Análise de Sobrevivência

PROFA. SUELY RUIZ GIOLO
Departamento de Estatística - UFPR

Exercícios - Capítulo 1

1. Suponha que seis ratos foram expostos a um material cancerígeno. Os tempos até o desenvolvimento do tumor de um determinado tamanho são registrados para os ratos. Os ratos A, B e C desenvolveram os tumores em 10, 15 e 25 semanas, respectivamente. O rato D morreu acidentalmente sem tumor na vigésima semana de observação. O estudo terminou com 30 semanas sem os ratos E e F apresentarem tumor.

   (a) Defina cuidadosamente a resposta do estudo.

   (b) Identifique o tipo de resposta (falha ou censura) observado para cada um dos ratos no estudo.

2. Um número grande de indivíduos foi acompanhado para estudar o aparecimento de um certo sintoma. Os indivíduos foram incluídos ao longo do estudo e foi considerada como resposta de interesse a idade em que este sintoma apareceu pela primeira vez. Para os seis indivíduos selecionados e descritos a seguir, identifique o tipo de censura apresentado.

   (a) O primeiro indivíduo entrou no estudo com 25 anos já apresentando o sintoma.

   (b) Outros dois indivíduos entraram no estudo com 20 e 28 anos e não apresentaram o sintoma até o encerramento do estudo.

   (c) Outros dois indivíduos entraram com 35 e 40 anos e apresentaram o sintoma no segundo e no sexto exames, respectivamente, após terem entrado no estudo. Os exames foram realizados a cada dois anos.

   (d) O último indivíduo selecionado entrou no estudo com 36 anos e mudou da cidade depois de 4 anos sem ter apresentado o sintoma.

3. Mostre que $\lambda(t) = \dfrac{f(t)}{S(t)} = -\dfrac{d}{dt}\Big(\log S(t)\Big)$.

4. Mostre que $\Lambda(t) = \int_0^t \lambda(u)du = -\log S(t)$.

5. Mostre que $\text{vmr}(t) = \dfrac{\int_t^{\infty}(u-t)f(u)du}{S(t)} = \dfrac{\int_t^{\infty} S(u)du}{S(t)}$.
$\Big($Sugestão: utilize uma integral por partes sabendo que $f(u)du = -\frac{d}{du}S(u)\Big)$.

6. Suponha que a taxa de falha da variável aleatória tempo de falha T seja expressa pela função linear $\lambda(t) = \beta_0 + \beta_1 t$, com β_0 e $\beta_1 > 0$. Obtenha $S(t)$ e $f(t)$.

7. Suponha que a vida média residual de T seja dada por $\text{vmr}(t) = t+10$. Obtenha $E(T)$, $\lambda(t)$ e $S(t)$.

8. Em cada um dos exemplos descritos na Seção 1.5, identifique o tempo inicial, a escala de medida e o evento de interesse.

Exercícios - Capítulo 2

1. Mostre que a partir da transformação $U(t) = \log[-\log S(t)]$ obtém-se o intervalo de 95% de confiança para $S(t)$ mostrado em (2.8).

2. 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 do 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	

A partir desses dados amostrais, deseja-se obter:

(a) uma estimativa para o tempo mediano de vida deste tipo de isolante elétrico funcionando a 35 Kvolts;

(b) uma estimativa (por ponto e por intervalo) para a fração de defeituosos esperada nos dois primeiros minutos de funcionamento;

(c) uma estimativa (por ponto) para o tempo médio de vida destes isolantes funcionando a 35 Kvolts (limitado em 40 minutos) e;

(d) o tempo necessário para 20% dos isolantes estarem fora de operação.

3. Os dados da Tabela 1 referem-se aos tempos de sobrevivência (em dias) de pacientes com câncer submetidos à radioterapia (o símbolo + indica censura).

Tabela 1: Tempos de pacientes submetidos à radioterapia.

