DSBD - Análise de Sobrevivência

José Luiz Padilha

20 de setembro de 2024

Exemplo 1: Estimador de Kaplan-Meier e Testes Não Paramétricos

Estimador de Kaplan-Meier

Considere como exemplo os dados de hepatite. O estudo tinha como objetivo investigar o efeito da terapia com esteroide em pacientes com hepatite viral aguda. O evento de interesse é a morte do paciente.

```
tempos <- c(1, 2, 3, 3, 3, 5, 5, 16, 16, 16, 16, 16, 16, 16, 16, 1, 1, 1, 1, 4, 5, 7, 8,
            10, 10, 12, 16, 16, 16)
cens <- c(0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1, 0, 0, 1, 1, 1, 1, 0, 0, 0,
          0, 0)
grupos <- c(rep("Controle", 15), rep("Esteroide", 14))
dados <- data.frame(tempos, cens, grupos)
```

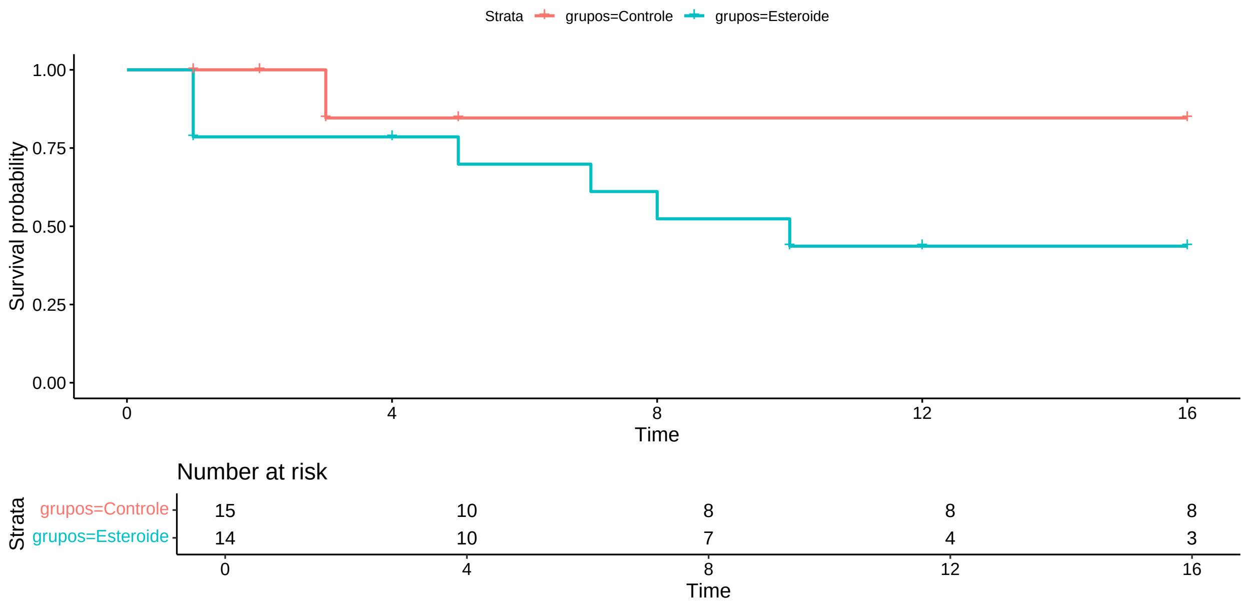
O estimador de Kaplan-Meier é obtido por meia da função `survfit` do pacote `survival`.

```
require(survival)
require(survminer)
ekm <- survfit(Surv(tempos, cens) ~ grupos, data = dados)
summary(ekm)
```

```
## Call: survfit(formula = Surv(tempos, cens) ~ grupos, data = dados)
##
##                 grupos=Controle
##         time       n.risk      n.event     survival      std.err lower 95% CI
##        3.000       13.000        2.000        0.846        0.100        0.671
## upper 95% CI
##        1.000
##
##                 grupos=Esteroide
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##     1     14       3    0.786   0.110        0.598        1.000
##     5      9       1    0.698   0.128        0.488        0.999
##     7      8       1    0.611   0.138        0.392        0.952
##     8      7       1    0.524   0.143        0.306        0.896
##    10      6       1    0.437   0.144        0.229        0.832
```

A seguir, temos uma representação gráfica das curvas de sobrevivência estimadas, o que nos permite uma comparação visual entre funções de sobrevivência para os dois grupos.

```
ggsurvplot(ekm, risk.table = TRUE)
```



Diferentes intervalos de confiança para a sobrevivência podem ser obtidos alterando o argumento `conf.type`. A opção padrão é `conf.type="log"`.

```
#fórmula anterior apresentada
ekm0 <- survfit(Surv(tempos, cens) ~ grupos, conf.type = "plain", data = dados)
summary(ekm0)
```

```
## Call: survfit(formula = Surv(tempos, cens) ~ grupos, data = dados,
##     conf.type = "plain")
##
##                 grupos=Controle
##         time       n.risk      n.event     survival      std.err lower 95% CI
##        3.000       13.000        2.000        0.846        0.100        0.650
## upper 95% CI
##        1.000
##
##                 grupos=Esteroide
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##     1     14       3    0.786   0.110        0.571        1.000
##     5      9       1    0.698   0.128        0.448        0.948
##     7      8       1    0.611   0.138        0.340        0.882
##     8      7       1    0.524   0.143        0.243        0.805
##    10      6       1    0.437   0.144        0.155        0.718
```

```
#transformação log-log para S(t)
ekm1 <- survfit(Surv(tempos, cens) ~ grupos, conf.type = "log-log", data = dados)
summary(ekm1)
```

```
## Call: survfit(formula = Surv(tempos, cens) ~ grupos, data = dados,
##     conf.type = "log-log")
##
##                 grupos=Controle
##         time       n.risk      n.event     survival      std.err lower 95% CI
##        3.000       13.000        2.000        0.846        0.100        0.512
## upper 95% CI
```

```
##        0.959
##
##                  grupos=Esteroide
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##     1     14       3    0.786   0.110        0.472        0.925
##     5      9       1    0.698   0.128        0.378        0.876
##     7      8       1    0.611   0.138        0.298        0.819
##     8      7       1    0.524   0.143        0.227        0.754
##    10      6       1    0.437   0.144        0.164        0.683
```

A probabilidade estimada de um paciente do grupo esteroide sobreviver a 6 semanas obtida diretamente da curva de Kaplan-Meier é de $0,698$. No entanto, se uma interpolação linear for utilizada, obtém-se:

$$\frac{7-5}{0,611-0,698} = \frac{6-5}{\hat{S}(6)-0,698} \Rightarrow \hat{S}(6) = 0,655.$$

Uma informação útil é o tempo mediano de vida. Como a curva de Kaplan-Meier é uma função escada, esta estimativa obtida por meio de uma interpolação linear é:

$$\frac{10-8}{0,437-0,524} = \frac{MED-8}{0,500-0,524} \Rightarrow MED = t_{0,50} = 8,55 \ semanas.$$

