

Atividade Avaliativa

Análise de Sobrevida

Ana Maria Alves da Silva

2025-05-24

O objetivo da atividade avaliativa é a análise do conjunto de df de pacientes diagnosticados com neoplasia maligna do estômago (CID C16), com diagnóstico entre os anos de 2013 a 2016, com acompanhamento até o ano de 2021. Os df foram obtidos da Fundação Oncocentro de São Paulo (FOSP).

As variáveis disponíveis para análise são:

- TOPOGRUP: grupo da topografia
- TEMPO: tempo em anos do diagnóstico até a falha ou censura
- CENSURA: variável que indica se o tempo é de falha (=1) ou de censura à direita (=0)
- ANODIAG: indica o ano do diagnóstico
- SEXO: 0 - Masculino ou 1 - Feminino
- CIRURGIA: 0 se não realizou cirurgia - 1 se realizou cirurgia
- RADIO: 0 se não realizou radioterapia - 1 se realizou radioterapia
- QUIMIO: 0 se não realizou quimioterapia - 1 se realizou quimioterapia

Questão 1. Faça a leitura do conjunto de df cancer_c16.csv e formate as variáveis ANODIAG, SEXO, CIRURGIA, RADIO e QUIMIO para fator.

Solução:

Primeiro vamos carregar o conjunto de df:

```
setwd <- "/Users/anamaria/especializacao/modulo_12/Atividade"
df <- read.csv("cancer_c16.csv", sep = ",")
print(dim(df))
```

```
## [1] 9376    8
```

```
print(is.data.frame(df))
```

```
## [1] TRUE
```

Agora, vamos transforma as variáveis em fator.

```
df$ANODIAG <- as.factor(df$ANODIAG)
df$SEXO <- as.factor(df$SEXO)
df$CIRURGIA <- as.factor(df$CIRURGIA)
df$RADIO <- as.factor(df$RADIO)
df$QUIMIO <- as.factor(df$QUIMIO)
```

```
# Verificando as transformações
str(df)
```

```
## 'data.frame': 9376 obs. of 8 variables:
## $ TOPOGRUP: chr "C16" "C16" "C16" "C16" ...
## $ TEMPO : num 3.696 3.668 0.512 0.953 1.041 ...
## $ CENSURA : int 1 1 0 1 1 1 1 0 0 1 ...
## $ ANODIAG : Factor w/ 4 levels "2013","2014",...: 3 3 1 4 2 4 2 1 1 4 ...
## $ SEXO : Factor w/ 2 levels "0","1": 1 1 2 2 2 1 1 1 1 2 ...
## $ CIRURGIA: Factor w/ 2 levels "0","1": 1 1 1 2 2 1 2 1 1 1 ...
## $ RADIO : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...
## $ QUIMIO : Factor w/ 2 levels "0","1": 1 2 2 2 2 2 2 2 2 2 ...
```

Questão 2. Faça as seguintes análises descritivas:

- (i) calcule a proporção e o valor absoluto para os possíveis valores das variáveis CENSURA, ANODIAG, SEXO, CIRURGIA, RADIO e QUIMIO. Comente os valores encontrados.

Solução:

```
table(df$CENSURA)
```

```
##
## 0 1
## 3372 6004
```

```
Censura_prop <- round(prop.table(table(df$CENSURA)),2)
print(Censura_prop)
```

```
##
## 0 1
## 0.36 0.64
```

Note o valor 1 indica falha (óbito) e 0 indica censura. Os valores indicam a proporção de pacientes que foram a óbito ou censurados ao final do período de acompanhamento onde 64% dos pacientes vieram a óbito.

```
table(df$ANODIAG)
```

```
##
## 2013 2014 2015 2016
## 2387 2432 2317 2240
```

```
ANODIAG_prop <- round(prop.table(table(df$ANODIAG)),2)
print(ANODIAG_prop)
```

```
##
## 2013 2014 2015 2016
## 0.25 0.26 0.25 0.24
```

Note que não há um ano isolado com grande predominância de diagnósticos, indicando que o número de novos casos de câncer de estômago foi relativamente estável ao longo do período de 2013 a 2016. Essa estabilidade sugere que a distribuição anual de diagnósticos não sofreu grandes alterações no intervalo estudado.

```
table(df$SEXO)
```

```
##
##      0      1
## 5967 3409
```

```
SEXO_prop <- round(prop.table(table(df$SEXO)),2)
print(SEXO_prop)
```

```
##
##      0      1
## 0.64 0.36
```

Note que temos uma proporção maior de registros de câncer de estômago para homens do que para mulheres.

```
table(df$CIRURGIA)
```

```
##
##      0      1
## 4834 4542
```

```
CIRURGIA_prop <- round(prop.table(table(df$CIRURGIA)),2)
print(CIRURGIA_prop)
```

```
##
##      0      1
## 0.52 0.48
```

Note que a proporção de casos no qual houve intervenção cirurgica é similar a quantidade de casos onde não houve cirurgias.

```
table(df$RADIO)
```

```
##
##      0      1
## 7848 1528
```

```
RADIO_prop <- round(prop.table(table(df$RADIO)),2)
print(RADIO_prop)
```

```
##
##      0      1
## 0.84 0.16
```

Note que a quantidade de pacientes que fizeram radiografia é muito menor do que a quantidade de pacientes que não fizeram.

```
table(df$QUIMIO)
```

```
##
##      0      1
## 4322 5054
```

```
QUIMIO_prop <- round(prop.table(table(df$QUIMIO)),2)
print(QUIMIO_prop)
```

```
##
##      0      1
## 0.46 0.54
```

Note que a quantidade de pacientes que fizeram radiografia é similar a quantidade de pacientes que não fizeram.

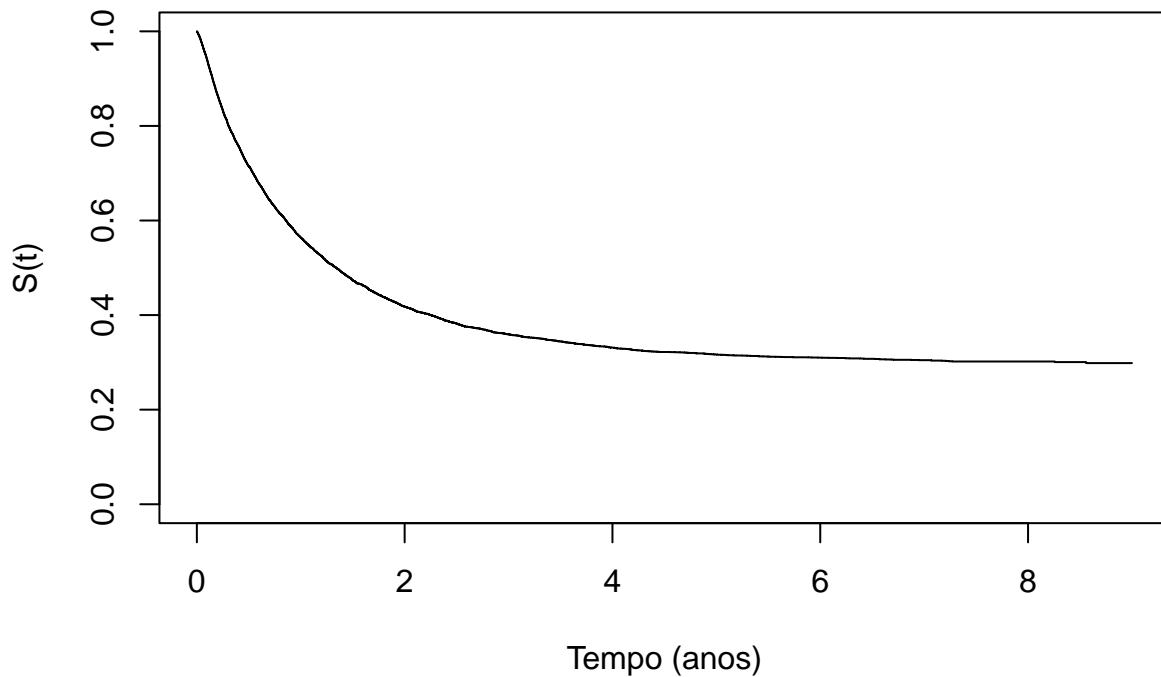
- (ii) faça o gráfico com a estimativa de Kaplan-Meier para a função de Sobrevida sem considerar covariáveis. Faça comentários sobre o gráfico.

