

- As listas PRECISAM ser entregues por grupos de 4 ou 5 alunos. Listas com menos autores NÃO serão aceitas.
- Prazo de entrega: 31/10/2018.

1. (100 pontos) O arquivo disponível em

<http://www.ime.usp.br/~jmsinger/MAE0560/esfregaco.xls>

contém dados de um estudo em que 3 avaliadores (A, B e C) classificaram 150 dentes quanto ao efeito de um tratamento na remoção do esfregaço de paredes dentinárias. Utilizou-se uma escala ordenada de 4 pontos em que 0 corresponde a uma superfície completamente coberta pelo esfregaço e 3 a uma superfície completamente livre do esfregaço.

- Proponha um modelo probabilístico conveniente para os dados e interprete os seus parâmetros.
- Compare a concordância entre cada par de avaliadores por intermédio de estatísticas Kappa.

2. (50 pontos) Sob o modelo Produto de Multinomiais cuja função de probabilidades é

$$P(\mathbf{n}|\boldsymbol{\theta}) = \prod_{i=1}^s n_i! \prod_{j=1}^r \theta_{ij}^{n_{ij}} / n_{ij}!, \quad \sum_{j=1}^r \theta_{ij} = 1, \quad i = 1, \dots, s,$$

modelos log-lineares podem ser expressos como

$$\log \boldsymbol{\theta} = (\mathbf{I}_s \otimes \mathbf{1}_r) \boldsymbol{\lambda} + \mathbf{X} \boldsymbol{\beta}$$

em que $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_s)^\top$, $\mathbf{X} = (\mathbf{X}_1^\top, \dots, \mathbf{X}_s^\top)^\top$ e $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ é um vetor de dimensão $p \times 1$ com $p < s(r-1)$. Mostre que essa formulação do modelo log-linear é equivalente à formulação

$$\mathbf{A} \log \boldsymbol{\theta} = \mathbf{X}_G \boldsymbol{\beta}$$

em que \mathbf{A} é uma matriz de dimensão $s(r-1) \times sr$ tal que $r(\mathbf{A}) = s(r-1)$ e $\mathbf{A}(\mathbf{I}_s \otimes \mathbf{1}_r) = \mathbf{0}$ com as relações $\mathbf{X}_G = \mathbf{A} \mathbf{X}$ e $\mathbf{X} = \mathbf{A}^\top (\mathbf{A} \mathbf{A}^\top)^{-1} \mathbf{X}_G$.

3. (30 pontos) Considere uma tabela 2x3 sob um modelo probabilístico Multinomial. Mostre que sob o modelo estrutural log-linear de efeito de linhas com escores 1, 2 e 3 para as colunas, o logaritmo da razão de chances envolvendo as colunas extremas é o dobro do logaritmo de cada uma das razões de chances envolvendo colunas adjacentes.

4. (20 pontos) Considere uma tabela 2×2 sob um modelo probabilístico Multinomial. Expresse o modelo de independência na formulação $\mathbf{A} \log(\boldsymbol{\theta}) = \mathbf{X} \boldsymbol{\beta}$.

5. (30 pontos) Considere o seguinte modelo log-linear para uma tabela 2×2 sob um esquema probabilístico Multinomial

$$\log(\theta_{ij}) = \lambda + \lambda_i^X + \lambda_i^Y,$$

com as restrições $\lambda_2^X = \lambda_2^Y = 0$ e $\lambda_1^X = \lambda_1^Y$.

a) Que relações estruturais são impostas por esse modelo?

b) Construa a tabela de frequências correspondente sabendo que $\lambda_1^X = \log(7)$ e que $n = 64$.

6. (30 pontos) Para avaliar a preferência de consumidores por dois produtos, A e B, uma amostra de n consumidores foi selecionada de uma população de interesse. Desses n consumidores, n_1 declararam preferir o produto A, n_2 declararam preferir o produto B e n_3 declararam-se indiferentes. Proponha um modelo probabilístico para os dados e obtenha o testes da razão de verossimilhanças e de Pearson para a hipótese de que a probabilidade de preferência pelo produto B é o dobro daquela de preferência pelo produto A.

7. (40 pontos) Os dados do arquivo disponível em

<http://www.ime.usp.br/~jmsinger/Dados/Singer&Correia&Paschoalinoto1989.xls>

são provenientes de um estudo realizado na Faculdade de Ciências Farmacêuticas da USP com a finalidade de comparar a técnica sorológica ELIEDA com outras quanto à capacidade de diagnóstico da esquistossomose. Estime a sensibilidade e especificidade de cada uma das técnicas sorológicas para detecção de esquistossomose e construa intervalos de confiança (95%) em cada caso.