ANÁLISE DE SOBREVIDA 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 individuos por câncer no pulmão O estudo teve como objetivo estudar o tempo até a reincidência de morte por cancer no pulmão no estado de Minas Gerais, buscando a identificação de possíveis fatores que podem estar relacionados ao tempo até a reincidência desses indivíduos.

## 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. Para essa análise foi utilizado um banco de dados contendo informações sobre pacientes com câncer no estado de Minas Gerais , após o tratamento dos dados foram adicionados duas novas variáveis **dias** e **falhas** , onde a variável **dias** é referente ao número de dias até a primeira falha acontecer ou seja , será nosso tempo observado, já a variável **falhas** contém a informação se nosso evento falhou ou foi censurado(variável binária, 0 ou 1 ) onde 0 é caso tenha ocorrido um censura e 1 caso tenha ocorrido uma falha. Foi utilizado o estimador de kaplan-meier , para fazer a comparação das curvas entre os grupos usamos neste projeto o teste log-rank .

# Introdução

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.

A análise de sobrevivência é uma das áreas da estatística que mais cresceu nas últimas duas décadas do século passado. A razão deste crescimento ´e 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.

## Representação dos Dados de Sobrevivência

Os dados de sobrevivência para o indivÍduo sob estudo, são representados, em geral, pelo par ( , ) sendo o tempo de falha ou de censura e a variável indicadora de falha ou censura, isto é

## Carregando os pacotes

Vamos iniciar carregando os pacotes que iremos utilizar

library(tidyverse)  
  
library(tidyverse)  
  
  
library(plotly)  
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"

#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

# Calculando a taxa de censura

Antes de mais nada ,iremos calcular a taxa de censura presente em nossos dados

round(1 - mean(pulmao$falhas),3)

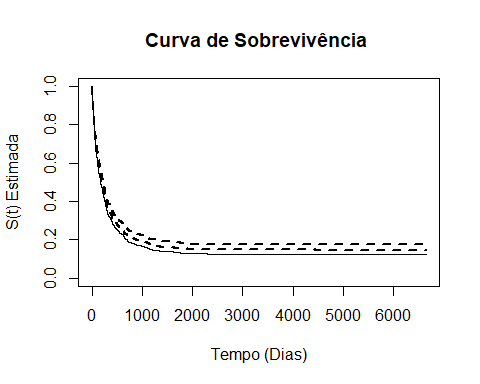
## [1] 0.186

nossa taxa de censura é de aproximadamente 18,4% .

# Estimador de Kaplan-Meier

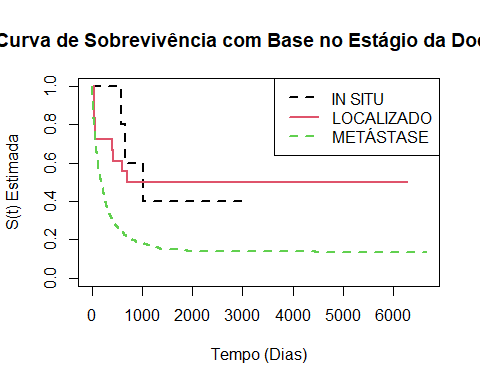
O Esttimador de Kaple-Meier é um estimador não-paramétrico onde a função de sobrevivência será dada da forma

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 ")



# 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)



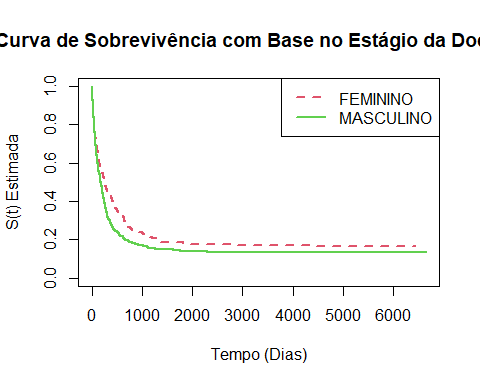
# d\_km <- tidy(ekm\_exp)  
# p<-ggplot(d\_km, aes(x = time, y = estimate, colour = strata)) +  
# geom\_ribbon(aes(ymin = conf.low, ymax = conf.high, fill = strata),  
# stat = 'stepribbon', alpha = .1, size = .1) +  
# geom\_step() +  
# theme\_bw(16)  
#   
#   
# library(plotly)  
#   
# ggplotly(p)

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 possuiem câncer em metafase pussui 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)



#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 cido feito o diagnóstico o grupo femino começa apresentar uma maior probabilidade de sobrevivência se comparado ao grupo masculino.

# vamos comparar as curvas de sobrevivencia dos grupos sexo

survdiff(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'  
survdiff(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

# Ajustes de modelos

# 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>