

Análise de Sobrevida

0.7 - Aula Prática

Prof. Dr. Eder Angelo Milani

25/04/2025

Leitura dos dados

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.

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

# 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)
```

Ajuste do modelo de Cox - variável sexo

As linhas de código a seguir executam as seguintes tarefas:

- modifica a variável SEXO de 1-Masculino e 2-Feminino para 0-Masculino e 1-Feminino;
- ajusta o modelo de Cox utilizando a variável sexo.

```
library(survival)

table(dados$SEXO)

##
##      1      2
## 5042 3867

dados$SEXO <- dados$SEXO -1

cox_sexo <- coxph(Surv(TEMPO, CENSURA) ~ as.factor(SEXO), data = dados, x=TRUE)
# x = TRUE salva a matriz das covariáveis para outras análises

summary(cox_sexo) # apresenta uma série de estatísticas de interesse

## Call:
## coxph(formula = Surv(TEMPO, CENSURA) ~ as.factor(SEXO), data = dados,
##       x = TRUE)
##
##      n= 8909, number of events= 6892
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## as.factor(SEXO)1 -0.28278   0.75369  0.02457 -11.51  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## as.factor(SEXO)1    0.7537      1.327    0.7183    0.7909
##
## Concordance= 0.533 (se = 0.003 )
## Likelihood ratio test= 134.3 on 1 df,  p=<2e-16
## Wald test               = 132.5 on 1 df,  p=<2e-16
## Score (logrank) test = 133.4 on 1 df,  p=<2e-16

# estimando a função de sobrevivência

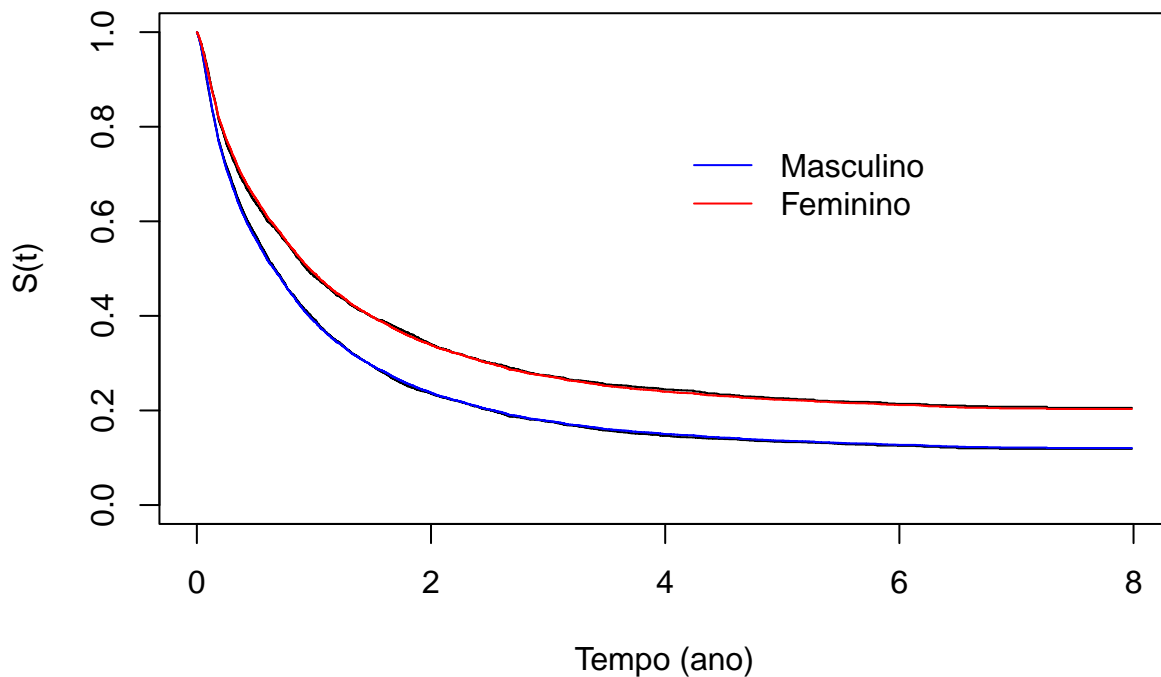
# usando Kaplan-Meier
ekm <- survfit(Surv(TEMPO, CENSURA) ~ as.factor(SEXO), data=dados)

