

Lista de Exercícios 2

Análise de Sobrevida

Márcio Roger Piaggio
RA: 67384

Universidade Estadual de Maringá
Prof. Dr. Brian Melo

Maringá-Pr
25/01/23

1º Questão

Mostre que vale a seguinte relação entre a Função de Risco Acumulado e a Função de Sobrevida.

$$H(t) = -\log(S(t))$$

$$\begin{aligned} H(t) &= \int_0^t h(u) du = \int_0^t \frac{f(u)}{S(u)} du \\ &= \frac{F(t)}{S(t)} - \int_0^t -\frac{F(u)}{S(u)^2} du \\ &= \frac{1-S(t)}{S(t)} + \int_0^t \frac{1-S(u)}{S(u)^2} du \\ &= \frac{1}{S(t)} - 1 + \int_0^t \left(\frac{1}{S(u)^2} - \frac{1}{S(u)} \right) du \\ &= \frac{1}{S(t)} - 1 - \frac{1}{S(t)} - \log(S(t)) + c \\ &= -\log(S(t)) + (c-1) \end{aligned}$$

■

2º Questão

(Seção 1.6 do Livro Análise de Sobrevida Aplicada) Suponha que a vida média residual de T seja dada por $vmr(t) = t + 10$. Obtenha $E(T)$, $h(t)$ e $S(t)$.

$$E(T) = vmr(0) = 0 + 10 = 10$$

$$h(t) = \frac{\left(\frac{d}{dt}(vmr(t)) + 1 \right)}{vmr(t)} = \frac{1+1}{t+10} = \frac{2}{t+10}$$

$$S(t) = \frac{vmr(0)}{vmr(t)} \exp \left\{ - \int_0^t \frac{du}{vmr(u)} \right\} = \frac{10}{t+10} \exp \left\{ - \log(u+10) \Big|_0^t \right\} = \frac{10}{t+10} \frac{10}{t+10} = \frac{100}{(t+10)^2}$$

3º Questão

Os dados da Tabela 2.9 referem-se aos tempos de sobrevida (em dias) de pacientes com câncer submetidos à radioterapia (o símbolo + indica censura).

Para estes dados, obtenha estimativas para:

```
tempos <- c(7,34,42,63,64,74,83,84,91,108,112,129,133,133,139,140,140,146,149,
            154,157,160,160,165,173,176,185,218,225,241,248,273,277,279,297,319,
            405,417,420,440,523,523,583,594,1101,1116,1146,1226,1349,1412,1417)

cens <- c(1,1,1,1,1,0, rep(1,20), 0,1,1,1,1,1,0,1,0,1,1,1,1,0,1,1,1,0,1,0,0,0,1)
```

