

## Lista 6 - MAE0217

23/06/2020

### Exercício 3 - Capítulo 8

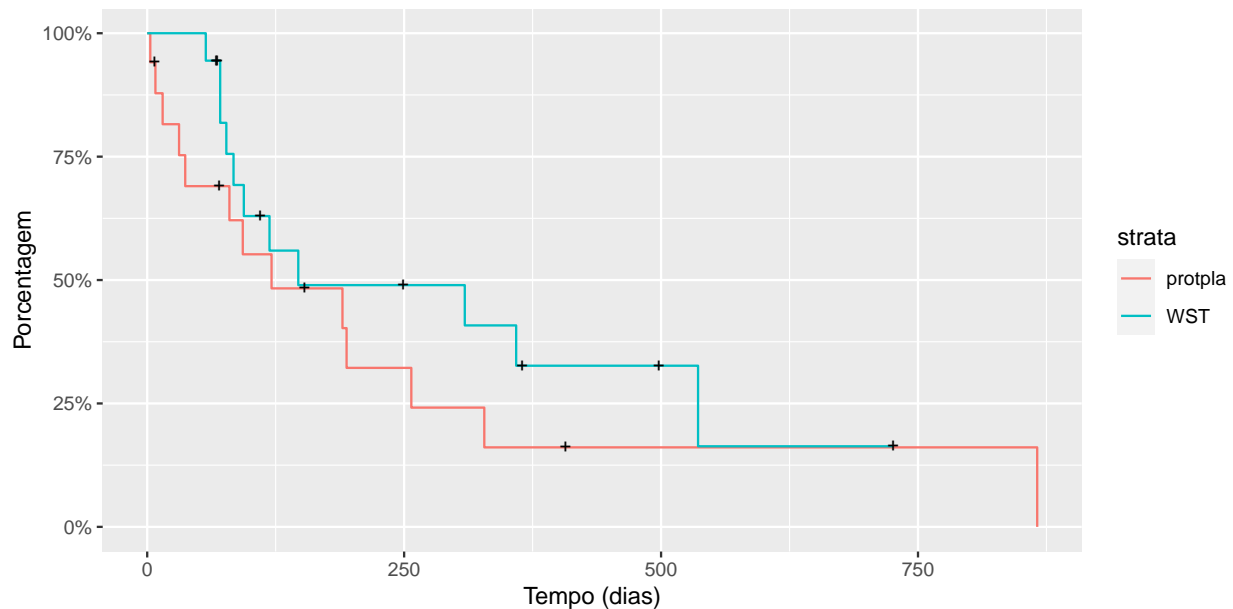
a)

A partir dos dados disponíveis no arquivo **sondas**, podemos construir curvas de sobrevivência para pacientes submetidos aos tipos de sondas protpla e WST com o auxílio do seguinte código do R:

```
sondas <- readxl::read_excel(path="/Users/kevin/Downloads/sondas.xls")
sondas <- subset(sondas,select = -paciente)
#foi removida a coluna que contem os pacientes pois a mesma é equivalente a própria indexação do R
tipo <- survfit(Surv(dias,evento)~sonda,data = sondas)
```

Resultando no seguinte gráfico da função de sobrevivência:

```
autoplot(tipo,xlab = "Tempo (dias)",ylab = "Porcentagem",conf.int = F)
```



Além disso, o tempo médio e o tempo mediano para obstrução em cada tipo de sonda estão indicados abaixo e podem ser calculados a partir do código do R a seguir.

```
print(tipo,print.rmean = T)
```

```
## Call: survfit(formula = Surv(dias, evento) ~ sonda, data = sondas)
##
##              n events *rmean *se(rmean) median 0.95LCL 0.95UCL
## sonda=protpla 17      13   232      71.4   121      37      NA
## sonda=WST     18      11   319      73.5   147      94      NA
##      * restricted mean with upper limit = 796
```

Pelo resultado do código “print(tipo, print.rmean = T)”, podemos afirmar que:

- Para a sonda *protpla*
  1. *média* = 232(71.4)
  2. *mediana* = 121
- Para a sonda *WST*
  1. *média* = 319(73.5)
  2. *mediana* = 147

As médias foram apresentadas com os erros padrões respectivos entre parênteses.

