

# homework 7

joseferson da silva barreto

2022-07-16

**Questão 1** - Baseado no conjunto de dados atribuído, envie um relatório (em pdf) contendo os seguintes apontamentos:

a) A variável resposta escolhida;

\*\* o tempo até a morte por Câncer no Pulmão b)\*\* O conjunto de dados pronto para análise;

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.6      v purrr   0.3.4
## v tibble  3.1.7      v dplyr   1.0.9
## v tidyr   1.2.0      v stringr 1.4.0
## v readr   2.1.2      v forcats 0.5.1
```

```
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

```
library(plotly)
```

```
##
```

```
## Attaching package: 'plotly'
```

```
## The following object is masked from 'package:ggplot2':
```

```
##
```

```
##     last_plot
```

```
## The following object is masked from 'package:stats':
```

```
##
```

```
##     filter
```

```
## The following object is masked from 'package:graphics':
```

```
##
```

```
##     layout
```

```

library(readxl)

dados<-read.csv("analise_de_sobrevivencia.csv",sep = ";",encoding = "latin1")

library(lubridate)

##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##      date, intersect, setdiff, union

# (day( as.Date(dados$Data.do.Obito)) - day(as.Date( dados$Data.de.Diagnostico)))
#
#
# (( as.Date(dados$Data.do.Obito)) - (as.Date( dados$Data.de.Diagnostico)))

# (month( as.Date(dados$Data.do.Obito)) - month(as.Date( dados$Data.de.Diagnostico)))
#
# floor_date(as.Date(dados$Data.do.Obito),"month")
#
# mdy(dados$Data.do.Obito)

# dados %>%
# # dplyr::select(where(is.factor)) %>%
# glimpse()

pulmao<-dados %>%
  dplyr::filter(Descrição.da.Topografia == "PULMAO, SOE")

pulmao$falhas<-pulmao$Tipo.do.Obito
pulmao$falhas[pulmao$falhas == "CÂNCER"] <- "1"
pulmao$falhas[pulmao$falhas == "NÃO CÂNCER"] <- "0"
pulmao$falhas[pulmao$falhas == "SEM INFORMAÇÃO"] <- "0"
pulmao$falhas[pulmao$falhas == ""] <- "0"

table(pulmao$falhas)

##
##      0      1
## 916 3511

```

fazendo a análise descritiva

```

pulmao %>%
  dplyr::select(where(is.character)) %>%
  glimpse()

# =====
#   posiveis fatores
# =====
#
# # fatores
#
#   sexo
#   grau de instrução
#   Estado civil
#   nome.profissão
#   raca.cor
#   nacionalidade
#   cidade.endereço
#   Descrição.da.Doenca.Infantil
#   Código.da.Doenca.Infantil
#   Indicador.de.Caso.Raro
#   Meio.de.Diagnostico
#   Status.Vital
#   Metástase.à.distância

# pulmao %>%
#   dplyr::select(where(is.numeric)) %>%
#   glimpse()

#   obs
#   =====
#   Código.do.Paciente que está sendo identificada como uma variável número na verdade
#   é o do tipo texto ,ou seja não é um valor quantitativo
#   vamos converter essa variável
pulmao$Código.do.Paciente<-as.character(pulmao$Código.do.Paciente)

# codigo da profissão tbm é apenas um codigo

pulmao$Código.Profissão<-as.character(pulmao$Código.Profissão)

pulmao %>%
  dplyr::select(where(is.numeric)) %>%
  glimpse()

# fazendo o mesmo para código da morfologia

pulmao$Código.da.Morfologia<-as.character(pulmao$Código.da.Morfologia)

pulmao %>%

```

```
dplyr::select(where(is.numeric)) %>%  
glimpse()
```

```
#pulmao %>% select(Idade) %>% table()  
#pulmao<-read.csv("pulmaonovot.csv", sep = ",")
```

