

MAE514 - Introdução a Análise de Sobrevivência

1º Semestre/2020

2ª Lista de Exercícios

Informações Importantes

- Data de entrega: a definir
- Forma de entrega: pelo sistema e-Disciplinas, em data a ser definida

1. Sejam T_1, T_2, \dots, T_n tempos de falha sujeitos a censura à direita, de forma que se observa $Z_i = \min(T_i, C_i)$ e $\delta_i = I(T_i \leq C_i)$, em que C_i são os tempos de censura, $i = 1, 2, \dots, n$. Sejam $t_1 < t_2 < \dots < t_D$ os instantes em que alguma falha foi observada e defina n_j como sendo o número de indivíduos em risco em t_j (ou seja, indivíduos que não falharam e não foram censurados até o instante imediatamente anterior a t_j) e d_j o número de falhas observadas em t_j . O estimador de Kaplan-Meier da função de sobrevivência associada aos tempos de falha é dado por

$$\hat{S}(t) = \begin{cases} 1, & \text{se } t < t_1 \\ \prod_{t_j \leq t} \left(1 - \frac{d_j}{n_j}\right), & \text{se } t_1 \leq t \end{cases}.$$

A variância de $\hat{S}(t)$ pode ser estimada pela fórmula de Greenwood, dada por

$$\widehat{\text{Var}}(\hat{S}(t)) = [\hat{S}(t)]^2 \sum_{j:t_j \leq t} \left(\frac{d_j}{(n_j - d_j)n_j} \right).$$

- (a) Mostre que o estimador de Kaplan-Meier se reduz à função de sobrevivência empírica se não há censuras, ou seja,

$$\hat{S}(t) = \frac{\text{nº obs. } > t}{n}.$$

- (b) Mostre que a fórmula de Greenwood se reduz à estimativa da variância de uma proporção, ou seja,

$$\widehat{\text{Var}}\left(\hat{S}(t)\right) = n^{-1}\hat{S}(t)\left(1 - \hat{S}(t)\right).$$

2. Considere um estudo sobre AZT, medicamento utilizado para tratar pacientes com HIV. Os dados são de 45 pacientes que foram acompanhados desde sua entrada no estudo até a morte. Os dados contêm informação sobre a idade do paciente quando entrou no estudo e a idade que tinha quando faleceu. Os dados estão disponíveis no arquivo `Lista2-HIV.txt` e a descrição dos dados está em `Lista2-HIV.des`.

- (a) Com base nos dados apresentados, calcule a variável **tempo**: número de meses entre a entrada no estudo e óbito (ou tempo de censura).
- (b) Calcule o estimador da tabela de vida, considerando as seguintes faixas de tempo:

Faixa 1:	60 meses ou menos
Faixa 2:	de 60 (exclusive) a 120 meses
Faixa 3:	de 120 (exclusive) a 240 meses
Faixa 4:	de 240 (exclusive) a 360 meses
Faixa 5:	de 360 (exclusive) a 480 meses
Faixa 6:	mais de 480 meses

Apresente o estimador de duas formas: através de tabelas e gráfico. O que pode ser dito?

- (c) Calcule o estimador Kaplan-Meier para os dados (você pode utilizar um *software*).
- (d) Coloque em um mesmo gráfico as duas curvas estimadas nos itens anteriores. Compare as curvas e comente.
3. Os dados deste exercício referem-se a um estudo em pacientes com leucemia. Os dados são referentes a tempos de remissão (período em que o paciente está sem tratamento e sem a doença, ou seja, período compreendido entre o fim do tratamento e a reincidência da leucemia). Os pacientes foram submetidos a dois diferentes tratamentos e os tempos, em dias, de remissão estão

apresentados na tabela abaixo. Os tempos censurados à direita são denotados por um sinal “+”.