Como pode ser notada no gráfico da função de sobrevivência, a curva de sobrevivência estimada referente à sonda *protpla* aparenta um decaimento mais acentuado em relação àquela referente à sonda *WST*. Desse modo, temos a sugestão de que os pacientes que receberam sondas do tipo *protpla* apresentam obstrução em um tempo inferior aos que receberam sondas do tipo *WST*. Numericamente, esse resultado é reforçado pelos valores de tempo médio e de tempo mediano para obstrução de cada tipo de sonda, em que obtivemos tempos médios de 232 e 319 dias e tempos medianos de 121 e 147 dias para as sondas *protpla* e *WST*, respectivamente. Portanto, a sonda *WST* aparenta ser mais eficiente no sentido de demorar mais para que ocorra sua obstrução, no entanto, as diferenças obtidas podem ter caráter amostral apenas, não podendo ser estendidas para a população, o que será analisado em sequência.

b)

Queremos saber se, com base nas curvas de Kaplan-Meier obtidas no item (a), podemos concluir que as curvas de sobrevivência associadas às populações de onde as amostras foram selecionadas são iguais ou não. Ou seja, se podemos extrapolar os resultados discutidos em (a) para a população (nesse caso, as curvas de sobrevivência populacionais de pacientes submetidos a cada tipo de sonda seriam diferentes), ou se podemos assumir que as diferenças obtidas apresentam caráter apenas amostral (nesse caso, as curvas de sobrevivência populacionais de pacientes submetidos a cada tipo de sonda seriam iguais). Com tal propósito, usaremos o teste log-rank, o qual se baseia na comparação de valores esperados e observados para testar a hipótese nula de que não há diferença entre os tempos de obstrução de cada tipo de sonda. Para tanto, segue o código do R com sua referente saída.

```
survdifff(Surv(dias,evento)~sonda,data=sondas,rho=0)
```

```
## Call:
## survdifff(formula = Surv(dias, evento) ~ sonda, data = sondas,
##      rho = 0)
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## sonda=protpla 17      13    10.6     0.539     1.04
## sonda=WST     18      11    13.4     0.427     1.04
##
## Chisq= 1  on 1 degrees of freedom, p= 0.3
```

O valor-p obtido para o teste foi de 0.3, de modo que não rejeitamos a hipótese nula e assumimos que as diferenças observadas na comparação das curvas de sobrevivência estimadas têm caráter essencialmente amostral, isto é, aceitamos que as curvas de sobrevivência associadas a cada tipo de sonda são iguais. Portanto, em análise populacional, não há diferença entre o tempo de obstrução da sonda em relação aos tipos protpla e WST.

c)

Vamos agora construir 4 estratos com base nos fatores idade e tipo de sonda. Primeiramente, analisemos, com o auxílio do R, os dados referentes à idade e adotemos a mediana como o valor que determina a divisão dos pacientes em grupos de “jovens” e “idosos”.

```
summary(sondas)
```

```
##      dias      evento      sonda      sexo
## Min.   : 3.0   Min.   :0.0000   Length:35   Length:35
## 1st Qu.: 69.0   1st Qu.:0.0000   Class :character   Class :character
## Median :110.0   Median :1.0000   Mode  :character   Mode  :character
## Mean   :196.2   Mean   :0.6857
## 3rd Qu.:283.0   3rd Qu.:1.0000
## Max.   :866.0   Max.   :1.0000
##      idade
## Min.   :53.00
## 1st Qu.:69.00
## Median :76.00
## Mean   :74.57
## 3rd Qu.:81.50
## Max.   :91.00
```

