

ME - 731 Métodos em Análise Multivariada
Prof. Caio Azevedo
Segundo semestre de 2009, Data: 29/10/2009
Prova II

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.

Obs: Quando falarmos do modelo de análise fatorial ortogonal (MAFO), com as suposições usuais, estamos nos referindo à:

$$\begin{aligned} \mathbf{X}_{(p \times 1)} &= \boldsymbol{\mu}_{(p \times 1)} + \mathbf{L}_{(p \times m)} \mathbf{F}_{(m \times 1)} + \boldsymbol{\xi}_{(p \times 1)} \\ \mathcal{E}(\mathbf{F}) &= \mathbf{0}_{(m \times 1)}; \text{Cov}(\mathbf{F}) = \mathbf{I}_m; \mathcal{E}(\boldsymbol{\xi}) = \mathbf{0}_{(p \times 1)}; \text{Cov}(\boldsymbol{\xi}) = \boldsymbol{\Psi}; \text{Cov}(\mathbf{F}, \boldsymbol{\xi}) = \mathbf{0}_{(m \times p)}; \\ \boldsymbol{\Psi} &= \begin{bmatrix} \psi_1 & 0 & \dots & 0 \\ 0 & \psi_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \psi_p \end{bmatrix} \end{aligned}$$

Faça uma excelente prova!!

Questões

1. Seja $\mathbf{X} \sim N_p(\mathbf{0}, \mathbf{\Sigma})$, $\mathbf{\Sigma}$ conhecida e positiva definida, e $\boldsymbol{\rho}$ a respectiva matriz de correlações. Sejam $(\lambda'_1, \mathbf{e}'_1)', (\lambda'_2, \mathbf{e}'_2)', \dots, (\lambda'_p, \mathbf{e}'_p)'$, os respectivos autovalores e autovetores ortonormalizados obtidos à partir de $\boldsymbol{\rho}$, $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$. Defina $\mathbf{Y}_{(p \times 1)}$ as componentes principais associadas à \mathbf{X} . Responda os itens:

- a) Escreva, em forma matricial, as equações que relacionam as componentes principais com as variáveis originais padronizadas, definindo adequadamente a matriz que determina tal relação (0,5 pontos).
- b) Qual a distribuição do vetor \mathbf{Y} ? Justifique adequadamente sua resposta (0,5 pontos).
- c) As componentes do vetor \mathbf{Y} são mutuamente independentes? Justifique adequadamente sua resposta (0,5 pontos).
- d) Qual a distribuição de $W = \sum_{i=1}^p \left(\frac{Y_i}{\sqrt{\lambda_i}} \right)^2$? Justifique adequadamente sua resposta (0,5 pontos).

2. Considere $\mathbf{X}_j \sim N_p(\boldsymbol{\mu}, \mathbf{\Sigma})$, $j = 1, \dots, n$, uma amostra aleatória, em que $(\boldsymbol{\mu}, \mathbf{\Sigma})$ são ambos desconhecidos e o MAFO, com as suposições usuais. Desejamos testar a validade do MAFO, ou seja, desejamos testar $H_0 : \mathbf{\Sigma} = \mathbf{L}\mathbf{L}' + \mathbf{\Psi}$ vs $H_1 : \mathbf{\Sigma} \neq \mathbf{L}\mathbf{L}' + \mathbf{\Psi}$. Responda os itens:

- a) Prove que a estatística do teste da razão de verossimilhanças pode ser escrita como (1 ponto):

$$\Lambda = \left(\frac{|\widehat{\mathbf{\Sigma}}_0|}{|\widehat{\mathbf{\Sigma}}_1|} \right)^{-n/2},$$

em que $\widehat{\mathbf{\Sigma}}_0 = \widehat{\mathbf{L}}\widehat{\mathbf{L}}' + \widehat{\mathbf{\Psi}}$ e $\widehat{\mathbf{\Sigma}}_1 = \frac{1}{n} \sum_{i=1}^n (\mathbf{X}_i - \overline{\mathbf{X}})(\mathbf{X}_i - \overline{\mathbf{X}})'$. Sugestão: admita que $tr(\widehat{\mathbf{\Sigma}}_0^{-1} \widehat{\mathbf{\Sigma}}_1) = p$.

- b) Qual a distribuição assintótica de $\Lambda^* = -2 \ln \Lambda$? Justifique adequadamente sua resposta (1 ponto).
- c) Desenhe a região do teste baseado em Λ^* , definindo, claramente, as regiões crítica e de aceitação (0,5 pontos).
- d) Obtenha o nível descritivo (aproximado) do teste em questão baseado em Λ^* (0,5 pontos).

3. Considere uma matriz de dados $\mathbf{X}_{(60 \times 6)}$, em que as variáveis representam as notas de indivíduos em testes (provas) nas seguintes áreas de conhecimento: Geometria, Estatística, Física, Química, Português e Inglês. Com base na matriz de correlações amostrais \mathbf{R} , calculada a partir de $\mathbf{X}_{(60 \times 6)}$, obtivemos o seguinte resultado no ajuste de um MAFO, através do método de máxima verossimilhança:

Tabela 1:

| Variável | Cargas Fatoriais (não rotacionadas) | | Comunalidade (h_i) | Variância específica (ψ_i) |
|-------------|--|---------|------------------------|-----------------------------------|
| | Fator 1 | Fator 2 | | |
| Geometria | 0,919 | 0,170 | | |
| Estatística | 0,863 | 0,241 | | |
| Física | 0,857 | 0,163 | | |
| Química | 0,920 | 0,230 | | |
| Português | 0,189 | 0,979 | | |
| Inglês | 0,263 | 0,848 | | |

Tabela 2:

| Estadística | Fator 1 | Fator 2 |
|--|---------|---------|
| Soma dos quadrados das cargas fatoriais | 3,28 | 1,84 |
| Proporção da variância explicada (%) | 54,58 | 30,73 |
| Proporção acumulada da variância explicada (%) | 54,58 | 85,32 |

A estatística Λ^* do TRV observada, para testar $H_0 : \Sigma = \mathbf{L}\mathbf{L}' + \Psi$ vs $H_1 : \Sigma \neq \mathbf{L}\mathbf{L}' + \Psi$, foi $\Lambda_c^* = 4,65$ com p-valor = 0,3260.

Responda os itens:

- Complete, adequadamente, a Tabela 1, apresentando os respectivos cálculos (1 pontos).
 - Escreva, adequadamente, as equações ajustadas do MAFO (0,5 pontos).
 - Você acha, com base nas informações apresentadas, que o MAFO se ajusta bem ao conjunto de dados? Justifique, adequadamente, seus comentários, utilizando o maior número de informações possíveis. Utilize $\alpha = 5\%$ (1 pontos).
 - Interprete as cargas fatoriais obtidas, em termos das variáveis originais. O que cada um dos fatores representa? (0,5 pontos)
 - Nesse caso, você acha que é necessário rotacionar os fatores? Justifique adequadamente sua resposta (0,5 pontos).
4. Explique, com suas palavras, os principais aspectos da metodologia de componentes principais. Quais são os principais objetivos, aspectos positivos, aspectos negativos e utilidades? (1,5 pontos)

Formulário

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