Comparação das curvas de sobrevivência

As curvas de Kaplan-Meier indicam que, possivelmente, a terapia com esteroide não é um tratamento adequado para pacientes com hepatite viral aguda. Procedemos com a aplicação dos testes logrank e de Wilcoxon para comparação das curvas de sobrevivência.

```r
#teste logrank
survdiff(Surv(tempos,cens) ~ grupos, rho = 0)
```

```
## Call:
## survdiff(formula = Surv(tempos, cens) ~ grupos, rho = 0)
##
##                   N Observed Expected (O-E)^2/E (O-E)^2/V
## grupos=Controle  15        2     4.81      1.64      3.67
## grupos=Esteroide 14        7     4.19      1.89      3.67
##
##  Chisq= 3.7  on 1 degrees of freedom, p= 0.06
```

```r
#teste wilcoxon
survdiff(Surv(tempos, cens) ~ grupos, rho = 1)
```

```
## Call:
## survdiff(formula = Surv(tempos, cens) ~ grupos, rho = 1)
##
##                   N Observed Expected (O-E)^2/E (O-E)^2/V
## grupos=Controle  15     1.79     4.16      1.35      3.43
## grupos=Esteroide 14     6.00     3.63      1.54      3.43
##
##  Chisq= 3.4  on 1 degrees of freedom, p= 0.06
```

Os testes apontam uma possível diferença entre os dois grupos. O pacote `coin` implementa uma série de testes para testar a igualdade de funções de sobrevivência em dois ou mais grupos independentes. Por exemplo, para realizarmos o teste de Tarone-Ware, podemos fazer:

```r
library(coin)
logrank_test(Surv(tempos, cens) ~ factor(grupos), type = "Tarone-Ware")
```

```
##
##  Asymptotic Two-Sample Tarone-Ware Test
##
## data:  Surv(tempos, cens) by factor(grupos) (Controle, Esteroide)
## Z = 1.859, p-value = 0.06303
## alternative hypothesis: true theta is not equal to 1
```

Exemplo 2: Estimador de Kaplan-Meier e Testes Não Paramétricos

Um produtor de requeijão deseja comparar dois tipos de embalagens (A e B) para o seu produto. Ele deseja saber se existe diferença na durabilidade de seu produto com relação às embalagens. O produto dele é vendido em temperatura ambiente e sem conservantes. O evento de interesse é o aparecimento de algum tipo de fungo no produto.

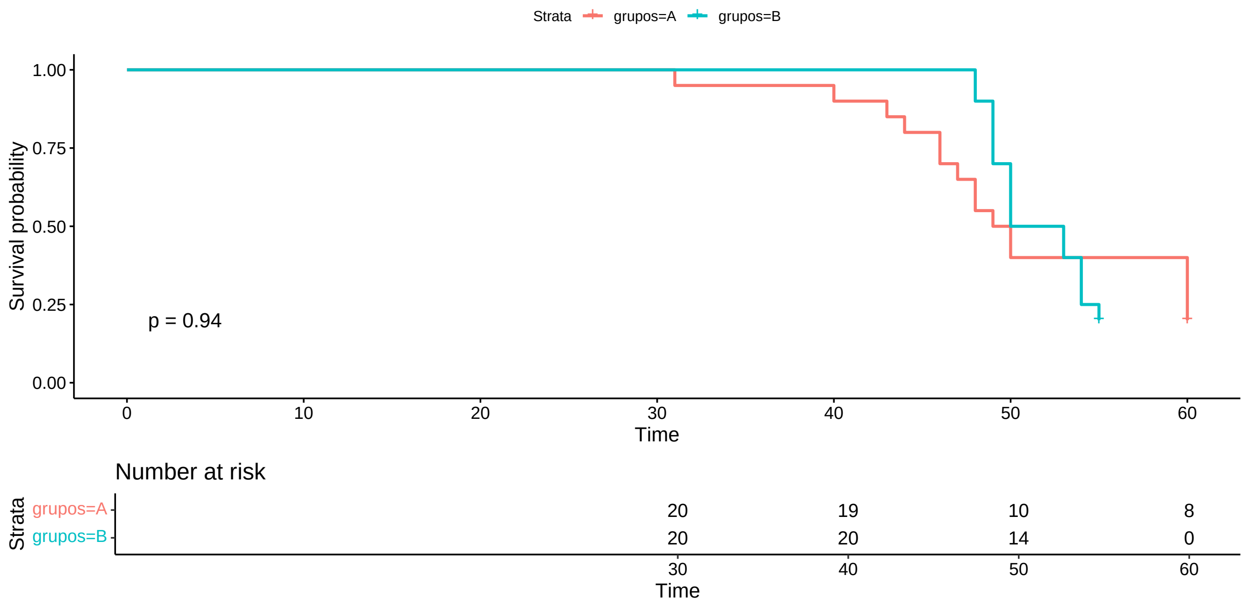
Vamos verificar se existe diferença entre as embalagens.

```
tempos.a <- c(31, 40, 43, 44, 46, 46, 47, 48, 48, 49, 50, 50, 60, 60, 60, 60, 60, 60, 60,
              60)
cens.a <- c(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0)
tempos.b <- c(48, 48, 49, 49, 49, 49, 50, 50, 50, 50, 53, 53, 54, 54, 54, 55, 55, 55, 55,
              55)
cens.b <- c(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0)
tempos <- c(tempos.a, tempos.b)
cens <- c(cens.a, cens.b)
grupos <- c(rep("A", length(tempos.a)), rep("B", length(tempos.b)))
dados <- data.frame(tempos, cens, grupos)
ekm <- survfit(Surv(tempos, cens) ~ grupos, data = dados); summary(ekm)
```

```
## Call: survfit(formula = Surv(tempos, cens) ~ grupos, data = dados)
##
##                 grupos=A
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    31     20       1     0.95  0.0487       0.8591        1.000
##    40     19       1     0.90  0.0671       0.7777        1.000
##    43     18       1     0.85  0.0798       0.7071        1.000
##    44     17       1     0.80  0.0894       0.6426        0.996
##    46     16       2     0.70  0.1025       0.5254        0.933
##    47     14       1     0.65  0.1067       0.4712        0.897
##    48     13       2     0.55  0.1112       0.3700        0.818
##    49     11       1     0.50  0.1118       0.3226        0.775
##    50     10       2     0.40  0.1095       0.2339        0.684
##    60      8       4     0.20  0.0894       0.0832        0.481
##
##                 grupos=B
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    48     20       2     0.90  0.0671       0.7777        1.000
##    49     18       4     0.70  0.1025       0.5254        0.933
##    50     14       4     0.50  0.1118       0.3226        0.775
##    53     10       2     0.40  0.1095       0.2339        0.684
##    54      8       3     0.25  0.0968       0.1170        0.534
##    55      5       1     0.20  0.0894       0.0832        0.481
```

A seguir, os gráficos com as curvas estimadas.

```
ggsurvplot(ekm, risk.table = TRUE, pval = TRUE)
```




```
survdiff(Surv(tempos,cens) ~ grupos, rho = 0)#teste logrank
```

