ME 714 - Análise de dados discretos
Primeiro semestre de 2014
Prova I
Data: 07/05/2014

Nome: ______________________________________ RA: ____________________

Leia atentamente as instruções abaixo:

- **Coloque seu nome completo e RA em todas as folhas que você recebeu, inclusive nesta.**
- **Utilize somente um dos lados de cada folha.**
- **Leia atentamente cada uma das questões.**
- **Enuncie, claramente, todos os resultados que você utilizar.**
- **Justifique, adequadamente, seus desenvolvimentos, sem, no entanto, escrever excessivamente.**
- **O(a) aluno(a) só poderá sair da sala após as 16h30, mesmo que já tenha finalizado a prova. Após a saída do(a) primeiro(a) aluno(a) não será permitido a entrada de nenhum(a) outro(a) aluno(a).**
- **Não é permitdo empréstimo de material.**
- **Não serão dirimidas dúvidas de quaisquer natureza, após os 20 minutos iniciais.**
- **Resolva a prova, preferencialmente, à caneta, e procure ser organizado(a). Se fizer à lápis, destaque, à caneta, sua resposta.**
- **O(a) aluno(a) deverá portar sua carteira de estudante e apresentá-la, quando for solicitada sua assinatura.**
- **Utilize somente um lado de cada folha de resolução.**
- **Contestações a respeito da nota, só serão consideradas se estiverem por escrito.**
- **A nota do aluno(a) será $\frac{NP}{NT} \times 10$, em que NP é o número de pontos obtidos na prova e NT é o numero total de pontos da prova.**
- **Os resultados numéricos finais devem ser apresentados com duas casas decimais, apenas.**
- **A prova terá duração de 120 minutos, das 16h às 18h, improrrogáveis.**

Faça uma excelente Prova!!

1. Seja uma amostra aleatória, de tamanho n, de $X \sim \text{Bernoulli}(\theta), \theta \in (0,1)$. Responda os itens:

   a) Obtenha o estimador de máxima verossimilhança (emv) de θ (provando que é ponto de máximo) e sua respectiva distribuição assintótica. (100 pontos)

   b) Considere o interesse em testar $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$, com $\theta_0 \in (0,1)$ conhecido. Proponha um teste para testar essas hipóteses, de sorte que a distribuição assintótica da estatística do teste, sob H_0, seja χ^2_1. Sugestão: utilize o emv e sua respectiva distribuição assintótica. (50 pontos)

   c) Obtenha a versão assintótica do teste da razão de verossimilhanças (TRV) para testar as hipóteses descritas no item b). (100 pontos)

   d) Considere agora uma única observação de um vetor aleatório $\boldsymbol{X} = (X_1, X_2) \sim \text{trinomial}(n, \boldsymbol{\theta}), \boldsymbol{\theta} = (\theta_1, \theta_2)'$. Assuma que $\theta_1 = \theta_2 = \theta, \theta \in (0,1)$. Obtenha o emv de θ (não precisa provar que é ponto de máximo) e suas respectivas esperança e variância exatas. (150 pontos)

2. Considere um estudo realizado, entre eleitores estadunidenses, para verificar a existência de uma possível associação entre a inclinação partidária e o gênero. Para tal, foram selecionados aleatoriamente um total de 1230 pessoas do gênero feminino e 961 do gênero masculino e, para cada uma delas, foi perguntada sua inclinação partidária. Os resultados se encontram resumidos na Tabela 1. Denote a frequência (populacional) de cada casela por N_{ij} e as respectivas probabilidades de ocorrência por θ_{ij}, em que i e j correspondem às linhas e colunas, respectivamente.

Tabela 1: Resultados da pesquisa sobre inclinação partidária

		Inclinação partidária		
		Democrata	Republicano	Total
Gênero	Feminino	762	468	1230
	Masculino	484	477	961
Total	-	1246	945	2191

Responda os itens:

   a) Escreva o modelo probabilístico (a função de probabilidade) gerador da Tabela 1. (50 pontos)

   b) Escreva as hipóteses (H_0) e (H_1) correspondentes à homogeneidade entre as distribuições binomiais relativas à Tabela 1, em termos das probabilidades θ_{ij} e em termos da razão de chances. Prove que as duas hipóteses nulas (H_0) (e consequentemente as hipóteses alternativas (H_1)), de fato, são equivalentes. (100 pontos)

c) Teste as hipóteses acima utilizando os testes de qui-quadrado e para a razão de chances, considerando um nível de significância $\alpha = 0,05$. Suas conclusões foram as mesmas em relação aos dois testes? A igualdade (ou diferença) que você observou, entre as conclusões, era esperada? Justifique, adequadamente, sua resposta. (200 pontos)

d) Escreva a hipótese de homogeneidade em termos do modelo $\boldsymbol{A\pi} = \boldsymbol{X\beta}$, em que $\boldsymbol{\pi}$ é o vetor de parâmetros completo. (50 pontos)

3. Considere um estudo para verificar a concordância entre dois métodos de classificação de risco de cárie. Um total de 97 indivíduos foi selecionado. Cada um deles foi classificado, através de cada um dos métodos, em relação aos risco de apresentar cáries dentárias. Os resultados se encontram na Tabela 2. Denote a frequência (populacional) de cada casela por N_{ij} e as respectivas probabilidades de ocorrência por θ_{ij}, em que i e j correspondem às linhas e colunas, respectivamente.