# usando o modelo de Cox
novo_dados <- data.frame(SEXO = factor(c(0, 1)))
is.factor(novo_dados$SEXO)

## [1] TRUE
```

```
sobre_est <- survfit(cox_sexo, newdata = novo_dados)
#summary(sobre_est)

plot(ekm, mark.time = F, conf.int = F, xlab="Tempo (ano)", ylab="S(t)")
lines(sobre_est, col = c("blue", "red"))
legend(4, 0.8, legend = c("Masculino", "Feminino"),
      col = c("blue", "red"), lty = 1, bty = "n")
```



Observe que obtemos várias estatísticas, vamos a interpretação dos resultados;

- a coluna **coef** apresenta a estimativa dos parâmetros β 's, podemos destacar que valor positivo para β indica que a variável aumenta o risco de óbito, enquanto que valores negativos de β indica que a variável contribui para a redução do risco de óbito, quando comparado com o risco basal;
- a coluna **se(coef)** mostra o erro padrão das estimativas;
- a coluna **z** exibe a estatística do teste de hipóteses de Wald, com hipótese nula dada por $H_0 : \beta_i = 0$;
- a coluna **Pr(>|z|)** expõe o p-valor do teste de hipótese de Wald;
- a coluna **exp(coef)** calcula o valor $\exp(\beta_i)$, que é interpretado como a razão dos riscos. A coluna **exp(-coef)** indica o inverso de **exp(coef)**;
- as colunas **lower .95** e **upper .95** indicam os limites inferior e superior do intervalo de confiança de 95% para a razão dos riscos.

Neste exemplo, temos que a estimativa do parâmetro β referente a covariável Sexo é igual a -0,28, indicando assim que se o paciente for do sexo feminino diminui o risco de óbito. Além disso, o teste de hipóteses apresenta evidências a favor de rejeitar a hipótese nula, ao nível de significância de 5%, ou seja, que $\hat{\beta} \neq 0$, pois o p-valor é menor que 5%.

Agora quando comparamos o sexo masculino com o feminino, a partir da razão dos riscos, temos que

$$\frac{\lambda(t|x = \text{Fem})}{\lambda(t|x = \text{Mas})} = \exp(\hat{\beta}) = 0,75;$$

ou seja, a razão dos riscos é igual a 0,75, indicando que o risco do sexo feminino de ir a óbito é 0,75 vezes o risco do sexo de masculino de ir a óbito, ou ainda, o risco do sexo feminino é 25% menor do que risco do sexo masculino.

Caso o interesse esteja em usar o sexo masculino como referência, observe que temos que inverter a razão dada acima, o que resulta em analisar **exp(-coef)**, assim temos que o sexo masculino tem risco de óbito de 1,33 vezes o risco do sexo feminino, ou ainda, 33% a mais do que o sexo feminino.

O intervalo de confiança de 95% é dado por (0,718; 0,791), observamos que o valor 1 não pertence ao intervalo de confiança, indicando assim que a razão é diferente de 1, logo podemos concluir que o risco atribuído ao sexo feminino é diferente do risco do sexo masculino.

Obs.: note que já realizamos a interpretação dos parâmetros estimados pelo modelo de Cox, no entanto, essa interpretação só deve ser realizada após a verificação da adequação do modelo aos dados.

Adequação do modelo de Cox

```
# verificacao de proporcionalidade

# metodo grafico

mod1 <- coxph(Surv(TEMPO[SEXO==0], CENSURA[SEXO==0])~1,
              data=dados, method="breslow")
# estimacao do modelo de Cox sem covariaveis
# Colosimo e Giolo sugerem a estimacao do
# risco utilizando o metodo de Nelson-Aalen-Breslow

summary(mod1)

## Call:  coxph(formula = Surv(TEMPO[SEXO == 0], CENSURA[SEXO == 0]) ~
##       1, data = dados, method = "breslow")
##
## Null model
##   log likelihood= -31758.8
##   n= 5042

ss <- survfit(mod1)
H01 <- -log(ss$surv)
# relacao entre a S(t) e H(t)
# S(t) = exp(-H(t))
# -log(S(t)) = -log(exp(-H(t))) = H(t)
time1 <- ss$time