Solução:

```
library(survival)
# Ajuste do modelo Kaplan-Meier sem covariáveis
ekm <- survfit(Surv(TEMPO, CENSURA) ~ 1, data = df)

# Gráfico
plot(ekm,
     ylab = "S(t)",
     xlab = "Tempo (anos)",
     main = "Estimativa de Sobrevida - Kaplan-Meier (Global)",
     mark.time = FALSE,
     conf.int = FALSE)
```

Estimativa de Sobrevivencia – Kaplan–Meier (Global)



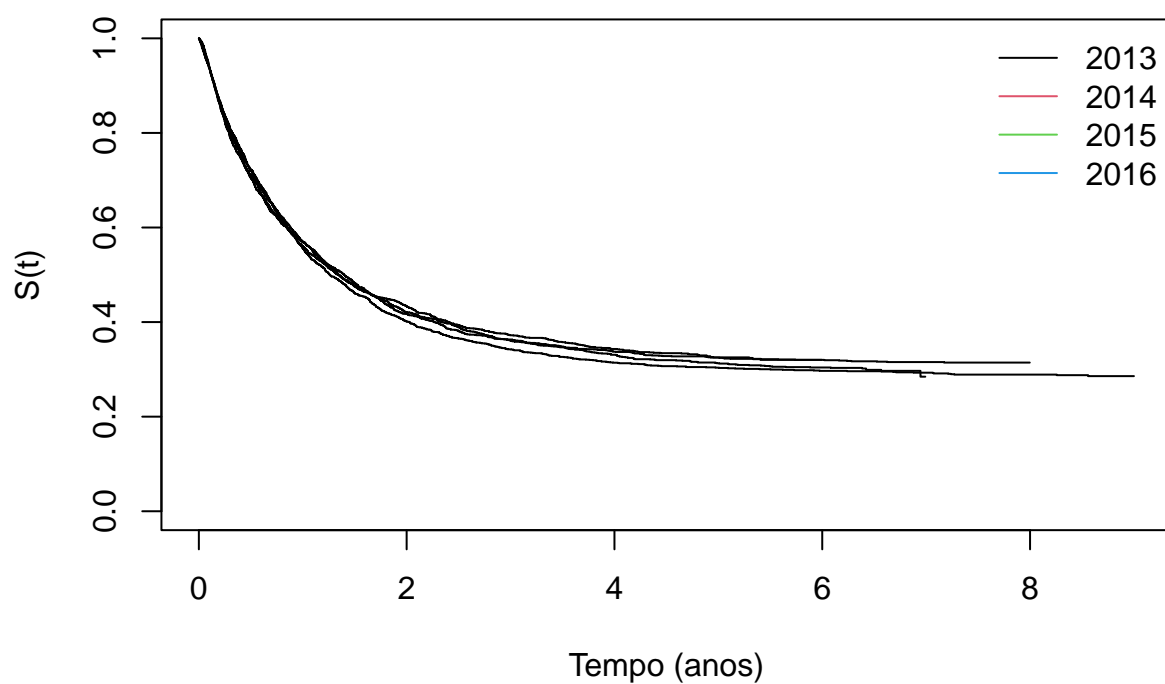
O gráfico acima representa a probabilidade acumulada de Sobrevivencia ao longo do tempo (em anos) para pacientes diagnosticados com neoplasia maligna do estômago (CID C16). A análise de Kaplan-Meier revela um alto risco de mortalidade nos primeiros anos após o diagnóstico de câncer de estômago. Observe que a curva se estabiliza com o tempo, indicando que pacientes que sobrevivem aos primeiros anos têm maiores chances de Sobrevivencia a longo prazo. Isso reforça a importância de intervenções precoces no tratamento e monitoramento dos pacientes com esse tipo de câncer.

- (iii) construa o gráfico com a estimativa de Kaplan-Meier para a função de Sobrevivencia considerando as covariáveis ANODIAG, SEXO, CIRURGIA, RADIO e QUIMIO, uma de cada vez. Faça comentários sobre os gráficos.

##Solução: 1. ANODIAG

```
ekm_anodiag <- survfit(Surv(TEMPO, CENSURA) ~ ANODIAG, data=df)
plot(ekm_anodiag, xlab= "Tempo (anos)", ylab="S(t)", mark.time = F,
     conf.int = F, main="Sobrevivencia - ANODIAG")
legend("topright", legend = levels(df$ANODIAG), lty = 1,
     col = 1:length(levels(df$ANODIAG)), bty = "n")
```

Sobrevivencia – ANODIAG

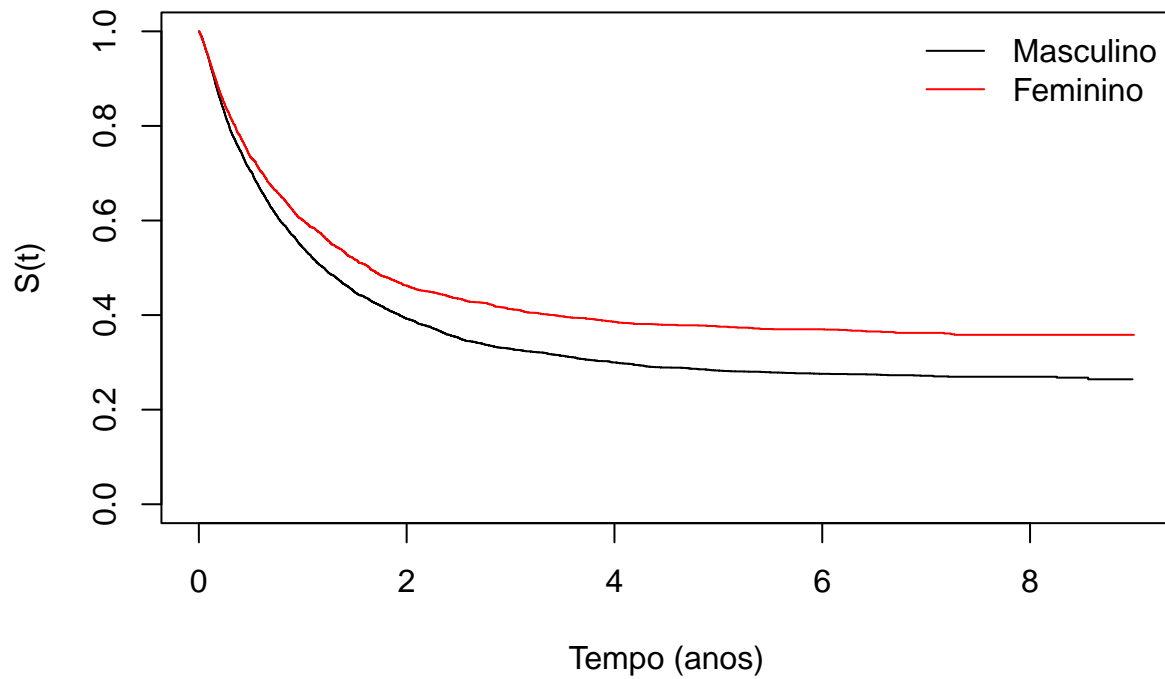


Com base no gráfico acima temos que o ano de diagnóstico não parece ter um impacto relevante na sobrevivência dos pacientes.

2. SEXO

```
ekm_sexo <- survfit(Surv(TEMPO, CENSURA) ~ SEXO, data=df)
plot(ekm_sexo, xlab= "Tempo (anos)", ylab="S(t)", mark.time = F,
     conf.int = F, col=c("black", "red"), main="Sobrevivencia - Sexo")
legend("topright", legend = c("Masculino", "Feminino"), lty = 1,
     col = c("black", "red"), bty = "n")
```