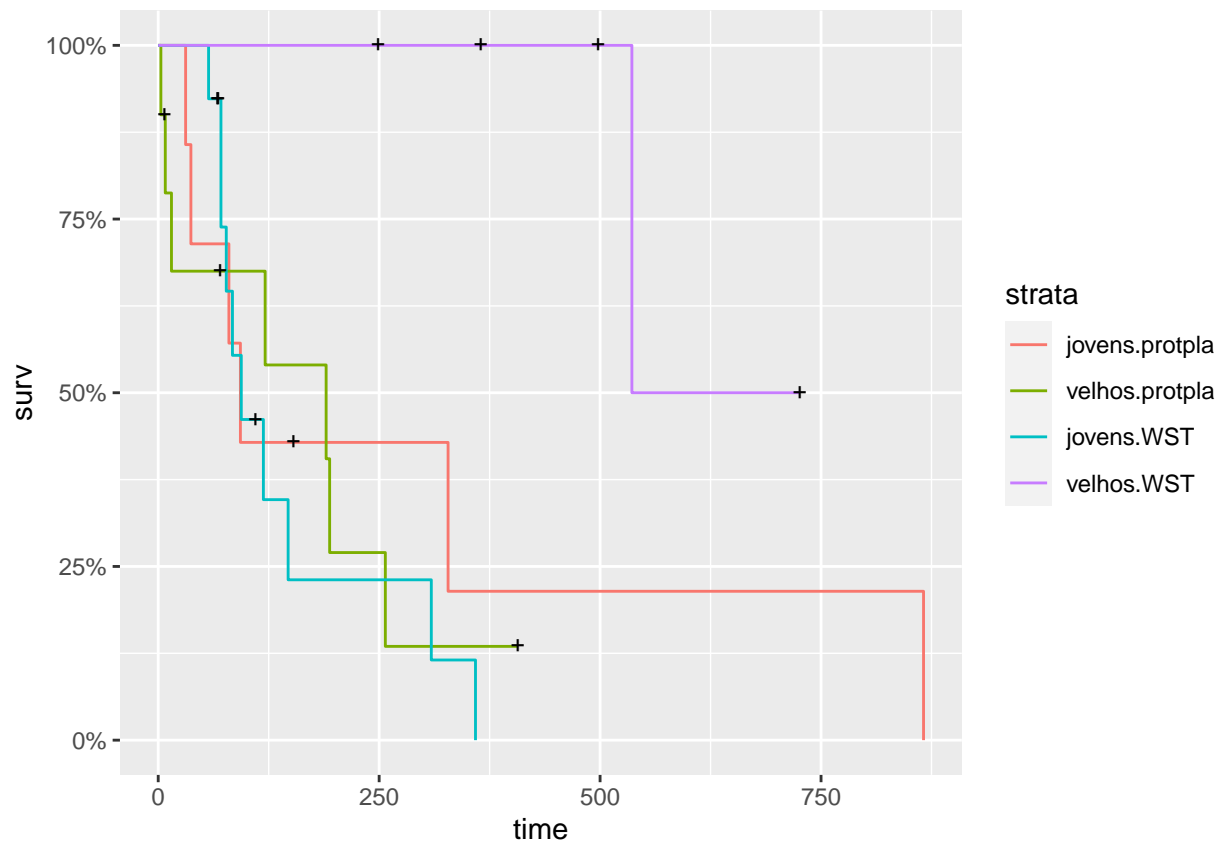
Como podemos observar, a idade mediana é 76 anos, de modo que aqueles pacientes cuja idade é inferior ou igual a este valor serão caracterizados como “jovens” e aqueles cuja idade é superior a 76 anos entrarão no grupo “idosos”. Assim, os pacientes foram divididos em 4 estratos, nomeadamente: “jovens com sonda tipo protpla”, “jovens com sonda tipo WST”, “idosos com sonda tipo protpla” e “idosos com sonda tipo WST”. Com base nessa estratificação, refaçamos as curvas de sobrevivência:

```
sondas$vivencia <- as.factor(ifelse(sondas$idade<=76,"jovens","velhos"))
#Com o comando acima eu crio mais uma coluna no data frame dos dados
#com a variável idade categorizada pela mediana como ponto de corte.
tipos <- survfit(Surv(dias,evento)~interaction(vivencia,sonda),data=sondas)
summary(tipos)
```

```
## Call: survfit(formula = Surv(dias, evento) ~ interaction(vivencia,
##      sonda), data = sondas)
##
##      interaction(vivencia, sonda)=jovens.protpla
## time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   31      7      1    0.857   0.132    0.633      1
##   37      6      1    0.714   0.171    0.447      1
##   80      5      1    0.571   0.187    0.301      1
##   93      4      1    0.429   0.187    0.182      1
##  328      2      1    0.214   0.178    0.042      1
##  866      1      1    0.000   NaN      NA      NA
```

```
##
##           interaction(vivencia, sonda)=velhos.protpla
## time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    3      10      1   0.900  0.0949   0.7320   1.000
##    8       8      1   0.787  0.1340   0.5641   1.000
##   15       7      1   0.675  0.1551   0.4303   1.000
##  121       5      1   0.540  0.1731   0.2881   1.000
##  190       4      1   0.405  0.1747   0.1739   0.943
##  194       3      1   0.270  0.1604   0.0843   0.865
##  257       2      1   0.135  0.1247   0.0221   0.825
##
##           interaction(vivencia, sonda)=jovens.WST
## time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   57      13      1   0.923  0.0739   0.7890   1.000
##   71      10      2   0.738  0.1309   0.5218   1.000
##   77       8      1   0.646  0.1434   0.4182   0.998
##   84       7      1   0.554  0.1497   0.3261   0.941
##   94       6      1   0.462  0.1506   0.2435   0.875
##  119       4      1   0.346  0.1508   0.1474   0.813
##  147       3      1   0.231  0.1378   0.0716   0.744
##  309       2      1   0.115  0.1068   0.0188   0.708
##  359       1      1   0.000   NaN      NA      NA
##
##           interaction(vivencia, sonda)=velhos.WST
##           time           n.risk           n.event           survival           std.err lower 95% CI
##           536.000           2.000           1.000           0.500           0.354           0.125
## upper 95% CI
##           1.000
```

```
autoplot(tipos, conf.int = F)
```



