MAE 0560 - $2o$ Semestre de 2018 **Lista 1**

- As listas PRECISAM ser entregues por grupos de 5 ou 6 alunos. Listas com menos autores NÃO serão aceitas.
- Prazo de entrega: 29/08/2018.

1. (40 pontos) Considere uma tabela de contingência $I \times J$ para a qual assumimos uma distribuição Multinomial cuja função de probabilidade é dada por:

$$P(\mathbf{n}|n, \boldsymbol{\theta}) = n! \prod_{i=1}^{I} \prod_{j=1}^{J} \theta_{ij}^{n_{ij}} / n_{ij}!, \ \sum_{i=1}^{I} \sum_{j=1}^{J} \theta_{ij} = 1$$

em que θ_{ij} representa a probabilidade de uma unidade amostral selecionada ao acaso ser classificada na i-ésima categoria de resposta de uma variável X e j-ésima categoria de resposta de uma variável Y.

a) Determine os estimadores de máxima verossimilhança (MV) dos parâmetros θ_{ij}, $i = 1, \ldots, I$, $j = 1, \ldots, J$.

b) Mostre que as distribuições dos vetores $\mathbf{n}_{*+} = (n_{1+}, \ldots, n_{I+})^\top$ e $\mathbf{n}_{+*} = (n_{+1}, \ldots, n_{+J})^\top$, em que $n_{i+} = \sum_{j=1}^{J} n_{ij}$ e $n_{+J} = \sum_{i=1}^{I} n_{ij}$, para $i = 1, \ldots, I$, $j = 1, \ldots, J$, também são distribuições Multinomiais;

c) Defina os parâmetros dessas distribuições e obtenha os respectivos estimadores de MV.

d) Determine o estimador de MV dos parâmetros θ_{ij}, $i = 1, \ldots, I$, $j = 1, \ldots, J$ sob a hipótese de independência das variáveis X e Y.

e) Determine a distribuição condicional de $\mathbf{n} = (n_{11}, \ldots, n_{IJ})^\top$ dado $\mathbf{n}_{*+}$; qual a importância prática desse resultado?

2. (40 pontos) Considere uma amostra aleatória $\mathbf{X}_1, \ldots, \mathbf{X}_n$ de uma distribuição Bernoulli Multivariada com vetor de parâmetros $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_r)^\top$, $0 < \pi_i < 1$, $\sum_{i=1}^{n} \pi_i = 1$.

a) Mostre que se $\mathbf{N} = (N_1, \ldots, N_r)^\top = \sum_{i=1}^{n} \mathbf{X}_i$ então $\mathbf{N}$ tem distribuição Multinomial com $\mathbb{E}(N_i) = n\pi_i$ e $\mathbb{C}\mathrm{ov}(N_i, N_j) = n\pi_i(1 - \pi_i)$ se $i = j$ ou $\mathbb{C}\mathrm{ov}(N_i, N_j) = -n\pi_i\pi_j$ se $i \neq j$.

b) Mostre que o estimador de máxima verossimilhança de $\boldsymbol{\pi}$ é $\mathbf{p} = (p_1, \ldots, p_r)^\top$ em que $p_i = N_i/n$. Além disso, mostre que $\mathbb{E}(\mathbf{p}) = \boldsymbol{\pi}$ e que $\mathbb{V}(\mathbf{p}) = n^{-1}[\mathbf{D}_{\boldsymbol{\pi}} - \boldsymbol{\pi}\boldsymbol{\pi}^\top]$ em que $\mathbf{D}_\pi$ denota uma matriz diagonal com os elementos de $\boldsymbol{\pi}$ ao longo da diagonal principal.

c) Como você aproximaria a distribuição de $\mathbf{p}$ para n grande? Justifique sua resposta.

3. (10 pontos) Considere a seguinte tabela 2x2

	Doente (D)	Não-doentes ($\overline{D}$)	Total
Exposto (E)	n_{11}	n_{12}	$n_{1\bullet}$
Não exposto ($\overline{E}$)	n_{21}	n_{22}	$n_{2\bullet}$
Total	$n_{\bullet 1}$	$n_{\bullet 2}$	$n_{\bullet\bullet}$

correspondente a um estudo em que o interesse é avaliar a associação entre a exposição de indivíduos a um certo fator de risco e a ocorrência de uma determinada moléstia. Em estudos prospectivos (*prospective, follow-up, cohort*) o planejamento envolve a escolha de amostras de tamanhos $n_{1\bullet}$ e $n_{2\bullet}$ de indivíduos expostos e não expostos ao fator de risco, respectivamente e a observação da ocorrência ou não da moléstia após um certo intervalo de tempo. A razão de chances é definida como:

$$\omega = \frac{P(D|E)P(\overline{D}|\overline{E})}{P(\overline{D}|E)P(D|\overline{E})}.$$

Em estudos retrospectivos ou caso-controle, o planejamento envolve a escolha de amostras de tamanhos $n_{\bullet 1}$ e $n_{\bullet 2}$ de indivíduos não-doentes (controles) e doentes (casos), respectivamente e a observação retrospectiva de sua exposição ou não ao fator de risco. Nesse caso a razão de chances é definida por:

$$\omega = \frac{P(E|D)P(\overline{E}|\overline{D})}{P(\overline{E}|D)P(E|\overline{D})}.$$

Utilize a definição de probabilidade condicional para mostrar que duas expressões são iguais e comente sobre a importância prática desse resultado.

