

# MAE0514 - Introdução a Análise de Sobrevivência - Lista 4

Bruno de Castro Paul Schultze<sup>1</sup>  
Rubens Santos Andrade Filho<sup>2</sup>

Julho de 2021

## Sumário

<b>Questão 1</b> . . . . .	<b>2</b>
<b>Questão 2</b> . . . . .	<b>2</b>
2.a . . . . .	2
<b>Questão 3</b> . . . . .	<b>2</b>
<b>Questão 4</b> . . . . .	<b>2</b>
a) . . . . .	2
b) . . . . .	3
c) . . . . .	4
d) . . . . .	5
<b>Questão 5</b> . . . . .	<b>5</b>
a) . . . . .	5
b) . . . . .	7
c) . . . . .	7
d) . . . . .	9
<b>Código Completo</b> . . . . .	<b>10</b>

---

<sup>1</sup>Número USP: 10736862

<sup>2</sup>Número USP: 10370336

## Questão 1

Breve resumo do artigo:

- (ii) Fabiani, M., Ramigni, M., Gobbetto, V., Mateo-Urdiales, A., Pezzotti, P., Piovesan, C. (2021). Effectiveness of the Comirnaty (BNT162b2, BioNTech/Pfizer) vaccine in preventing SARS-CoV-2 infection among healthcare workers, Treviso province, Veneto region, Italy, 27 December 2020 to 24 March 2021, *Euro Surveillance*, 26(17).

O objetivo do artigo foi estimar a eficácia da vacina Comirnaty (BNT162b2, BioNTech/Pfizer) na prevenção de infecção por SARS-CoV-2. Para isso, foi realizado um estudo de coorte retrospectivo no qual foram utilizados dados demográficos, características profissionais e datas de vacinação de 6.423 trabalhadores da área de saúde empregados na unidade local de saúde da província de Treviso em Veneto na Itália.

Primeiro, foram realizadas curvas de Kaplan-Meier para o tempo desde o início da campanha de vacinação até a infecção pelo vírus ou o fim do estudo para indivíduos não vacinados, que tomaram a primeira dose e ambas as doses da vacina. As curvas mostram que os pacientes que tomaram ambas as doses, e depois os que tomaram a primeira dose, apresentaram consistentemente uma menor probabilidade acumulada de infecção.

Também foi feita uma análise para o número de dias desde a administração da vacina para medir a duração do acompanhamento. A eficácia da vacina em diferentes intervalos de tempo foi estimada usando um modelo multivariável de risco proporcional de Cox, incluindo sexo, faixa etária, categoria profissional, contexto de trabalho e semana inicial de exposição como covariáveis.

A análise sugeriu que a vacina Comirnaty teve uma alta eficácia na prevenção da infecção por SARS-CoV-2 em HCW durante os intervalos de tempo após a administração em que a proteção pode ser esperada.

## Questão 2

2.a

## Questão 3

## Questão 4

a)

```
# Questão 4 ----
```

```
# t<0 representa censura à direita
```

```
t = c(28,59,89,175,195,309,-377, -393, -421, 447, 462,-709, -744, -1106, -1206,  
      34,88,137,-199, -280, 291,-299, -300, 308,351, 358,369,-370, 371,375,  
      -382, 392,-429, 451,-1119, 49,87,187,293,-302, -300, 353,-360, 402,-480,
```