Com base no gráfico, notamos um comportamento diferenciado da curva de sobrevivência estimada associada ao estrato “idosos com sonda tipo WST” em relação às curvas dos demais estratos. Nela notamos um decaimento mais lento da probabilidade de sobrevivência, sugerindo que pacientes idosos (mais de 76 anos) submetidos à sonda tipo WST formam o grupo em que o tempo até a obstrução da sonda é maior. Já entre as demais curvas de sobrevivência aparentam ser mais similares, de modo que as diferenças poderiam ser apenas amostrais assim como àquelas notadas em (a).

A fim de verificar tais impressões, façamos também o teste log-rank, cuja hipótese nula é que as curvas de sobrevivência populacionais associadas aos quatro estratos são iguais, não havendo diferença em relação aos tempos até a ocorrência de obstrução da sonda. Segue, portanto, o código do R relacionado ao teste e sua saída:

```
survdif(Surv(dias,evento)~vivencia+sonda,data=sondas)
```

```
## Call:
## survdiff(formula = Surv(dias, evento) ~ vivencia + sonda, data = sondas)
##
##
##           N Observed Expected (O-E)^2/E (O-E)^2/V
## vivencia=jovens, sonda=protpla  7         6      5.66    0.021    0.0325
## vivencia=jovens, sonda=WST     13        10      6.52    1.856    2.7965
## vivencia=velhos, sonda=protpla 10         7      4.95    0.846    1.1133
## vivencia=velhos, sonda=WST      5         1      6.87    5.016    7.9350
##
## Chisq= 8.8  on 3 degrees of freedom, p= 0.03
```

Diferentemente do resultado obtido em (b), aqui temos um p-valor pequeno (0.03), de modo que rejeitamos a

hipótese nula, isto é, há evidências amostrais de que as curvas de sobrevivência populacionais são diferentes entre si ou, em outras palavras, temos diferenças quanto ao tempo até a obstrução da sonda.

## Exercício 4 - Capítulo 8

Baseando-se nos dados do arquivo **rehabcardio**, queremos analisar a sobrevivência de pacientes infartados submetidos a tratamento clínico ou a um programa de exercícios. Temos como evento de interesse a revascularização miocárdica, a ocorrência de um segundo infarto ou o óbito pós admissão no estudo, sendo que os pacientes foram acompanhados até 31/12/2000. A fim de adequação dos dados, foram adotadas tais modificações:

