Outline

Riscos Proporcionais

O modelo de regressão mais amplamente utilizado para dados de sobrevida ajusta a função de risco $\lambda(t)$, considerando um risco basal $\lambda_0(t)$ e incluindo o vetor de covariáveis $\boldsymbol{x}$, de forma que:

$$\lambda(t|\boldsymbol{x}) = \lambda_0(t)\exp(x_1\beta_1 + x_2\beta_2 + \cdots + x_p\beta_p) = \lambda_0(t)\exp(\boldsymbol{x}\boldsymbol{\beta})$$

Ou seja, as covariáveis têm um efeito multiplicativo na função de risco.

Riscos Proporcionais

A razão entre os riscos de ocorrência do evento de dois indivíduos i e j, com covariáveis $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \cdots, x_{ip})$ e $\boldsymbol{x}_j = (x_{j1}, x_{j2}, \cdots, x_{jp})$ é:

$$\frac{\lambda_i(t|\boldsymbol{x}_i)}{\lambda_j(t|\boldsymbol{x}_j)} = \frac{\exp(\boldsymbol{x}_i\boldsymbol{\beta})}{\exp(\boldsymbol{x}_j\boldsymbol{\beta})}$$

Observe que esta razão de riscos NÃO varia ao longo do tempo –> Modelo de Riscos Porporcionais

Riscos Proporcionais

O modelo RP também pode ser escrito em termos da função de risco acumulado ou da função de sobrevida:

$$\Lambda(t|\boldsymbol{x}) = \Lambda_0(t)\exp(\boldsymbol{x\beta})$$
$$S(t|\boldsymbol{x}) = [S_0(t)]^{\exp(\boldsymbol{x\beta})}$$

O risco acumulado basal é $\Lambda_0(t) = \sum_{i:\, t_i \leq t} \frac{N_i(t)}{\sum_{j \in R(t_i)} \exp(\boldsymbol{x}_j \boldsymbol{\beta})}$ e a sobrevida basal é dada por $S_0(t) = \exp[-\Lambda_0(t)]$

Modelo de Cox

Partindo desta proporcionalidade, é possível estimar os efeitos das covariáveis sem qualquer suposição a respeito da distribuição do tempo de sobrevida, e por isso o modelo de Cox é dito semi-paramétrico.
Não se assume qualquer distribuição estatística para a função de risco basal, $\lambda_0(t)$, apenas que as covariáveis agem multiplicativamente sobre o risco e esta é a parte paramétrica do modelo.

Pressupostos do modelo de Cox

- As covariáveis agem multiplicativamente sobre o risco –> parte paramétrica do modelo.
- A razão de riscos é constante ao longo de tempo –> riscos proporcionais.
- Os tempos de ocorrência do evento são independentes.

Pressupostos do modelo de Cox

- As covariáveis agem multiplicativamente sobre o risco –> parte paramétrica do modelo.
- A razão de riscos é constante ao longo de tempo –> riscos proporcionais.
- Os tempos de ocorrência do evento são independentes.

Pressupostos do modelo de Cox

- As covariáveis agem multiplicativamente sobre o risco −> parte paramétrica do modelo.
- A razão de riscos é constante ao longo de tempo −> riscos proporcionais.
- Os tempos de ocorrência do evento são independentes.

Pressupostos do modelo de Cox

- As covariáveis agem multiplicativamente sobre o risco −> parte paramétrica do modelo.
- A razão de riscos é constante ao longo de tempo −> riscos proporcionais.
- Os tempos de ocorrência do evento são independentes.

Estimativa dos coeficientes

Para estimar os coeficientes da regressão paramétrica, a função de verossimilhança foi construída a partir da função de densidade de probabilidade calculada nos tempos de ocorrência do evento, multiplicada pela função de sobrevida calculada nos tempos de censura.

No Modelo de Cox o vetor de parâmetros $\boldsymbol{\beta}$ é estimado a partir de uma verossimilhança parcial.
De forma semelhante ao Kaplan Meier, considera-se apenas, a cada tempo t, a informação dos indivíduos sob risco, estimando os efeitos das covariáveis no tempo de sobrevida.

Verossimilhança parcial

- Considere m diferentes tempos até a ocorrência de um evento (sem empate), ordenados assim: $t_1 < t_2 < \ldots < t_m$.
- A verossimilhança individual, L_i, é a razão entre o risco $\lambda_i(t_i)$ do indivíduo i falhar em t_i e a soma dos riscos de ocorrência de evento de todos os indivíduos em risco:

$$
\begin{aligned}
L_i &= \frac{\lambda_i(t_i)}{\sum_{j \in R(t_i)} \lambda_j(t_j)} \\
&= \frac{\exp(\boldsymbol{x}_i \boldsymbol{\beta})}{\sum_{j \in R(t_i)} \exp(\boldsymbol{x}_j \boldsymbol{\beta})}
\end{aligned}
$$

Verossimilhança parcial

- Sob o processo de contagem a verossimilhança individual é igual a

$$L_i = \frac{\exp(\boldsymbol{x}_i\boldsymbol{\beta})}{\sum_{t\geq 0} Y_j(t)\exp(\boldsymbol{x}_j\boldsymbol{\beta})},$$

- com $Y_j(t)$ igual a 1 se o indivíduo j estiver em risco no tempo t e 0, caso contrário.