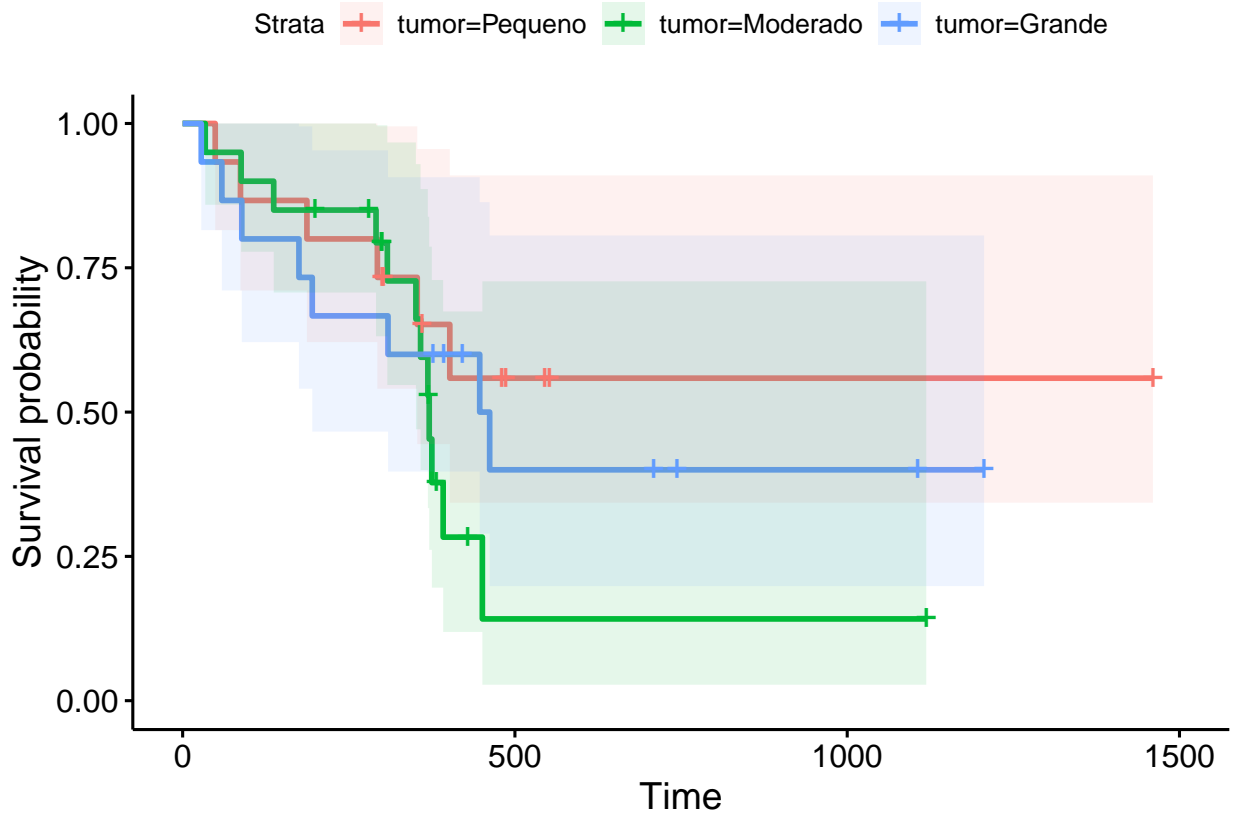
```

-486, -545, -552, -1460, -1460)

dados <- tibble(
  tempo=abs(t),
  delta=as.numeric(t>0),
  tumor=c(rep("Grande",15), rep("Moderado",20), rep("Pequeno",15))) %>%
  mutate(tumor=factor(tumor, levels = c("Pequeno","Moderado","Grande") ))

## 4a ----
library(survival)
library(survminer)
fit <- survfit(Surv(tempo, delta) ~ tumor, data = dados)
p=ggsurvplot(fit, data = dados, conf.int = T, conf.int.alpha=0.1)
p$plot

```



As curvas de Kaplan-Meier mostram que as estimativas da probabilidade de sobrevivência para o grupo com tumor moderado são menores a partir do dia 350. Entretanto, as curvas se cruzam por volta do dia 300 e todas as estimativas apresentam intervalos de confiança sobrepostos, o que indica que as curvas são próximas.