## Convertendo os possíveis fatores

```
#  sexo  
#  grau de instrução  
#  Estado civil  
#  nome.profissão  
#  raca.cor  
#  nacionalidade  
#  cidade.endereço  
#  Descrição.da.Doenca.Infantil  
#  Código.da.Doenca.Infantil  
#  Indicador.de.Caso.Raro  
#  Meio.de.Diagnostico  
#  Status.Vital  
#  Metástase.à.distância  
  
pulmao$Sexo<- as.factor(pulmao$Sexo)  
  
# glimpse(pulmao)  
  
# c(pulmao$Sexo, pulmao$Nacionalidade)<-lapply(c(pulmao$Sexo, pulmao$Nacionalidade), as.factor)  
  
pulmao$Nacionalidade<- as.factor(pulmao$Nacionalidade)  
pulmao$Raca.Cor<- as.factor(pulmao$Raca.Cor)  
pulmao$Naturalidade<- as.factor(pulmao$Naturalidade)  
  
pulmao$Grau.de.Instrução<- as.factor(pulmao$Grau.de.Instrução)  
  
pulmao$Estado.Civil<- as.factor(pulmao$Estado.Civil)  
  
pulmao %>% select(  
  Descrição.da.Morfologia  
) %>% table()  
  
pulmao$Descrição.da.Morfologia<- as.factor(pulmao$Descrição.da.Morfologia)  
#  
#  Descrição.da.Doenca  
# ) %>% table()  
  
pulmao$Descrição.da.Doenca<- as.factor(pulmao$Descrição.da.Doenca)  
pulmao$Descrição.da.Doenca.Adulto.Jovem<- as.factor(pulmao$Descrição.da.Doenca.Adulto.Jovem)  
pulmao$Descrição.da.Doenca.Infantil<- as.factor(pulmao$Descrição.da.Doenca.Infantil)  
pulmao$Indicador.de.Caso.Raro<- as.factor(pulmao$Indicador.de.Caso.Raro)
```

```
pulmao$Meio.de.Diagnostico<- as.factor(pulmao$Meio.de.Diagnostico)
pulmao$Extensão<- as.factor(pulmao$Extensão)
pulmao$Tipo.do.Obito<- as.factor(pulmao$Tipo.do.Obito)
pulmao$Nome.Profissão<- as.factor(pulmao$Nome.Profissão)
```

## Mostrando os dados

```
pulmao <- pulmao %>% filter(pulmao$Extensão != 'SEM INFORMAÇÃO' )
# pulmao %>%
#   head(10)
```

c) tipo de câncer escolhido para ser estudado;

**câncer no pulmão**

d) A estimativa da curva de Kaplan-Meier em formato gráfico para os tempos até a ocorrência do evento e também considerando cada uma das variáveis categóricas.

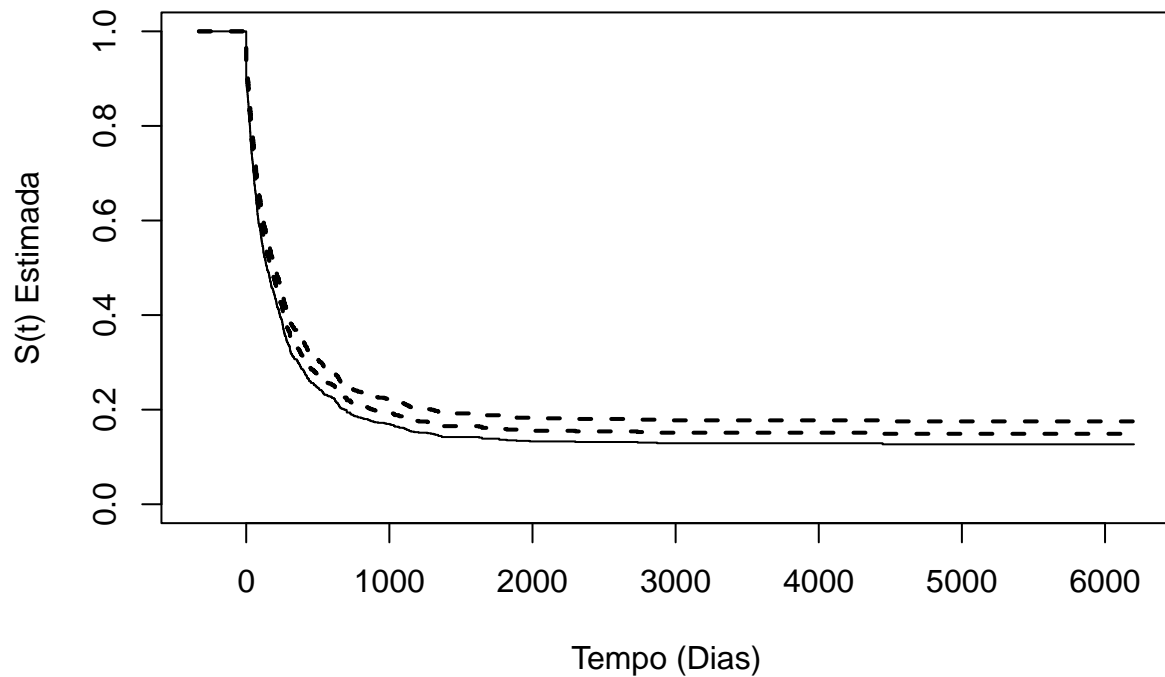
## grafico geral

```
library(survival)
pulmao$falhas<-as.numeric(pulmao$falhas)

ek <- survfit(Surv(pulmao$dias,pulmao$falhas)~1)

plot(ek, lty=c(2,1), xlab="Tempo (Dias)",
      ylab="S(t) Estimada", col = 1:1,
      lwd = c(2,1),main = "Curva de Sobrevida ")
```

