Análise de Sobrevida
Teoria e Aplicações em Saúde

Caderno de Respostas

Marilia Sá Carvalho
Valeska Lima Andreozzi
Claudia Torres Codeço
Maria Tereza Serrano Barbosa
Silvia Emiko Shimakura

5

Estimação paramétrica

Exercícios

Exercício 5.1: Em um estudo, ajustou-se um modelo exponencial aos tempos de sobrevida observados nos grupos controle e tratamento. Os modelos encontrados foram:

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

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

Com base nesses modelos, responda:

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

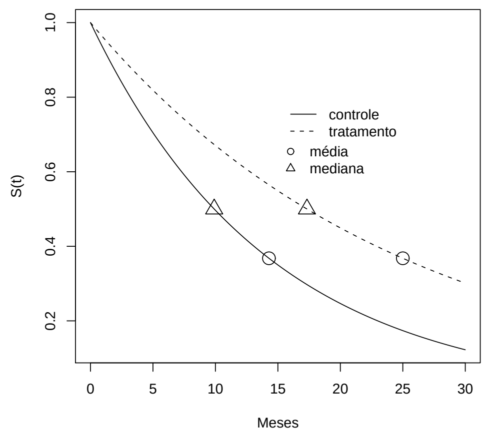
   Resposta: Segundo os resultados acima a estimativa do parâmetro da distribuição exponencial para o grupo controle, digamos $\hat{\lambda}_c$, foi de $\hat{\lambda}_c = 0,07$, e para o grupo tratamento foi de $\hat{\lambda}_{tr} = 0,04$. Assim, o risco estimado em qualquer tempo sob a distribuição exponencial para o grupo controle é 0,07 enquanto que para o grupo tratamento é 0,04.

2. Qual foi a sobrevida média e mediana no grupo controle? E no grupo recebendo tratamento?

   Resposta: O tempo médio de sobrevida é dado por $\bar{T} = \frac{1}{\alpha}$. Logo, para o grupo controle temos $\bar{T} = \dfrac{1}{0,07} = 14,28$ e para o grupo placebo temos $\bar{T} = \dfrac{1}{0,04} = 25$. Já o tempo mediano de sobrevida é dado por $T_{mediano} =$

$\dfrac{\ln(2)}{\alpha}$. Sendo assim, temos que o tempo mediano para o grupo controle igual a $T_{mediano} = \dfrac{\ln(2)}{0,07} = 9,90$ e $T_{mediano} = \dfrac{\ln(2)}{0,04} = 17,32$

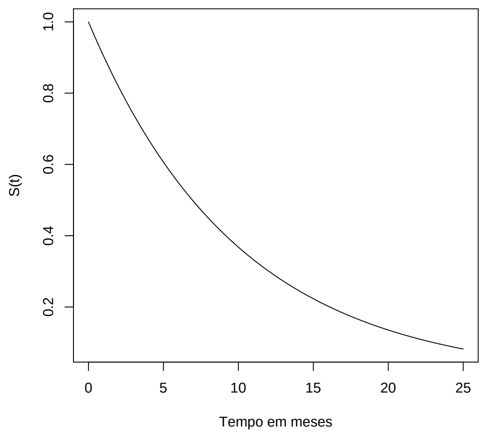
3. As duas curvas estimadas de sobrevida são apresentadas na figura que segue. Localize, nesta, o tempo mediano e médio que você calculou. Com base neste gráfico, você acha que o tratamento teve efeito na sobrevida desses pacientes?



Resposta: O gráfico sugere que o tratamento teve um efeito significativo no aumento do tempo de sobrevida dos pacientes. No entanto para que possamos tirar conclusões estatísticamente conclusivas é importante considerar tanto a variabilidade nas curvas estimadas quanto nos tempos médios e medianos de sobrevida estimados.

