

MAE514 - Introdução a Análise de Sobrevida

Segunda Atividade Avaliativa

1º Semestre de 2020

Informações Importantes

- A resolução desta atividade deverá ser enviada via e-disciplinas até o dia 03/08 às 23:55.
- No dia 29/07 às 12:00 (meio-dia), a professora estará disponível para dúvidas no *google meet*, no mesmo link das aulas.

ATIVIDADE 1 (2,5 pontos)

Seja T uma variável aleatória que, condicionalmente a variável aleatória U , tem distribuição Weibull com parâmetros $\lambda = Ue^{\mathbf{x}^\top \boldsymbol{\beta}}$ e γ , ou seja, tem função de risco (ou taxa de falha) dada por

$$\alpha(t|U, \mathbf{x}) = Ue^{\mathbf{x}^\top \boldsymbol{\beta}} \gamma t^{\gamma-1},$$

em que \mathbf{x} é um vetor de covariáveis conhecidas (que inclui o intercepto), $\boldsymbol{\beta}$ e γ são parâmetros (desconhecidos). Assuma que a variável aleatória U tenha distribuição estável positiva, com função densidade de probabilidade dada por

$$f_U(u) = \frac{1}{\pi} \sum_{\kappa=1}^{\infty} (-1)^{\kappa+1} \frac{\Gamma(\kappa\theta + 1)}{\kappa!} u^{-\kappa\theta-1} \sin(\kappa\theta\pi), u \geq 0,$$

em que θ é o parâmetro da distribuição com $0 < \theta \leq 1$. É interessante observar que o caso $\theta = 1$ corresponde a uma variável aleatória degenerada no ponto $U = 1$. Essa distribuição tem a propriedade de apresentar todos os momentos infinitos e, conseqüentemente, o valor esperado é infinito e a variância não existe. Apesar da expressão da função densidade de probabilidade poder ser expressa apenas em termos de uma série, a transformada de Laplace tem uma forma bastante simples:

$$\mathcal{L}(s) = E[e^{-sU}] = e^{-s^\theta}. \quad (1)$$

- (a) Qual é a distribuição marginal de T (considerando-se dado o vetor de covariáveis \mathbf{x})?
- (b) Assuma agora que a variável aleatória T tenha distribuição condicional a U especificada pela seguinte função de risco:

$$\alpha(t|U, \mathbf{x}) = \alpha_o(t)Ue^{\mathbf{x}^\top \boldsymbol{\beta}},$$

em que $\alpha_o(t)$ é a função de risco basal. Obtenha a função de sobrevivência marginal de T (considerando-se conhecido o vetor \mathbf{x} de covariáveis) e também a função de risco.

- (c) Com base nos itens (a) e (b), discuta o efeito de ignorarmos a existência da variável de fragilidade U no modelo nas estimativas dos parâmetros associados com as covariáveis.
- (d) Assuma agora que temos um par de variáveis aleatórias (T_1, T_2) satisfazendo um modelo de fragilidade compartilhada, ou seja:

- T_1 e T_2 são condicionalmente independentes dada a variável U ;
- a distribuição de T_j condicionalmente a U é dada por

$$\alpha_j(t|U) = \alpha_o(t)U\lambda_j, j = 1, 2,$$

em que λ_1 e λ_2 são parâmetros desconhecidos (que podem eventualmente depender de covariáveis);

- U tem distribuição estável positiva, com transformada de Laplace dada por (1).

Obtenha a função de sobrevivência conjunta de (T_1, T_2) marginal (integrando-se com relação a U).

Instruções para Atividade 1

Para esta atividade, você deverá enviar dois arquivos:

- Resolução por escrito: o arquivo com a resolução escrita pode ser digitado em algum editor de texto ou você pode escrever manualmente a resolução e escanear (ou tirar foto). Dê preferência para arquivos no formato pdf para o envio. A clareza e redação do arquivo por escrito serão também avaliadas.
- **Arquivo de áudio:** você deve gravar um áudio em que você explica a sua resolução. O arquivo de áudio pode ser gravado no celular e enviado ou pode ser gravado no *google meet* (com imagens) e, neste caso, basta enviar o link da gravação. É importante ressaltar que o arquivo com a resolução escrita é necessário mesmo com as imagens do *google meet*.

