

Análise de Sobrevida

1.0 - Aula Prática

Prof. Dr. Eder Angelo Milani

26/04/2025

Modelo de Cox estratificado

A análise dos dados de câncer CID-C34 apresentou violação da suposição de taxas de falha proporcionais para as covariáveis radioterapia e quimioterapia. Assim, uma possibilidade para a análise desses dados é a de estratificá-los de acordo com as categorias das covariáveis que violaram a suposição de riscos proporcionais.

Para exemplificar, vamos considerar o ajuste considerando a variável radioterapia, neste caso, o modelo de Cox estratificado fica representado por

$$\lambda(t|x_{ij}) = \lambda_{0j}(t) \exp(x'_{ij}\beta),$$

para $j = 1, 2$, $i = 1, \dots, n_j$, com n_j denotando o número de pacientes no j -ésimo estrato.

As linhas de código a seguir executam as seguintes tarefas:

- leitura dos dados filtrados para CID C34;
- breve visualização do conjunto de dados;
- modifica a escala de tempo para ano, em vez de dia;
- excluir os tempos iguais a zero.
- modifica a variável SEXO de 1-Masculino e 2-Feminino para 0-Masculino e 1-Feminino;
- ajusta o modelo de Cox estratificado.

```
# limpando o que tem na memoria
rm(list=ls())
library(survival)

# local onde esta o arquivo com os dados
setwd("G:\\Meu Drive\\UFG\\Especializacao\\Aulas Análise Sobrevida\\Códigos")

### leitura
dados <- read.csv("cancer_c34.csv")
head(dados)
```

##	TOPOGRUP	TEMPO	CENSURA	ANODIAG	IDADE	SEXO	CIRURGIA	RADIO	QUIMIO	ECGRUP
## 1	C34	292	1	2014	63	1	0	1	1	III
## 2	C34	132	1	2016	58	2	0	0	0	I
## 3	C34	3	0	2016	61	2	0	0	0	IV
## 4	C34	17	1	2016	67	1	0	0	0	IV
## 5	C34	182	1	2015	57	1	0	0	1	III
## 6	C34	287	1	2015	69	1	0	0	1	IV

```

# mudança na variável tempo - de dias para ano
dados$TEMPO <- dados$TEMPO/365

# excluir os tempos iguais a zero
ind_tempo_zero <- which(dados$TEMPO == 0)
ind_tempo_zero

## [1] 256 297 322 374 865 996 1010 1049 1083 1165 1514 1665 1754 2090 5830
## [16] 6012 6196 7079 8030 8049 8383 8495

dados$TEMPO[256]

## [1] 0

dados <- dados[-ind_tempo_zero,]

# outra forma de filtrar
# dados <- dados %>% filter(TEMPO != 0)

table(dados$SEXO)

##
##      1      2
## 5042 3867

dados$SEXO <- dados$SEXO -1

dados$SEXO <- factor(dados$SEXO)
dados$CIRURGIA <- factor(dados$CIRURGIA)
dados$RADIO <- factor(dados$RADIO)

cox_est <- coxph(Surv(TEMPO, CENSURA) ~ SEXO + CIRURGIA
                 + strata(RADIO), data = dados, x=TRUE)

summary(cox_est) # apresenta uma serie de estatisticas de interesse

## Call:
## coxph(formula = Surv(TEMPO, CENSURA) ~ SEXO + CIRURGIA + strata(RADIO),
##       data = dados, x = TRUE)
##
##      n= 8909, number of events= 6892
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## SEX01      -0.23203   0.79292  0.02460  -9.43  <2e-16 ***
## CIRURGIA1  -1.39642   0.24748  0.04185 -33.37  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## SEX01           0.7929      1.261   0.7556   0.8321
## CIRURGIA1       0.2475      4.041   0.2280   0.2686
##
## Concordance= 0.613 (se = 0.004 )
## Likelihood ratio test= 1663 on 2 df,  p=<2e-16
## Wald test               = 1224 on 2 df,  p=<2e-16
## Score (logrank) test = 1387 on 2 df,  p=<2e-16

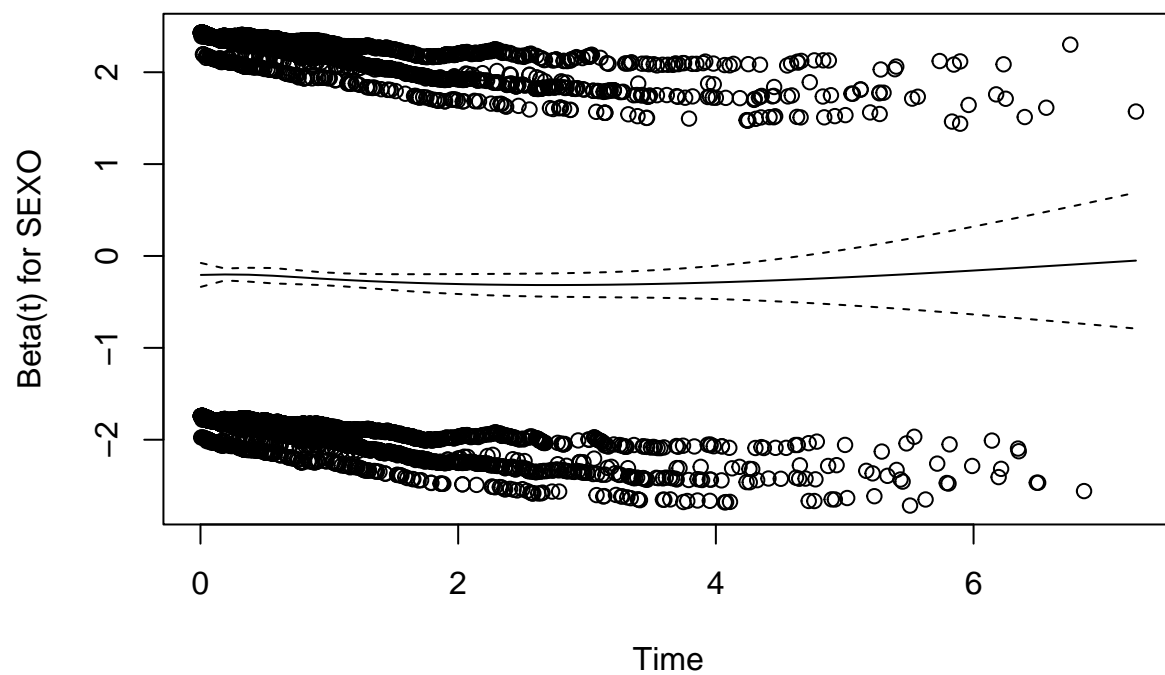
```