Tratam.	Tempo de remissão								
1	5	5	9	10	12	12	10	23	28
	28	28	29	32	32	37	41	41	57
	62	74	100	139	20 ⁺	258 ⁺	269 ⁺		
2	8	10	10	12	14	20	48	70	75
	99	103	162	169	195	220	161 ⁺	199 ⁺	217 ⁺
	245 ⁺								

Usando esses dados, calcule **à mão** o seguinte:

- O estimador Kaplan-Meier para cada tratamento.
 - Estimativas pontuais e intervalares (use coeficiente de 90%) para a mediana de-- cada tratamento.
 - O tempo médio de sobrevivência para cada tratamento.
 - Faça (à mão) um gráfico com as funções de sobrevivência estimadas (para cada tratamento).
 - O estimador de Nelson-Aalen. Faça um gráfico com as duas curvas estimadas.
 - Utilizando a estatística de *log-rank*, teste a igualdade dos tratamentos. Apresente os cálculos realizados para a obtenção da estatística em forma de uma tabela (você pode utilizar alguma planilha eletrônica para os cálculos, mas todos os passos devem estar bem explicados).
4. Os dados mostrados a seguir representam o tempo até a ruptura de um tipo de isolante elétrico sujeito a uma tensão de estresse de 35 Kvolts. O teste consistiu em deixar 25 destes isolantes funcionando até que 15 deles falhassem (censura tipo II), obtendo-se os seguintes resultados (em minutos):

0,19	0,78	0,96	1,31	2,78	3,16	4,67	4,85
6,50	7,35	8,27	12,07	32,52	33,91	36,71	

Observe que 10 observações foram censuradas. Para este exercício, os cálculos podem ser feitos *à mão* ou com auxílio computacional, porém a ideia é **não** utilizar uma função pronta que calcule o que for pedido. Você deve usar uma planilha ou escrever o código que façam as contas no R ou outro *software* de sua preferência. A partir desses dados amostrais, deseja-se obter:

- a função de sobrevivência estimada por Kaplan-Meier;

- (b) uma estimativa para o tempo mediano de vida deste tipo de isolante elétrico funcionando a essa tensão;
 - (c) uma estimativa (pontual e intervalar) para a fração de defeituosos esperada nos dois primeiros minutos de funcionamento;
 - (d) o tempo necessário para 20% dos isolantes estarem fora de operação.
5. Os dados disponíveis no arquivo `Lista2-hodgkins.xlsx` são referentes 60 pacientes com doença de Hodgkins que receberam tratamento padrão para a doença. O tempo de vida (em meses), bem como idade, sexo, histologia e estágio da doença de cada paciente foi observado. Em todos os itens a seguir, apresente os resultados em forma de relatório (você pode utilizar o *software* de sua preferência), explicando e interpretando os resultados. Acrescente o código do programa utilizado no final, como um apêndice.
- (a) Construa, no mesmo gráfico, as curvas de Kaplan Meier para pacientes do sexo masculino e feminino. Teste a igualdade das curvas de sobrevivência (obtenha duas estatísticas de teste, sendo uma delas a de *log-rank*). Comente.
 - (b) Divida os pacientes em quatro grupos etários: menos de 25 anos; de 25 anos (inclusive) até menos de 38 anos; de 38 anos (inclusive) até 53 anos; 53 anos ou mais. Obtenha as curvas de Kaplan Meier e teste a igualdade das curvas de sobrevivência (obtenha duas estatísticas de teste, sendo uma delas a de *log-rank*). Comente.
 - (c) Repita o item (a) para as variáveis estágio da doença e histologia. Comente os resultados.
6. Considere os dados do arquivo `pharmacoSmoking.csv` com 125 pacientes e 14 variáveis. Esse arquivo está disponível na biblioteca `asaur` do R, cuja documentação está disponibilizada. A descrição dos dados está na documentação e um dos principais objetivos do estudo era comparar o tempo até o fumante voltar a fumar após o início de um dentre dois diferentes tratamentos. Utilizando o *software* de sua preferência:
- (a) Obtenha as curvas de Kaplan-Meier dos dois tratamentos;
 - (b) Compare os tratamentos utilizando o teste de log-rank e também utilizando diferentes ponderações (escolha pelo menos três diferentes);
 - (c) Compare as curvas de Kaplan-Meier utilizando o teste de log-rank, estratificado por situação de trabalho (variável `employment`) e discuta os resultados.
- Observação:** Consulte o livro do Klein e Moeschberger, página 219, para mais detalhes sobre testes estratificados.