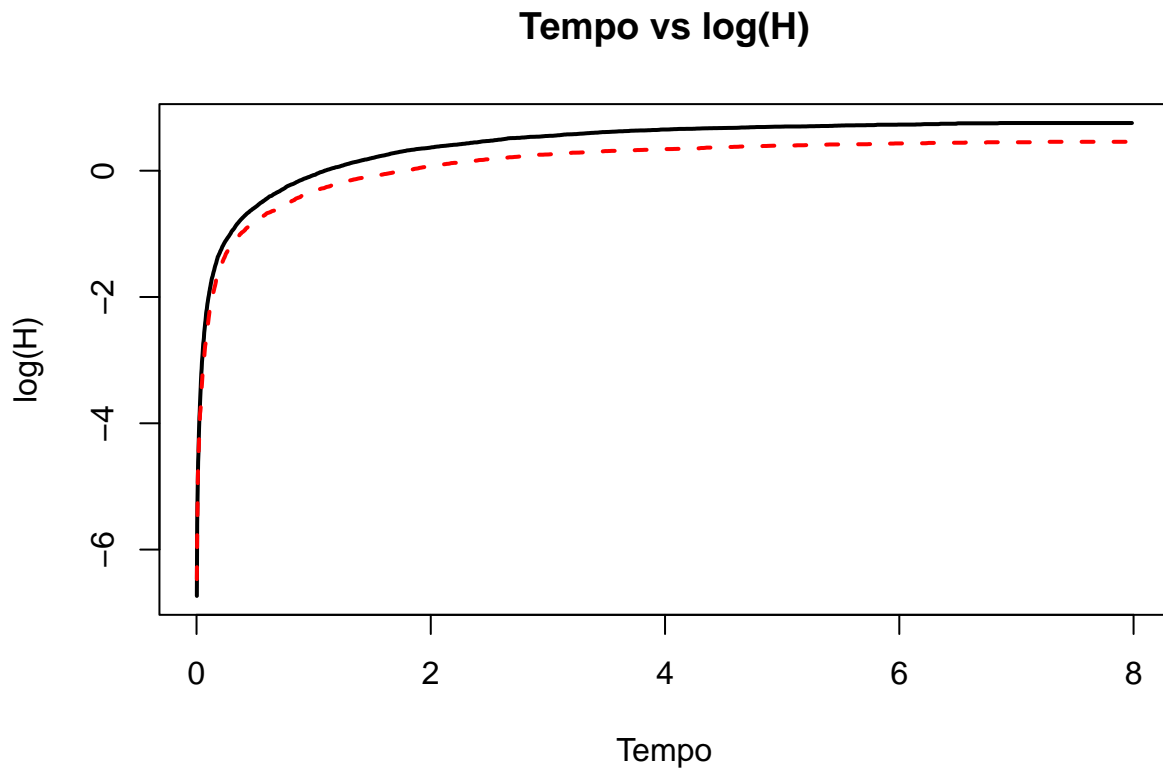
mod2 <- coxph(Surv(TEMPO[SEXO==1], CENSURA[SEXO==1])~1,
              data=dados, method="breslow")
summary(mod2)

## Call:  coxph(formula = Surv(TEMPO[SEXO == 1], CENSURA[SEXO == 1]) ~
##       1, data = dados, method = "breslow")
##
## Null model
```

```
## log likelihood= -21543.33
## n= 3867
```

```
ss <- survfit(mod2)
H02 <- -log(ss$surv)
time2 <- ss$time
```

```
plot(time1, log(H01), type='l', main="Tempo vs log(H)", lwd=2, xlab="Tempo", ylab="log(H)")
lines(time2, log(H02), lty=2, col="red", lwd=2)
```



```
# observe que existe indícios de proporcionalidade
```

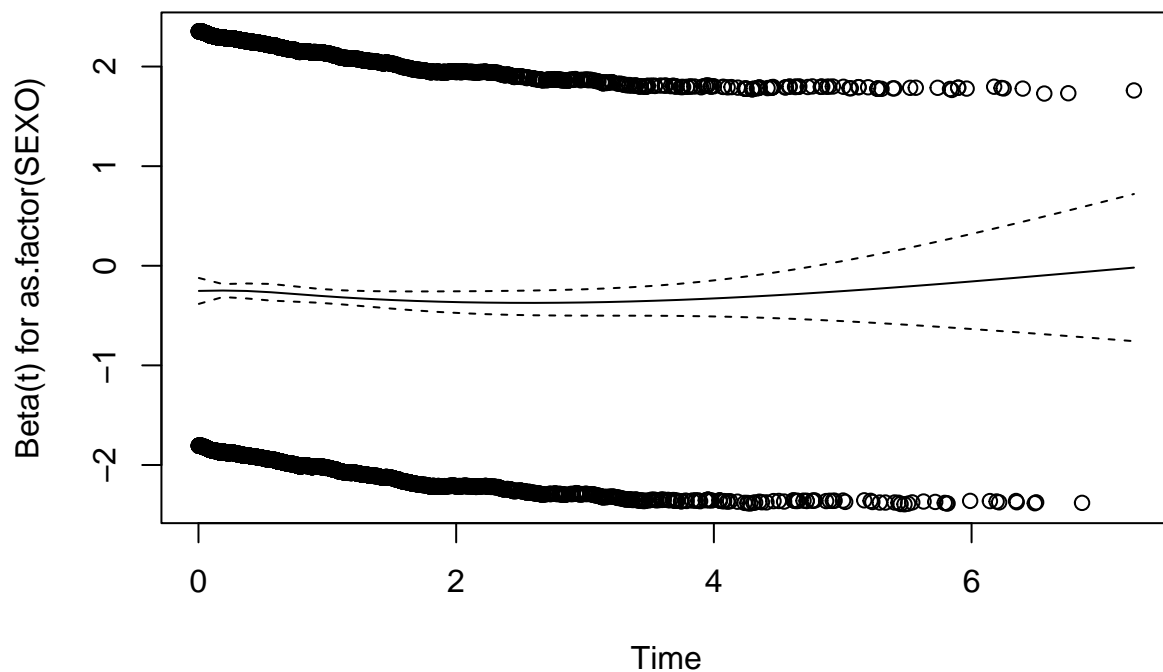
```
# resíduos de Schoenfeld
```

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