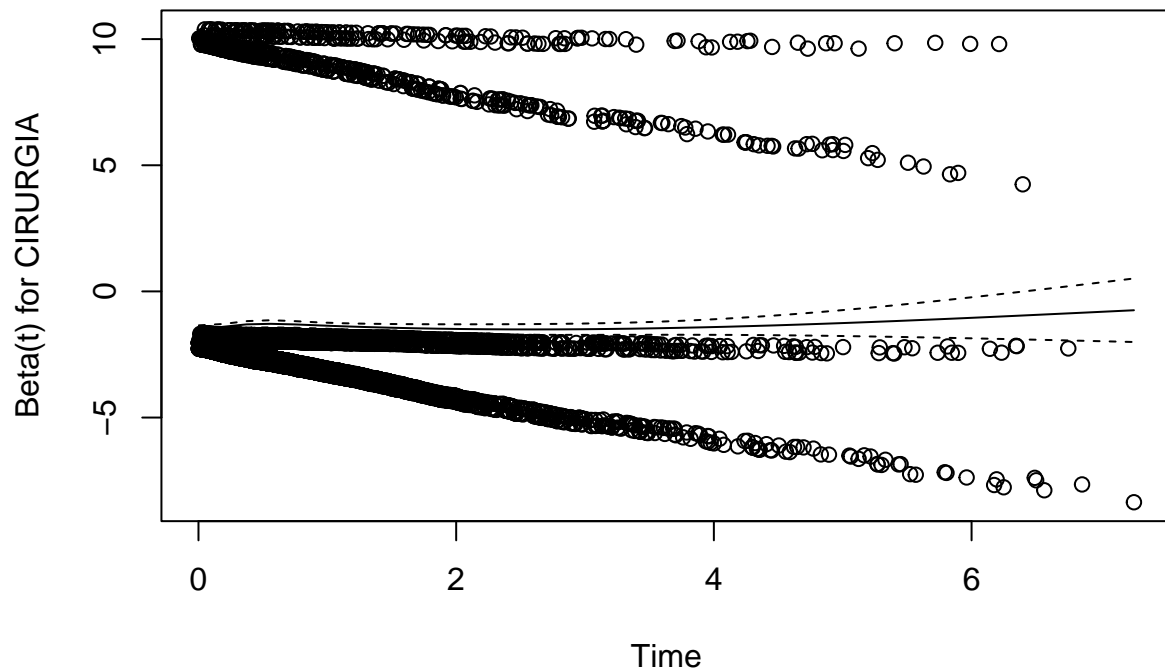
```
# avaliacao do modelo

# residuo de Schoenfeld
cox.zph(cox_est, transform = "identity")

##           chisq df    p
## SEXO      1.265326  1 0.26
## CIRURGIA  0.000076  1 0.99
## GLOBAL    1.265442  2 0.53

plot(cox.zph(cox_est, transform = "identity"))
```





```
# residuo de Cox-Snell
```

