# MAE514 - Introdução a Análise de Sobrevivência

**Segunda Prova**

$1^o$ Semestre de 2021

## Informações Importantes

- Esta prova está dividida em duas partes:

  - **PARTE 1:** deve ser feita em horário de aula e entregue no **dia 19/07**, via e-disciplinas exclusivamente;

  - **PARTE 2:** deve ser entregue no **dia 26/07**, via e-disciplinas exclusivamente.

- Ambas partes devem ser feitas **individualmente**, sem consulta a colegas de classe, amigos ou profissionais.
- Leia com atenção as instruções contidas em cada parte da prova.

# Segunda Prova: PARTE 1 (5,0 pontos)

**Instruções para a PARTE 1**

- Esta parte da prova deverá ser realizada em horário de aula.
- A entrega deverá ser via e-disciplinas, exclusivamente, no **dia 19/07**. - A entrega deverá ser realizada até às **21h00** do dia 19/07.
- Haverá uma tolerância de algumas horas na entrega da prova, porém não serão aceitas de forma nenhuma provas entregues após a tolerância. Você poderá entregar a prova até 02h00 do dia 20/07 (2 horas da madrugada, não da tarde). Caso você tenha dificuldade de conexão, por favor entre em contato com a professora *antes do fim do prazo de entrega* da prova.
- A prova deverá ser feita **à mão** e escaneada (você pode tirar fotos também) para envio para o e-disciplinas. Não serão aceitas provas da parte I feita em latex, Libreoffice, Word ou outro editor de texto. Por favor, escreva de forma legível e verifique se a versão escaneada está legível antes da entrega.

## QUESTÃO 1 (2,5 pontos)

Considere um estudo com pacientes com câncer de ovário, em estágio mais avançado. Todos os pacientes foram submetidos a um tratamento padrão e observou-se, para cada paciente, o tempo até sua morte, em dias. No momento do início do tratamento, os pacientes foram separados em dois grupos: pacientes com tumor grande e pacientes com tumor moderado. Deseja-se avaliar o efeito do tamanho do tumor na sobrevida dos pacientes.

(a) Crie uma variável binária adequada para indicação do grupo e escreva a forma do modelo semi-paramétrico de Cox.

(b) Assumindo que não há empates, escreva a forma da verossimilhança parcial de Cox para este problema.

(c) Utilizando a expressão obtida em (b), encontre o escore.

(d) Utilizando a expressão obtida em (b), encontre a expressão para a informação observada.

(e) Suponha que deseja-se testar a hipótese $\mathrm{H}_0 : \beta = 0$, em que $\beta$ é o parâmetro no modelo de Cox associado à covariável binária criada anteriormente. Para realizar este teste, pode-se utilizar a estatística do teste do escore. Escreva a expressão para a estatística do escore seguindo os seguintes passos:

  (a) Com base no item (c), obtenha $U(0)$, ou seja, a expressão para o escore avaliado no ponto $\beta = 0$. Simplifique a expressão obtida, escrevendo-a em função do número de indivíduos em risco em cada grupo.

  (b) Obtenha $I(0)$, ou seja, a expressão para a informação avaliada no ponto $\beta = 0$. Simplifique a expressão obtida, escrevendo-a em função do número de indivíduos em risco em cada grupo.

  (c) Obtenha a estatística do teste do escore, dada por

$$S = \frac{U^2(0)}{I(0)},$$

  que, sob $\mathrm{H}_0 : \beta = 0$, tem distribuição de qui-quadrado com 1 grau de liberdade.

## QUESTÃO 2 (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 $\gamma$, 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 $\gamma$ são parâmetros (desconhecidos). Assuma que a variável aleatória $U$ tenha distribuição estável positiva, com função densidade de probabilidade dada por

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

em que $\theta$ 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, consequentemente, 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:

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

(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:

$$\alpha(t|U, \mathbf{x}) = \alpha_\circ(t) U e^{\mathbf{x}^\top \boldsymbol{\beta}},$$

em que $\alpha_\circ(t)$ 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

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

  em que $\lambda_1$ e $\lambda_2$ 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 $U$).

# Segunda Prova: PARTE 2 (5,0 pontos)

**Instruções para a PARTE 2**

- A entrega deverá ser via e-disciplinas, exclusivamente, no **dia 26/07**.
- A entrega deverá ser realizada até às **23h59** do dia 26/07. Não haverá tolerância para atrasos.
- Para a atividade 1, você poderá enviar um arquivo gravado ou enviar o link da gravação feita utilizando o *google meet*.
- Para a atividade 2, você deverá entregar um arquivo com a resolução da questão (você pode usar o editor de texto de desua preferência) contendo os códigos utilizados.

## ATIVIDADE 1 (2,5 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 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.

## ATIVIDADE 2 (2,5 pontos)

O arquivo `FRTCS.dat` contém dados sobre um estudo realizado na França (French Three Cities Study), com o objetivo de estudar o efeito da pressão e uso de drogas para hipertensão na sobrevida de pessoas de uma determinada população. Os indivíduos na amostra foram acompanhados até óbito ou perda de acompanhamento, e os valores da pressão diastólica, sistólica e o uso de drogas anti-hipertensivas foram avaliados no momento de entrada no estudo e em mais dois outros instantes.

(a) Faça uma análise descritiva dos dados. Para as variáveis que variam no tempo, faça gráficos de Kaplan-Meier segundo os valores da variável no instante inicial.

(b) Organize os dados de maneira apropriada para a ajustar o modelo semiparamétrico de Cox (não esqueça de incluir os códigos utilizados no final da prova) com variáveis dependentes do tempo. Ajuste o modelo semiparamétrico de Cox, incluindo como variáveis explicativas: sexo, idade, pressão diastólica, sistólica e o uso de drogas anti-hipertensivas.

(c) Selecione as variáveis explicativas significativas (deixe claro qual foi a metodologia utilizada) e apresente o modelo final ajustado.

(d) Interprete o modelo ajustado no item anterior e discuta os resultados.

**Importante:** Em todos os itens, os resultados apresentados devem ser interpretados. A redação será também avaliada. Não se esqueça de apresentar códigos dos programas utilizados.