

# **ANÁLISE DE SOBREVIVÊNCIA**

## **Teoria e aplicações em saúde**

### **Caderno de Respostas**

---

#### **Capítulo 5**

#### **Modelo paramétrico de sobrevivência**

**Exercício 5.1:** Em um estudo, ajustou-se um modelo exponencial aos tempos de sobrevivência observados (em meses) nos grupos controle e tratamento. Os modelos encontrados foram:

$$S_c(t) = \exp(-0,07t) \text{ para o grupo controle}$$

$$S_{tr}(t) = \exp(-0,04t) \text{ para o grupo tratamento}$$

Com base nesses modelos, responda:

- a. Qual foi o risco instantâneo estimado para o grupo controle? E para o grupo recebendo tratamento?

**Resposta:**

Segundo os resultados acima a estimativa do parâmetro da distribuição exponencial para o grupo controle, digamos  $\lambda_c$ , foi igual a 0,07, e para o grupo tratamento ( $\lambda_{tr}$ ) foi igual a 0,04. Isto é, o risco estimado em qualquer tempo sob a distribuição exponencial para o grupo controle é 0,07 enquanto que para o grupo tratamento é 0,04.

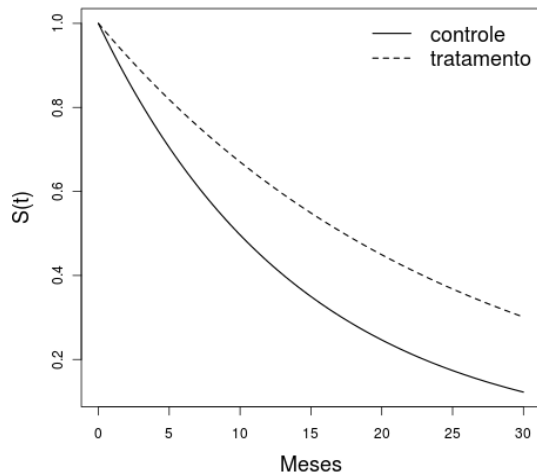
- b. Qual foi a sobrevivência média e mediana no grupo controle? E no grupo recebendo tratamento?

**Resposta:**

O tempo médio de sobrevivência é dado por  $T_{\text{médio}} = 1/\alpha$ . Logo, para o grupo controle temos  $T_{\text{médio}_c} = 1/0,07 = 14,3$  meses e para o grupo tratamento temos  $T_{\text{médio}_{tr}} = 1/0,04 = 25,0$  meses.

Já o tempo mediano de sobrevivência é dado por  $T_{\text{mediano}} = \ln(2)/\alpha$ . Sendo assim, temos que o tempo mediano para os grupos controle e tratamento são iguais à, 9,9 e 17,3 meses respectivamente ( $T_{\text{mediano}_c} = \ln(2)/0,07 = 9,9$  e  $T_{\text{mediano}_{tr}} = \ln(2)/0,04 = 17,3$ ).

- c. As duas curvas estimadas de sobrevivência são apresentadas na figura a seguir. Localize, nessa figura, o tempo mediano e médio calculado. Analisando o gráfico, você acha que o tratamento teve efeito na sobrevivência desses pacientes?



**Resposta:**

O gráfico sugere que o tratamento teve um efeito significativo no aumento do tempo de sobrevivência dos pacientes. No entanto para que possamos tirar conclusões estatisticamente conclusivas é importante considerar além da variabilidade nas curvas estimadas, os tempos médios e medianos de sobrevivência estimados e realizar os testes estatísticos adequados.

**Exercício 5.2:** Estude no R o efeito do parâmetro  $\alpha$  no modelo exponencial:

- a. Construa gráficos da função de sobrevivência utilizando  $\alpha= 0,1$ ,  $\alpha= 0,5$  e  $\alpha= 0,7$ .

**Resposta:**

# Exercício 5.2

# Gráficos da função de sobrevivência:  $\alpha= 0,1$ ,  $\alpha= 0,5$  e  $\alpha= 0,7$ .

#  $\alpha= 0,1$

alfa <- 0.1

curve(exp(-alfa \* x), from = 0, to = 25, ylab = "S(t)", xlab = "Tempo")

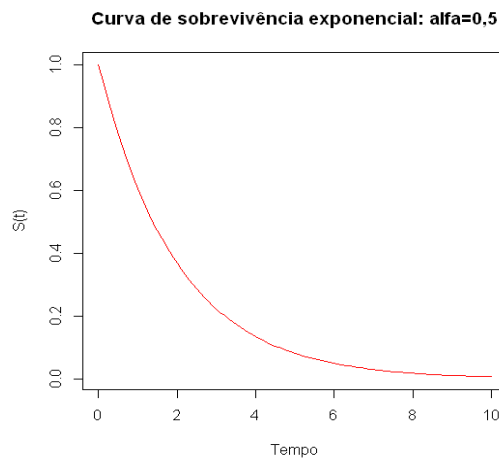
title("Curva de sobrevivência exponencial: alfa=0,1")



#  $\alpha= 0,5$

```
alfa <- 0.5
```

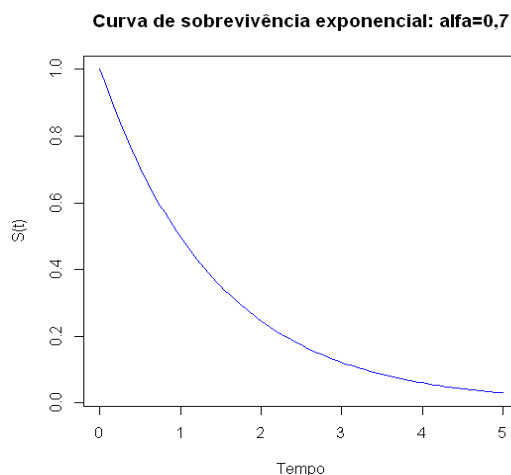
```
curve(exp(-alfa * x), from = 0, to = 10, ylab = "S(t)", xlab = "Tempo", lty = 1, col = "red")  
title("Curva de sobrevivência exponencial: alfa=0,5")
```



```
#  $\alpha = 0,7$ 
```

```
alfa <- 0.7
```

```
curve(exp(-alfa * x), from = 0, to = 5, ylab = "S(t)", xlab = "Tempo", lty = 1, col =  
"blue")  
title("Curva de sobrevivência exponencial: alfa=0,7")
```



- b. Calcule o tempo mediano de sobrevivência para modelos com  $\alpha = 0,1$ ,  $\alpha = 0,5$  e  $\alpha = 0,7$ .

**Resposta:**

```
#  $\alpha = 0,1$ 
```

```
alfa <- 0.1
```

```
# Percentil 50 (Mediana) para  $\alpha = 0,1$ 
```

```
p50 <- log(1/0.5)/alfa
```

```
p50
```

```
[1] 6.931472
```

```
#  $\alpha = 0,5$ 
```

```
alfa <- 0.5
```

```
# Percentil 50 (Mediana) para  $\alpha = 0,5$ 
p50 <- log(1/0.5)/alfa
p50
[1] 1.386294
```

```
#  $\alpha = 0,7$ 
alfa <- 0.7
# Percentil 50 (Mediana) para  $\alpha = 0,7$ 
p50 <- log(1/0.5)/alfa
p50
[1] 0.9902103
```

- c. Calcule os percentis 90% (P90) e 10% (P10), isto é, os tempos em que 90% e 10% dos pacientes ainda não tinham sofrido o evento.

**Resposta:**

```
# Percentil 90 para  $\alpha = 0,1$ 
p90 <- log(1/0.9)/alfa
p90
[1] 1.053605
```

```
# Percentil 10 para  $\alpha = 0,1$ 
p10 <- log(1/0.1)/alfa
p10
[1] 23.02585
```

```
# Percentil 90 para  $\alpha = 0,5$ 
p90 <- log(1/0.9)/alfa
p90
[1] 0.2107210
```

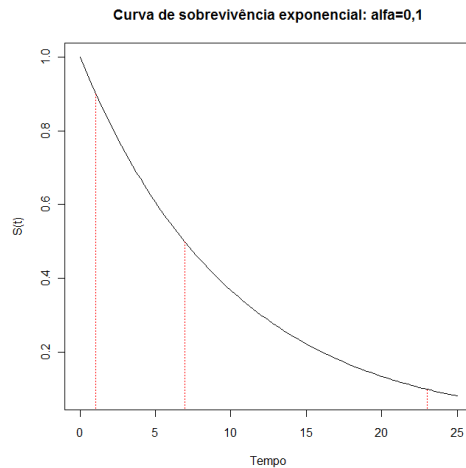
```
# Percentil 10 para  $\alpha = 0,5$ 
p10 <- log(1/0.1)/alfa
p10
[1] 4.60517
```

```
# Percentil 90 para  $\alpha = 0,7$ 
p90 <- log(1/0.9)/alfa
p90
[1] 0.1505150
```

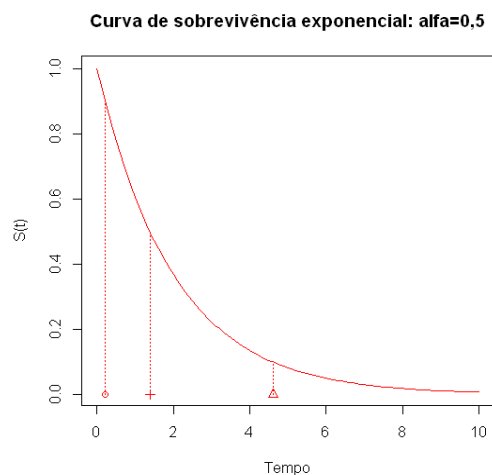
```
# Percentil 10 para  $\alpha = 0,7$ 
p10 <- log(1/0.1)/alfa
p10
[1] 3.289407
```

```
# Refazendo os gráficos com todos os percentis
alfa <- 0.1
curve(exp(-alfa * x), from = 0, to = 25, ylab = "S(t)", xlab = "Tempo")
title("Curva de sobrevivência exponencial: alfa=0,1")
```

```
# Incluindo os pontos na curva
segments(p90, 0, p90, exp(-alfa * p90), lty = 3, col = "red")
segments(p50, 0, p50, exp(-alfa * p50), lty = 3, col = "red")
segments(p10, 0, p10, exp(-alfa * p10), lty = 3, col = "red")
```



```
alfa <- 0.5
curve(exp(-alfa * x), from = 0, to = 10, ylab = "S(t)", xlab = "Tempo", lty = 1, col = "red")
title("Curva de sobrevivência exponencial: alfa=0,5")
# Incluindo os pontos na curva
segments(p90, 0, p90, exp(-alfa * p90), lty = 3, col = "red")
points(p90, 0, pch = 1, col = "red")
segments(p50, 0, p50, exp(-alfa * p50), lty = 3, col = "red")
points(p50, 0, pch = 3, col = "red")
segments(p10, 0, p10, exp(-alfa * p10), lty = 3, col = "red")
points(p10, 0, pch = 2, col = "red")
```

