

Lista 4

Profa. Agatha Rodrigues

Julho de 2025

1. Sejam T_1, \dots, T_n tempos de falha i.i.d. e seja C_i o tempo de censura da i -ésima observação, com $i = 1, \dots, n$, e considere $\delta_i = I(T_i < C_i)$. Além disso, considere o vetor de covariáveis para a i -ésima observação $\mathbf{x}_i = (1, x_{i1}, \dots, x_{ip})^\top$. Assuma censura independente e não informativa e, além disso, considere o modelo: $Y = \log T = \exp(\mathbf{x}^\top \beta) + \sigma \nu$, em que $\beta = (\beta_0, \dots, \beta_p)^\top$ e σ é o parâmetro de escala. Construa a função de verossimilhança e veja que não é possível encontrar o valor de $\theta = (\beta^\top, \sigma)^\top$ que maximiza a função de verossimilhança de forma fechada em cada um dos casos seguintes. Discuta também métodos numéricos para resolver o problema.
 - (a) Considere que ν segue distribuição valor extremo padrão.
 - (b) Considere que ν segue distribuição logística padrão.
 - (c) Considere que ν segue distribuição normal padrão.
2. Os dados no arquivo `HOD_NHL.csv` são referentes a 43 pacientes com linfoma de Hodgking ou linfoma não Hodgking, submetidos a transplante de medula óssea. Alguns pacientes receberam transplante de doador aparentado compatível (transplante alogênico) e outros receberam transplante autólogo (ou seja, a própria medula óssea do paciente é coletada e posteriormente infundida). Nos dados também há informação sobre o escore de Karnofsky, que é uma medida de performance que classifica os pacientes segundo o bem-estar dos pacientes (quanto menor a classificação na escala, pior a qualidade de vida). O objetivo principal do estudo é a comparação dos tipos de transplante, considerando-se o tempo (dias) livre de doença (i.e., tempo antes de ocorrência da recorrência ou óbito).
 - (a) Construa curvas de Kaplan-Meier para o tempo de sobrevivência dos pacientes com linfoma para cada um dos grupos de tratamento. Categorize a variável escore de Karnofsky, criando uma variável com duas categorias: escore < 80 ou escore ≥ 80 . Construa gráficos de Kaplan-Meier para essa nova variável categorizada. Comente.
 - (b) Ajuste um modelo de regressão Weibull com a variável tratamento e a variável escore de Karnofsky categorizada conforme o item (a). Apresente as estimativas de máxima verossimilhança dos parâmetros do modelo Weibull.

- (c) Encontre uma estimativa pontual para a razão de taxas de falha de pacientes que receberam transplante autólogo e alogênico. Faça o mesmo para a outra variável (escore de Karnofsky) incluída no modelo.
 - (d) Teste a hipótese de igualdade dos tipos de transplante e também das categorias do escore de Karnofsky, utilizando a estatística de Wald, com um nível de significância de 10%. Comente.
3. Considere os dados do exercício 1.
- (a) Refaça o item (b) utilizando a distribuição log-logística. Especifique claramente qual foi o modelo utilizado e quais foram os parâmetros estimados.
 - (b) A chance de sobrevivência após t é definida como

$$\frac{S(t|\mathbf{X})}{1 - S(t|\mathbf{X})}.$$

No modelo logístico, mostre que

$$\frac{S(t|\mathbf{X})}{1 - S(t|\mathbf{X})} = \exp[-\beta^T \mathbf{X}] \frac{S(t|\mathbf{X} = \mathbf{0})}{1 - S(t|\mathbf{X} = \mathbf{0})}.$$

- (c) Obtenha uma estimativa da razão de chances de sobrevivência após t de pacientes com transplante autólogo e pacientes com transplante alogênico. Interprete.
 - (d) Repita o item (d) do exercício 1, utilizando o modelo log-logístico. Compare os resultados e comente.
4. Considerando ainda os dados do exercício 1, faça gráficos apropriados da taxa de falha acumulada para verificar a adequabilidade dos modelos.
- (a) Weibull;
 - (b) Log-logístico.

Em todos os casos, utilize o estimador de Nelson-Aalen da função de taxa de falha acumulada considerando cada grupo separadamente (ou seja, obtenha estimativas da função de taxa de falha acumulada para cada grupo).

5. Considere ainda os dados do exercício 1. Obtenha os resíduos de Cox-Snell e *deviance* para os modelos de regressão Weibull (ajustado no exercício 1) e log-logístico (ajustado no exercício 2). Faça gráficos dos resíduos em função do tempo e comente. Com base em todas as análises feitas, discuta se os modelos (Weibull ou log-logístico) parecem ser adequados para os dados trabalhados.

6. Os dados do arquivo *viarapida.xls* referem-se a um estudo feito pelo Instituto do Coração, do Hospital das Clínicas - FMUSP, para comparar dois tipos de protocolo de recuperação: via-rápida e convencional. O protocolo via-rápida tem como filosofia uma maior integração entre as várias equipes que atendem os pacientes com a finalidade de tentar diminuir o tempo de internação, melhorando a recuperação e diminuindo custos. Considere, para as questões que seguem, apenas as seguintes variáveis:
- Delta - indicador de censura (0 - censura, 1 - saída do centro cirúrgico);
 - Idade - idade do paciente na internação;
 - Protocolo - sendo 0 convencional e 1 via rápida;
 - Raça - sendo 1 Branco, 2 negro e 3 amarelo;
 - Sexo - sendo 0 feminino e 1 masculino;
 - Tipac - sendo 0 congênito e 1 coronariano; e
 - T-CC - tempo entre a entrada e a saída do centro cirúrgico (horas).
- (a) Construa curvas de Kaplan-Meier para o T-CC com a finalidade de comparar, descritivamente, o sexo, a raça, o tipo de paciente (Tipac) e o protocolo com relação ao tempo de permanência no centro cirúrgico.
- (b) Construa testes estatísticos não paramétricos para averiguar a significância estatística dos fatores do item anterior. Interprete os resultados.
- (c) Considerando apenas o protocolo e métodos gráficos baseados na função de taxa de falha acumulada não paramétrica, procure averiguar qual dos modelos, exponencial, Weibull ou log-normal é mais adequado aos dados.
- (d) Compare o ajuste dos modelos exponencial, Weibull e log-normal utilizando os resíduos de Cox-Snell. Qual o modelo que você recomenda?
- (e) Com base no modelo obtido no item anterior, faça o teste paramétrico para comparar os protocolos com relação ao tempo de permanência no centro cirúrgico.