

Análise de Sobrevida

0.8 - Aula Prática

Prof. Dr. Eder Angelo Milani

26/04/2025

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 quimioterapia

Vamos repetir a análise mas agora utilizando a covariável quimioterapia - Quimio - em vez da variável sexo.

```
require(survival)
```

```
## Carregando pacotes exigidos: survival
table(dados$QUIMIO)

## 
##      0      1
## 4029 4880

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

summary(cox_quimio) # apresenta uma série de estatísticas de interesse
```

```
## Call:
## coxph(formula = Surv(TEMPO, CENSURA) ~ as.factor(QUIMIO), data = dados,
##       x = TRUE)
## 
##      n= 8909, number of events= 6892
## 
##              coef exp(coef) se(coef)      z Pr(>|z|)    
## as.factor(QUIMIO)1 -0.22985   0.79465  0.02453 -9.369   <2e-16 ***
## ---      
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
##          exp(coef) exp(-coef) lower .95 upper .95    
## as.factor(QUIMIO)1    0.7947      1.258    0.7573    0.8338  
## 
## Concordance= 0.583  (se = 0.003 )    
## Likelihood ratio test= 86.67  on 1 df,   p=<2e-16    
## Wald test            = 87.79  on 1 df,   p=<2e-16    
## Score (logrank) test = 88.16  on 1 df,   p=<2e-16
```

resíduo de Cox-Snell

```
res_cs_quimio <- dados$CENSURA - resid(cox_quimio, type = "martingale")

summary(res_cs_quimio)
```

```
##      Min.    1st Qu.    Median     Mean    3rd Qu.    Max.    
## 0.0008224 0.2805063 0.6239537 0.7735997 1.1840098 2.1088068
```

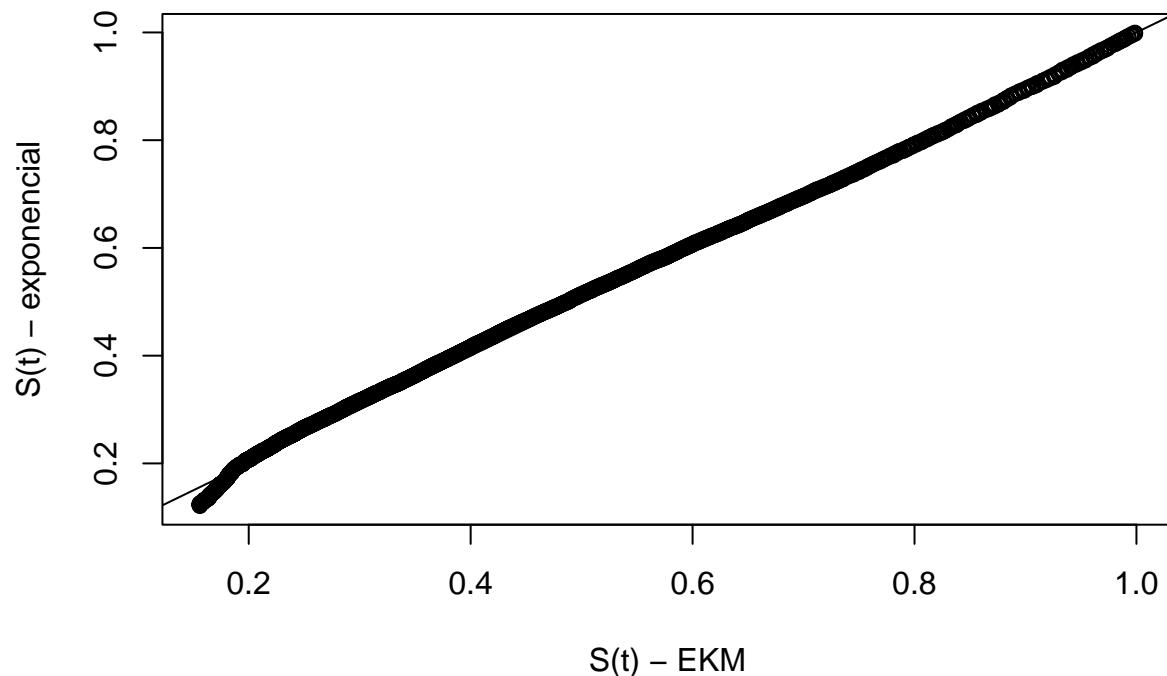
gráfico

```
ekm_cox_snell_quimio <- survfit(Surv(res_cs_quimio, dados$CENSURA) ~ 1)

exp_res_cs_quimio <- exp(- ekm_cox_snell_quimio$time)
```

```
# grafico de curvas de sobrevida sobrepostas
plot(ekm_cox_snell_quimio$surv, exp_res_cs_quimio, main= "Resíduos de Cox-Snell", ylab="S(t) - exponencial",
abline(a=0, b=1)
```

Resíduos de Cox-Snell



```
# verificacao de proporcionalidade

# metodo grafico

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

summary(mod1)

## Call: coxph(formula = Surv(TEMPO[QUIMIO == 0], CENSURA[QUIMIO == 0]) ~
##               1, data = dados, method = "breslow")
## 
## Null model
##   log likelihood= -22725.9
##   n= 4029

ss <- survfit(mod1)
H01 <- -log(ss$surv)
time1 <- ss$time

mod2 <- coxph(Surv(TEMPO[QUIMIO==1], CENSURA[QUIMIO==1])~1,
```

```

    data=dados, method="breslow")
summary(mod2)

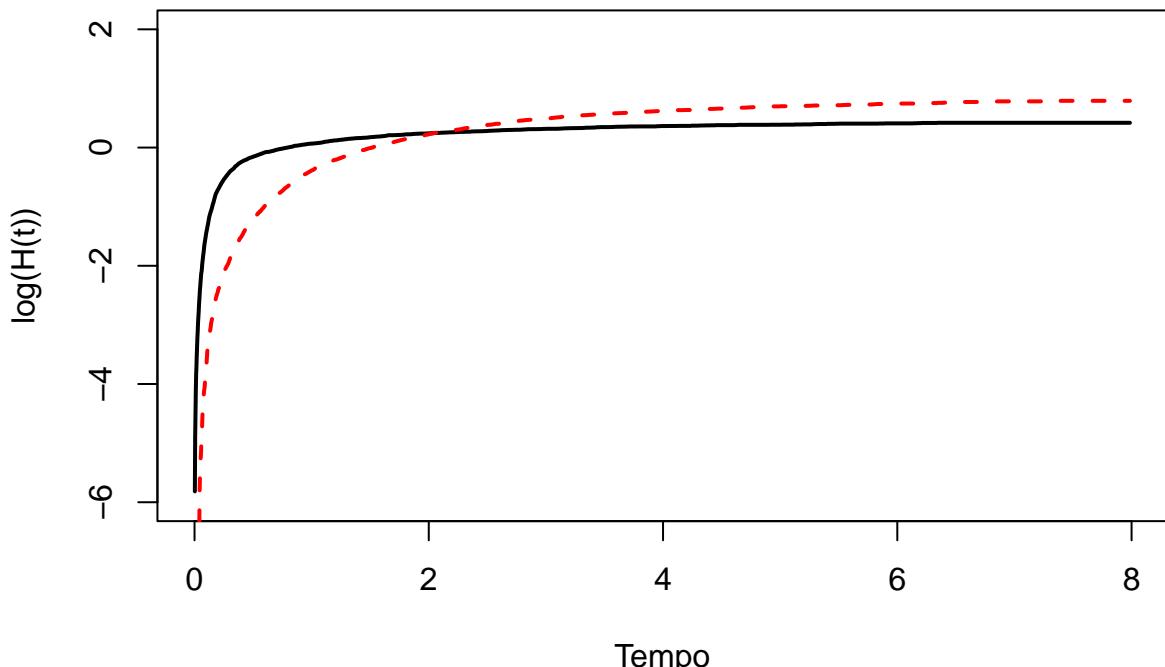
## Call: coxph(formula = Surv(TEMPO[QUIMIO == 1], CENSURA[QUIMIO == 1]) ~
##      1, data = dados, method = "breslow")
##
## Null model
##  log likelihood= -30566.25
##  n= 4880

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",
      ylim=c(-6, 2), ylab="log(H(t))")
lines(time2, log(H02), lty=2, col="red", lwd=2)

