

Trabalho final: Função SOBREVIVE

Arquivos

Código da função: [funcao_sobrevive](#)

Página de ajuda: [help_sobrevive](#)

Arquivos para os exemplos:

- Exemplo 1: [teste](#)
- Exemplo 2: [dados_reais_controle](#)
- Exemplo 3: [dados_reais_longdon_2015](#)

Código da função

```
##### FUNCAO SOBREVIVE #####

sobrevive<-function(dataframe, na.rm=TRUE, grupos="x", dias="y", z=TRUE)
#Dataframe: o objeto de input eh um data.frame no qual conste a
identificacao dos grupos
#analizados, as replicas, os dias de observacao e os numeros de individuos
perdidos e/ou
#mortos em cada dia.
#Argumento 1: omissao de dados faltantes (TRUE ou FALSE).
{
  if(na.rm==TRUE)
  {
    dados=read.table(dataframe,header=T) #carregando o arquivo de input.
    old=length(dados$Grupo)
    dados=na.omit(dados) #omitindo os dados NA do dataframe.
    n.na=old - length(dados$Grupo) #calcula a quantidade de na que foram
excluidos.
    cat("\t", n.na, "valores NA excluidos\n") #avisando o usuario quantos NA
foram excluidos.
  }
  else
  {
    dados=read.table(dataframe,header=T)
  }
#Argumento 2: escolha do numero de grupos de estudo (minimo de 1).
#0 usuario pode escolher quantos grupos da sua planilha de dados ele quer
analizar.
#Atencao: aqui serao considerados apenas os "x" primeiros grupos.
if(grupos < 1)
{
  stop("Numero de grupos deve ser maior ou igual a um")
}
```

```
#Argumento 3: numero de dias de observacao (minimo de 2).
#0 usuario pode escolher quantos dias da sua planilha de dados ele quer
analizar.
#Atencao: aqui serao considerados apenas os "y" primeiros dias.
if(dias < 2)
{
  stop("Numero de períodos de observacao deve ser maior ou igual a dois")
}

#Criando o objeto para guardar as probabilidades de sobrevivencia (S(t))
de cada replica.
St<-matrix(NA, ncol=4, nrow = length(dados$Grupo))
#Calculando a estimativa de Kaplan-Meier (probabilidades de sobrevivencia
(S(t))) e
#guardando os resultados na matriz St.
for (i in 1:length(dados$Grupo))
{
  n<-(dados$Total[i]+dados$Mortes[i]) #n eh o total de individuos em
risco.
  d<-dados$Mortes[i] #d eh o numero de individuos mortos.
  if (dados$Dia[i]>1) #calculo da produtividade.
  {
    if (n==0) #corrigindo a divisao por zero presente na estimativa.
    {
      prob=0 #calculo do estimador Kaplan-Meier.
    }
    else
    {
      prob=prob*((n-d)/n) #calculo do estimador Kaplan-Meier.
    }
  }
  if (dados$Dia[i]==1)
  {
    prob=(n-d)/n #calculo do estimador Kaplan-Meier.
  }
  #salvando os resultados na matriz St.
  St[i,4]<-prob
  St[i,3]<-dados$Dia[i]
  St[i,2]<-dados$Replica[i]
  St[i,1]<-dados$Grupo[i]
}
#Criando o objeto para guardar as probabilidades médias de sobrevivencia
(S(t)medio) entre
#as replicas para um mesmo dia.
Stmedio<-matrix(NA, ncol = dias, nrow = grupos)
nomes_grupos<-levels(as.factor(dados$Grupo)) #guardando os nomes dos
grupos (para o grafico).
nomes_dias<-levels(as.factor(dados$Dia)) #guardando os dias das
observacoes (para o grafico).
```

```
#Calculando o Stmedio entre as réplicas para um mesmo dia e guardando os
resultados na matriz
#Stmedio.
for (i in 1:grupos)
  {
  for (j in 1:dias)
    {
    Stmedio[i,j]<-mean(St[St[,3]==j & St[,1]==i,4])
    }
  }

#Realizando os graficos.
x11()
if (grupos%%2==1) #checando se o numero de grupos analisados eh impar.
  {
  par(mfrow=c(grupos/2+0.5,2))
  }
else
  {
  par(mfrow = c(grupos/2,2))
  }
for (i in 1:grupos) #realizando os plots de todos os grupos.
  {
  plot(nomes_dias[1:dias], Stmedio[i,], type = "s", main=
paste("Estimativa de Kaplan-Meier do grupo", nomes_grupos[i], sep=" ",
collapse = NULL), xlab="Tempo", ylab="Probabilidade de sobrevivência")
  }
#Realizando o teste estatistico log-rank (Z).
#Atencao: o teste Z sera realizado entre todos os grupos disponiveis no
dataframe. O teste eh
#realizado de dois em dois grupos. Em caso de numero impar de grupos, o
ultimo grupo nao sera
#comparado.
if (z==TRUE)
  {
  Z<-matrix(NA, nrow=1, ncol=trunc(grupos/2))
  for(i in 1:trunc(grupos/2))
    {
    controle = dados[dados$Grupo == nomes_grupos[2*i-1],]
    experimental = dados[dados$Grupo == nomes_grupos[2*i],]
    O = controle$Mortes + experimental$Mortes
    N = controle$Total + experimental$Total
    E1 = O/N*controle$Total
    V = (O*(controle$Total/N)*(1-controle$Total/N)*(N-O))/(N-1)
    Z[i] <- sum(controle$Mortes - E1)/(sqrt(sum(V)))
    }
  }
#Extraindo o parametro meia vida dos dados.
meiavida<-matrix(NA, nrow = 1, ncol=grupos)
for (i in 1:grupos)
  {
```

```
#Como a meia vida eh um parametro extraido dos dados, a funcao emitira
uma mensagem caso
#o grupo analisado nao tenha Stmedio minimo de 0.5.
if (min(Stmedio[i,])>0.5)
{
  cat("\tParametro meia vida nao pode ser extraido dos dados\n")
}
#Caso o grupo tenha Stmedio minimo de 0.5, a funcao fara uma media entre
os extremos do
#intervalo no qual 0.5 esteja contido.
else
{
  temp<-Stmedio[i,]
  dia = St[,3]
  meiavida[i]<-
(dia[length(temp[temp>0.5])+dia[length(temp[temp>0.5])+1])/2
}
}
return(c(Z, meiavida)) #retornando a estatistica Z e o parametro meia vida
para o usuario.
}
```

