# Coxův model proporcionálního rizika

Iveta Selingerová

Ústav matematiky a statistiky
Přírodovědecká fakulta
Masarykova univerzita

12.4.2012

# Úvod

- testujeme vliv faktoru na přežití ⇒ porovnání křivek přežití přežití (log-rank či Gehan-Wilcoxonův test)
  - např. srovnáváme přežití mužů a žen
- chceme studovat více faktorů najednou nebo máme kvantitativní proměnné ⇒ regresní model
  - např. přežití může záviset na pohlaví, věku, výsledcích vyšetření, typu léčby, ...

## Regresní model

- Parametrický model - předpokládáme, že známe rozdělení přežití (Normální, exponenciální, lognormální, ...)
- Semiparametrický model - založen pouze na poměru rizik (Coxův model)

# Úvod

- testujeme vliv faktoru na přežití ⇒ porovnání křivek přežití přežití (log-rank či Gehan-Wilcoxonův test)
  - např. srovnáváme přežití mužů a žen
- chceme studovat více faktorů najednou nebo máme kvantitativní proměnné ⇒ regresní model
  - např. přežití může záviset na pohlaví, věku, výsledcích vyšetření, typu léčby, . . .

Regresní model

- Parametrický model - předpokládáme, že známe rozdělení přežití (Normální, exponenciální, lognormální, . . . )
- Semiparametrický model - založen pouze na poměru rizik (Coxův model)

# Definice Coxova modelu

$(T_i, \delta_i, \mathbf{Z}_i(t))$

- $T_i$ pozorovaný čas pro i-tého jedince
- $\delta_i$ indikátor cezorování pro i-tého jedince
- $\mathbf{Z}_i(t)$ vektor kovariátů nebo rizikových faktorů pro i-tého jedince, které mohou mít efekt na přežití
  - časově závislý, např. výsledek stejného vyšetření při jednotlivých kontrolách
  - konstantní - známý v čase 0, např. pohlaví $\mathbf{Z}_i(t) = \mathbf{Z}_i$

Máme $p$ nezávisle proměnných $\mathbf{Z} = (Z_1, \ldots, Z_p)$.

# Definice Coxova modelu

Coxův model má tvar

$$\lambda(t|\mathbf{Z}) = \lambda_0(t)\exp(\beta^{\mathsf{T}}\mathbf{Z}) = \lambda_0(t)\exp\left(\sum_{k=1}^{p}\beta_k Z_k\right)$$

$\lambda(t|\mathbf{Z})$ riziková funkce pro jedince v čase $t$ v závislosti na proměnných $\mathbf{Z}$

$\lambda_0(t)$ základní riziková funkce

$\beta^T = (\beta_1, \ldots, \beta_k)$ vektor parametrů

$$\frac{\lambda(t|\mathbf{Z})}{\lambda(t|\mathbf{Z}^*)} = \frac{\lambda_0(t)\exp\left(\sum_{k=1}^{p}\beta_k Z_k\right)}{\lambda_0(t)\exp\left(\sum_{k=1}^{p}\beta_k Z_k^*\right)} = \exp\left(\sum_{k=1}^{p}\beta_k(Z_k - Z_k^*)\right)$$

např. $Z_1$ léčebný efekt ($Z_1 = 1$ pacient je léčen, $Z_1 = 0$ použito placebo), $\frac{\lambda(t|\mathbf{Z})}{\lambda(t|\mathbf{Z}^*)} = \exp(\beta_1)$

# Kódování proměnných

- **Dichotomická proměnná** např. věk $Z = 1$ pro muže, $Z = 0$ pro ženy,
- **Kvalitativní proměnné - faktory** $n$ skupin kódujeme pomocí $n$ proměnných (přeparametrizovaný model) nebo pomocí $n - 1$ proměnných,
  např. barva vlasů $Z_1 = 1$ pro blonďaté, 0 jinak, $Z_2 = 1$ pro černé, 0 jinak, *($Z_3 = 1$ pro hnědé, 0 jinak)*
- **Spojité proměnné** např. věk $Z$
- **Interakce faktorů**
  např. pohlaví a barva pleti: $Z_1 = 1$ černý muž, 0 jinak, $Z_2 = 1$ bílý muž, 0 jinak, $Z_3 = 1$ černá žena, 0 jinak,
  nebo $Z_1 = 1$ muž, 0 jinak, $Z_2 = 1$ černý, 0 jinak,
  $Z_3 = Z_1 \times Z_2 = 1$ černý muž, 0 jinak

# Odhad a testování parametrů

Metoda maximální věrohodnosti

$$\frac{\partial L}{\partial L \beta_k} = 0, k = 1, \dots, p$$

Řešení se obvykle provádí numericky (Newton-Raphsonova či jiné iterační metody)

Testování hypotézy

$$H_0 : \beta_1 = \beta_{10}, \dots, \beta_q = \beta_{q0}$$

- Waldův test
- Test věrohodnostním poměrem
- Skórový test

Za platnosti nulové hypotézy mají statistiky těchto testů rozdělení $\chi^2$ s $q$ stupni volnosti

# Výstavba modelu

Věrohodnostní poměr $LR = -2 \log L$, $L$ je hodnota věrohodnostní funkce pro odhadnuté parametry
Akeikeho informační kritérium $AIC = -2 \log L + kp$, $p$ je počet regresních koeficientů, $k$ je nějaká konstanta (většinou 2)
Schwarzovo informační kritérium $SBC = -2 \log L + k \log n$

- krokový výstavbový princip dopředu
- krokový výstavbový princip dozadu
- krokový výstavbový princip kombinovaný

# Odhad funkce přežití

Coxova regrese poskytuje odhad rizikové funkce

$$\hat{\lambda}(t|\mathbf{Z}) = \widehat{\lambda}_0(t) \exp\left(\sum_{k=1}^{p} \hat{\beta}_k Z_k\right)$$

Vztah mezi funkcí přežití a rizikovou funkcí
$S(t) = \exp(-\int_0^t \lambda(s)\mathrm{d}s)$

Odhad funkce přežití

$$\hat{S}(t|\mathbf{Z}) = \widehat{S}_0(t)^{\exp(\sum_{k=1}^{p} \hat{\beta}_k Z_k)}$$