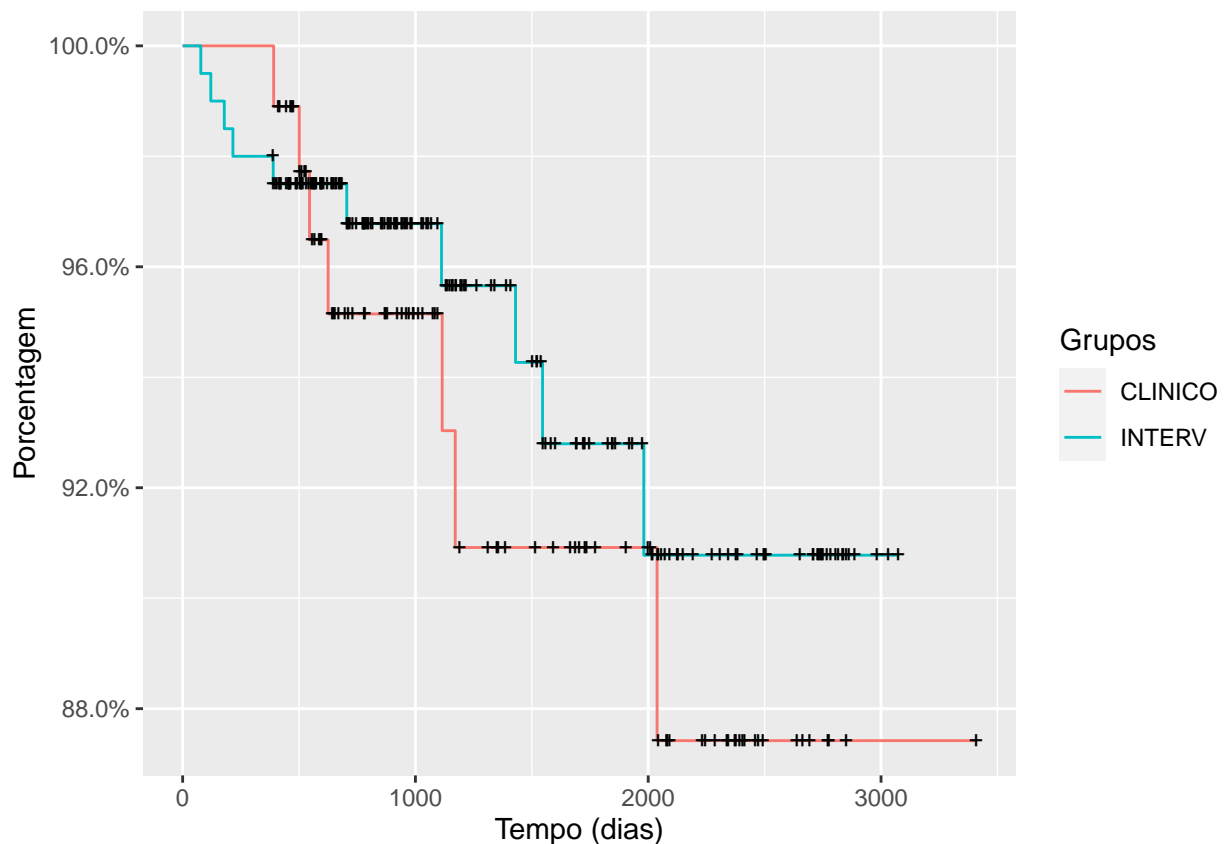
1. Todas as linhas com dados faltantes nas variáveis *Programa* e *Nascimento*, foram descartadas.
2. As linhas que continham dados de datas referentes as variáveis *Revascularização miocárdica*(**DataRM2**), *Ocorrência de um segundo infarto*(**DataIAM2**) ou *Óbito*, incompatíveis com a data de nascimento foram excluídas.
3. Os dados que apresentados, revelaram que existem casos onde o paciente pode ter passado por uma revascularização miocárdica ou pela ocorrência de um segundo infarto e em seguida vir a óbito(não necessariamente no mesmo dia), e portanto, dada a gravidade dos eventos de interesse, nesses casos, a data considerada é a data de óbito.
4. Para todos os pacientes não excuídos da tabela foi calculado o tempo de permanência no estudo, de acordo com os respectivos tempos de entrada no programa.
5. Para todos os pacientes nos quais não ocorrem nenhum dos eventos de interesse foram considerados como censuras.

```
card <- readxl::read_excel(path="/Users/kevin/Downloads/rehabcardio.xls")
lim <- as.Date(dmy("31/12/2000"))
card1 <- card %>%
  select(DataRM2,DataIAM2,Obito,Nascimento,Programa,Grupo) %>%
  #seleciona apenas as variáveis de interesse
  filter(!is.na(Nascimento)|is.na(Programa)) %>%
  filter(Nascimento%--DataRM2/ddays(1)>=0|is.na(Nascimento%--DataRM2/ddays(1)),
         Nascimento%--DataIAM2/ddays(1)>=0|is.na(Nascimento%--DataIAM2/ddays(1)),
         Nascimento%--Obito/ddays(1)>=0|is.na(Nascimento%--DataIAM2/ddays(1))) %>%
  #remove as incoerências com a data de nascimento
  mutate(DataRM2=(Programa%--DataRM2)/ddays(1)) %>% #
  mutate(DataIAM2=(Programa%--DataIAM2)/ddays(1)) %>% # Converte as datas em dias
  mutate(Obito=(Programa%--Obito)/ddays(1)) %>% #
  mutate(duracao=(Programa%--lim)/ddays(1)) %>%
  mutate(vida=(Nascimento%--lim)/ddays(1)) %>%
  filter(vida-duracao>0) %>%
  #remove os paciente que entraram no programa antes mesmo de nascer
  replace(is.na(.),0) %>%
  # Interpreta os dados faltantes como 0, para poder juntar as variaveis em uma coluna
  filter(Obito>=0,DataRM2>=0,DataIAM2>=0) %>%
  mutate(DataRM2=ifelse((DataRM2>0&Obito>0),0,DataRM2)) %>%
  # Considera o Óbito como prioridade nos eventos
  mutate(DataIAM2=ifelse(DataIAM2>0&Obito>0,0,DataIAM2)) %>% #
  mutate(datas=DataIAM2+DataRM2+Obito) %>%
  #junta as variaveis em uma unica variavel resposta
```

```
mutate(evento=ifelse(datas>0,1,0)) %>%
#cria uma variavel de evento (sim ou não)
mutate(datas=ifelse(datas==0,duracao,datas))
#coloca na variavel resposta os tempos de censura
card1 <- na.omit(card1) #remove possíveis dados faltantes
```

