

MAE514 - Introdução a Análise de Sobrevivência

Primeira Atividade Avaliativa

1º Semestre de 2020

Informações Importantes

- A resolução desta atividade deverá ser enviada via e-disciplinas até o dia 22/06 às 23:55.
- No dia 15/06 às 15:30, a professora estará disponível para dúvidas no *google meet*, no mesmo link das aulas.

ATIVIDADE 1 (2,5 pontos)

Algumas vezes, dados de sobrevivência são reduzidos a respostas binárias. Nestes casos, a informação utilizada é se o tempo de falha T é maior ou não do que t_0 fixado. Em geral, utiliza-se um instante t_0 com relevância clínica para aquele problema particular.

- (a) Considere um modelo de regressão Weibull (utilizando a parametrização de riscos proporcionais) e defina $p_0(\mathbf{x}) = P(T > t_0 | \mathbf{x})$, em que \mathbf{x} representa o vetor de covariáveis. Obtenha uma expressão para $p_0(\mathbf{x})$ em termos dos parâmetros do modelo Weibull utilizado. Defina um modelo de regressão adequado para a variável resposta binária, especificando a função de ligação apropriada para este caso.

Dica: Calcule $\log(-\log p_0(\mathbf{x}))$.

- (b) Repita o item (a) considerando um modelo de regressão log-logístico.
- (c) Assuma que os dados estão sujeitos a censura à direita tipo I, com o mesmo tempo de acompanhamento para todas as observações no estudo. Discuta em que situações é possível utilizar um modelo binário e em que situações o modelo binário não é adequado.

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.
- **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 a função de verossimilhança parcial de Cox. É preciso entender a derivação mais intuitiva (primeira derivação dada em aula, que consta nos livros adotados na disciplina), a obtenção da função de log-verossimilhança e cálculo do vetor escore. Você deverá me explicar o conteúdo estudado e, para isso, você deverá gravar um vídeo em que esses tópicos são explicados. Este vídeo pode ser feito com imagens de suas notas (ou slides), no formato em que as aulas estão sendo ministradas, ou com você explicando o conteúdo. 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 de um tópico dado no curso **a derivação intuitiva e informal da função de verossimilhança parcial de Cox, obtenção a função de log-verossimilhança parcial e cálculo do vetor escore**. Você deverá enviar:

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

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

ATIVIDADE 3 (4,0 pontos)

Um estudo foi conduzido em uma comunidade de Tartu, segunda maior cidade da Estônia, para avaliar a sobrevivência de pacientes que sofreram infarto no período de 1991 a 1993 (com seguimento até 1996). Os dados considerados são de 824 pacientes com 18 anos ou mais que tiveram infarto no período de 1991 a 1993 e algumas variáveis foram observadas:

- Tempo de vida, em meses, após o infarto (os dados estão sujeitos a censura à direita);
- Gênero (feminino ou masculino);
- Idade (em anos);
- Diagnóstico (Isquêmico/Hemorragia intracranial/Não identificado/hemorragia subaracnóide);
- Coma - variável binária (Não/Sim) indicando se o paciente entrou em coma após o infarto;
- Infarto prévio do miocárdio - variável binária (Não/Sim) indicando se o paciente tem histórico de infarto prévio do miocárdio.

Os dados estão disponíveis no arquivo **stroke-final.csv**, com as seguintes variáveis:

- **sex**: sexo do paciente
- **died**: data do óbito
- **dstr**: data o infarto
- **age**: idade na data do infarto
- **dgn**: diagnóstico
- **coma**: indicadora de coma após infarto
- **minf**: infarto prévio do miocárdio
- **obsmonths**: tempo, em meses, decorrido entre infarto e óbito ou censura (optou-se por imputar 0,1 para pacientes que morreram no mesmo dia do infarto)
- **dead**: indica ocorrência de óbito ou não.

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 um modelo Weibull aos dados. 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. Você pode apresentar os resultados do modelo na parametrização de locação-escala.
- (c) Interprete os parâmetros do modelo final obtido em (b).
- (d) Faça análise de resíduos do modelo final obtido em (b).
- (e) De forma semelhante ao item (b), ajuste um modelo log-logístico aos dados. Faça da mesma forma (porém utilizando a distribuição log-logística) e apresente os resultados do modelo completo e do modelo final.
- (f) Interprete os parâmetros do modelo final obtido em (e).
- (g) Faça análise de resíduos do modelo final obtido em (e).
- (h) Compare os ajustes e os gráficos de resíduos dos dois modelos finais obtidos (com a distribuição Weibull e log-logística). Escolha um dos modelos para apresentar ao pesquisador como modelo final e justifique sua resposta.

Instruções para Atividade 3

Você deverá enviar um arquivo no formato pdf com a resolução da questão. Escreva em formato de relatório e acrescente o código utilizado no final do arquivo. A redação e clareza do texto serão também avaliados.