Exercício 5.2: No R, faça gráficos da função de sobrevida de acordo com um modelo exponencial utilizando $\alpha = 0,1$. Calcule o tempo mediano de sobrevida de acordo com este modelo. Calcule também o percentil 90 (P90) e o percentil 10 (P10), isto é, o tempo em que 90% e 10% dos pacientes, respectivamente, ainda não tinham sofrido o evento.

```
> alfa <- 0.1
> curve(exp(-alfa * x), from = 0, to = 25, ylab = "S(t)", xlab = "Tempo em meses")
```



Percentil 90

```
> p90 <- log(1/0.9)/alfa
> p90
```

```
[1] 1.053605
```

Percentil 10

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

```
[1] 23.02585
```

Com base nesses comando do R:

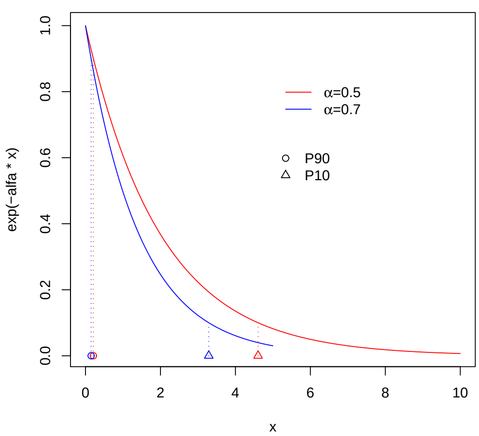
1. Troque o valor do parâmetro para $\alpha = 0,5$ e $\alpha = 0,7$.

   ```
   > alfa <- 0.5
   > curve(exp(-alfa * x), from = 0, to = 10, lty = 1, col = "red")
   > p90 <- log(1/0.9)/alfa
   > p90
   ```

   ```
   [1] 0.2107210
   ```

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

```
[1] 4.60517

> segments(p90, 0, p90, exp(-alfa * p90), lty = 3, col = "red")
> points(p90, 0, pch = 1, 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, add = T, lty = 1, col = "blue")
> p90 <- log(1/0.9)/alfa
> p90

[1] 0.1505150

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

[1] 3.289407

> segments(p90, 0, p90, exp(-alfa * p90), lty = 3, col = "blue")
> points(p90, 0, pch = 1, col = "blue")
> segments(p10, 0, p10, exp(-alfa * p10), lty = 3, col = "blue")
> points(p10, 0, pch = 2, col = "blue")
> legend(5, 0.85, c(expression(paste(alpha, "=0.5")), expression(paste(alpha,
+     "=0.7"))), lty = 1, col = c("red", "blue"), bty = "n")
> legend(5, 0.65, c("P90", "P10"), pch = 1:2, bty = "n")
```



2. Observe o comportamento da função.

Resposta: Note que a função de sobrevivência cai mais rapidamente a medida que aumenta o valor do parâmetro α, 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 α.

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

1. 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.

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

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

3. Qual a relação entre o parâmetro γ e o comportamento da função de risco?

   Resposta: Quando $\gamma = 1$ a função de risco é constante, ou seja, o risco intantâneo 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.

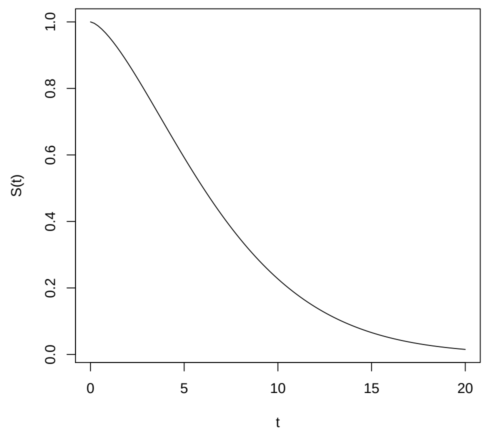
4. Quais das curvas de risco apresentadas na Figura 3.3 não poderiam ser modeladas pela função Weibull, nem mesmo aproximadamente?

   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 descrescente. O que se vê nos quadros D, E e F são misturas destes comportamentos.

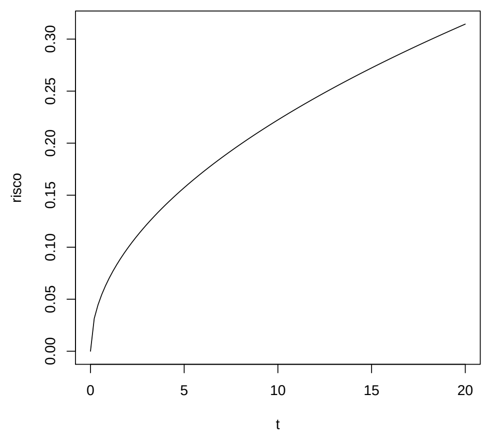
Exercício 5.4: Seja T o tempo de sobrevida até a ocorrência de um evento, que segue uma distribuição Weibull com parâmetros $\gamma = 1,5$ e $\alpha = 0,13$.