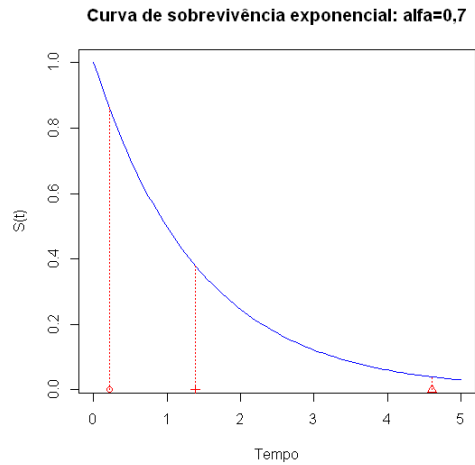


```
alfa <- 0.7
curve(exp(-alfa * x), from = 0, to = 5, ylab = "S(t)", xlab = "Tempo", lty = 1, col = "blue")
title("Curva de sobrevivência exponencial: alfa=0,7")
# Incluindo os pontos na curva
segments(p90, 0, p90, exp(-alfa * p90), lty = 3, col = "red")
points(p90, 0, pch = 1, col = "red")
```

```

segments(p50, 0, p50, exp(-alfa * p50), lty = 3, col = "red")
points(p50, 0, pch = 3, col = "red")
segments(p10, 0, p10, exp(-alfa * p10), lty = 3, col = "red")
points(p10, 0, pch = 2, col = "red")

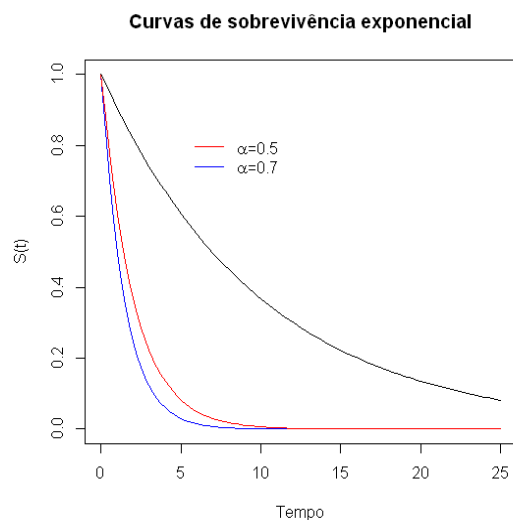
```



```

# Gráfico único com todos os valores de alfa
alfa <- 0.7
curve(exp(-alfa * x), from = 0, to = 25, lty = 1, col = "blue", ylab = "S(t)", xlab =
"Tempo")
alfa <- 0.5
curve(exp(-alfa * x), from = 0, to = 25, add=T, lty = 1, col = "red")
alfa <- 0.1
curve(exp(-alfa * x), from = 0, to = 25, add=T)
title("Curvas de sobrevivência exponencial")
legend(5, 0.85, c(expression(paste(alpha, "=0.5")), expression(paste(alpha, "=0.7"))),
lty = 1, col = c("red", "blue"), bty = "n")

```



- d. O que você conclui sobre o parâmetro  $\alpha$  e seu efeito no comportamento do no modelo exponencial de sobrevivência?

**Resposta:**

Nota-se que a função de sobrevivência cai mais rapidamente à medida que aumenta o valor do parâmetro  $\alpha$ , e, portanto também decrescem os percentis 10% e 90%. Este comportamento é esperado já que o risco instantâneo de falha em qualquer tempo, sob o modelo exponencial, aumenta com o aumento de  $\alpha$ .

**Exercício 5.3:** Com relação ao modelo paramétrico Weibull, responda:

- a. Por que o modelo Weibull é considerado mais flexível do que o modelo exponencial?

**Resposta:**

Porque possui um parâmetro adicional que permite ajustar diferentes formas para a função risco, daí o nome parâmetro de forma, representado por  $\gamma$ .

- b. Em que situação particular o modelo Weibull é equivalente ao exponencial?

**Resposta:**

Na situação em que o parâmetro de forma  $\gamma = 1$ .

- c. Qual a relação entre o parâmetro de forma  $\gamma$  e o comportamento da função de risco?

**Resposta:**

Quando  $\gamma = 1$  a função de risco é constante, ou seja, o risco instantâneo de ocorrência do evento não varia com o passar do tempo; quando  $\gamma > 1$  o risco cresce no tempo; e  $\gamma < 1$  o risco decresce no tempo.

- d. Quais das curvas de risco apresentadas na Figura 3.3 não poderiam ser modeladas pela função Weibull, nem mesmo aproximadamente?

**Resposta:**

As curvas D, E e F não poderiam ser modeladas pela função Weibull pois o comportamento da função risco ao longo de tempo deve ser monotônico: somente crescente ou somente decrescente. O que se vê nos quadros D, E e F são misturas destes comportamentos.

**Exercício 5.4:** Seja  $T$  o tempo de sobrevivência até a ocorrência de um evento, que segue uma distribuição Weibull com parâmetros  $\gamma = 1,5$  e  $\alpha = 0,13$ .

- a. Escreva as funções de sobrevivência -  $S(t)$ , de risco -  $\lambda(t)$  e de risco acumulado  $\Lambda(t)$

**Resposta:**



$$S(t) = \exp(-(\alpha t)^{\gamma}) = \exp(-(0,13t)^{1,5})$$

$$\lambda(t) = \gamma \alpha^{\gamma} t^{\gamma-1} = 1,5 \times 0,13^{1,5} t^{(1,5-1)} = 1,5 \times 0,046872 t^{0,5} = 0,070308 t^{0,5}$$

$$\Lambda(t) = -\ln S(t) = -\ln(\exp(-(\alpha t)^{\gamma})) = (\alpha t)^{\gamma-1} = (0,13t)^{1,5-1} = (0,13t)^{0,5}$$

- b. Use o R para construir o gráfico das funções de sobrevivência -  $S(t)$ , de risco -  $\lambda(t)$  e de risco acumulado  $\Lambda(t)$ .

**Resposta:**

# Exercício 5.4

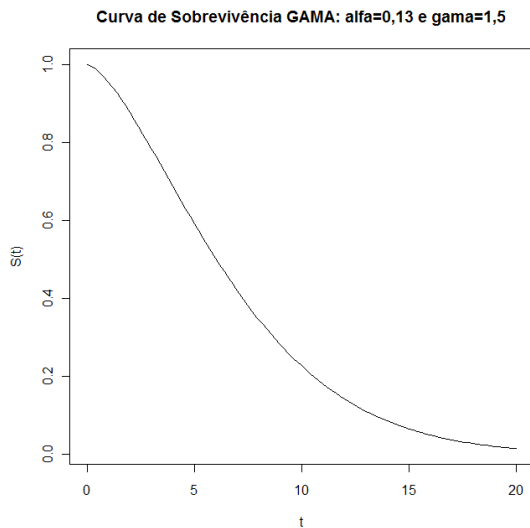
# item b

alfa <- 0.13

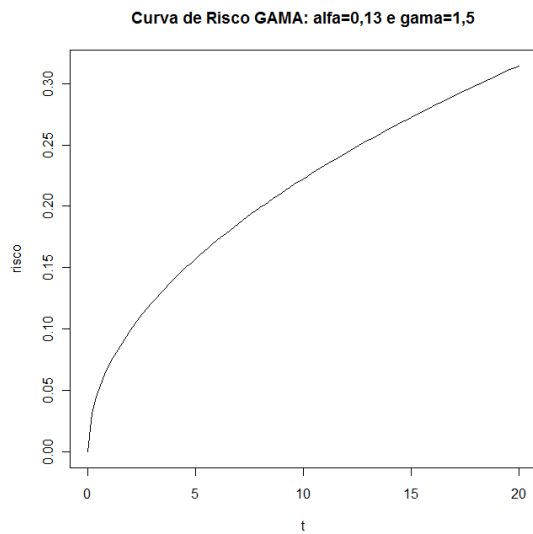
gama <- 1.5

curve(exp(-(alfa \* x)^gama), from = 0, to = 20, ylab = "S(t)", xlab = "t")

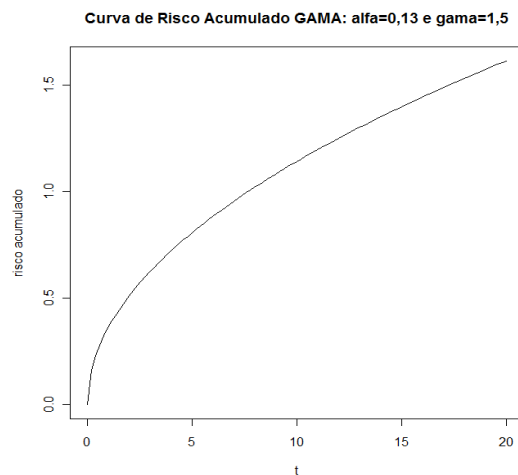
title("Curva de Sobrevivência GAMA: alfa=0,13 e gama=1,5")



curve(alfa \* gama \* (alfa \* x)^(gama - 1), from = 0, to = 20, ylab = "risco", xlab = "t")  
title("Curva de Risco GAMA: alfa=0,13 e gama=1,5")



```
curve((alfa * x)^(gama - 1), from = 0, to = 20, ylab = "risco acumulado", xlab = "t")
title("Curva de Risco Acumulado GAMA: alfa=0,13 e gama=1,5")
```



- c. Calcule o tempo mediano de sobrevivência, e os percentis 80 e 10 dessa distribuição.

**Resposta:**

# item c

# Tempo mediano ( $S(t) = 0.5$ )

```
tmediano <- log(1/0.5)^(1/gama)/alfa
```

```
tmediano
```

```
[1] 6.024767
```

# Percentil 80 ( $S(t) = 0.80$ )

```
p80 <- log(1/0.8)^(1/gama)/alfa
```

```
p80
```

```
[1] 2.829955
```

# Percentil 10 ( $S(t) = 0.10$ )

```
p10 <- log(1/0.1)^(1/gama)/alfa
```

```
p10
```

[1] 13.41324

- d. Fixe o valor do parâmetro  $\alpha = 0,13$  e faça gráficos da função de risco e da função de sobrevivência para diversos valores do parâmetro de forma  $\gamma$ :

- i.  $0 < \gamma < 1$
- ii.  $\gamma = 1$
- iii.  $\gamma > 1$

**Resposta:**

```
# item d. Fixando  $\alpha = 0,13$ 
```

```
# sub-item (i)  $0 < \gamma < 1$ 
```

```
par(mfrow = c(3, 2))
```

```
alfa <- 0.13
```

```
gama <- 0.5
```

```
curve(exp(-(alfa * x)^gama), from = 0, to = 20, ylab = "S(t)", xlab = "t", main =  
expression(paste(gamma, "=0.5")))
```

```
curve(alfa * gama * (alfa * x)^(gama - 1), from = 0, to = 20, ylab = "risco", xlab = "t",  
main = expression(paste(gamma, "=0.5")))
```

```
# sub-item (ii)  $\gamma = 1$ 
```

```
gama <- 1
```

```
curve(exp(-(alfa * x)^gama), from = 0, to = 20, ylab = "S(t)", xlab = "t", main =  
expression(paste(gamma, "=1")))
```