b)

```
## 4b ----
# logrank
survdifff(formula(fit),data=dados)

## Call:
## survdifff(formula = formula(fit), data = dados)
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## tumor=Pequeno 15         6    8.58    0.7773    1.177
## tumor=Moderado 20        12    8.87    1.1007    1.748
## tumor=Grande   15         8    8.54    0.0344    0.052
##
##  Chisq= 2  on 2 degrees of freedom, p= 0.4

# Harrington and Fleming com rho = 0.5
survdifff(formula(fit),data=dados, rho=0.5)
```

```
## Call:
## survdifff(formula = formula(fit), data = dados, rho = 0.5)
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## tumor=Pequeno 15     5.25    7.19  0.52342    0.9139
## tumor=Moderado 20     9.96    7.78  0.60816    1.1100
## tumor=Grande   15     6.90    7.14  0.00779    0.0136
##
##  Chisq= 1.3  on 2 degrees of freedom, p= 0.5
```

Realizamos os testes de *log-rank* e o teste da família Harrington and Fleming com  $\rho = 0.5$  e obtemos os valores p 0.4 e 0.5 respectivamente. Em ambos os testes, não rejeitamos a hipótese nula de que as curvas de sobrevivência são iguais entre os pacientes com tumor pequeno, moderado e grande, a um nível de significância de 5%.

c)

```
## 4c ----
fit2 <- coxph(Surv(tempo, delta) ~ tumor, data = dados)
summary(fit2)

## Call:
## coxph(formula = Surv(tempo, delta) ~ tumor, data = dados)
##
##      n= 50, number of events= 26
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## tumorModerado 0.6878   1.9893   0.5100  1.349   0.177
## tumorGrande   0.2938   1.3415   0.5408  0.543   0.587
##
##              exp(coef) exp(-coef) lower .95 upper .95
```

```
## tumorModerado      1.989      0.5027      0.7321      5.405
## tumorGrande        1.342      0.7454      0.4648      3.872
##
## Concordance= 0.54 (se = 0.056 )
## Likelihood ratio test= 1.98 on 2 df, p=0.4
## Wald test           = 1.94 on 2 df, p=0.4
## Score (logrank) test = 1.99 on 2 df, p=0.4
```

A partir da saída do modelo, interpretamos que

- O risco de óbito no grupo com tumor moderado é  $e^{\hat{\beta}_M} \approx 2$  vezes o risco no grupo com tumor pequeno.
- O risco de óbito no grupo com tumor moderado é  $(e^{\hat{\beta}_M} - 1)100\% \approx 34\%$  maior que o risco no grupo com tumor pequeno.

Entretanto as estimativas para o erro padrão das estimativas dos coeficientes de tumor moderado e grande são altas, e isso reflete que as estimativas intervalares de 95% de confiança para  $e^{\beta_M}$  e  $e^{\beta_G}$  contém o 1, ou, de forma equivalente, os valores p não são significativos a um nível de 5%. Dessa forma, não rejeitamos as hipóteses de que  $\beta_M = 0$  e  $\beta_G = 0$ .

d)

Testamos a hipótese  $H_0 : \beta = \mathbf{0}$  utilizando a estatística de teste da razão de verossimilhanças. A saída do teste se encontra no item c). O valor da estatística obtido foi de 1,98 com dois graus de liberdade e um valor-p de 0,4. Com isso, não rejeitamos a hipótese nula a um nível de 5%.

## Questão 5

a)

Primeiro, ajustando o modelo de Cox sem usar a idade:

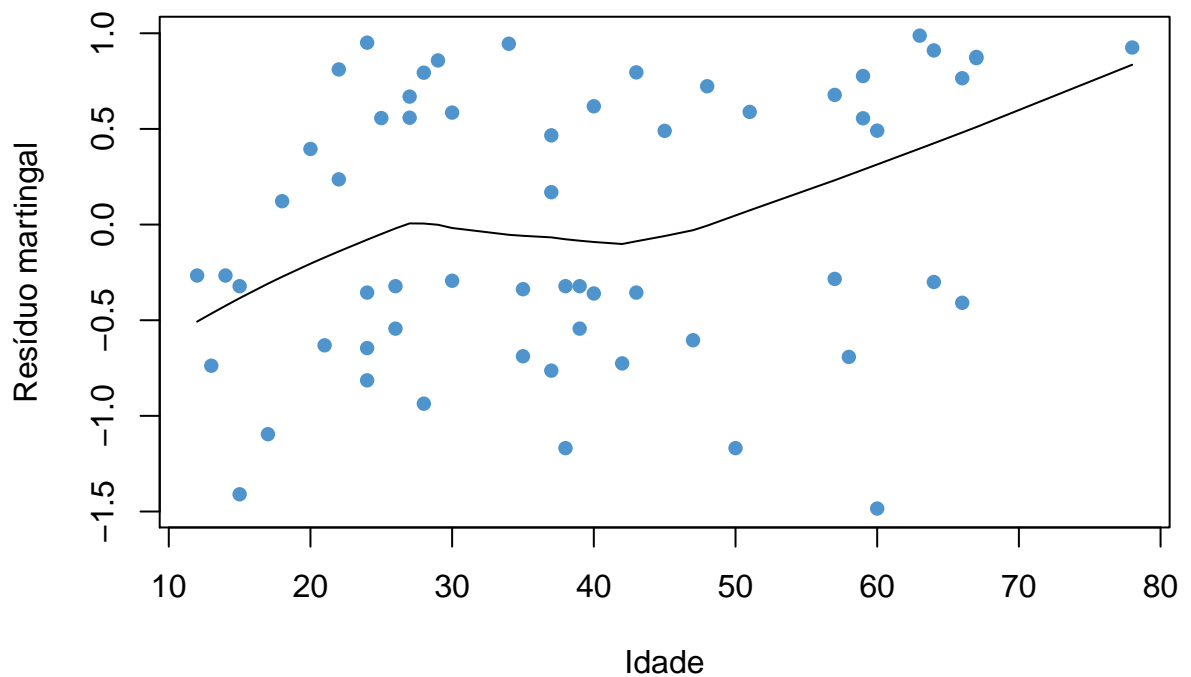
```
q5_cox_a = coxph(Surv(survivaltime, dead)~sex+stage+hist, data = q5)
summary(q5_cox_a)
```

```
## Call:
## coxph(formula = Surv(survivaltime, dead) ~ sex + stage + hist,
##       data = q5)
##
##      n= 60, number of events= 30
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## sex1      0.19046   1.20981  0.45891  0.415  0.67812
## stage1    0.76467   2.14828  0.40826  1.873  0.06107 .
## hist2     0.06433   1.06644  0.42182  0.153  0.87879
## hist3     1.66692   5.29586  0.55367  3.011  0.00261 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
##      exp(coef) exp(-coef) lower .95 upper .95
## sex1      1.210      0.8266   0.4921   2.974
## stage1     2.148      0.4655   0.9651   4.782
## hist2      1.066      0.9377   0.4665   2.438
## hist3      5.296      0.1888   1.7892  15.675
##
## Concordance= 0.667  (se = 0.051 )
## Likelihood ratio test= 12.43  on 4 df,   p=0.01
## Wald test               = 14.51  on 4 df,   p=0.006
## Score (logrank) test = 17.01  on 4 df,   p=0.002
```

Em seguida os resíduos martingal:

```
plot(q5$age, resid(q5_cox_a), xlab = 'Idade',
     ylab = 'Resíduo martingal', pch = 16, col = 'steelblue3')
smooth = lowess(q5$age, resid(q5_cox_a), iter = 1)
lines(smooth)
```



Observemos então que parece ser realmente uma tendência linear, indicando que transformações não são necessárias para essa variável no modelo.