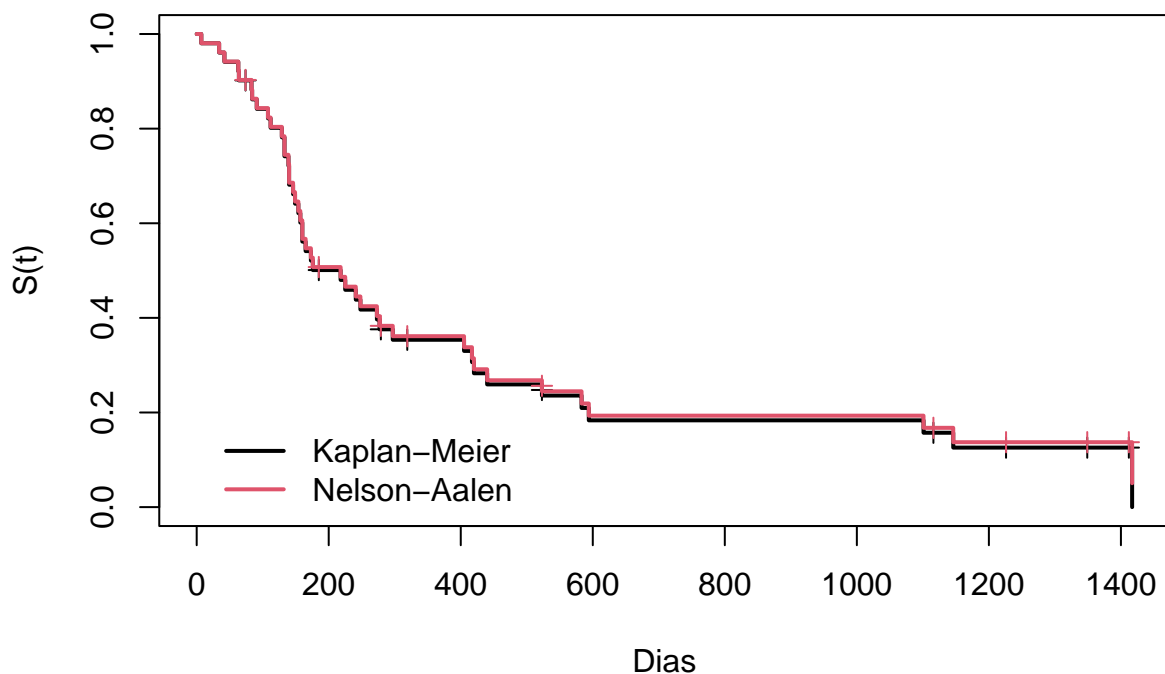
a) a função de sobrevida por meio dos estimadores de Kaplan-Meier e de Nelson-Aalen. Apresente-as em tabelas e gráficos;

```
# Estimador de Kaplan-Meier.
fit_KM <- survfit(Surv(tempos,cens)~1, type=c("kaplan-meier"), conf.type="log-log")
summary(fit_KM)
```

```
## Call: survfit(formula = Surv(tempos, cens) ~ 1, type = c("kaplan-meier"),
##      conf.type = "log-log")
##
##      time n.risk n.event survival std.err lower 95% CI upper 95% CI
##      7      51      1    0.980  0.0194    0.8689    0.997
##     34      50      1    0.961  0.0272    0.8522    0.990
##     42      49      1    0.941  0.0329    0.8286    0.981
##     63      48      1    0.922  0.0376    0.8044    0.970
##     64      47      1    0.902  0.0416    0.7804    0.958
##     83      45      1    0.882  0.0453    0.7559    0.945
##     84      44      1    0.862  0.0485    0.7319    0.932
##     91      43      1    0.842  0.0513    0.7084    0.918
##    108      42      1    0.822  0.0539    0.6854    0.903
##    112      41      1    0.802  0.0562    0.6627    0.888
##    129      40      1    0.782  0.0582    0.6405    0.873
##    133      39      2    0.742  0.0618    0.5969    0.841
##    139      37      1    0.722  0.0633    0.5755    0.825
##    140      36      2    0.681  0.0658    0.5336    0.791
##    146      34      1    0.661  0.0668    0.5130    0.774
##    149      33      1    0.641  0.0678    0.4927    0.757
##    154      32      1    0.621  0.0685    0.4726    0.739
##    157      31      1    0.601  0.0692    0.4527    0.721
##    160      30      2    0.561  0.0701    0.4136    0.685
##    165      28      1    0.541  0.0704    0.3944    0.667
##    173      27      1    0.521  0.0706    0.3754    0.648
##    176      26      1    0.501  0.0707    0.3566    0.629
##    218      24      1    0.480  0.0708    0.3370    0.610
##    225      23      1    0.459  0.0707    0.3177    0.590
##    241      22      1    0.438  0.0705    0.2987    0.570
##    248      21      1    0.418  0.0702    0.2799    0.549
##    273      20      1    0.397  0.0697    0.2614    0.529
##    277      19      1    0.376  0.0691    0.2431    0.508
##    297      17      1    0.354  0.0685    0.2238    0.486
##    405      15      1    0.330  0.0678    0.2031    0.463
##    417      14      1    0.307  0.0670    0.1830    0.439
##    420      13      1    0.283  0.0658    0.1634    0.415
##    440      12      1    0.259  0.0644    0.1443    0.390
##    523      11      1    0.236  0.0627    0.1258    0.365
##    583       9      1    0.210  0.0610    0.1053    0.338
##    594       8      1    0.183  0.0587    0.0858    0.310
##   1101       7      1    0.157  0.0559    0.0675    0.281
##   1146       5      1    0.126  0.0528    0.0458    0.248
##   1417       1      1    0.000      NaN      NA      NA
# Estimador de Nelson-Aalen.
fit_NeA <- survfit(coxph(Surv(tempos,cens)~1, method = "breslow"), data = surv_data, conf.type="log-log")
