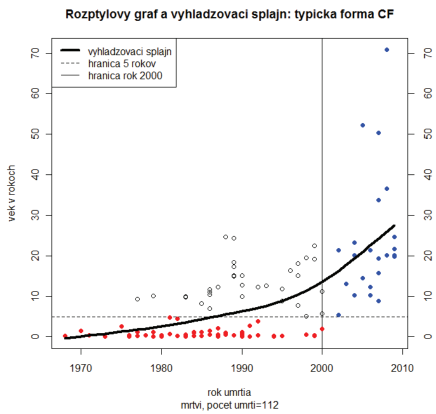
- priemerný vek prežívania pre všetkých pacientov bez rozdielu typu CF je 53.94 ± 2.10 rokov, kde 95% IS je rovný $(49.82, 58.06)$
- medián prežívania pre všetkých pacientov bez rozdielu typu CF je 70.82 roka; 95% **IS pre medián zatiaľ nie je možné vypočítať**
- Priemerný vek prežívania pre pacientov s typickou formou CF je 45.05 ± 2.47 rokov, kde 95% IS je $(40.21, 49.89)$
- Median prežívania je 52.26 roka, dolna hranica 95% IS pre medián je 36.43 roka; **Hornú hranicu IS pre medián zatiaľ nie je možné vypočítať**



Tabuľka: Početnosti zomrelých v päťročných vekových intervaloch ($n = 112$). Označenia: vekové intervaly (zdola je interval otvorený a zhora uzavretý, okrem prvého, ktorý je aj zdola uzavretý): $I_1 = \langle 0, 5 \rangle$, $I_2 = (5, 10\rangle$, $I_3 = (10, 15\rangle$, $I_4 = (15, 20\rangle$, $I_5 = (20, 25\rangle$, $I_6 = 25, \max(vek)\rangle$.

vekové intervaly	I_1	I_2	I_3	I_4	I_5	I_6
početnosti	56	11	16	13	11	5
percentá	50%	9.82%	14.29%	11.61%	9.82%	4.46%

Klasický prístup vs. analýza prežívania
Príklady



Klasický prístup vs. analýza prežívania
Príklady

Ďalšie možné otázky
Zlyhanie: smrť
Adjustujúce (prognostické) premenné: *antropologické ukazovatele*, *funkčné charakteristiky pľúc* a pod.

Udalosť
Úvodné definície

Udalosť: *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) chodoby
- **time-to-progression** – prvé znaky progresie choroby

Cenúrovanie
Úvodné definície

Cenzúra: *ukončenie pozorovania z dôvodu iného ako je zlyhanie alebo smrť pacienta* – do konca sledovaného obdobia dôjde k úmrtiu len niektorých pacientov, zatiaľ čo u ostatných k úmrtiu do konca sledovaného obdobia buď nedôjde alebo sa títo pacienti z pozorovania stratia

Príklady cenúr:

- **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

Základné princípy:

1. predpoklad – všetkých n jedincov vstupuje do experimentu súčasne
2. príčina cenzúrovania – plánované ukončenie experimentu
3. ide o **cenzúrovanie časom** – zvolíme pevné číslo t_c, ktoré nazveme *fixovaný cenzurujúci čas*
4. $T^{(1)} < T^{(2)} < \ldots < T^{(d)}$, kde $T^{(d)} < t_c < T^{(d+1)}$
5. **náhodná veličina** – počet skutočne pozorovaných zlyhaní $d \in \{0, 1, \ldots, n\}$
6. pozorujeme $X_1, X_2, \ldots, X_n$, kde

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

7. skutočnému pozorovaniu potom zodpovedá náhodný vektor (X_i, δ_i), kde

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

Základné princípy:

1. predpoklad – všetkých n jedincov vstupuje do experimentu súčasne
2. príčina cenzúrovania – plánované ukončenie experimentu
3. ide o **cenzúrovanie zlyhaním** – zvolíme si pevné číslo d, ktoré nazveme *fixovaný počet zlyhaní*; ukončenie teda nastáva po vopred zvolenom počte d zlyhaní, kde $d = [np] + 1, p \in (0, 1)$
4. $X_1 = T^{(1)}, X_2 = T^{(2)}, \ldots, X_d = T^{(d)}, X_{d+1} = T^{(d)}, \ldots X_n = T^{(d)}$
5. **náhodná veličina** – čas trvania experimentu
6. pozorujeme $X_1, X_2, \ldots, X_n$, kde

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

7. skutočnému pozorovaniu potom zodpovedá náhodný vektor (X_i, δ_i), kde

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

Základné princípy:

1. predpoklad – všetkých n jedincov vstupuje do experimentu súčasne
2. príčina cenzúrovania – plánované ukončenie experimentu
3. ide o **cenzúrovanie zlyhaním** – zvolíme čísla $t_{ci}, i = 1, 2, \ldots, k$, ktoré nazveme *fixované cenzurujúce časy*, v čase t_{ci} vyradíme m_i subjektov
4. $t_{c1} < t_{c2} < \ldots < t_{ck}$
5. v čase t_{c1} vyradíme m_1 subjektov, v čase t_{c2} vyradíme m_2 subjektov, ..., v čase t_{ck} vyradíme m_k subjektov
6. po k-tom kroku máme vyradených $m_1 + m_2 + \ldots + m_k$ subjektov
7. **náhodná veličina** – počet skutočne pozorovaných zlyhaní $d \in \{0, 1, \ldots, n\}$

Základné princípy:

1. predpoklad – všetkých n jedincov vstupuje do experimentu súčasne
2. príčina cenzúrovania – plánované ukončenie experimentu
3. ide o **cenzúrovanie časom** – zvolíme čísla d_i, ktoré nazveme *fixované počty zlyhaní*; vyradenie teda nastáva po vopred zvolenom počte d_i zlyhaní, kde $d_i = [np_i] + 1, p_i \in (0, 1)$
4. po d_1 zlyhaniach vyradíme m_1 subjektov, po d_2 zlyhaniach vyradíme m_2 subjektov, ..., po d_k zlyhaniach vyradíme m_k subjektov
5. po k-tom kroku máme vyradených $m_1 + m_2 + \ldots + m_k$ subjektov
6. **náhodná veličina** – čas trvania experimentu

Základné princípy:

1. predpoklad – n jedincov nevstupuje do experimentu súčasne
2. **čas do zlyhania** $T_1, T_2, \ldots, T_n$ sú nezávislé, rovnako rozdelené náhodné premenné, kde náhodná veličina T_i $(i = 1, \ldots, n)$ má hustotu $f(t)$ a distribučnú funkciu $F(t)$
3. **čas do cenzúrovania** $C_1, C_2, \ldots, C_n$ sú nezávislé, rovnako rozdelené náhodné premenné, kde náhodná veličina C_i $(i = 1, \ldots, n)$ má hustotu $g(t)$ a distribučnú funkciu $G(t)$
4. pozorujeme $X_1, X_2, \ldots, X_n$, kde
$$X_i = \min(T_i, C_i) = \begin{cases} T_i, T_i \le C_i, & \text{pre necenzúrované } X_i \\ C_i, T_i > C_i, & \text{pre cenzúrované } X_i \end{cases}$$
5. skutočnému pozorovaniu potom zodpovedá náhodný vektor (X, δ), kde $X_i = \min(T_i, C_i)$ a
$$\delta_i = \begin{cases} 1, T_i \le C_i, & \text{pre necenzúrované } X \\ 0, T_i > C_i, & \text{pre cenzúrované } X \end{cases}$$
6. **náhodná veličina** – čas trvania experimentu a čas do cenzúry (ak $C_i = c$, ide o **ľubovoľné cenzúrovanie**)

Základné princípy:

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

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

Example (nádor pľúc, animálny model)

Laboratórne myši sú injektované látkou, ktorá spôsobuje nádor. Kedže 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.