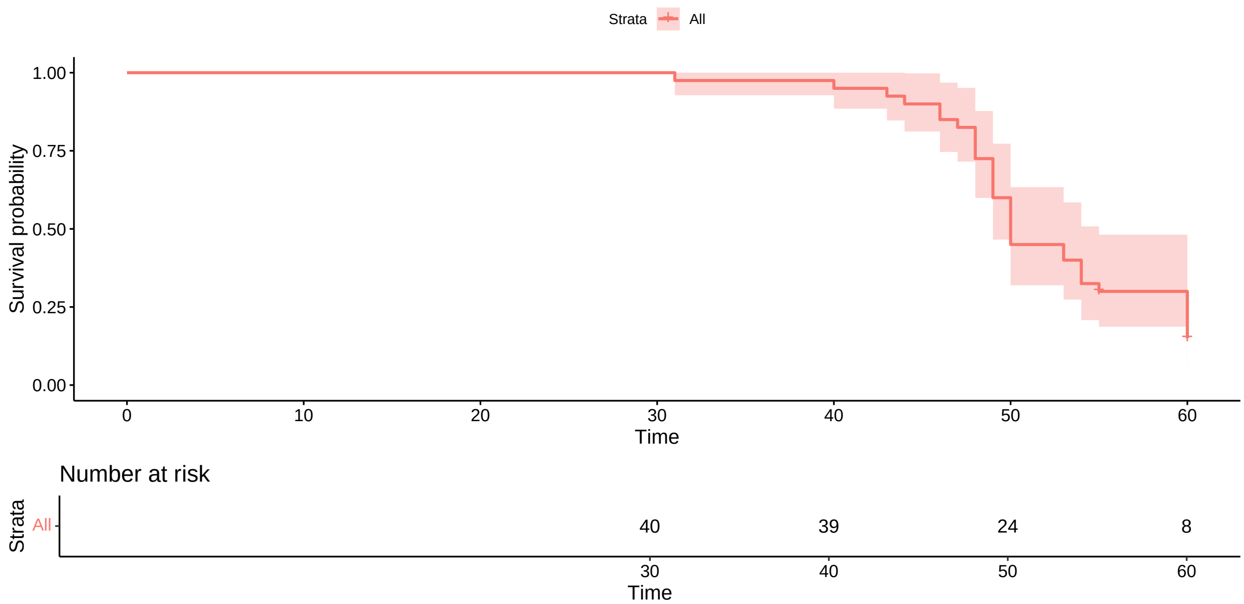
```
## Call:
## survdiff(formula = Surv(tempos, cens) ~ grupos, rho = 0)
##
##             N Observed Expected (O-E)^2/E (O-E)^2/V
## grupos=A 20       16     16.2   0.00241   0.00641
## grupos=B 20       16     15.8   0.00247   0.00641
##
##  Chisq= 0  on 1 degrees of freedom, p= 0.9
```

O teste logrank não apresenta diferença entre as curvas. Vamos construir uma curva de sobrevivência combinando todos os tempos de vida e plotá-la juntamente com o intervalo de confiança de 95%.

```
ekm <- survfit(Surv(tempos,cens) ~ 1, data = dados); summary(ekm)
```

```
## Call: survfit(formula = Surv(tempos, cens) ~ 1, data = dados)
##
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    31     40       1    0.975  0.0247       0.9278        1.000
##    40     39       1    0.950  0.0345       0.8848        1.000
##    43     38       1    0.925  0.0416       0.8469        1.000
##    44     37       1    0.900  0.0474       0.8117        0.998
##    46     36       2    0.850  0.0565       0.7462        0.968
##    47     34       1    0.825  0.0601       0.7153        0.952
##    48     33       4    0.725  0.0706       0.5990        0.877
##    49     29       5    0.600  0.0775       0.4659        0.773
##    50     24       6    0.450  0.0787       0.3195        0.634
##    53     18       2    0.400  0.0775       0.2737        0.585
##    54     16       3    0.325  0.0741       0.2079        0.508
##    55     13       1    0.300  0.0725       0.1869        0.482
##    60      8       4    0.150  0.0642       0.0648        0.347
```

```
ggsurvplot(ekm, risk.table = TRUE)
```


Strata
All
Survival probability
1.00
0.75
0.50
0.25
0.00
0
10
20
30
40
50
60
Time
Number at risk
Strata
All
40
39
24
8
30
40
50
60
Time

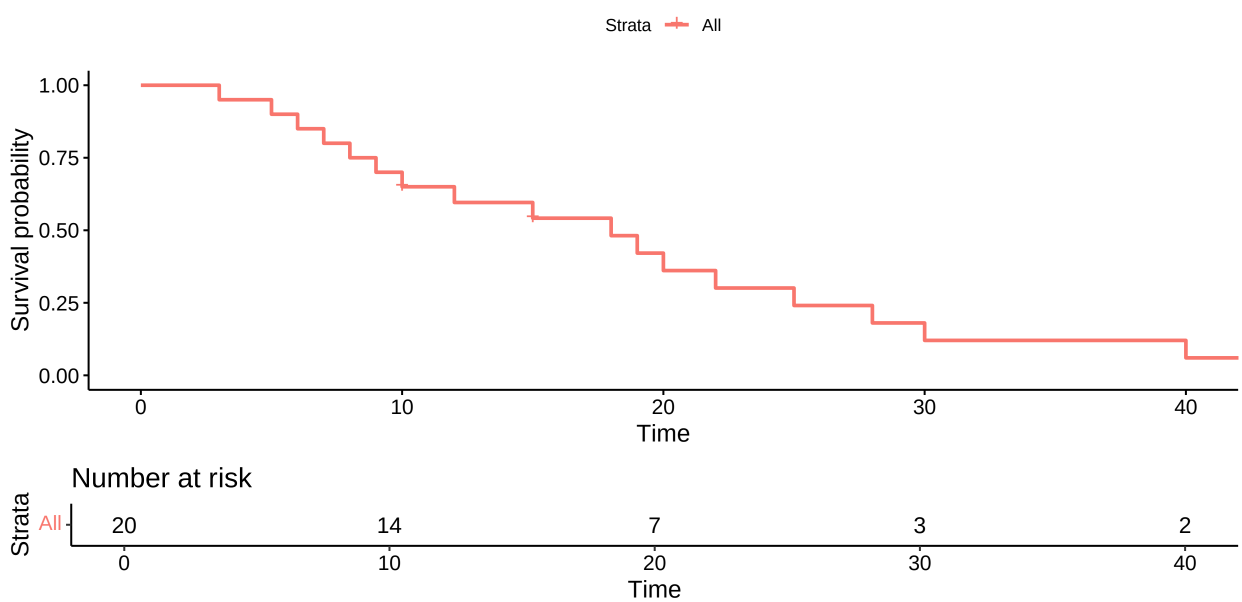
Exemplo 3: Modelos Probabilísticos

Vamos considerar os tempos de reincidência, em meses, de um grupo de 20 pacientes com câncer de bexiga que foram submetidos a um procedimento cirúrgico realizado por laser.

```
tempos <- c(3, 5, 6, 7, 8, 9, 10, 10, 12, 15, 15, 18, 19, 20, 22, 25, 28, 30, 40, 45)
cens <- c(1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 0)
dados <- data.frame(tempos, cens)
ekm <- survfit(Surv(tempos, cens) ~ 1, data = dados)
summary(ekm)
```

```
## Call: survfit(formula = Surv(tempos, cens) ~ 1, data = dados)
##
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##     3     20       1   0.9500  0.0487      0.85913        1.000
##     5     19       1   0.9000  0.0671      0.77767        1.000
##     6     18       1   0.8500  0.0798      0.70707        1.000
##     7     17       1   0.8000  0.0894      0.64257        0.996
##     8     16       1   0.7500  0.0968      0.58233        0.966
##     9     15       1   0.7000  0.1025      0.52541        0.933
##    10     14       1   0.6500  0.1067      0.47124        0.897
##    12     12       1   0.5958  0.1107      0.41402        0.857
##    15     11       1   0.5417  0.1131      0.35976        0.816
##    18      9       1   0.4815  0.1154      0.30096        0.770
##    19      8       1   0.4213  0.1156      0.24601        0.721
##    20      7       1   0.3611  0.1137      0.19481        0.669
##    22      6       1   0.3009  0.1095      0.14745        0.614
##    25      5       1   0.2407  0.1028      0.10422        0.556
##    28      4       1   0.1806  0.0931      0.06573        0.496
##    30      3       1   0.1204  0.0792      0.03317        0.437
##    40      2       1   0.0602  0.0581      0.00907        0.399
```

```
ggsurvplot(ekm, conf.int = FALSE, risk.table = TRUE)
```