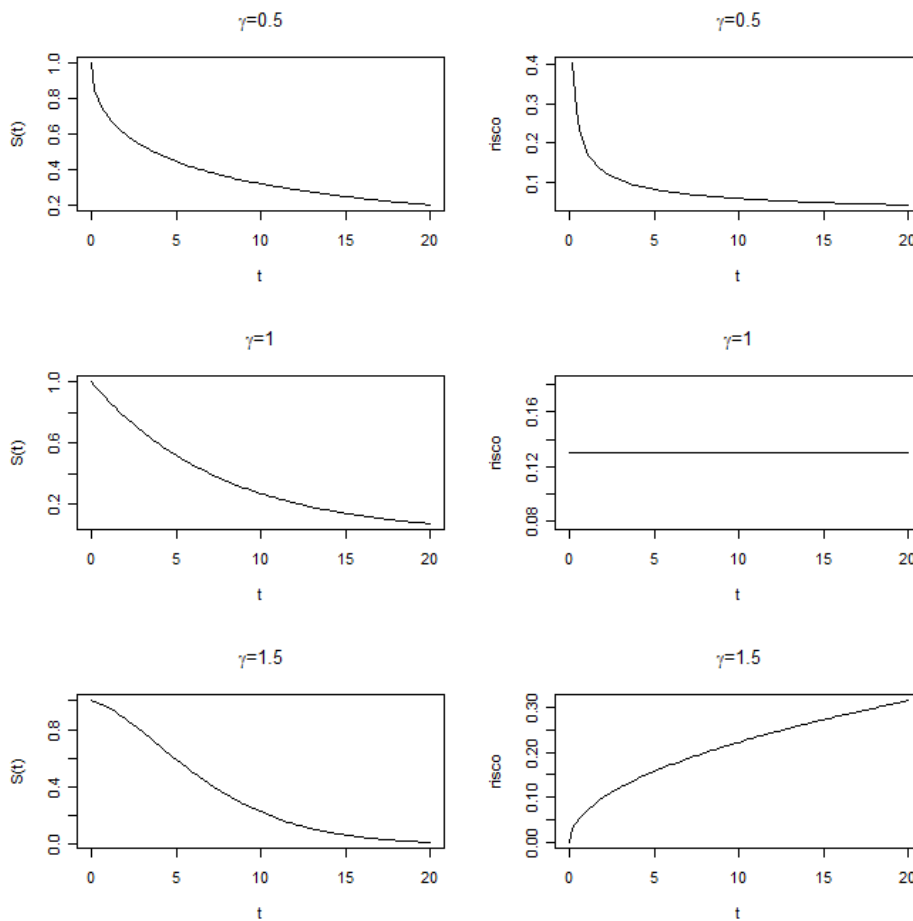
```
curve(alfa * gama * (alfa * x)^(gama - 1), from = 0, to = 20, ylab = "risco", xlab = "t",  
main = expression(paste(gamma, "=1")))
```

```
# sub-item (iii)  $\gamma > 1$ 
```

```
gama <- 1.5
```

```
curve(exp(-(alfa * x)^gama), from = 0, to = 20, ylab = "S(t)", xlab = "t", main =  
expression(paste(gamma, "=1.5")))
```

```
curve(alfa * gama * (alfa * x)^(gama - 1), from = 0, to = 20, ylab = "risco", xlab = "t",  
main = expression(paste(gamma, "=1.5")))
```



- e. Interprete como o parâmetro de forma  $\gamma$  afeta o comportamento do risco e da sobrevivência.

**Resposta:**

Para  $0 < \gamma < 1$  (por exemplo,  $\gamma=0,5$ ) a sobrevivência cai rapidamente e o risco é decrescente.

Para  $\gamma=1$  a sobrevivência cai mais suavemente e o risco é constante.

Para  $\gamma > 1$  (por exemplo,  $\gamma=1,5$ ) a sobrevivência cai suavemente no início do período e o risco é crescente.

**Exercício 5.5:** Em um estudo sobre o tempo de incubação de uma infecção, verificou-se que  $T$  é adequadamente descrito por uma função Weibull com parâmetros  $\gamma = 1,2$  e  $\alpha = 0,07$ .

- a. Calcule o tempo mediano de incubação desta infecção.

**Resposta:**

# Exercício 5.5

# item a

alfa <- 0.07

gama <- 1.2

```
tmediano <- log(1/0.5)^(1/gama)/alfa
tmediano
[1] 10.52583
```

O tempo mediano de incubação é de aproximadamente 10 horas e meia, em outras palavras, segundo este modelo espera-se que 50% das pessoas infectadas comecem a apresentar sintomas depois de 10 horas e meia do contato com o agente infeccioso.

b. É correto dizer que em 10 horas após a infecção, espera-se que 80% das pessoas já tenham desenvolvido sintomas?

**Resposta:**

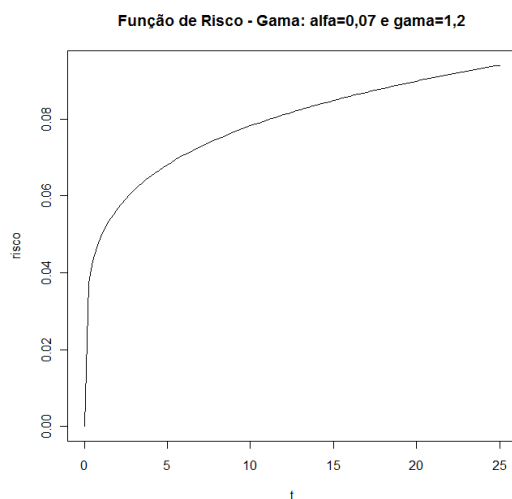
```
# item b
t <- 10
S10 <- exp(-(alfa * t)^gama)
S10
[1] 0.5211044
```

Não, em 10 horas, espera-se que aproximadamente 52% das pessoas não tenham desenvolvido sintomas, ou alternativamente, espera-se que 48% tenham desenvolvido sintomas.

c. O risco de surgimento de sintomas é crescente ou decrescente ao longo do tempo?

**Resposta:**

```
# item c
par(mfrow = c(1, 1))
curve(gama * alfa^gama * x^(gama - 1), from = 0, to = 25, ylab = "risco", xlab = "t")
title("Função de Risco")
```



Na verdade não é nem preciso traçar o gráfico da função risco; basta observar que o parâmetro de forma  $\gamma$  é, neste caso, maior do que 1, ou seja, risco crescente.

**Exercício 5.6:** Mil crianças não vacinadas são acompanhadas, a partir do nascimento, em um estudo cujo objetivo é identificar a idade em que adquirem hepatite A. Os resultados do

estudo indicam que a idade média de soroconversão das crianças foi de 4,5 anos e que o risco de contrair hepatite A foi constante e independente da idade.

- a. Proponha um modelo paramétrico para o tempo até a aquisição de hepatite A.

**Resposta:**

Neste caso, como o risco de contrair hepatite A é constante no tempo, um modelo simples (isto é, parcimonioso) e adequado (pois possui função risco constante) seria o modelo paramétrico exponencial, com alfa assumindo a idade média (4,5 anos) .

- b. Faça no R o gráfico da função de sobrevivência, de acordo com esse modelo.

**Resposta:**

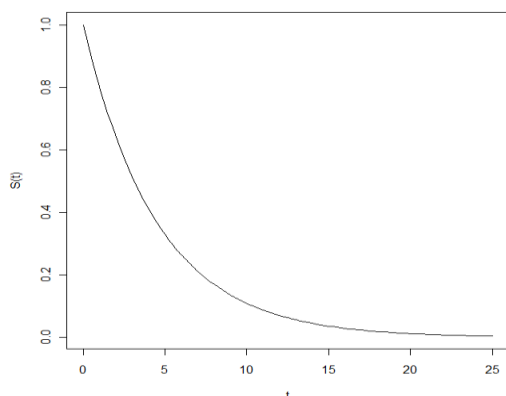
```
# Exercício 5.6
```

```
# item b
```

```
tm <- 4.5
```

```
alfa <- 1/tm
```

```
curve(exp(-alfa * x), from = 0, to = 25, ylab = "S(t)", xlab = "t")
```



- c. Com base nesse modelo, em que idade espera-se ter 90% das crianças soropositivas?

**Resposta:**

```
# item c
```

```
p10 <- log(1/0.1)/alfa
```

```
p10
```

```
[1] 23.02585
```

Segundo este modelo espera-se que aos 10 anos e 4 meses 90% das crianças sejam soropositivas, ou alternativamente que nesta idade apenas 10% ainda não sejam soropositivas.

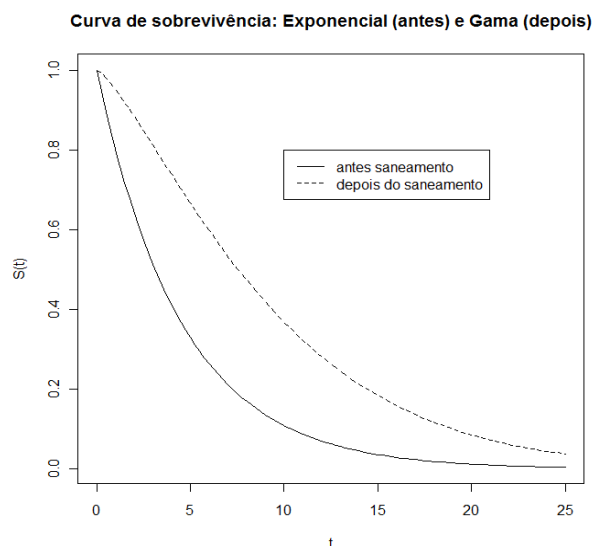
- d. Após este estudo, um projeto de saneamento é implementado nesta comunidade. Para avaliar o efeito do saneamento na transmissão de hepatite A, uma nova coorte é montada, semelhante à anterior. Ao analisar os dados dessa nova coorte, encontramos que um modelo Weibull com parâmetros  $\gamma = 1,3$  e  $\alpha = 0,1$  descreve bem

a curva de sobrevivência. Com base nessa informação, avalie qual foi o efeito do saneamento no risco de contrair hepatite A nessa comunidade. Sugestão: compare os gráficos das funções de sobrevivência.

**Resposta:**

```
# item d
par(mfrow = c(1, 2))
tm <- 4.5
alfa <- 1/tm
curve(exp(-alfa * x), from = 0, to = 25, ylab = "S(t)", xlab = "t")
```

```
alfa <- 0.1
gama <- 1.3
curve(exp(-(alfa * x)^gama), from = 0, to = 25, add = T, lty = 2)
title ("Curva de sobrevivência: Exponencial (antes) e Gama (depois)")
legend(10, 0.8, c("antes saneamento", "depois do saneamento"), lty = 1:2)
```



```
# Percentil 10
p10 <- log(1/0.1)^(1/gama)/alfa
p10
[1] 18.99448
```

Note que a sobrevivência aumentou consideravelmente após implantação do projeto de saneamento. Por exemplo, segundo o modelo pós-saneamento espera-se que somente aos 19 anos 10% não sejam soropositivas.

**Exercício 5.7:** O banco de dados leite2.txt contém dados do tempo de aleitamento de crianças de quatro comunidades.

- α. Ajuste uma distribuição Weibull ao tempo de aleitamento. Existe evidência de que o modelo Weibull seja mais adequado do que o exponencial?

**Resposta:**

# Exercício 5.7

```

# item a
library(survival)
leite <- read.table("leite.txt", header = T, sep = "")
# Modelo Weibull
modeloweib <- survreg(Surv(tempo, status) ~ 1, data = leite, dist = "weib")
summary(modeloweib)

Call:
survreg(formula = Surv(tempo, status) ~ 1, data = leite, dist = "weib")
Value Std. Error z p
(Intercept) 1.713 0.180 9.54 1.38e-21
Log(scale) -0.415 0.209 -1.99 4.70e-02

Scale= 0.66
Weibull distribution
Loglik(model)= -37.5 Loglik(intercept only)= -37.5
Number of Newton-Raphson Iterations: 6
n= 15

```

Considerando o parâmetro de escala (Scale = 0.66), e que  $\gamma = 1/\text{Scale}$ , então  $\gamma = 1.515$ , ligeiramente maior do que um, ou seja, risco crescente. O parâmetro de escala é marginalmente significativo:  $p=0,047$ . Usando o modelo exponencial assumiríamos que o risco é constante. Ajustando então o modelo exponencial:

```

# Modelo Exponencial
modeloexp <- survreg(Surv(tempo, status) ~ 1, data = leite, dist = "exp")
summary(modeloexp)

Call:
survreg(formula = Surv(tempo, status) ~ 1, data = leite, dist = "exp")
Value Std. Error z p
(Intercept) 1.61 0.258 6.23 4.57e-10
Scale fixed at 1
Exponential distribution
Loglik(model)= -39.1 Loglik(intercept only)= -39.1
Number of Newton-Raphson Iterations: 4
n= 15

```

Comparando os dois modelos, que diferem por um grau de liberdade, baseado nas razões de verossimilhança dos dois modelos (loglik=-39.1 para exponencial e loglik=-37.5 para Weibull), conclui-se que o modelo Weibull é de fato melhor.

