

EPI10 - Análise de Sobrevida

Modelos de Cox estratificado e com covariáveis dependentes do tempo

Rodrigo Citton P. dos Reis
citton.padilha@ufrgs.br

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE MEDICINA
PROGRAMA DE PÓS-GRADUAÇÃO EM EPIDEMIOLOGIA

Porto Alegre, 2021

Introdução

Introdução

- ▶ Algumas situações práticas envolvendo medidas longitudinais não são ajustadas adequadamente usando o modelo de Cox na sua forma original como apresentado nas últimas aulas.
- ▶ Existem **covariáveis** que são monitoradas durante o estudo, e seus valores **podem mudar** ao longo desse período.
 - ▶ Por exemplo, pacientes podem mudar de grupo durante o tratamento ou a dose de quimioterapia aplicada em pacientes com câncer pode sofrer alterações durante o curso do tratamento.

Introdução

- ▶ Se estes valores forem incorporados na análise estatística, resultados mais precisos podem ser obtidos comparados àqueles em que utilizam somente as mesmas medidas registradas no início do estudo.
 - ▶ Em outros exemplos, a não inclusão destes valores pode acarretar sérios vícios.
- ▶ Este tipo de **covariável** é chamada de **dependente do tempo** e o **modelo de Cox** pode ser **estendido** para incorporar as informações longitudinais registradas para esta variável.

Introdução

- ▶ Em outras situações a **suposição de riscos proporcionais é violada** e o modelo de Cox não é adequado.
- ▶ Modelos alternativos existem para enfrentar esta situação.
- ▶ Um deles é uma extensão do próprio modelo de Cox chamado de **modelo de riscos proporcionais estratificado**.
 - ▶ Neste caso supõe-se que os riscos proporcionais valem em cada estrato mas não valem entre estratos.

Modelo de Cox estratificado

Modelo de Cox estratificado

- ▶ Na aula 7 foram apresentadas técnicas estatísticas para avaliar a adequação do modelo de Cox.
- ▶ Essencialmente, estas técnicas avaliam a **suposição de riscos proporcionais**.
- ▶ O modelo

$$\lambda(t) = \lambda_0(t) \exp\{\beta_1 x_1 + \dots + \beta_p x_p\}$$

não pode ser usado se esta suposição for violada.

Modelo de Cox estratificado

- ▶ Nestes casos, uma solução para o problema é **estratificar os dados** de modo que a suposição seja válida em cada estrato.
- ▶ Por exemplo, os riscos podem não ser proporcionais entre homens e mulheres, mas esta suposição pode valer no estrato formado somente por homens e naquele formado somente por mulheres.

Modelo de Cox estratificado

- ▶ A análise estratificada consiste em dividir os dados de sobrevivência em m estratos, de acordo com uma indicação de violação da suposição.
- ▶ O modelo de riscos proporcionais é então expresso como

$$\lambda_{ij}(t) = \lambda_{0j}(t) \exp\{\beta_1 x_{1ij} + \dots + \beta_p x_{pij}\}$$

para $i = 1, \dots, m$ e $j = 1, \dots, n_j$, em que n_j é o número de observações no j -ésimo estrato.

- ▶ As funções de base $\lambda_{01}(t), \dots, \lambda_{0m}(t)$, são arbitrárias e completamente não relacionadas.

Modelo de Cox estratificado

- ▶ A estratificação não cria nenhuma complicação na estimação do vetor de parâmetros β .
- ▶ Uma **função de verossimilhança parcial**, como a apresentada na aula 4, é construída para **cada estrato** e a estimação dos coeficientes de regressão β é baseada na soma dos logaritmos das funções de verossimilhança parciais:

$$\ell(\beta) = \ell_1(\beta) + \dots + \ell_m(\beta),$$

em que $\ell_j(\beta) = \log[L_j(\beta)]$ obtida usando-se somente os dados dos indivíduos no j -ésimo estrato.

- ▶ As propriedades assintóticas dos estimadores são obtidas a partir dos estimadores do modelo não estratificado¹.

¹Colosimo, E. A. (1997). A note on the stratified proportional hazards model. *International Journal of Math. Statist. Sciences*, 6, 201-209.

