Lista 1 de Análise de Sobrevivência

Profa. Agatha Rodrigues

Maio de 2025

1. Leia todos os exemplos do Capítulo 1 do livro do Colosimo e Giolo.

2. Defina e dê exemplos:

   (a) censura;

   (b) censura à esquerda;

   (c) censura intervalar;

   (d) censura à direita e os três tipos.

3. Um número grande de indivíduos foi acompanhado para estudar o aparecimento de um certo sistema. Os indivíduos foram incluídos ao longo do estudo e foi considerada como resposta de interesse a idade em que este sintoma apareceu pela primeira vez. Para os seis indivíduos selecionados e descritos a seguir, identifique o tipo de censura apresentado.

   (a) O primeiro indivíduo entrou no estudo com 25 anos já apresentando o sintoma.

   (b) Outros dois indivíduos entraram no estudo com 20 e 28 anos e não apresentaram o sintoma até o encerramento do estudo.

   (c) Outros dois indivíduos entraram com 35 e 40 anos e apresentaram o sintoma no segundo e no sexto exames, respectivamente, após terem entrado no estudo. Os exames foram realizados a cada dois anos.

   (d) O último indivíduo selecionado entrou no estudo com 36 anos e mudou da cidade depois de 4 anos sem ter apresentado o sintoma.

4. Mostre que $\lambda(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt}(\log S(t))$.

5. Mostre que $\Lambda(t) = \int_0^t \lambda(u)du = -\log S(t)$. Sugestão: utilize o exercício 4.

6. Seja T uma variável aleatória não negativa e contínua. Mostre que $E(T) = \int_0^\infty S(t)d(t)$.

7. Mostre que $vmr(t) = \frac{\int_t^\infty (u-t)f(u)du}{S(t)} = \frac{\int_t^\infty S(u)du}{S(t)}$. Sugestão: utilize integral por partes sabendo que $f(u)du = -dS(u)$.

8. Suponha que a taxa de falha da variável aleatória tempo de falha T seja expressa pela função linear $\lambda(t) = \beta_0 + \beta_1 t$, com $\beta_0 > 0$ e $\beta_1 \geq 0$. Obtenha $S(t)$ e $f(t)$.

9. Suponha que a vida média residual associada à variável T seja dada por $vmr(t) = t + 10$. Obtenha $E(T)$, $\lambda(t)$ e $S(t)$.

10. Para uma variável aleatória T discreta, mostre que:

    (a) $S(t) = \prod_{t_j < t} \frac{S(t_j)}{S(t_{j-1})}$.

    (b) $S(t) = \prod_{t_j < t} [1 - \lambda(t_j)]$.

11. Suponha que o tempo até o óbito de pacientes submetidos a transplante de rim (em dias) segue uma distribuição log-logística, com função de sobrevivência dada por

$$S(t) = \frac{1}{1 + \lambda t^{\alpha}}, t \geq 0.$$

    Se $\alpha = 1,5$ e $\lambda = 0,001$:

    (a) Encontre a probabilidade de sobrevivência nos dias 50, 100 e 150.

    (b) Encontre o tempo mediano de vida dos pacientes após o transplante.

    (c) Mostre que a função de taxa de falha é inicialmente crescente e depois descrescente com o tempo. Encontre o ponto em que a taxa de falha muda de crescente para decrescente.

    (d) Encontre o tempo médio de vida dos pacientes após o transplante (você pode consultar o livro do Klein e Moeschberger, pg. 38).

12. Suponha que a variável aleatória Y tenha distribuição normal com média μ e variância σ^2 e assuma que $T = e^Y$.

    (a) Obtenha uma expressão para a função de sobrevivência de T, ou seja, $P(T > t)$.

    (b) Calcule $\mathrm{E}(T^r)$, a média e a variância de T.

13. Considere a distribuição Weibull, com função de sobrevivência dada por

$$S(t) = \exp\{-\lambda t^{\rho}\}, t \geq 0.$$

    Utilizando qualquer *software* ou pacote estatístico de sua preferência, construa gráficos da função de taxa de falha da distribuição Weibull, variando-se os valores dos parâmetros λ e ρ. Considere 6 combinações diferentes de valores de λ e ρ: utilize dois valores diferentes de λ e três valores diferentes de ρ, sendo um deles necessariamente igual a 1 (ou seja, $\rho = 1$). Construa também gráficos das respectivas funções de sobrevivência.

14. Considere a distribuição Weibull do exercício anterior e escolha uma das combinações de λ e ρ utilizadas. Utilizando qualquer *software* ou pacote estatístico de sua preferência, gere dados com a distribuição Weibull escolhida, com tamanho amostral $n = 100$.

    `IMPORTANTE`: Escreva claramente qual foi o *software* ou pacote estatístico utilizado e inclua necessariamente os códigos utilizados para a resolução do exercício.

    (a) Obtenha um boxplot dos dados e um histograma (com a curva da densidade teórica também no gráfico). Obtenha também a curva de sobrevivência empírica dos dados e coloque num mesmo gráfico a curva empírica e a curva teórica utilizada para gerar os gráficos.
    (b) Faça um gráfico de quantis (*QQ Plot*) comparando os quantis empíricos com os quantis teóricos da distribuição Weibull utilizada para gerar os dados.
    (c) Padronize os dados (ou seja, subtraia a média amostral e divida pelo desvio padrão) e faça um gráfico de quantis (*QQ Plot*) comparando os quantis empíricos da variável padronizada com os da normal padrão. Discuta a adequabilidade da distribuição normal aos dados.
    (d) Repita os itens (b) e (c) para $n = 40$, $n = 300$ e $n = 1200$.

15. Considere que o tempo até o evento T segue uma distribuição exponencial com parâmetro $\lambda > 0$, e que o tempo até a censura C também segue uma distribuição exponencial com parâmetro $\theta > 0$, de forma independente. Seja $Y = \min(T, C)$ o tempo observado e $\delta = 1(T \leq C)$ a variável indicadora de falha.

    (a) Seja p_c a proporção de dados censurados. Escreva p_c como função de λ e θ.
    (b) Suponha que $\lambda = 1$ e que desejamos uma proporção de censura esperada de 30% (isto é, $p_c = 0{,}30$). Determine o valor de θ que atende a essa condição.
    (c) Simule amostras de tamanho $n \in \{100, 1000, 10000\}$ com $T \sim Exp(1)$ e $C \sim Exp(\theta)$, usando o valor de θ obtido no item b (para ter 30% de dados censurados). Lembre-se que os dados observados são $\{t_i, \delta_i,\ i = 1, \ldots, n\}$.
    (d) Calcule a proporção de censura empírica observada nas amostras geradas. Compare com o valor teórico.