Sobrevivencia – Sexo

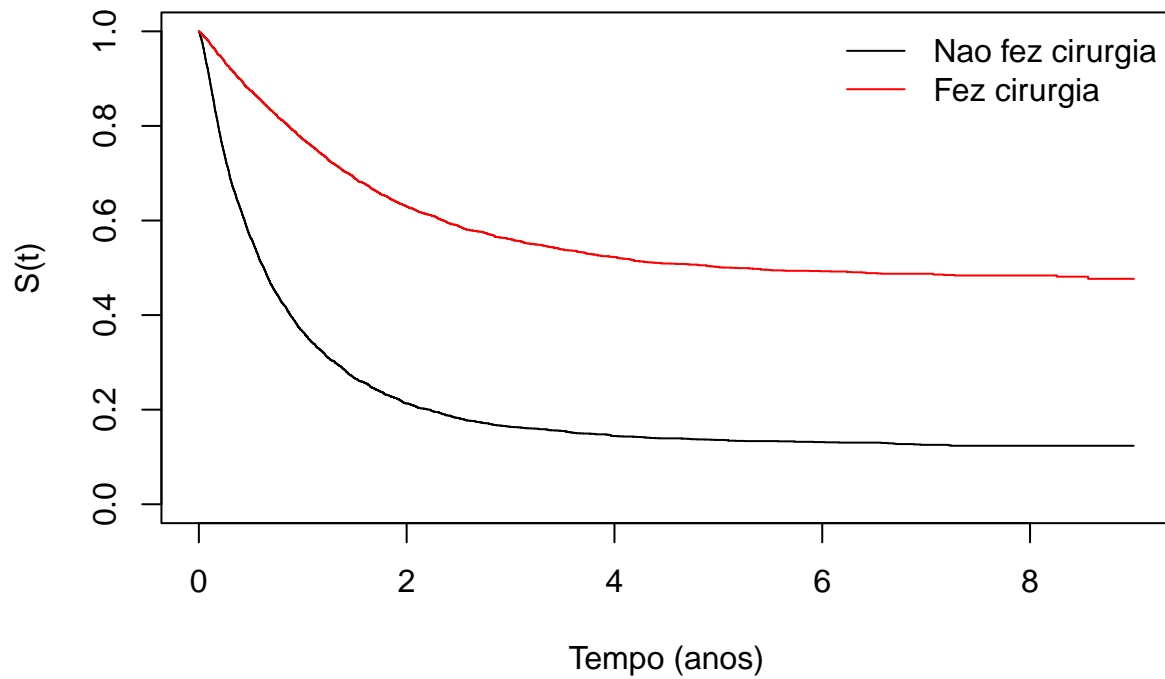


Com base no gráfico acima ser do sexo feminino está associado a uma maior chance de sobrevivência.

3. CIRURGIA

```
ekm_cirurgia <- survfit(Surv(TEMPO, CENSURA) ~ CIRURGIA, data=df)
plot(ekm_cirurgia, xlab= "Tempo (anos)", ylab="S(t)", mark.time = F,
     conf.int = F, col=c("black", "red"), main="Sobrevivencia - Cirurgia")
legend("topright", legend = c("Nao fez cirurgia", "Fez cirurgia"),
     lty = 1, col = c("black", "red"), bty = "n")
```

Sobrevivencia – Cirurgia

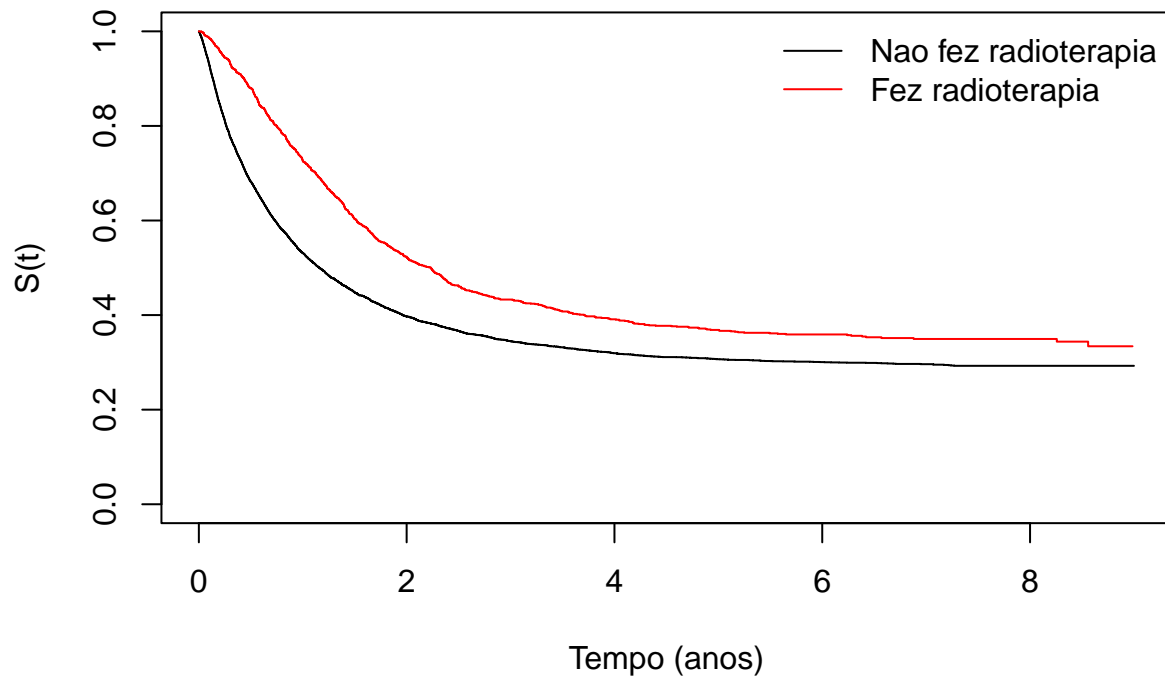


Note que a cirurgia é um fator protetor importante para a sobrevida dos pacientes.

4. RADIO

```
ekm_radio <- survfit(Surv(TEMPO, CENSURA) ~ RADIO, data=df)
plot(ekm_radio, xlab= "Tempo (anos)", ylab="S(t)", mark.time = F, conf.int = F,
     col=c("black", "red"), main="Sobrevivencia - Radioterapia")
legend("topright", legend = c("Nao fez radioterapia", "Fez radioterapia"),
     lty = 1, col = c("black", "red"), bty = "n")
```


Sobrevivencia – Radioterapia

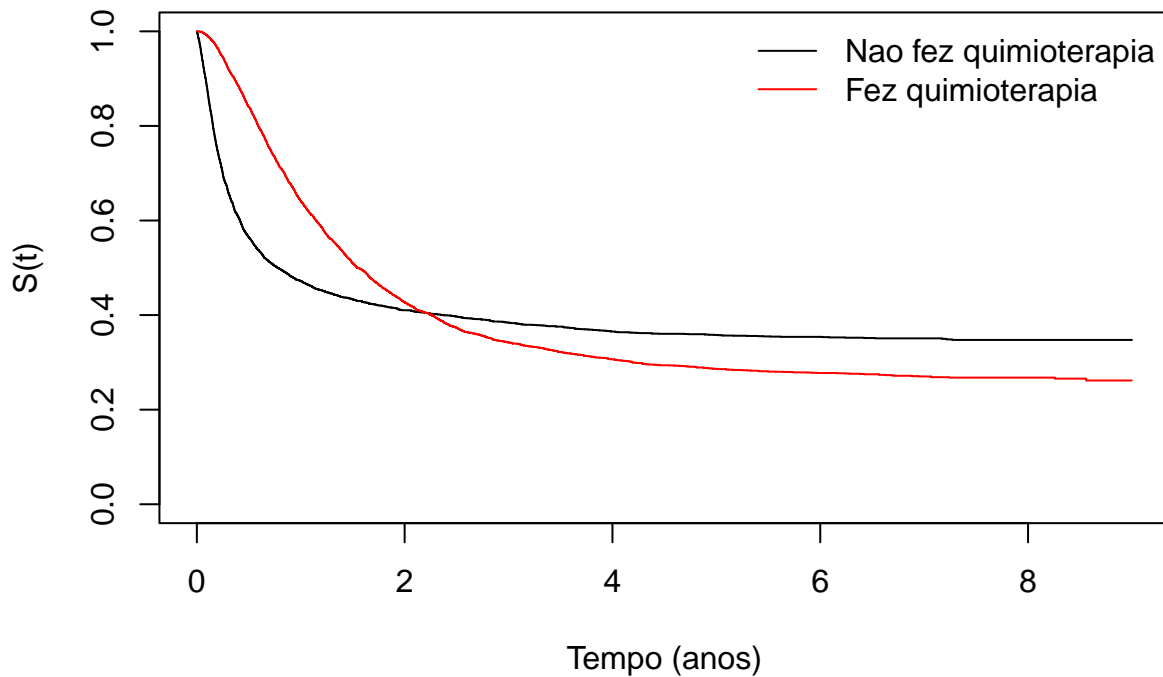


Note que a radioterapia pode ser benéfica para a sobrevida, mas é necessário avaliar em conjunto com outros fatores.

5. QUIMIO

```
ekm_quimio <- survfit(Surv(TEMPO, CENSURA) ~ QUIMIO, data=df)
plot(ekm_quimio, xlab= "Tempo (anos)", ylab="S(t)", mark.time = F,
     conf.int = F, col=c("black", "red"), main="Sobrevivencia - Quimioterapia")
legend("topright", legend = c("Nao fez quimioterapia", "Fez quimioterapia"),
     lty = 1, col = c("black", "red"), bty = "n")
```

Sobrevivencia – Quimioterapia



Note que a quimioterapia aparece associada a menor sobrevivência, mas isso pode estar relacionado a fatores de confusão, como o estágio da doença.

- (iv) realize o teste logrank considerando as covariáveis ANODIAG, SEXO, CIRURGIA, RADIO e QUIMIO, uma de cada vez. Descreva a conclusão do teste utilizando o p-valor.

