

## Lista 4

Profa. Agatha Rodrigues

Maio de 2024

1. Os dados no arquivo `HOD_NHL.csv` são referentes a 43 pacientes com linfoma de Hodgking ou linfoma não Hodgking, submetidos a transplante de medula óssea. Alguns pacientes receberam transplante de doador aparentado compatível (transplante alogênico) e outros receberam transplante autólogo (ou seja, a própria medula óssea do paciente é coletada e posteriormente infundida). Nos dados também há informação sobre o escore de Karnofsky, que é uma medida de performance que classifica os pacientes segundo o bem-estar dos pacientes (quanto menor a classificação na escala, pior a qualidade de vida). O objetivo principal do estudo é a comparação dos tipos de transplante, considerando-se o tempo (dias) livre de doença (i.e., tempo antes de ocorrência da recorrência ou óbito).
  - (a) Construa curvas de Kaplan-Meier para o tempo de sobrevivência dos pacientes com linfoma para cada um dos grupos de tratamento. Categorize a variável escore de Karnofsky, criando uma variável com duas categorias:  $\text{escore} < 80$  ou  $\text{escore} \geq 80$ . Construa gráficos de Kaplan-Meier para essa nova variável categorizada. Comente.
  - (b) Ajuste um modelo de regressão Weibull com a variável tratamento e a variável escore de Karnofsky categorizada conforme o item (a). Apresente as estimativas de máxima verossimilhança dos parâmetros do modelo Weibull.
  - (c) Encontre uma estimativa pontual para a razão de taxas de falha de pacientes que receberam transplante autólogo e alogênico. Faça o mesmo para a outra variável (escore de Karnofsky) incluída no modelo.
  - (d) Teste a hipótese de igualdade dos tipos de transplante e também das categorias do escore de Karnofsky, utilizando a estatística de Wald, com um nível de significância de 10%. Comente.
2. Considere os dados do exercício 1.
  - (a) Refaça o item (b) utilizando a distribuição log-logística. Especifique claramente qual foi o modelo utilizado e quais foram os parâmetros estimados.

(b) A chance de sobrevivência após  $t$  é definida como

$$\frac{S(t|\mathbf{X})}{1 - S(t|\mathbf{X})}.$$

No modelo logístico, mostre que

$$\frac{S(t|\mathbf{X})}{1 - S(t|\mathbf{X})} = \exp[-\beta^T \mathbf{X}] \frac{S(t|\mathbf{X} = \mathbf{0})}{1 - S(t|\mathbf{X} = \mathbf{0})}.$$

- (c) Obtenha uma estimativa da razão de chances de sobrevivência após  $t$  de pacientes com transplante autólogo e pacientes com transplante alogênico. Interprete.
- (d) Repita o item (d) do exercício 1, utilizando o modelo log-logístico. Compare os resultados e comente.
3. Considerando ainda os dados do exercício 1, faça gráficos apropriados da taxa de falha acumulada para verificar a adequabilidade dos modelos.
- (a) Weibull;
- (b) Log-logístico.

Em todos os casos, utilize o estimador de Nelson-Aalen da função de taxa de falha acumulada considerando cada grupo separadamente (ou seja, obtenha estimativas da função de taxa de falha acumulada para cada grupo).

4. Considere ainda os dados do exercício 1. Obtenha os resíduos de Cox-Snell e *deviance* para os modelos de regressão Weibull (ajustado no exercício 1) e log-logístico (ajustado no exercício 2). Faça gráficos dos resíduos em função do tempo e comente. Com base em todas as análises feitas, discuta se os modelos (Weibull ou log-logístico) parecem ser adequados para os dados trabalhados.
5. Os dados do arquivo *viarapida.xls* referem-se a um estudo feito pelo Instituto do Coração, do Hospital das Clínicas - FMUSP, para comparar dois tipos de protocolo de recuperação: via-rápida e convencional. O protocolo via-rápida tem como filosofia uma maior integração entre as várias equipes que atendem os pacientes com a finalidade de tentar diminuir o tempo de internação, melhorando a recuperação e diminuindo custos. Considere, para as questões que seguem, apenas as seguintes variáveis:
- Delta - indicador de censura (0 - censura, 1 - saída do centro cirúrgico);
  - Idade - idade do paciente na internação;
  - Protocolo - sendo 0 convencional e 1 via rápida;
  - Raça - sendo 1 Branco, 2 negro e 3 amarelo;
  - Sexo - sendo 0 feminino e 1 masculino;

- Tipac - sendo 0 congênito e 1 coronariano; e
  - T-CC - tempo entre a entrada e a saída do centro cirúrgico (horas).
- (a) Construa curvas de Kaplan-Meier para o T-CC com a finalidade de comparar, descritivamente, o sexo, a raça, o tipo de paciente (Tipac) e o protocolo com relação ao tempo de permanência no centro cirúrgico.
  - (b) Construa testes estatísticos não paramétricos para averiguar a significância estatística dos fatores do item anterior. Interprete os resultados.
  - (c) Considerando apenas o protocolo e métodos gráficos baseados na função de taxa de falha acumulada não paramétrica, procure averiguar qual dos modelos, exponencial, Weibull ou log-normal é mais adequado aos dados.
  - (d) Compare o ajuste dos modelos exponencial, Weibull e log-normal utilizando os resíduos de Cox-Snell. Qual o modelo que você recomenda?
  - (e) Com base no modelo obtido no item anterior, faça o teste paramétrico para comparar os protocolos com relação ao tempo de permanência no centro cirúrgico.
6. 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 semiparamétrico 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  $H_0 : \beta = 0$ , 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:
    - i. 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.

- ii. 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.
- iii. Obtenha a estatística do teste do escore, dada por

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

que, sob  $H_0 : \beta = 0$ , tem distribuição de qui-quadrado com 1 grau de liberdade.

7. Considere a situação descrita no exercício 6. Esse problema corresponde a um estudo realizado com 35 mulheres, sendo observados os seguintes tempos (tempos censurados à direita estão denotados por um sinal “+”):

Grupo	Tempos
Tumor grande	28, 89, 175, 195, 309, 377 <sup>+</sup> , 393 <sup>+</sup> , 421 <sup>+</sup> , 447 <sup>+</sup> , 462, 709 <sup>+</sup> , 744 <sup>+</sup> , 770 <sup>+</sup> , 1106 <sup>+</sup> , 1206 <sup>+</sup>
Tumor moderado	34, 88, 137, 199, 280, 291, 299 <sup>+</sup> , 300 <sup>+</sup> , 308, 351, 358, 369, 370, 371, 375, 382, 392, 429 <sup>+</sup> , 451, 1119 <sup>+</sup>

- (a) Utilizando um pacote estatístico, obtenha as curvas de Kaplan-Meier de cada um dos grupos. Apresente as duas curvas no mesmo gráfico e compare.
- (b) Teste a igualdade das curvas utilizando o teste de *log-rank*. Conclua.
- (c) Ajuste um modelo de riscos proporcionais de Cox. Obtenha a estimativa de  $\beta$ , bem como o seu erro padrão. Interprete o coeficiente  $\beta$ .
- (d) Teste a hipótese  $H_0 : \beta = 0$  utilizando a estatística do teste de Wald, calculada com base nas estimativas de  $\beta$  e de seu erro-padrão obtidos em (c).
- (e) Teste a hipótese  $H_0 : \beta = 0$  utilizando a estatística do teste do escore obtida no exercício 6 (sem a utilização de *software* estatístico). Compare com os itens anteriores, comente e conclua.