Verossimilhança Parcial

- A verossimilhança parcial $L(\boldsymbol{\beta})$ = produto das L_i

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \prod_{t \geq 0} \left\{ \frac{Y_i(t) \exp(\boldsymbol{x}_i \boldsymbol{\beta})}{\sum_j Y_j(t) \exp(\boldsymbol{x}_j \boldsymbol{\beta})} \right\}^{dN_i(t)}$$

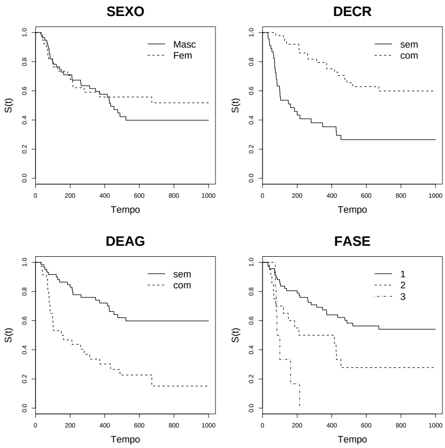
- $dN_i(t)$ = diferença entre a contagem de eventos até o instante t e a contagem no momento imediatamente anterior a t.
- Numerador depende apenas da informação dos indivíduos que experimentam o evento
- Denominador utiliza informações a respeito de todos os indivíduos que ainda não experimentaram o evento, incluindo aqueles que serão censurados mais tarde.

Exemplo TMO

- Avaliar os fatores prognósticos associados ao tempo de transplante de medula óssea TMO até o óbito nos pacientes com leucemia mielóide crônica tratados no INCA.
- covariáveis:
  - sexo,
  - idade,
  - fase da doença no momento do transplante (`fase`),
  - a ocorrência ou não de doença enxerto contra hospedeiro aguda (`deag`) ou crônica (`decr`).

Proporcionalidade

Curvas de KM para avaliar o pressuposto de proporcionalidade



No R

```
> tmocens <- read.table("tmoclas.dat", header=T, sep=",")
> mod1 <- coxph(Surv(os,status)~idade+factor(sexo),data=tmocens, x=TRUE)
> summary(mod1)

Call:
coxph(formula = Surv(os, status) ~ idade + factor(sexo), data = tmocens,
    x = TRUE)
  n= 96
                 coef exp(coef) se(coef)     z    p
idade         -0.0186     0.982   0.0141 -1.32 0.19
factor(sexo)2 -0.3299     0.719   0.3219 -1.02 0.31
              exp(coef) exp(-coef) lower .95 upper .95
idade             0.982       1.02     0.955      1.01
factor(sexo)2     0.719       1.39     0.383      1.35

Rsquare= 0.022   (max possible= 0.984 )
Likelihood ratio test= 2.16  on 2 df,   p=0.34
Wald test            = 2.11  on 2 df,   p=0.348
Score (logrank) test = 2.11  on 2 df,   p=0.348
```

Selecionando modelos

- Teste de Wald
- Análise da função desvio

Comparando quatro modelos

```
> anova(mod1,mod2,mod3,mod4,test='Chisq')

Analysis of Deviance Table

Model 1: Surv(os, status) ~ idade + factor(sexo)
Model 2: Surv(os, status) ~ idade + factor(sexo) + factor(fase)
Model 3: Surv(os, status) ~ idade + factor(sexo) + factor(fase) + deag
Model 4: Surv(os, status) ~ idade + factor(sexo) + factor(fase) + deag +
    decr
  Resid. Df Resid. Dev Df Deviance P(>|Chi|)
1         94     395.93
2         92     380.78  2    15.14 0.0005146
3         91     366.67  1    14.11 0.0001726
4         90     358.20  1     8.47 0.0036015
```

Selecionando Modelos

- A função desvio é assintoticamente semelhante à estatística de Wald quando o número de observações é grande.
- Para número de observações pequenos, a análise da função desvio é mais robusta.
- Outra ressalva a respeito de valores ausentes. Caso eles existam para algumas variáveis incluídas em alguns modelos, mesmo que aninhados, os modelos perdem a comparabilidade.

