Analýza přežití

Iveta Selingerová

Ústav matematiky a statistiky
Přírodovědecká fakulta
Masarykova univerzita

15.3.2012

Obsah

Analýza přežití

Popis dat odpovídajících času od vstupní události (**počáteční bod**) do výskytu nějaké sledované události (**koncový bod**).

Vstupní událost	Koncová událost
narození	úmrtí
počátek léčby	uzdravení
počátek onemocnění	výskyt onemocnění
vstup jedince do studie	návrat onemocnění
zavedení nového přístroje do provozu	porucha přístroje

Doba mezi počáteční a koncovou událostí je označována jako **doba přežití**.

Vlastnosti dat v analýze přežití

Proč nelze použít standardní statistické metody?

- data přežití nejsou obecně symetricky rozdělena
- koncová událost nebyla zpozorována
  - jedinec stále naživu
  - s pozorovaným jedincem ztratíme kontakt
  - jedinec zemřel na jinou nemoc

CENZOROVÁNÍ

Vlastnosti dat v analýze přežití

Proč nelze použít standardní statistické metody?

- data přežití nejsou obecně symetricky rozdělena
- koncová událost nebyla zpozorována
  - jedinec stále naživu
  - s pozorovaným jedincem ztratíme kontakt
  - jedinec zemřel na jinou nemoc

CENZOROVÁNÍ

Cenzorování

- **Cenzorování zprava**

  skutečná doba přežití je vyšší než doba pozorování

- **Cenzorování zleva**

  doba přežití jedince je menší než sledovaná

- **Intervalové cenzorování**

  jedince je možné sledovat jen v určitých okamžicích

Funkce přežití

Funkce přežití $\overline{F}(t)$ je definována jako pravděpodobnost, že doba přežití je větší nebo rovna t.

$$\overline{F}(t) = P(T \geq t) = \int_t^{\infty} f(u)\mathrm{d}u = 1 - F(t).$$

- zleva spojitá
- klesající
- platí

$$\overline{F}(0) = 1,$$
$$\lim_{t \to \infty} \overline{F}(t) = 0.$$

Riziková funkce

Riziková funkce $\lambda(t)$ vyjadřuje pravděpodobnost, že jedinec nepřežije krátký časový interval Δt za předpokladu, že se dožil času t.

$$\lambda(t)\Delta t = P(t \leq T < t + \Delta t | T \geq t)$$

$$\lambda(t) = \frac{f(t)}{\overline{F}(t)}$$

$$\overline{F}(t) = \mathrm{e}^{-\int_0^t \lambda(u)\mathrm{d}u}$$

Integrál v tomto výrazu se nazývá *kumulativní riziková funkce*

$$\Lambda(t) = \int_0^t \lambda(u)\mathrm{d}u.$$

Riziková funkce je nezáporná funkce, která může mít různý průběh.

Náhodný model cenzorování

Nechť $T_1, T_2, \ldots, T_n$ jsou nezávislé a identicky rozdělené časy přežití pro n pozorování v rámci studie s distribuční funkcí F. Dále nechť $C_1, C_2, \ldots, C_n$ jsou nezávislé a identicky rozdělené časy cenzorování s distribuční funkcí G. Časy cenzorování se obvykle předpokládají nezávislé na časech přežití. Není možné pozorovat jak T_i tak C_i. Místo toho budeme pozorovat dvojici $(X_i, \delta_i), i = 1, \ldots, n$, kde $X_i = \min(T_i, C_i)$ a $\delta_i = I_{T_i \leq C_i}$. Funkce I je indikátorová funkce, tzn.

$$I_{\{T_i \leq C_i\}} = \begin{cases} 1 & T_i \leq C_i, \\ 0 & T_i > C_i. \end{cases}$$

δ_i udává, zda je i-té pozorování cenzorováno, či nikoliv.