```
##               chisq df    p
## as.factor(SEX0)  1.23  1 0.27
## GLOBAL           1.23  1 0.27
```

```
# observe que o teste de hipóteses não rejeita
# a hipótese nula de proporcionalidade
```

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



*# observe que a curva estimada fica quase que constante ao longo do tempo
e proximo do valor estimado do beta*

residuo de Cox-Snell

```
res_cs_sexo <- dados$CENSURA - resid(cox_sexo, type = "martingale")
```

```
summary(res_cs_sexo)
```

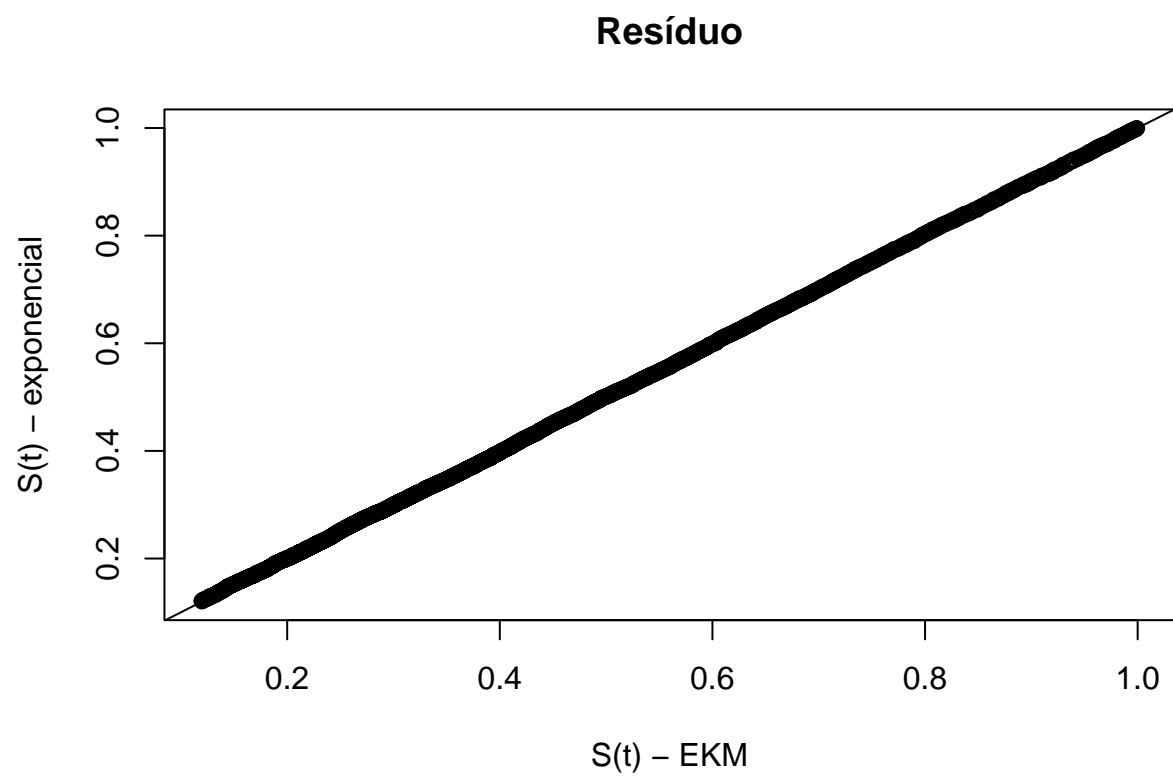
```
##      Min.   1st Qu.   Median     Mean  3rd Qu.     Max.
## 0.000616 0.267722 0.644540 0.773600 1.224780 2.115041
```

grafico

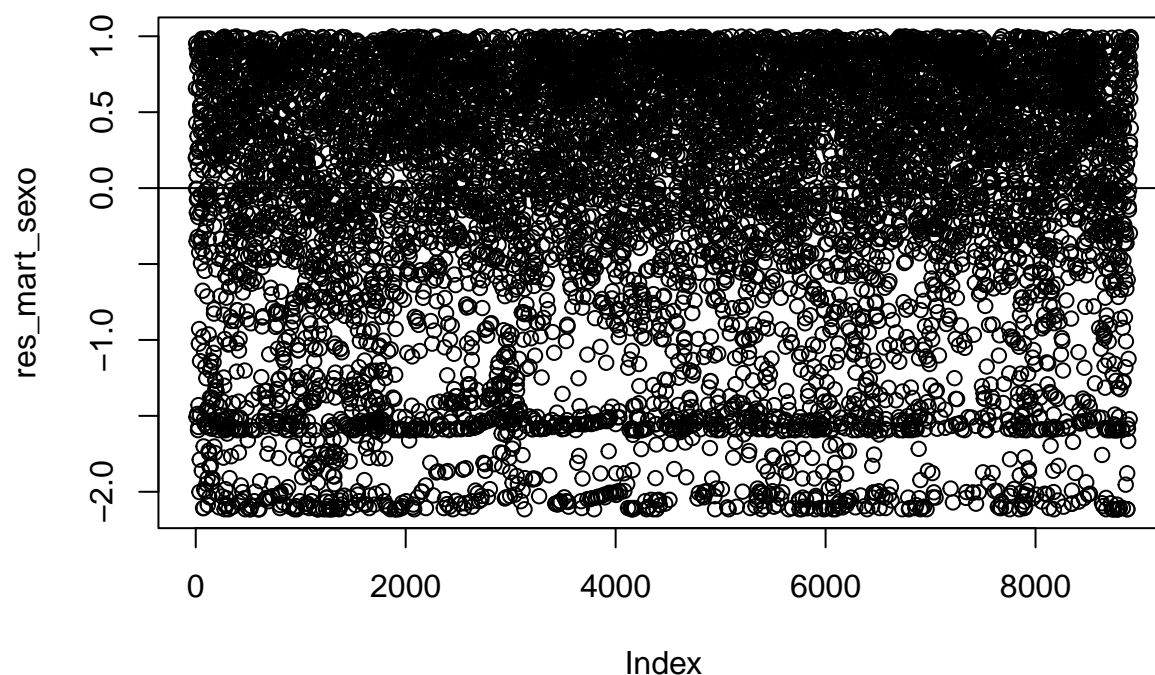
```
ekm_cox_snell_sexo <- survfit(Surv(res_cs_sexo, dados$CENSURA) ~ 1)
```