summary(fit_NeA)
## Call: survfit(formula = coxph(Surv(tempos, cens) ~ 1, method = "breslow"),
##      conf.type = "log-log", data = surv_data)
##
##      time n.risk n.event survival std.err lower 95% CI upper 95% CI
##      7      51      1    0.9806  0.0192    0.87006    0.997
##     34      50      1    0.9612  0.0269    0.85353    0.990
##     42      49      1    0.9417  0.0326    0.83018    0.981
##     63      48      1    0.9223  0.0373    0.80616    0.970
##     64      47      1    0.9029  0.0413    0.78235    0.958
##     83      45      1    0.8831  0.0449    0.75807    0.946
##     84      44      1    0.8632  0.0480    0.73433    0.932
##     91      43      1    0.8434  0.0509    0.71106    0.918
##    108      42      1    0.8235  0.0534    0.68822    0.904
##    112      41      1    0.8037  0.0557    0.66577    0.889
##    129      40      1    0.7839  0.0577    0.64367    0.874
##    133      39      2    0.7447  0.0611    0.60118    0.843
##    139      37      1    0.7248  0.0626    0.57995    0.827
##    140      36      2    0.6856  0.0651    0.53906    0.794
##    146      34      1    0.6658  0.0662    0.51856    0.777
```

```
## 149 33 1 0.6459 0.0671 0.49831 0.760
## 154 32 1 0.6260 0.0679 0.47829 0.743
## 157 31 1 0.6062 0.0686 0.45850 0.725
## 160 30 2 0.5671 0.0695 0.42040 0.690
## 165 28 1 0.5472 0.0699 0.40120 0.672
## 173 27 1 0.5273 0.0701 0.38222 0.653
## 176 26 1 0.5074 0.0702 0.36346 0.635
## 218 24 1 0.4867 0.0704 0.34392 0.615
## 225 23 1 0.4660 0.0703 0.32464 0.596
## 241 22 1 0.4453 0.0702 0.30563 0.576
## 248 21 1 0.4246 0.0699 0.28687 0.556
## 273 20 1 0.4038 0.0695 0.26837 0.535
## 277 19 1 0.3831 0.0690 0.25015 0.515
## 297 17 1 0.3613 0.0684 0.23086 0.493
## 405 15 1 0.3380 0.0678 0.21032 0.470
## 417 14 1 0.3147 0.0670 0.19028 0.447
## 420 13 1 0.2914 0.0660 0.17074 0.423
## 440 12 1 0.2681 0.0647 0.15171 0.399
## 523 11 1 0.2448 0.0631 0.13323 0.374
## 583 9 1 0.2190 0.0615 0.11282 0.348
## 594 8 1 0.1933 0.0594 0.09336 0.320
## 1101 7 1 0.1676 0.0568 0.07493 0.292
## 1146 5 1 0.1372 0.0540 0.05345 0.260
## 1417 1 1 0.0505 0.0542 0.00237 0.229
# Gráfico
plot(fit_KM, mark.time=T, conf.int=F, lwd=2, xlab='Dias', ylab='S(t)')
lines(fit_NeA, mark.time=T, conf.int=F, lwd=2, col=2)
legend(1,0.2, c('Kaplan-Meier', 'Nelson-Aalen'), lwd=2, col=1:2, bty='n')
title("Curvas de sobrevivência segundo tipo de estimador")
```