```

Tempo vs log(H)



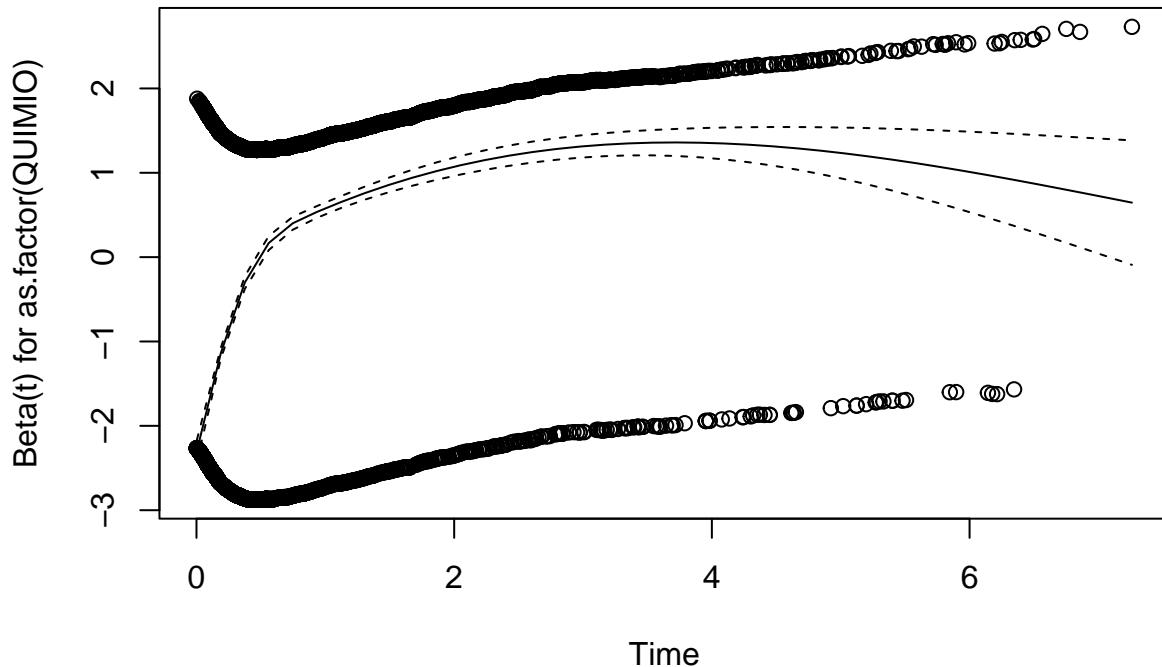
```
# observe que existe indicios de NAO proporcionalidade
```

```
# residuos de Schoenfeld
cox.zph(cox_quimio, transform = "identity")

##                      chisq df      p
## as.factor(QUIMIO) 1116   1 <2e-16
## GLOBAL            1116   1 <2e-16
```

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

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



```
# observe que a curva estimada NAO é constante ao longo do tempo
# indicando assim ausência de proporcionalidade
```

Observe que não verificamos proporcionalidade na covariável quimioterapia, indicando assim que o modelo de Cox não pode ser usado.

E na prática como funciona?

Como na prática temos várias covariáveis para proceder com o ajuste, vamos utilizar rotinas para a seleção de covariáveis, como por exemplo o stepwise. E na sequência verificar a questão da adequação do modelo de Cox aos dados.

```
head(dados)
```

```
##   TOPOGRUP      TEMPO CENSURA ANODIAG IDADE SEXO CIRURGIA RADIO QUIMIO ECGRUP
## 1    C34 0.8000000000      1  2014    63    1      0      1      1     III
## 2    C34 0.361643836      1  2016    58    2      0      0      0      0      I
## 3    C34 0.008219178      0  2016    61    2      0      0      0      0     IV
## 4    C34 0.046575342      1  2016    67    1      0      0      0      0     IV
## 5    C34 0.498630137      1  2015    57    1      0      0      0      1     III
## 6    C34 0.786301370      1  2015    69    1      0      0      0      1     IV
```

```

# TOPOGRUP nao sera analisada pois e 100% preenchida com C34

# ANODIAG sera analisada como uma variavel categorica
dados$ANODIAG <- factor(dados$ANODIAG)
table(dados$ANODIAG)

## 
## 2014 2015 2016
## 2985 2926 2998
is.factor(dados$ANODIAG)

## [1] TRUE
# IDADE e quantitativa
is.numeric(dados$IDADE)

## [1] TRUE
# SEXO e qualitativa
dados$SEXO <- factor(dados$SEXO)
table(dados$SEXO)

## 
## 1 2
## 5042 3867
# CIRURGIA e qualitativa
dados$CIRURGIA <- factor(dados$CIRURGIA)
table(dados$CIRURGIA)

## 
## 0 1
## 7302 1607
# RADIO e qualitativa
dados$RADIO <- factor(dados$RADIO)
table(dados$RADIO)

## 
## 0 1
## 6217 2692
# QUIMIO e qualitativa
dados$QUIMIO <- factor(dados$QUIMIO)
table(dados$QUIMIO)

## 
## 0 1
## 4029 4880
# todas as variaveis ok, estamos prontos para seguir com o ajuste
#install.packages("MASS")
library(MASS)

## Warning: pacote 'MASS' foi compilado no R versão 4.4.3
# modelo inicial com todas as variaveis usando a distribuicao exponencial
modelo_inicial_cox <- coxph(Surv(TEMPO, CENSURA) ~ ANODIAG + IDADE + SEXO+
                           CIRURGIA + RADIO + QUIMIO,
                           data = dados)