β. Qual o tempo mediano de amamentação estimado por esse modelo?

**Resposta:**

```

# Os parâmetros da distribuição Weibull no R são  $\alpha = \exp(-\text{intercept})$  e  $\gamma = 1/\text{Scale}$ .
# item b
alfa <- as.vector(exp(-modeloweib$coef[1]))
alfa
[1] 0.1802502
gama <- 1/modeloweib$scale

```



```

gama
[1] 1.514787
tmediano <- log(1/0.5)^(1/gama)/alfa
tmediano
[1] 4.355558

```

O tempo mediano de amamentação estimado por este modelo é de 4.4 meses.

- χ. Faça um gráfico da curva de sobrevivência ajustada pelo modelo Weibull, junto com o gráfico de Kaplan-Meier. O modelo paramétrico representa bem os dados?

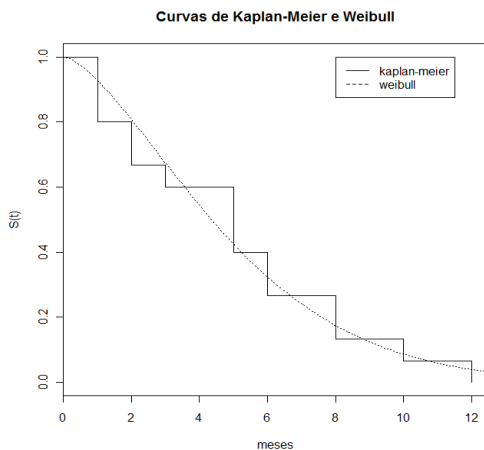
**Resposta:**

```

# item c
km <- survfit(Surv(tempo, status) ~ 1, data = leite)
plot(km, ylab = "S(t)", xlab = "meses", conf.int = F)
alfa <- exp(-1.713)
gama <- 1/0.66
curve(exp(-(alfa * x)^gama), from = 0, to = 15, lty = 3, add = T)
legend(8, 1, c("kaplan-meier", "weibull"), lty = c(1:2))
title ("Curvas de Kaplan-Meier e Weibull")
# Adicionando a linha da mediana
abline(v=tmediano,h=0.5,col=2,lty=3)

```

O modelo de Weibull parece se ajustar muito bem aos dados.



**Exercício 5.8:** O banco de dados leite2.txt contém dados de tempo de aleitamento de crianças de quatro comunidades. No ajuste não-paramétrico a esses dados, observamos que pertencer a uma comunidade não teve efeito no período de aleitamento. Confirme este achado, ajustando um modelo paramétrico a esses dados. Experimente com as distribuições exponencial e o Weibull.

**Resposta:**

```

# Exercício 5.8
library(survival)

```

```

leite2 <- read.table("leite2.txt", header = T, sep = "")
y <- Surv(leite2$tempo, leite2$status)
modeloE1 <- survreg(y ~ factor(grupo), data = leite2, dist = "exponential")
modeloW1 <- survreg(y ~ factor(grupo), data = leite2, dist = "weib")

```

```
summary(modeloE1)
```

```

Call:
survreg(formula = y ~ factor(grupo), data = leite2, dist = "exponential")
Value Std. Error z p
(Intercept) 1.609 0.258 6.233 4.57e-10
factor(grupo)2 0.113 0.365 0.310 7.56e-01
factor(grupo)3 0.410 0.365 1.123 2.62e-01
factor(grupo)4 0.052 0.365 0.142 8.87e-01
Scale fixed at 1

```

```

Exponential distribution
Loglik(model)= -165.2 Loglik(intercept only)= -166
Chisq= 1.58 on 3 degrees of freedom, p= 0.66
Number of Newton-Raphson Iterations: 5
n= 60

```

```
summary(modeloW1)
```

```

Call:
survreg(formula = y ~ factor(grupo), data = leite2, dist = "weib")
Value Std. Error z p
(Intercept) 1.651 0.218 7.568 3.78e-14
factor(grupo)2 0.105 0.306 0.344 7.31e-01
factor(grupo)3 0.426 0.307 1.391 1.64e-01
factor(grupo)4 0.126 0.310 0.408 6.83e-01
Log(scale) -0.175 0.103 -1.703 8.86e-02
Scale= 0.84
Weibull distribution
Loglik(model)= -163.8 Loglik(intercept only)= -164.9
Chisq= 2.19 on 3 degrees of freedom, p= 0.53
Number of Newton-Raphson Iterations: 7
n= 60

```

Note que realmente tanto sob o modelo exponencial quanto o Weibull a variável comunidade (aqui chamada 'grupo') não foi significativa. A estimativa do parâmetro de escala é marginalmente significativo (p-valor=0,089), ou seja, existe alguma indicação de que entre o modelo exponencial e o modelo Weibull o segundo pode ser mais adequado para estes dados.

**Exercício 5.9:** Um estudo foi realizado para estimar o efeito do transplante de medula óssea na sobrevivência de pacientes com leucemia. As covariáveis analisadas foram: idade, fase da doença, ter ou não desenvolvido doença do enxerto crônica e ter ou não desenvolvido doença do enxerto aguda (para mais detalhes acerca desse estudo, refira-se ao Apêndice 12.4). Um modelo exponencial ajustado aos dados, apresenta a seguinte saída do R:

```

              Value      Std. Error    z          p
(Intercept)  7.13536    0.4992    14.293  2.44e-46

```

idade	-0.00179	0.0146	-0.122	9.03e-01
fase interm	-0.79363	0.3651	-2.174	2.97e-02
fase avançada	-1.29759	0.4995	-2.598	9.39e-03
doençacronica	0.92521	0.3335	2.775	5.53e-03
doençaaguda	-1.43654	0.3158	-4.549	5.40e-06

Scale fixed at 1

Exponential distribution

Loglik(model)= -348.3 Loglik(intercept only)= -374.2

Chisq= 51.96 on 5 degrees of freedom, p= 5.5e-10

Number of Newton-Raphson Iterations: 5

Observe a saída do R e responda:

- a. O modelo com covariáveis é melhor do que o modelo nulo (sem covariáveis)?

**Resposta:**

Sim, o modelo com covariáveis é melhor. Note que a log-verossimilhança do modelo com covariáveis é muito maior do que a do modelo nulo e o teste da Razão da Verossimilhança entre o modelo com covariáveis e o modelo nulo resultou um p-valor praticamente nulo ( $p=5.5e-10$ ). Portanto temos evidências altamente significativas contra o modelo nulo.

- b. Que covariáveis estão associadas com maior sobrevivência? Quais estão associadas com redução da sobrevivência?

**Resposta:**

As covariáveis associadas com a redução da sobrevivência são aquelas em que o efeito estimado é negativo: fase intermediária, fase avançada e doença aguda. A covariável idade tem efeito estimado negativo, porém este efeito é não significativo. A única covariável associada a um prognóstico favorável da sobrevivência é doença crônica. Cabe observar que como o próprio nome indica, doença crônica necessariamente tem que evoluir durante um tempo razoável, ou seja, o paciente tem que sobreviver por um tempo razoável para apresentá-la.

- c. Escreva a função de risco,  $\lambda(t)$ , estimada para esta coorte.

**Resposta:**

$\lambda(t) = \exp(- (7.13536 - 0.00179 \times \text{idade} - 0.79363 \times \text{fase interm} - 1.29759 \times \text{fase avançada} + 0.92521 \times \text{doençacronica} - 1.43654 \times \text{doençaaguda}))$

- d. Qual seria o risco de óbito de um paciente de 30 anos, em fase intermediária, com doença crônica?

**Resposta:**

# Exercício 5.9

# item d

```
lambdac <- exp(-(7.13536 - 0.00179 * 30 - 0.79363 + 0.92521))
```

```
lambdac
```

```
[1] 0.0007367662
```

- e. Qual seria o risco de óbito de um paciente de 30 anos, em fase intermediária, com doença aguda?

**Resposta:**

```
# item e
```

```
lambdaa <- exp(-(7.13536 - 0.00179 * 30 - 0.79363 - 1.43654))
```

```
lambdaa
```

```
[1] 0.007816722
```

O risco de óbito é 10,6 vezes maior para o paciente com doença aguda.

- f. Um segundo modelo, mais simples, foi ajustado aos dados, contendo apenas a covariável fase. O logaritmo da função de verossimilhança deste modelo simples foi de -363.6. Compare este modelo com o mais completo acima e indique se o completo resultou em melhor ajuste.

**Resposta:**

l1 e l2 são as log das verossimilhanças dos modelos completo e reduzido, respectivamente.

```
# item f
```

```
l1 <- -348.3
```

```
l2 <- -363.6
```

Calculando a deviance (dev) e os graus de liberdade que é a diferença entre o número de parâmetros do modelo

```
dev <- 2 * (l1 - l2)
```

```
dev
```

```
[1] 30.6
```

E, por último, calcula-se o p-valor da distribuição<sub>2</sub> sob a hipótese nula de que o modelo reduzido é melhor.

```
gl <- 6 - 4
```

```
pvalor <- 1 - pchisq(dev, gl)
```

```
pvalor
```

```
[1] 2.26618e-07
```

Rejeitamos o modelo mais simples com p-valor=0,00000023, ou seja, a redução no valor da verossimilhança dada pelo modelo mais completo foi significativa. Em outras palavras, doença crônica ou aguda é um fator prognóstico importante para o tempo de sobrevivência.

**Exercício 5.10:** A Aids passou a ter tratamento apenas em 1991, desde então a terapia anti-retroviral evoluiu da monoterapia para a terapia combinada (2 ou mais componentes) e, por fim, para a terapia de alta potência (no mínimo 3 componentes, sendo um inibidor de

protease). Espera-se que as terapias mais recentes sejam mais efetivas em aumentar a sobrevivência. Teste esta hipótese, ajustando um modelo exponencial aos dados da coorte de Aids (ipec.csv).

- a. Ajuste um modelo com a variável tratamento apenas. O modelo com a variável tratamento é melhor do que o modelo sem covariáveis? Interprete o efeito dos tratamentos na sobrevivência (lembrando-se que os efeitos dos tratamentos estão sendo estimados em relação à ausência de tratamento).