Modelo de Cox estratificado

- ▶ Note que o modelo de Cox estratificado assume que as covariáveis atuam de modo similar na função de risco de base de cada estrato.
 - ▶ Ou seja, β é assumido ser comum para todos os estratos.
- ▶ Esta suposição pode ser testada, por exemplo, o teste da razão de verossimilhanças, cuja estatística de teste é dada, nesse caso, por:

$$TRV = -2 \left[\ell(\hat{\beta}) - \sum_{j=1}^m \ell_j(\hat{\beta}_j) \right],$$

em que $\ell(\hat{\beta})$ é o logaritmo da função de verossimilhança parcial sob o modelo que assume β 's comuns em cada estrato e $\sum_{j=1}^m \ell_j(\hat{\beta}_j)$, o logaritmo da função de verossimilhança parcial sob o modelo que assume β 's distintos em cada estrato.

Modelo de Cox estratificado

- ▶ Sob a hipótese nula e para grandes amostras, a estatística TRV segue uma distribuição qui-quadrado com $(m - 1)p$ graus de liberdade em que m é o número de estratos e p a dimensão do vetor β .
- ▶ O modelo estratificado deve ser usado somente caso realmente necessário.
 - ▶ Ou seja, quando a suposição de riscos proporcionais é violada no modelo de Cox usual.
- ▶ O uso desnecessário da estratificação acarreta em uma perda de eficiência das estimativas obtidas.

Exemplo

Estudo de Leucemia Pediátrica

- ▶ 103 crianças com leucemia pediátrica.
- ▶ O desfecho de interesse é o tempo a partir da remissão (ausência da doença) até a recidiva ou óbito (o que ocorrer primeiro).

Exemplo

Tabela 5.8: Descrição das covariáveis utilizadas no estudo sobre leucemia pediátrica.

Código	Descrição	Categorias
LEUINI	Número de leucócitos no sangue periférico	0 se ≤ 75000 leucócitos/mm ³ 1 se > 75000 leucócitos/mm ³
IDADE	Idade em meses	0 se ≤ 96 meses 1 se > 96 meses
ZPESO	Peso padronizado pela idade e sexo	0 se ≤ -2 e 1 se > -2
ZEST	Altura padronizada pela idade e sexo	0 se ≤ -2 e 1 se > -2
PAS	Porcentagem de linfoblastos medulares que reagiram positivamente ao ácido periódico de Schiff	0 se $\leq 5\%$ e 1 se $> 5\%$
VAC	Porcentagem de vacúolos no citoplasma dos linfoblastos	0 se $\leq 15\%$ e 1 se $> 15\%$
RISK	Fator de risco obtido a partir de uma fórmula que é função dos tamanhos do fígado e do baço e do número de blastos	0 se $\leq 1,7\%$ e 1 se $> 1,7\%$
R6	Remissão na sexta semana de tratamento	0 se não e 1 se sim

Exemplo

```
df.leucemia <- read.table(file = "../dados/leucemia.txt",  
                           header = TRUE)
```

```
head(df.leucemia)
```

##	leuini	tempos	cens	idade	zpeso	zest	pas	vac	risk	r6
## 1	380.0	1.760	1	60.52	-0.97	-0.48	0.1	5.7	1.58	1
## 2	328.0	0.260	1	68.04	0.36	1.44	0.6	1.5	1.64	0
## 3	84.7	0.129	1	159.93	-1.84	-2.17	0.6	20.4	1.26	1
## 4	2.9	3.639	1	92.91	-1.06	-0.69	0.7	1.5	0.96	1
## 5	400.0	4.331	0	156.98	-0.84	-0.82	13.7	1.0	1.32	1
## 6	64.0	4.252	0	69.62	-0.20	-0.19	2.3	2.0	1.40	1

Exemplo

```
df.leucemia$idadec <- ifelse(df.leucemia$idade > 96,1,0)
df.leucemia$leuinic <- ifelse(df.leucemia$leuini > 75,1,0)
df.leucemia$zpesoc <- ifelse(df.leucemia$zpeso > -2,1,0)
df.leucemia$zestc <- ifelse(df.leucemia$zest > -2,1,0)
df.leucemia$pasc <- ifelse(df.leucemia$pas > 0.05,1,0)
df.leucemia$ vacc <- ifelse(df.leucemia$vac > 15,1,0)
df.leucemia$pasc <- ifelse(df.leucemia$pas > 5,1,0)
df.leucemia$riskc <- ifelse(df.leucemia$risk > 1.7,1,0)
```


Exemplo

```
library(survival)

mod1 <- coxph(Surv(time = tempos, event = cens) ~ idadec +
              zpesoc + pasc + vacc + strata(leuinic),
              data = df.leucemia, method = "breslow")

summary(mod1)
```

