

ANÁLISE DE SOBREVIVÊNCIA DE PACIENTES COM CÂNCER DE PULMÃO

joseferson da silva barreto

2022-07-25

.

Professor: Cleanderson

Projeto final de análise de sobrevivência apresentado na
última aula da disciplina .

julho, 25, 2022

Sumário

- Objetivo
- Metodos
- Introdução
- Resultados
- Conclusão

Objetivo

O objetivo é avaliar e observar o tempo até a morte dos indivíduos por câncer no pulmão

Metodo

Foi Feito a Análise exploratória e os tratamentos dos dados para buscar relações entre as variáveis , a limpeza dos dados foi feita no software Rstudio ,utilizando a linguagem R para as demais análises. Iniciamos o tratamentos dos dados substituindo nas variáveis quantitativas onde tinha valores ausentes colocamos os valores da média de cada variavel respectivamente , em seguida foi feito alguns boxplot com o objetivo de observar principalmente o comportamento da média e mediana , por fim foi criado um modelo de regressão,executamos o teste tukey , criamos alguns gráficos para observar o comportamento dos residuos .

Introdução

A análise de sobrevivência é uma das áreas da estatística que mais cresceu nas últimas duas décadas do século passado. A razão deste crescimento é o desenvolvimento e aprimoramento de técnicas estatísticas combinado com computadores cada vez mais velozes. Uma evidência quantitativa deste sucesso é o número de aplicações de análise de sobrevivência em medicina.

O câncer de pulmão é uma das maiores causas de morte no mundo, sendo a primeira causa em morte por câncer entre homens no Brasil. A história natural da doença inclui elevada letalidade e evolução agressiva, quase sempre com o paciente chegando ao médico quando a doença já se encontra em fase avançada. Nos últimos 40 anos a taxa de sobrevida em neoplasias de pulmão melhorou em 8 %, nos Estados Unidos, apesar do avanço nas tecnologias de detecção precoce e no refinamento das técnicas de tratamento. Um desafio contínuo é descrever o papel das variáveis clínicas, assistenciais e demográficas dos pacientes com câncer de pulmão. Neste sentido, este trabalho avaliou uma amostra da população de pacientes acompanhados pelo Programa de Oncopneumologia, do Instituto de Doenças do Tórax e do Hospital Universitário Clementino Fraga Filho (UFRJ).

Carregando os pacotes

Vamos iniciar carregando os pacotes que iremos utilizar

```
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(tidyverse)

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")

# table(dados$Descrição.da.Doenca)
```

```
# 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"
```

```
##
## Attaching package: 'lubridate'

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

```
#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)