7, 34, 42, 63, 64, 74^+, 83, 84, 91, 108, 112, 129, 133, 133, 139, 140, 140, 146, 149, 154, 157, 160, 160, 165, 173, 176, 185^+, 218, 225, 241, 248, 273, 277, 279^+, 297, 319^+, 405, 417, 420, 440, 523, 523^+, 583, 594, 1101, 1116^+, 1146, 1226^+, 1349^+, 1412^+, 1417

Fonte: Louzada Neto et al. (2002)

Para estes dados, obtenha estimativas para:

(a) a função de sobrevivência por meio dos estimadores de Kaplan-Meier e de Nelson-Aalen. Apresente-as em tabelas e gráficos;

(b) os tempos mediano e médio;

(c) as probabilidades de um paciente com câncer sobreviver a:
i) 42 dias, ii) 100 dias, iii) 300 dias e iv) 1000 dias;

(d) o tempo médio de vida restante dos pacientes que sobreviverem 1000 dias;

(e) interprete as estimativas obtidas nos três itens anteriores.

(f) para quais tempos tem-se: i) $\widehat{S}(t) = 0,80$, ii) $\widehat{S}(t) = 0,30$ e $\widehat{S}(t) = 0,10$? Interprete.

4. Os dados apresentados na Tabela 2 representam o tempo (em dias) até a morte de pacientes com câncer de ovário tratados na Mayo Clinic (Fleming et al., 1980). O símbolo + indica censura.

Tabela 2: Tempos dos pacientes no estudo de câncer de ovário.

Amostras	Tempos de sobrevivência em dias
1. Tumor Grande	28, 89, 175, 195, 309, 377+, 393+, 421+, 447+, 462, 709+, 744+, 770+, 1106+, 1206+
2. Tumor Pequeno	34, 88, 137, 199, 280, 291, 299+, 300+, 309, 351, 358, 369, 369, 370, 375, 382, 392, 429+, 451, 1119+

(a) Obtenha as estimativas de Kaplan-Meier para as funções de sobrevivência de ambos os grupos e apresente-as no mesmo gráfico.

(b) Repita a letra (a) utilizando, agora, o estimador de Nelson-Aalen.

(c) Usando os intervalos de confiança assintóticos das estimativas de Kaplan-Meier, teste a hipótese de igualdade das funções de sobrevivência dos dois grupos em $t = 6$ meses e $t = 15$ meses.

(d) Teste a hipótese de igualdade das funções de sobrevivência dos dois grupos usando dois testes diferentes. Os resultados dos testes são consistentes? Em caso negativo, explique a razão da diferença dos resultados.

5. Um estudo de sobrevivência foi realizado para comparar dois métodos para a realização de transplante de medula em pacientes com leucemia. A resposta de interesse era o tempo contado a partir do transplante até a morte do paciente.

   (a) Os seguintes resultados foram obtidos:

$$\sum_{j=1}^{k}(d_{2j} - w_{2j}) = 3,964 \qquad \text{e} \qquad \sum_{j=1}^{k}(V_j)_2 = 6,211.$$

   Estabeleça as hipóteses, obtenha o teste *logrank* e conclua. Use o nível de significância de 5% (3,84).

   (b) Neste estudo os pesquisadores não têm interesse em detectar diferenças entre os métodos nos tempos iniciais devido à toxicidade dos medicamentos. Você usaria o teste *logrank* ou de Wilcoxon nesta situação? Justifique sua resposta.

6. Um produtor de requeijão deseja comparar dois tipos de embalagens (A e B) para o seu produto. Ele deseja saber se existe diferença na durabilidade de seu produto com relação às embalagens. O produto dele é vendido a temperatura ambiente e sem conservantes. O evento de interesse é o aparecimento de algum tipo de fungo no produto. Os dados estão apresentados na Tabela 3, em que o tempo foi medido em horas. O símbolo + indica censura.

Tabela 3: Tempos dos requeijões no estudo das embalagens.