Ajustaremos três distribuições: exponencial, Weibull e lognormal.

```r
ajust1 <- survreg(Surv(tempos, cens) ~ 1, dist = 'exponential')
ajust1
```

```
## Call:
## survreg(formula = Surv(tempos, cens) ~ 1, dist = "exponential")
##
## Coefficients:
## (Intercept)
##    3.016111
##
## Scale fixed at 1
##
## Loglik(model)= -68.3   Loglik(intercept only)= -68.3
## n= 20
```

```r
alpha <- exp(ajust1$coefficients[1])
alpha
```

```
## (Intercept)
##    20.41176
```

```r
#
ajust2 <- survreg(Surv(tempos, cens) ~ 1, dist = 'weibull')
ajust2
```

```
## Call:
## survreg(formula = Surv(tempos, cens) ~ 1, dist = "weibull")
##
## Coefficients:
## (Intercept)
##    3.060529
##
## Scale= 0.647922
##
## Loglik(model)= -66.1   Loglik(intercept only)= -66.1
## n= 20
```

```r
alpha <- exp(ajust2$coefficients[1])
gama <- 1/ajust2$scale
cbind(gama, alpha)
```

```
##                 gama    alpha
## (Intercept) 1.543396 21.33885
```

```r
#
ajust3 <- survreg(Surv(tempos, cens) ~ 1, dist = 'lognormal')
ajust3
```

```
## Call:
## survreg(formula = Surv(tempos, cens) ~ 1, dist = "lognormal")
##
## Coefficients:
## (Intercept)
##    2.717176
##
## Scale= 0.7648167
##
```

```
## Loglik(model)= -65.7   Loglik(intercept only)= -65.7
## n= 20
```

Com base nos valores acima, podemos calcular as sobrevivências estimadas. Por exemplo, para o tempo $t = 10$ meses, obtemos:

```r
exp(-10/20.41)#exponencial
```

```
## [1] 0.6126534
```

```r
exp(-(10/21.34)^1.54)#Weibull
```

```
## [1] 0.7325668
```

```r
pnorm((-log(10) + 2.72)/0.76)#lognormal
```

```
## [1] 0.7085762
```

As estimativas obtidas por meio dos modelos Weibull e lognormal são bem próximas. O mesmo não ocorre com o modelo exponencial, que apresenta um valor ligeiramente diferente dos obtidos pelos outros dois modelos.

A seguir construímos as estimativas das funções de sobrevivência para os tempos de reincidência usando-se os modelos exponencial, Weibull e lognormal, além do Kaplan-Meier.

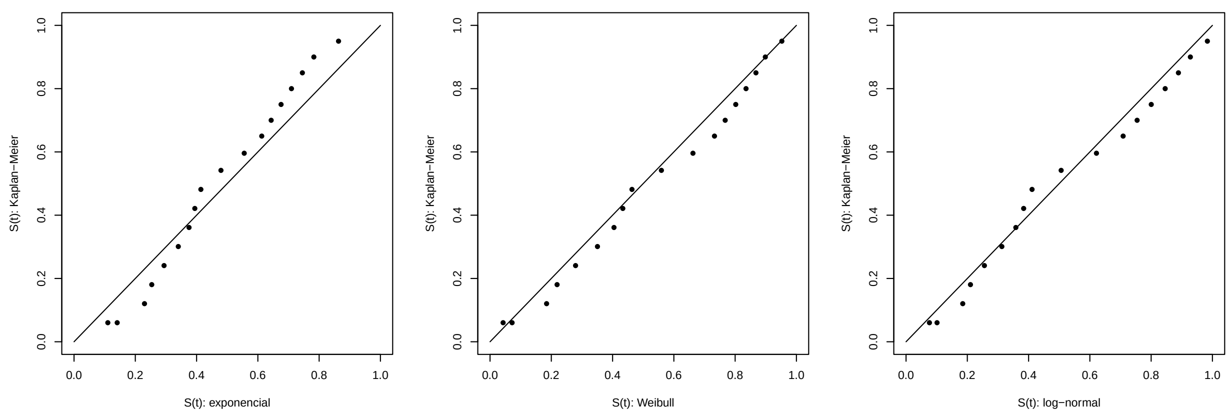
```r
time <- ekm$time
st <- ekm$surv
ste <- exp(-time/20.41)
stw <- exp(-(time/21.34)^1.54)
stln <- pnorm((-log(time) + 2.72)/0.76)
round(cbind(time, st, ste, stw, stln), 3)
```

```
##       time    st   ste   stw  stln
##  [1,]    3 0.950 0.863 0.952 0.984
##  [2,]    5 0.900 0.783 0.899 0.928
##  [3,]    6 0.850 0.745 0.868 0.889
##  [4,]    7 0.800 0.710 0.836 0.846
##  [5,]    8 0.750 0.676 0.802 0.800
##  [6,]    9 0.700 0.643 0.768 0.754
##  [7,]   10 0.650 0.613 0.733 0.709
##  [8,]   12 0.596 0.555 0.662 0.621
##  [9,]   15 0.542 0.480 0.559 0.506
## [10,]   18 0.481 0.414 0.463 0.411
## [11,]   19 0.421 0.394 0.433 0.384
## [12,]   20 0.361 0.375 0.405 0.358
## [13,]   22 0.301 0.340 0.351 0.313
## [14,]   25 0.241 0.294 0.279 0.256
## [15,]   28 0.181 0.254 0.219 0.210
## [16,]   30 0.120 0.230 0.185 0.185
## [17,]   40 0.060 0.141 0.072 0.101
## [18,]   45 0.060 0.110 0.043 0.076
```

Para escolha dos modelos, vamos inicialmente usar o método gráfico comparando as estimativas de sobrevivência pelo estimador de Kaplan Meier versus as respectivas estimavas por cada um destes modelos.