Call:

```
## coxph(formula = Surv(time = tempos, event = cens) ~ idadec +
##       zpesoc + pasc + vacc + strata(leuinic), data = df.leucemia,
##       method = "breslow")
##
## n= 103, number of events= 39
##
##          coef exp(coef) se(coef)      z Pr(>|z|)
## idadec  0.7993   2.2240   0.3836  2.084  0.03716 *
## zpesoc  -2.4091   0.0899   0.5210 -4.624  3.77e-06 ***
## pasc    -1.2490   0.2868   0.4649 -2.686  0.00723 **
## vacc     1.3589   3.8920   0.4194  3.240  0.00119 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Exemplo

```
##
##      exp(coef) exp(-coef) lower .95 upper .95
## idadec      2.2240      0.4496   1.04874   4.7165
## zpesoc      0.0899     11.1239   0.03238   0.2496
## pasc        0.2868      3.4868   0.11529   0.7134
## vacc        3.8920      0.2569   1.71085   8.8538
##
## Concordance= 0.713  (se = 0.049 )
## Likelihood ratio test= 28.67  on 4 df,   p=9e-06
## Wald test              = 30.11  on 4 df,   p=5e-06
## Score (logrank) test = 36.69  on 4 df,   p=2e-07
```

Exemplo

```
# TRV
mod20 <- coxph(Surv(time = tempos, event = cens) ~ idadec +
               zpesoc + pasc + vacc,
               data = df.leucemia, subset = leuinic == 0,
               method = "breslow")

mod21 <- coxph(Surv(time = tempos, event = cens) ~ idadec +
               zpesoc + pasc + vacc,
               data = df.leucemia, subset = leuinic == 1,
               method = "breslow")

TRV <- as.numeric(-2 * (logLik(mod1) - (logLik(mod20) + logLik(mod21))))
TRV

## [1] 5.37051

1 - pchisq(TRV, 4)

## [1] 0.2513484

cox.zph(mod1, transform = "identity")
```

Exemplo

```
##           chisq df      p
## idadec 0.852963  1 0.356
## zpesoc 0.143580  1 0.705
## pasc  2.903555  1 0.088
## vacc  0.000494  1 0.982
## GLOBAL 4.249793  4 0.373