1. Escreva as funções $S(t)$, $\lambda(t)$ e $\Lambda(t)$ e use o R para fazer o respectivos gráficos.

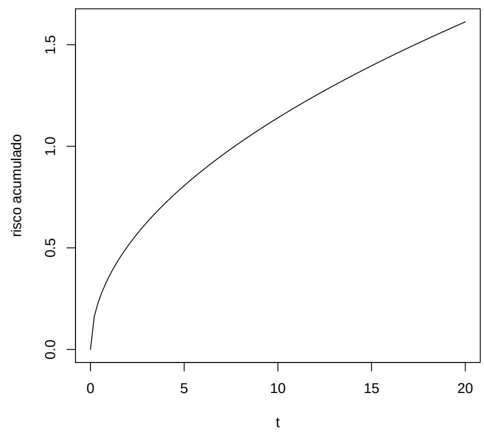
```
> alfa <- 0.13
> gama <- 1.5
> curve(exp(-(alfa * x)^gama), from = 0, to = 20, ylab = "S(t)",
+     xlab = "t")
```




```
> curve(alfa * gama * (alfa * x)^(gama - 1), from = 0, to = 20,
+     ylab = "risco", xlab = "t")
```




```
> curve((alfa * x)^(gama - 1), from = 0, to = 20, ylab = "risco acumulado",
+     xlab = "t")
```



2. Calcule o tempo mediano de sobrevida. Calcule o percentil 80 e o percentil 10 dessa distribuição.

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

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

[1] 6.024767
```

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

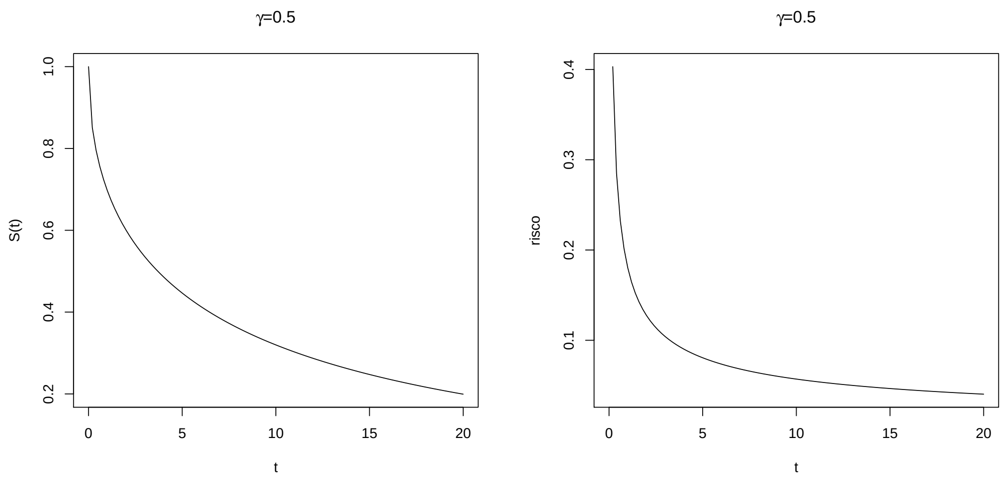
```
> p80 <- log(1/0.8)^(1/gama)/alfa
> p80

[1] 2.829955
```

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

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

[1] 13.41324
```

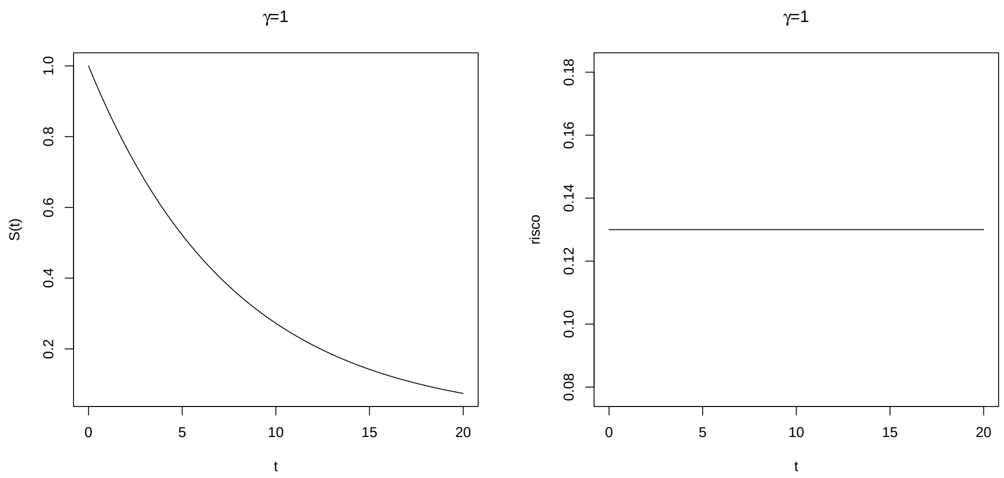
3. Fixe o valor do parâmetro $\alpha = 0,13$ e faça gráficos da função de risco e da função de sobrevida para diversos valores de γ: $0 < \gamma < 1$, $\gamma = 1$ e $\gamma > 1$. Visualize como o parâmetro γ afeta o comportamento do risco e da sobrevida.

```
> 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")))
```