b)

```
# 5.b
q5_cox_b = coxph(Surv(survivaltime, dead)~sex+stage+hist+age, data = q5)
summary(q5_cox_b)

## Call:
## coxph(formula = Surv(survivaltime, dead) ~ sex + stage + hist +
##       age, data = q5)
##
##      n= 60, number of events= 30
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## sex1      0.24416   1.27654  0.45492   0.537  0.5915
## stage1    0.84294   2.32320  0.41313   2.040  0.0413 *
## hist2    -0.13859   0.87058  0.42942  -0.323  0.7469
## hist3     1.30686   3.69455  0.59721   2.188  0.0287 *
## age       0.03028   1.03075  0.01241   2.441  0.0147 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## sex1            1.2765      0.7834   0.5234    3.114
## stage1          2.3232      0.4304   1.0338    5.221
## hist2           0.8706      1.1487   0.3752    2.020
## hist3           3.6946      0.2707   1.1461   11.910
## age             1.0307      0.9702   1.0060    1.056
##
## Concordance= 0.74 (se = 0.049 )
## Likelihood ratio test= 18.64 on 5 df,  p=0.002
## Wald test              = 21.75 on 5 df,  p=6e-04
## Score (logrank) test = 25.04 on 5 df,  p=1e-04
```

Notemos que pelos testes de Wald apenas *sex* não se mostra significativa ao nível de 5%.

c)

Aplicando o Teste de Razão de Verossimilhanças para avaliar a exclusão de cada variável individualmente temos, para Idade:

```
# 5.c
q5_cox_c_age = coxph(Surv(survivaltime, dead)~sex+stage+hist, data = q5)
anova(q5_cox_b, q5_cox_c_age)

## Analysis of Deviance Table
## Cox model: response is Surv(survivaltime, dead)
## Model 1: ~ sex + stage + hist + age
## Model 2: ~ sex + stage + hist
##      loglik  Chisq Df P(>|Chi|)
```

```
## 1 -101.29
## 2 -104.40 6.2124 1 0.01269 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Para Sexo:

```
q5_cox_c_sex = coxph(Surv(survivaltime, dead)~age+stage+hist, data = q5)
anova(q5_cox_b, q5_cox_c_sex)
```

```
## Analysis of Deviance Table
## Cox model: response is Surv(survivaltime, dead)
## Model 1: ~ sex + stage + hist + age
## Model 2: ~ age + stage + hist
##      loglik  Chisq Df P(>|Chi|)
## 1 -101.29
## 2 -101.44 0.2972 1 0.5856
```

Para Estágio:

```
q5_cox_c_stage = coxph(Surv(survivaltime, dead)~age+sex+hist, data = q5)
anova(q5_cox_b, q5_cox_c_stage)
```