# plot(cox.zph(mod1))
```

Modelo de Cox com covariáveis dependentes do tempo

Covariáveis dependentes do tempo

- ▶ As covariáveis no modelo de Cox consideradas nas aulas 4-8, foram medidas no início do estudo ou na origem do tempo.
- ▶ Entretanto, existem covariáveis que são monitoradas durante o estudo e seus valores podem mudar ao longo do período de acompanhamento.

Covariáveis dependentes do tempo

Estudo de transplante de coração

- ▶ Um estudo bastante analisado na literatura é o do programa de transplante de coração de Stanford (Crowley e Hu, 1977).
- ▶ Neste estudo os pacientes eram aceitos no programa quando se tornavam candidatos a um transplante de coração.
- ▶ Quando surgia um doador, os médicos escolhiam, de acordo com alguns critérios, o candidato que iria receber o coração.
- ▶ Alguns pacientes morreram sem receber o transplante.
- ▶ A forma de alocação estava fortemente viciada na direção daqueles pacientes com maior tempo de sobrevivência pois somente estes pacientes viveram o suficiente para receber o (transplante de) coração.
- ▶ O uso de uma covariável assumindo o valor zero para aqueles esperando o transplante e um para aqueles com coração novo, serve para **minimizar esse vício** e ela muda de valor assim que o transplante é realizado e é, portanto, dependente do tempo.

Covariáveis dependentes do tempo

Exemplos:

- ▶ Medidas clínicas que são aferidas diversas vezes no mesmo indivíduo:
 - ▶ IMC, pressão arterial, etc;

intervalo (em anos)	IMC (kg/m ²)	status
(0,5]	21,7	0
(5,8]	23,7	0
(8,12]	26,9	0
(12,18]	30,4	0
(18,25]	32,6	0
(25,28]	35,8	1

Covariáveis dependentes do tempo

- ▶ Alteração de alguma condição clínica ou socioeconômica:
 - ▶ entrada no climatério;
 - ▶ promoção no emprego;
 - ▶ mudança de endereço/vizinhança.

Covariáveis dependentes do tempo

- ▶ Realização de algum procedimento ou mudança de tratamento durante o período de observação:
 - ▶ início de hemodiálise em pacientes com doença renal.

paciente	intervalo	hemodiálise	status
1	(0,30]	0	0
1	(30,90]	1	1
2	(20,150]	0	0
2	(150,300]	1	0
3	(15,400]	0	1

Covariáveis dependentes do tempo

- ▶ Temos ainda as situações em que se deseja estudar o efeito de múltiplas covariáveis mudando no tempo.
- ▶ Também nesses casos precisamos apenas construir um objeto/arquivo/banco de dados adequado para que as mudanças nas diversas combinações de atributos ao longo do tempo possam ser incorporadas corretamente no modelo de Cox.

Covariáveis dependentes do tempo

- ▶ Tais variáveis, quando presentes em um estudo, podem ser incorporadas ao modelo de regressão de Cox, generalizando-o como:

$$\lambda(t) = \lambda_0(t) \exp\{\beta_1 x_1(t) + \dots + \beta_p x_p(t)\}.$$

- ▶ Definido desta forma, o modelo acima já não mais possui a propriedade de riscos proporcionais, pois a razão das funções de risco no tempo t para dois indivíduos i e j fica

$$\frac{\lambda_i(t)}{\lambda_j(t)} = \exp\{x'_i(t)\beta - x'_j(t)\beta\},$$

que é dependente do tempo.

Estudos de Transplante de medula óssea

- ▶ Estes dados provêm de uma coorte de 96 pacientes submetidos a transplante de medula óssea (TMO) para tratamento de leucemia mieloide crônica, no período de junho de 1986 a junho de 1998, no Centro de Transplante de Medula Óssea do Instituto Nacional do Câncer (Cemo - Inca).
- ▶ O acompanhamento dessa coorte possibilitou o estudo do efeito de fatores prognósticos para a ocorrência de doença do enxerto contra hospedeiro aguda e crônica, da sobrevivência livre de doença e da sobrevivência global.

Estudos de Transplante de medula óssea

Variável	Descrição
id	identificador do paciente
sexo	1 = masculino, 2 = feminino
idade	idade na data do transplante (5 a 53 anos)
status	0 = censura, 1 = óbito
os	data do óbito ou censura
plaq	recuperação de plaquetas: 0 = não, 1 = sim
tempplaq	dias até a recuperação de plaquetas
deag	doença enxerto aguda: 0 = não, 1 = sim
tempdeag	dias até a ocorrência da doença enxerto aguda
decr	doença enxerto crônica: 0 = não, 1 = sim
tempdecr	tempo até a ocorrência da doença enxerto crônica
fase	fase da doença na data do transplante: 1 = crônica, 2 = aguda, 3 = crise blástica
recplaq	0 = plaquetas não recuperadas, 1 = recuperadas
fasegr	fase da doença na data do transplante agrupada: CP1 = 1a crônica e Other = outras
inicio	data do transplante ou da mudança de covariável
fim	data de mudança de covariável ou do fim do estudo

Estrutura dos dados mudando no tempo

- ▶ No estudo da sobrevivência de pacientes submetidos ao TMO consideramos, além das variáveis medidas somente no início do estudo, **as que mudam ao longo do tempo**, como:
 - ▶ a ocorrência de doença do enxerto contra o hospedeiro (DECH) em suas formas aguda (deag) e crônica (decr);
 - ▶ a recuperação no número de plaquetas a um nível acima de $20.000/\text{mm}^3$ ('recplaq').
- ▶ É necessário construir um banco de dados contemplando o momento de mudança de cada uma dessas três características.

Estrutura dos dados mudando no tempo

```
df.tmo <- read.table(file = here::here("dados","tmopc.csv"),
                     header = TRUE, sep = ";")
```