Základné princípy:

Majme opäť n subjektov. Označme $T_i, i = 1, 2, \ldots, n$, **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 $\delta_{1i} = I(T_i \le C_{1i})$, $\delta_{2i} = I(C_{1i} < T_i \le C_{2i})$ a $\delta_{3i} = I(T_i > C_{2i})$, t.j.

$$\delta_{1i} = \begin{cases} 1, T_i \le C_{1i}, & \text{pre necenzúrované } X_i \\ 0, T_i > C_{1i}, & \text{pre cenzúrované } X_i \end{cases},$$

$$\delta_{2i} = \begin{cases} 1, C_{1i} < T_i \le C_{2i}, & \text{pre necenzúrované } X_i \\ 0, T_i > C_{2i}, & \text{pre cenzúrované } X_i \end{cases}$$

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

Example (nádor pľúc, pacienti)

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

Základné princípy:

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 $(C_{1i}, C_{2i}\rangle$, 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 zopovedá náhodný vektor
$$(X_{1i}, X_{2i}, \delta_{1i}, \delta_{2i}).$$

Všimnime si, že δ_{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**.

Cenúrovanie
Funkcia vierohodnosti – pravé typy cenzúrovania

1. cenzúrovanie I. typu

$$L = \prod_{i=1}^{n} f(x_i)^{\delta_i} \times S_f(t_c)^{1-\delta_i}$$

2. cenzúrovanie II. typu

$$L = \frac{n!}{(n-d)!} f(t_{(1)}) f(t_{(2)}) \ldots f(t_{(d)}) \times S_f(t_{(d)})^{n-d}$$

3. náhodné cenzúrovanie

$$L = \prod_{i=1}^{n} f(x_i)^{\delta_i} S_f(x_i)^{1-\delta_i} = \prod_{i=1}^{n} \lambda(x_i)^{\delta_i} S_f(x_i)$$

Cenúrovanie
Funkcia vierohodnosti – intervalové cenzúrovanie

1. intervalové cenzúrovanie I. typu

$$L = \prod_{i=1}^{n} [S_f(x_i)]^{1-\delta_i} [F(x_i)]^{\delta_i}$$

2. intervalové cenzúrovanie II. typu

$$L = \prod_{i=1}^{n} [F(x_{1i})]^{\delta_{1i}} [F(x_{2i}) - F(x_{1i})]^{\delta_{2i}} [S_f(x_{2i})]^{\delta_{3i}},$$

kde $\delta_{3i} = 1 - \delta_{1i} - \delta_{2i}$

Označenia
Časy do zlyhania

Definition

Majme neusporiadané časy $t_1, t_2, \ldots, t_n$. Zoradené časy zapíšeme ako $t_{(1)}, t_{(2)}, \ldots, t_{(n)}$. Pokiaľ predpokladáme, že $t_1, t_2, \ldots, t_n$ sú už zoradené, t.j. $t_1 < t_2 < \ldots < t_n$, označenia v ďalšom texte sa týmto preznačením zjednodušia. Potom $t_n = t_{\max}$. Ak $t_{\max} < c_{\max}$, potom bez straty na všeobecnosti bude $t_n = c_{\max}$ (pozri aj výpočet strednej hodnoty času prežívania, kde je potrebné situáciu $t_{\max} < c_{\max}$ zohľadniť). V časoch cenzúr c sú hodnoty $S(c)$ a $\Lambda(c)$ – ako aj ostatných charakteristík – identické ich hodnotám v najbližšom čase zlyhania t, ktorý predchádza c. Preto, bez straty na všeobecnosti, uvažujeme n zoradených časov, v ktorých sa charakteristiky prežívania počítajú. Tieto časy označujeme $t_1, t_2, \ldots, t_n$. Ak máme v časoch t_i zhody, t.j. $t_1 \leq t_2 \leq \ldots \leq t_n$, potom počet rôznych časov bude $l \leq n$ a $t_l = t_{\max}$.

Riziko
Príklady

Example (zadanie z prednášky)

Závislosť hodnoty rizika na jednotkách času.

$\lambda(t) = \lim_{\Delta t \to 0^+} \frac{\Pr(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$

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

Ak $\Delta t = \frac{1}{3}$ dňa, potom $\lambda(t) = \frac{\frac{1}{4}}{\frac{1}{3}} = 0.75$ na deň

Ak $\Delta t = \frac{1}{21}$ týždňa, potom $\lambda(t) = \frac{\frac{1}{4}}{\frac{1}{21}} = 5.25$ na týždeň

Definície

Príklady

Example (zadanie z prednášky)

Majme náhodný vektor (X_i, δ_i), definovaný nasledovne (pre nejaku 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: $(N_i(0.5), Y_i(0.5)) = (1, 1)$, $(N_i(2), Y_i(2)) = (2, 1)$

Zoznam zadaní príkladov

Príklady

Example (domáca úloha)

Nech nezáporná náhodná veličina T je charakterizovaná funkciou prežívania $S(T)$. Nech je k-ty moment, $\mathbb{E}(T^k)$, konečný, $\mathbb{E}(T^k) < \infty, k \in \mathbb{N}$. (a) Ukážte, že platí $\mathbb{E}(T) = \sum_{t \in \mathbb{N}_0} \Pr(T > t) = \sum_{t \in \mathbb{N}_0} 1 - F(T) = \sum_{t \in \mathbb{N}_0} S(T)$. Použite pri tom definíciu strednej hodnoty $\mathbb{E}(T) = \sum_{t \in \mathbb{N}_0} t \Pr(t)$ 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 $\mathbb{E}(T) = \int_0^\infty S(t)\, dt$. Použite pri tom definíciu strednej hodnoty $\mathbb{E}(T) = \int_0^\infty t f(t) dt$, aplikujte vlastnosti súm z DÚ 1A 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 $\mathbb{E}(T^k) = k \int_0^\infty t^{k-1} S(t) dt$.

Zoznam zadaní príkladov

Príklady

Example (zadanie z cvičenia)

AML (pokrač.) Vypočítajte empirickú funkciu prežívania pre skupinu A.

Skupina	čas po kompletnú remisiu (v týždňoch)	n	udalostí	cenzúr
skupina A	$9, 13, 13+, 18, 23, 28+, 31, 34, 45+, 48, 161+$	11	7	4

Example (zadanie z cvičenia)

AML (pokrač.) Naprogramujte v R algoritmus na výpočet empirickej funkcie prežívania a aplikujte ho na skupinu A.