Solução:

```
# Teste Logrank - ANODIAG
logrank_anodiag <- survdiff(Surv(TEMPO, CENSURA) ~ ANODIAG, data = df)
print(logrank_anodiag)
```

```
## Call:
## survdiff(formula = Surv(TEMPO, CENSURA) ~ ANODIAG, data = df)
##
##               N Observed Expected (O-E)^2/E (O-E)^2/V
## ANODIAG=2013 2387      1581      1529    1.7543    2.36140
## ANODIAG=2014 2432      1549      1580    0.6025    0.81930
## ANODIAG=2015 2317      1487      1484    0.0052    0.00692
## ANODIAG=2016 2240      1387      1411    0.3988    0.52292
##
## Chisq= 2.8  on 3 degrees of freedom, p= 0.4
```

```
# Teste Logrank - SEXO
```

```
logrank_sexo <- survdiff(Surv(TEMPO, CENSURA) ~ SEXO, data = df)
print(logrank_sexo)
```

```
## Call:
```

```
## survdiff(formula = Surv(TEMPO, CENSURA) ~ SEXO, data = df)
```

```
##
```

```
##           N Observed Expected (O-E)^2/E (O-E)^2/V
```

```
## SEXO=0 5967      4002      3704      23.9      62.6
```

```
## SEXO=1 3409      2002      2300      38.5      62.6
```

```
##
```

```
## Chisq= 62.6 on 1 degrees of freedom, p= 2e-15
```

```
# Teste Logrank - CIRURGIA
```

```
logrank_cirurgia <- survdiff(Surv(TEMPO, CENSURA) ~ CIRURGIA, data = df)
print(logrank_cirurgia)
```

```
## Call:
```

```
## survdiff(formula = Surv(TEMPO, CENSURA) ~ CIRURGIA, data = df)
```

```
##
```

```
##           N Observed Expected (O-E)^2/E (O-E)^2/V
```

```
## CIRURGIA=0 4834      3896      2240      1224      2031
```

```
## CIRURGIA=1 4542      2108      3764       728      2031
```

```
##
```

```
## Chisq= 2031 on 1 degrees of freedom, p= <2e-16
```

```
# Teste Logrank - RADIO
```

```
logrank_radio <- survdiff(Surv(TEMPO, CENSURA) ~ RADIO, data = df)
print(logrank_radio)
```

```
## Call:
```

```
## survdiff(formula = Surv(TEMPO, CENSURA) ~ RADIO, data = df)
```

```
##
```

```
##           N Observed Expected (O-E)^2/E (O-E)^2/V
```

```
## RADIO=0 7848      5090      4802      17.2      86.4
```

```
## RADIO=1 1528       914      1202      68.9      86.4
```

```
##
```

```
## Chisq= 86.4 on 1 degrees of freedom, p= <2e-16
```

```
# Teste Logrank - QUIMIO
```

```
logrank_quimo <- survdiff(Surv(TEMPO, CENSURA) ~ QUIMIO, data = df)
print(logrank_quimo)
```

```
## Call:
```

```
## survdiff(formula = Surv(TEMPO, CENSURA) ~ QUIMIO, data = df)
```

```
##
```

```
##           N Observed Expected (O-E)^2/E (O-E)^2/V
```

```
## QUIMIO=0 4322      2610      2385      21.2      35.4
```

```
## QUIMIO=1 5054      3394      3619      13.9      35.4
```

```
##
```

```
## Chisq= 35.4 on 1 degrees of freedom, p= 3e-09
```

Note que como não há diferença estatisticamente significativa entre as curvas de sobrevivência dos anos de diagnóstico 2013, 2014, 2015 e 2016 podemos interpretar que o ano em que o paciente foi diagnosticado não influencia significativamente a sua sobrevida no período estudado. Por outro lado, existe diferença estatisticamente significativa entre as curvas de sobrevivência de homens e mulheres indicando que o sexo influencia a sobrevida.

Existe diferença estatisticamente significativa nas curvas de sobrevivência entre pacientes que realizaram cirurgia e os que não realizaram. O mesmo acontece para Radiografia e Quimioterapia, indicando que pacientes que realizaram cirurgia têm uma probabilidade de sobrevivência significativamente maior do que os que não realizaram e fazer radioterapia está associado a maior sobrevida. No entanto, pacientes que não fizeram quimioterapia tiveram maior sobrevida. Isso pode indicar que os pacientes que receberam quimioterapia tinham casos mais graves, necessitando de tratamentos mais agressivos, o que não significa que a quimioterapia seja ineficaz, mas que os casos tratados com ela provavelmente tinham piores prognósticos.

Questão 3 Utilizando as covariáveis ANODIAG, SEXO, CIRURGIA, RADIO e QUIMIO, e os modelos exponencial, Weibull e log-normal, responda: (i) adotando o modelo exponencial, qual é o conjunto de covariáveis que apresenta melhor AIC, considerando a rotina *stepAIC* com *direction* = “both”?

Solução:

```
# Carregar os pacotes necessários
library(survival)
library(MASS)

# Ajuste do modelo inicial (com todas as variáveis)
modelo_inicial_exp <- survreg(Surv(TEMPO, CENSURA) ~ ANODIAG + SEXO + CIRURGIA + RADIO + QUIMIO,
                             data = df, dist = "exponential")

# Aplicar a seleção stepwise baseada no AIC
modelo_final_exp <- stepAIC(modelo_inicial_exp, direction = "both")

## Start:  AIC=23625.24
## Surv(TEMPO, CENSURA) ~ ANODIAG + SEXO + CIRURGIA + RADIO + QUIMIO
##
##           Df   AIC
## <none>      23625
## - QUIMIO    1 23639
## - ANODIAG    3 23639
## - SEXO       1 23720
## - RADIO      1 23742
## - CIRURGIA   1 26871

# Resumo do modelo final
summary(modelo_final_exp)

##
## Call:
## survreg(formula = Surv(TEMPO, CENSURA) ~ ANODIAG + SEXO + CIRURGIA +
##        RADIO + QUIMIO, data = df, dist = "exponential")
##           Value Std. Error      z      p
## (Intercept)  0.2963      0.0330  8.98 < 2e-16
## ANODIAG2014 -0.0266      0.0358 -0.74  0.4564
```

```
## ANODIAG2015 -0.0956      0.0361 -2.64  0.0082
## ANODIAG2016 -0.1467      0.0368 -3.98 6.8e-05
## SEX01       0.2672      0.0276  9.69 < 2e-16
## CIRURGIA1   1.5052      0.0272 55.35 < 2e-16
## RADIO1      0.3793      0.0362 10.47 < 2e-16
## QUIMIO1     0.1043      0.0266  3.92 8.7e-05
##
## Scale fixed at 1
##
## Exponential distribution
## Loglik(model)= -11804.6   Loglik(intercept only)= -13567.5
##  Chisq= 3525.85 on 7 degrees of freedom, p= 0
## Number of Newton-Raphson Iterations: 5
## n= 9376
```

```
# Verificar o AIC
AIC(modelo_final_exp)
```

```
## [1] 23625.24
```

Note que pelo resultados do Teste Logrank todas as variáveis, exceto ANODIAG, apresentaram impacto significativo na sobrevivência. Os resultados do modelo inicial compelto nos diz que pacientes diagnosticados em anos mais recentes (2015 e 2016) tendem a ter melhor sobrevida. Mulheres, quem fez cirurgia, e quem fez radioterapia têm melhor prognóstico. Quimioterapia está associada a pior sobrevida, mas provavelmente por estar ligada a casos mais graves. Note ainda que o modelo completo foi o selecionado pelo stepAIC, ou seja, nenhuma variável foi eliminada.

- (ii) adotando o modelo Weibull, qual é o conjunto de covariáveis que apresenta melhor AIC, considerando a rotina *stepAIC* com *direction* = “both”?