4. (10 pontos) De uma tabela construída para avaliar a associação entre tratamento (ativo e placebo) e cura (sim ou não) de uma certa moléstia obteve-se uma razão de chances igual a 2.0. Explique por que não se pode concluir daí que a probabilidade de cura para pacientes submetidos ao tratamento ativo é 2 vezes a probabilidade de cura para pacientes submetidos ao placebo.

5. (25 pontos) Numa pesquisa sexológica, ambos os parceiros de 36 casais foram inquiridos sobre sua opinião relativa a vários aspectos de sua vida sexual conjunta. À questão "Você acredita que seu parceiro tem problemas sexuais?" foram dadas as respostas indicadas na tabela abaixo:

Número de casais com			
		Resposta da mulher	
		sim	não
Resposta	sim	4	10
do marido	não	7	15

a) Proponha um modelo probabilístico adequado para descrever a situação e traduza as seguintes hipóteses em termos dos parâmetros do modelo

b) a crença de que o respectivo parceiro tem problemas sexuais é tão provável entre os homens quanto entre as mulheres;

c) a probabilidade de uma resposta afirmativa da mulher é a mesma tanto sob uma resposta afirmativa quanto sob uma resposta negativa do homem;

d) a ocorrência de opiniões divergentes em cada casal é duas vezes mais provável do que a ocorrência de uma resposta negativa por parte da mulher e afirmativa por parte do homem;

e) a opinião da mulher é independente daquela de seu companheiro.

6. (30 pontos) Num estudo em que se deseja comparar homens e mulheres com relação à probabilidade de apoio a um determinado projeto social, obteve-se a seguinte tabela após a seleção de n_{1+} homens e n_{2+} mulheres de uma população de interesse.

	Favorável (F)	Desfavorável ($\overline{F}$)	Total
Homem (H)	n_{11}	n_{12}	n_{1+}
Mulher (M)	n_{21}	n_{22}	n_{2+t}
Total	n_{+1}	n_{+t2}	n_{++}

Mostre que a distribuição da estatística de Pearson utilizada para esse fim, nomeadamente

$$Q_P = \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{(n_{ij} - n_{i+}n_{+j}/n_{++})^2}{n_{i\bullet}n_{+j}/n_{++}}$$

pode ser aproximada por uma distribuição qui-quadrado com 1 grau de liberdade.

Sugestão: Mostre que a estatística de Pearson pode ser escrita como

$$Q_P = \left[\frac{n_{11}/n_{1+} - n_{21}/n_{2+}}{\sqrt{(n_{+1}/n_{++})(1 - n_{+1}/n_{++})(1/n_{1+} + 1/n_{2+})}}\right]^2$$

e use o Teorema Limite Central.

7. (30 pontos) Um tipo de equipamento eletrônico tem dois componentes que podem falhar ou não. Com o objetivo de estudar a distribuição do número de falhas, cada um de n equipamentos desse tipo foi avaliado durante um certo tempo e classificado segundo o número de componentes com falha. As frequências observadas estão representadas na tabela seguinte.

Número de componentes com falha	0	1	2	Total
Frequência observada	n_0	n_1	n_2	n

a) Proponha um modelo probabilístico para os dados e interprete os seus parâmetros.

b) Admitindo que a probabilidade de falha de cada um dos dois componentes é θ e que eles atuam independentemente, proponha um modelo estrutural para os parâmetros do modelo proposto em a) sem recorrer à notação matricial.

c) Obtenha o estimador de máxima verossimilhança de θ.

d) Mostre que o estimador obtido no item anterior é não-enviesado e que sua variância é $\theta(1-\theta)/(2n)$.

e) Proponha um teste para a hipótese de que $\theta = \theta_0$.

f) Teste a hipótese do item e) com $\theta_0 = 0.10$ sabendo que as frequências observadas para 0, 1 e 2 falhas foram, respectivamente 153, 17 e 9.

8. (30 pontos) Mostre como uma expansão de Taylor pode ser usada para obter expressões aproximadas para a variância de funções de variáveis aleatórias (método Delta). Considere os casos uni e multivariados.

9. (20 pontos) Considere um modelo binomial com parâmetros n e p. Seja $\widehat{p}$ a proporção amostral. Obtenha a variância aproximada de $\log[\widehat{p}/(1-\widehat{p})]$.

10. (40 pontos) Considere os estimadores de mínimos quadrados $\widehat{\boldsymbol{\beta}} = (\widehat{\alpha}\ \ \widehat{\beta}\ \ \widehat{\gamma})^\top$ dos parâmetros do modelo de regressão polinomial $y_i = \alpha + \beta x_i + \gamma x_i^2 + e_i$ com $\mathbb{E}(e_i) = 0$, $\mathbb{V}(e_i) = \sigma^2$, independentes, $i = 1, \ldots, n$ e seja

$$\mathbb{V}(\widehat{\boldsymbol{\beta}}) = \begin{pmatrix} s_\alpha^2 & s_{\alpha\beta} & s_{\alpha\gamma} \\ s_{\alpha\beta} & s_\beta^2 & s_{\beta\gamma} \\ s_{\alpha\gamma} & s_{\beta\gamma} & s_\gamma^2 \end{pmatrix}.$$

respectiva matriz de covariâncias. Obtenha uma expressão para a variância de $-\widehat{\beta}/(2\widehat{\alpha})$, estimador do ponto em que o polinômio quadrático atinge o máximo (ou mínimo).