

Introdução a Análise de Sobrevida

1º Semestre/2025

2ª Lista de Exercícios

1. Considere a transformação $\widehat{U}(t) = \log[-\log(\widehat{S}(t))]$, em que $\widehat{S}(t)$ é o estimador de Kaplan-Meier de $S(t)$, apresentada na aula sobre o estimador de Kaplan-Meier, slide 12.
 - (a) Demonstre a expressão da $\widehat{Var}[\widehat{U}(t)]$.
 - (b) Mostre como obter a expressão do intervalo de confiança para $S(t)$ a partir da transformação log-log.
2. 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

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

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

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

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

$$\widehat{S}(t) = \frac{1}{n} \sum_{i=1}^n I(T_i > t).$$

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

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

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 o seguinte:

- O estimador Kaplan-Meier para cada tratamento.
 - Estimativas pontuais e intervalares (use coeficiente de 90%) para a mediana para cada tratamento.
 - O tempo médio de sobrevivência para cada tratamento.
 - Faça 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.
4. Um estudo com pacientes diagnosticados com câncer de língua foi conduzido para verificar o efeito prognóstico de alterações no número de cromossomos (aneuploidia) no DNA das células do tumor. Tecidos do tumor dos pacientes foram analisados segundo uma determinada técnica para determinar se o tumor continha células com aneuploidia (anormal) ou diploidia (normal). Os dados estão disponíveis no arquivo `Lista2-cancerLingua.txt` e a descrição dos dados no arquivo `Lista2-cancerLingua.des`. Para este exercício, considere apenas os pacientes com **aneuploidia**.

- (a) Obtenha o estimador de Kaplan-Meier dos dados dos pacientes considerados (com aneuploidia). Apresente o gráfico do estimador. Obtenha o valor estimado da função de sobrevivência no instante $t = 5$ anos (ou 60 meses) e seu respectivo erro padrão.
 - (b) Obtenha a função de taxa de falha acumulada estimada no instante igual a 5 anos. Encontre também o erro padrão de $\hat{\Lambda}(60 \text{ meses})$. Estime $S(60 \text{ meses})$ usando $\tilde{S}(t) = \exp\{-\hat{\Lambda}(t)\}$, com $t = 60$ meses. Compare com a estimativa obtida em (a).
 - (c) Obtenha um intervalo de confiança para $S(60 \text{ meses})$.
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) Encontre um ponto de corte para a variável *idade* utilizando o pacote *maxstat* do R (ou equivalente em outra linguagem). Explique como esse ponto de corte é encontrado. 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) 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.
 - (d) Repita o item (a) para as variáveis estágio da doença e histologia. Comente os resultados.