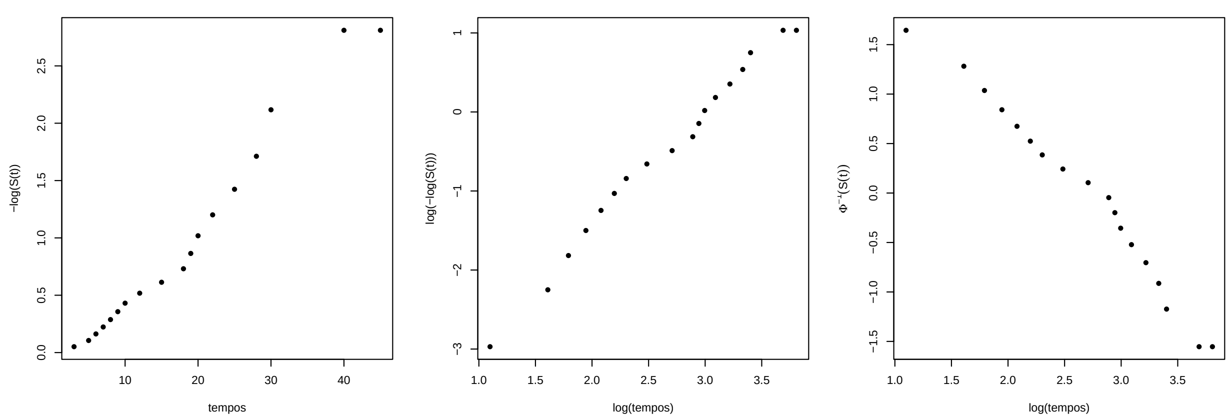
```r
par(mfrow = c(1, 3))
plot(ste, st, pch = 16, ylim = range(c(0.0, 1)), xlim = range(c(0, 1)),
     ylab = "S(t): Kaplan-Meier", xlab = "S(t): exponencial")
lines(c(0, 1), c(0, 1), type = "l", lty = 1)
```

```r
plot(stw, st, pch = 16, ylim = range(c(0.0, 1)), xlim = range(c(0, 1)),
     ylab = "S(t): Kaplan-Meier", xlab = "S(t): Weibull")
lines(c(0, 1), c(0, 1), type = "l", lty = 1)
plot(stln, st, pch = 16, ylim = range(c(0.0, 1)), xlim = range(c(0, 1)),
     ylab = "S(t): Kaplan-Meier", xlab = "S(t): log-normal")
lines(c(0, 1), c(0, 1), type = "l", lty = 1)
```



O modelo exponencial não parece ser adequado, pois há afastamentos da reta $y = x$. Os modelos Weibull e lognormal acompanham melhor a reta de igualdade. Vamos proceder com a comparação dos gráficos linearizados.

```r
par(mfrow = c(1, 3))
invst <- qnorm(st)
plot(time, -log(st), pch = 16, xlab = "tempos", ylab = "-log(S(t))")
plot(log(time), log(-log(st)), pch = 16, xlab = "log(tempos)", ylab = "log(-log(S(t)))")
plot(log(time), invst, pch = 16, xlab = "log(tempos)", ylab = expression(Phi^-1 * (S(t))))
```

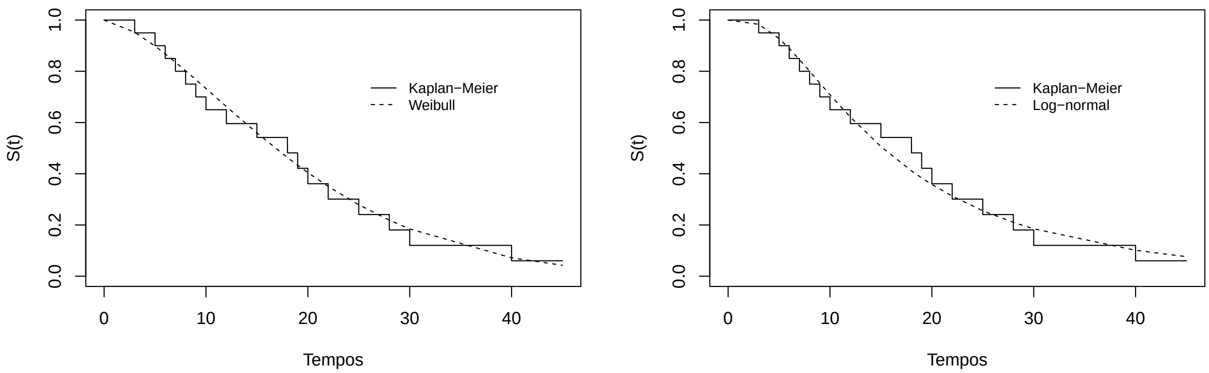


Novamente vemos que o modelo exponencial não apresenta bom ajuste, enquanto os outros dois modelos mostram comportamentos similares.

As curvas de sobrevivência estimadas pelos modelos Weibull e lognormal são mostradas a seguir, juntamente com a sobrevivência estimada por Kaplan-Meier.

```r
par(mfrow = c(1, 2))
plot(ekm, conf.int = F, xlab = "Tempos", ylab = "S(t)")
```

```r
lines(c(0, time), c(1, stw), lty = 2)
legend(25, 0.8, lty = c(1, 2), c("Kaplan-Meier", "Weibull"), bty = "n", cex = 0.8)
plot(ekm, conf.int = F, xlab = "Tempos", ylab = "S(t)")
lines(c(0, time), c(1, stln), lty = 2)
legend(25, 0.8, lty = c(1,2), c("Kaplan-Meier", "Log-normal"), bty = "n", cex = 0.8)
```



A figura indica que ambos os modelos apresentam ajustes satisfatórios. O logaritmo da verossimilhança dos modelos ajustados e o correspondente AIC são dadas por:

```r
cbind(loglik = c(ajust1$loglik[2], ajust2$loglik[2], ajust3$loglik[2]),
      aic = c(extractAIC(ajust1)[2], extractAIC(ajust2)[2], extractAIC(ajust3)[2]))
```

```
##         loglik      aic
## [1,] -68.27389 138.5478
## [2,] -66.13336 136.2667
## [3,] -65.73990 135.4798
```

Estimativas para o tempo médio são obtidas para os modelos Weibull e lognormal como, respectivamente:

$$\hat{E}(T) = 21,34[\Gamma(1 + (1/1,54))] = 19,206 \text{ meses},$$
$$\hat{E}(T) = \exp\left\{2,72 + (0,76^2/2)\right\} = 20,263 \text{ meses}.$$

Ainda, uma estimativa para o tempo mediano, obtida a partir da expressão dos percentis do modelo lognormal é

$$\hat{t}_{0,5} = \exp(z_{0,5}0,76 + 2,72) = 15,18 \text{ meses}.$$

O estimador de Kaplan Meier, fazendo uso de interpolação linear, fornece um valor de 17,05 meses como estimativa para o tempo mediano.

Exemplo 4: Modelos de Regressão Paramétricos

Os dados utilizados provêm do Registro Hospitalar do Câncer (RHC) do Hospital Erasto Gaertner. O RHC foi implantado em novembro de 1992. A amostra é formada por pacientes com câncer maligno de localização topográfica C71 (encéfalo). Os registros vão de 17/05/1990 a 30/12/2001.

A amostra é composta por 397 pacientes. As variáveis disponíveis são:

- Sexo: 246 do sexo masculino e 151 do sexo feminino.
- Idade: Varia de 0 a 77 anos, sem grandes concentrações
- Tratamento realizado:
  - Radioterapia: 257
  - Radioterapia+Cirurgia: 71
  - Cirurgia: 21
  - Outros: 48 (Quimioterapia, Hormonioterapia ou combinações de tratamentos).
- Estadiamento da doença:
  - I:6 II: 40 III: 28 IV: 12
  - Não pode ser aplicado: 2
  - Não codificado: 308
- AED: Avaliação da extensão da doença:
  - Localizado: 350
    * Extensão direta: 33
    * Metástese: 5
    * Não aplicável: 2
    * Ignorado: 8