Example (zadanie z cvičenia)

AML (pokrač.) Naprogramujte v R algoritmus na výpočet empirickej funkcie prežívania len pre zlyhania (cenzúry nemeberime do úvahy) a aplikujte ho na skupinu A.

Zoznam zadaní príkladov

Príklady

Example (zadanie z prednásky)

Odvoďte maximálne vierohodný odhad funkcie prežívania $\widehat{S}_{KM}(t)$.

Example (zadanie z cvičenia)

AML (pokrač.) Naprogramujte v R algoritmus na výpočet KM odhadu funkcie prežívania a aplikujte ho na skupinu A.

Example (zadanie z cvičenia)

AML (pokrač.). Výpočtom a graficky porovnajte empirickú funkciu prežívania $S_n(t)$ s KM odhadom funkcie prežívania $\widehat{S}_{KM}(t)$ pre skupinu A.

Example (zadanie z cvičenie)

AML (pokrač.). Výpočtom a graficky porovnajte empirickú funkciu prežívania $S_n(t)$ len pre časy zlyhania (cenzúry nemeberime do úvahy) s KM odhadom funkcie prežívania $\widehat{S}_{KM}(t)$ pre skupinu A.

Example (domáca úloha)

Použitím funkcie vierohodnosti $L = \prod_{i=1}^{I} \lambda_i^{d_i}(1-\lambda_i)^{n_i-d_i}$ odvoďte maximálne vierohodný odhad $\widehat{Var}\left[\widehat{\lambda}_i\right]$, $i = 1, 2, \ldots, I$ a $\widehat{Var}\left[\widehat{\Lambda}(t)\right]$.

Example (zadanie z cvičenie)

AML (pokrač.). Vypočítajte rozptyl KM odhadu funkcie prežívania v čase 13 (pre skupinu A). Využite Greenwoodovu formulu.

Example (zadanie z cvičenie)

AML (pokrač.). Vypočítajte odhad rizika $\widehat{\lambda}$, odhad rizika v intervale $t_i \leq t < t_{i+1}$, odhad kumulatívneho rizika $\widehat{\Lambda}_{KM}$ a $\widehat{\Lambda}_{NA}$ spolu s ich rozptylmi $\widehat{Var}[\widehat{\Lambda}_{KM}]$ a $\widehat{Var}[\widehat{\Lambda}_{NA}]$ v čase 26 (pre skupinu A).

Example (zadanie z cvičenie)

AML (pokrač.). Nakreslite a porovnajte odhady kriviek prežívania $\widehat{S}_{KM}(t)$, $\widehat{S}_B(t)$ a $\widehat{S}_{FHmodB}(t)$ pre skupinu A.

Example (zadanie z cvičenie)

AML (pokrač.). Vypočítajte KM odhad funkcie prežívania $\widehat{S}_{KM}(t)$ a 95% IS pre $S(t)$ vo všetkých bodoch t v 1) plain škále, 2) log-škále a 3) log-log škále (pre skupinu A).

Example (domáca úloha)

Ak v náhodnom výbere nie sú cenzúry, skóre testovacia štatistika Z_S má za platnosti H_0 štandardizované normálne rozdelenie $Z_S = \frac{\widehat{S}(t)-S(t)}{\sqrt{\frac{S(t)(1-S(t))}{n}}} \sim N(0,1)$, kde platí $\{S(t) : |z| \leq z_{\alpha/2}\}$. Potom riešením kvadratickej rovnice $\{S(t) : (1 + \frac{z^2_{\alpha/2}}{n})S^2(t) - (2\widehat{S}(t) + \frac{z^2_{\alpha/2}}{n})S(t) + \widehat{S}(t) \leq 0\}$ bude $100 \times (1-\alpha)\%$ IS pre $S(t)$. Odvoďte tento interval a upravte ho do podoby: vzorec pre stred IS $\pm z_{\alpha/2}\sqrt{\text{vzorec}}$.

Example (zadanie z cvičenie)

AML (pokrač.) Naprogramujte v R algoritmus na výpočet obsahu pod $\widehat{S}_{KM}(t)$ krivkou. Aplikujte ho na skupinu A. Porovnajte s aritmetickým priemerom časov do zlyhania a aritmetickým priemerom časov do zlyhania a časov cenzúr.

Example (zadanie z cvičenie)

AML (pokrač.). Vypočítajte priemerný čas prežívania $\widehat{\mu}$ a jeho rozptyl $\widehat{Var}(\widehat{\mu})$, medián času prežívania $\widetilde{\mu}$ a jeho rozptyl $\widehat{Var}[\widetilde{\mu}]$ (pre skupinu A). Porovnajte s necenzurovaným mediánom.

Example (zadanie z cvičenie)

Naprogramujte v R funkciu na výpočet kvantilov času prežívania t_p a ich $100 \times (1-\alpha)\%$ intervalov spoľahlivosti.

Example (zadanie z cvičenie)

(a) Naprogramujte v R funkcie na výpočet nasledovných odhadov funkcií prežívania $\widehat{S}_{KM}(t)$, $\widehat{S}_{KMmod}(t)$, $\widehat{S}_{B}(t)$ a $\widehat{S}_{FHmodB}(t)$, kde

1. $\widehat{S}_{KMmod}(t) = \prod_{i:t_i \leq t}(1 - \frac{1}{n_i})$ [pre nerozsekané a aj rozsekané zhody]
2. $\widehat{Var}\left[\widehat{S}_{KM}(t)\right]$, dolnú (DH) a hornú (HH) hranicu 95% *IS* v log-škále,
3. $\widehat{S}_{B}(t) = \exp\left(-\widehat{\Lambda}_{NA}(t)\right)$ [pre nerozsekané zhody]
4. $\widehat{S}_{FHmodB}(t) = \exp\left(-\widehat{\Lambda}_{FHmodB}(t)\right)$ [pre rozsekané zhody]

(b) Vypočítajte tieto odhady pre dáta (pozri tabuľku)

Example (zadanie z cvičenie; pokrač.)

(c) Naprogramujte v R výpočet počtu cenzúr v čase t_i, ak poznáte d_i a n_i (pozri tabuľku).
(d) Naprogramujte v R funkciu na výpočet odhadu rozptylu funkcie prežívania $\widehat{Var}\left[\widehat{S}_{KM}(t)\right]$, dolnú a hornú hranicu 95% IS pre $S(t)$ v log-škále.
(e) Vypočítajte tento odhad a IS pre $S(t)$ pre dáta (pozri tabuľku).

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

Example (zadanie z cvičenie)

AML (pokrač.) Naprogramujte v R algoritmus na výpočet $100 \times (1-\alpha)\%$ pásov spoľahlivosti pre funkciu prežívania v škále $S(t)$ – (a) Nairov a (b) Hall-Walnerov. Aplikujte na skupinu A. Výsledok porovnajte s IS pre $S(t)$ v škále $S(t)$. Na obrázku zobrazte IS bodovo (zobrazenie, ktoré je prednastavené v R je nesprávne).