Tabela 2: Resultados do estudo de comparação dos métodos de classificação de risco de cárie

		Risco de cárie segundo o método convencional			
		Baixo	Médio	Alto	Total
Risco de cárie segundo	Baixo	11	5	0	16
o método simplificado	Médio	14	34	7	55
	Alto	2	13	11	26
Total	-	27	52	18	97

a) Escreva o modelo probabilístico (a função de probabilidade) gerador da Tabela 2. (50 pontos)

b) Escreva as hipóteses (H_0) e (H_1) correspondentes à independência entre os métodos de classificação de cárie. Teste estas hipóteses utilizando o testes de qui-quadrado, considerando um nível de significância $\alpha = 0,01$.(150 pontos)

c) Escreva as hipóteses (H_0) e (H_1) correspondentes à existência de simetria (concordância) marginal, ou seja, correspondentes à hipótese de igualdade entre as distribuições marginais, de modo escalar e na forma $\boldsymbol{B\pi} = \boldsymbol{D}$, em que $\boldsymbol{\pi}$ é o vetor de parâmetros completo. (150 pontos)

Formulário

1. Se $X \sim \text{Bernoulli}(\theta), \theta \in (0,1)$, então
$p(x) = \theta^x(1-\theta)^{1-x}\mathbb{1}_{\{0,1\}}(x)$, $\mathcal{E}(X) = \theta, \mathcal{V}(X) = \theta(1-\theta)$.

2. Se $\boldsymbol{X} = (X_1, X_2, ..., X_{r-1}) \sim \text{multinomial}_r(n, \boldsymbol{\theta}), \boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_{r-1})'$, $\boldsymbol{\pi} = (\theta_1, \theta_2, ..., \theta_{r-1}, \theta_r)'$, $\sum_{i=1}^{r} \theta_i = 1$, então

$$p(\boldsymbol{x}) = \frac{n!}{\prod_{i=1}^{r} x_i!} \prod_{I=1}^{r} {\theta_i}^{x_i} \left(\prod_{i=1}^{r} \mathbb{1}_{\{0,1,..,n\}}(x_i) \right) \mathbb{1}_{\{n\}} \left(\sum_{i=1}^{r} x_i \right),$$

$\mathcal{E}(X_i) = n\theta_i$, $\mathcal{V}(X) = n\theta_i(1-\theta_i)$ e $\text{Cov}(X_i, X_j) = -n\theta_i\theta_j$.

3. Teste da razão de verossimilhanças: Sejam $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_1$, $\Theta = \Theta_0 \cup \Theta_1$, $\Theta_0 \cap \Theta_1 = \varnothing$ e Θ o espaço paramétrico. Estatística do teste (versão assintótica): $-2\ln\Lambda = -2(l(\widehat{\theta}_0) - l(\widehat{\theta}))$, em que $\widehat{\theta}_0$ e $\widehat{\theta}$ são os emv de θ, sob H_0 e irrestrito, respectivamente e $l(.)$ denota a logverossimilhança. Sob H_0, para n suficientemente grande, $-2\ln\Lambda \approx \chi^2_q$, q= número total de parâmetros - número total de parâmetros sob H_0.

4. Razão de chances $\pi = \dfrac{\lambda_1}{\lambda_2}$, $\lambda_i = \dfrac{\theta_{i1}}{1-\theta_{i1}}, i = 1, 2$. Tabela (2×2): Estimador de MV: $\widehat{\pi} = \frac{N_{11}N_{22}}{N_{12}N_{21}}$. Distribuição assintótica: Seja $\widehat{\eta} = \ln\widehat{\pi}$, então $\widehat{\eta} \approx N(\eta, \sigma^2_\eta)$, para os tamanhos amostrais relacionados, suficientemente grandes, em que $\eta = \ln\pi$ e $\sigma^2_\eta = \dfrac{1}{n_{11}} + \dfrac{1}{n_{12}} + \dfrac{1}{n_{21}} + \dfrac{1}{n_{22}}$. A estatística do teste para testar as hipóteses $H_0 : \eta = \eta_0$ vs $H_1 : \eta \neq \eta_0$ é facilmente dedutível do resultado assintótico.

5. Teste qui-quadrado para homegeneidade e independência (tabela $r \times s$). Estatística do teste: $Q_H = \displaystyle\sum_{i=1}^{r}\sum_{j=1}^{s} \frac{(N_{ij} - E_{ij})^2}{E_{ij}}$, N_{ij} : frequências observadas de cada casela, $E_{ij} = \dfrac{N_{i.}N_{.j}}{N_{..}}$: frequências esperadas, $N_{i.} = \sum_{j=1}^{s} N_{ij}, N_{.j} = \sum_{i=1}^{r} N_{ij}, N_{..} = \sum_{i=1}^{r}\sum_{j=1}^{s} N_{ij}$; sob H_0, para os tamanhos amostrais relacionados, suficientemente grandes, $Q_H \approx \chi^2_{(r-1)(s-1)}$.