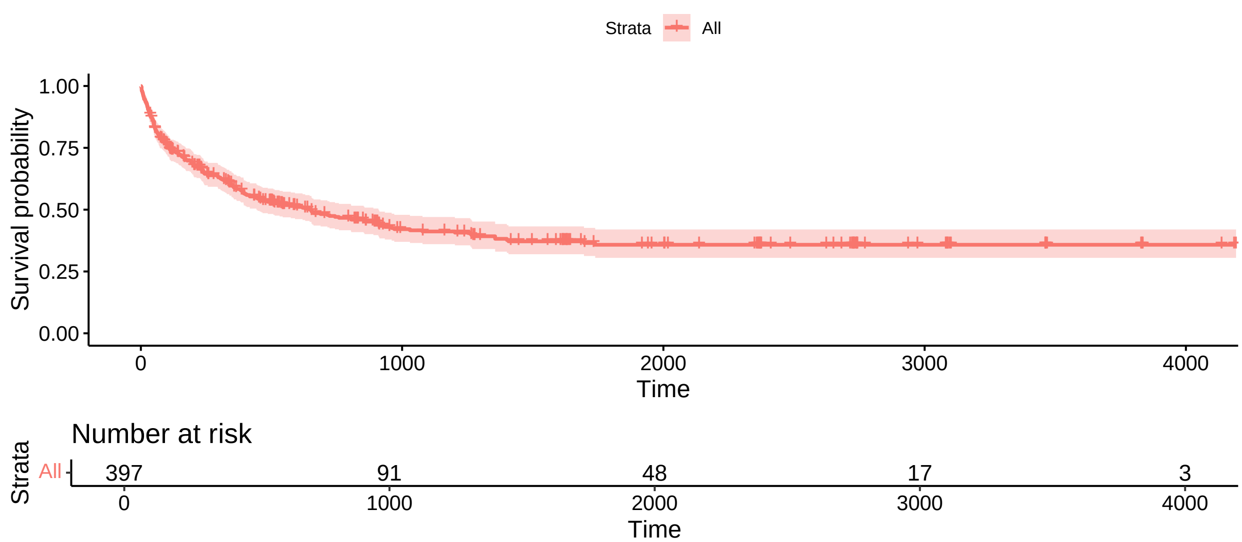
Foram observadas 216 falhas e 181 censuras. As covariáveis utilizadas nesta aplicação foram: idade do paciente, sexo e tipo de tratamento realizado.

```
encefalo <- read.table("enc.txt", h = T, dec = ",")
```

Gráfico de Kaplan-Meier e testes logrank

A seguir o gráfico marginal de sobrevivência estimada.

```
ekm <- survfit(Surv(tempo, cens.1) ~ 1, data = encefalo)
ggsurvplot(ekm, risk.table = TRUE)
```



Uma estimativa do tempo mediano é dada por

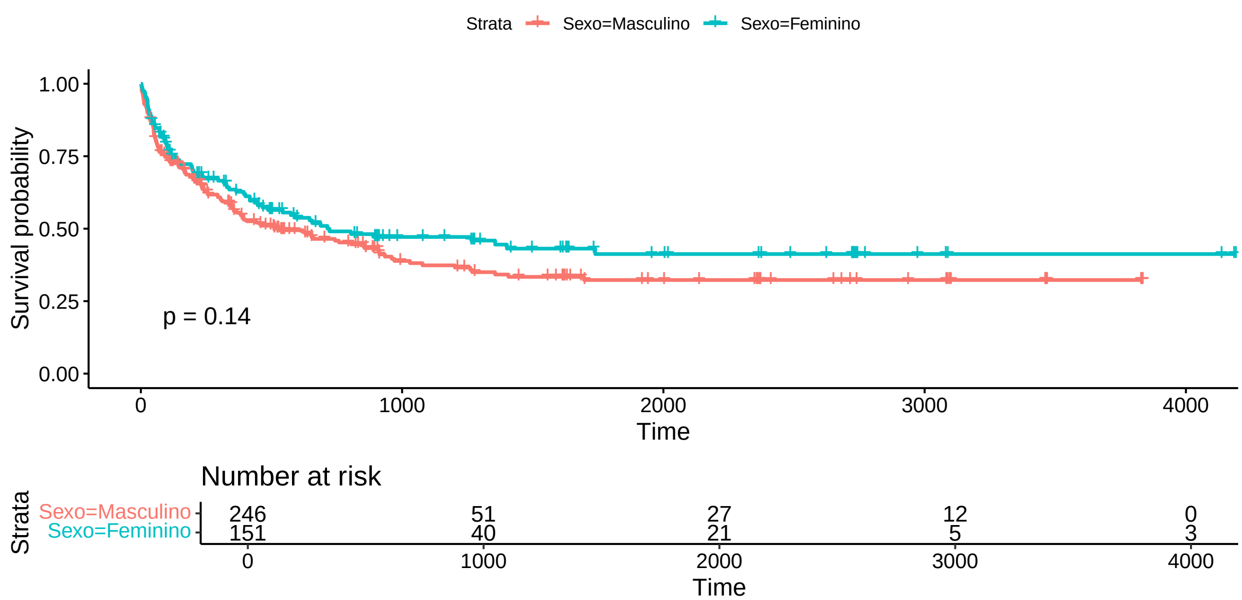
```
ekm
```

```
## Call: survfit(formula = Surv(tempo, cens.1) ~ 1, data = encefalo)
##
##         n events median 0.95LCL 0.95UCL
## [1,] 397    216    653     443     911
```

Vamos agora comparar as curvas de sobrevivência por meio do teste logrank. As covariáveis idade e tratamento foram dicotomizadas para aplicação do teste.

- Sexo

```
levels(encefalo$Sexo) <- c("Masculino", "Feminino")
ekm1 <- survfit(Surv(tempo, cens.1) ~ Sexo, data = encefalo)
ggsurvplot(ekm1, risk.table = TRUE, pval = TRUE)
```



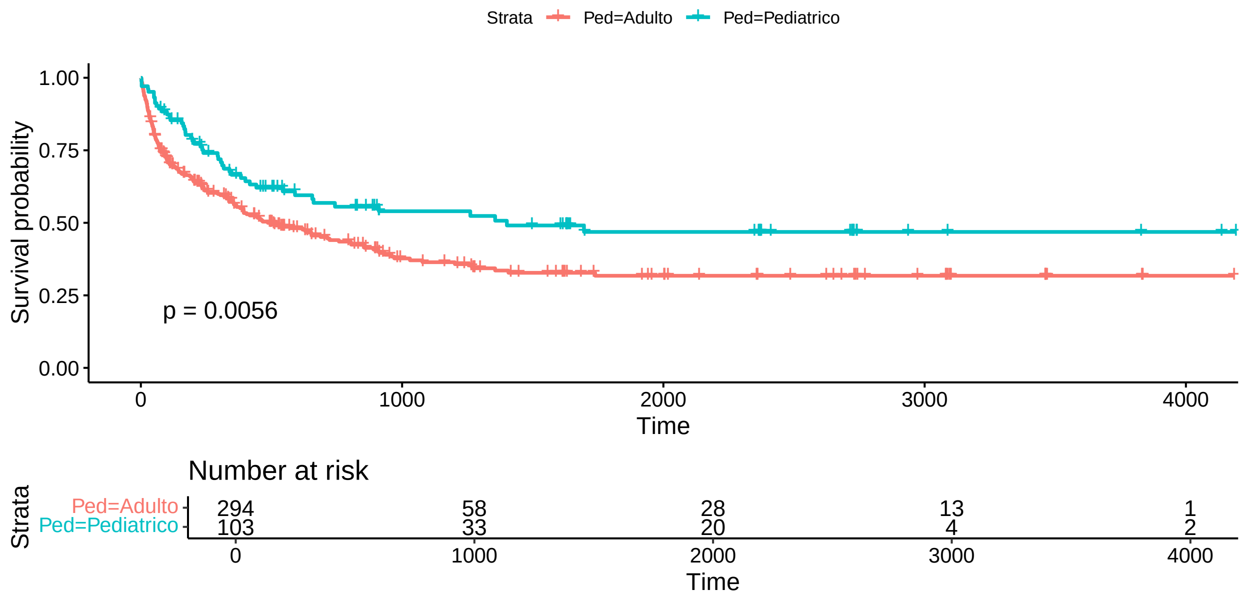

```
survdiff(Surv(tempo, cens.1) ~ Sexo, rho = 0, data = encefalo)
```

```
## Call:
## survdiff(formula = Surv(tempo, cens.1) ~ Sexo, data = encefalo,
##     rho = 0)
##
##                  N Observed Expected (O-E)^2/E (O-E)^2/V
## Sexo=Masculino 246      140    129.4     0.873      2.18
## Sexo=Feminino  151       76     86.6     1.303      2.18
##
##  Chisq= 2.2  on 1 degrees of freedom, p= 0.1
```

O teste não aponta diferença entre os sexos.