Curvas de sobrevivência segundo tipo de estimador



b) os tempos mediano e médio;

```
kable(t(summary(fit_KM)$tab1))
```

records	n.max	n.start	events	rmean	se(rmean)	median	0.95LCL	0.95UCL
51	51	51	42	422.3037	66.97964	218	149	297

O tempo médio de vida dos pacientes é de 422 dias e o tempo mediano de vida é de 218 dias

c) as probabilidades de um paciente com câncer sobreviver a:

d) 42 dias

$$\hat{S}(42) = \prod_{i/t_i < t} \left(1 - \frac{d_i}{n_i}\right) = 0.978$$

ii) 100 dias

$$\hat{S}(100) = 0.939$$

iii) 300 dias

$$\hat{S}(300) = 0.887$$

iv) 1000 dias

$$\hat{S}(1000) = 0.756$$

d) o tempo médio de vida restante dos pacientes que sobreviverem 1000 dias;

$$vmr(1000) = 326 \text{ dias}$$

f) para quais tempos tem-se: i) $\hat{S}(t) = 0$, 80, ii) $\hat{S}(t) = 0,30$ e $\hat{S}(t) = 0$, 10? Interprete.

(i) $t = 114$; (ii) $t = 418$ e $t = 1202$ dias

4º Questão

Os dados apresentados na Tabela 2.10 representam o tempo (em dias) até a morte de pacientes com câncer de ovário tratados na Mayo Clinic (Fleming et al., 1980).

```
tempos <- c(28,89,175,195,309,377,393,421,447,462,709,744,770,1106,1206,34,88,137,
            199,280,291,299,
            300,309,351,358,369,369,370,375,382,392,429,451,1119)
cens <- c(1,1,1,1,1,0,0,0,0,1,0,0,0,0,0,1,1,1,1,1,0,0,1,1,1,1,1,1,1,1,0,1,0)
Tumor <- c(rep("Grande",15), rep("Pequeno",20))
```

```
surv_data <- data.frame(tempos,cens,Tumor)
```

```
km <- survfit(Surv(tempos,cens)~Tumor, conf.type="log-log")
```

```
summary(km)
```

```
## Call: survfit(formula = Surv(tempos, cens) ~ Tumor, conf.type = "log-log")
```

```
##
```

```
##           Tumor=Grande
```

##	time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
##	28	15	1	0.933	0.0644	0.613	0.990
##	89	14	1	0.867	0.0878	0.564	0.965
##	175	13	1	0.800	0.1033	0.500	0.931
##	195	12	1	0.733	0.1142	0.436	0.891
##	309	11	1	0.667	0.1217	0.375	0.846
##	462	6	1	0.556	0.1434	0.249	0.780

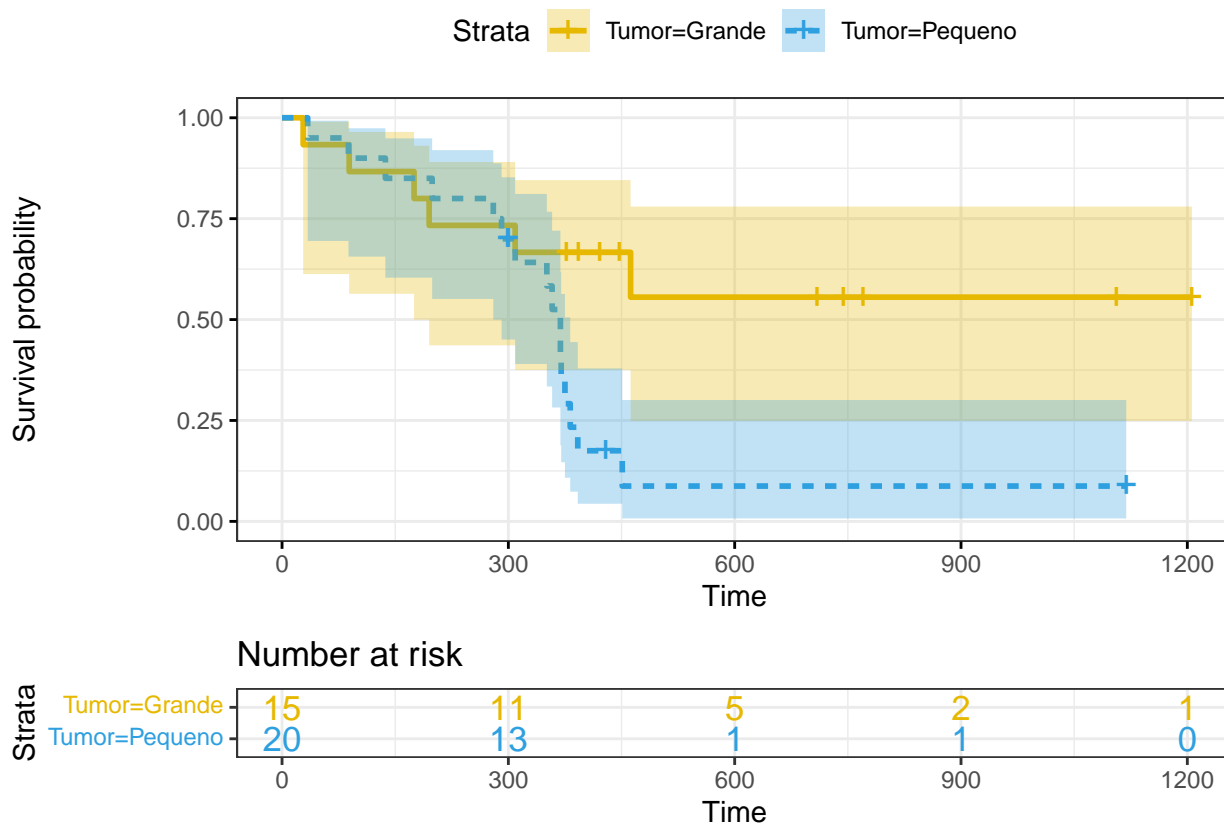
```
##
```

```
##           Tumor=Pequeno
```