Embalagens	Tempos de sobrevivência em horas
A	31, 40, 43, 44, 46, 46, 47, 48, 48, 49, 50, 50, 60, 60, 60, 60, 60+, 60+, 60+, 60+
B	48, 48, 49, 49, 49, 49, 50, 50, 50, 50, 53, 53, 54, 54, 54, 55, 55+, 55+, 55+, 55+

(a) Existe diferença entre as duas embalagens?

(b) Caracterize a durabilidade do produto (percentil 10 e tempo médio de vida) para cada embalagem, se houver diferença entre elas. Caso contrário, faça o mesmo, mas combinando todos os tempos de vida.

7. Vinte e oito cães com leishmaniose foram selecionados para comparar quatro diferentes tipos de tratamentos (A, B, C e D) e um grupo controle. O evento de interese foi a morte do animal. Os dados estão apresentados na Tabela 4 em que o tempo foi medido em meses. O símbolo + indica censura.

Tabela 4: Tempos dos cães no estudo de leishmaniose.

Grupos	Tempos de sobrevivência em meses
A	1, 4, 5, 6, 7+, 7+
B	3, 5, 5+, 5+, 6
C	1, 3, 4+, 7, 7, 7+
D	3, 5+, 7, 7+, 7+
Controle	3, 5, 5, 7+, 7+, 7+

Há diferenças entre os grupos?

Exercícios - Capítulo 3

1. O tempo em dias para o desenvolvimento de tumor em ratos expostos a uma substância cancerígena segue uma distribuição de Weibull tal que:

$$S(t) = \exp\left\{ -\left(\frac{t}{\alpha}\right)^{\gamma}\right\},$$

com $\alpha = 100$ e $\gamma = 2$.

(a) Qual é a probabilidade de um rato sobreviver sem tumor aos primeiros 30 dias? E aos primeiros 45 dias?

(b) Qual é o tempo médio até o aparecimento do tumor?

(c) Qual é o tempo mediano até o aparecimento do tumor?

(d) Encontre a taxa de falha de aparecimento de tumor aos 30, 45 e 60 dias. Interprete estes valores.

2. Deseja-se comparar duas populações de tempos de vida. Uma amostra de tamanho n ($r \leq n$ falhas) foi obtida da população 1 que tem distribuição exponencial com média α. Uma amostra de tamanho m ($s \leq m$ falhas) foi obtida da população 2 que tem distribuição exponencial com média $\alpha + \Delta$.

   (a) Estabeleça as hipóteses que se deseja testar.

   (b) Apresente a função de verossimilhança para $\theta = (\alpha, \Delta)'$.

   (c) Apresente o vetor escore ($U(\theta)$) e a matriz de informação observada ($-\mathcal{F}(\theta)$).

   (d) Obtenha as expressões dos testes de Wald e da razão de verossimilhanças para as hipóteses apresentadas em (a).

3. 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 do 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	

   Este exercício foi proposto no Capítulo 2 para ser resolvido utilizando-se métodos não-paramétricos. O que se deseja aqui é que o exercício seja repetido utilizando-se modelos paramétricos. Inicialmente, deve-se identificar um modelo paramétrico para explicar estes dados e, em seguida, responder novamente às mesmas perguntas. Isto é, a partir destes dados amostrais, deseja-se obter as seguintes informações:

   (a) Uma estimativa para o tempo mediano de vida deste tipo de isolante elétrico funcionando a 35 Kvolts.

   (b) Uma estimativa (por ponto e por intervalo) para a fração de defeituosos esperada nos dois primeiros minutos de funcionamento.

   (c) Uma estimativa (por ponto e por intervalo) para o tempo médio de vida destes isoladores funcionando a 35 Kvolts.

   (d) O tempo necessário para 20% dos isolantes estarem fora de operação.

