ME - 731 Análise Multivariada
Segundo semestre de 2009
Lista de Exercícios IV
Entrega: Exercícios 2, 4 e 5 em 29/10/2009

Obs1: Não é necessário digitar a resolução da lista, os exercícios podem ser entregues feitos à mão.

Obs2: Quando falarmos do modelo de análise fatorial ortogonal (MAFO), com as suposições usuais, estamos nos referindo à:

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

Exercícios

1. Considere $\boldsymbol{X}_j \sim N_p(\boldsymbol{\mu},\boldsymbol{\Sigma}), j = 1,...,n$, uma amostra aleatória em que $(\boldsymbol{\mu},\boldsymbol{\Sigma})$ são ambos desconhecidos. Mostre que o e.m.v de $\boldsymbol{\Sigma}$, digamos $\widehat{\boldsymbol{\Sigma}}$, satisfaz à seguinte equação:

$$tr[\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{S}_n] = p$$

em que $\boldsymbol{S}_n = \frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{X}_i - \overline{\boldsymbol{X}}\right)\left(\boldsymbol{X}_i - \overline{\boldsymbol{X}}\right)'$.

2. Considere $\boldsymbol{X}_j \sim N_p(\boldsymbol{\mu},\boldsymbol{\Sigma}), j = 1,...,n$, uma amostra aleatória em que $(\boldsymbol{\mu},\boldsymbol{\Sigma})$ são ambos desconhecidos e o MAFO, com as suposições usuais. Obtenha a estatística do TRV, da forma mais resumida possível, para testar $H_0 : \boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}' + \boldsymbol{\Psi}$ vs $H_1 : \boldsymbol{\Sigma} \neq \boldsymbol{L}\boldsymbol{L}' + \boldsymbol{\Psi}$. Sugestão: Denote, em seu problema, por $\widehat{\boldsymbol{L}}$ e $\widehat{\boldsymbol{\Psi}}$ os estimadores de máxima verossimilhança de $\boldsymbol{L}$ e $\boldsymbol{\Psi}$, respectivamente. Além, denote por $\widehat{\boldsymbol{\Sigma}} = \widehat{\boldsymbol{L}}\widehat{\boldsymbol{L}}' + \widehat{\boldsymbol{\Psi}}$, o e.m.v de $\boldsymbol{\Sigma}$.

3. Considere o TRV (o teste que você obteve na Questão anterior) para testar $H_0 : \boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}' + \boldsymbol{\Psi}$ vs $H_1 : \boldsymbol{\Sigma} \neq \boldsymbol{L}\boldsymbol{L}' + \boldsymbol{\Psi}$, no MAFO. Sabemos que a estatística associada ao

TRV, digamos Λ, é tal que $\Lambda \approx \chi^2_r$. Prove que $r = \frac{1}{2}\left[(p-m)^2 - p - m\right]$. Além disso, ontenha o nível descritivo (valor p) e a respectiva função poder do teste, à um nível de significância de α.

4. Considere o MAFO com as suposições usuais ($\boldsymbol{L}$ e $\boldsymbol{\Psi}$ conhecidos) e os preditores dos escores fatoriais vistos em sala de aula. Ou seja, $\widehat{\boldsymbol{F}}_{MQP} = \left(\boldsymbol{L}'\boldsymbol{\Psi}^{-1}\boldsymbol{L}\right)^{-1}\boldsymbol{L}'\boldsymbol{\Psi}^{-1}\left(\boldsymbol{X} - \boldsymbol{\mu}\right)$ e $\widehat{\boldsymbol{F}}_R = \boldsymbol{L}\left(\boldsymbol{L}\boldsymbol{L}' + \boldsymbol{\Psi}\right)^{-1}\left(\boldsymbol{X} - \boldsymbol{\mu}\right)$.

   Responda os itens:

   a) Calcule, no MAFO, $\mathcal{E}(\boldsymbol{X}|\boldsymbol{F})$ e $Cov(X|\boldsymbol{F})$.

   b) Calcule os valores esperados de $\widehat{\boldsymbol{F}}_{MQP}$ e $\widehat{\boldsymbol{F}}_R$ bem como suas matrizes de covariâncias condicionadas ao vetor $\boldsymbol{F}$ $[\mathcal{E}(\widehat{\boldsymbol{F}}|\boldsymbol{F})$ e $Cov(\widehat{\boldsymbol{F}}|\boldsymbol{F})]$.

   c) Compare os dois estimadores (preditores) considerados utilizando as quantidades que você calculou no item b). Qual dos estimadores você escolheria? Justifique, adequadamente, sua resposta.

   d) Refaça os itens a), b) e c), considerando o vetor de variáveis padronizadas $\boldsymbol{Z} = \boldsymbol{D}^{-1}\left(\boldsymbol{X} - \boldsymbol{\mu}\right)$.

5. Com base no conjunto de dados sobre informações nutricionais de cereais (o arquivo está disponível na página do curso sob o nome de *cereal.txt*), e considerando a matriz de correlações, utilize a metodologia do MAFO para responder os seguintes itens:

   a) Calcule os autovalores e autovetores (ortonormalizados) associados à $\boldsymbol{\Sigma}$.

   b) Considere 2 fatores : calcule a matriz de cargas fatoriais (a partir do método das componentes principais, veja item a)), as comunalidades e as variâncias específicas. Você acha que 2 fatores são suficientes para explicar de modo razoável a estrutura de covariâncias? Utilize os autovalores e o cálculo da matriz $\widetilde{\boldsymbol{L}}\widetilde{\boldsymbol{L}}' + \widetilde{\boldsymbol{\Psi}}$. Interprete os fatores que você obteve.

   c) Através do método de máxima verossimilhança, ajuste um MAFO com dois fatores. Calcule a matriz de cargas fatoriais, as comunalidades e as variâncias específicas. Você acha que 2 fatores são suficientes para explicar de modo razoável a estrutura de covariâncias? Utilize o cálculo da matriz $\widetilde{\boldsymbol{L}}\widetilde{\boldsymbol{L}}' + \widetilde{\boldsymbol{\Psi}}$ e a estatística do TRV. Interprete os fatores que você obteve.

   d) Considere a rotação varimax dos fatores obtidos no item c). Interprete os fatores que você obteve.

e) Calcule os escores fatoriais, através do método dos mínimos quadrados ponderados (Bartlett). Faça um gráfico de dipersão entre as componentes do item d). Você conseguiria identificar grupos distintos entre as fabricantes de cereal? Como se caraterizam estes grupos, em termos das componentes principais e em termos das variáveis originais? Sugestão: Calcule medidas descritivas para os escores fatoriais e para as variáveis originais em cada um desses grupos.