##	time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
##	34	20	1	0.9500	0.0487	0.69474	0.993
##	88	19	1	0.9000	0.0671	0.65603	0.974
##	137	18	1	0.8500	0.0798	0.60379	0.949
##	199	17	1	0.8000	0.0894	0.55115	0.920
##	280	16	1	0.7500	0.0968	0.49994	0.887
##	291	15	1	0.7000	0.1025	0.45055	0.853
##	309	12	1	0.6417	0.1093	0.39006	0.811
##	351	11	1	0.5833	0.1139	0.33420	0.767
##	358	10	1	0.5250	0.1165	0.28218	0.720
##	369	9	2	0.4083	0.1162	0.18837	0.618
##	370	7	1	0.3500	0.1133	0.14645	0.563
##	375	6	1	0.2917	0.1084	0.10804	0.505
##	382	5	1	0.2333	0.1012	0.07354	0.444
##	392	4	1	0.1750	0.0912	0.04366	0.379
##	451	2	1	0.0875	0.0769	0.00717	0.301

a) Obtenha as estimativas de Kaplan-Meier para as funções de so-brevivência de ambos os grupos e apresente-as no mesmo gráfico.

```
ggsurvplot(km, data = surv_data,
  pval = F, conf.int = TRUE,
  risk.table = TRUE,
  risk.table.col = "strata",
  linetype = "strata",
  ggtheme = theme_bw(),
  palette = c("#E7B800", "#2E9FDF"))
```