```

```

summary(modelo_inicial_cox)

## Call:
## coxph(formula = Surv(TEMPO, CENSURA) ~ ANODIAG + IDADE + SEXO +
##        CIRURGIA + RADIO + QUIMIO, data = dados)
##
##    n= 8909, number of events= 6892
##
##              coef exp(coef)   se(coef)      z Pr(>|z|)
## ANODIAG2015 -0.039641  0.961134  0.029447 -1.346 0.178240
## ANODIAG2016 -0.044219  0.956745  0.029477 -1.500 0.133588
## IDADE       -0.004092  0.995916  0.001110 -3.685 0.000228 ***
## SEX02        -0.236483  0.789399  0.024761 -9.551 < 2e-16 ***
## CIRURGIA1   -1.700512  0.182590  0.043391 -39.191 < 2e-16 ***
## RADIO1       -0.298125  0.742209  0.026410 -11.288 < 2e-16 ***
## QUIMIO1      -0.602149  0.547634  0.025919 -23.232 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## ANODIAG2015   0.9611      1.040    0.9072   1.0182
## ANODIAG2016   0.9567      1.045    0.9030   1.0136
## IDADE         0.9959      1.004    0.9938   0.9981
## SEX02         0.7894      1.267    0.7520   0.8287
## CIRURGIA1    0.1826      5.477    0.1677   0.1988
## RADIO1        0.7422      1.347    0.7048   0.7816
## QUIMIO1       0.5476      1.826    0.5205   0.5762
##
## Concordance= 0.701 (se = 0.003 )
## Likelihood ratio test= 2358 on 7 df,  p=<2e-16
## Wald test        = 1920 on 7 df,  p=<2e-16
## Score (logrank) test = 2060 on 7 df,  p=<2e-16
# vamos agora aplicar a selecao stepwise baseada no AIC
modelo_cox <- stepAIC(modelo_inicial_cox, direction = "both")

## Start:  AIC=113681.9
## Surv(TEMPO, CENSURA) ~ ANODIAG + IDADE + SEXO + CIRURGIA + RADIO +
##        QUIMIO
##
##              Df      AIC
## - ANODIAG   2  113681
## <none>        113682
## - IDADE     1  113693
## - SEXO      1  113772
## - RADIO     1  113811
## - QUIMIO    1  114200
## - CIRURGIA 1  115791
##
## Step:  AIC=113680.6
## Surv(TEMPO, CENSURA) ~ IDADE + SEXO + CIRURGIA + RADIO + QUIMIO
##
##              Df      AIC
## <none>        113681

```

```

## + ANODIAG 2 113682
## - IDADE 1 113692
## - SEXO 1 113771
## - RADIO 1 113809
## - QUIMIO 1 114200
## - CIRURGIA 1 115787

# Resumo do modelo final
summary(modelo_cox)

## Call:
## coxph(formula = Surv(TEMPO, CENSURA) ~ IDADE + SEXO + CIRURGIA +
##        RADIO + QUIMIO, data = dados)
##
## n= 8909, number of events= 6892
##
##          coef exp(coef)  se(coef)      z Pr(>|z|)
## IDADE     -0.004121  0.995888  0.001110  -3.711 0.000206 ***
## SEXO2     -0.236616  0.789294  0.024760  -9.556 < 2e-16 ***
## CIRURGIA1 -1.698821  0.182899  0.043378 -39.163 < 2e-16 ***
## RADIO1     -0.297250  0.742858  0.026404 -11.258 < 2e-16 ***
## QUIMIO1    -0.602888  0.547229  0.025910 -23.268 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##          exp(coef) exp(-coef) lower .95 upper .95
## IDADE      0.9959      1.004    0.9937    0.9981
## SEXO2      0.7893      1.267    0.7519    0.8285
## CIRURGIA1  0.1829      5.467    0.1680    0.1991
## RADIO1     0.7429      1.346    0.7054    0.7823
## QUIMIO1    0.5472      1.827    0.5201    0.5757
##
## Concordance= 0.701 (se = 0.003 )
## Likelihood ratio test= 2355 on 5 df,  p=<2e-16
## Wald test      = 1918 on 5 df,  p=<2e-16
## Score (logrank) test = 2057 on 5 df,  p=<2e-16

# Adequacao do modelo ajustado

# verificacao de proporcionalidade

cox.zph(modelo_cox, transform = "identity")

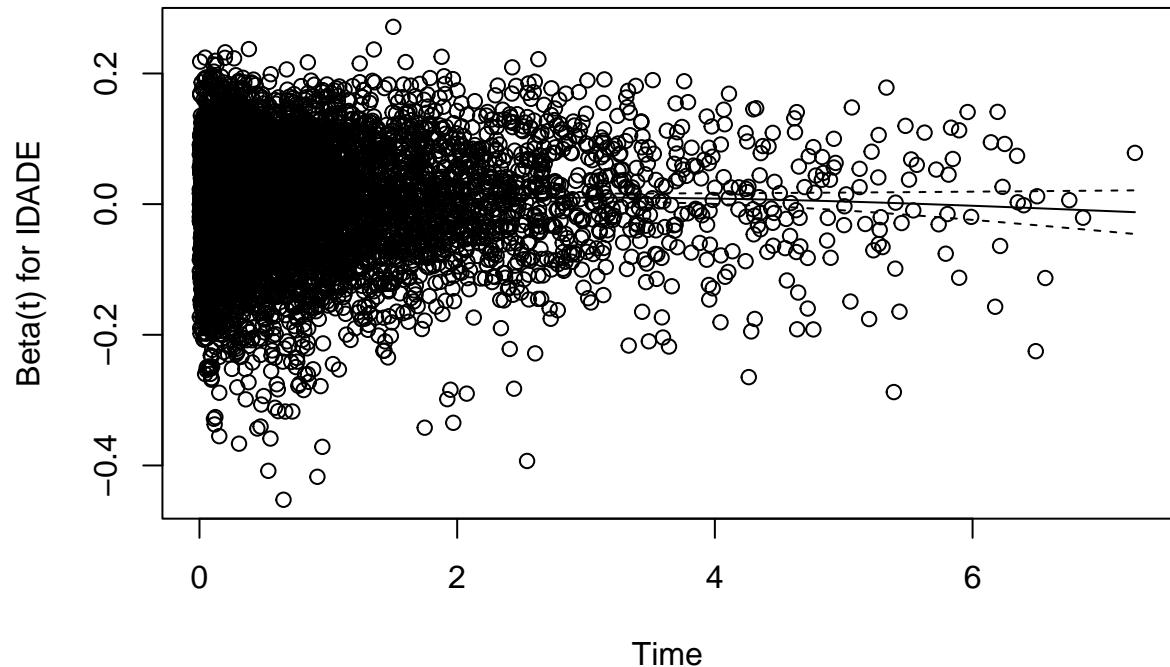
##          chisq df      p
## IDADE     0.0252  1  0.874
## SEXO      3.1305  1  0.077
## CIRURGIA  3.8496  1  0.050
## RADIO     108.8549  1 <2e-16
## QUIMIO    790.8282  1 <2e-16
## GLOBAL    942.2898  5 <2e-16

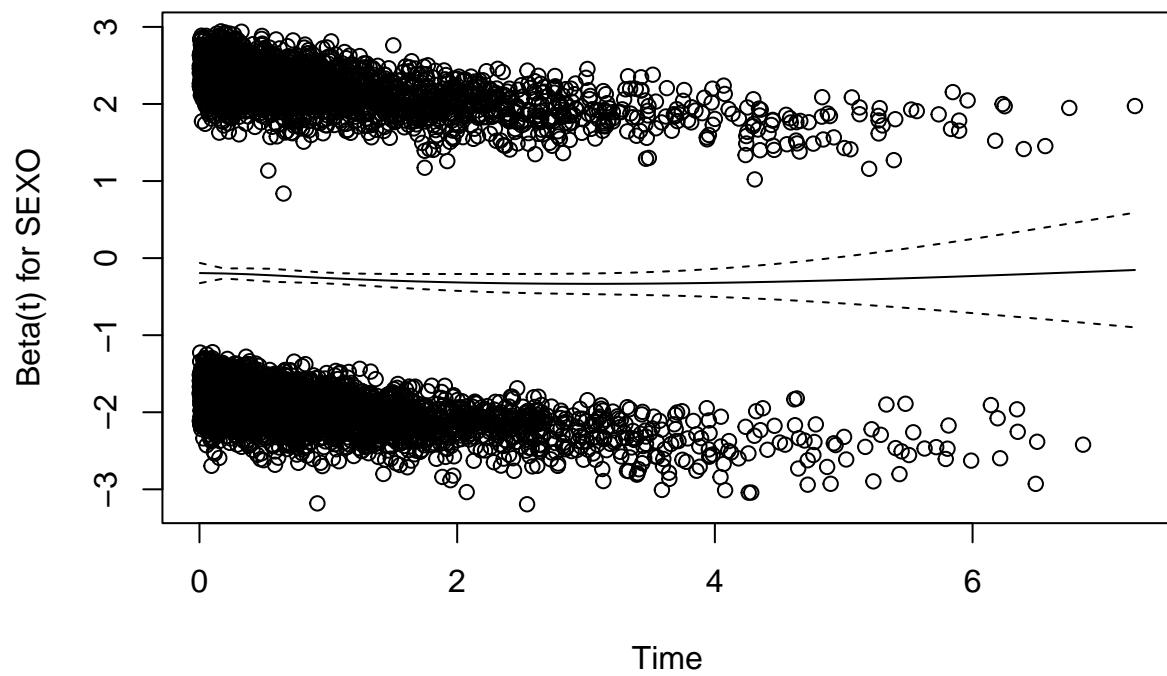
# notemos que:
# idade - aceita a hipotese de proporcionalidade
# sexo - aceita a hipotese de proporcionalidade
# cirurgia - aceita a hipotese de proporcionalidade
# note que no limite de 5% de significancia

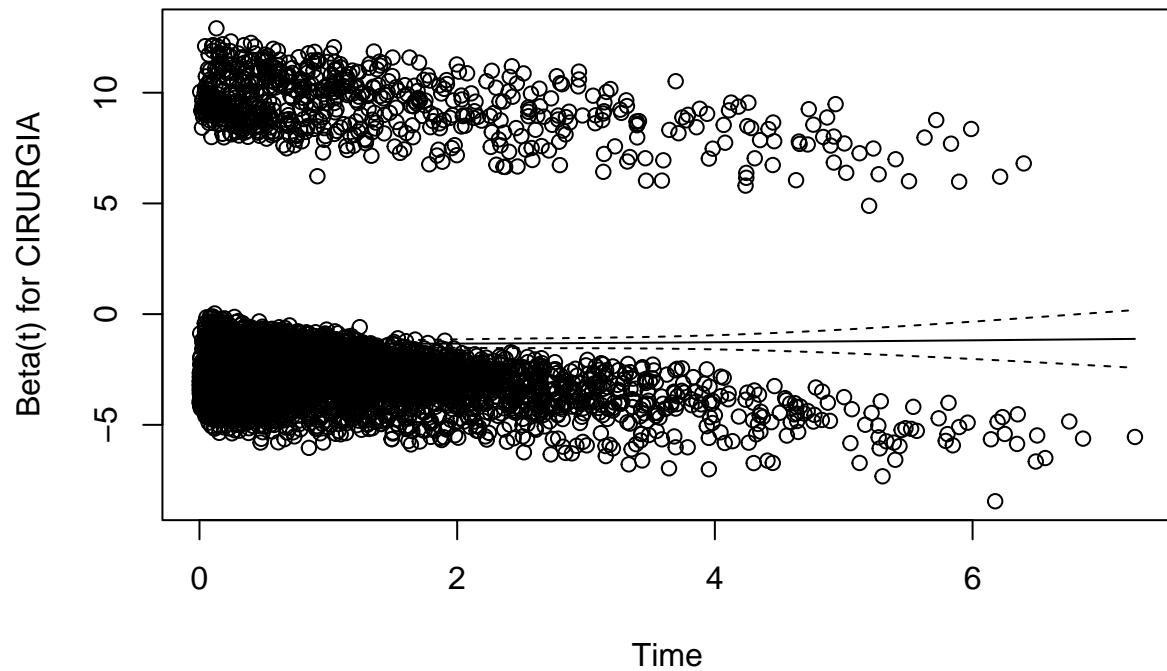
```