```
## Analysis of Deviance Table
## Cox model: response is Surv(survivaltime, dead)
## Model 1: ~ sex + stage + hist + age
## Model 2: ~ age + sex + hist
##      loglik  Chisq Df P(>|Chi|)
## 1 -101.29
## 2 -103.55 4.5201 1 0.0335 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Para Histologia:

```
q5_cox_c_hist = coxph(Surv(survivaltime, dead)~age+stage+sex, data = q5)
anova(q5_cox_b, q5_cox_c_hist)
```

```
## Analysis of Deviance Table
## Cox model: response is Surv(survivaltime, dead)
## Model 1: ~ sex + stage + hist + age
## Model 2: ~ age + stage + sex
##      loglik Chisq Df P(>|Chi|)
## 1 -101.29
## 2 -104.03 5.478 2 0.06463 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Notemos então que ao nível de 5% somente as variáveis Idade e Estágio serão utilizadas no modelo final.



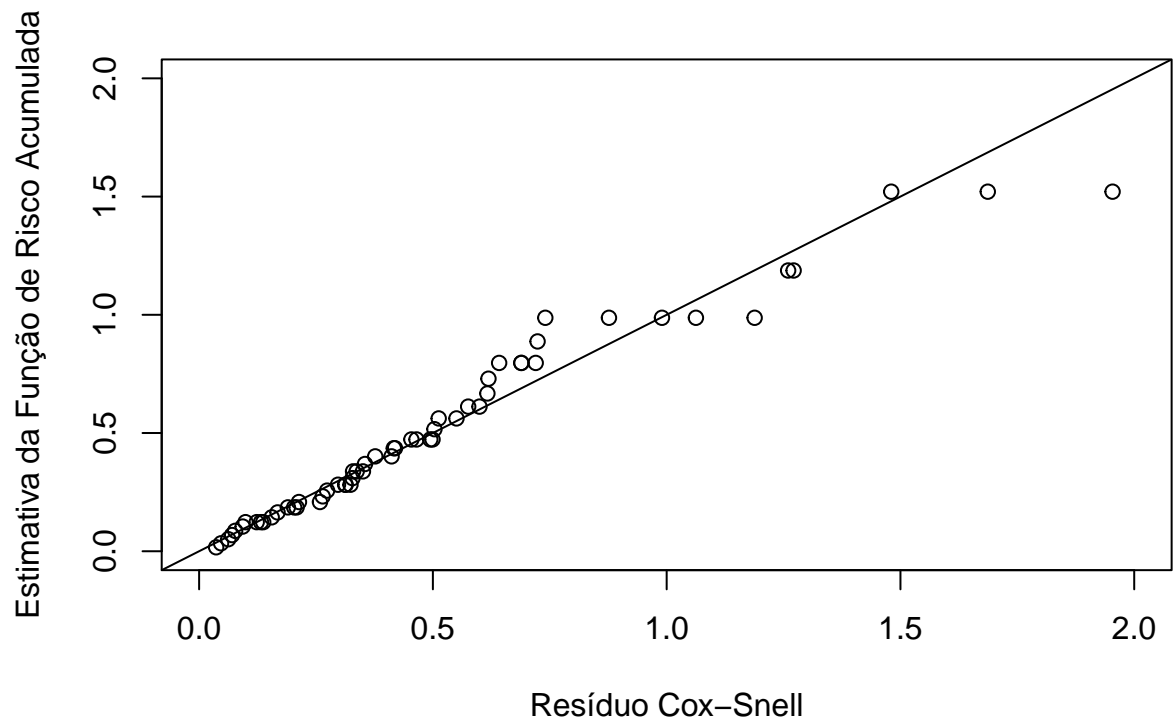
```
q5_cox = coxph(Surv(survivaltime, dead)~age+stage, data = q5)
summary(q5_cox)
```

```
## Call:
## coxph(formula = Surv(survivaltime, dead) ~ age + stage, data = q5)
##
##      n= 60, number of events= 30
##
##              coef exp(coef) se(coef)      z Pr(>|z|)
## age      0.03579   1.03644  0.01222  2.929   0.0034 **
## stage1  0.97067   2.63970  0.40653  2.388   0.0170 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## age              1.036      0.9648      1.012      1.062
## stage1           2.640      0.3788      1.190      5.856
##
## Concordance= 0.711 (se = 0.047 )
## Likelihood ratio test= 13.09 on 2 df,  p=0.001
## Wald test              = 12.1 on 2 df,  p=0.002
## Score (logrank) test = 12.71 on 2 df,  p=0.002
```

Notemos que ter Estágio igual a 1 leva a ter uma taxa de falha 164% maior do que pacientes com Estágio igual a 0. E a cada ano a mais na idade a taxa de falha é 3.6% maior comparada com o ano anterior.

d)

```
# 5.d
q5$cox_snell = - (resid(q5_cox) - q5$dead)
q5_cox_snell = coxph(Surv(cox_snell, dead)~1, data = q5)
q5_base = basehaz(q5_cox_snell, centered = FALSE)
plot(q5_base$time, q5_base$hazard, xlab = 'Resíduo Cox-Snell' ,
     ylab = 'Estimativa da Função de Risco Acumulada', xlim = c(0,2), ylim = c(0,2))
abline(0,1)
```



Notemos então que, com exceção dos último ponto à direita, os pontos se mantêm de forma considerável sobre a reta com intercepto 0 e inclinação 1 ( $x=y$ ), mostrando que o ajuste parece estar razoável.

## Código Completo

