ANÁLISE DE SOBREVIVÊNCIA - COX

Airlane P. Alencar – IME-USP

Alessandra C. Gourlart – FM-USP

Gisela Tunes da Silva – IME-USP

Taxa de falha e Curva de Sobrevida

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t}$$

$$S(t) = \exp\left[-\int_0^t h(u) du \right]$$

$$h(t) = -\left[\frac{dS(t)/dt}{S(t)} \right]$$

- h(t) mede o quanto varia S(t)

Modelo de Cox

- Modelo de riscos proporcionais
- O risco no tempo t com variável explicativa X é

$$h(t \mid X) = h_0(t) e^{\beta X}$$

- Exemplo para X=1 se fuma e 0 se não fuma

$$h(t \mid X) = \begin{cases} h_0(t), \ \mathrm{X} = 0 \\ h_0(t)\exp(\beta), \ \mathrm{X} = 1 \end{cases}$$

- Os riscos são proporcionais e a razão entre os riscos é

$$\frac{h(t \mid 1)}{h(t \mid 0)} = \exp(\beta)$$

- Se beta=1.2, exp(1,2)=3.32, então o risco de ...
- Interprete se X é quantitativa como IMC

Múltiplo

- Para p variáveis explicativas

$$h(t) = h_0(t) e^{\sum_{i=1}^{p} \beta_i X_i}$$

- As variáveis explicativas não dependem do tempo.
- $h_0(t)$ é o risco basal e sua forma não especificada => modelo semiparamétrico
- Interpretação dos parâmetros: ceteris paribus

Exemplo – Leucemia (KK)

- Leucemia

```
d <-
read.csv("https://www.ime.usp.br/~lane/home/MCM5916/anders
on.csv", sep=";")

f<- coxph(formula = Surv(t, obito) ~ Grupo, data = d, method =
"breslow")
summary(f)
coxph(formula = Surv(t, obito) ~ Grupo*lwbc, data = d, method =
"breslow")
model2<-coxph(formula = Surv(t, obito) ~ Grupo + lwbc, data =
d, method = "breslow")
summary(model2)
```

Estimação

- A verossimilhança depende do risco basal $h_0(t)$.
- Verossimilhança Parcial – Cox (1975)
- Probabilidades condicionais
- P(i falhar no tempo t(i)| uma falha em t(i), história até t(i))=

$$= \frac{h(t_i \mid X_i)}{\sum_{j \in R(t_i)} h(t_j \mid X_j)} = \frac{h_0(t_i)\exp^{x_i^T \beta}}{\sum_{j \in R(t_i)} h_0(t_j)\exp^{x_j^T \beta}} = \frac{\exp^{x_i^T \beta}}{\sum_{j \in R(t_i)} \exp^{x_j^T \beta}}$$

t_i são os tempos de falha e $R(t_i)$ são os em risco no tempo t_i

Exemplo Kleinbaum e Klein

	t	falha	fuma
Barry	2	1	1
Gary	3	1	0
Harry	5	0	0
Larry	8	1	1

- Risco = $h(t \mid X) = h_0(t)e^{\beta X}$

$$L(\beta) = \frac{h_0(t)e^{\beta}}{h_0(t)e^{\beta} + h_0(t)e^{0} + h_0(t)e^{0} + h_0(t)e^{\beta}} \times$$

$$\times \frac{h_0(t)e^{0}}{h_0(t)e^{0} + h_0(t)e^{0} + h_0(t)e^{\beta}} \times \frac{h_0(t)e^{\beta}}{h_0(t)e^{\beta}}$$

Empates e Distribuição

- Detalhes em Colosimo e Giolo pg. 162

- Assume-se que os tempos são contínuos e não há empates.
- Convenção: Censuras ocorrem após as falhas para definir as observações em risco
- Verossimilhança aprox. Breslow e Peto

- Os estimadores são assintoticamente normais e consistentes sob condições de regularidade
- Testes assintóticos como de Wald e de Razão de Verossimilhanças valem com dist. Qui quadrado sob H0.