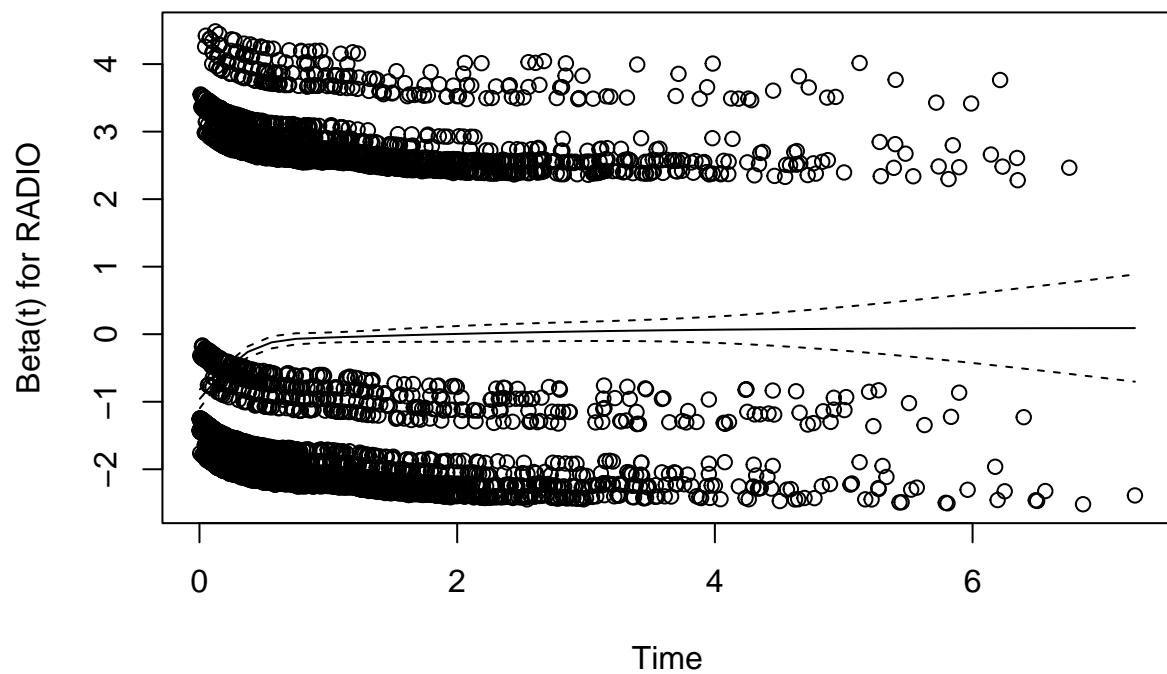
```
# radio - rejeita a hipotese de proporcionalidade
# quimio - rejeita a hipotese de proporcionalidade
# global - rejeita a hipotese de proporcionalidade

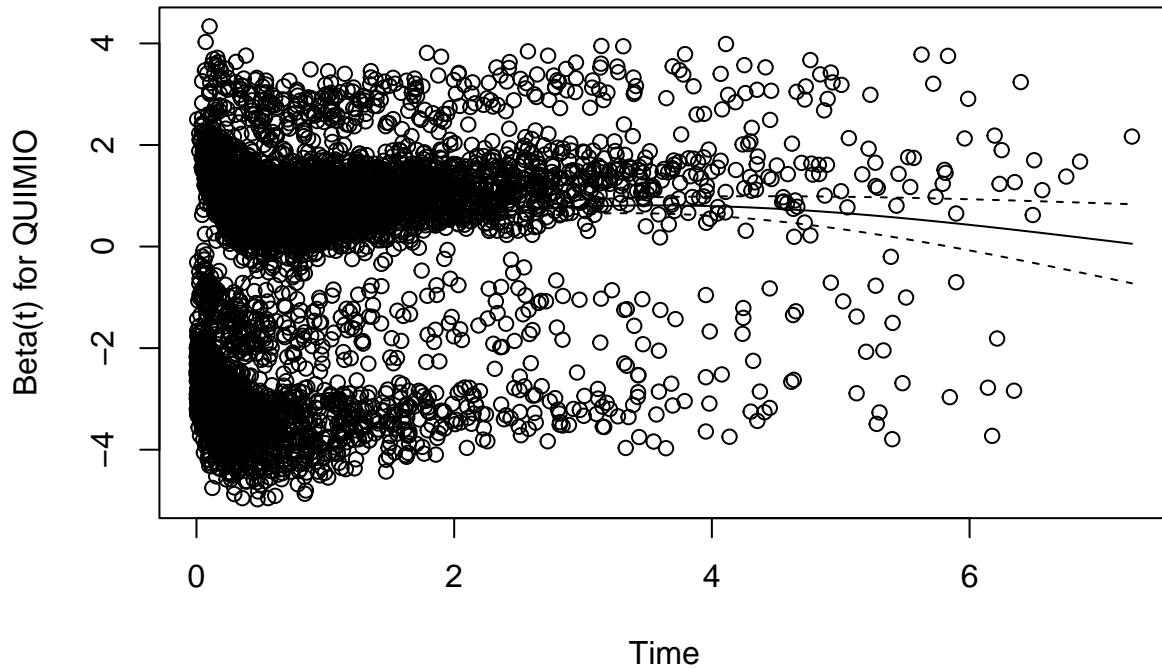
plot(cox.zph(modelo_cox, transform = "identity"))
```