### Resposta:

```
# Exercício 5.10
```

```
# item a
```

```
library(survival)
```

```
ipec <- read.table("ipec.csv", header = T, sep = ";")
```

```
# A variável tratamento deve ser modelada como um fator, e não como numérica,  
# lembrando que os efeitos dos tratamentos devem ser estimados em relação à  
# ausência de tratamento (tratam=0).
```

```
ipec$tratam <- factor(ipec$tratam, labels =  
c("Nenhum", "Mono", "Combinada", "Potente"))
```

```
# Modelo Exponencial com tratamento
```

```
mod.ipec <- survreg(Surv(tempo, status) ~ factor(tratam), data = ipec, dist = "exp")  
summary(mod.ipec)
```

```
Call:
```

```
survreg(formula = Surv(tempo, status) ~ factor(tratam), data = ipec,  
dist = "exp")
```

	Value	Std. Error	z	p
(Intercept)	6.14	0.177	34.73	2.91e-264
factor(tratam)1	1.59	0.226	7.07	1.58e-12
factor(tratam)2	<b>2.68</b>	0.445	6.01	1.80e-09
factor(tratam)3	3.01	1.016	2.97	3.00e-03

```
Scale fixed at 1
```

```
Exponential distribution
```

```
Loglik(model)= -742.9 Loglik(intercept only)= -774.6
```

```
Chisq= 63.49 on 3 degrees of freedom, p= 1.1e-13
```

```
Number of Newton-Raphson Iterations: 6
```

```
n= 193
```

```
# Para calcular o risco substituir os valores das variáveis DUMMIES No risco
```

```
# do modelo exponencial:  $\lambda(t|x) = \alpha(x) = \exp(x\alpha)$ .
```

```
# Para interpretar lembrar que o R parametriza as distribuições de forma diferente
```

```
# Trocar o sinal dos coeficientes para interpretar de acordo com o texto do capítulo.
```

```
# Calculando o risco de um paciente sem nenhum tratamento:
```

```
trat1 <- 0
```

```
trat2 <- 0
```

```
trat3 <- 0
```

```
lambda0 <- exp(-(mod.ipec$coef[1] + mod.ipec$coef[2] * trat1 + mod.ipec$coef[3] *  
trat2 + mod.ipec$coef[4] * trat3))
```

```
lambda0
```

```
(Intercept)
0.002156625
```

```
# Calculando o risco de um paciente com monoterapia (tratam = 1):
```

```
trat1 <- 1
```

```
trat2 <- 0
```

```
trat3 <- 0
```

```
lambda1 <- exp(-(mod.ipec$coef[1] + mod.ipec$coef[2] * trat1 + mod.ipec$coef[3] *  
trat2 + mod.ipec$coef[4] * trat3))
```

```
lambda1
```

```
(Intercept)
0.0004381632
```

```
# Calculando o risco de um paciente com terapia combinada (tratam = 2):
```

```
trat1 <- 0
```

```
trat2 <- 1
```

```
trat3 <- 0
```

```
lambda2 <- exp(-(mod.ipec$coef[1] + mod.ipec$coef[2] * trat1 + mod.ipec$coef[3] *  
trat2 + mod.ipec$coef[4] * trat3))
```

```
lambda2
```

```
(Intercept)
0.0001485001
```

```
# Calculando o risco de um paciente com terapia potente (tratam = 3):
```

```
trat1 <- 0
```

```
trat2 <- 0
```

```
trat3 <- 1
```

```
lambda3 <- exp(-(mod.ipec$coef[1] + mod.ipec$coef[2] * trat1 + mod.ipec$coef[3] *  
trat2 + mod.ipec$coef[4] * trat3))
```

```
lambda3
```

```
(Intercept)
0.0001058985
```

```
# Calculando os riscos relativos em relação ao paciente sem nenhum tratamento:
```

```
lambda0/lambda1
```

```
(Intercept)
4.921968
```

```
lambda0/lambda2
```

```
(Intercept)
14.52271
```

```
lambda0/lambda3
```

```
(Intercept)
20.36501
```

Rejeitamos a hipótese nula de que o modelo nulo é melhor através da estatística de deviance igual a **63,49** que segue uma distribuição  $\chi^2$  com 3 graus de liberdade e p-valor menor que 0,001 ( $p = 1.1e-13$ ). Conclusão, o modelo com a covariável tratamento é melhor.

Todos os tratamentos são altamente significativos no aumento da sobrevivência, mas a terapia potente aumenta mais a sobrevivência do que a terapia combinada e esta tem um melhor efeito do que a monoterapia. Em termos do risco de óbito, a razão dos riscos de pacientes sem tratamento e com a terapia potente é aproximadamente **20,4** (valor extremamente alto).

- b. Faça uma análise gráfica do ajuste do modelo, comparando-o com a curva de Kaplan-Meier estratificada por tratamento. O que você tem a dizer sobre a adequação do modelo exponencial?

**Resposta:**

```
# item b
# Ajustando Kaplan-Meier estratificado por tratamento
km <- survfit(Surv(tempo, status) ~ factor(tratam), data = ipec)
plot(km, ylab = "S(t)", xlab = "dias", conf.int = F, col = 1:4, mark.time = F)
title("Tratamento em Aids")

# Pacientes sem tratamento
# Atribuindo o valor estimado para paciente sem tratamento (tratam=0) no modelo
exponencial
alpha0 <- exp(-6.14)
sobre0 <- function(x) {exp(-alpha0 * x)}
# Adicionando a nova curva ao gráfico existente (add=T)
curve(sobre0, from = 0, to = 3500, lty = 2, add = T, col = 1)

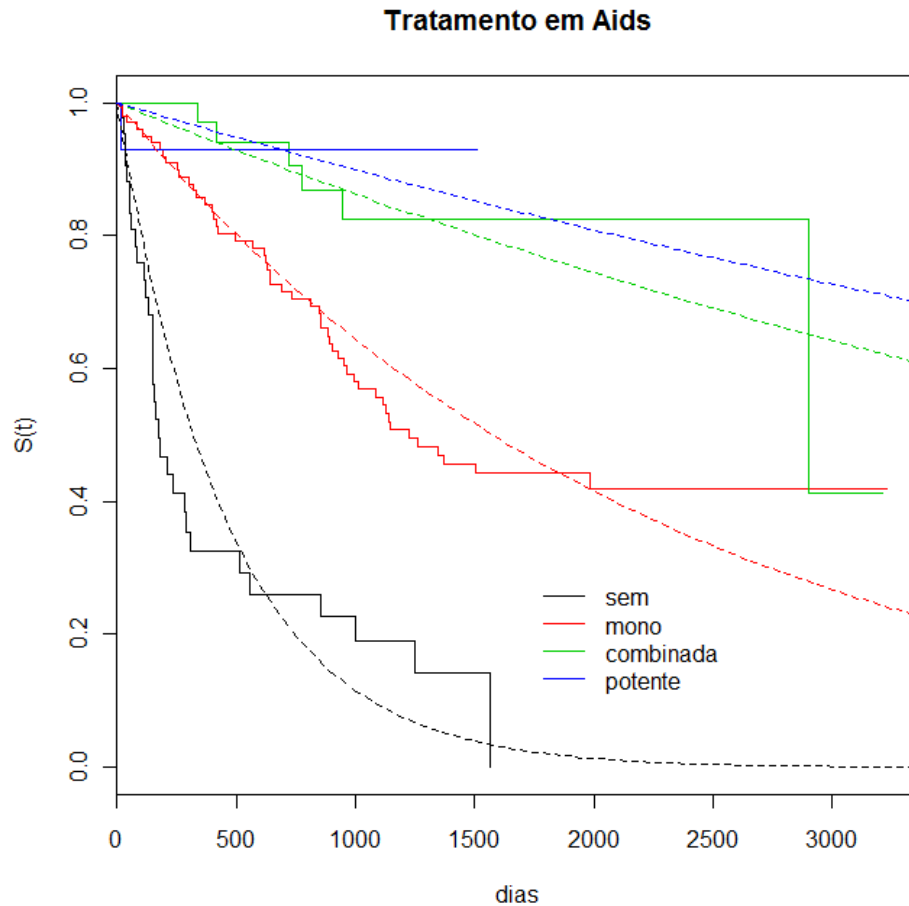
# Pacientes em monoterapia
# Atribuindo o valor estimado para paciente com monoterapia (tratam=1) no modelo
exponencial
alpha1 <- exp(-6.14 - 1.59)
sobre1 <- function(x) {exp(-alpha1 * x)}
curve(sobre1, from = 0, to = 3500, lty = 2, add = T, col = 2)

# Pacientes em terapia combinada
# Atribuindo o valor estimado para paciente com terapia combinada (tratam=2) no
modelo exponencial
alpha2 <- exp(-6.14 - 2.68)
sobre2 <- function(x) {exp(-alpha2 * x)}
curve(sobre2, from = 0, to = 3500, lty = 2, add = T, col = 3)

# Pacientes em terapia potente
# Atribuindo o valor estimado para paciente com terapia potente (tratam=3) no modelo
exponencial
alpha3 <- exp(-6.14 - 3.01)
sobre3 <- function(x) {exp(-alpha3 * x)}
curve(sobre3, from = 0, to = 3500, lty = 2, add = T, col = 4)

# Identificando as curvas
```

```
legend(1700, 0.3, c("sem", "mono", "combinada", "potente"), bty = "n", col = 1:4, lty = 1)
```



O modelo exponencial se ajusta razoavelmente bem para os grupos em que foram observados mais óbitos (Sem Terapia e Monoterapia).

```
# Verificando a Qualidade do Ajuste (Deviance)
# Deviance – Ajuste global
modipec.dev <- sum(resid(mod.ipeec, type='deviance')^2)
modipec.dev
[1] 264.3447
# p-valor da estatística de teste com 189 graus de liberdade
1-pchisq(modipec.dev,189)
[1] 0.0002434806
```

Rejeita-se a hipótese nula, logo se conclui que o modelo não é um bom ajuste para os tempos observados.



- c. Faça a análise de resíduos do modelo estimado. Existe algum ponto influente, sobre a estimativa dos parâmetros (ldcase), sobre os valores preditos (ldresp) ou sobre o parâmetro de forma?

**Resposta:**

# Item c

# Análise de Resíduos do Modelo Exponencial estimado (mod.ipec)

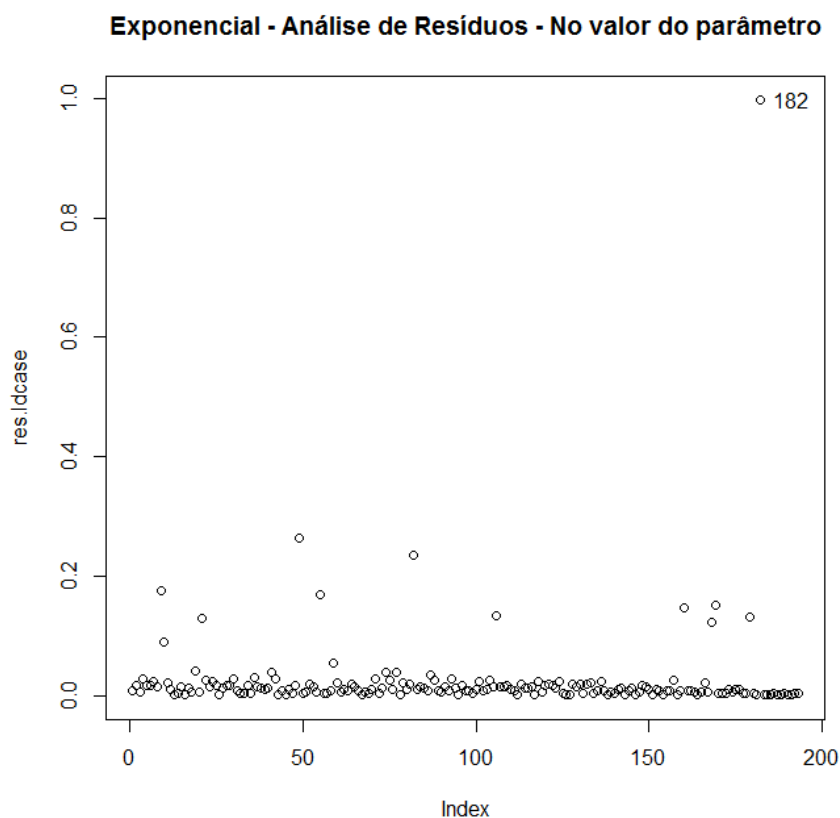
# Identificando observações mal ajustados