Taxa de falha e Curva de Sobrevida

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \le T < t + \Delta t | T \ge t)}{\Delta t}$$

$$S(t) = \exp\left[-\int_0^t h(u) du \right]$$

$$h(t) = -\left[\frac{dS(t)/dt}{S(t)} \right]$$

- h(t) mede o quanto varia S(t)

Estimação da função de risco basal

- Para estimar S(t) temos que estimar H_0(t).

$$S(t) = \exp[-H(t)] = \exp\left[-\int_0^u h_0(u)\exp(x'\beta)du\right] =$$

$$= \exp\left[-\int_0^u h_0(u)du \cdot \exp(x'\beta)\right] =$$

$$= \exp[-H_0(t)]^{\exp(x'\beta)} = S_0(t)^{\exp(x'\beta)}$$

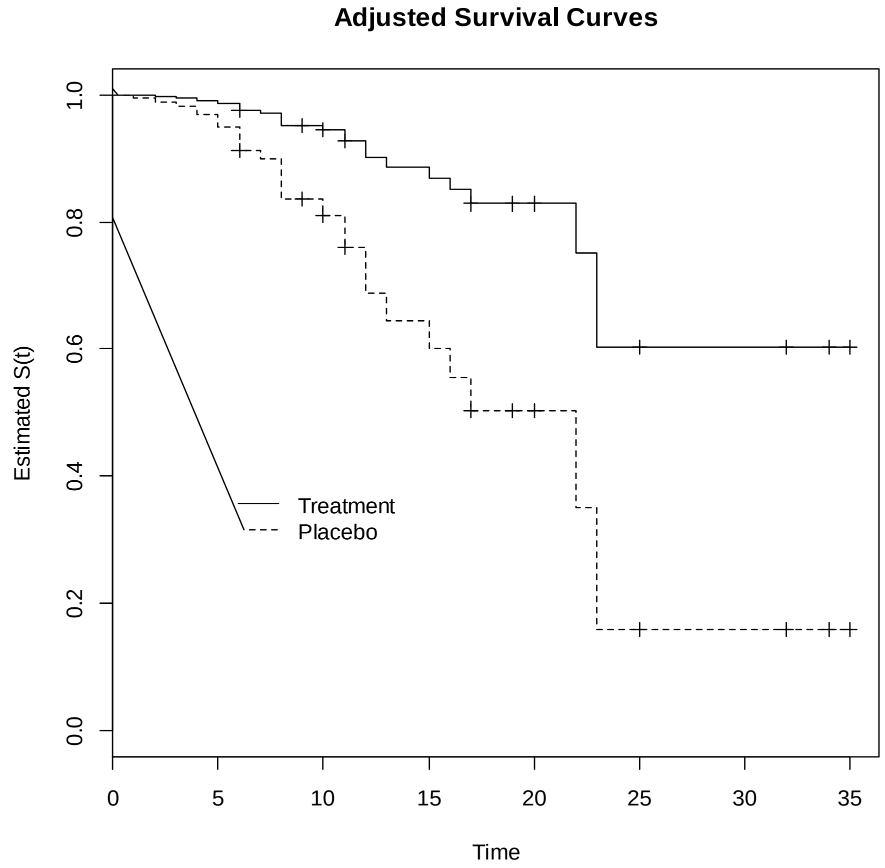
- O estimador de Breslow(1972) é $\hat{H}_0(t) = \sum_{j:t_j<t} \frac{m_j}{\sum_{l\in R(t_j)} \exp(x_l'\hat{\beta})}$

Exemplo

```
coxph(formula = Surv(t, obito) ~ Grupo, data = d, method = "breslow")


      coef exp(coef) se(coef)    z       p
Grupo 1.51      4.52     0.41 3.68 0.00023

Likelihood ratio test=15.2  on 1 df, p=9.61e-05  n= 42, number of events= 30
> coxph(formula = Surv(t, obito) ~ Grupo*lwbc, data = d, method = "breslow")
Call:
coxph(formula = Surv(t, obito) ~ Grupo * lwbc, data = d, method = "breslow")


             coef exp(coef) se(coef)      z     p
Grupo       2.355     10.54    1.681  1.401 0.160
lwbc        2.145      8.54    0.888  2.416 0.016
Grupo:lwbc -0.342      0.71    0.520 -0.658 0.510

Likelihood ratio test=43.8  on 3 df, p=1.63e-09  n= 42, number of events= 30
```