pulmao <- pulmao %>% filter(pulmao$Extensão !='SEM INFORMAÇÃO' )

```

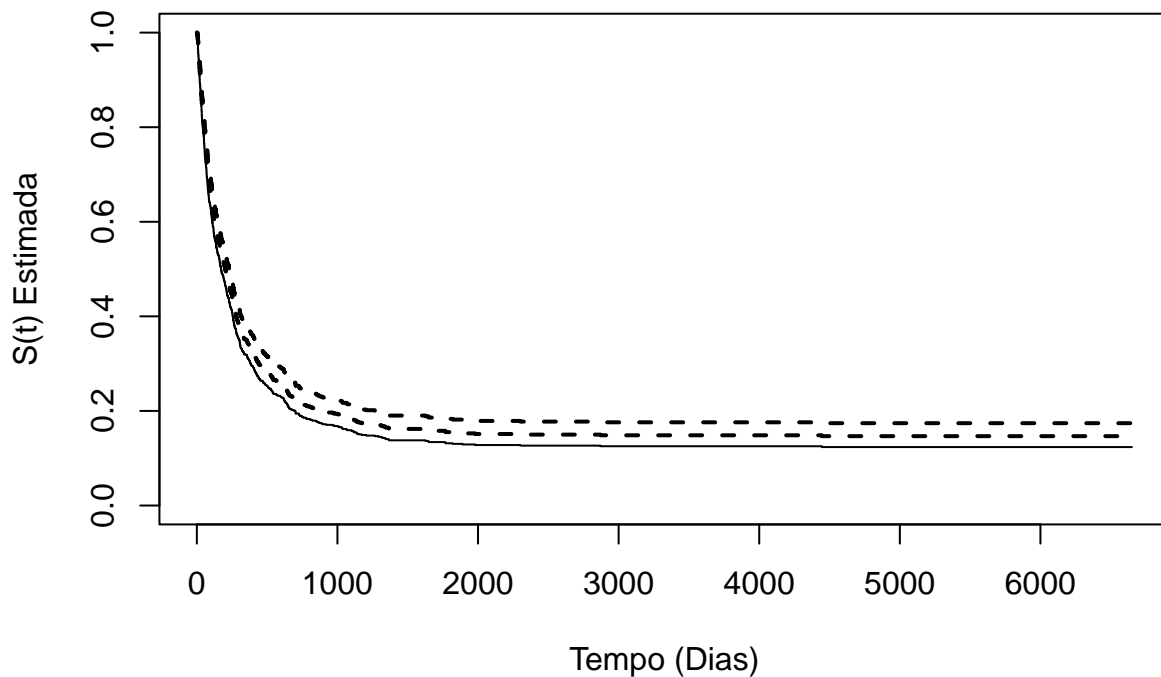
Selecionando as Colunas que Iremos Utilizar

```
library(lubridate)

pulmao<- pulmao %>% filter(Extensão != "NÃO SE APLICA")
pulmao$falhas<-as.numeric(pulmao$falhas)
library(survival)

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 Sobrevivência ")
```

Curva 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

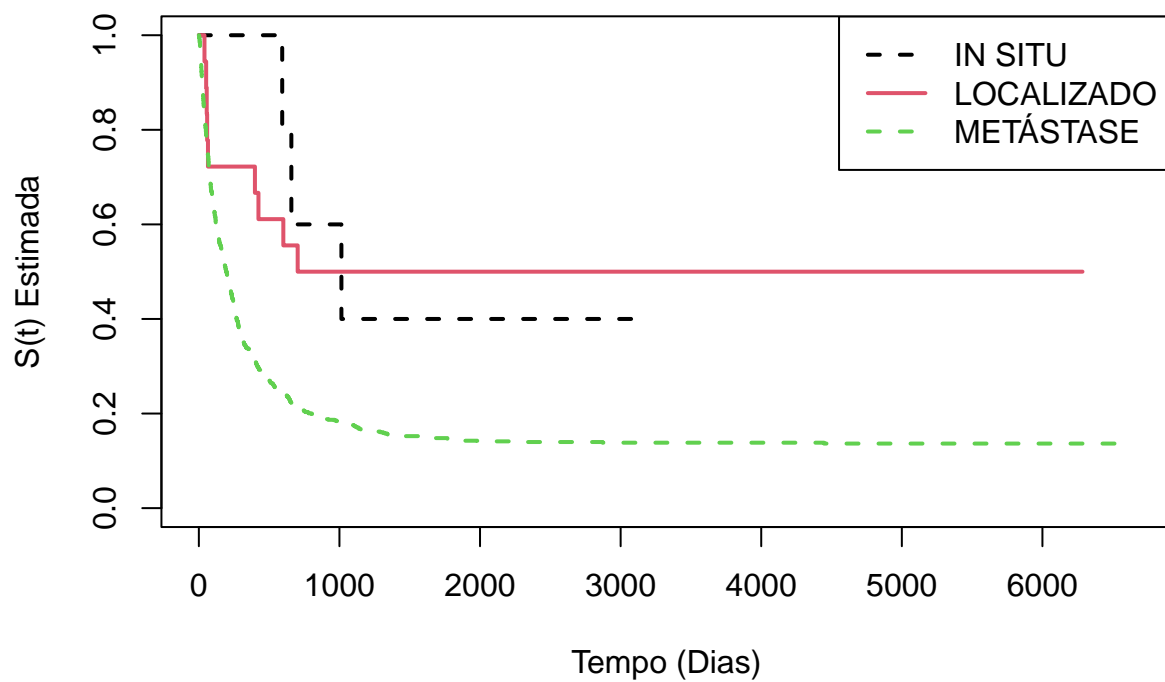
# pulmao %>%
#   dplyr::select(where(is.factor)) %>%
#   glimpse()

# table(pulmao$Extensão)

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"), # Legend texts
       lty = c(2,1),           # Line types
       col = c(1:4),           # Line colors
       lwd = 2)