Example (domáca úloha)

AML (pokrač.) Naprogramujte v R algoritmus na výpočet $100 \times (1-\alpha)\%$ pásov spoľahlivosti pre funkciu prežívania v log-log škále – (a) Nairov a (b) Hall-Walnerov. Aplikujte na skupinu A. Výsledok porovnajte s IS pre $S(t)$ v log-log škále. Na obrázku zobrazte IS bodovo (zobrazenie, ktoré je prednastavené v R je nesprávne).

Example (domáca úloha)

AML (pokrač.) Naprogramujte v R algoritmus na výpočet $100 \times (1-\alpha)\%$ pásov spoľahlivosti pre kumulatívne riziko v log-log škále – (a) Nairov a (b) Hall-Walnerov. Aplikujte na skupinu A. Výsledok porovnajte s IS pre kumulatívne riziko v log-log škále. Na obrázku zobrazte IS bodovo.

Example (domáca úloha)

AML (pokrač.) Naprogramujte v R algoritmus na výpočet odhadu strednej hodnoty zostatkového života a jej rozptylu a aplikujte ho na skupinu A v čase $t = 30$ týždňov.

Empirická funkcia prežívania
Príklady

Example

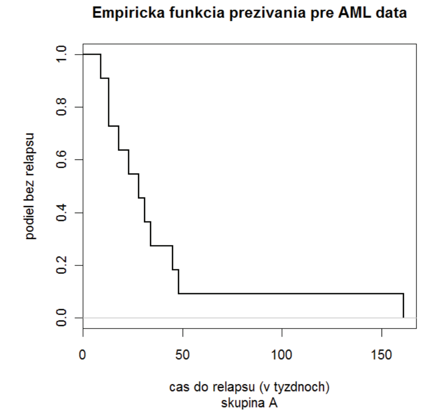
AML (pokrač.) Vypočítajte empirickú funkciu prežívania pre skupinu A.

	čas po kompletnú remisiu (v týždňoch)	n	udalostí	cenzúr
A	$9, 13, 13+, 18, 23, 28+, 31, 34, 45+, 48, 161+$	11	7	4

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

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

Empirická funkcia prežívania
Príklady



Empirická funkcia prežívania a KM odhad
Príklady

Example

AML (pokrač.). Porovnajte empirickú funkciu prežívania $S_n(t)$ s KM odhadom funkcie prežívania $\widehat{S}_{KM}(t)$ pre skupinu A.

	čas po kompletnú remisiu (v týždňoch)	n	udalostí	cenzúr
A	$9, 13, 13+, 18, 23, 28+, 31, 34, 45+, 48, 161+$	11	7	4

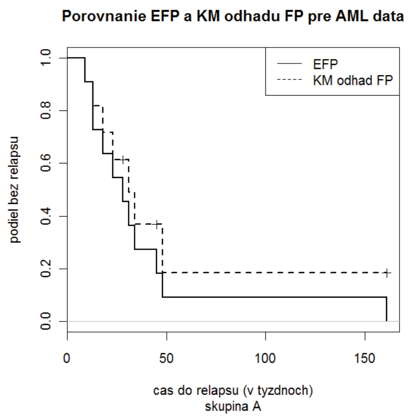
$$\widehat{S}_{KM}(t) = \prod_{i:t_i \le t} \left[1 - \widehat{\lambda}_i\right], \text{ kde } \widehat{\lambda}_i = \frac{d_i}{n_i}$$

Empirická funkcia prežívania a KM odhad
Príklady

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

Empirická funkcia prežívania a KM odhad
Príklady

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



Odhady
Príklady

Example

AML (pokrač.). Vypočítajte rozptyl KM odhadu funkcie prežívania v čase 13 (pre skupinu A). Využite Greenwoodovu formulu.

$$\widehat{Var_G[\widehat{S}_{KM}(t)]} = \widehat{S}^2_{KM}(t) Var\left[\ln \widehat{S}_{KM}(t)\right] = \widehat{S}^2_{KM}(t) \sum_{i:t_i \leq t} \frac{d_i}{n_i(n_i-d_i)}$$

$$\widehat{Var_G[\widehat{S}_{KM}(t)]} = \widehat{S}^2_{KM}(t) \sum_{i:t_i \leq t} \frac{d_i}{n_i(n_i-d_i)}$$

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

$$\widehat{SE_G[\widehat{S}_{KM}(13)]} = 0.1166$$

Odhady
Príklady

Example

AML (pokrač.). Vypočítajte riziko $\widehat{\lambda}$ a kumulatívne riziko $\widehat{\Lambda}_{KM}$ a $\widehat{\Lambda}_{NA}$ spolu s ich rozptylmi $\widehat{Var[\widehat{\Lambda}_{KM}]}$ a $\widehat{Var[\widehat{\Lambda}_{NA}]}$ v čase 26 (pre skupinu A) [viď R kód v prílohe, Príklad 2].

$\widehat{\lambda}(t_i) = \frac{d_i}{n_i}$

Odhad rizika v intervale $t_i \leq t < t_{i+1}$ je rovný $\widehat{\lambda}(t) = \frac{d_i}{n_i(t_{i+1}-t_i)}$; hovoríme mu aj KM typ odhadu; odhad rizika zlyhania na jednotku času v intervale $\langle t_i, t_{i+1})$

$$\widehat{\Lambda}_{KM}(t) = -\ln(\widehat{S}_{KM}(t)) = -\ln\left(\prod_{i:t_i \leq t} \frac{n_i-d_i}{n_i}\right)$$

$$\widehat{Var[\widehat{\Lambda}_{KM}(t)]} = \sum_{i:t_i \leq t} \frac{d_i}{n_i(n_i-d_i)}$$

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

Odhady
Príklady

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

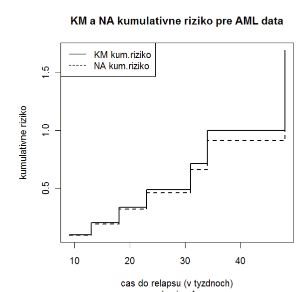
$$\widehat{\lambda}(26) = \widehat{\lambda}(23) = \frac{1}{7(31-23)} = 0.018$$

$$\widehat{\Lambda}_{KM}(26) = -\ln(\widehat{S}_{KM}(26)) = -\ln(0.61) = 0.49,$$

$$\widehat{\Lambda}_{NA}(26) = \frac{1}{11} + \frac{1}{10} + \frac{1}{8} + \frac{1}{7} = 0.4588$$

$$\widehat{Var[\widehat{\Lambda}_{KM}(t)]} = \frac{1}{11(11-1)} + \frac{1}{10(10-1)} + \frac{1}{8(8-1)} + \frac{1}{7(7-1)} = 0.0619$$

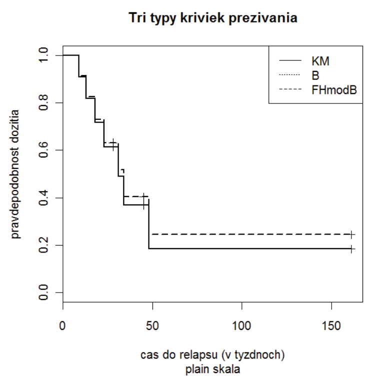
$$\widehat{Var[\widehat{\Lambda}_{NA}(t)]} = \frac{1}{11^2} + \frac{1}{10^2} + \frac{1}{8^2} + \frac{1}{7^2} = 0.0543$$



Odhady
Príklady

Example

AML (pokrač.). Nakreslite a porovnajte odhady kriviek prežívania $\widehat{S}_{KM}(t)$, $\widehat{S}_B(t)$ a $\widehat{S}_{FHmodB}(t)$ [viď R kód v prílohe, Príklad 1].