ATIVIDADE 2 (3,5 pontos)

Para esta atividade, você deverá estudar o modelo de Cox estratificado. Você deverá gravar um vídeo em que você define o modelo, justifique e explique quando deve ser utilizado e obtenha a função de verossimilhança parcial para o modelo de Cox estratificado. Este vídeo pode ser feito com imagens de suas notas (ou slides), mas não com as imagens das notas de aula fornecidas pela professora. Você deverá criar as suas próprias anotações (à mão ou slides). Você pode usar qualquer recurso ou ferramenta que tenha disponível, mas a avaliação será pelo domínio do conteúdo explicado. Você deverá enviar o arquivo com o vídeo ou o link do *google meet*, caso a gravação tenha sido feita usando essa ferramenta.

Instruções para Atividade 2

Esta atividade envolve o estudo do **modelo de Cox estratificado**. Você deverá enviar:

- arquivo com vídeo ou
- link do *google meet*.

Não é necessário enviar o arquivo E o link, basta enviar um dos dois.

ATIVIDADE 3 (4,0 pontos)

Considere o conjunto de dados sobre câncer de mama, obtidos de um grupo de estudos de câncer de mama da Alemanha. Os dados são referentes a 686 mulheres diagnosticadas com câncer da mama entre julho de 1984 e dezembro de 1989. Algumas variáveis disponíveis são:

- Tempo (**survtime**): tempo de vida desde o diagnóstico até o óbito (em dias);
- Delta (**censdead**): indicadora de óbito ou censura à direita;
- Idade (**age**): idade na data do diagnóstico;
- Terapia hormonal (**hormone**): indicadora de terapia hormonal (1 = Sim / 0= Não);
- Estadiamento (**grade**): estadiamento do tumor (1, 2 e 3), que é uma classificação que indica a gravidade do tumor encontrado;
- Tamanho (**size**): tamanho do tumor primário, em milímetros;
- Linfonodos (**nodes**): número de linfonodos comprometidos;
- Monopausa (**menopause**): indicadora de ocorrência de menopausa no diagnóstico (1 = Sim / 0= Não).

Os dados estão disponíveis no arquivo **Dados-ex3-prova2.csv** e a descrição está no arquivo **Dados-ex3-prova2.txt**. Utilizando esses dados, responda os itens descritos a seguir:

- (a) Faça uma análise descritiva dos dados. Essa análise descritiva deve envolver curvas de Kaplan-Meier segundo as covariáveis descritas, bem como testes para comparação das curvas obtidas.
- (b) Ajuste o modelo de riscos proporcionais de Cox aos dados. Observe que as variáveis idade, linfonodos e tamanho podem ser consideradas como variáveis contínuas. Apresente os resultados do modelo completo, com todas as covariáveis incluídas. Faça um processo de seleção de variáveis e apresente o resultado do modelo final obtido. Você precisa descrever claramente o processo de seleção das variáveis adotado, mas deve apresentar apenas as estimativas e resultados de dois modelos: modelo completo e modelo final.
- (c) Interprete os parâmetros do modelo final obtido em (b).
- (d) Utilizando o estimador de Breslow para função de taxa de falha acumulada do modelo final em (b), obtenha gráficos da função de taxa de falha acumulada e da função de sobrevivência para pacientes nas seguintes situações:

- Idade de 53 anos, sem terapia hormonal, na menopausa, tumor de 15 milímetros, com estadiamento 1 e apenas 1 linfonodo comprometido;
- Idade de 53 anos, sem terapia hormonal, na menopausa, tumor de 30 milímetros, com estadiamento 1 e apenas 1 linfonodo comprometido;
- Idade de 53 anos, sem terapia hormonal, na menopausa, tumor de 30 milímetros, com estadiamento 1 e 5 linfonodos comprometidos;
- Idade de 53 anos, sem terapia hormonal, na menopausa, tumor de 30 milímetros, com estadiamento 2 e 5 linfonodos comprometidos;
- Idade de 53 anos, sem terapia hormonal, na menopausa, tumor de 30 milímetros, com estadiamento 3 e 5 linfonodos comprometidos.

Você deve fazer um gráfico para função de taxa de falha acumulada, comparando as 5 curvas referentes a cada paciente, e um gráfico com as 5 curvas para a função de sobrevivência.

- (e) Faça análise de resíduos do modelo final obtido em (b). Obtenha todos os resíduos discutidos em aula, bem como os testes de Schoenfeld para a proporcionalidade das taxas de falha.

Instruções para Atividade 3

Você deverá enviar um arquivo no formato pdf com a resolução da questão. Escreva em formato de relatório e **não** esqueça de acrescentar o código utilizado no final do arquivo (sua nota poderá ser penalizada caso o código não seja anexado). A redação e clareza do texto serão também avaliados.