Do fato que as variáveis radioterapia e quimioterapia não apresentaram riscos proporcionais, vamos eliminá-las do ajuste, pois ainda não sabemos como lidar com covariáveis com riscos não proporcionais e modelo de Cox.

modelo final sem as covariáveis radioterapia e quimioterapia

```
modelo_final_cox <- coxph(Surv(TEMPO, CENSURA) ~ IDADE + SEXO + CIRURGIA,
                           data = dados)
```

Resumo do modelo final
`summary(modelo_final_cox)`

```
## Call:
## coxph(formula = Surv(TEMPO, CENSURA) ~ IDADE + SEXO + CIRURGIA,
##        data = dados)
##
##      n= 8909, number of events= 6892
##
##              coef  exp(coef)  se(coef)      z Pr(>|z|)
## IDADE      0.001113  1.001114  0.001117  0.996   0.319
## SEX02     -0.228378  0.795823  0.024720 -9.239  <2e-16 ***
## CIRURGIA1 -1.359678  0.256743  0.041277 -32.941 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef)  exp(-coef) lower .95 upper .95
## IDADE      1.0011      0.9989   0.9989    1.0033
## SEX02      0.7958      1.2566   0.7582    0.8353
```

```

## CIRURGIA1    0.2567    3.8949    0.2368    0.2784
##
## Concordance= 0.602  (se = 0.004 )
## Likelihood ratio test= 1661  on 3 df,  p=<2e-16
## Wald test          = 1210  on 3 df,  p=<2e-16
## Score (logrank) test = 1379  on 3 df,  p=<2e-16
# Adequacao do modelo ajustado

# verificacao de proporcionalidade

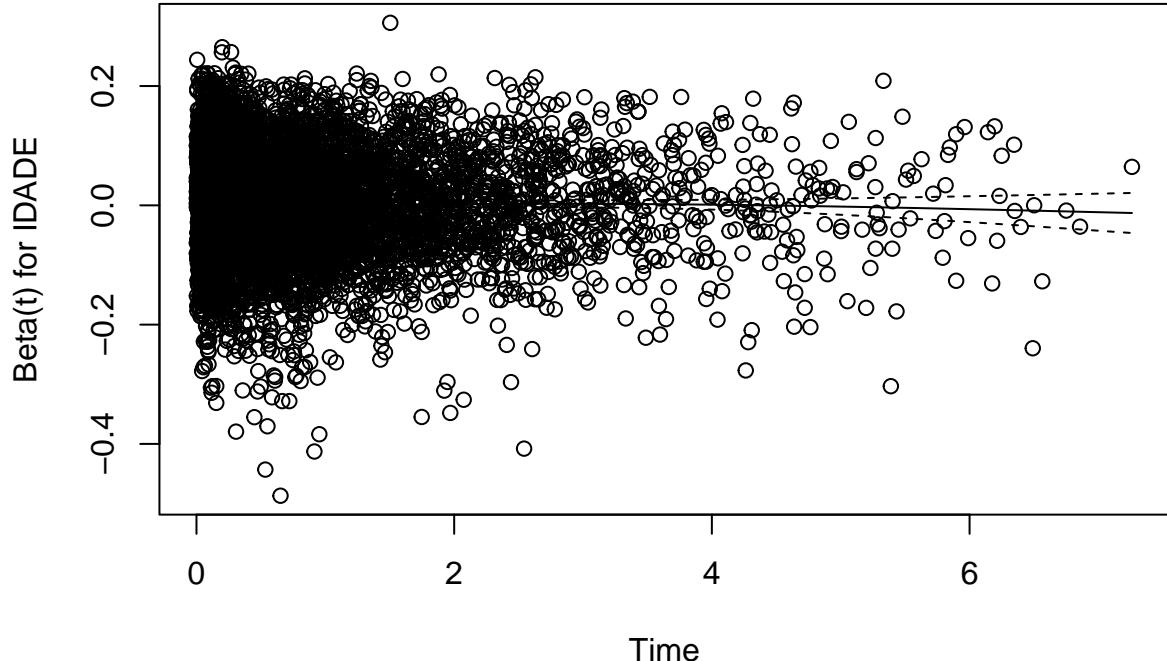
cox.zph(modelo_final_cox, transform = "identity")

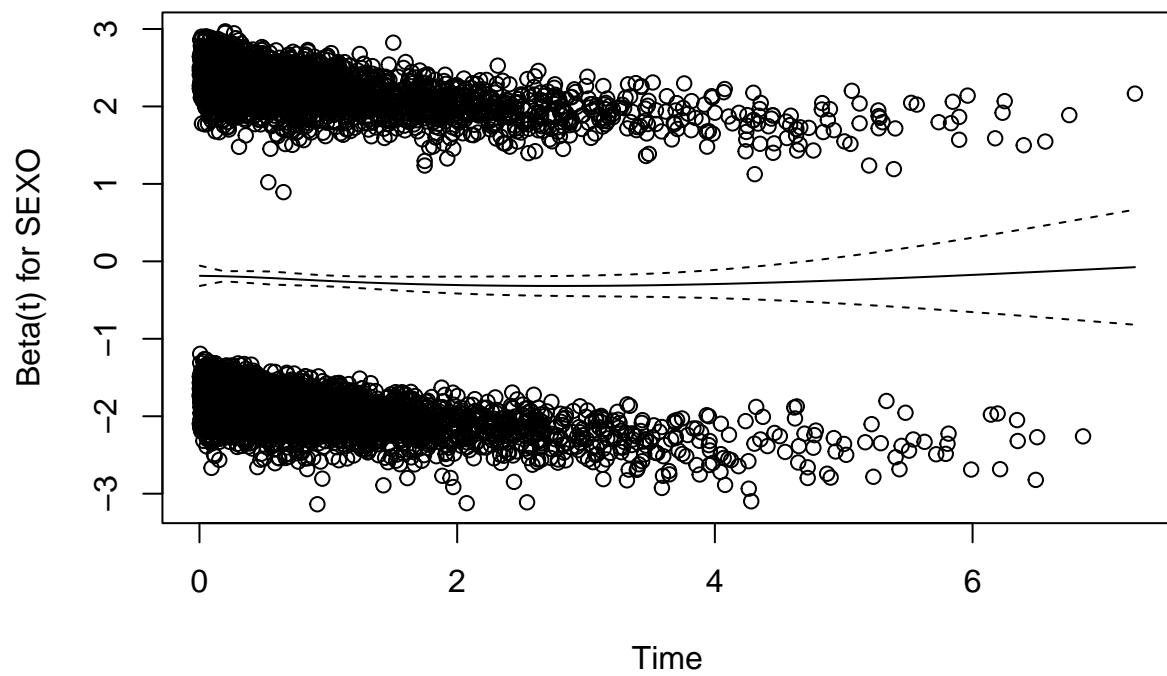
##          chisq df      p
## IDADE    0.363  1 0.55
## SEXO     1.616  1 0.20
## CIRURGIA 1.298  1 0.25
## GLOBAL   3.519  3 0.32

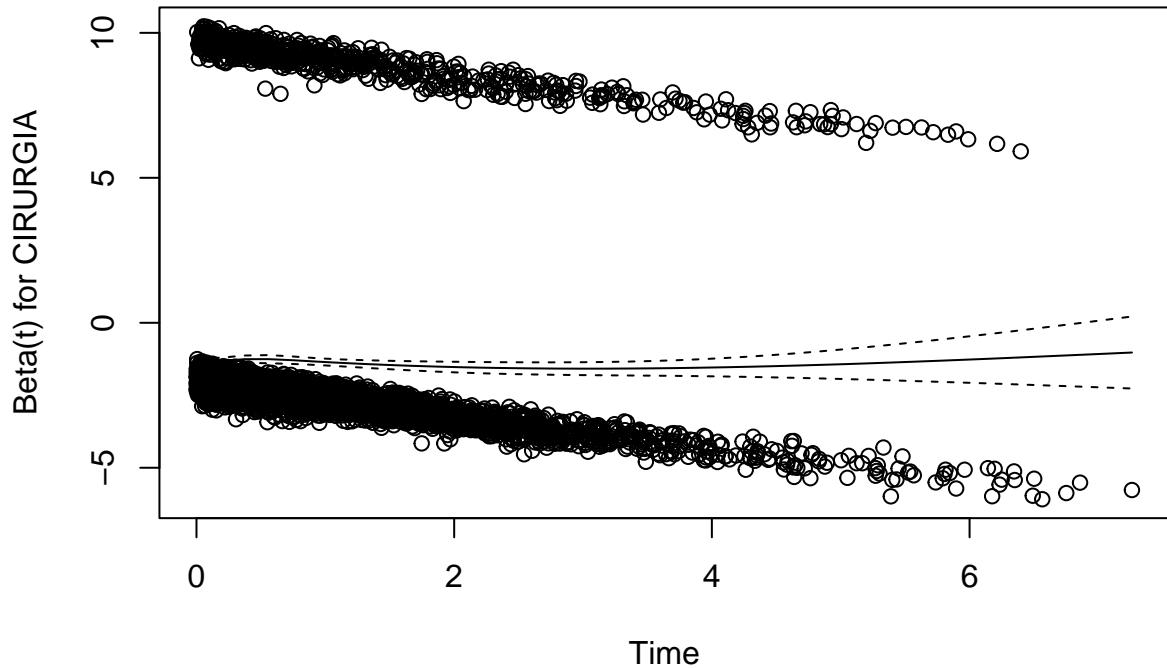
# todas as variaveis apresentaram riscos proporcionais
# o teste global tambem indicou riscos proporcionais

plot(cox.zph(modelo_final_cox, transform = "identity"))