- Idade

```
ekm2 <- survfit(Surv(tempo, cens.1) ~ Ped, data = encefalo)
ggsurvplot(ekm2, risk.table = TRUE, pval = TRUE)
```



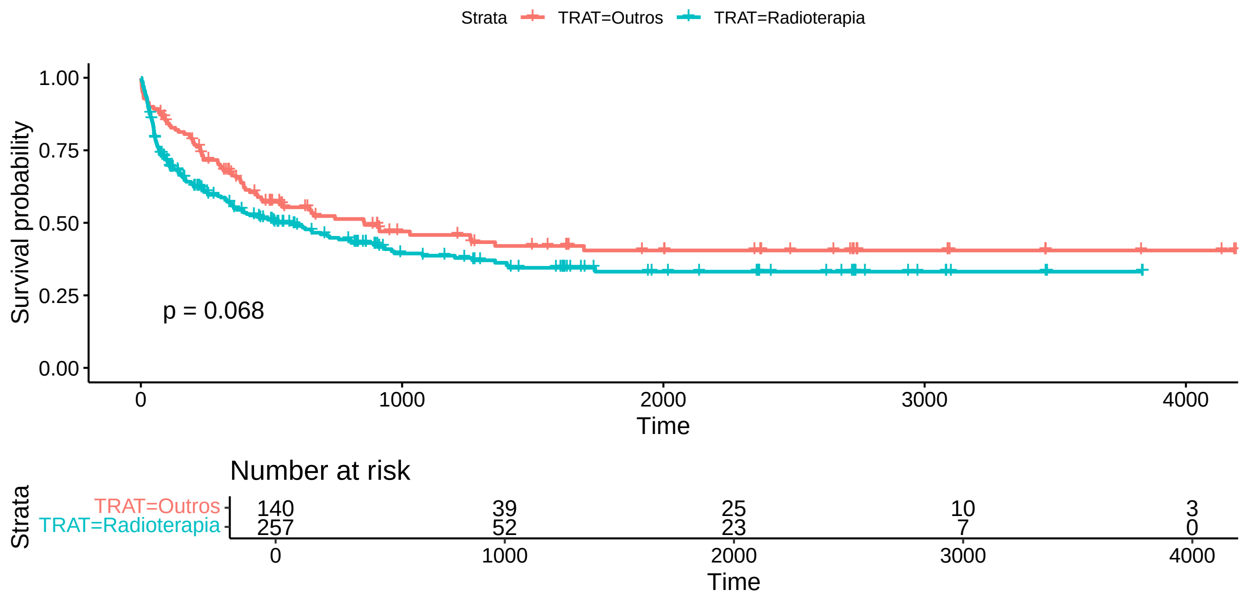

```r
survdiff(Surv(tempo, cens.1) ~ Ped, rho = 0, data = encefalo)
```

```
## Call:
## survdiff(formula = Surv(tempo, cens.1) ~ Ped, data = encefalo,
##     rho = 0)
##
##                  N Observed Expected (O-E)^2/E (O-E)^2/V
## Ped=Adulto     294      169    150.3      2.32      7.68
## Ped=Pediatrico 103       47     65.7      5.31      7.68
##
##  Chisq= 7.7  on 1 degrees of freedom, p= 0.006
```

Uma diferença estatisticamente significativa foi encontrada entre os grupos de idade. Os pacientes mais velhos apresentaram sobrevida pior.

- Tratamento

```r
ekm3 <- survfit(Surv(tempo, cens.1) ~ TRAT, data = encefalo)
ggsurvplot(ekm3, risk.table = TRUE, pval = TRUE)
```




```r
survdiff(Surv(tempo, cens.1) ~ TRAT, rho = 0, data = encefalo)
```

```
## Call:
## survdiff(formula = Surv(tempo, cens.1) ~ TRAT, data = encefalo,
##     rho = 0)
##
##                      N Observed Expected (O-E)^2/E (O-E)^2/V
## TRAT=Outros        140       72     85.1      2.01      3.34
## TRAT=Radioterapia 257      144    130.9      1.31      3.34
##
##  Chisq= 3.3  on 1 degrees of freedom, p= 0.07
```

Com relação ao grupo tratamento dicotomizado, uma diferença marginal foi obtida. Os pacientes do grupo radioterapia apresentaram sobrevida estimada menor.

Ajuste dos modelos de regressão

Vamos avaliar o ajuste dos modelos paramétricos exponencial, Weibull e lognormal.

```r
m0e <- survreg(Surv(tempo, cens.1) ~ Idade + Sexo + TRAT, dis = "exponential",
               data = encefalo)
summary(m0e)
```

```
##
## Call:
## survreg(formula = Surv(tempo, cens.1) ~ Idade + Sexo + TRAT,
##     data = encefalo, dist = "exponential")
##                      Value Std. Error     z       p
## (Intercept)        7.87033    0.25911 30.37 < 2e-16
## Idade             -0.02664    0.00352 -7.56 4.1e-14
## Sexo               0.21937    0.14319  1.53    0.13
## TRATRadioterapia -0.14262    0.14905 -0.96    0.34
##
## Scale fixed at 1
##
```

```
## Exponential distribution
## Loglik(model)= -1741   Loglik(intercept only)= -1775.3
##  Chisq= 68.49 on 3 degrees of freedom, p= 9e-15
## Number of Newton-Raphson Iterations: 5
## n= 397
```

```
#
m0w <- survreg(Surv(tempo, cens.1) ~ Idade + Sexo + TRAT, dis = "weibull",
               data = encefalo)
summary(m0w)
```

```
##
## Call:
## survreg(formula = Surv(tempo, cens.1) ~ Idade + Sexo + TRAT,
##     data = encefalo, dist = "weibull")
##                      Value Std. Error     z       p
## (Intercept)        8.15448    0.47224 17.27 < 2e-16
## Idade             -0.03466    0.00638 -5.44 5.5e-08
## Sexo               0.34607    0.26159  1.32    0.19
## TRATRadioterapia -0.13564    0.27367 -0.50    0.62
## Log(scale)         0.60088    0.05714 10.52 < 2e-16
##
## Scale= 1.82
##
## Weibull distribution
## Loglik(model)= -1666.8   Loglik(intercept only)= -1684.7
##  Chisq= 35.73 on 3 degrees of freedom, p= 8.5e-08
## Number of Newton-Raphson Iterations: 5
## n= 397
```

```
#
m0l <- survreg(Surv(tempo, cens.1) ~ Idade + Sexo + TRAT, dis = "lognormal",
               data = encefalo)
summary(m0l)
```

```
##
## Call:
## survreg(formula = Surv(tempo, cens.1) ~ Idade + Sexo + TRAT,
##     data = encefalo, dist = "lognormal")
##                      Value Std. Error     z       p
## (Intercept)        7.12391    0.48611 14.65 < 2e-16
## Idade             -0.03133    0.00656 -4.78 1.8e-06
## Sexo               0.42578    0.27695  1.54    0.12
## TRATRadioterapia -0.12610    0.29305 -0.43    0.67
## Log(scale)         0.88086    0.05207 16.92 < 2e-16
##
## Scale= 2.41
##
## Log Normal distribution
## Loglik(model)= -1654   Loglik(intercept only)= -1668.2
##  Chisq= 28.33 on 3 degrees of freedom, p= 3.1e-06
## Number of Newton-Raphson Iterations: 3
## n= 397
```