# ldcase: Identificando os indivíduos influentes na estimativa dos parâmetros

```
res.ldcase <- residuals(mod.ipec, type="ldcase")
```

```
plot (res.ldcase, main="Exponencial - Análise de Resíduos - No valor do parâmetro")
```

```
identify(res.ldcase)
```



# Retirando o paciente 182 mal ajustado

```
mod.ipec <- survreg(Surv(tempo, status) ~ factor(tratam), data = ipec, dist = "exp",  
subset= -182)
```

```
summary(mod.ipec)
```

Call:

```
survreg(formula = Surv(tempo, status) ~ factor(tratam), data = ipec,  
subset = -182, dist = "exp")
```

	Value	Std. Error	z	p
(Intercept)	6.14	0.177	34.72862	2.91e-264
factor(tratam)Mono	1.59	0.226	7.06692	1.58e-12

```
factor(tratam)Combinada 2.68      0.445  6.01449  1.80e-09
factor(tratam)Potente  19.32    3481.762  0.00555  9.96e-01
```

Scale fixed at 1

Exponential distribution

```
Loglik(model)= -732.7  Loglik(intercept only)= -767
      Chisq= 68.55 on 3 degrees of freedom, p= 8.7e-15
Number of Newton-Raphson Iterations: 18
n= 192
```

# Repetindo a verificação da Qualidade do Ajuste (Deviance) no modelo modificado

# Deviance – Ajuste global

```
modipec.dev <- sum(resid(mod.ipec, type='deviance')^2)
```

```
modipec.dev
```

```
[1] 251.5838
```

```
1-pchisq(modipec.dev,189)
```

```
[1] 0.001573104
```

O modelo ainda não se ajusta bem aos dados.

# Como o modelo Exponencial mesmo sem o indivíduo 182 não apresentou um bom ajuste, vamos ajustar uma Weibull

# Ajustando uma Weibull com Sexo, Idade e tipo de acompanhamento

```
modipec.wei <- survreg(Surv(tempo, status) ~ factor(tratam) + sexo + idade, data = ipec, dist = "weibull")
```

```
summary(modipec.wei)
```

Call:

```
survreg(formula = Surv(tempo, status) ~ factor(tratam) + sexo +
      idade, data = ipec, dist = "weibull")
```

	Value	Std. Error	z	p
(Intercept)	5.97802	0.5618	10.640	1.93e-26
factor(tratam)Mono	1.64172	0.2678	6.130	8.76e-10
factor(tratam)Combinada	2.83399	0.5426	5.223	1.76e-07
factor(tratam)Potente	3.37902	1.1804	2.863	4.20e-03
sexoM	-0.20476	0.3247	-0.631	5.28e-01
idade	0.00903	0.0127	0.709	4.78e-01
Log(scale)	0.12519	0.0867	1.443	1.49e-01

Scale= 1.13

Weibull distribution

```
Loglik(model)= -741.4  Loglik(intercept only)= -770.3
      Chisq= 57.96 on 5 degrees of freedom, p= 3.2e-11
Number of Newton-Raphson Iterations: 5
n= 193
```

# Verificando a Qualidade do Ajuste (Deviance)

# Deviance – Ajuste global

```
modipec.dev <- sum(resid(modipec.wei, type= 'deviance'))^2)
```

```
modipec.dev
```

```
[1] 238.7927
```

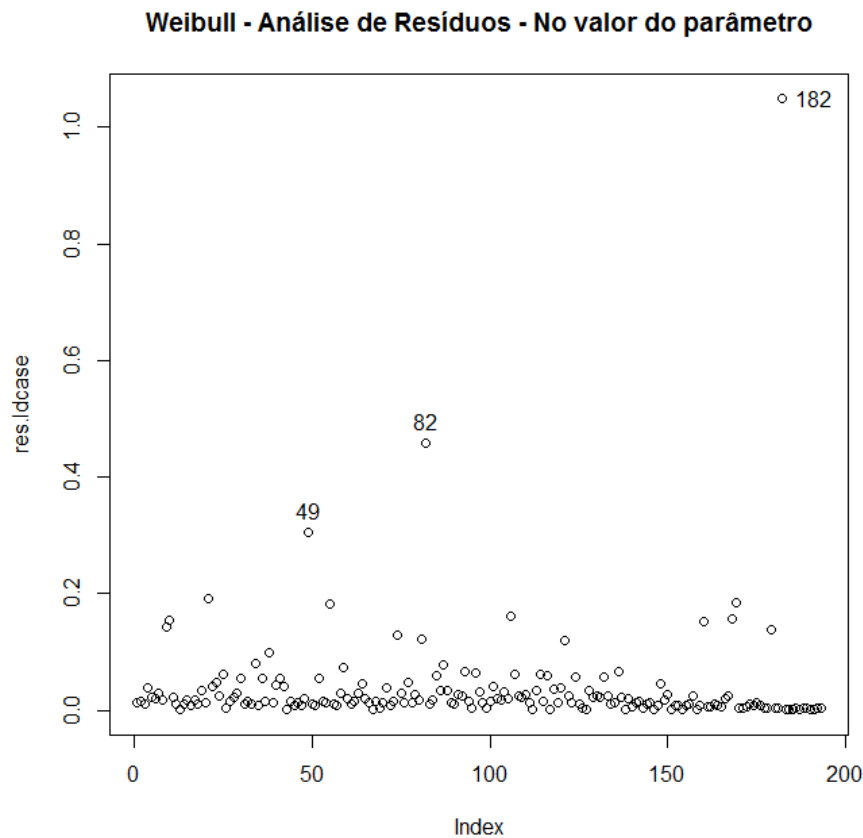
```
1-pchisq(modipec.dev,189)
```

```
[1] 0.00821915
```

```
res.lcase <- residuals(modipec.wei, type='lcase')
```

```
plot (res.lcase, main="Weibull - Análise de Resíduos - No valor do parâmetro")
```

```
identify(res.lcase)
```



Os indivíduos mal ajustados são, em ordem 182 e 82.

```
# Vendo os dados dos indivíduos mal ajustados
```

```
ipec[c(182,82),c(4,5,6,8,13)]
```

```
      tempo status sexo idade  tratam
182     16      1    M    42 Potente
82    1272      0    M    22  Nenhum
```

Pode ser avaliado se é adequado ou não manter esses indivíduos no modelo. Vamos retirar apenas o indivíduo 182:

```
# Retirando o paciente 182 mal ajustado
```

```
modipec.wei2 <- survreg(Surv(tempo, status) ~ factor(tratam) + sexo + idade, data =  
ipec, dist = "weibull", subset= -182)
```

## summary(modipec.wei2)

```
Call:
survreg(formula = Surv(tempo, status) ~ factor(tratam) + sexo +
  idade, data = ipec, subset = -182, dist = "weibull")

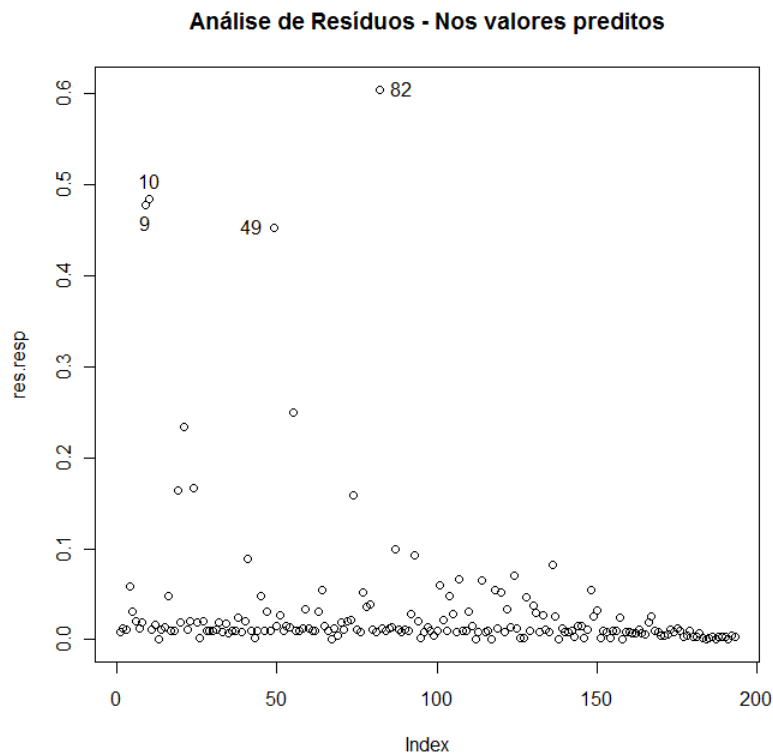
              Value Std. Error      z      p
(Intercept)    5.93673    0.5527 10.741 6.54e-27
factor(tratam)Mono    1.62916    0.2632  6.189 6.06e-10
factor(tratam)Combinada  2.79794    0.5330  5.250 1.52e-07
factor(tratam)Potente 16.46655    0.0000   Inf 0.00e+00
sexoM           -0.19167    0.3192 -0.600 5.48e-01
idade            0.00998    0.0126  0.793 4.28e-01
Log(scale)      0.10638    0.0868  1.225 2.20e-01

Scale= 1.11

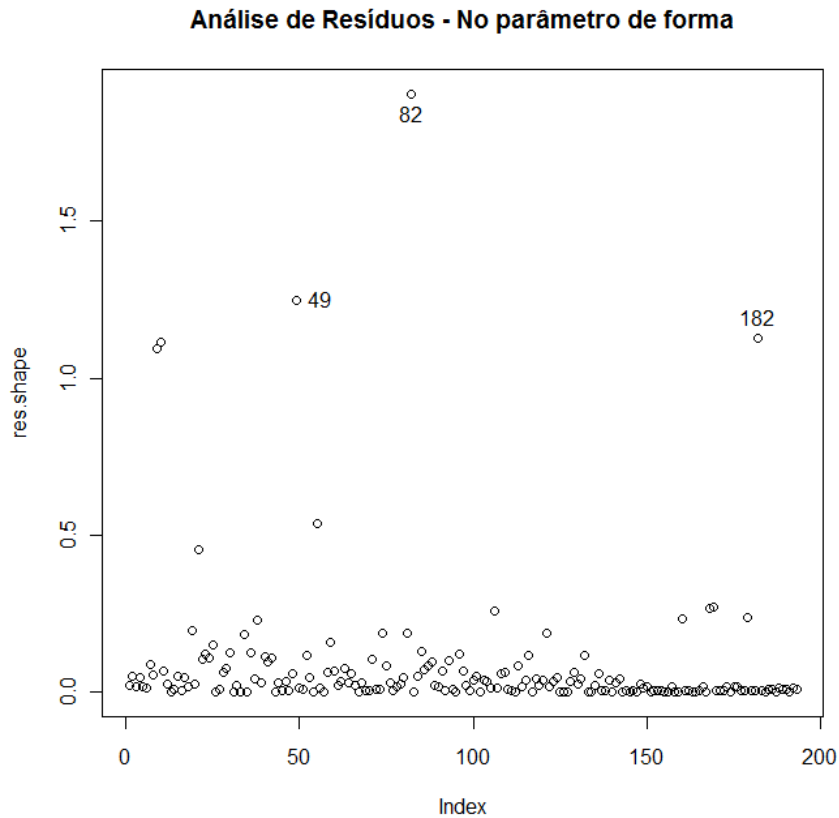
Weibull distribution
Loglik(model)= -731.5  Loglik(intercept only)= -763.4
  Chisq= 63.86 on 5 degrees of freedom, p= 1.9e-12
Number of Newton-Raphson Iterations: 14
n= 192
```

Observe que a maior alteração foi do coeficiente do tratamento potente, que passou de 16,4 para 3,8.

```
# Resíduo nos Valores Preditos no modelo original (sem retirar o ind 182)
res.resp <- residuals(modipec.wei, type='ldresp')
plot (res.lcase, main="Análise de Resíduos - Nos valores preditos")
identify(res.lcase)
```



```
# Idresp: Identificando os indivíduos influentes no parâmetro de forma
res.shape <- residuals(modipec.wei, type='ldshape')
plot (res.Idcase, main="Análise de Resíduos - No parâmetro de forma")
identify(res.shape)
```



d. Caso considere algum ponto muito influente, retire-o e refaça a análise.

