ME - 731 Métodos em Análise Multivariada
Prof. Caio Azevedo
Segundo semestre de 2009, Data: 29/09/2009
Prova I

Leia atentamente as instruções abaixo:

- **Tenha em mãos somente: lápis, borracha e caneta.**
- **Leia atentamente cada uma das questões.**
- **Enuncie, claramente, todos os resultados que você utilizar.**
- **Em caso de dúvida, levante-se e dirija-se ao professor. Pergunte somente o que for imprescindível.**
- **Entregue todas as folhas que você recebeu, inclusive os rascunhos e a prova propriamente, informando o que deve ser corrigido.**
- **Faça a prova, preferencialmente, à caneta, e procure ser organizado. Se fizer à lápis, destaque, à caneta, sua resposta.**
- **Não proceda de maneira indevida como: conversar durante a prova, utilizar-se de material que não permitido, emprestar material à colegas, sem autorização do professor e atender ao telefone celular (a não ser em casos de EXTREMA URGÊNCIA). Isso acarretará em nota 0 na prova.**
- **Se precisar de algum material, inclusive calculadora, levante-se e dirija-se ao professor.**
- **A prova terá duração de 2 horas, improrrogáveis, das 14h às 16h.**

Faça uma excelente prova!!

Questões

1. Seja $\boldsymbol{Z} \sim N_p(\mathbf{0}_p, \boldsymbol{I}_p)$, em que $\mathbf{0}_p$ é um vetor coluna de zeros e $\boldsymbol{I}_p$ é uma matriz identidade de ordem p. Seja $\boldsymbol{\mu} \in \mathcal{R}^p$ e $\boldsymbol{\Sigma}_{(p\times p)}$ uma matriz positiva definida, tal que $\boldsymbol{\Sigma} = \boldsymbol{\Psi}\boldsymbol{\Psi}'$, em que $\boldsymbol{\Psi}$ é a decomposição de Cholesky de $\boldsymbol{\Sigma}$. Qual é a distribuição de $\boldsymbol{X} = \boldsymbol{\Psi}\boldsymbol{Z} + \boldsymbol{\mu}$? Justifique adequadamente sua resposta (2 pontos).

2. Seja $\boldsymbol{X}_1, ..., \boldsymbol{X}_n$ uma amostra aleatória de $\boldsymbol{X} \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\boldsymbol{\mu} = (\mu_1, \mu_2)'$ e $\boldsymbol{\Sigma}$ conhecida. Suponha que desejamos testar as hipóteses $H_0 : \boldsymbol{R\mu} = \boldsymbol{b}_0$ vs $H_1 : \boldsymbol{R\mu} \neq \boldsymbol{b}_0$, em que $\boldsymbol{R}_{(c\times 2)}$ é uma matriz de posto coluna completo, $c \leq 2$, $posto(R) = r \leq 2$, conhecida e $\boldsymbol{b}_0$ é um vetor $(c \times 1)$, conhecido. Para tal, propomos a seguinte estatística:

$$Q = n\left(\boldsymbol{R}\overline{\boldsymbol{X}} - \boldsymbol{b}_0\right)'\left(\boldsymbol{R}\Sigma\boldsymbol{R}'\right)^{-1}\left(\boldsymbol{R}\overline{\boldsymbol{X}} - \boldsymbol{b}_0\right) \tag{1}$$

Responda aos itens:

a) Qual é a distribuição de Q sob H_0? Jusitifique adequadamente (0,5 pontos).

b) Para um determinado conjunto de dados, temos os seguintes resultados: $\overline{\boldsymbol{x}} = (0,7;2,5)'$, $\boldsymbol{\Sigma} = \begin{bmatrix} 2 & 1 \\ 1 & 2.5 \end{bmatrix}$ e $n = 50$. Para testar $H_0 : \boldsymbol{\mu} = (1,2)'$ vs $H_1 : \boldsymbol{\mu} \neq (1,2)'$ obtivemos o seguinte resultado $Q = 12,81$. Qual sua conclusão utilizando $\alpha = 5\%$? Neste caso, quem seriam as matrizes $\boldsymbol{R}$ e $\boldsymbol{b}_0$ (0,5 pontos)?

c) Caso, no item anterior, você tenha rejeitado H_0, descubra onde reside(m) a(s) diferença(s). Para tal, formule adequadamente as duas hipóteses de interesse, defina-as em termos da hipótese: $\boldsymbol{R}\boldsymbol{\mu} = \boldsymbol{b}_0$, escrevendo, adequadamente as matrizes $\boldsymbol{R}$ e $\boldsymbol{b}_0$, construa as regiões do testes, apresente os valores críticos e escreva, adequadamente, suas conclusões, utilizando $\alpha = 5\%$ (1 ponto).

d) Calcule o nível descritivo do teste $H_0 : \boldsymbol{R}\boldsymbol{\mu} = \boldsymbol{b}_0$ vs $H_1 : \boldsymbol{R}\boldsymbol{\mu} \neq \boldsymbol{b}_0$, utilizando a estatística definida em (1) (1 ponto).

e) Obtenha a função poder do teste no contexto descrito no item d) (1 ponto).

3. Seja $\boldsymbol{X}_1, ..., \boldsymbol{X}_n$ uma amostra aleatória de $\boldsymbol{X} \sim N_2(\boldsymbol{0}, \boldsymbol{\Sigma})$, $\boldsymbol{\Sigma} = \sigma^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$, em que $\rho \in (-1,1)$ é conhecido e $\sigma^2 \in (0, +\infty)$, desconhecido. Desejamos testar $H_0 : \sigma^2 = \sigma_0^2$ vs $H_1 : \sigma^2 \neq \sigma_0^2$, σ_0^2 conhecido. Para tal, utilizaremos o teste da razão de verossimilhanças. Responda os itens:

   a) Obtenha os estimadores de máxima verossilhança de σ^2 sob H_0 e sob H_1. Sob H_1, você terá que encontrar o valor que maximiza a verossimilhança e provar que ele é ponto de máximo (1,5 pontos).

   b) Obtenha a estatística do teste da razão de verossimilhanças, do modo mais simplificado possível, e a denote por Λ (1 ponto).

   c) Qual a distribuição assintótica de $-2\ln\Lambda$ sob H_0 (0,5 pontos)?

4. Considere a situação em que temos vários grupos independentes, todos com distribuição normal p variada. Suponha que queremos testar a igualdade simultânea dos seus respectivos vetores de médias e, posteriormente, se for o caso, a investigação de onde reside(m) a(s) diferença(s). Para tal propomos duas metodologias a saber: o teste baseado na estatística de Wilks (que utiliza a decomposição das matrizes de somas de quadrados) e modelo linear normal multivariado em sua forma vetorial. Discuta as vantagens e desvantagens de cada um, comparando-os da melhor forma possível (1 ponto).

Formulário

- Se $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ então $f_{\boldsymbol{X}}(\boldsymbol{x}) = (2\pi)^{-p/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}-\boldsymbol{\mu})\right]$ e $M_{\boldsymbol{X}}(\boldsymbol{t}) = \exp\left\{\boldsymbol{\mu}'\boldsymbol{t} + \frac{1}{2}\boldsymbol{t}'\boldsymbol{\Sigma}\boldsymbol{t}\right\}$