```
head(df.tmo, 14)
```

##	id	sexo	idade	status	inicio	fim	deag	decr	recplaq	fasegr
## 1	1	2	31	0	0	9	0	0	0	CP1
## 2	1	2	31	0	9	1000	0	0	1	CP1
## 3	2	2	38	0	0	28	0	0	0	CP1
## 4	2	2	38	1	28	39	1	0	0	CP1
## 5	3	1	23	0	0	27	0	0	0	CP1
## 6	3	1	23	0	27	36	0	0	1	CP1
## 7	3	1	23	0	36	268	1	0	1	CP1
## 8	3	1	23	1	268	434	1	1	1	CP1
## 9	4	2	5	0	0	24	0	0	0	CP1
## 10	4	2	5	1	24	69	1	0	0	CP1
## 11	5	2	15	0	0	22	0	0	0	CP1
## 12	5	2	15	0	22	83	1	0	0	CP1
## 13	5	2	15	0	83	446	1	0	1	CP1
## 14	5	2	15	1	446	672	1	1	1	CP1

Estrutura dos dados mudando no tempo

Observamos que cinco pacientes constituem 14 linhas de registro no banco de dados.

▶ **Paciente número 1 (2 linhas):**

- ▶ recuperou as plaquetas $> 20.000/\text{mm}^3$;
- ▶ não apresentou DECH aguda ou crônica;
- ▶ não faleceu até o fim do acompanhamento.

▶ **Paciente número 3 (4 linhas):**

- ▶ recuperou as plaquetas $> 20.000/\text{mm}^3$ (dia 27);
- ▶ apresentou DECH aguda (dia 36);
- ▶ apresentou DECH crônica (dia 268);
- ▶ faleceu no dia 434.

Estrutura dos dados mudando no tempo

- Observe que nos tempos (início e fim), a variável indicadora de censura (status) e as variáveis que sofrem alteração ao longo do estudo mudam de uma linha para outra (deag, decr e recplaq);
- Os valores das variáveis medidas no início do estudo (idade e fasegr) são repetidos em todas as linhas referentes a um mesmo indivíduo.

Estrutura dos dados mudando no tempo

- O tempo até a ocorrência do evento pode ser representado da seguinte forma:

paciente 1 $(0,9+]$ $(9,1000+]$

paciente 2 $(0,28+]$ $(28,39]$

paciente 3 $(0,27+]$ $(27,36+]$ $(36,268+]$ $(268,434]$

paciente 4 $(0,24+]$ $(24,69]$

paciente 5 $(0,22+]$ $(22,83+]$ $(83,446+]$ $(446,672]$

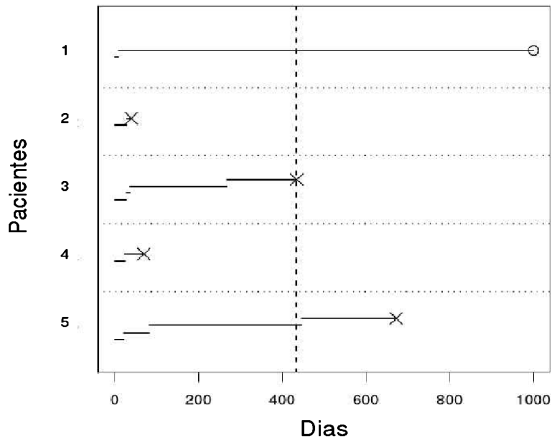
Estrutura dos dados mudando no tempo

- ▶ Uma preocupação que pode surgir nesse caso (em que múltiplas linhas referem-se a um mesmo paciente) é que estas são **correlacionadas** e, portanto, os procedimentos usuais de estimação e inferência sobre parâmetros do modelo não poderiam ser utilizados.
- ▶ No entanto, como os **intervalos de tempo** definidos nas diferentes linhas de cada paciente são **disjuntos**, os resultados do modelo ajustado são inteiramente válidos.

Estrutura dos dados mudando no tempo

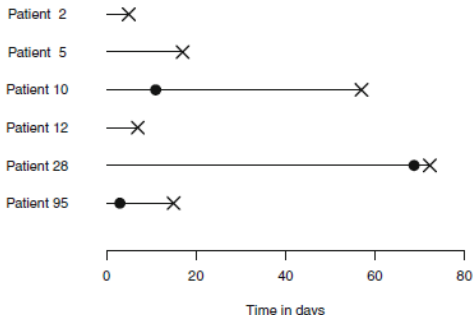
- ▶ Isso porque o cálculo da verossimilhança parcial, que envolve a soma das observações de indivíduos sob risco num dado momento em que ocorre um evento, utilizará no máximo uma observação de cada paciente em qualquer momento.
- ▶ Ou seja, a soma ainda é feita sobre um conjunto de **observações independentes**.

Estrutura dos dados mudando no tempo



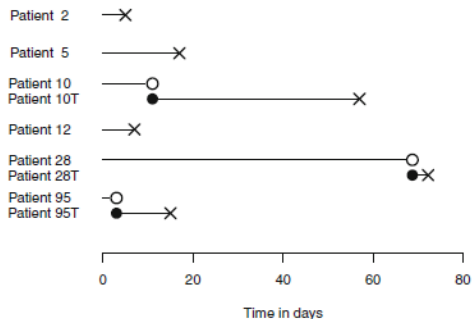
Estrutura dos dados mudando no tempo

Fig. 8.1 Sample of six patients from the Stanford heart transplant data set. In this plot, death is denoted by an “X”, and the time of transplant (for Patients 10, 28, and 95) by a *solid dot*. In the plot on the right, the timelines of patients who received a transplant are split into pre- and post-transplant components



Estrutura dos dados mudando no tempo

Fig. 8.2 Plot of sample of heart transplant patients in start-stop counting process format



Estrutura dos dados mudando no tempo

- ▶ A construção de bancos de dados com múltiplas linhas por indivíduo é simples quando o número de covariáveis variando no tempo é pequeno;
 - ▶ mas pode tornar-se bastante complexa e sujeita a erros à medida que o número de covariáveis aumenta.
- ▶ Ver **criabanco.pdf** e **tmerge.pdf** (*Moodle*).

Estrutura dos dados mudando no tempo

► Retornando ao exemplo.

