

MI616 - Análise de Sobrevivência

Profa.: Hildete Prisco Pinheiro

2ª Lista de Exercícios

1. Prove que se T tem uma distribuição contínua arbitrária, então a função risco acumulada de T , $\Lambda(t)$, tem distribuição exponencial de parâmetro unitário.

2. Prove que o gráfico de $\log(-\log(\widehat{S}(t)))$ versus $\log(t)$ para dados de uma distribuição Weibull com parâmetros ρ e κ será aproximadamente uma reta com coeficiente angular κ e intercepto $-\kappa \log(\rho)$.

3. Considere os tempos de sobrevivência de 16 pacientes em semanas:

4, 20, 22, 25, 27, 38, 38, 40, 44, 56, 83, 89, 98, 110, 138 e 145.

Assumindo que eles seguem uma distribuição exponencial de parâmetro ρ , obtenha:

(a) O EMV de ρ .

(b) O EMV de $\mu = E(T)$

(c) Um intervalo de confiança (95%) exato para ρ e μ .

4. Assuma que X_i ($i = 1, \dots, n$) são i.i.d. com distribuição exponencial com parâmetro ρ . Se há censura do tipo II, prove que $2d\rho/\hat{\rho} \sim \chi_{2d}^2$, em que $\hat{\rho}$ é o EMV de ρ e d é o número de falhas. Dica: Use a distribuição conjunta dos tempos de falha ordenados $T_{(1)} < T_{(2)} < \dots < T_{(n)}$ e a definição de censura do tipo II.

5. Mostre que, se a função risco é dada por $\kappa\rho(\rho t)^{\kappa-1} \exp[(\rho t)^\kappa]$, a função de sobrevivência é dada por $\exp\{-[\exp((\rho t)^\kappa) - 1]\}$.

6. Mostre que na ausência de censura a matriz de informação esperada para a distribuição Weibull tem elementos:

$$\begin{aligned} -E\left(\frac{\partial^2 l}{\partial \rho^2}\right) &= \frac{n\kappa^2}{\rho^2}, & -E\left(\frac{\partial^2 l}{\partial \rho \partial \kappa}\right) &= \frac{n\psi(2)}{\rho}, \\ -E\left(\frac{\partial^2 l}{\partial \kappa^2}\right) &= \frac{n}{\kappa^2}\{1 + \psi'(2) + [\psi(2)]^2\}, \end{aligned}$$

em que $\psi(\alpha) = \frac{\partial}{\partial \alpha} \log \Gamma(\alpha)$ é a função digama.

7. Mostre que na ausência de censura, a distribuição gama tem estatística suficiente minimal $(\sum t_i, \sum \log t_i)$ para ρ e κ , e que o EMV de κ com ρ desconhecido é a solução de

$$\psi(\kappa) - \log(\kappa) - \log(R) = 0,$$

em que R é a razão entre a média geométrica e a média aritmética da amostra.

8. Os dados abaixo referem-se a um estudo feito por um produtor de requeijão, que deseja estudar dois tipos diferentes de embalagens para o seu produto (A e B). O produto dele é fabricado sem a utilização de conservantes e o interesse do estudo é avaliar o efeito das diferentes embalagens no tempo até o aparecimento de fungos. Os tempos medidos, em horas, até o aparecimento dos fungos. Os tempos medidos, em horas, até o aparecimento dos fungos em 40 potes (20 de cada embalagem) estão apresentados na tabela abaixo. Os tempos censurados à direita são denotados por um sinal "+".

| Embalagem | Tempo até o aparecimento de fungo, em horas | | | | | | | | | |
|-----------|---|----|----|----|----|----|-----------------|-----------------|-----------------|-----------------|
| A | 31 | 40 | 43 | 44 | 46 | 46 | 47 | 48 | 48 | 49 |
| | 50 | 50 | 60 | 60 | 60 | 60 | 60 ⁺ | 60 ⁺ | 60 ⁺ | 60 ⁺ |
| B | 48 | 48 | 49 | 49 | 49 | 49 | 50 | 50 | 50 | 50 |
| | 53 | 53 | 54 | 54 | 55 | 55 | 55 ⁺ | 55 ⁺ | 55 ⁺ | 55 ⁺ |

(a) Considerando cada grupo separadamente e supondo que os tempos de falha seguem uma distribuição exponencial, obtenha estimativas de máxima verossimilhança para os parâmetros envolvidos e escreva as correspondentes expressões para a densidade, função de sobrevivência e função risco.

(b) Encontre o erro padrão associado à estimativa do parâmetro obtido em (a).

(c) Utilizando as expressões das funções de sobrevivência para cada um dos casos acima, desenhe em um mesmo gráfico as curvas de sobrevivência para os dois grupos.

9. Para comparar dois tratamentos para tumores sólidos, dois grupos de ratos foram inoculados com um neoplasma altamente maligno, e depois da aleatorização entre os dois tratamentos, o tempo de sobrevivência (dias) foi observado:

Droga A: 12, 16, 16, 18+, 19+, 20, 28

Droga B: 10, 14, 15, 18, 18, 20+, 21.

(a) Calcule o estimador dde Kaplan-Meier das duas curvas de sobrevivência e os seus correspondentes erros padrão para cada tempo observado.

(b) Faça o gráfico das curvas de sobrevivência para cada grupo.

(c) Compare as duas curvas e comente sobre os resultados.

10. Compare os tempos de remissão nos dois tratamentos recebidos pelos pacientes com Melanoma, ajustando pelo sexo. Compare também os tempos de sobrevivência nos dois tratamentos ajustando pelo estágio inicial da doença. Os dados estão no arquivo *melanoma.dat* e a descrição das variáveis estão no arquivo *melanoma.txt*.