```
knitr::opts_chunk$set(warning=FALSE,
  # fig.dim = c(5,5),
  # out.height = '40%',
  # fig.align = 'center',
  message=FALSE
)

library(tidyverse)
library(ggplot2)
library(knitr)
library(readr)
library(dplyr)

# Questão 4 ----

# t<0 representa censura à direita
```

```

t = c(28,59,89,175,195,309,-377, -393, -421, 447, 462,-709, -744, -1106, -1206,
      34,88,137,-199, -280, 291,-299, -300, 308,351, 358,369,-370, 371,375,
      -382, 392,-429, 451,-1119, 49,87,187,293,-302, -300, 353,-360, 402,-480,
      -486, -545, -552, -1460, -1460)

dados <- tibble(
  tempo=abs(t),
  delta=as.numeric(t>0),
  tumor=c(rep("Grande",15), rep("Moderado",20), rep("Pequeno",15))) %>%
  mutate(tumor=factor(tumor, levels = c("Pequeno","Moderado","Grande") ))

## 4a ----
library(survival)
library(survminer)
fit <- survfit(Surv(tempo, delta) ~ tumor, data = dados)
p=ggsurvplot(fit, data = dados, conf.int = T, conf.int.alpha=0.1)
p$plot

## 4b ----
# logrank
survdifff(formula(fit),data=dados)
# Harrington and Fleming com rho = 0.5
survdifff(formula(fit),data=dados, rho=0.5)

## 4c ----
fit2 <- coxph(Surv(tempo, delta) ~ tumor, data = dados)
summary(fit2)

# QUESTAO 5 ----
# QUESTAO 5a ----
library(survival)
library(survminer)
library(readxl)

q5 = read_excel('data/List4_Hodgkins.xlsx')
q5$sex = as.factor(q5$sex)
q5$stage = as.factor(q5$stage)
q5$hist = as.factor(q5$hist)
q5
q5_cox_a = coxph(Surv(survivaltime, dead)~sex+stage+hist, data = q5)
summary(q5_cox_a)
plot(q5$age, resid(q5_cox_a), xlab = 'Idade',
     ylab = 'Resíduo martingal', pch = 16, col = 'steelblue3')
smooth = lowess(q5$age, resid(q5_cox_a), iter = 1)
lines(smooth)

# 5.b
q5_cox_b = coxph(Surv(survivaltime, dead)~sex+stage+hist+age, data = q5)
summary(q5_cox_b)

# 5.c
q5_cox_c_age = coxph(Surv(survivaltime, dead)~sex+stage+hist, data = q5)
anova(q5_cox_b, q5_cox_c_age)
q5_cox_c_sex = coxph(Surv(survivaltime, dead)~age+stage+hist, data = q5)
anova(q5_cox_b, q5_cox_c_sex)
q5_cox_c_stage = coxph(Surv(survivaltime, dead)~age+sex+hist, data = q5)
anova(q5_cox_b, q5_cox_c_stage)
q5_cox_c_hist = coxph(Surv(survivaltime, dead)~age+stage+sex, data = q5)

```

```

anova(q5_cox_b, q5_cox_c_hist)
q5_cox = coxph(Surv(survivaltime, dead)~age+stage, data = q5)
summary(q5_cox)
# 5.d
q5$cox_snell = - (resid(q5_cox) - q5$dead)
q5_cox_snell = coxph(Surv(cox_snell, dead)~1, data = q5)
q5_base = basehaz(q5_cox_snell, centered = FALSE)
plot(q5_base$time, q5_base$hazard, xlab = 'Resíduo Cox-Snell' ,
      ylab = 'Estimativa da Função de Risco Acumulada', xlim = c(0,2), ylim = c(0,2))
abline(0,1)

```