Medián a priemerný čas prežívania
Príklady

Example

AML (pokrač.). Vypočítajte priemerný čas prežívania $\widehat{\mu}$ a jeho rozptyl $\widehat{Var(\widehat{\mu})}$, medián času prežívania $\widetilde{\mu}$ a jeho rozptyl $\widehat{Var(\widetilde{\mu})}$ (pre skupinu A). Porovnajte s necenzurovaným mediánom.

$\widehat{\mu} = \sum_{i=1}^{l}(t_i - t_{i-1})\widehat{S}(t_{i-1}) = \sum_{i=0}^{l} \Delta t_i \widehat{S}(t_i)$, kde $\Delta t_i = t_{i+1} - t_i$, $l \leq n$ je počet rôznych zlyhaní, $t_0 = 0$, $\widehat{S}(t_0) = 1$ a $\widehat{S}(t_{i-1})$ je výška funkcie v bode t_{i-1}.

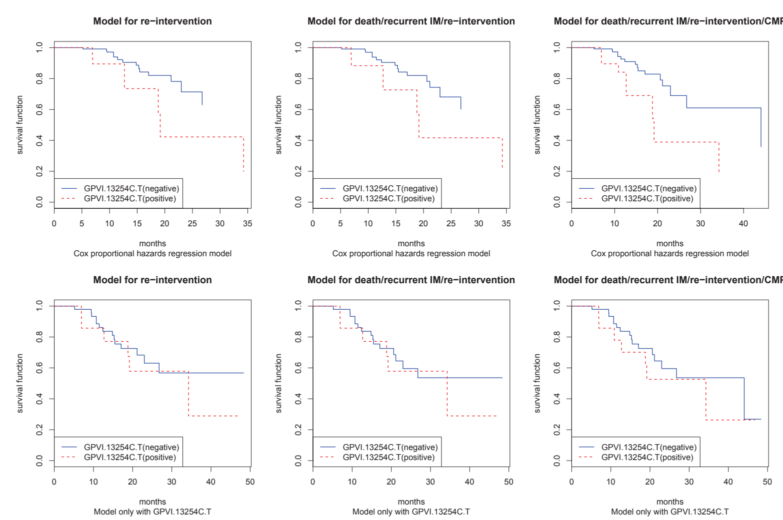
$$\widehat{Var}\,[\widehat{\mu}] = \sum_{i=1}^{l}\left[\sum_{t_i \leq t_j \leq t_n} \widehat{S}\,(t_j)\right]^2 \frac{d_i}{n_i(n_i - d_i)}$$

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

Medián a priemerný čas prežívania
Príklady

$t_{(n)} = 161$
$\widehat{\mu} = 52.6$ týždňa
$\widehat{Var}\,[\widehat{\mu}] = 19.8^2$
95% IS $= (13.792, 91.408)$ týždňa
$\widehat{t}_{0.5} = 31$ týždňov
$\widehat{u}_{0.5} = \max\{t_i : \widehat{S}(t_i) \geq 0.55\} = 23$
$\widehat{l}_{0.5} = \min\{t_i : \widehat{S}(t_i) \leq 0.45\} = 34$
$\widehat{f}(31) = \frac{\widehat{S}(\widehat{u}_{0.5}) - \widehat{S}(\widehat{l}_{0.5})}{\widehat{l}_{0.5} - \widehat{u}_{0.5}} = \frac{\widehat{S}(23) - \widehat{S}(34)}{34 - 23} = \frac{0.6136364 - 0.3681818}{11} = 0.022$
$\widehat{Var}[31] = (\frac{0.16419327}{0.02231405})^2 = 54.144$
$\sqrt{\widehat{Var}[31]} = 7.358$
95% IS $= (16.578, 45.422)$ týždňa

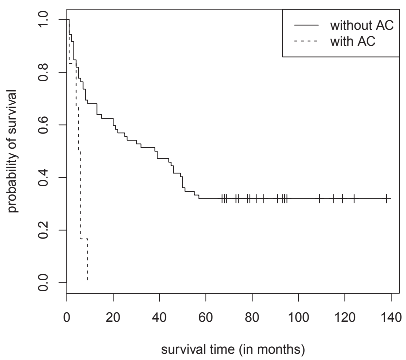
Prežívanie pacientov po infarkte myokardu
Príklad: MACE (rôzne kombinácie; s a bez adjustácie) [funkcia prežívania]



Chronická myeloidná leukémia

Príklad: zmeny po transplantácii (typické a netypické) [funkcia prežívania]

- typické zmeny: $\widehat{\mu} = 58.08$ (± 6.70) mesiaca ($\widetilde{\mu} = 38.00$) a pravdepodobnosť úmrtia $49/72$
- netypické zmeny: $\widehat{\mu} = 5.17$ (± 0.98) mesiaca ($\widetilde{\mu} = 5.50$) pravdepodobnosť úmrtia 6/6



Odhady

Príklady

Example

Vypočítajte odhady 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).

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

Odhady

Príklady

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

- $\widehat{S}_{B}(12) = \exp\left[-(1/70 + 2/68)\right] = 0.9572$
- $\widehat{S}_{FHmodB}(12) = \exp\left[-(1/70 + 1/68 + 1/67)\right] = 0.9570$

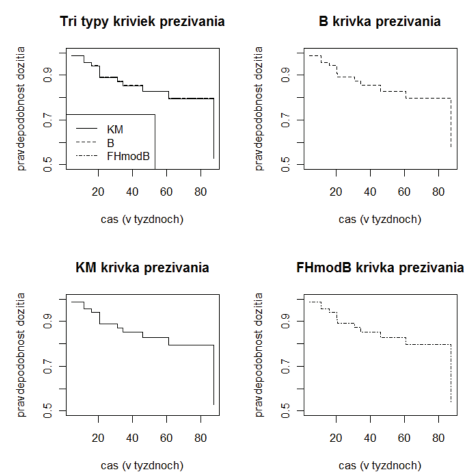
Pre 5 zhôd v čase 88 platí:

- $\widehat{S}_{KM}(88) = 0.5294$
- $\widehat{S}_{FHmodB}(88) = 0.5395$
- $\widehat{S}_{B}(88) = 0.5709$