**Resposta:**

# item d

# retirando o paciente identificado em todas as análises de resíduos: 82

```
modipec.wei3 <- survreg(Surv(tempo, status) ~ factor(tratam) + sexo + idade, data =
ipec, dist = "weibull", subset= -82)
summary(modipec.wei3)
```

Call:

```
survreg(formula = Surv(tempo, status) ~ factor(tratam) + sexo +
idade, data = ipec, subset = -82, dist = "weibull")
```

	Value	Std. Error	z	p
(Intercept)	5.6806	0.5688	9.986	1.75e-23
factor(tratam)Mono	1.7320	0.2628	6.590	4.40e-11
factor(tratam)Combinada	2.8897	0.5316	5.436	5.46e-08
factor(tratam)Potente	3.4432	1.1598	2.969	2.99e-03
sexoM	-0.2272	0.3188	-0.713	4.76e-01
idade	0.0151	0.0130	1.156	2.48e-01

```
Log(scale)                0.1071    0.0862  1.243 2.14e-01
```

```
Scale= 1.11
```

```
Weibull distribution
```

```
Loglik(model)= -738.3    Loglik(intercept only)= -769.7
```

```
    Chisq= 62.87 on 5 degrees of freedom, p= 3.1e-12
```

```
Number of Newton-Raphson Iterations: 5
```

```
n= 192
```

```
# Verificando a Qualidade do Ajuste (Deviance)
```

```
# Deviance – Ajuste global
```

```
Modipecwei3.dev <- sum(resid(modipec.wei, type='deviance')^2)
```

```
Modipecwei3.dev
```

```
[1] 235.8687
```

```
1-pchisq(modipec.dev,189)
```

```
[1] 0.01161369
```

Observamos que apesar do modelo estar mais bem ajustado aos dados ainda não tem um bom ajuste global.

- e. Ajuste um outro modelo exponencial, adicionando variáveis de controle (sexo, idade e tipo de atendimento). Quais variáveis tiveram efeito significativo? Quais tiveram efeito protetor?

**Resposta:**

```
# Item e
```

```
# Novo modelo com covariáveis sexo, idade e tipo de atendimento.
```

```
Mod3.ipec <- survreg(Surv(tempo, status) ~ factor(tratam) + sexo + idade +
```

```
factor(acompan), data = ipec, dist = "exp")
```

```
summary(mod3.ipec)
```

```
Call:
```

```
survreg(formula = Surv(tempo, status) ~ factor(tratam) + sexo +  
         idade + factor(acompan), data = ipec, dist = "exp")
```

	Value	Std. Error	z	p
(Intercept)	7.95467	0.6554	12.137	6.69e-34
factor(tratam)1	1.38695	0.2972	4.667	3.06e-06
factor(tratam)2	2.21397	0.4656	4.755	1.99e-06
factor(tratam)3	2.98559	1.0165	2.937	3.31e-03
sexoM	-0.07670	0.2833	-0.271	7.87e-01
idade	-0.00292	0.0120	-0.242	8.09e-01
factor(acompan)1	-1.70869	0.4064	-4.205	2.61e-05
factor(acompan)2	-2.23186	0.4664	-4.785	1.71e-06

```
Scale fixed at 1
```

```
Exponential distribution
```

```
Loglik(model)= -723.5    Loglik(intercept only)= -774.6
```

```
    Chisq= 102.25 on 7 degrees of freedom, p= 0
```

```
Number of Newton-Raphson Iterations: 6
```

```
n= 193
```

Os fatores sexo e idade são não significativos ( $p=0,0787$  e  $p=0,0809$ , respectivamente). Dentre os fatores significativos somente tratamento teve um efeito protetor. O tipo de atendimento – internação hospitalar posterior ou imediata, comparadas a tratamento apenas ambulatorial – é significativo, e indica maior risco para pacientes que necessitaram internação.

**Exercício 5.11:** Use a distribuição Weibull, para ajustar um modelo para os pacientes em diálise (arquivo diálise.csv) descrito no Apêndice 12.4..

- a. Avalie qual é o efeito da doença de base na sobrevivência, controlado por idade.

**Resposta:**

```
# Exercício 5.11
dialise <- read.table("dialise.csv", header = T, sep = ",")
# item a
# A distribuição default é Weibull
# Modelo Weibull com idade + cdiab + crim + congenita
modelo1 <- survreg(Surv(tempo, status) ~ idade + cdiab + crim + congenita, data =
dialise)
summary(modelo1)
```

```
Call:
survreg(formula = Surv(tempo, status) ~ idade + cdiab + crim +
congenita, data = dialise)

              Value Std. Error      z      p
(Intercept)  6.7737    0.14999  45.161 0.00e+00
idade        -0.0428    0.00225 -19.017 1.24e-80
cdiab        -0.3605    0.07353  -4.903 9.44e-07
crim         -0.0384    0.08139  -0.472 6.37e-01
congenita     0.8855    0.27529   3.217 1.30e-03
Log(scale)   0.1951    0.02082   9.373 7.04e-21

Scale= 1.22

Weibull distribution
Loglik(model)= -7857.3   Loglik(intercept only)= -8104.2
      Chisq= 493.87 on 4 degrees of freedom, p= 0
Number of Newton-Raphson Iterations: 7
n= 6805
```

A diabetes aumenta o risco e as doenças congênitas têm efeito protetor.

- b. Existe evidência a favor da utilização de um modelo mais simples (exponencial)?

**Resposta:**

```
# Ajustando um modelo exponencial
# item b
modeloE <- survreg(Surv(tempo, status) ~ idade + cdiab + crim + congenita, dist =
"exp", data = dialise)
summary(modeloE)
```

```
Call:
```

```

survreg(formula = Surv(tempo, status) ~ idade + cdiab + crim +
        congenita, data = dialise, dist = "exp")
      Value Std. Error      z      p
(Intercept)  6.1643    0.10897  56.568 0.00e+00
idade       -0.0365    0.00176 -20.676 5.66e-95
cdiab       -0.3092    0.06029  -5.127 2.94e-07
crim        -0.0313    0.06696  -0.467 6.41e-01
congenita    0.7550    0.22616   3.338 8.43e-04

Scale fixed at 1

Exponential distribution
Loglik(model)= -7905.3  Loglik(intercept only)= -8169
      Chisq= 527.4 on 4 degrees of freedom, p= 0
Number of Newton-Raphson Iterations: 6
n= 6805)

```

```

# Comparando os modelos Weibull e exponencial, isto é,
# testando a hipótese que o parâmetro de forma é igual a 1,
# através da estatística de deviance

```

```

dev <- 2 * (modelo1$loglik[2] - modeloE$loglik[2])
dev
[1] 95.95186

```

```

gl <- 1
pvalor <- 1 - pchisq(dev, gl)
pvalor
[1] 0

```

```

# Alternativamente pode-se testar a redução do modelo usando o comando anova().
# Observe que na coluna Deviance temos o mesmo valor que calculamos
anteriormente (dev = 95.95)

```

```

anova(modeloE, modelo1)

```

Terms	Resid.	Df	-2*LL	Test	Df	Deviance	P(> Chi )
1 idade + cdiab + crim + congenita			6800	15810.56		NA	NA
2 idade + cdiab + crim + congenita			6799	15714.61	= 1	95.95186	1.177112e-22

Segundo o teste apresentado no sumário do modelo1 (modelo Weibull) e no teste de razão de verossimilhança, a estimativa do parâmetro de escala é significativamente diferente de 1 e portanto a redução para um modelo com menos parâmetros (exponencial) não seria adequada.

- c. Existe evidência a favor da utilização de um modelo com menos variáveis?

**Resposta:**

```

# Ajustando um modelo sem a covariável causas renais (crim)

```

```

# item c

```

```

# Modelo 2 - Weibull com idade + cdiab + congenita

```

```

modelo2 <- survreg(Surv(tempo, status) ~ idade + cdiab + congenita, data = dialise)

```



## summary(modelo2)

```
Call:
survreg(formula = Surv(tempo, status) ~ idade + cdiab + congenita,
        data = dialise)
```

```
              Value Std. Error      z      p
(Intercept)  6.7623    0.14798  45.70 0.00e+00
idade        -0.0428    0.00225 -19.00 1.70e-80
cdiab        -0.3510    0.07061  -4.97 6.69e-07
congenita     0.8951    0.27454   3.26 1.11e-03
Log(scale)   0.1951    0.02082   9.37 7.05e-21
```

```
Scale= 1.22
```

```
Weibull distribution
```

```
Loglik(model)= -7857.4   Loglik(intercept only)= -8104.2
      Chisq= 493.64 on 3 degrees of freedom, p= 0
Number of Newton-Raphson Iterations: 7
n= 6805
```

```
# Testando a hipótese nula de que o modelo reduzido é melhor,
# ou em outras palavras, que o coeficiente da covariável crim não é significativo.
```

```
dev <- 2 * (modelo1$loglik[2] - modelo2$loglik[2])
```

```
dev
```

```
[1] 0.2218873
```

```
gl <- modelo1$df - modelo2$df
```

```
gl
```

```
[1] 1
```

```
pvalor <- 1 - pchisq(dev, gl)
```

```
pvalor
```

```
[1] 0.6376056
```

```
# Alternativamente pode-se testar a redução do modelo usando o comando anova
anova(modelo2, modelo1)
```

```
Terms Resid. Df    -2*LL  Test Df  Deviance P(>|Chi|)
1      idade + cdiab + congenita      6800 15714.83      NA      NA      NA
2  idade + cdiab + crim + congenita      6799 15714.61 +crim  1 0.2218873 0.6376056
```

Adotando-se o modelo Weibull, testou-se a redução do modelo por exclusão da variável crim, e o modelo reduzido não pode ser rejeitado sendo possível a retirada de tal variável do modelo. Isto significa que, segundo o modelo Weibull, a variável crim não é um fator prognóstico importante para a sobrevivência.