Medida Global de Ajuste

- R^2 – poder explicativo das covariáveis no tempo de ocorrência do evento em estudo.

$$
\begin{aligned}
R^2_{LR} &= 1 - \{L(0)/L(\hat{\boldsymbol{\beta}})\}^{2/n} \\
&= 1 - \exp(2\{l(0) - l(\hat{\boldsymbol{\beta}})\}/n)
\end{aligned}
$$

- Valor mínimo possível de R^2 é zero quando $L(0) = L(\hat{\boldsymbol{\beta}})$
- Valor máximo não é 1 (ou 100%), mas a razão entre as verossimilhanças do modelo saturado e do modelo nulo.

Medida Global de Ajuste

Modelo	$\ln$(Verossimil.)	R^2	% Var. Explicada*
Nulo	-199,0424	0,000	0,0%
Saturado	-0,2670	0,984	100,0%
M1: Idade+Sexo	-197,9626	0,022	2,2%
M2: Mod1+Fase	-190,3905	0,165	16,8%
M3: Mod2+*deag*	-183,3364	0,279	28,4%
M4: Mod3+*decr*	-179,0992	0,340	34,6%

* $R^2_{modelo}/R^2_{saturado}$

Medida Global de Ajuste

Gráfico de sobrevida estratificado por índice de prognóstico (IP)

- IP é o preditor linear do modelo de Cox, $\boldsymbol{x\beta}$, calculado para cada indivíduo usando as covariáveis observadas e as estimativas dos coeficientes de regressão do modelo ajustado.
- Os indivíduos são estratificados em grupos de tamanhos aproximadamente iguais (grupos de alto, médio e baixo IP)
- Os valores médios de cada uma das covariáveis dentro de cada grupo são utilizados para obtenção de curvas de sobrevida sob o modelo ajustado.
- Espera-se, se o modelo for razoável, que o gráfico das curvas ajustadas pelo modelo em cada estrato sejam próximas das estimadas por Kaplan-Meier.

Medida Global de Ajuste

- Assumindo modelo *mod4*
- Indivíduo 1: sexo masculino (`sexo = 0`) com 56 anos (`idade = 56`), na fase intermediária (`fase2 = 1` e `fase3 = 0`), com manifestação de doença do enxerto aguda (`deag=1`, `decr=0`)

$$
\begin{array}{llll}
\beta_{idade} & \times 56 = -0,0044 \times 56 & = -0,2469 \\
\beta_{sexo} & \times 0 = -0,2260 \times 0 & = 0 \\
\beta_{fase2} & \times 1 = 0,6413 \quad \times 1 & = 0,6413 \\
\beta_{fase3} & \times 0 = 1,0279 \quad \times 0 & = 0 \\
\beta_{deag} & \times 1 = 1,2530 \quad \times 1 & = 1,2530 \\
\beta_{decr} & \times 0 = -0,9775 \times 0 & = 0
\end{array}
$$

Soma $= 1,6474$

Medida Global de Ajuste

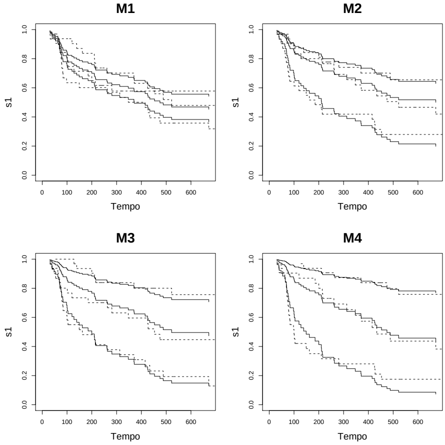
- Assumindo modelo *mod4*
- Indivíduo 2: sexo feminino (`sexo = 1`) com 20 anos (`idade = 20`), na fase avançada (`fase2 = 0` e `fase3 = 1`) com manifestação de doença do enxerto aguda (`deag=1`, `decr=0`)

$$
\begin{array}{llllll}
\beta_{idade} & \times 20 & = -0,0044 & \times 20 & = -0,0882 \\
\beta_{sexo} & \times 1 & = -0,2260 & \times 1 & = -0,2260 \\
\beta_{fase2} & \times 0 & = 0,6413 & \times 0 & = 0 \\
\beta_{fase3} & \times 1 & = 1,0279 & \times 1 & = 1,0279 \\
\beta_{deag} & \times 1 & = 1,2530 & \times 1 & = 1,2530 \\
\beta_{decr} & \times 0 & = -0,9775 & \times 0 & = 0
\end{array}
$$

Soma $= 1,9667$

Medida Global de Ajuste

Gráfico de sobrevida estratificado por índice de prognóstico.



Linha sólida representa o modelo ajustado e linha pontilhada a estimativa de Kaplan-Meier.