Página de ajuda

sobrevive package:nenhum R Documentation

Análise de sobrevivência utilizando a estimativa de Kaplan-Meier

Description:

A função calcula, por meio da estimativa de Kaplan-Meier, a probabilidade de sobrevivência em um dado tempo para cada um dos grupos considerados em um conjunto de dados e realiza o teste estatístico log-rank (Z) para comparar os grupos dois a dois. A função retorna a estatística Z, um gráfico de cada grupo com sua respectiva curva de sobrevivência e, quando possível, o parâmetro meia vida de cada grupo.

Usage:

```
sobrevive<-function(dataframe, na.rm=TRUE, grupos="x", dias="y", z=TRUE)
```

Arguments:

dataframe nome do arquivo que contém os dados. O arquivo é uma tabela que precisa conter as colunas "Grupo", "Replica",

"Dia", "Mortes" e "Total".

na.rm omissão de dados ausentes na tabela de entrada. O default é na.rm=TRUE, mas deve se ter cuidado para que a omissão de na's não deixe as réplicas com quantidades diferentes de intervalos de observação.

grupos quantidade de grupos, dentre os disponíveis na tabela de entrada, que o usuário deseja analisar.

dias
disponíveis quantidade de intervalos de observação, dentre os disponíveis na tabela de entrada, que o usuário deseja analisar.

z cálculo do teste de hipótese de log-rank. O default é z=TRUE.

Details:

Análise de sobrevivência é um ramo da estatística que analisa o tempo até a ocorrência de determinado fenômeno, como morte ou diagnóstico de doença.

Análises de sobrevivência obedecem modelos de regressão logística e têm um ponto de partida - geralmente definido quando as observações se iniciam - e um término

- quando o evento de interesse acontece. O período entre o início das observações

e a ocorrência do evento de interesse compreende no tempo de observação, que é

a variável independente (preditora) do modelo. Contudo, o tempo para a ocorrência

de um dado evento pode ser maior que o tempo disponível para observação e coleta

de dados, gerando observações incompletas. Processos produzindo esse tipo de observação incompleta é chamado de "censoring", ao passo que a observação incompleta em si é chamada "censored data".

"Censored data" são comuns de acontecer em diversos tipos de trabalho devido à limitação de recursos para a realização do estudo, como tempo, equipe

e financiamento. Desse modo, análises de sobrevivência devem utilizar modelos

que estimam a probabilidade de sobrevivência de um determinado grupo considerando

"censored data". A estimativa padrão para funções de sobrevivência é a de Kaplan

-Meier (Kaplan & Meier, 1958), a qual utiliza os dados observados para estimar

a probabilidade de sobrevivência em cada grupo analisado e para extrair os parâmetros de interesse do pesquisador.