4. O fabricante de um tipo de isolador elétrico quer conhecer o comportamento de seu produto funcionando a uma temperatura de 200^{o}C. Um teste de vida foi realizado nestas condições usando-se 60 isoladores elétricos. O teste terminou quando 45 deles havia falhado (censura do tipo II). As 15 unidades que não haviam falhado ao final do teste foram, desta forma, censuradas no tempo $t =$ 2729 horas. O fabricante tem interesse em estimar o tempo médio e mediano de vida do isolador e o percentual de falhas após 500 horas de uso. Os tempos (em horas) obtidos são apresentados na Tabela 3.5.

   Responda às questões de interesse do fabricante fazendo uso do modelo paramétrico que se apresentar mais apropriado para descrever os dados.

   Tabela 5: Tempos (horas) dos isolantes elétricos funcionando a 200^{o}C.

151	164	336	365	403	454	455	473	538	577	592	628	632	647	675	727	785
801	811	816	867	893	930	937	976	1008	1040	1051	1060	1183	1329	1334		
1379	1380	1633	1769	1827	1831	1849	2016	2282	2415	2430	2686	2729				
2729^+	2729^+	2729^+	2729^+	2729^+	2729^+	2729^+	2729^+	2729^+	2729^+							
2729^+	2729^+	2729^+	2729^+	2729^+												

5. Ajuste um modelo paramétrico aos dados do Exercício 3 do Capítulo 2. Compare os resultados com aqueles obtidos no Capítulo 2.

Exercícios - Capítulo 4

1. Os dados apresentados na Tabela 6 referem-se aos tempos de sobrevivência, em meses, de dois grupos de pacientes com a mesma doença que foram submetidos a um de dois tratamentos alternativos (A ou B). (+ indica censura)

Tabela 6: Tempos de pacientes submetidos aos tratamentos A ou B.

Tratamentos	Tempos de sobrevivência (em meses)
A	1 2 2 2 2^+ 6 8 8 9 9^+ 13 13^+ 16 17 22^+ 25^+ 29 34 36 43^+ 45^+
B	1 2 5 7 7^+ 11^+ 12 19 22 30 35^+ 39 42 46 55

Considerando a covariável X = tratamento recebido em que:

$$X = \begin{cases} 0 & \text{se tratamento A} \\ 1 & \text{se tratamento B,} \end{cases}$$

(a) Ajuste os modelos de regressão exponencial, Weibull e log-normal e verifique qual é o mais adequado para esses dados.

(b) Utilizando o modelo escolhido no item anterior, use o teste da razão de verossimilhanças para testar se os tratamentos A e B diferem.

(c) Para o modelo final, apresente graficamente a análise dos resíduos e a(s) curva(s) de sobrevivência estimada(s).

(d) Utilizando o modelo ajustado, obtenha e interprete a sobrevivência estimada em $t = 40$ meses.

2. Ajuste um modelo de regressão paramétrico aos dados do Exercício 6 do Capítulo 2.

3. Ajuste um modelo de regressão paramétrico aos dados do Exercício 7 do Capítulo 2. Compare os resultados com aqueles obtidos no Capítulo 2.

Exercícios - Capítulo 5

1. Os dados a seguir representam o tempo (em dias) até a morte de pacientes com câncer de ovário tratados na Mayo Clinic (Fleming et al., 1980). O símbolo + indica censura.

   Amostra 1 (tumor grande): 28, 89, 175, 195, 309, 377+, 393+, 421+, 447+, 462, 709+, 744+, 770+, 1106+, 1206+

   Amostra 2 (tumor pequeno): 34, 88, 137, 199, 280, 291, 299+, 300+, 309, 351, 358, 369, 369, 370, 375, 382, 392, 429+, 451, 1119+.

   (a) Escreva o modelo de Cox para esses dados.

   (b) Escreva a função de verossimilhança parcial.

   (c) Ajuste o modelo de Cox e construa um intervalo de confiança para o parâmetro do modelo.