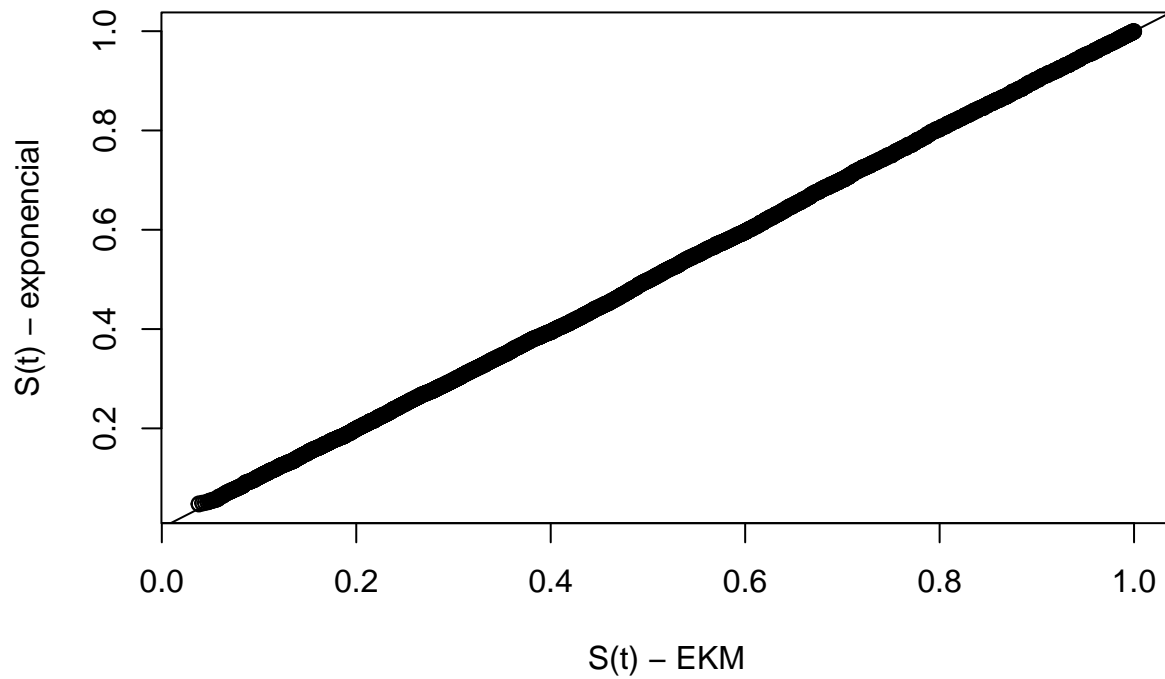
```
res_cs_cox_est <- dados$CENSURA - resid(cox_est, type = "martingale")
summary(res_cs_cox_est)
```

```
##      Min.   1st Qu.   Median     Mean   3rd Qu.     Max.
## 0.0003536 0.2622100 0.5762587 0.7735997 1.1061700 3.0534271
```

```
# grafico
```

```
cox_snell <- survfit(Surv(res_cs_cox_est, dados$CENSURA) ~ 1)
exp_res_cs_cox_est <- exp(- cox_snell$time)
plot(cox_snell$surv, exp_res_cs_cox_est, ylab="S(t) - exponencial",
      xlab="S(t) - EKM", main = "Resíduos de Cox-Snell")
abline(a=0, b=1)
```