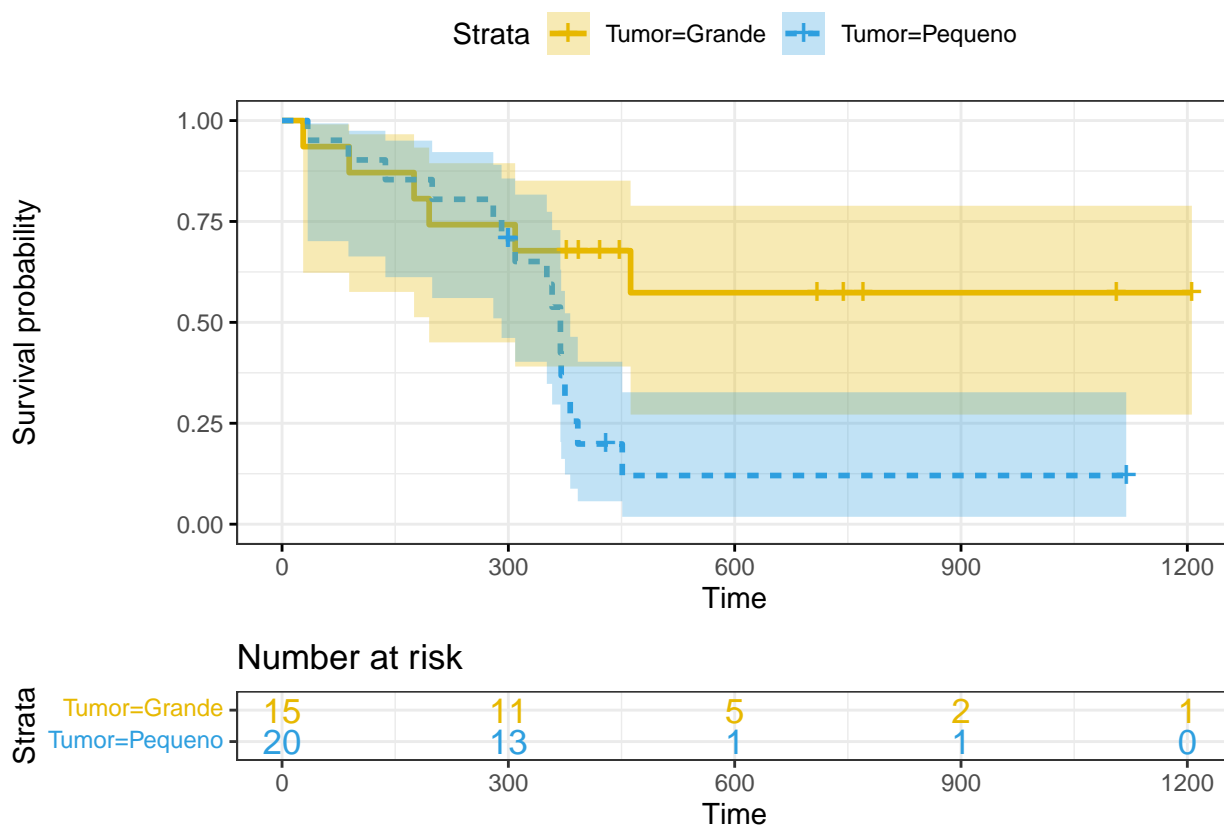


b) Repita a letra (a) utilizando, agora, o estimador de Nelson-Aalen.

```
na<- survfit(Surv(tempo,cens)~Tumor, conf.type="log-log",type='fh')
summary(na)
## Call: survfit(formula = Surv(tempo, cens) ~ Tumor, conf.type = "log-log",
## type = "fh")
##
##
##           Tumor=Grande
## time  n.risk  n.event survival std.err lower 95% CI upper 95% CI
##  28      15       1    0.936  0.0624    0.623    0.991
##  89      14       1    0.871  0.0851    0.575    0.966
## 175      13       1    0.807  0.1003    0.513    0.933
## 195      12       1    0.742  0.1111    0.450    0.894
## 309      11       1    0.678  0.1187    0.391    0.851
## 462       6       1    0.574  0.1387    0.272    0.789
##
##
##           Tumor=Pequeno
## time  n.risk  n.event survival std.err lower 95% CI upper 95% CI
##  34      20       1    0.951  0.0476    0.7012    0.993
##  88      19       1    0.902  0.0655    0.6633    0.975
## 137      18       1    0.854  0.0780    0.6120    0.950
## 199      17       1    0.805  0.0875    0.5603    0.922
## 280      16       1    0.756  0.0948    0.5100    0.890
## 291      15       1    0.707  0.1005    0.4614    0.856
## 309      12       1    0.651  0.1072    0.4023    0.817
## 351      11       1    0.594  0.1118    0.3476    0.774
## 358      10       1    0.538  0.1145    0.2964    0.729
## 369       9       2    0.425  0.1150    0.2036    0.631
## 370       7       1    0.368  0.1127    0.1617    0.578
```

```
## 375 6 1 0.312 0.1086 0.1231 0.523
## 382 5 1 0.255 0.1025 0.0879 0.464
## 392 4 1 0.199 0.0940 0.0567 0.402
## 451 2 1 0.121 0.0830 0.0182 0.327
```

```
ggsurvplot(na, data = surv_data,
  pval = F, conf.int = TRUE,
  risk.table = TRUE,
  risk.table.col = "strata",
  linetype = "strata",
  ggtheme = theme_bw(),
  palette = c("#E7B800", "#2E9FDF"))
```



- c) Usando os intervalos de confiança assintóticos das estimativas de Kaplan-Meier, teste a hipótese de igualdade das funções de sobrevivência dos dois grupos em $t = 6$ meses e $t = 15$ meses.
- d) Teste a hipótese de igualdade das funções de sobrevivência dos dois grupos usando dois testes diferentes. Os resultados dos testes são consistentes? Em caso negativo, explique a razão da diferença dos resultados.

```
# teste de log-rank
surv_diff <- survdiff(Surv(tempo,cens)~Tumor, data = surv_data)
surv_diff
## Call:
## survdiff(formula = Surv(tempo, cens) ~ Tumor, data = surv_data)
##
##              N Observed Expected (O-E)^2/E (O-E)^2/V
## Tumor=Grande 15      6      11.3    2.51    5.57
## Tumor=Pequeno 20     16     10.7    2.67    5.57
##
## Chisq= 5.6 on 1 degrees of freedom, p= 0.02
```

A estatística do teste qui-quadrado é 5,6 com 1 grau de liberdade e o valor p correspondente é 0,02. Como esse valor-p é menor que 0,05, rejeitamos a hipótese nula.

```
# teste de Wilcoxon
wilcox.test(tempo[Tumor=="Grande"], tempo[Tumor=="Pequeno"])
##
## Wilcoxon rank sum test with continuity correction
##
## data:  tempo[Tumor == "Grande"] and tempo[Tumor == "Pequeno"]
```

```
## W = 198.5, p-value = 0.1095
```

```
## alternative hypothesis: true location shift is not equal to 0
```