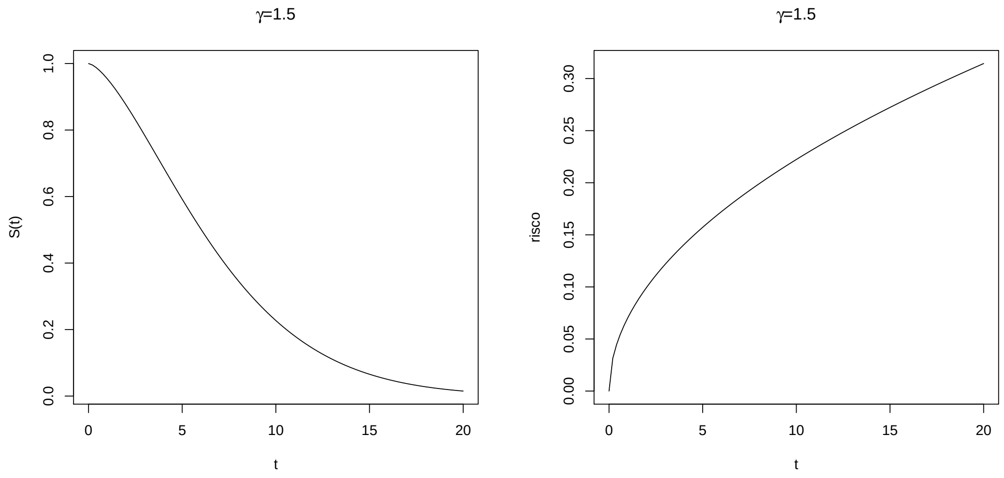
Sobrevida cai rapidamente no início do acompanhamento, e risco é decrescente.

```
> 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")))
```



Sobrevida cai mais suavemente, risco constante.

```
> 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")))
```



Sobrevida cai suavemente no início do período, risco aumenta com o tempo decorrido.

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 $\alpha = 0,07$.

1. Calcule o tempo mediano de incubação desta infecção.

   ```
   > alfa <- 0.07
   > gama <- 1.2
   > tmediano <- log(1/0.5)^(1/gama)/alfa
   > tmediano

   [1] 10.52583
   ```

   Resposta: 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.

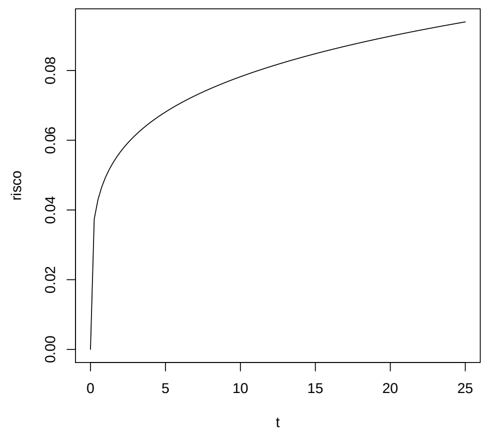
2. É correto dizer que em 10 horas do momento da infecção, espera-se que 80% das pessoas já tenham desenvolvido sintomas?