```







```

# residuo de Cox-Snell

res_cs_final_cox <- dados$CENSURA - resid(modelo_final_cox, type = "martingale")
summary(res_cs_final_cox)

##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## 0.0007256 0.2588521 0.5734939 0.7735997 1.1141334 2.9706098

# grafico

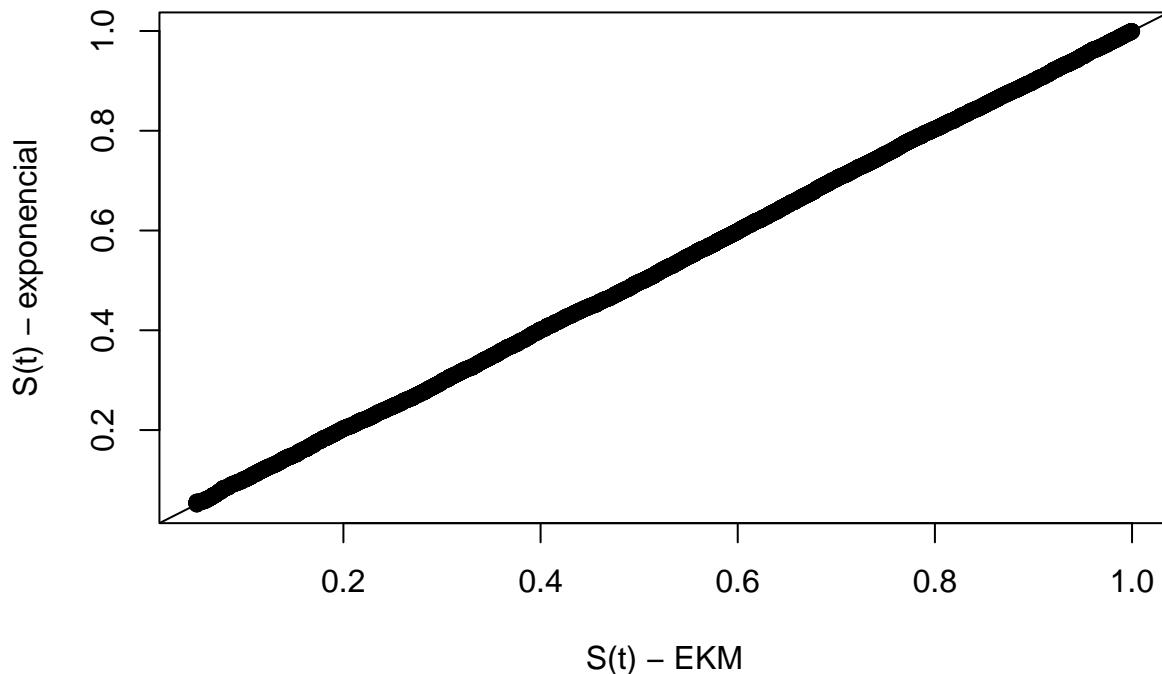
ekm_cox_snell_final <- survfit(Surv(res_cs_final_cox, dados$CENSURA) ~ 1)

exp_res_cs_final_cox <- exp( - ekm_cox_snell_final$time)

plot(ekm_cox_snell_final$surv, exp_res_cs_final_cox, main = "Resíduo de Cox-Snell", ylab="S(t) - exponencial")
abline(a=0, b=1)

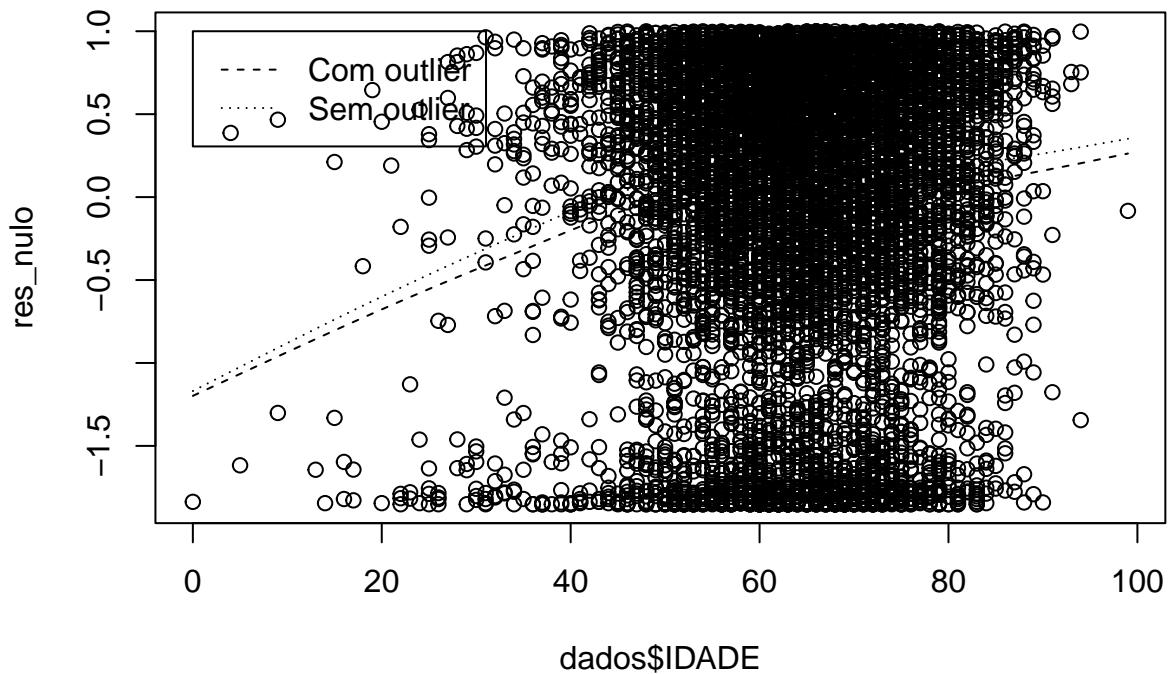
```