```

Curva de Sobrevivência com Base no Estágio da Doença



```
# d_km <- tidy(ekm_exp)
# p<-ggplot(d_km, aes(x = time, y = estimate, colour = strata)) +
#   geom_ribbon(aes(ymax = conf.low, ymin = conf.high, fill = strata),
#             stat = 'stepribbon', alpha = .1, size = .1) +
#   geom_step() +
#   theme_bw(16)
#
#
# library(plotly)
#
# ggplotly(p)
```

Neste Gráfico podemos perceber que após mais de 1000 dias após o diagnostico que o paciente estava com câncer é melhor pertencer ao grupo onde o câncer no pulmão está localizado, pois após os 1000 dias é o grupo que possui a maior probabilidade de sobrevivência, uma vez que quando o câncer é localizado isso significa que não existe sinal de disseminação da doença.

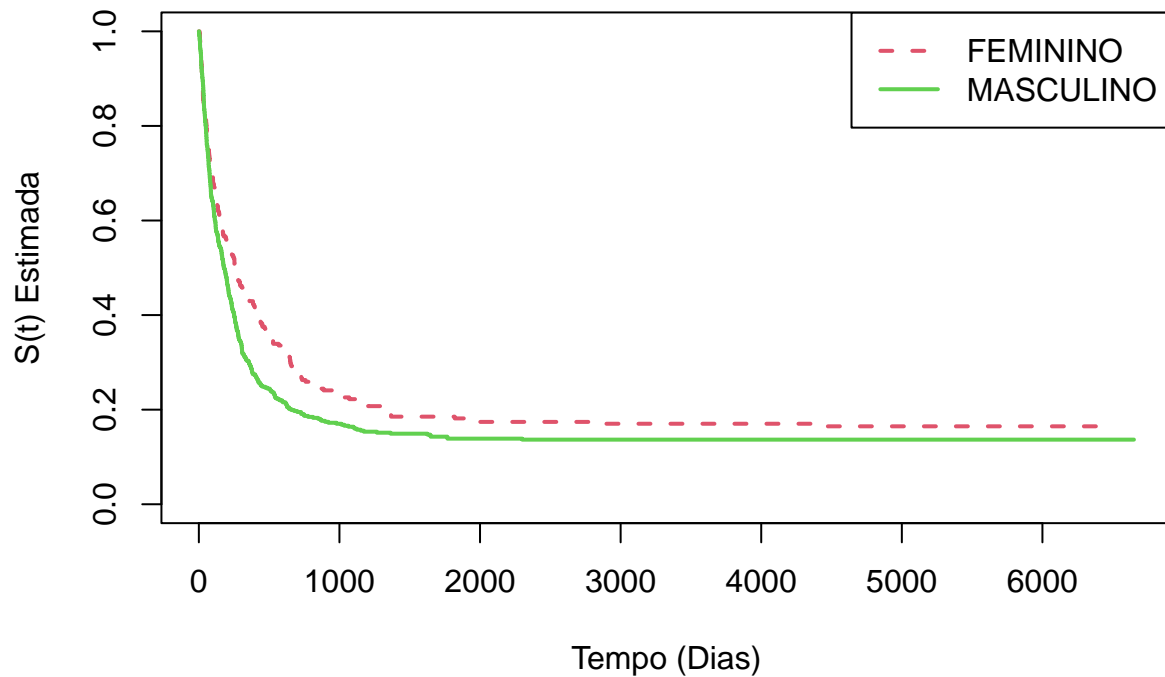
Além disso, os pacientes que possuem câncer em metáfase possuem a menor probabilidade de sobrevivência, tendo em vista que as células cancerosas se destacam do tumor original e viajam através da circulação sanguínea ou linfática, formando novos tumores. Esse processo é conhecido como metástase.