```
df.tmo$sexo <- factor(df.tmo$sexo,  
                      labels = c("masculino", "feminino"))  
df.tmo$deag <- factor(df.tmo$deag,  
                      labels = c("não", "sim"))  
df.tmo$decr <- factor(df.tmo$decr,  
                      labels = c("não", "sim"))  
df.tmo$recplaq <- factor(df.tmo$recplaq,  
                         labels = c("nao rec", "rec"))  
df.tmo$recplaq <- relevel(df.tmo$recplaq, ref = "rec")
```

Estrutura dos dados mudando no tempo

```
library(survival)

mod1 <- coxph(
  Surv(time = inicio, time2 = fim , event = status) ~ idade +
    sexo + fasegr + deag + decr + recplaq,
  data = df.tmo, method = "breslow")

summary(mod1)
```

```
## Call:
## coxph(formula = Surv(time = inicio, time2 = fim, event = status) ~
##      idade + sexo + fasegr + deag + decr + recplaq, data = df.tmo,
##      method = "breslow")
##
##      n= 259, number of events= 53
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## idade          -0.02069   0.97952  0.01402 -1.475 0.140128
## sexofeminino    -0.17604   0.83859  0.30934 -0.569 0.569306
## fasegr0ther      0.92768   2.52864  0.31082  2.985 0.002839 **
## deagsim          1.05480   2.87139  0.29171  3.616 0.000299 ***
## decrsim          0.43741   1.54868  0.38586  1.134 0.256970
```

Estrutura dos dados mudando no tempo

```
## recplaqnao rec  1.96214    7.11456  0.46714  4.200 2.67e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##               exp(coef) exp(-coef) lower .95 upper .95
## idade                0.9795      1.0209    0.9530    1.007
## sexofeminino          0.8386      1.1925    0.4573    1.538
## fasegrOther           2.5286      0.3955    1.3750    4.650
## deagsim               2.8714      0.3483    1.6210    5.086
## decrsim               1.5487      0.6457    0.7270    3.299
## recplaqnao rec       7.1146      0.1406    2.8478   17.774
##
## Concordance= 0.752  (se = 0.032 )
## Likelihood ratio test= 50.32  on 6 df,   p=4e-09
## Wald test               = 44.36  on 6 df,   p=6e-08
## Score (logrank) test = 57.04  on 6 df,   p=2e-10
```

Estrutura dos dados mudando no tempo

- ▶ A interpretação dos coeficientes do modelo de Cox estendido é feita da mesma forma que nos modelos semiparamétricos usuais em que a exponencial do coeficiente estima a razão de riscos (*hazard ratio*).
- ▶ No exemplo da TMO, através do teste de Wald, temos que idade, sexo, decr não têm efeito significativo, mas a fase da doença (*fasegr*), a presença de doença de enxerto aguda (*deag*) e a recuperação no número de plaquetas (*reqplaq*) são significativas.

Estrutura dos dados mudando no tempo

- ▶ O risco aumenta 2,5 e 2,8 vezes, respectivamente, para quem fez o transplante fora da primeira fase crônica da doença e para quem apresentou doença do enxerto aguda.
- ▶ Quem não recuperou o número de plaquetas, apresenta risco de óbito sete vezes maior.

Próxima aula

- ▶ Eventos (riscos) concorrentes (competitivos).
- ▶ Análise de sobrevivência multivariada (modelos de fragilidade).

Por hoje é só!

Bons estudos!