Solução:

```
# Ajuste do modelo inicial Weibull (com todas as covariáveis)
modelo_inicial_weibull <- survreg(Surv(TEMPO, CENSURA) ~ ANODIAG + SEXO + CIRURGIA + RADIO + QUIMIO,
                                data = df, dist = "weibull")
```

```
# Aplicar a seleção stepwise baseada no AIC
modelo_final_weibull <- stepAIC(modelo_inicial_weibull, direction = "both")
```

```
## Start:  AIC=22042.13
## Surv(TEMPO, CENSURA) ~ ANODIAG + SEXO + CIRURGIA + RADIO + QUIMIO
##
##           Df    AIC
## - ANODIAG   3 22039
## <none>      22042
## - QUIMIO    1 22061
## - SEXO      1 22107
## - RADIO     1 22117
## - CIRURGIA  1 24287
##
```

```
## Step: AIC=22038.99
## Surv(TEMPO, CENSURA) ~ SEXO + CIRURGIA + RADIO + QUIMIO
##
##           Df    AIC
## <none>      22039
## + ANODIAG   3 22042
## - QUIMIO    1 22058
## - SEXO      1 22103
## - RADIO     1 22115
## - CIRURGIA  1 24285
```

```
# Resumo do modelo final
summary(modelo_final_weibull)
```

```
##
## Call:
## survreg(formula = Surv(TEMPO, CENSURA) ~ SEXO + CIRURGIA + RADIO +
##           QUIMIO, data = df, dist = "weibull")
##           Value Std. Error      z      p
## (Intercept) 0.0251      0.0365  0.69    0.49
## SEXO1       0.3234      0.0402  8.05 8.2e-16
## CIRURGIA1   1.8678      0.0410 45.54 < 2e-16
## RADIO1      0.4509      0.0529  8.52 < 2e-16
## QUIMIO1     0.1763      0.0387  4.56 5.1e-06
## Log(scale)  0.3774      0.0104 36.32 < 2e-16
##
## Scale= 1.46
##
## Weibull distribution
## Loglik(model)= -11013.5 Loglik(intercept only)= -12224.6
## Chisq= 2422.15 on 4 degrees of freedom, p= 0
## Number of Newton-Raphson Iterations: 5
## n= 9376
```

```
# Verificar o AIC
AIC(modelo_final_weibull)
```

```
## [1] 22038.99
```

Veja que o resultado do stepAIC removeu a variável ANODIAG. Com isso o melhor conjunto de covariáveis para o modelo Weibull, de acordo com o AIC, é: SEXO, CIRURGIA, RADIO e QUIMIO.

- (iii) adotando o modelo log-normal, qual é o conjunto de covariáveis que apresenta melhor AIC, considerando a rotina *stepAIC* com *direction* = “both”?

Solução:

```
# Ajuste do modelo inicial Log-Normal (com todas as covariáveis)
modelo_inicial_lognorm <- survreg(Surv(TEMPO, CENSURA) ~ ANODIAG + SEXO + CIRURGIA + RADIO + QUIMIO,
                                data = df, dist = "lognormal")
```

```
# Aplicar a seleção stepwise baseada no AIC
```

```
modelo_final_lognorm <- stepAIC(modelo_inicial_lognorm, direction = "both")
```

```
## Start: AIC=20943.4
```

```
## Surv(TEMPO, CENSURA) ~ ANODIAG + SEXO + CIRURGIA + RADIO + QUIMIO
```

```
##
```

```
##           Df    AIC
```

```
## - ANODIAG   3 20939
```

```
## <none>      20943
```

```
## - SEXO      1 20990
```

```
## - RADIO     1 21014
```

```
## - QUIMIO    1 21159
```

```
## - CIRURGIA  1 22929
```

```
##
```

```
## Step: AIC=20938.84
```

```
## Surv(TEMPO, CENSURA) ~ SEXO + CIRURGIA + RADIO + QUIMIO
```

```
##
```

```
##           Df    AIC
```

```
## <none>      20939
```

```
## + ANODIAG   3 20943
```

```
## - SEXO      1 20985
```

```
## - RADIO     1 21009
```

```
## - QUIMIO    1 21154
```

```
## - CIRURGIA  1 22924
```

```
# Resumo do modelo final
```

```
summary(modelo_final_lognorm)
```

```
##
```

```
## Call:
```

```
## survreg(formula = Surv(TEMPO, CENSURA) ~ SEXO + CIRURGIA + RADIO +
```

```
##   QUIMIO, data = df, dist = "lognormal")
```

```
##           Value Std. Error      z      p
```

```
## (Intercept) -0.84387    0.03577 -23.59 < 2e-16
```

```
## SEX01        0.27573    0.03970   6.95 3.8e-12
```

```
## CIRURGIA1    1.76697    0.03869  45.67 < 2e-16
```

```
## RADIO1       0.45074    0.05312   8.49 < 2e-16
```

```
## QUIMIO1      0.58143    0.03887  14.96 < 2e-16
```

```
## Log(scale)   0.53719    0.00975  55.10 < 2e-16
```

```
##
```

```
## Scale= 1.71
```

```
##
```

```
## Log Normal distribution
```

```
## Loglik(model)= -10463.4   Loglik(intercept only)= -11655.5
```

```
##   Chisq= 2384.1 on 4 degrees of freedom, p= 0
```

```
## Number of Newton-Raphson Iterations: 4
```

```
## n= 9376
```

```
# Verificar o AIC
```

```
AIC(modelo_final_lognorm)
```

```
## [1] 20938.84
```

Observe que o melhor conjunto de covariáveis para o modelo log-normal, segundo o stepAIC, é: SEXO, CIRURGIA, RADIO e QUIMIO. Assim como no teste anterior, a variável ANODIAG foi eliminada, pois sua inclusão não contribui significativamente para explicar a variação na sobrevivência.

(iv) usando o critério AIC, qual o melhor modelo?

Solução:

Comparando os valores de AIC dos três modelos (Exponencial, Weibull e Log-Normal), observamos que:

- O **modelo log-normal** apresentou o menor AIC (**20938.84**), indicando o melhor ajuste aos dados.
- O modelo Weibull apresentou AIC intermediário (**22038.99**).
- O modelo Exponencial teve o maior AIC (**23625.24**).

Portanto, **o modelo log-normal é o mais adequado** para explicar a sobrevivência dos pacientes com câncer de estômago (CID C16), considerando as covariáveis selecionadas.

(v) faça o gráfico dos resíduos de Cox-Snell. O que se pode afirmar sobre a qualidade do ajuste do modelo aos dados?

Solução:

```
# Obter matriz de design
matriz_modelo <- model.matrix(~ SEXO + CIRURGIA + RADIO + QUIMIO, data = df)

# Calcular o preditor linear
x_beta <- matriz_modelo %*% modelo_final_lognorm$coefficients

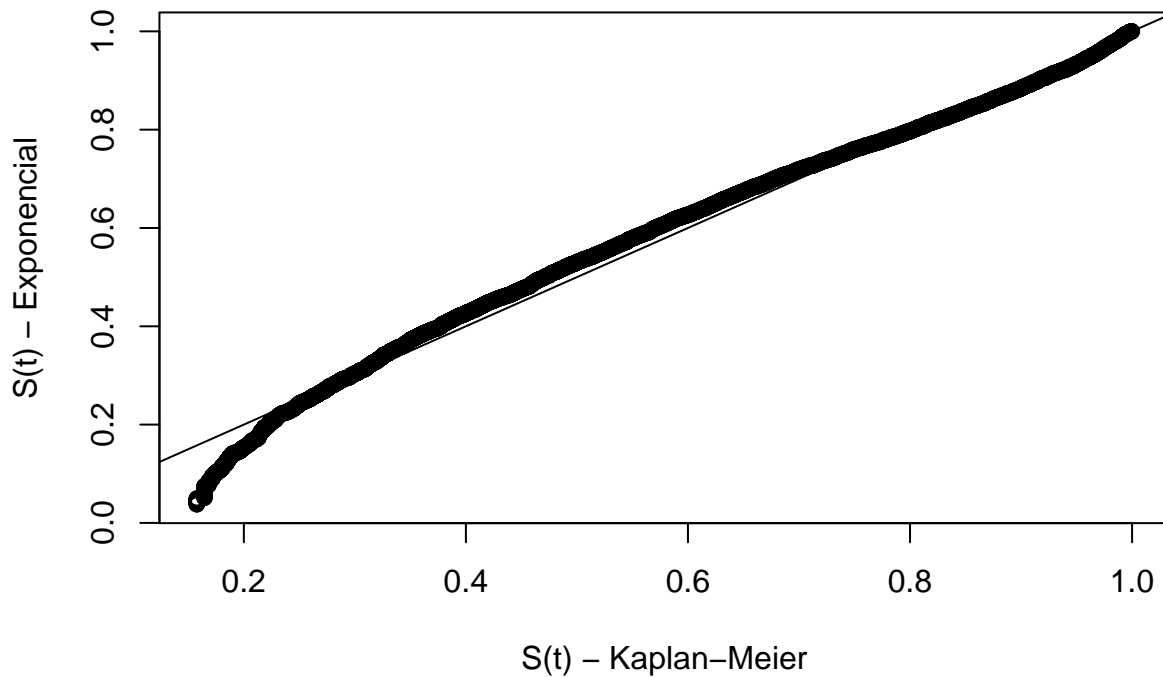
# Calcular o resíduo de Cox-Snell para o modelo log-normal
residuo_coxsnell <- -log(1 - pnorm((log(df$TEMPO) - x_beta) / modelo_final_lognorm$scale))

# Estimar a curva de sobrevivência dos resíduos
ekm_coxsnell <- survfit(Surv(residuo_coxsnell, df$CENSURA) ~ 1)

# Calcular a curva teórica da exponencial(1)
exp_coxsnell <- exp(- ekm_coxsnell$time)

# Gráfico
plot(ekm_coxsnell$surv, exp_coxsnell,
     main= "Resíduos de Cox-Snell - Modelo Log-Normal",
     xlab="S(t) - Kaplan-Meier", ylab="S(t) - Exponencial")
abline(0,1)
```


Resíduos de Cox-Snell – Modelo Log-Normal



O gráfico dos resíduos de Cox-Snell para o modelo log-normal mostra que os pontos estão alinhados de forma satisfatória com a linha de referência (reta de 45 graus). Portanto, podemos afirmar que o modelo log-normal apresenta um bom ajuste global aos dados de sobrevivência.