Kaplan-Meierův odhad funkce přežití

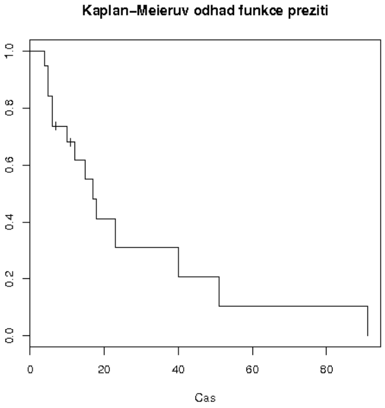
Uspořádaný náhodný výběr $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ a příslušné indikátory cenzorování $\delta_{(1)}, \delta_{(2)}, \ldots, \delta_{(n)}$.

$$\widehat{\overline{F}}(x) = \prod_{i:X_{(i)}<x} \left(\frac{n-i}{n-i+1}\right)^{\delta_{(i)}}$$

Pro odhad $\widehat{\overline{F}}(x)$ platí $\widehat{\overline{F}}(x) = 1$ pro $x \leq X_{(1)}$.

Kaplan-Meierův odhad funkce přežití

čas	6	7	4	11	10	5	51	5	40	91
cenzorování	1	0	1	0	1	1	1	1	1	1
čas	18	18	6	23	15	18	12	12	17	
cenzorování	0	0	1	1	1	1	0	1	1	



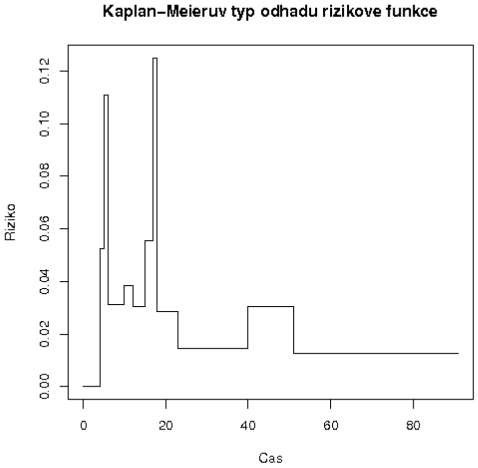
Odhad rizikové funkce

Nechť $x_1, x_2 \ldots, x_m$ jsou všechny body nespojitosti Kaplan-Meierova odhadu funkce přežití takové, že $x_0 = 0 < x_1 < x_2 < \cdots < x_m \leq X_{(n)}$.
Počet úmrtí v čase x_j je $\sum\limits_{X_{(i)}=x_j} \delta_{(i)}$ a počet jedinců, kteří jsou v čase x_j v riziku je $\sum_{i=1}^n I_{\{X_{(i)} \geq x_j\}}$.

$$\hat{\lambda}(x) = \frac{\sum_{X_{(i)}=x_j} \delta_{(i)}}{(x_{j+1} - x_j) \sum_{i=1}^n I_{\{X_{(i)} \geq x_j\}}},$$

pro $x \in [x_j, x_{j+1})$, $j = 0, 1, \ldots, m-1$.

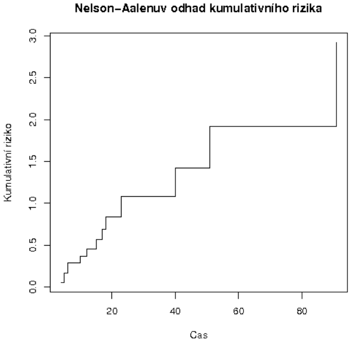
Odhad rizikové funkce



Odhad kumulativní rizikové funkce

Nelson-Aalenův odhad

$$\hat{\Lambda}_n(x) = \sum_{X_{(i)} \le x} \frac{\delta_{(i)}}{n - i + 1}$$



Testy pro srovnání funkcí přežití pro dvě skupiny

Nulová hypotéza: $\overline{F}_1(x) = \overline{F}_2(x) \; \forall x$
Alternativní hypotéza: $\overline{F}_1(x) \neq \overline{F}_2(x) \; \forall x$

- **Cox-Mantelův (log-rank) test**
- **Gehan-Wilcoxonův test**
- Peto-Peto test
- Tarone-Wareův test

Pokud se funkce přežití kříží je vhodné použít Gehan-Wilcoxonův test, jinak je vhodnější log-rank test.