Resíduo de Cox–Snell



```
# Note que a idade nao foi significativa, vamos realizar a analise do residual martingale.

modelo_nulo <- coxph(Surv(TEMPO, CENSURA) ~ 1, data = dados)
res_nulo <- resid(modelo_nulo, type="martingale")
plot(dados$IDADE, res_nulo)
lines(lowess(dados$IDADE, res_nulo, iter = 0), lty=2)
# iter = 0 -> para que a curva estimada não elimine os valores extremos
# e não suavize demais os pontos aberrantes (outliers)
lines(lowess(dados$IDADE, res_nulo), lty=3)
legend(0, 1, lty=c(2,3), c("Com outlier", "Sem outlier"))
```



```

# a estimativa da funcao suavizada indica uma linearizacao
# mas a covariavel nao foi estatisticamente significativa,
# sendo assim, vamos excluir-la da modelagem

# modelo final v2 sem as covariaveis radioterapia, quimioterapia e idade

modelo_final_v2_cox <- coxph(Surv(TEMPO, CENSURA) ~ SEXO + CIRURGIA,
                                data = dados)

# Resumo do modelo final v2
summary(modelo_final_v2_cox)

## Call:
## coxph(formula = Surv(TEMPO, CENSURA) ~ SEXO + CIRURGIA, data = dados)
##
##     n= 8909, number of events= 6892
##
##             coef  exp(coef)  se(coef)      z Pr(>|z|)    
## SEX02      -0.23092   0.79380  0.02459  -9.392  <2e-16 ***
## CIRURGIA1 -1.36256   0.25601  0.04118 -33.090  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##             exp(coef)  exp(-coef) lower .95 upper .95    
## SEX02      0.7938      1.260     0.7565     0.8330

```

```

## CIRURGIA1      0.2560      3.906      0.2362      0.2775
##
## Concordance= 0.601  (se = 0.004 )
## Likelihood ratio test= 1660  on 2 df,  p=<2e-16
## Wald test          = 1209  on 2 df,  p=<2e-16
## Score (logrank) test = 1378  on 2 df,  p=<2e-16
# Adequacao do modelo ajustado

# verificacao de proporcionalidade

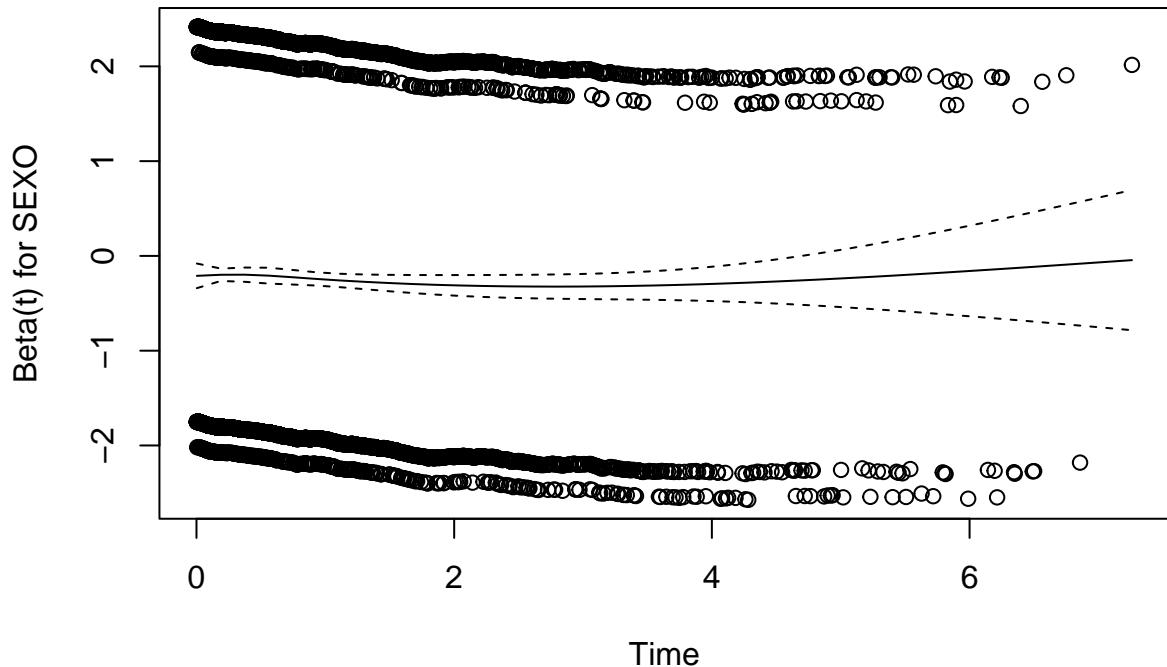
cox.zph(modelo_final_v2_cox, transform = "identity")

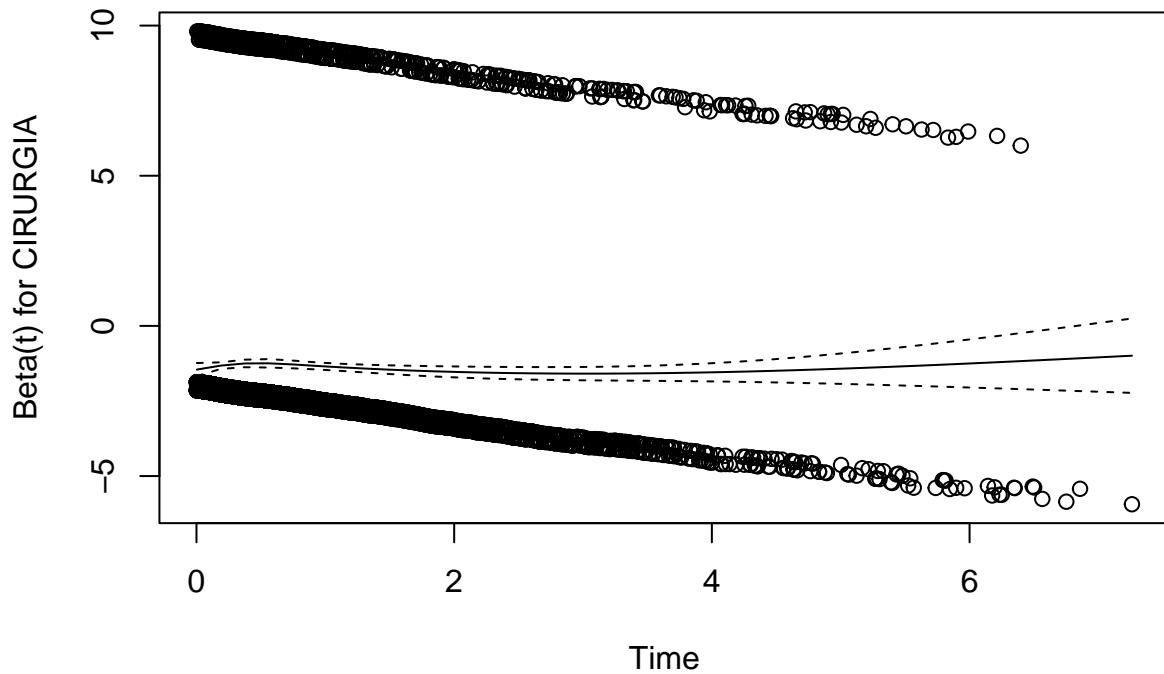
##          chisq df      p
## SEXO      1.61  1 0.20
## CIRURGIA  1.31  1 0.25
## GLOBAL    2.85  2 0.24

# todas as variaveis apresentaram riscos proporcionais
# o teste global tambem indicou riscos proporcionais

plot(cox.zph(modelo_final_v2_cox, transform = "identity"))