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

Odhady

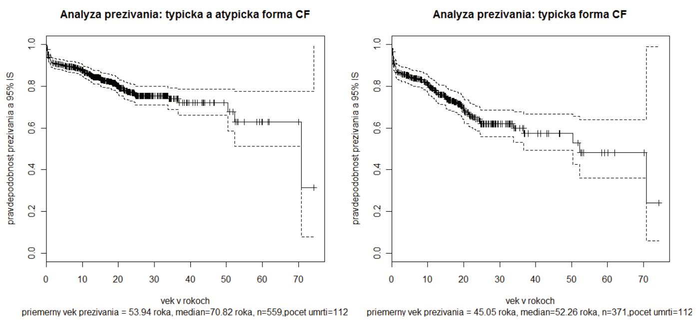
Príklady



Odhady
Príklady

Example

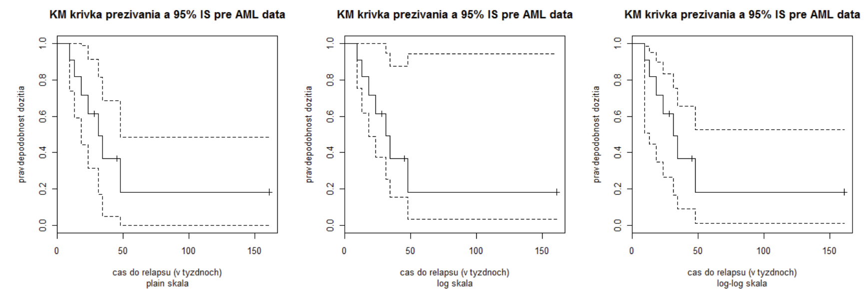
CF (pokrač.) Nakreslite KM odhad funkcie prežívania $\widehat{S}_{KM}(t)$ a 95% IS pre $S(t)$ vo všetkých bodoch t pre pacientov s CF (všetci pacienti a typická forma).



Odhady
Príklady

Example

AML (pokrač.). Vypočítajte KM odhad funkcie prežívania $\widehat{S}_{KM}(t)$ a 95% IS pre $S(t)$ vo všetkých bodoch t v 1) plain škále, 2) log-škále a 3) log-log škále (pre skupinu A).



Prehľad vzorcov
Odhady kumulatívneho rizika

- *Nelson-Aalenov (NA) odhad kumulatívneho rizika*

$$\widehat{\Lambda}_{NA}(t) = \int_0^t \frac{d\overline{N}(s)}{\overline{Y}(s)} ds \approx \sum_{i:t_i \le t} \frac{\Delta \overline{N}(t_i)}{\overline{Y}(t_i)}$$

- *Flemingom a Harringtonom (FH) modifikovaný NA odhad kumulatívneho rizika*

$$\begin{aligned}\widehat{\Lambda}_{FHmodNA}(t) &= \int_0^t \left[\sum_{j=0}^{\Delta\overline{N}(s)-1} \frac{1}{\overline{Y}(s) - j} \right] ds \\ &\approx \sum_{i:t_i \le t} \left[\sum_{j=0}^{\Delta\overline{N}(t_i)-1} \frac{1}{\overline{Y}(t_i) - j} \right]\end{aligned}$$

Prehľad vzorcov
Odhady kumulatívneho rizika

- *Nelson-Aalenov (NA) odhad kumulatívneho rizika*

$$\widehat{\Lambda}_{NA}(t) = \sum_{i:t_i \le t} \widehat{\lambda}(t_i) = \sum_{i:t_i \le t} \frac{d_i}{n_i},$$

- *Flemingom a Harringtonom (FH) modifikovaný NA odhad kumulatívneho rizika*

$$\widehat{\Lambda}_{FHmodNA}(t) = \sum_{i:t_i \le t} \left[\sum_{j=0}^{d_i - 1} \frac{1}{n_i - j} \right]$$

Prehľad vzorcov
Odhady funkcie prežívania

- *Kamplan-Meierov odhad funkcie prežívania*

$$\widehat{S}_{KM}(t) = \prod_{i:t_i\le t}\left[1-\Delta\widehat{\Lambda}(t_i)\right], \Delta\widehat{\Lambda}(t_i)=\frac{\Delta\overline{N}(t_i)}{\overline{Y}(t_i)}$$

- *Breslowov odhad funkcie prežívania*

$$\widehat{S}_B(t)=\exp\left(-\widehat{\Lambda}(t)\right)=\prod_{i:t_i\le t} e^{-\Delta\widehat{\Lambda}(t_i)}, \Delta\widehat{\Lambda}(t_i)=\frac{\Delta\overline{N}(t_i)}{\overline{Y}(t_i)}$$

- *Flemingom a Harringtonom modifikovaný Breslowov odhad funkcie prežívania*

$$\widehat{S}_{FHmodB}(t)=\exp\left(-\widehat{\Lambda}_{FHmodNA}(t)\right)=\prod_{i:t_i\le t} e^{-\Delta\widehat{\Lambda}_{FHmodNA}(t_i)}$$

Prehľad vzorcov
Odhady funkcie prežívania

- *Kamplan-Meierov odhad funkcie prežívania*

$$\widehat{S}_{KM}(t)=\prod_{i:t_i\le t}\left[1-\frac{d_i}{n_i}\right]=\widehat{S}_{KMmod}=\prod_{i:t_i\le t}\left[1-\sum_{j=0}^{d_i-1}\frac{1}{n_i-j}\right]$$

- *Breslowov odhad funkcie prežívania*

$$\widehat{S}_B(t)=\exp\left(-\widehat{\Lambda}_{NA}(t)\right)=\prod_{i:t_i\le t} e^{-\frac{d_i}{n_i}}=e^{-\sum_{i:t_i\le t}\frac{d_i}{n_i}}$$

- *Flemingom a Harringtonom modifikovaný Breslowov odhad funkcie prežívania*

$$\widehat{S}_{FHmodB}(t)=\exp\left(-\widehat{\Lambda}_{FHmodNA}(t)\right)=e^{-\sum_{i:t_i\le t}\left[\sum_{j=0}^{d_i-1}\frac{1}{n_i-j}\right]}$$

Prehľad vzorcov
Odhady rozptylu kumulatívneho rizika

- *Greenwoodov odhad rozptylu kumulatívneho rizika*

$$\widehat{Var_G\left[\widehat{\Lambda}(t)\right]}=\widehat{Var_G\left[-\ln\widehat{S}_{KM}(t)\right]}=\int_0^t\frac{d\overline{N}(s)}{\overline{Y}(s)\left[\overline{Y}(s)-d\overline{N}(s)\right]}ds$$

- *NA odhad rozptylu kumulatívneho rizika*

$$\widehat{Var\left[\widehat{\Lambda}_{NA}(t)\right]}=\int_0^t\frac{d\overline{N}(s)}{\overline{Y}^2(s)}ds$$

- *Flemingom a Harringtonom modifikovaný NA odhad rozptylu kumulatívneho rizika*

$$\widehat{Var\left[\widehat{\Lambda}_{FHmodNA}(t)\right]}=\int_0^t\left[\sum_{j=0}^{\Delta\overline{N}(s)-1}\frac{1}{\left[\overline{Y}(s)-j\right]^2}\right]ds$$

