MI616 - Análise de Sobrevivência

Profa.: Hildete Prisco Pinheiro

$3^{\underline{a}}$ Lista de Exercícios - Modelos paramétricos em análise de sobrevivência

1. **(a)** Sejam X_1 e X_2 duas variáveis aleatórias independentes com distribuições Binomiais $B(n_1, p)$ e $B(n_2, p)$, respectivamente. Qual a distribuição de $Y = X_1 + X_2$? Mostre que:

$$X_1 \mid Y = m \sim H(N, m, n_1), \text{ isto é,}$$

$$\mathbb{P}(X_1 = x \mid Y = m) = \frac{\binom{m}{x}\binom{N-m}{n_1-x}}{\binom{N}{n_1}}.$$

(b) Se X é uma v.a. com distribuição $H(N, m, n_1)$ dada acima, em que $x = 0, 1, \ldots, n_1$, se $n_1 \leq m$ e $x = 0, 1, \ldots, m$, se $n_1 > m$.

Mostre que:

$$(i) \quad E(X) = \frac{n_1 m}{N}$$

$$(ii) \quad Var(X) = \frac{n_1 m}{N}\left(1 - \frac{m}{N}\right)\left(\frac{N - n_1}{N - 1}\right) = \frac{n_1 m(N-m)(N-n_1)}{N^2(N-1)}.$$

2. Utilizando o "R" ou algum outro pacote estatístico e os dados *oc.dat* e *oc.txt*. (Dica: Você pode modificar o programa *lifetest.s*).

(a) Faça o gráfico da curva de sobrevivência, usando o estimador de Kaplan-Meier, para cada um dos três estágios (grupos) em um só gráfico. O que você percebe sobre a associação entre censura e grupo?

(b) Ache a estimativa do tempo mediano de recorrência de 2LL (*"Second Look Laparatomy"*) para cada um dos três estágios (grupos). Será mais fácil olhar a lista de valores da função estimada do que olhar no gráfico (use o comando *print(nome da variável)*).

(c) Faça o gráfico da estimativa da função risco por *kernel* usando a largura da banda de 120 para cada um dos três grupos no mesmo gráfico. (Calcule as estimativas por *kernel* no intervalo [0,00001; 5400] com distância igual a 500 entre os instantes de tempo).

(d) Repita o item (c) com largura de banda igual a 60 e 120.

(e) Compare os gráficos gerados em (c) e (d). Qual é o motivo das diferenças?

(f) Baseado nos valores de $\hat{S}(t)$ e seus I.C. (95%), o que você poderia dizer sobre a diferença entre os 3 grupos?

3. Utilizando o SAS (ou algum outro pacote que possa resolver as questões abaixo) e os dados descritos nos arquivos *oc.dat* e *oc.txt*:

(a) Faça o gráfico da função de sobrevivência para cada um dos dois tratamentos em um só gráfico.

(b) Faça o gráfico de $log(-logS(t))$ e $-logS(t)$ para cada um dos tratamentos contra o tempo.

(c) Faça o teste de logrank para comparar a sobrevivência dos dois tratamentos.

(d) Faça o teste generalizado de Wilcoxon - Peto/Prentice para comparar os dois tratamentos.

(e) O que você pode concluir sobre os dois tratamentos?

4. Utilizando o SAS (ou algum outro pacote que possa resolver as questões abaixo) e os dados descritos nos arquivos *oc.dat* e *oc.txt*:

(a) Estime a função de sobrevivência para cada grau (*grade*). Faça gráficos para checar graficamente se as distribuições exponencial ou Weilbull são razoáveis para o ajuste dos dados. Comente. Qual grau parece ter um melhor prognóstico e qual parece ser ter o pior, em termos de tempo de recorrência.

(b) Ajuste um modelo exponencial separadamente para cada grau. Diga quais as estimativas das funções risco, log-risco e seus erros padrão. Interprete as estimativas dos riscos.

OBS: Para o cálculo dos erros padrão você precisa usar o método delta.

(c) Sob o modelo exponencial, teste a hipótese nula de que os três graus têm a mesma distribuição do tempo de recorrência. Use $\alpha = 0,05$. Ache os p-valores.

Use o teste da razão de verossimilhança.

(d) Repita o item (c) para um modelo Weibull. Comente.

(e) Para o grau=3, teste a hipótese nula de que os tempos de recorrência têm uma distribuição exponencial contra a hipótese alternativa de que eles têm distribuição Weibull. Use o teste da razão de verossimilhança. Use $\alpha = 0,05$. Ache o p-valor.

5. Pacientes em tratamento para câncer de ovário foram assinalados para tratamento com radioterapia (P32) ou controle (nenhum tratamento). O tempo de recorrência é a resposta de interesse. Se o tratamento é eficaz, espera-se que o tempo de recorrência seja prolongado. Utilizando ainda o arquivo *oc.dat* e *oc.txt*, analise:

1. Para cada grupo (P32, controle).

   (a) Ajuste um modelo exponencial (ρ) usando o SAS proc lifereg. Obtenha a estimativa de ρ e um intervalo de 95% de confiança.