```





```

# residuo de Cox-Snell

res_cs_final_v2_cox <- dados$CENSURA - resid(modelo_final_v2_cox, type = "martingale")
summary(res_cs_final_v2_cox)

##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## 0.0007333 0.2582639 0.5748390 0.7735997 1.1131134 2.9129895

# grafico

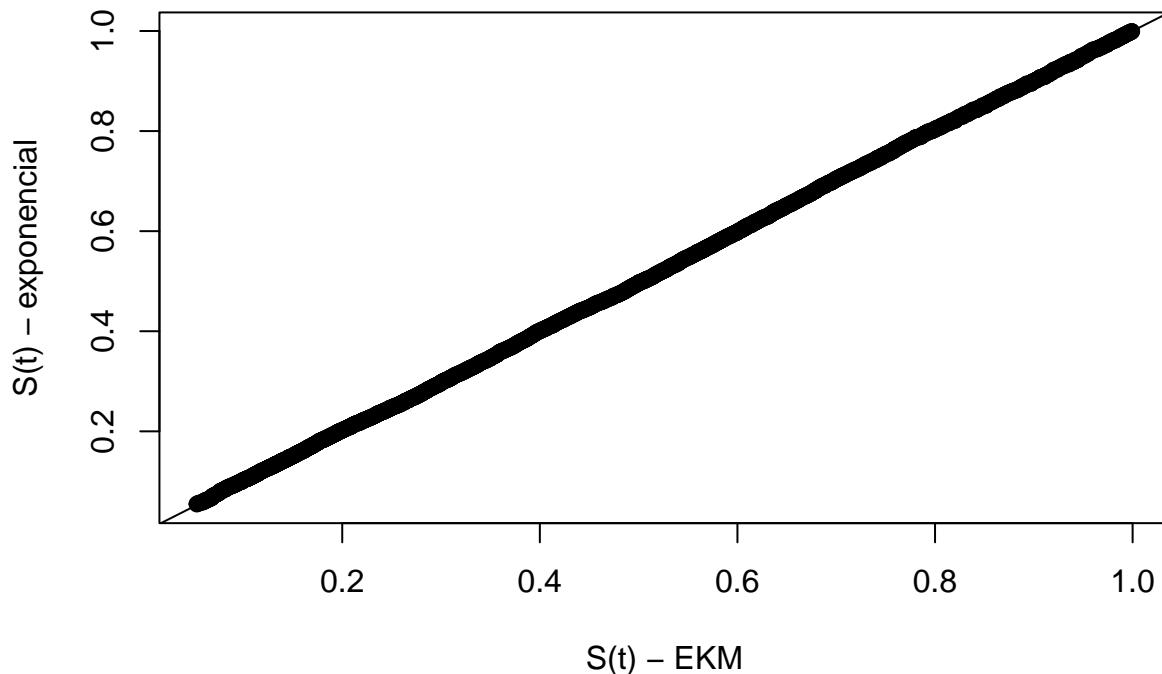
ekm_cox_snell_final <- survfit(Surv(res_cs_final_v2_cox, dados$CENSURA) ~ 1)

exp_res_cs_final_v2_cox <- exp(-ekm_cox_snell_final$time)

plot(ekm_cox_snell_final$surv, exp_res_cs_final_v2_cox, main= "Resíduo de Cox-Snell", ylab="S(t) - exp")
abline(a=0, b=1)

```

Resíduo de Cox–Snell



```

# retomando o modelo final
# para analisar os resultados

summary(modelo_final_v2_cox)

## Call:
## coxph(formula = Surv(TEMPO, CENSURA) ~ SEXO + CIRURGIA, data = dados)
##
## n= 8909, number of events= 6892
##
##          coef  exp(coef)  se(coef)      z Pr(>|z|)
## SEX02     -0.23092   0.79380  0.02459  -9.392  <2e-16 ***
## CIRURGIA1 -1.36256   0.25601  0.04118 -33.090  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##          exp(coef)  exp(-coef) lower .95 upper .95
## SEX02      0.7938      1.260    0.7565    0.8330
## CIRURGIA1     0.2560      3.906    0.2362    0.2775
##
## Concordance= 0.601  (se = 0.004 )
## Likelihood ratio test= 1660  on 2 df,  p=<2e-16
## Wald test          = 1209  on 2 df,  p=<2e-16
## Score (logrank) test = 1378  on 2 df,  p=<2e-16

```

Do fato que o modelo de Cox se ajustou bem aos dados, podemos afirmar que:

- pacientes do sexo feminino e que realizaram cirurgia são fatores que diminuem o risco de morte (observe que a estimativa dos β 's são negativas);
- pacientes do sexo feminino apresentam 21% a menos do risco de morte do que pacientes do sexo masculino;
- pacientes que realizaram cirurgia apresentam risco de óbito igual a 0,25 do risco daqueles que não realizaram cirurgia;
- o modelo apresenta concordância igual a 0,601;
- todos os testes de hipóteses indicam que os parâmetros β 's são diferentes de zero.