Resíduos de Cox-Snell

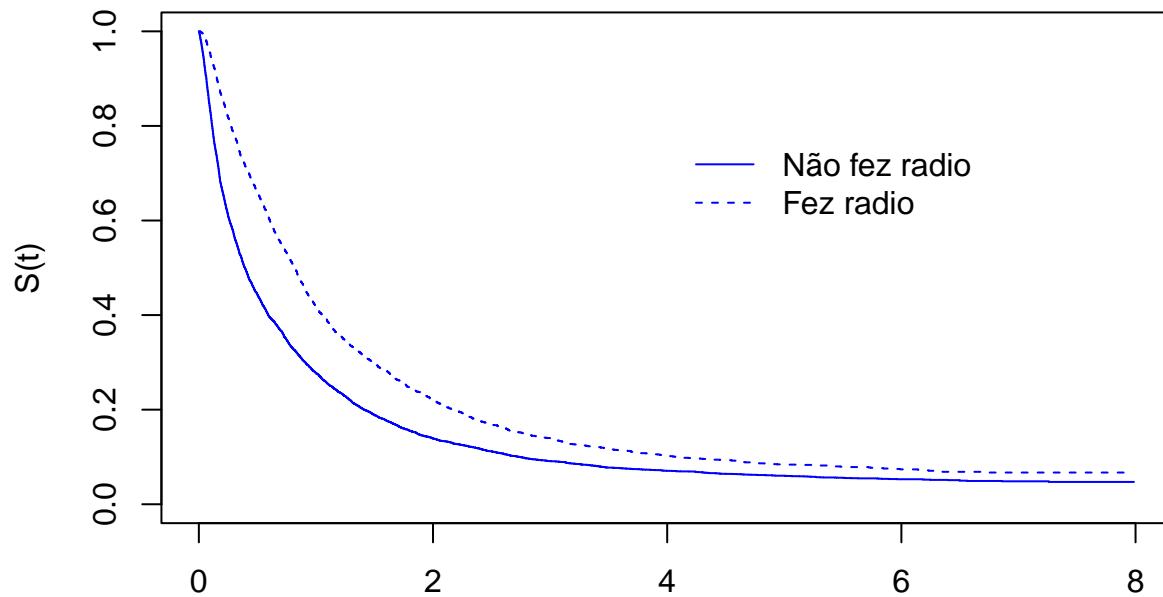


```
# sobrevivencia estimada

# Obter a função de sobrevivência basal
sobre_basal <- survfit(cox_est)
#summary(sobre_basal)

# Plotar a função de sobrevivência basal
plot(sobre_basal, main = "Função de Sobrevivência Basal",
      xlab = "Tempo", ylab = "S(t)", col=c("blue", "blue"), lty=1:2)
title(sub = "Covariáveis: Sexo = 0 - Masc e Cirurgia = 0 - Não")
legend(4, 0.8, legend = c("Não fez radio", "Fez radio"),
      col = c("blue", "blue"), bty="n", lty=1:2)
```

Função de Sobrevivência Basal



Tempo

Covariáveis: Sexo = 0 – Masc e Cirurgia = 0 – Não

Quero agora as 8 curvas de sobrevivencia estimada

```
covs <- expand.grid(  
  SEXO = c("0", "1"),  
  CIRURGIA = c("0", "1"),  
  RADIO = c("0", "1")  
)
```

variaveis como fatores

```
covs$SEXO <- factor(covs$SEXO, levels = levels(dados$SEXO))  
covs$CIRURGIA <- factor(covs$CIRURGIA, levels = levels(dados$CIRURGIA))  
covs$RADIO <- factor(covs$RADIO, levels = levels(dados$RADIO))
```

covs

```
##   SEXO CIRURGIA RADIO  
## 1    0         0     0  
## 2    1         0     0  
## 3    0         1     0  
## 4    1         1     0  
## 5    0         0     1  
## 6    1         0     1  
## 7    0         1     1  
## 8    1         1     1
```