Fazendo o Kaplan-Meier para o variável Sexo

```
KPS <- survfit(Surv(pulmao$dias,pulmao$falhas)~pulmao$Sexo)

plot(KPS, lty=c(2,1), xlab="Tempo (Dias)",
      ylab="S(t) Estimada", col = 2:4,
      lwd = c(2,2), main = "Curva de Sobrevivência com Base no Estágio da Doença")
legend(x = "topright",           # Position
       c("FEMININO", "MASCULINO"), # Legend texts
       lty = c(2,1),             # Line types
       col = c(2:4),             # Line colors
       lwd = 2)
```


Curva de Sobrevivência com Base no Estágio da Doença



```
#install.packages("broom")  
library(ggplot2)  
library(ggalt)
```

```
## Registered S3 methods overwritten by 'ggalt':  
##   method                      from  
##   grid.draw.absoluteGrob      ggplot2  
##   grobHeight.absoluteGrob     ggplot2  
##   grobWidth.absoluteGrob      ggplot2  
##   grobX.absoluteGrob          ggplot2  
##   grobY.absoluteGrob          ggplot2
```

```
library(broom)
```

```
d_km <- tidy(KPS)  
# p<-ggplot(d_km, aes(x = time, y = estimate, colour = strata)) +  
#   geom_ribbon(aes(ymin = conf.low, ymax = conf.high, fill = strata),  
#               stat = 'stepribbon', alpha = .1, size = .1) +  
#   geom_step() +  
#   theme_bw(16)  
#  
#  
# library(plotly)
```

```
#
# ggplotly(p)
```

Pelos Gráficos acima podemos observar que após OS 150 dias após ter sido feito o diagnóstico o grupo feminino começa a apresentar uma maior probabilidade de sobrevivência se comparado ao grupo masculino.

vamos comparar as curvas de sobrevivencia dos grupos sexo

```
survdif(Surv(pulmao$dias,pulmao$falhas)~pulmao$Sexo,rho=0)
```

```
## Call:
## survdiff(formula = Surv(pulmao$dias, pulmao$falhas) ~ pulmao$Sexo,
##      rho = 0)
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## pulmao$Sexo=FEMININO 294      234      265      3.71      6.19
## pulmao$Sexo=MASCULINO 528      437      406      2.43      6.19
##
## Chisq= 6.2  on 1 degrees of freedom, p= 0.01
```

calculando o Teste Log Rank para os Grupos de Extensão

```
colnames(pulmao)[14]<-'Extensão'
survdif(Surv(pulmao$dias,pulmao$falhas)~pulmao$Extensão,rho=0)
```

```
## Call:
## survdiff(formula = Surv(pulmao$dias, pulmao$falhas) ~ pulmao$Extensão,
##      rho = 0)
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## pulmao$Extensão=IN SITU      5          3      8.25      3.341      3.40
## pulmao$Extensão=LOCALIZADO 18          9     23.58      9.014      9.42
## pulmao$Extensão=METÁSTASE 799        659    639.17      0.615     13.10
##
## Chisq= 13.1  on 2 degrees of freedom, p= 0.001
```

logo podemos ver que a 5% de significância as curvas de sobrevivencia apresentaram diferença significativa

Referências

Análise de Sobrevivência Aplicada . An extension of Shapiro and Wilks WW test for normality to large samples. Applied Statistics, 31, 115–124. doi:10.2307/2347973