```
exp_res_cs_sexo <- exp(- ekm_cox_snell_sexo$time)
```

```
plot(ekm_cox_snell_sexo$surv, exp_res_cs_sexo, main = "Resíduo", ylab="S(t) - exponencial", xlab="S(t) - exponencial", abline(a=0, b=1))
```



```
# residuo martigale  
  
res_mart_sexo <- resid(cox_sexo, type = "martingale")  
  
plot(res_mart_sexo)  
abline(h=0)
```



```
# note que nao vemos pontos aberrantes/outliers
```

```
# probabilidade de concordancia
```

```
# teste de hipoteses
```

```
summary(cox_sexo)
```

```
## Call:
## coxph(formula = Surv(TEMPO, CENSURA) ~ as.factor(SEX0), data = dados,
##       x = TRUE)
##
## n= 8909, number of events= 6892
##
##               coef exp(coef) se(coef)      z Pr(>|z|)
## as.factor(SEX0)1 -0.28278   0.75369  0.02457 -11.51  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##               exp(coef) exp(-coef) lower .95 upper .95
## as.factor(SEX0)1    0.7537      1.327    0.7183    0.7909
##
## Concordance= 0.533 (se = 0.003 )
## Likelihood ratio test= 134.3 on 1 df,  p=<2e-16
## Wald test               = 132.5 on 1 df,  p=<2e-16
## Score (logrank) test = 133.4 on 1 df,  p=<2e-16
```



```
# concordancia de 0.533 - significa que a concordancia pode ser por acaso  
# teste de hipoteses para  $\beta_{\text{sexo}} = 0$   
# TRV - rejeita a hipotese nula  
# Wald - rejeita a hipotese nula  
# Score - rejeita a hipotese nula
```