### d. Existe evidência de pontos influentes

**Resposta:**

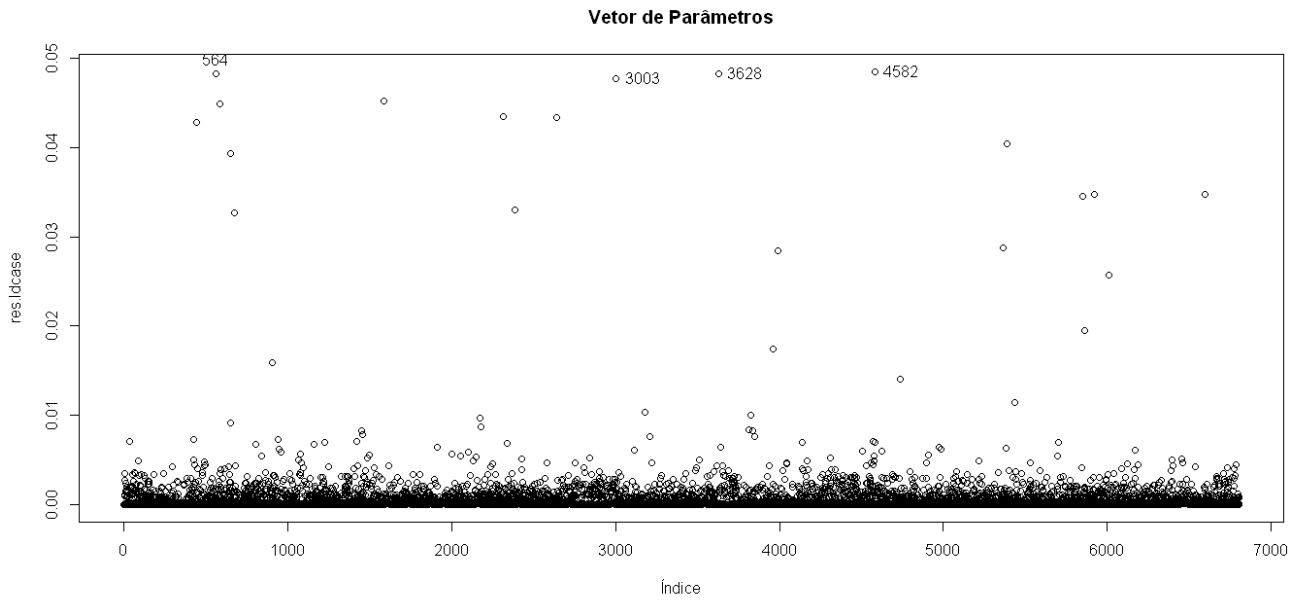
```
# Item d
```

```
# Verificando se existe evidência de pontos influentes
```

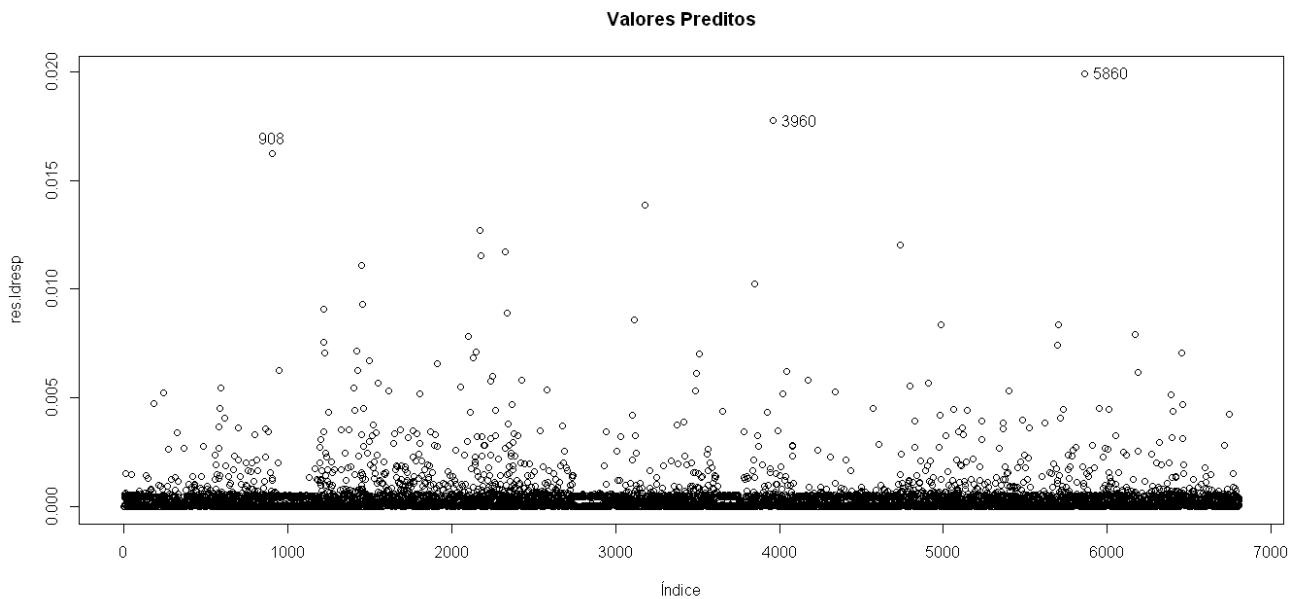
```
# Avaliando os resíduos ditos de “perturbação” do Modelo 2
```

```
# Avaliando as mudanças no vetor estimado de parâmetros
```

```
res.lcase<- residuals(modelo2, type="ldcase")
plot(res.lcase, main="Vetor de Parâmetros", xlab="Índice")
identify(res.lcase)
```

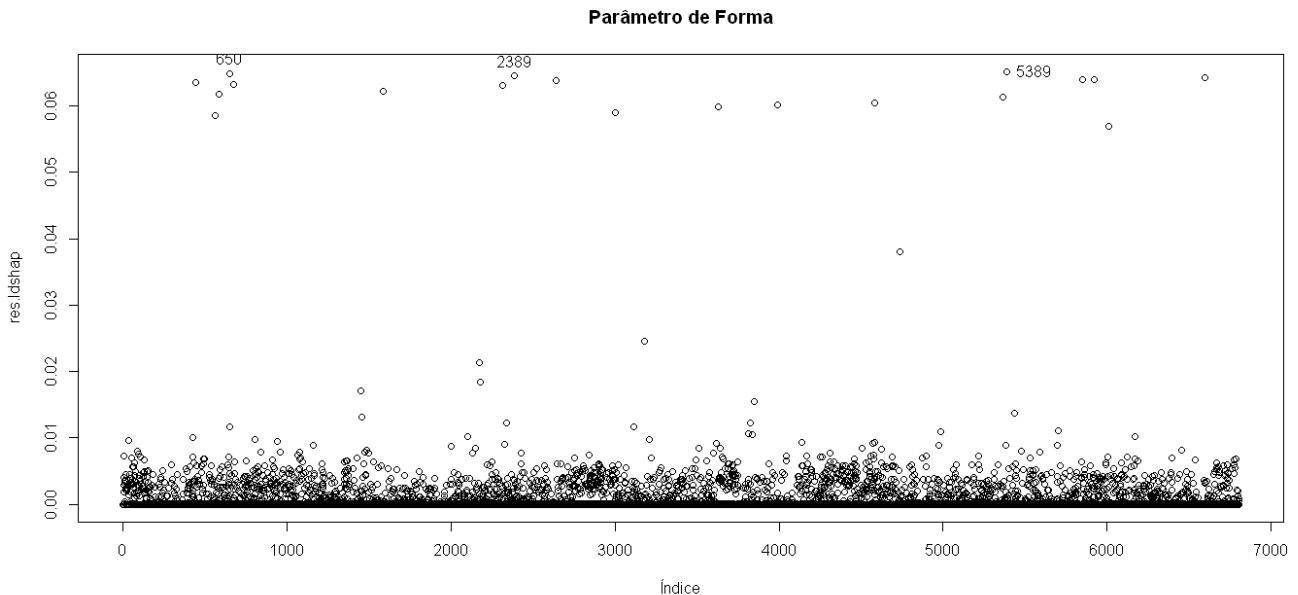


```
# Avaliando as mudanças nos valores preditos (em unidades de desvio padrão)
res.lcase<- residuals(modelo1, type="ldresp")
plot(res.lcase, main="Valores Preditos", xlab="Índice")
identify(res.lcase)
```



```
# Avaliando as mudanças no parâmetro de forma estimado
res.lcase<- residuals(modelo1, type="ldshap")
```

```
plot(res.lidcase, main="Parâmetro de Forma", xlab="Índice")
```



Não existe nenhuma evidência da existência de pontos influentes de fato, dado que a escala é muito pequena (0,02 a 0,06) e principalmente devido ao grande tamanho da amostra.

```
# Listando os casos identificados como os mais influentes (res.lidshap>0,065)
dialise[(650,2359,5389),c(2,3,4,5,6,9,11)]
```

	idade	inicio	fim	status	tempo	cdiab	congenita
650	50	2	21	1	19	0	1
2359	15	32	45	0	13	0	0
5389	73	4	10	1	6	0	1

Observamos que os pacientes 650 e 5389 sobreviveram por pouco tempo (19 e 6 meses, respectivamente) e faziam hemodiálise devido a doença congênita embora ambos fossem adultos. Por outro lado, o paciente 2359 era um jovem de 15 anos, acompanhado por apenas 13 meses e censura.

- e. Exclua do modelo as variáveis com p-valor > 0,1 e compare o novo modelo com o anterior, calculando a razão de verossimilhança entre os dois.

### Resposta:

```
# Item e
# Excluindo do modelo as variáveis com p-valor > 0,1
# Observando o modelo 2
# Modelo Weibull com idade + cdiab + congenita
Modelo2 <- survreg(Surv(tempo, status) ~ idade + cdiab + congenita, data = dialise)
summary(modelo2)
```

Call:

```

survreg(formula = Surv(tempo, status) ~ idade + cdiab + congenita,
        data = dialise)
      Value Std. Error      z      p
(Intercept)  6.7623    0.14798  45.70 0.00e+00
idade       -0.0428    0.00225 -19.00 1.70e-80
cdiab       -0.3510    0.07061  -4.97 6.69e-07
congenita    0.8951    0.27454   3.26 1.11e-03
Log(scale)  0.1951    0.02082   9.37 7.05e-21

Scale= 1.22

Weibull distribution
Loglik(model)= -7857.4  Loglik(intercept only)= -8104.2
      Chisq= 493.64 on 3 degrees of freedom, p= 0
Number of Newton-Raphson Iterations: 7
n= 6805

```

Já não há nenhuma variável com p-valor >0,1. Assim este pode ser considerado o melhor modelo ajustado. Podemos retirar um dos indivíduos que poderiam ser considerados como pontos influentes na estimação do parâmetro de forma.

```

# Modelo3 - Weibull sem o indivíduo 650 (idade + cdiab + congênita)
modelo3 <- survreg(Surv(tempo, status) ~ idade + cdiab + congenita, data = dialise,
subset=-650)
summary(modelo3)

```

```

Call:
survreg(formula = Surv(tempo, status) ~ idade + cdiab + congenita,
        data = dialise, subset = -650)
      Value Std. Error      z      p
(Intercept)  6.7657    0.14812  45.68 0.00e+00
idade       -0.0429    0.00226 -19.01 1.49e-80
cdiab       -0.3510    0.07064  -4.97 6.74e-07
congenita    0.9505    0.28171   3.37 7.41e-04
Log(scale)  0.1955    0.02083   9.39 6.11e-21

Scale= 1.22

Weibull distribution
Loglik(model)= -7852  Loglik(intercept only)= -8099.6
      Chisq= 495.16 on 3 degrees of freedom, p= 0
Number of Newton-Raphson Iterations: 7
n= 6804

```

Podemos observar que a retirada do indivíduo 650 não alterou em muito as estimativas dos coeficientes e do parâmetro de forma:

Coefficientes da regressão (ldcase): sem alteração em cdiab e pequenas alterações (3ª ou 4ª casa decimal) nos demais.

Verossimilhança (ldresp): passou de -7857 para -7852.

Parâmetro de Forma (ldshap): Sem alteração:  $1/1,22 = 0,8197$