11. A tabela abaixo mostra os tempos de sobrevivência em semanas de 30 pacientes com tumor de cérebro recebendo quatro diferentes tratamentos. Será que os efeitos dos quatro tratamentos são iguais?

| | | | |
|------|----|-----|-----|
| 1 | 2 | 3 | 4 |
| 4 | 1 | 3 | 5 |
| 5 | 4 | 7 | 15 |
| 9 | 9 | 14 | 20 |
| 12 | 12 | 200 | 31 |
| 20+ | 15 | 27 | 39 |
| 25 | 23 | 30 | 47 |
| 30 + | 30 | 32+ | 55+ |
| | | 50+ | 67+ |

12. Os dados abaixo representam dados de sobrevivência para um estudo coorte comparando dois fatores de exposição. Os tempos representam os anos desde a entrada no estudo; o indicador de censura nos diz se o indivíduo morreu ($\delta = 1$) ou ainda está vivo ($\delta = 0$) na última observação.

| | | | | | | |
|--------------|------------|----|----|----|----|----|
| Exposição A: | Y: | 3 | 5 | 7 | 9 | 18 |
| | δ : | 1 | 1 | 1 | 0 | 1 |
| Exposição B: | Y: | 12 | 19 | 20 | 20 | 33 |
| | δ : | 1 | 1 | 1 | 0 | 0 |

(a) Calcule, a mão, o estimador de Kaplan-Meir (produto limite) da curva de sobrevivência $S(t)$ para cada um dos grupos de exposição. Calcule a variância de $\hat{S}(t)$ para cada observação não censurada.

(b) Faça um gráfico das duas curvas de sobrevivência colocando legendas para os dois grupos.

(c) Dê a estimativa do tempo mediano de sobrevivência para cada um dos grupos.

(d) Faça o teste de *logrank* para comparar os tempos de sobrevivência dos dois grupos.

(e) Faça o teste de Gehan generalizado de Wilcoxon para comparar os tempos de sobrevivência dos dois grupos.

(f) Faça o teste de Peto/Prentice generalizado de Wilcoxon para comparar os tempos de sobrevivência dos dois grupos.

(g) O que você concluiria sobre os dados?

13. Sejam X_1 e X_2 duas variáveis aleatórias independentes com distribuições Binomiais $B(n_1, p)$ e $B(n_2, p)$, respectivamente. Qual a distribuição de $Y = X_1 + X_2$? Mostre que:

$$X_1 | Y = m \sim H(N, m, n_1), \text{ isto é,}$$

$$\mathbb{P}(X_1 = x | Y = m) = \frac{\binom{m}{x} \binom{N-m}{n_1-x}}{\binom{N}{n_1}}.$$

14. Em um estudo para avaliar o nível de ruído e eficiência, foram dados a 18 estudantes um teste simples sob três diferentes níveis de ruído. Sabe-se que sob circunstâncias normais, eles seriam capazes de terminar o teste em 10 minutos. Os estudantes foram aleatoriamente assinalados para cada um dos três níveis de ruído. A tabela abaixo mostra o tempo requerido para terminar o teste para cada um dos estudantes para cada um dos níveis de ruído. Os níveis de ruído são significativamente diferentes? Se são, quais são os que diferem uns dos outros ?

| 1 | 2 | 3 |
|------|------|------|
| 10,5 | 10,0 | 12,0 |
| 9,0 | 12,0 | 13,0 |
| 9,5 | 12,5 | 15,5 |
| 9,0 | 11,0 | 14,0 |
| 8,5 | 12,0 | 12,5 |
| 10,0 | 10,5 | 15,0 |

15. Utilizando o SAS (ou algum outro pacote que possa resolver as questões abaixo) e os dados descritos nos arquivos *oc.dat* e *oc.txt*:

- (a) Faça um gráfico das curvas de sobrevivência de Kaplan-Meier para cada um dos dois grupos de tratamentos.
- (b) Faça o gráfico do $\log(-\log S(t))$ e $-\log(S(t))$ para cada um dos grupos de tratamentos.
- (c) Faça o teste de *logrank* para comparar os tempos de sobrevivência dos dois grupos.
- (d) Faça o teste de Peto/Prentice generalizado de Wilcoxon para comparar os tempos de sobrevivência dos dois grupos.
- (e) O que você concluiria sobre os dois tratamentos?

16. Num estudo feito para a comparação de tratamentos de pacientes com hipernefroma e, para relacionar a resposta e sobrevivência à cirurgia, metástase e tempo de tratamento foram coletados dados de 58 pacientes. As variáveis coletadas foram:

sexo

idade

nefrectomia (1=sim, 0=não)

tempo de nefrectomia (número de anos antes do tratamento)

tratamento (1=quimioterapia, 0=outros)

resposta (0=nenhuma, 1=completa, 2=parcial, 3=estável, 4=doença crescente, 9=desconhecida)

tempo de sobrevivência

status (1=morto, 0=vivo, 9=desconhecido)

metástase no pulmão (1=sim, 0=não)

metástase nos ossos (1=sim, 0=não)

Como você analisaria os dados para responder às seguintes questões?

- (a) Os pacientes que tiveram nefrectomia têm uma taxa de resposta maior do que os que não tiveram nefrectomia?
- (b) O tempo de nefrectomia é relacionado com resposta e com sobrevivência?
- (c) Existe diferença significativa entre os tratamentos?
- (d) Quais as variáveis mais importantes relacionadas com resposta e com sobrevivência?