A função implementa a estimativa de Kaplan-Meier para curvas de sobrevivência

provenientes de diferentes conjuntos de dados, considerando "censored data". De modo geral, a função (a) calcula a estimativa Kaplan-Meier para cada grupo de estudo do conjunto de dados disponível, (b) constrói um gráfico com a curva de sobrevivência para cada grupo de estudo, (c) extração dos dados o parâmetro meia-vida da curva de sobrevivência (tempo t em que $S(t) = 50$) e (d) calcula Z do teste de log-rank para a comparação estatística entre os grupos.

Value:

A função gera um gráfico para cada grupo analisado, com sua respectiva curva de interesse, e retorna uma lista com as seguintes posições:

Z resultado do teste log-rank. A estatística Z será calculada pareando os grupos dois a dois. Caso seja escolhida a opção $z=FALSE$, serão retornados NA's.

meiavida parâmetro meia-vida. O parâmetro meia vida só poderá ser extraído dos dados de um grupo que tenha apresentado probabilidade de sobrevivência média mínima de 0.5. Caso contrário, serão retornados NA's.

Warning:

A função é interrompida quando, no argumento "grupos", não houver no mínimo um grupo a ser analisado; quando, no argumento "dias", não houver no mínimo dois períodos de observação; e quando a quantidade de períodos de observação entre as réplicas for diferente.

Note:

Detalhes importantes para a formatação correta do arquivo de entrada:

A tabela precisa conter as colunas "Grupo", "Replica", "Dia", "Mortes" e "Total".

(a) A coluna "Grupo" contém a descrição dos grupos a serem analisados e pode conter números ("1", "2") ou palavras ("Especie.A", "Especie.B").

(b) A coluna "Replica" contém a identificação das replicagens feitas em cada um dos grupos e também pode conter números ("1", "2") ou palavras ("rep.A",

"rep.B"). A função realizará uma média entre as probabilidades de sobrevivência calculadas para cada uma das réplicas. Assim, os grupos podem ter quantidades diferentes de réplicas.

(c) A coluna "Dia" contém a identificação dos intervalos de observação. Os intervalos entre as observações podem ser irregulares (ex. dias 1, 3, 4, 10,

23), mas precisam ser os mesmos entre todas as réplicas.

(d) A coluna "Mortes" deve conter a quantidade de indivíduos mortos em

em cada um dos intervalos de observação.

(e) A coluna "Total" contém a quantidade de indivíduos vivos em cada um dos dias.

(f) A ordem das colunas não importa.

(g) A tabela pode conter outras colunas de interesse para o usuário.

Limitações da função:

(a) A função permite escolher quantos grupos, do total disponível na tabela, o usuário quer analisar. Contudo, a função não permite escolher quais

grupos analisar. Assim, a função analisará os "x" primeiros grupos disponíveis

na tabela, sendo "x" o valor atribuído ao argumento "grupos".

(b) A tabela precisa ter os mesmos intervalos de observação entre as rélicas.

(c) Apesar de o usuário poder escolher os grupos e os dias a serem analisados, a estatística Z será calculada entre todos os grupos disponíveis na tabela.

(d) A estatística Z será calculada pareando os grupos dois a dois. Por exemplo, o grupo na primeira posição será comparado com o da segunda posição, e

o grupo na terceira posição será comparado com o grupo na quarta posição. Desse

modo, o usuário deverá organizar a tabela adequadamente de acordo com as comparações que deseja realizar.

(e) Para a correta interpretação da estatística Z, o usuário deverá consultar uma tabela de qui-quadrado.

(f) O parâmetro meia vida só poderá ser extraído dos dados de um grupo caso o grupo tenha apresentado probabilidade de sobrevivência média (Stmedio)

mínima de 0.5. Caso contrário, aparecerá um aviso para o usuário dizendo que o

parâmetro não pode ser extraído dos dados e o valor da meia vida será retornado

como NA.

Author(s):

Camila Souza Beraldo
ca.berald@gmail.com

References:

Para análise de sobrevivência, ver:

Hosmer, DW; Lemeshow, S; May, S. Applied survival analysis : regression modeling of time-to-event data - 2nd ed. 411p.

Examples:

```
###Exemplo "teste.txt"
```

```
sobrevive("teste.txt", na.rm=TRUE, grupos=2, dias=5)
```

```
###Exemplo "dados_reais_controle.txt"
```

```
sobrevive("dados_reais_controle.txt", na.rm=TRUE, grupos=6, dias=20)
```

```
###Exemplo "dados_reais_longdon_2015.txt"
```

```
sobrevive("dados_reais_longdon_2015.txt", na.rm=TRUE, grupos=6, dias=20)
```

From:

<http://ecor.ib.usp.br/> - **ecoR**

Permanent link:

http://ecor.ib.usp.br/doku.php?id=05_curso_antigo:r2016:alunos:trabalho_final:camila.beraldo:trabalho_final:funcao_sobrevive 

Last update: **2020/08/12 06:04**