   (d) Teste a hipótese de igualdade dos dois grupos. Caso exista diferença entre os grupos, interprete o coeficiente estimado.

   (e) Sabendo-se que o teste *logrank* **coincide** com o teste escore associado ao modelo de Cox, use este teste para testar a hipótese estabelecida em (d).

2. Um estudo realizado para comparar dois tratamentos pós-cirúrgicos de câncer de ovário, envolveu o acompanhamento de 26 mulheres após a cirurgia de remoção do tumor. A resposta foi o tempo (em dias), contado a partir do início do tratamento (aleatorização) até a morte do paciente. As seguintes covariáveis foram registradas: X_1 = tratamento, X_2 = idade, X_3 = resíduo (1 se o resíduo da doença foi parcialmente removido e 2 se foi completamente removido) e X_4 = status (1 se a condição do doente no início do estudo era boa e 2 se ruim). Os dados encontram-se na Tabela 7.

   a) Encontre o modelo de Cox que melhor se ajuste a esses dados.

   b) Use uma das técnicas de adequação apresentadas para verificar a suposição de taxas de falha proporcionais.

   c) Caso a suposição de proporcionalidade seja válida, utilize o modelo ajustado no item (a) para verificar se existe diferença entre os tratamentos.

   d) Estime a probabilidade de uma paciente com 45 anos, resíduo = 1 e status = 2, sobreviver aos primeiros dois anos após o uso do tratamento 2.

3. Utilizando o modelo de Cox, reanalise o exercício 7 do Capítulo 2.

Tabela 7: Conjunto de dados referente ao Exercício 2.

Paciente	tempo	ind. falha	tratamento	idade	resíduo	status
1	156	1	1	66	2	2
2	1040	0	1	38	2	2
3	59	1	1	72	2	1
4	421	0	2	53	2	1
5	329	1	1	43	2	1
6	769	0	2	59	2	2
7	365	1	2	64	2	1
8	770	0	2	57	2	1
9	1227	0	2	59	1	2
10	268	1	1	74	2	2
11	475	1	2	59	2	2
12	1129	0	2	53	1	1
13	464	1	2	56	2	2
14	1206	0	2	44	2	1
15	638	1	1	56	1	2
16	563	1	2	55	1	2
17	1106	0	1	44	1	1
18	431	1	1	50	2	1
19	855	0	1	43	1	2
20	803	0	1	39	1	1
21	115	1	1	74	2	1
22	744	0	2	50	1	1
23	477	0	1	64	2	1
24	448	0	1	56	1	2
25	353	1	2	63	1	2
26	377	0	2	58	1	1

Exercícios - Capítulo 6

1. A covariável altura inicial foi dicotomizada na mediana para realizar a análise apresentada na Seção 6.6. Dicotomize esta covariável no primeiro quartil e refaça a análise.

2. Um teste alternativo ao de proporcionalidade das taxas de falha baseado nos resíduos padronizados de Schoenfeld é aquele devido a Cox (1979) e apresentado na Seção 5.6.2 do Capítulo 5. Este teste introduz uma covariável dependente do tempo no modelo. Faça este teste para verificar a existência de taxas de falha proporcionais nos dados de leucemia pediátrica devido à covariável leucócitos iniciais (LEUINI).

3. Repita o teste do Exercício 2 para verificar a suposição de taxas de falha proporcionais nos dados do hormônio de crescimento devido à altura inicial.

Exercícios - Capítulo 7

1. Ajuste o modelo aditivo de Aalen para os dados do Exercício 6 do Capítulo 2.

2. Ajuste o modelo aditivo de Aalen para os dados de leucemia pediátrica descritos na Seção 1.5.3. Compare os resultados com aqueles do modelo de Cox apresentados na Seção 5.7.3.

3. Ajuste o modelo aditivo de Aalen para os dados de leucemia aguda apresentados na Seção 4.5.2. Compare os resultados com aqueles do modelo de regressão exponencial apresentados nesta mesma seção.