(vi) obtenha a estimativa da sobrevivência nos instantes de 1 e 10 anos, para dois pacientes com:

- paciente 1 - sexo masculino, não fez cirurgia, não fez radio, não fez quimio - $x=(1, 0, 0, 0, 0)$
- paciente 2 - sexo feminino, fez cirurgia, fez radio, fez quimio - $x=(1, 1, 1, 1, 1)$

Solução:

```
# Parâmetros do modelo log-normal final
coef_log <- modelo_final_lognorm$coefficients
scale_log <- modelo_final_lognorm$scale

# Paciente 1: Sexo masculino (0), não fez cirurgia (0), não fez rádio (0), não fez quimio (0)
x_p1 <- c(1, 0, 0, 0, 0)

# Paciente 2: Sexo feminino (1), fez cirurgia (1), fez rádio (1), fez quimio (1)
x_p2 <- c(1, 1, 1, 1, 1)

# Calcular o preditor linear (x_beta)
```

```

x_beta_p1 <- sum(x_p1 * coef_log)
x_beta_p2 <- sum(x_p2 * coef_log)

# Definir os tempos de interesse
tempos <- c(1, 10)

# Calcular a sobrevivência para cada tempo
S_p1 <- 1 - pnorm((log(tempos) - x_beta_p1) / scale_log)
S_p2 <- 1 - pnorm((log(tempos) - x_beta_p2) / scale_log)

# Exibir os resultados
data.frame(
  Tempo = tempos,
  Sobrevivencia_P1 = S_p1,
  Sobrevivencia_P2 = S_p2
)

```

```

##      Tempo Sobrevivencia_P1 Sobrevivencia_P2
## 1      1      0.31095496      0.9038444
## 2     10      0.03297634      0.4833181

```

O Paciente 1 tem uma baixa chance de sobrevivência a longo prazo. Apenas 3,3% de chance de sobreviver até 10 anos. No primeiro ano, a sobrevida é de 31,1%. Enquanto a Paciente 2 tem uma alta probabilidade de sobrevivência aos 1 ano (90,4%) e ainda mantém uma chance considerável aos 10 anos (48,3%).

(vii) calcule para os dois pacientes do item (vi) o MTTF.

Solução:

```

# Calcular o MTTF para Paciente 1 e Paciente 2 no modelo log-normal

# Parâmetros do modelo log-normal
coef_log <- modelo_final_lognorm$coefficients
scale_log <- modelo_final_lognorm$scale

# Paciente 1: Sexo masculino (0), não fez cirurgia (0), não fez rádio (0), não fez quimio (0)
x_p1 <- c(1, 0, 0, 0, 0)
mu_p1 <- sum(x_p1 * coef_log)
MTTF_p1 <- exp(mu_p1 + (scale_log^2) / 2)

# Paciente 2: Sexo feminino (1), fez cirurgia (1), fez rádio (1), fez quimio (1)
x_p2 <- c(1, 1, 1, 1, 1)
mu_p2 <- sum(x_p2 * coef_log)
MTTF_p2 <- exp(mu_p2 + (scale_log^2) / 2)

# Exibir resultados
data.frame(
  Paciente = c("Paciente 1", "Paciente 2"),
  MTTF = c(MTTF_p1, MTTF_p2)
)

```

```
##      Paciente      MTF
## 1 Paciente 1  1.859351
## 2 Paciente 2 40.249832
```

- Paciente 1: Sexo masculino, não fez cirurgia, não fez rádio, não fez quimio.

– Expectativa média de tempo até o evento (MTTF): 1.86 anos.

– Baixa expectativa de sobrevivência, refletindo a ausência de tratamentos e o fator de risco associado ao sexo masculino.

- Paciente 2: Sexo feminino, fez cirurgia, fez rádio, fez quimio.

– Expectativa média de tempo até o evento (MTTF): 40.25 anos.

– Expectativa de sobrevivência muito maior, indicando o impacto positivo do tratamento (cirurgia, radioterapia, quimioterapia) e o efeito protetor do sexo feminino.

Questão 4 Utilizando as covariáveis ANODIAG, SEXO, CIRURGIA, RADIO e QUIMIO, e o modelo de Cox, responda:

- faça o ajuste do modelo de Cox considerando todas as covariáveis.

Solução:

```
# Ajuste do modelo de Cox com todas as covariáveis
modelo_cox <- coxph(Surv(TEMPO, CENSURA) ~ ANODIAG + SEXO + CIRURGIA + RADIO + QUIMIO, data = df)

# Resumo do modelo
summary(modelo_cox)
```

```
## Call:
## coxph(formula = Surv(TEMPO, CENSURA) ~ ANODIAG + SEXO + CIRURGIA +
##      RADIO + QUIMIO, data = df)
##
##      n= 9376, number of events= 6004
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## ANODIAG2014 -0.03684   0.96383  0.03577  -1.030   0.303
## ANODIAG2015 -0.03050   0.96996  0.03616  -0.843   0.399
## ANODIAG2016 -0.03998   0.96081  0.03685  -1.085   0.278
## SEX01       -0.20721   0.81285  0.02754  -7.524 5.31e-14 ***
## CIRURGIA1    -1.19532   0.30261  0.02774 -43.097 < 2e-16 ***
## RADIO1       -0.29532   0.74429  0.03627  -8.143 3.87e-16 ***
## QUIMIO1      -0.22669   0.79717  0.02661  -8.519 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## ANODIAG2014   0.9638      1.038   0.8986   1.0338
## ANODIAG2015   0.9700      1.031   0.9036   1.0412
## ANODIAG2016   0.9608      1.041   0.8939   1.0328
```

```
## SEX01          0.8128      1.230      0.7701      0.8579
## CIRURGIA1      0.3026      3.305      0.2866      0.3195
## RADIO1         0.7443      1.344      0.6932      0.7991
## QUIMIO1        0.7972      1.254      0.7567      0.8398
##
## Concordance= 0.691 (se = 0.003 )
## Likelihood ratio test= 2157 on 7 df,  p=<2e-16
## Wald test          = 2016 on 7 df,  p=<2e-16
## Score (logrank) test = 2213 on 7 df,  p=<2e-16
```

O modelo de Cox ajustado com todas as covariáveis mostra que o SEXO (feminino), CIRURGIA, RADIO e QUIMIO são significativos e reduzem o risco de óbito. As categorias de ANODIAG não são estatisticamente significativas ($p > 0.05$), ou seja, o ano de diagnóstico não influencia significativamente o risco de óbito neste modelo.

- (ii) utilizando da rotina *stepAIC* com *direction* = “both”, obtenha o conjunto de covariáveis que apresenta melhor AIC.