Prehľad vzorcov
Odhady rozptylu kumulatívneho rizika

- *Greenwoodov odhad rozptylu kumulatívneho rizika*

$$\widehat{Var_G\left[\widehat{\Lambda}(t)\right]}=\sum_{i:t_i\le t}\frac{\Delta\overline{N}(t_i)}{\overline{Y}(t_i)\left[\overline{Y}(t_i)-\Delta\overline{N}(t_i)\right]}$$

- *NA odhad rozptylu kumulatívneho rizika*

$$\widehat{Var\left[\widehat{\Lambda}_{NA}(t)\right]}=\sum_{i:t_i\le t}\frac{\Delta\overline{N}(t_i)}{\overline{Y}^2(t_i)}$$

- *Flemingom a Harringtonom modifikovaný NA odhad rozptylu kumulatívneho rizika*

$$\widehat{Var\left[\widehat{\Lambda}_{FHmodNA}(t)\right]}=\sum_{i:t_i\le t}\left[\sum_{j=0}^{\Delta\overline{N}(t_i)-1}\frac{1}{\left[\overline{Y}(t_i)-j\right]^2}\right]$$

Prehľad vzorcov
Odhady rozptylu kumulatívneho rizika

- *Greenwoodov odhad rozptylu kumulatívneho rizika*

$$\widehat{Var_G\left[\widehat{\Lambda}(t)\right]} = \sum_{i:t_i \leq t} \frac{\Delta \overline{N}(t_i)}{\overline{Y}(t_i)\left[\overline{Y}(t_i) - \Delta \overline{N}(t_i)\right]} = \sum_{i:t_i \leq t} \frac{d_i}{n_i(n_i - d_i)}$$

- *NA odhad rozptylu kumulatívneho rizika*

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

- *Flemingom a Harringtonom modifikovaný NA odhad rozptylu kumulatívneho rizika*

$$Var\,\widehat{\left[\widehat{\Lambda}_{FHmodNA}(t)\right]} = \sum_{i:t_i \leq t} \left[\sum_{j=0}^{d_i - 1} \frac{1}{(n_i - j)^2}\right]$$

Prehľad vzorcov
$(1-\alpha)100\%$ IS $S(t)$ v t

- *škála* $S(t)$ (**survival (plane) scale**)

$$\widehat{S}(t) \pm u_{\alpha/2}\sqrt{Var\,\widehat{\left[\widehat{S}(t)\right]}}, \text{ kde } Var\left[\widehat{S}(t)\right] = \widehat{S}^2(t)\,Var\left[\widehat{\Lambda}(t)\right]$$

- *škála kumulatívneho rizika* (**log-survival scale**)

$$\ln \widehat{S}(t) \pm u_{\alpha/2}\sqrt{Var\,\widehat{\left[\ln \widehat{S}(t)\right]}}, \widehat{S}(t)\exp\left(\pm u_{\alpha/2}\sqrt{Var\,\widehat{\left[\widehat{\Lambda}(t)\right]}}\right)$$

$$\widehat{\Lambda}(t)\exp\left(\pm u_{\alpha/2}\sqrt{Var\,\widehat{\left[\ln \widehat{\Lambda}(t)\right]}}\right)$$

Prehľad vzorcov
$(1-\alpha)100\%$ IS $S(t)$ v t

$$\text{kde } Var\,\widehat{\left[\ln \widehat{\Lambda}(t)\right]} \approx Var\,\widehat{\left[\widehat{\Lambda}(t)\right]} / \left[\widehat{\Lambda}(t)\right]^2$$

- *log-log škála* (**log-log (survival) scale**; škála log-kumulatívneho rizika)

$$\ln(-\ln \widehat{S}(t)) \pm u_{\alpha/2}\sqrt{\widehat{Var[W]}},$$

$$\text{kde } W = \ln(-\ln \widehat{S}(t))), \widehat{Var[W]} = Var\,\widehat{\left[-\ln \widehat{S}(t)\right]}/(\ln \widehat{S}(t))^2,$$

$$(\widehat{S}(t))^{\exp(\pm u_{\alpha/2}\sqrt{\widehat{Var[W]}})},$$

a

$$\exp\left[-\widehat{\Lambda}(t)\exp\left(\pm u_{\alpha/2}\sqrt{Var\,\widehat{\left[\ln \widehat{\Lambda}(t)\right]}}\right)\right]$$

Prehľad vzorcov
Odhad strednej hodnoty do času prežívania a jej rozptylu

(Urezaný) odhad strednej hodnoty času do zlyhania $E[T]$

$$\widehat{\mu} = \int_0^{t_{\max}} \widehat{S}(t)\,dt, \widehat{\mu} = \sum_{i=0}^{I} \Delta t_i \widehat{S}(t_i), \Delta t_i = t_{i+1} - t_i, I \leq n,$$

kde $\widehat{S}(t)$ je *KM* odhad funkcie prežívania, $t_{\max}$ je maximum pozorovaných časov, I počet rôznych zlyhaní

Odhad rozptylu $Var[\widehat{\mu}]$

$$\widehat{Var[\widehat{\mu}]} = \int_0^T \left[\int_t^T \widehat{S}(u)du\right]^2 \frac{d\overline{N}(t)}{\overline{Y}(t)\left[\overline{Y}(t) - d\overline{N}(t)\right]}dt$$

$$\widehat{Var[\widehat{\mu}]} = \sum_{i:t_i \leq t_n} \left[\sum_{t_i \leq t_j \leq t_n} \widehat{S}(t_j)\right]^2 \frac{\Delta \overline{N}(t_i)}{\overline{Y}(t_i)\left[\overline{Y}(t_i) - \Delta \overline{N}(t_i)\right]}$$

Prehľad vzorcov
Odhad mediánu do času prežívania a jeho IS

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

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

Horná a dolná hranica IS pre medián – 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
- dolná hranica IS pre medián je prvý čas, v ktorom je dolná hranica IS pre $S(t)$ menšia alebo rovná 0.5

To korešponduje s narysovaním horizontálnej úsečky na grafe krivky prežívania, tj. pretnutím tejto úsečky s krivkou prežívania, dolnou a hornou hranicou IS pre $S(t)$ v danom čase

Prehľad vzorcov
Odhad kvantilov do času prežívania a ich rozptylov

Nech t_p je *p-ty kvantil* rozdelenia T ($100 \times$ *p-ty percentil*), teda $F(t_p) = \Pr(T < t_p) = p, t_p = F^{-1}(p)$. Potom

$$S(t_p) = \Pr(T \geq t_p) = 1 - p, t_p \leq S^{-1}(1-p)$$

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 $\widehat{Var_G(\widehat{S}(\widehat{t}_p))}$ dostaneme

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

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 $i = 1, 2, \ldots, I \leq n$, I je počet rozdielnych časov zlyhania, ϵ je veľmi malé číslo; vo všeobecnosti je $\epsilon = 0.05$ akceptovateľné, ale musí byť veľké, ak $|\widehat{l}_p - \widehat{u}_p| \approx 0$

