MAE514 - Introdução a Análise de Sobrevivência
1º Semestre/2020

3ª Lista de Exercícios

Informações Importantes

- Data de entrega: 10/06, quarta-feira
- Forma de entrega: pelo sistema e-Disciplinas

1. Em um estudo com pacientes com mieloma múltiplo, pacientes foram tratados com tramento padrão e os tempos de sobrevivência dos 25 pacientes, contados a partir do início do tratmanto, estão disponíveis na tabela a seguir (tempos censurados à direita são denotados por um sinal "+"). Suponha que deseja-se fazer um estudo para comparar o tratamento padrão com um novo tratamento. Esse novo estudo terá um período de acompanhamento total igual a 4 anos e meio e espera-se que a mortalidade inicial seja reduzida e, após um ano e meio de tratamento, ainda se tenha em torno de 65% de pacientes vivos.

Tempo de sobrevivência, em meses												
0,3	5,9	20,8	$28{,}0^+$	1,7	$73{,}6^+$	7,2	2,1	6,4	2,5	2,3	0,3	0,4
$65{,}4^+$	$64{,}9^+$	0,6	23,0	$42{,}6^+$	$48{,}0^+$	6,9	2,1	$43{,}6^+$	42,6	$12{,}0^+$	0,8	

(a) Obtenha o tamanho da amostra necessário se o recrutamento for feito por um período de 2 anos e, nos próximos dois anos e meio, for feito apenas o acompanhamento desses pacientes. Considere 5% e 8% de nível de significância do teste e poder igual a 80%, 85% e 90%.

(b) Repita o item (a) assumindo que o recrutamento poderá ser feito ao longo dos 4 anos e meio. Compare com os resultados do item (a).

2. Considere $a_1 < \ldots < a_J$ uma partição do eixo do tempo. A função de sobrevivência da distribuição Exponencial Segmentada (*Piecewise Exponential*) é dada por

$$S(t) = \exp\left\{-\sum_{j=1}^{J} \lambda_j \nabla_j(t)\right\}, \quad t > 0,$$

em que

$$\nabla_j(t) = \begin{cases} 0, & \text{se } t < a_{j-1} \\ t - a_{j-1}, & \text{se } a_{j-1} \leq t < a_j \\ a_j - a_{j-1}, & \text{se } t \geq a_j \end{cases}$$

para $j = 1, \ldots, J$. Defina $a_0 = 0$ e $a_{J+1} = \infty$.

(a) Mostre que a função de taxa de falha desta distribuição é constante no intervalo (a_j, a_{j+1}), para todo $j = 0, 1, \ldots, J$.

(b) Para uma amostra (censurada) de tamanho n desta distribuição, mostre que o estimador de máxima verossimilhança do vetor $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_J)$ é
dado por $\widehat{\boldsymbol{\lambda}} = (\hat{\lambda}_1, \ldots, \hat{\lambda}_J)$, com

$$\hat{\lambda}_j = \frac{d_j}{\sum_{i=1}^{n} \nabla_j(t_i)}, \quad j = 1, \ldots, J,$$

em que d_j representa o número de falhas no j-ésimo intervalo.

(c) Para introduzir covariáveis no modelo, considere a matriz de desenho $\mathbf{x}$ sem intercepto. Mostre que esta distribuição pertence à classe de modelos de riscos proporcionais em um modelo de regressão com $\lambda_j = \exp\{\beta_{0j} + \mathbf{x}^\top \beta\}$.

3. 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 perfomance que classifica os pacientes segundo o bem-estar dos pacientes. 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), utilizando algum pacote estatístico (R, Splus, SAS, etc.). Apresente as estimativas de máxima verossimilhança dos parâmetros considerando a representação do modelo Weibull como um modelo de locação-escala e também como um modelo de riscos proporcionais.

(c) Encontre uma estimativa pontual para a razão de taxas de falha de pacientes que receberam transplante autólogo e alogênico. Encontre também uma estimativa do fator de aceleração, deixando claro como foi calculado. 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.

4. Considere os dados do exercício 3.

(a) Refaça o item (b) utilizando a distribuição log-logística. Especifique claramente qual foi o modelo utilizado e quais foram os parâmetros estimados.

(b) Encontre uma estimativa pontual para ao fator de aceleração e interprete o resultado.

(c) A chance de sobrevivência após t é definida como

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

No modelo logístico, mostre que

$$\frac{S(t|\mathbf{x})}{1 - S(t|\mathbf{x})} = \exp\left[-\mathbf{x}^\top \boldsymbol{\beta}\right] \frac{S(t|\mathbf{x} = \mathbf{0})}{1 - S(t|\mathbf{x} = \mathbf{0})}.$$

(d) Obtenha uma estimativa da razão de chances de sobrevivência após t de pacientes com células anormais e pacientes com células normais. Interprete.

(e) Repita o item (d) do exercício 3, utilizando o modelo log-logístico. Compare os resultados e comente.

5. Considerando ainda os dados do exercício 3, 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).

6. Considere ainda os dados do exercício 3. Obtenha os resíduos de Cox-Snell e *deviance* para os modelos de regressão Weibull (ajustado no exercício 4) e log-logístico (ajustado no exercício 3). 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.