   ```
   > t <- 10
   > S10 <- exp(-(alfa * t)^gama)
   > S10

   [1] 0.5211044
   ```

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

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

   ```
   > par(mfrow = c(1, 1))
   > curve(gama * alfa^gama * x^(gama - 1), from = 0, to = 25, ylab = "risco",
   +     xlab = "t")
   ```



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

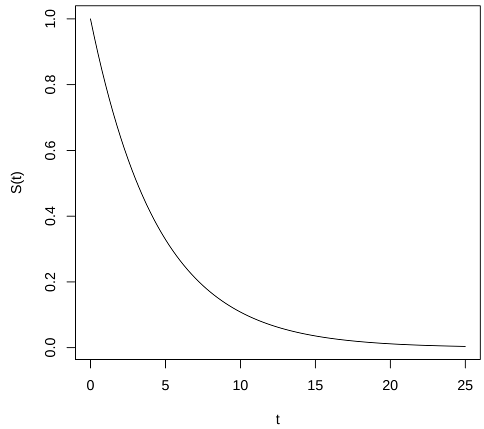
Exercício 5.6: Mil crianças não vacinadas são acompanhadas, a partir do nascimento, em um estudo cujo objetivo é identificar a idade em que adquirem hepatite A. Os resultados do estudo indicam que a idade média de soroconversão das crianças foi de 4,5 anos e que o risco de contrair hepatite A foi constante e independente da idade.

1. 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 (i.e, parcimonioso) e adequado (pois possui função risco constante) seria o modelo paramétrico exponencial.

2. Faça no R o gráfico da função de sobrevida, de acordo com esse modelo.

```
> tm <- 4.5
> alfa <- 1/tm
> curve(exp(-alfa * x), from = 0, to = 25, ylab = "S(t)", xlab = "t")
```



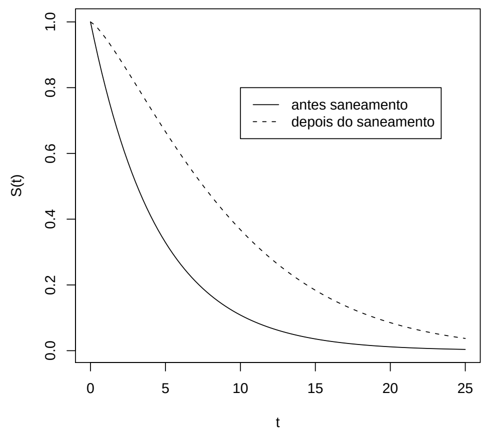
3. Com base nesse modelo, em que idade espera-se ter 90% das crianças soropositivas?

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

[1] 10.36163
```

Resposta: 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.

4. 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 sobrevida. 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 sobrevida.

```
> 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)
> 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 sobrevida 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: Retorne ao exemplo do Exercício 4.1, sobre tempo de aleitamento de crianças (arquivo: `leite.txt`).

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

```
> require(survival)
> leite <- read.table("leite.txt", header = T, sep = "")
> 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 $\gamma = 1/Scale$, 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:

```
> 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
```

Existem técnicas para comparar os dois modelos. Baseiam-se na razão de verossimilhança dos dois modelos, que diferem por um grau de liberdade. O modelo Weibull é de fato melhor.

2. Qual o tempo mediano de amamentação, estimado por esse modelo? (Dica: não se esqueça de que a parametrização das distribuições no R difere da vista no texto). Os parâmetros da distribuição Weibul são $alpha = \exp(-intercept)$ e $\gamma = 1/Scale$.

```
> alfa <- as.vector(exp(-modeloweib$coef[1]))
> alfa

[1] 0.1802502

> gama <- 1/modeloweib$scale
> gama
```

```
[1] 1.514787
```

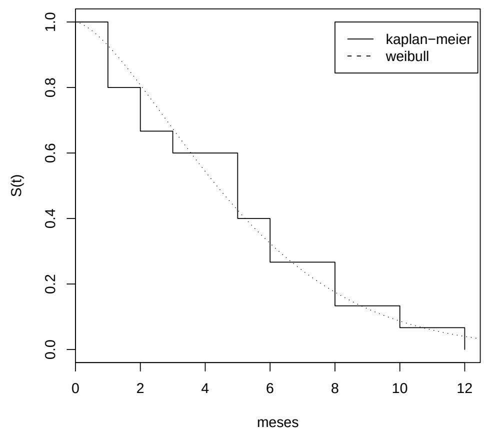
```
> tmediano <- log(1/0.5)^(1/gama)/alfa
> tmediano
```

```
[1] 4.355558
```

Resposta: O tempo mediano de amamentação estimado por este modelo é de 4.36 meses.

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

```
> 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))
```



Resposta: O modelo parece se ajustar muito bem aos dados.