Como visto, apenas a variável Idade foi significativa. Vamos considerar então modelos apenas com este preditor.

```r
m1e <- survreg(Surv(tempo,cens.1)~Idade,dis="exponential",data=encefalo)
m1w <- survreg(Surv(tempo,cens.1)~Idade,dis="weibull",data=encefalo)
m1l <- survreg(Surv(tempo,cens.1)~Idade,dis="lognormal",data=encefalo)
```

A seguir a comparação dos modelos via AIC:

```r
extractAIC(m1e)[2]
```

```
## [1] 3489.708
```

```r
extractAIC(m1w)[2]
```

```
## [1] 3341.762
```

```r
extractAIC(m1l)[2]
```

```
## [1] 3316.697
```

O modelo lognormal resulta no menor valor de AIC. Para o modelo lognormal, temos as estimativas:

```r
summary(m1l)
```

```
##
## Call:
## survreg(formula = Surv(tempo, cens.1) ~ Idade, data = encefalo,
##     dist = "lognormal")
##               Value Std. Error    z       p
## (Intercept)  7.6594     0.2715 28.2 < 2e-16
## Idade       -0.0321     0.0063 -5.1 3.4e-07
## Log(scale)   0.8853     0.0521 17.0 < 2e-16
##
## Scale= 2.42
##
## Log Normal distribution
## Loglik(model)= -1655.3   Loglik(intercept only)= -1668.2
##  Chisq= 25.66 on 1 degrees of freedom, p= 4.1e-07
## Number of Newton-Raphson Iterations: 3
## n= 397
```

A razão de tempos medianos entre dois indivíduos com diferença de um ano (pacientes com 26 e 25 anos de idade, por exemplo) é dada por $e^{-0.0321} = 0,968$. Isso significa que o tempo mediano de vida vai diminuindo com a idade: pacientes mais jovens apresentam sobrevida superior àquela de pacientes mais velhos.

Exemplo 5: Modelo de Cox

O modelo de Cox, que assume que os riscos são proporcionais, também foi ajustado aos dados desse estudo.

```r
cox0 <- coxph(Surv(tempo, cens.1) ~ Idade + Sexo + TRAT, data = encefalo)
round(summary(cox0)$coef, 4)
```

```
##                     coef exp(coef) se(coef)       z Pr(>|z|)
## Idade             0.0169    1.0170   0.0035  4.8491   0.0000
## Sexo             -0.1860    0.8303   0.1433 -1.2977   0.1944
## TRATRadioterapia  0.0473    1.0484   0.1505  0.3141   0.7534
```

Novamente, apenas a variável idade foi significativa. O modelo final fica então dado por:

```r
cox1 <- coxph(Surv(tempo, cens.1) ~ Idade, data = encefalo)
summary(cox1)
```

```
## Call:
## coxph(formula = Surv(tempo, cens.1) ~ Idade, data = encefalo)
##
##   n= 397, number of events= 216
##
##           coef exp(coef) se(coef)     z Pr(>|z|)
## Idade 0.017268  1.017418 0.003354 5.148 2.63e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##       exp(coef) exp(-coef) lower .95 upper .95
## Idade     1.017     0.9829     1.011     1.024
##
## Concordance= 0.606  (se = 0.02 )
## Likelihood ratio test= 26.72  on 1 df,   p=2e-07
## Wald test            = 26.5  on 1 df,   p=3e-07
## Score (logrank) test = 27.08  on 1 df,   p=2e-07
```

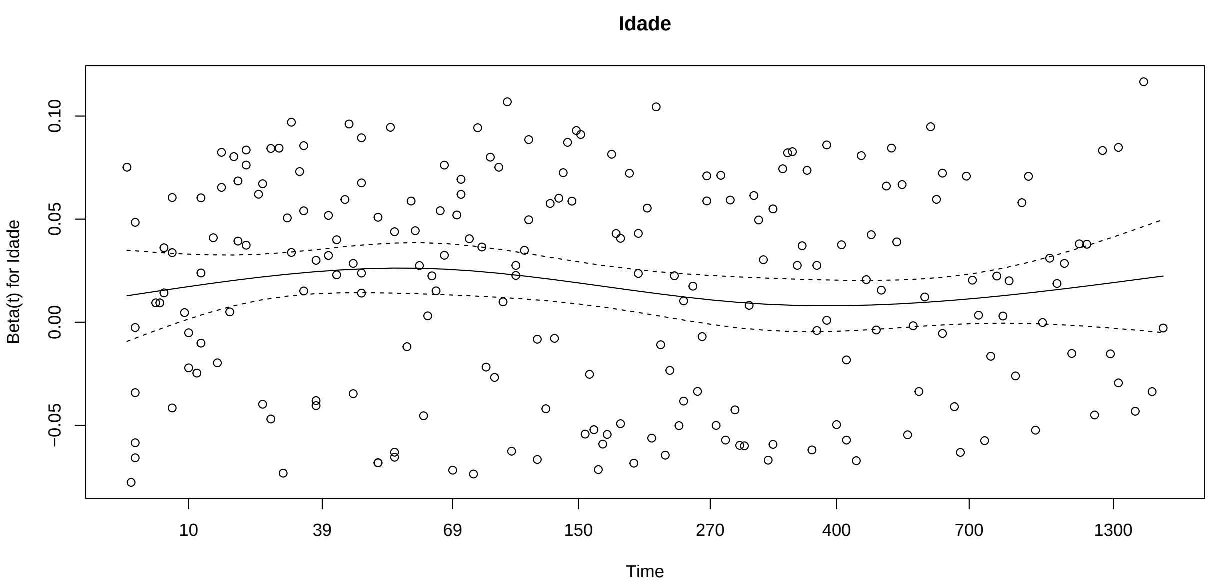
A verificação de proporcionalidade dos riscos pode ser testada via:

```r
zph <- cox.zph(cox1)
zph
```

```
##        chisq df    p
## Idade   1.02  1 0.31
## GLOBAL  1.02  1 0.31
```

O teste não rejeita a hipótese nula de riscos proporcionais. Tal conclusão também é obtida pelo exame dos resíduos de Schoenfeld.

```r
plot(zph, main = "Idade")
```



No modelo de Cox, a exponencial do parâmetro estimado é interpretada como razão de riscos:

- O aumento em 1 ano de idade aumenta o risco em 1,7% de o paciente ir a óbito ($e^{0,0173}) = 1,0174$).
- O aumento de 10 anos na idade (indivíduos com 40 e 30 anos, por exemplo) aumenta o risco de ir a óbito em 18,8% ($e^{10\times 0,0173} = 1,1888$).