Com os dados ajustados, façamos um gráfico das curvas de sobrevivência, obtido a partir do código disponível a seguir.

```
grupos <- survfit(Surv(datas,evento)~Grupo,data=card1)
#ajusta as curvas de sobrevida de acordo com o grupo de cada paciente
autoplot(grupos,conf.int = F) + labs(x = "Tempo (dias)",y = "Porcentagem",colour="Grupos")
```



Além disso, com o auxílio do R, estimemos também o primeiro quartil e o tempo médio de sobrevivência.

```
print(grupos,print.rmean = T) # Mostra a média
```

```
## Call: survfit(formula = Surv(datas, evento) ~ Grupo, data = card1)
##
##              n events *rmean *se(rmean) median 0.95LCL 0.95UCL
## Grupo=CLINICO   91      7  2979      95.1    NA      NA      NA
## Grupo=INTERV   200     10  3048      60.7    NA      NA      NA
##      * restricted mean with upper limit = 3242
```

```
quantile(grupos)
```

```
## $quantile
##           25 50 75
## Grupo=CLINICO NA NA NA
## Grupo=INTERV  NA NA NA
##
## $lower
##           25 50 75
## Grupo=CLINICO NA NA NA
## Grupo=INTERV  NA NA NA
##
## $upper
##           25 50 75
## Grupo=CLINICO NA NA NA
## Grupo=INTERV  NA NA NA
```

Analisando apenas o gráfico, não temos indicativos de diferença entre as curvas de sobrevivência de cada grupo, o que é reforçado pela pequena diferença entre os tempos médios de sobrevivência que foram de 2979 e 3048 dias para pacientes submetidos a tratamento clínico e a um programa de exercícios, respectivamente. Notamos que não foi possível estimar o primeiro quartil, uma vez que seu estimador pode ser dado por  $T_{0.25} = \{\inf t : S(t) \leq 0.75\}$ , porém não temos valor de  $t$  que satisfaça tal condição.

Desejamos saber se podemos concluir se as curvas de sobrevivência associadas às populações de onde as amostras foram selecionadas são iguais ou não, tendo como base as curvas de Kaplan-Meier. Ou seja, queremos saber se podemos extrapolar os resultados já discutidos para a população (neste caso, as curvas de sobrevivência populacionais de pacientes submetidos a cada tipo de tratamento seriam diferentes), ou se podemos assumir que as diferenças encontradas apresentam-se apenas na amostra (nesse caso, as curvas de sobrevivência populacionais de pacientes submetidos a cada tipo de tratamento seriam iguais). De tal forma, usa-se o teste *log – rank*, que é baseado na comparação de valores esperados e observados, para testar a hipótese nula de que não há diferença entre os tempos de sobrevivência de cada paciente, com relação aos tipos de tratamento. Para tanto, segue o código do R, e a saída obtida.

```
survdif(Surv(datas,evento)~Grupo,data=card1)
```

```
## Call:
## survdiff(formula = Surv(datas, evento) ~ Grupo, data = card1)
##
##           N Observed Expected (O-E)^2/E (O-E)^2/V
## Grupo=CLINICO  91         7      5.7      0.296      0.446
## Grupo=INTERV  200        10     11.3      0.149      0.446
##
## Chisq= 0.4  on 1 degrees of freedom, p= 0.5
```

O valor-p obtido para o teste foi de 0,5. Assim, concluímos que não rejeita-se a hipótese nula e assumimos que as diferenças observadas na comparação das curvas de sobrevivência estimadas têm caráter essencialmente amostral, isto é, aceitamos que as curvas de sobrevivência associadas a cada tipo de tratamento são iguais. Portanto, em análise populacional, não teríamos diferenças entre o tempo de sobrevivência dos pacientes com relação ao tipo de tratamento a eles deferidos.