Solução:

```
# Ajuste inicial do modelo de Cox (com todas as covariáveis)
modelo_inicial_cox <- coxph(Surv(TEMPO, CENSURA) ~ ANODIAG + SEXO + CIRURGIA + RADIO + QUIMIO, data = d)

# Aplicar a seleção stepwise baseada no AIC
modelo_final_cox <- stepAIC(modelo_inicial_cox, direction = "both")

## Start:  AIC=101317.5
## Surv(TEMPO, CENSURA) ~ ANODIAG + SEXO + CIRURGIA + RADIO + QUIMIO
##
##           Df    AIC
## - ANODIAG   3 101313
## <none>       101317
## - SEXO       1 101373
## - RADIO      1 101386
## - QUIMIO     1 101387
## - CIRURGIA   1 103287
##
## Step:  AIC=101313
## Surv(TEMPO, CENSURA) ~ SEXO + CIRURGIA + RADIO + QUIMIO
##
##           Df    AIC
## <none>       101313
## + ANODIAG   3 101317
## - SEXO       1 101369
## - RADIO      1 101381
## - QUIMIO     1 101383
## - CIRURGIA   1 103284

# Resumo do modelo final
summary(modelo_final_cox)
```

```
## Call:
## coxph(formula = Surv(TEMPO, CENSURA) ~ SEXO + CIRURGIA + RADIO +
##       QUIMIO, data = df)
##
##      n= 9376, number of events= 6004
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## SEX01      -0.20730   0.81277  0.02753  -7.529 5.10e-14 ***
## CIRURGIA1  -1.19577   0.30247  0.02773 -43.117 < 2e-16 ***
## RADIO1     -0.29463   0.74481  0.03626  -8.126 4.45e-16 ***
## QUIMIO1    -0.22697   0.79694  0.02660  -8.534 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## SEX01              0.8128      1.230    0.7701    0.8578
## CIRURGIA1          0.3025      3.306    0.2865    0.3194
## RADIO1             0.7448      1.343    0.6937    0.7997
## QUIMIO1            0.7969      1.255    0.7565    0.8396
##
## Concordance= 0.692 (se = 0.003 )
## Likelihood ratio test= 2155  on 4 df,  p=<2e-16
## Wald test              = 2014  on 4 df,  p=<2e-16
## Score (logrank) test = 2211  on 4 df,  p=<2e-16
```

O melhor conjunto de covariáveis no Modelo de Cox, considerando o critério AIC e a seleção stepAIC, inclui: SEXO, CIRURGIA, RADIO e QUIMIO. A variável ANODIAG foi eliminada por não contribuir significativamente para o modelo. Note que modelo final é mais simples e apresenta um AIC reduzido, indicando melhor ajuste aos dados.

- (iii) a partir do conjunto de covariáveis obtido do item (ii), faça o gráfico com os resíduos de Cox-Snell. Comente o resultado.

Solução:

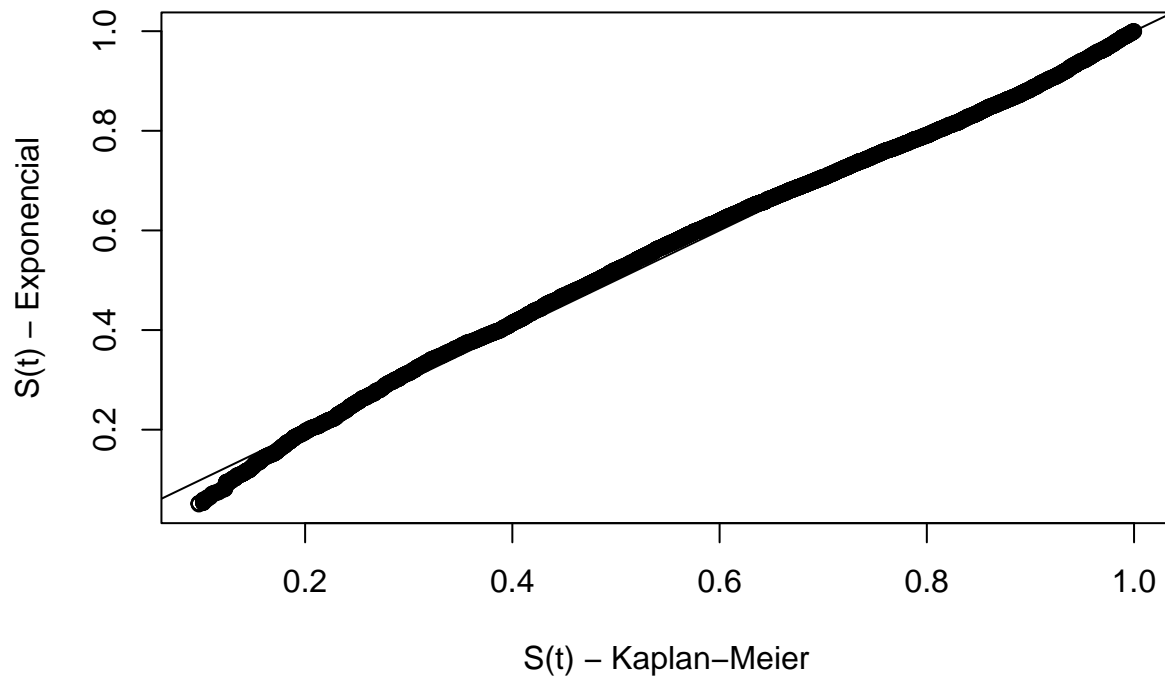
```
res_coxsnell <- df$CENSURA - residuals(modelo_final_cox, type = "martingale")

# Ajustar a curva de sobrevivência para os resíduos
ekm_coxsnell <- survfit(Surv(res_coxsnell, df$CENSURA) ~ 1)

# Calcular a curva teórica de referência (exponencial(1))
exp_coxsnell <- exp(-ekm_coxsnell$time)

# Gráfico
plot(ekm_coxsnell$surv, exp_coxsnell,
     main = "Resíduos de Cox-Snell - Modelo de Cox Final",
     xlab = "S(t) - Kaplan-Meier", ylab = "S(t) - Exponencial")
abline(0, 1)
```

Resíduos de Cox–Snell – Modelo de Cox Final



Note que como não há indícios de grandes desvios sistêmicos, e o modelo é adequado para explicar a sobrevivência dos pacientes com neoplasia maligna do estômago (CID C16).

- (iv) também utilizando do conjunto de covariáveis obtido do item (ii), verifique a adequação do modelo utilizando o teste de hipóteses do Resíduo de Schoenfeld. Comente o resultado.

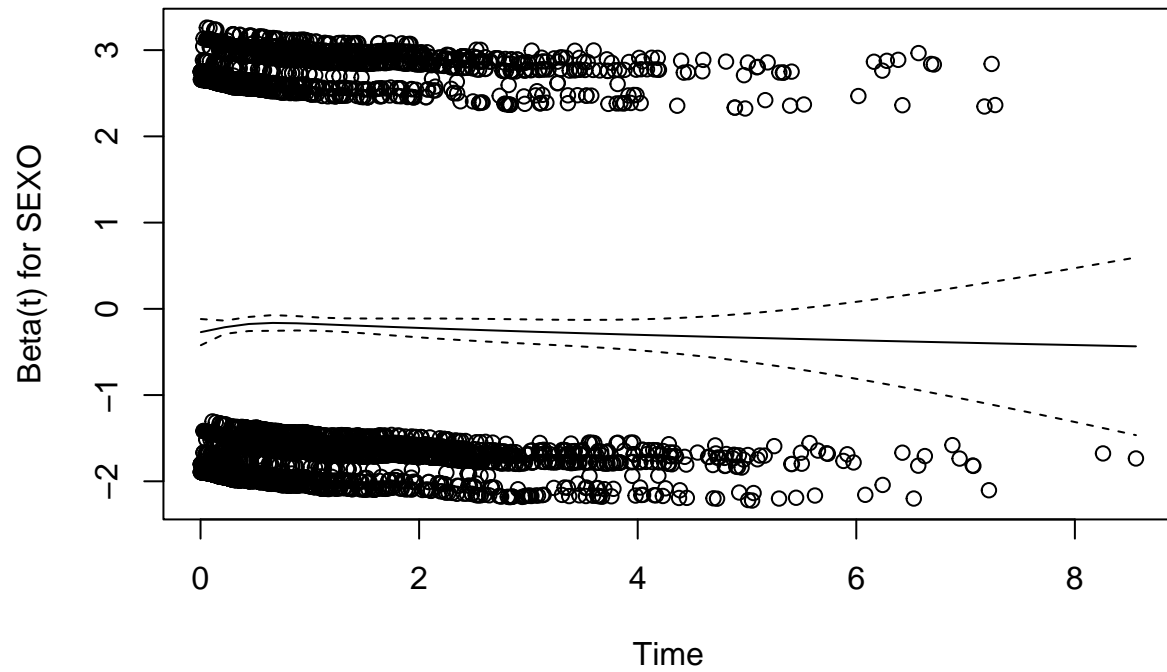
Solução:

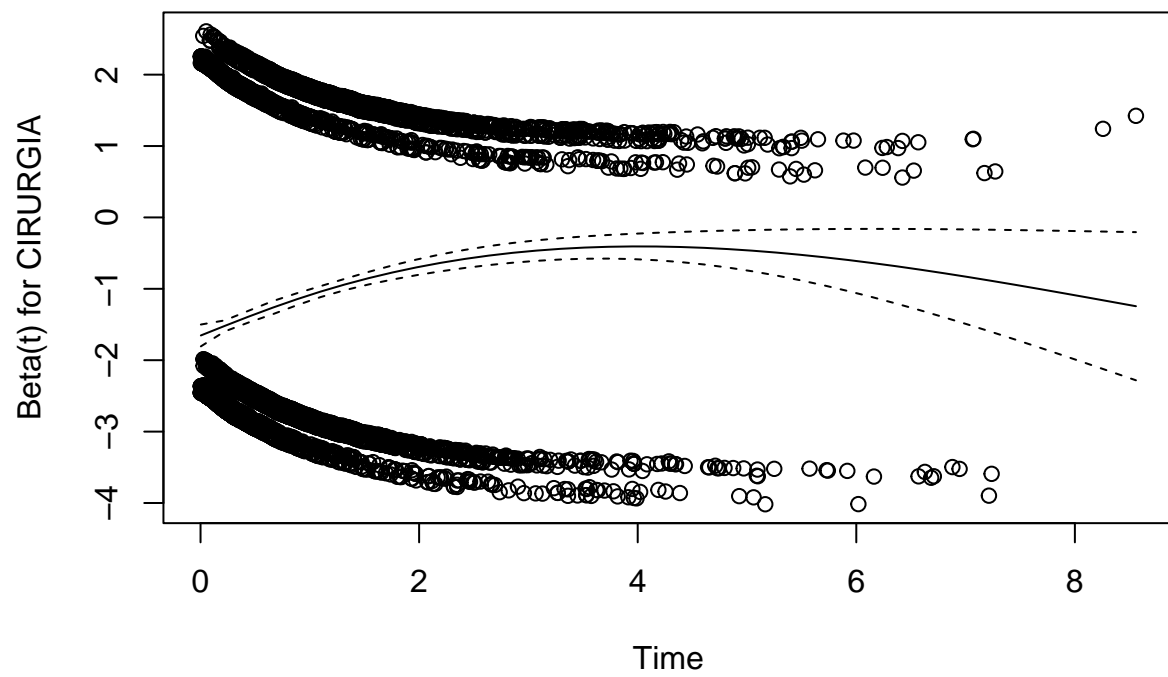
```
# Teste de proporcionalidade (Schoenfeld) para o modelo de Cox final
teste_schoenfeld <- cox.zph(modelo_final_cox, transform = "identity")

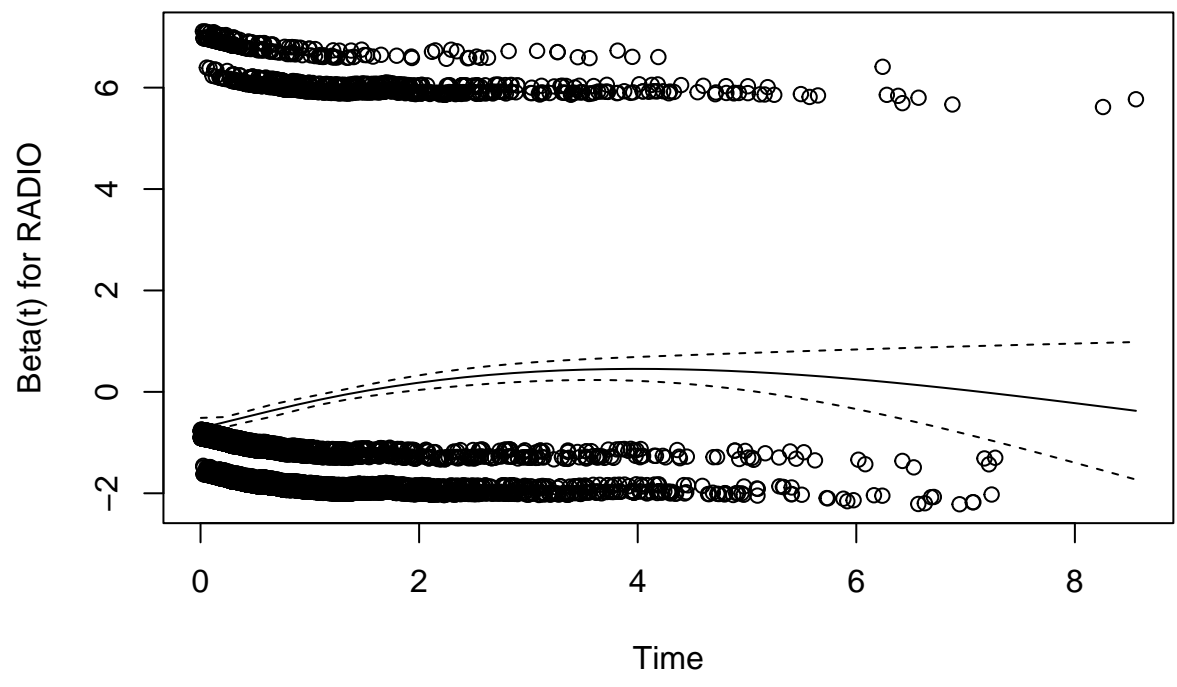
# Exibir o resumo do teste
print(teste_schoenfeld)
```

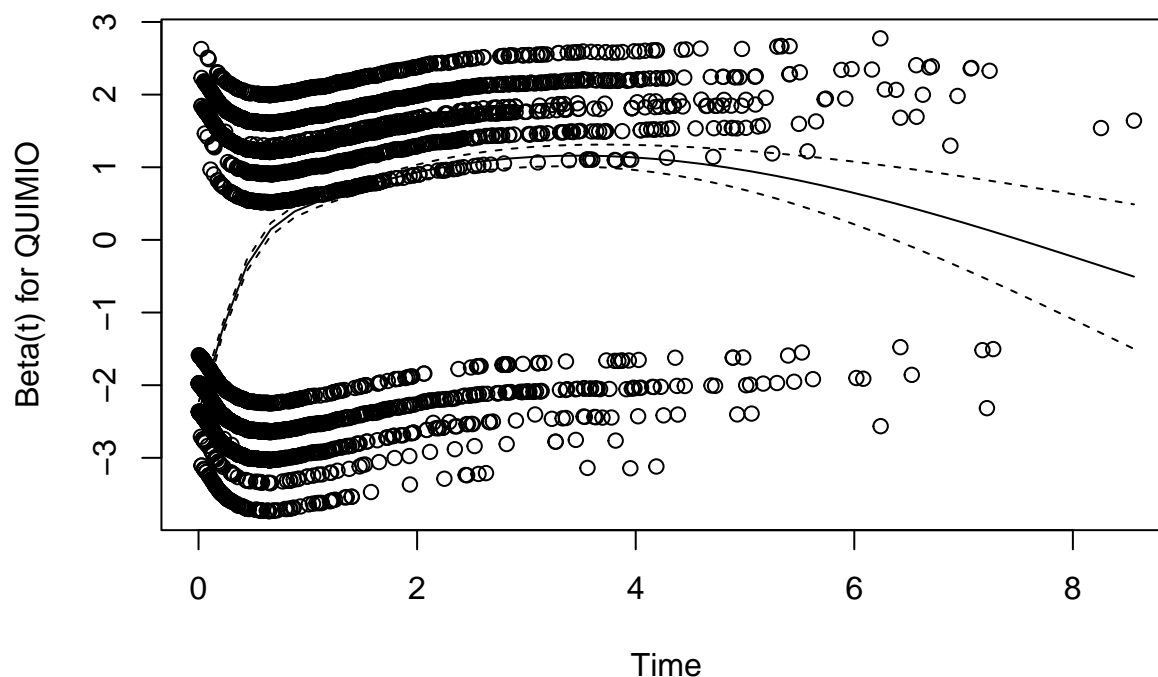
```
##          chisq df      p
## SEXO      11.0  1 0.00094
## CIRURGIA   74.6  1 < 2e-16
## RADIO     138.9  1 < 2e-16
## QUIMIO     835.0  1 < 2e-16
## GLOBAL    986.2  4 < 2e-16
```

```
# Plotar os gráficos para cada covariável  
plot(teste_schoenfeld)
```









Note que pelo Teste de Hipóteses do Resíduo de Schoenfeld temos que todos os p-valores são menores que 0.05, incluindo o teste global. Isso significa que a hipótese de riscos proporcionais é rejeitada para todas as covariáveis do modelo. Portanto, o modelo de Cox final não atende ao pressuposto de proporcionalidade dos riscos.

Gráficos dos Resíduos de Schoenfeld mostram que as linhas contínuas (estimativas) não são planas e apresentam tendência ao longo do tempo. Isso reforça a rejeição da hipótese de proporcionalidade. Curvas inclinadas ou com padrões sistemáticos indicam que o efeito da covariável varia com o tempo, violando o pressuposto de riscos proporcionais.

Com isso podemos concluir que o modelo de Cox final (SEXO, CIRURGIA, RADIO, QUIMIO) viola o pressuposto de riscos proporcionais, tanto pelo teste de Schoenfeld quanto pelos gráficos.