   (b) Ajuste um modelo Weibull(ρ, κ). Obtenha as estimativas de ρ e κ.

   (c) Teste a hipótese de que o tempo de sobrevivência tem uma distribuição exponencial contra uma distribuição Weibull.

   (d) Baseado no modelo Weibull, ache a estimativa da média e mediana dos tempos de sobrevivência.

2. Tome o grupo controle como referência. Assuma que os tempos de sobrevivência têm uma distribuição Weibull. Ajuste um modelo em que o tratamento acelera (ou desacelera) o tempo de recorrência. Interprete as estimativas dos parâmetros (como você explicaria os resultados para um médico). Obtenha um intervalo de 95% de confiança para o fator de aceleração ($\psi(\mathbf{Z}) = exp(-\beta^{\star})$).

3. Tome o grupo controle como referência. Assuma que os tempos de sobrevivência têm uma distribuição Weibull. Ajuste um modelo de riscos proporcionais. Interprete as estimativas dos parâmetros. Obtenha um intervalo de 95% de confiança para a razão dos riscos.

6. Os dados abaixo são de pacientes com câncer de ovário que se submeterem a uma segunda cirurgia. Os tempos (em dias) da cirurgia a recorrência são dados. Este conjunto de dados inclui pacientes com grau=3 da doença, que fizeram tratamento com radioterapia. Os dados são:

120	140	181	183	250	301	313+	358	395	400
420+	596	599	649	685	729	762	830+	866+	914
937+	1097+	1223+	1261+	1273	1608	1699	2435+	2647	2746+
3032+	3053+	3125+	3995+	4097+	4164+.				

(a) A distribuição exponencial parece um modelo razoável para os dados? Faça um gráfico que mostre essa questão.
Para um modelo exponencial com risco β:
(b) Escreva a função de verossimilhança.
(c) Faça um gráfico da verossimilhança versus β.
(d) Qual é o estimador de máxima verossimilhança do risco, a média e a mediana dos tempos de sobrevivência?
(e) Teste a hipótese de que a verdadeira mediana do tempo de sobrevivência é um ano contra a hipótese alternativa de que é maior do que um ano.

7. Suponha que a função risco, $\lambda(.)$, para uma variável aleatória contínua positiva começa em 0 e cresce linearmente com o tempo, i.e.,

$$\lambda(t) = \beta t, \qquad \beta > 0,\ t > 0$$

Ache expressões para as funções densidade e de sobrevivência. Interprete.

8. Considere os dados do arquivo *6-mp.dat*. Assuma que os tempos de remissão dos dois grupos de tratamento seguem uma distribuição exponencial. Teste a hipótese de que os dois tratamentos são igualmente eficazes usando o teste da razão de verossimilhança.

9. Para os mesmos dados do arquivo *6-mp.dat*, teste a hipótese de que $\lambda_2 = 5\lambda_1$.

10. Suponha que os tempos de sobrevivência de dois grupos de pacientes de câncer de pulmão seguem uma distribuição Weibull. Uma amostra de 30 pacientes (15 para cada grupo) foi estudada. As estimativas de máxima verossimilhança obtidas para os dois grupos foram, respectivamente, $\hat{\gamma}_1 = 3$, $\hat{\lambda}_1 = 1,2$ e $\hat{\gamma}_2 = 2$, $\hat{\lambda}_2 = 0,5$. Teste a hipótese de que os dois grupos têm a mesma distribuição Weibull.

11. Prove que se T tem distribuição exponencial com parâmetro ρ e a censura é em um tempo fixado c, então a probabilidade de censura é $\pi = e^{-\rho c}$ e se $X = min(T, c)$,

$$\rho E(X) = 1 - \pi.$$

Então mostre que o mesmo resultado é válido quando o tempo de censura é uma variável aleatória com uma distribuição arbitrária. (Dica: $T \mid C \sim exp(\rho)$)

12. Considere os dados de pacientes de leucemia abaixo. Assuma que o tempo de remissão de dois grupos de tratamentos seguem uma distribuição exponencial. Teste a hipótese de que os dois tratamentos são igualmente eficazes usando:

(a) O teste da razão de verossimilhança.

(b) Obtenha um intervalo de confiança (95%) para a razão dos riscos.

6-mp:	6	6	6	7	10	13	16	22	23	6+	9+
	10+	11+	17+	19+	20+	25+	32+	32+	34+	35+	
Placebo:	1	1	2	2	3	4	4	5	5	8	8
	8	8	11	11	12	12	15	17	22	23	

13. Para os mesmos dados do exercício anterior, teste a hipótese de que $\lambda_2 = 5\lambda_1$.