## Curva de Sobrevivência



Podemos perceber que quanto maior for o número dias apos o diagnóstico do Câncer, menor a probabilidade de sobrevivência

## Fazendo com Base o fator Extensão

```
# Estimação da função de sobrevivência pelo método de Kaplan-Meier
library(survival)
pulmao$falhas<-as.numeric(pulmao$falhas)
ekm_exp <- survfit(Surv(pulmao$dias,pulmao$falhas)~pulmao$Extensão)

# pulmao[28]<-as.factor(pulmao[28])
# resumo do Kaplan-Meier

#summary(ekm_exp)

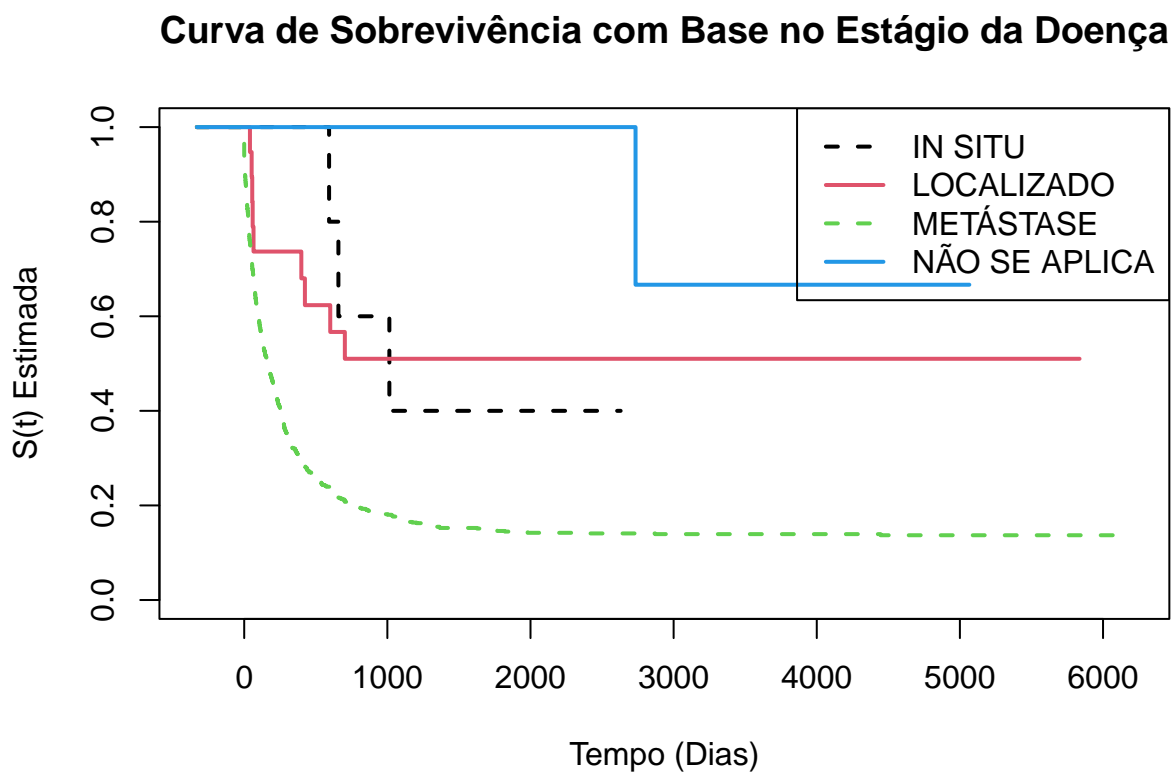
# Gráfico de Kaplan-Meier

# pulmao %>%
#   dplyr::select(where(is.factor)) %>%
```

```
# glimpse()

# table(pulmao$Extensao)

plot(ekm_exp, lty=c(2,1), xlab="Tempo (Dias)",
      ylab="S(t) Estimada", col = 1:4,
      lwd = c(2,2), main = "Curva de Sobrevivência com Base no Estágio da Doença")
legend(x = "topright", # Position
       c("IN SITU", "LOCALIZADO", "METÁSTASE", "NÃO SE APLICA"), # Legend texts
       lty = c(2,1), # Line types
       col = c(1:4), # Line colors
       lwd = 2)
```