Probabilidades de sobrevida ajustadas

- Fixando lwbc= média de lwbc



Checando a proporcionalidade dos riscos

- Resíduos de Schoenfeld
- Para o índivíduo i (falha) e variável explicativa q= 1,..., p:

$$r_{iq} = x_{iq} - \frac{\sum_{j \in R(t_i)} x_{jq} \exp\left(x'_j \hat{\beta}\right)}{\sum_{j \in R(t_i)} \exp\left(x'_j \hat{\beta}\right)}$$

- Esses resíduos são padronizados, s^*_{iq}.
- $s^*_{iq} + \hat{\beta}_q$ mede o efeito da q-ésima explicativa no risco.
- O gráfico dessa quantidade em função do tempo (ou de g(t)) deve ser constante para termos riscos proporcionais.
- Nesse gráfico é incluida uma curva suavizada e bandas de confiança.

Detalhes em Colosimo e Giolo

Teste de proporcionalidade de riscos

- Um teste para a hipótese de proporcionalidade dos riscos para todas as covariáveis com g(t) usa

$$T = \frac{(g-\bar{g})' S^* I (S^*)' (g-\bar{g})}{d \sum_k (g-\bar{g})^2}$$

sendo d= número de falhas, $S^*=dRI^{-1}$, R= resíduos de Schoenfeld não padronizados e I a matriz informação observada.

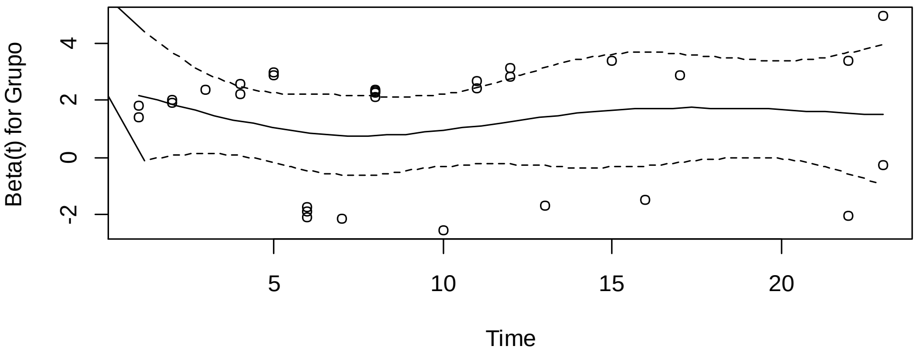
- Sob H0 (proporcionalidade), T ~ χ^2_p assintoticamente
- Para cada covariável x_q, temos

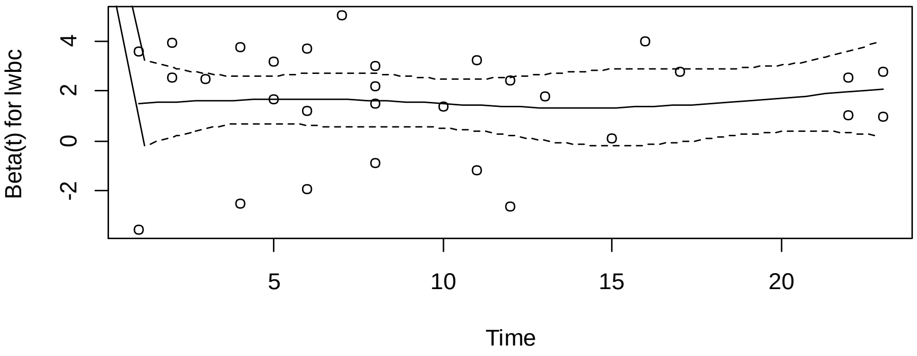
$$T_q = \frac{d\left(\sum_k (g_k - \bar{g}) s^*_{qk}\right)^2}{I_q^{-1} \sum_k (g-\bar{g})^2} \underset{n\to\infty}{\overset{H_0}{\sim}} \chi^2_1$$

Exemplo Leucemia

```
cox.zph(model2, transform="identity")
# default transforme=km
            rho    chisq      p
Grupo   0.0116 0.00339 0.954
lwbc    0.0391 0.06682 0.796
GLOBAL      NA 0.06893 0.966

Conclusão: … (para n grande)

par(mfrow=c(2,1))
plot(cox.zph(model2, transform="identity"))
```

Exemplo Leucemia





Referências

- Kleinbaum e Klein. Survival Analysis – a self learning text. Springer.

- Colosimo e Giolo. Análise de sobrevivência aplicada. Blucher.