sobrevivencia estimada para cada combinacao

```
curvas <- survfit(cox_est, newdata = covs)
```

```

# Cores e linhas para identificar
cores <- c("blue", "red", "blue", "red", "blue", "red", "blue", "red")
tipos_linha <- c(1, 1, 2, 2, 3, 3, 4, 4)

# Criar rotulos para legenda
rotulos <- apply(covs, 1, function(linha) {
  paste0("S", linha[["SEXO"]],
        ", C", linha[["CIRURGIA"]],
        ", R", linha[["RADIO"]])
})

# rotulos

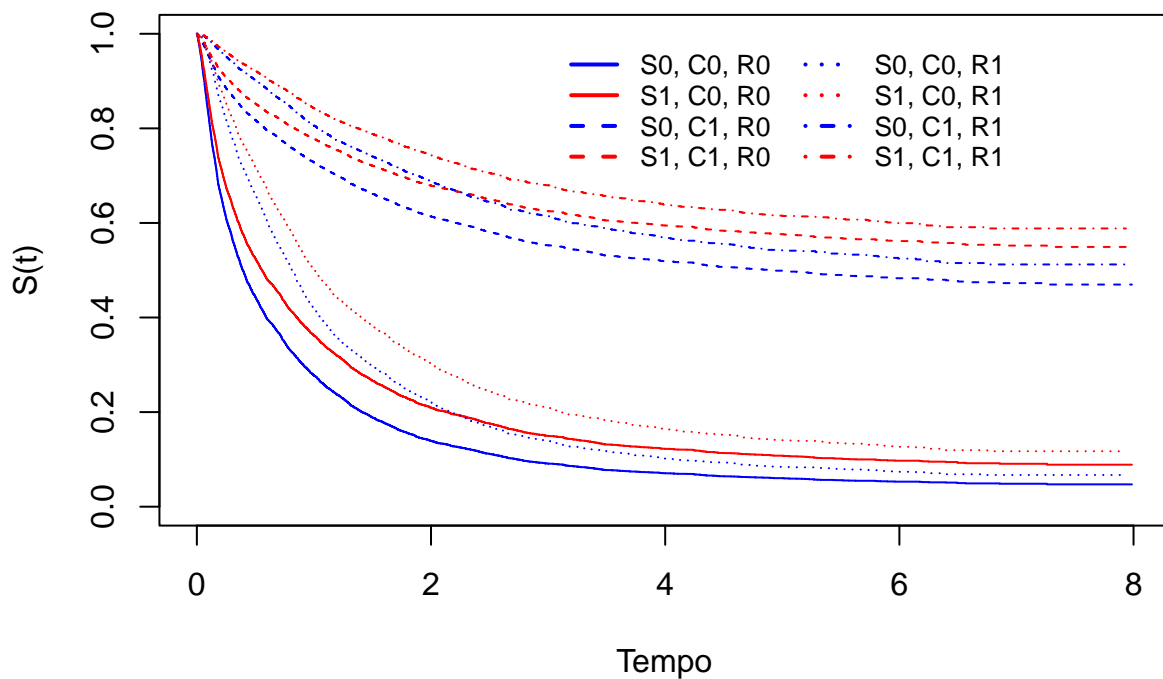
plot(curvas, col = cores, lty = tipos_linha,
     xlab = "Tempo", ylab = "S(t)",
     main = "Funções de Sobrevivência para todas as combinações")

legend(3, 1, legend = rotulos[1:4], col = cores[1:4],
      lty = tipos_linha[1:4], lwd = 2, horiz = F, cex = 0.8, bty = "n")

legend(5, 1, legend = rotulos[5:8], col = cores[5:8],
      lty = tipos_linha[5:8], lwd = 2, horiz = F, cex = 0.8, bty = "n")

```

Funções de Sobrevivência para todas as combinações



Observações:

- (i) Note que as estimativas obtidas para os β 's são muito similares as obtidas sem considerar a estratificação.

Com isso, as interpretações são também similares.

- (ii) Na formulação do modelo, independentemente da configuração das covariáveis, a razão das taxas de falhas sempre são constantes.

Mas observe que as curvas de sobrevivência para as covariáveis S1, C0, R0 e S0, C0, R1 se cruzam, isso só acontece se as taxas de falhas não são proporcionais, o mesmo também acontece para S1, C1, R0 e S0, C1, R1, isso ocorreu devido a utilização dos estratos que foram adotados seguindo a covariável RADIO.

- (iii) Quando a quantidade de combinações das covariáveis são muitas, a construção do gráfico com todas as possíveis combinações é pouco informativo. A sugestão nesse caso é a construção do gráfico para algumas configurações.