Prehľad vzorcov
Odhady strednej hodnoty zostatkového života a jej rozptyl

$$\widehat{\mathrm{mrl}}(t) = E(\widehat{T - t|T} > t) = \frac{\int_t^{t_{\max}} \widehat{S}(u)du}{\widehat{S}(t)}$$

$$\widehat{\mathrm{mrl}(t)} = \frac{(t_{i+1} - t)\widehat{S}(t_i) + \sum_{j \geq i+1}(t_{j+1} - t_j)\widehat{S}(t_j)}{\widehat{S}(t)}, t_i \leq t < t_{i+1}$$

$$\begin{aligned} \widehat{Var[\widehat{\mathrm{mrl}}(t)]} &= \frac{1}{\widehat{S}^2(t)}\left(\sum_{i:t \leq t_i \leq t_{\max}}\left(\int_{t_i}^{t_{\max}} \widehat{S}(u)du\right)^2 \frac{d_i}{n_i(n_i - d_i)}\right. \\ &+ \left.\left(\int_t^{t_{\max}} \widehat{S}(u)du\right)^2 \sum_{i:t_i \leq t} \frac{d_i}{n_i(n_i - d_i)}\right). \end{aligned}$$

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

R
R príkazy [R je voľne dostupné na http://cran.r-project.org/]

Voľby argumentov fcie `Surfit` v knižnici `library(survival)`:

```
Surfit(surv(time,status)~1, type="...",error="...",conf.type
="...")
```

1. $\widehat{S}_{KM}(t)$: `type="kaplan-meier"` (prednastavené)
2. $\widehat{S}_{B}(t)$: `type="fleming-harrington"`
3. $\widehat{S}_{FHmodB}(t)$: `type="fh2"`
4. $\widehat{Var_G\left[\widehat{\Lambda}_{KM}(t)\right]}$: `error = "greenwood"` (prednastavené)
5. $\widehat{Var\left[\widehat{\Lambda}_{NA}(t)\right]} = \widehat{\sigma}_T^2$: `error = "tsiatis"`
6. **žiadny**: `conf.type ="none"`
7. **survival (plane) scale**: `conf.type ="plain"`
8. **log-survival scale**: `conf.type ="log"` (prednastavené)
9. **log-log (survival) scale**: `conf.type="log-log"`
10. koeficient spoľahlivosti `conf.int=0.95` (prednastavené)

Označme `surv.obj <- survfit(Surv(cas,status)~1)`. Priemerný vek prežívania a jeho smerodajná odchýlka (medián a jeho smerodajná odchýlka je súčasťou výstupu) sa vypočíta 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 `type="right"` (pravý typ), `type="interval"` (intervalový typ cenzúrovania I. typu; kde interval $(-\infty, t_i\rangle$ označujeme $(NA, t_i\rangle$), `type="interval2"` (intervalový typ cenzúrovania II. typu; kde interval je typu $(t_{1i}, t_{2i}\rangle$ alebo interval (t_i, ∞), ktorý označujeme (t_i, NA)). Dolnou hranicou intervalu môže byť aj 0 a hornou hranicou t_{max}.

Example

Intervalový typ cenzúrovania pre dáta `heart` – intervaly sa nachádzajú v stĺpcoch `heart$start` a `heart$stop`, status (udalosť) je v stĺpci `heart$event`. Význam premenných pozri v `help(heart)`.

Príklad 1 (poznámky sú uvedené za znakom #):

```
library(survival)
attach(aml)
names(aml)
aml.A <- aml[x=="Maintained",1]
status.A <- aml[x=="Maintained",2]
KM.aml.A.KM <- survfit(Surv(aml.A,status.A)~1,conf.type =
"plain",type="kaplan-meier") # Ŝ_KM(t)
KM.aml.A.B <- survfit(Surv(aml.A,status.A)~1,conf.type =
"plain",type="fleming-harrington") # Ŝ_B(t)
KM.aml.A.FHmodB <- survfit(Surv(aml.A,status.A)~1,conf.type =
"plain",type="fh2") # Ŝ_FHmodB(t)
# obrazok (lwd: hrubka ciary, lty: typ ciary)
plot(KM.aml.A.KM,xlab="cas do relapsu (v tyzdnoch)",
ylab="pravdepodobnost dozitia",conf.int=FALSE,lwd=2)
lines(KM.aml.A.KM,lty=1,lwd=2)
lines(KM.aml.A.B,lty=3,lwd=2)
lines(KM.aml.A.FHmodB,lty=2,lwd=2)
legend("topright",c("KM","B","FHmodB"),lty=c(1,3,2))
title(main="Tri typy kriviek prezivania",sub="plain skala")
```


Príklad 2 (poznámky sú uvedené za znakom #):

```
cas <- summary(KM.aml.A)$time # casy zlyhania t_i
n.i <- summary(KM.aml.A)$n.risk # pocet jedincov v riziku n_i
d.i <- summary(KM.aml.A)$n.event # pocet zlyhani d_i
KM <- summary(KM.aml.A)$surv # Ŝ_KM(t)
SE.KM <- summary(KM.aml.A)$std.err # SE(Ŝ_KM(t))
lambda.KM <- d.i/n.i
DIFF <- diff(cas, lag = 1) # dlzka intervalu napravo od t_i
DIFF[length(DIFF) + 1] <- NA # NA su chybajuce hodnoty
lambda.INT <- lambda.KM/DIFF
Lambda.KM <- -log(KM) # Λ̂_KM(t)
Lambda.NA <- cumsum(lambda.KM) # Λ̂_NA(t)
sumand <- d.i/(n.i*n.i)
se.Lambda.KM <- SE.KM/lambda.KM # SE(Λ̂_KM(t))
se.Lambda.NA <- sqrt(cumsum(sumand)) # SE(Λ̂_NA(t))
# round(cislo,kolko.des.miest)
# data.frame: datovy ramec
RIZ <- round(data.frame(cas, n.i, d.i, lambda.KM,lambda.INT,
Lambda.KM, se.Lambda.KM, Lambda.NA,se.Lambda.NA),4)
```


Príklad 2 (pokrač.):

```
# obrazok
# schodovita funkcia: type="s", prazdny obrazok: type="n"
# $ znamena indexaciu stlpca z datoveho ramca v podobe
# RIZ$nazov.stlpca
plot(RIZ$cas,RIZ$Lambda.KM,xlab="cas do relapsu (v
tyzdnoch)",ylab="kumulativne riziko",type="n")
lines(RIZ$cas,RIZ$Lambda.KM,lty=1,lwd=2,type="s")
lines(RIZ$cas,RIZ$Lambda.NA,lty=2,lwd=2,type="s")
abline(h=0,col="gray")
title(main="KM a NA kumulativne riziko pre AML
data",sub="skupina A")
legend("topleft",c("KM kum.riziko","NA kum.riziko"),lty=c(1,2))
```

Viac sa o R dozviete
1) v mojich skriptách na http://www.iam.fmph.uniba.sk/skripta/katina/, kde "podtržník _"vo význame priradenia výpočtu nejakému názvu treba nahradiť "`<-`"(ide o v minulosti používanú syntax v komerčnej verzii `R`, programe `S-PLUS`)
2) v knižke **An Introduction to R** na http://cran.r-project.org/ v časti *Manuals*

Inštalácia: *R Binaries*→*windows*→*base*→Download R 3.0.1 for Windows