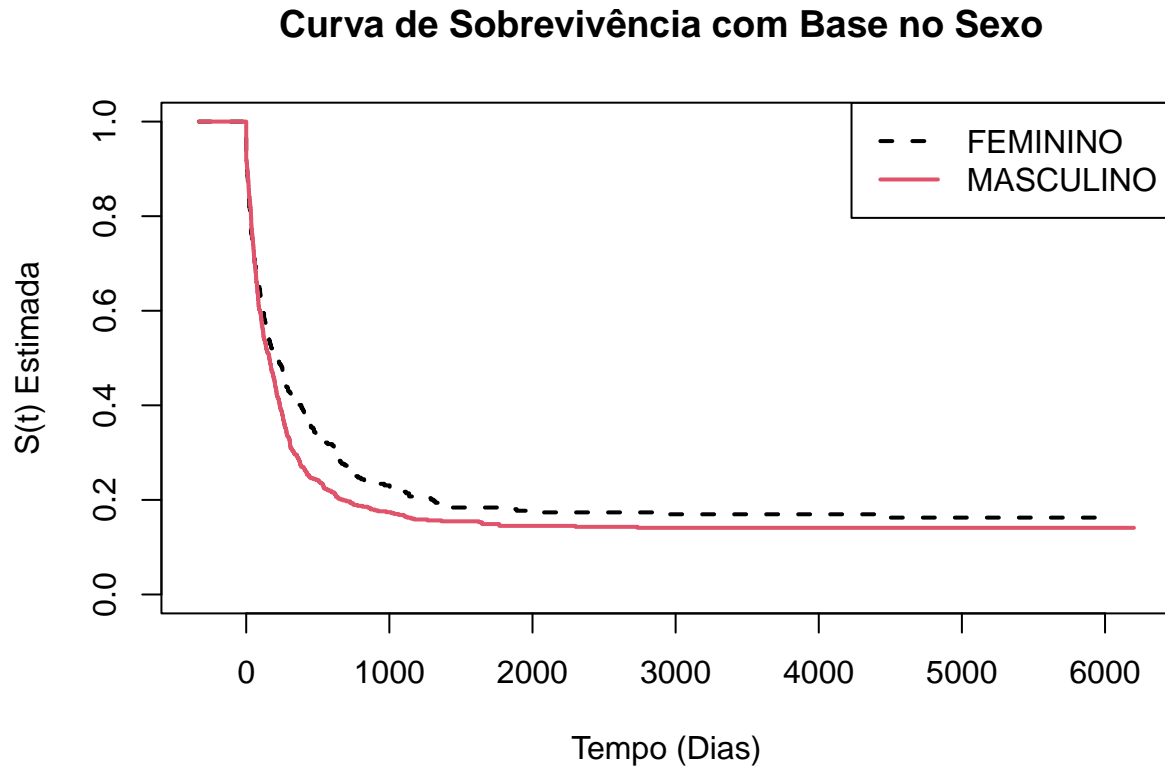


Logo podemos perceber que o grupo onde não se aplica o nível de estágio da doença tem maior chance de sobrevivência

## Fazendo com Base no Fator Sexo

```
pulmao$dias<-as.numeric(pulmao$dias)
pulmao$falhas<-as.numeric(pulmao$falhas)
ekm <- survfit(Surv(pulmao$dias,pulmao$falhas)~pulmao$Sexo)
#summary(ekm)
```

```
plot(ekm, lty=c(2,1), xlab="Tempo (Dias)",
     ylab="S(t) Estimada", col = 1:2,
     lwd = c(2,2), main = "Curva de Sobrevivência com Base no Sexo ")
legend(x = "topright",          # Position
       c("FEMININO", "MASCULINO"), # Legend texts
       lty = c(2,1),           # Line types
       col = c(1:2),           # Line colors
       lwd = 2)
```



Podemos perceber que as Pessoas que do sexo feminino após os 50-100 dias após o diagnóstico do câncer possuem maior probabilidade de sobrevivência que as pessoas do sexo masculino.