1ª prova – 2º/2015

1. Os autovalores e autovetores de uma matriz de correlações amostral são

$$\begin{aligned}\widehat{\lambda}_1 &= 1,96, \quad \widehat{\boldsymbol{e}}_1 = (0,625;\ 0,593;\ 0,507)',\\ \widehat{\lambda}_2 &= 0,68, \quad \widehat{\boldsymbol{e}}_2 = (-0,219;\ -0,491;\ 0,843)',\\ \widehat{\lambda}_3 &= 0,36 \quad \text{e} \quad \widehat{\boldsymbol{e}}_3 = (0,749;\ -0,638;\ -0,177)'.\end{aligned}$$

   Adotando um modelo com um único fator comum, apresente estimativas das cargas fatoriais, das comunalidades e das variâncias específicas.
   Solução. Aplicando o método dos componentes principais a um modelo com $k = 1$ fator, a estimativa da matriz (neste caso, vetor $p \times 1$) de cargas fatoriais é dada por

$$\widehat{\boldsymbol{\Lambda}} = \widehat{\lambda}_1^{1/2}\,\widehat{\boldsymbol{e}}_1 = \sqrt{1,96} \times \begin{pmatrix}0,625\\0,593\\0,507\end{pmatrix} = \begin{pmatrix}0,875\\0,830\\0,710\end{pmatrix}. \qquad (1)$$

   A partir de (1) obtemos as estimativas das comunalidades, dadas por $\widehat{h}_1^2 = 0,875^2 = 0,766$, $\widehat{h}_2^2 = 0,830^2 = 0,689$ e $\widehat{h}_3^2 = 0,710^2 = 0,504$. Uma vez que foi utilizada a matriz de correlações amostrais, as estimativas das variâncias específicas são $\widehat{\psi}_1 = 1 - \widehat{h}_1^2 = 0,234$, $\widehat{\psi}_2 = 1 - \widehat{h}_2^2 = 0,311$ e $\widehat{\psi}_3 = 1 - \widehat{h}_3^2 = 0,496$.

2. O vetor $\boldsymbol{X}$ com distribuição $\mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ é dividido como

$$\boldsymbol{X} = \begin{pmatrix}\boldsymbol{Y}\\ \boldsymbol{Z}\end{pmatrix}, \text{ com dimensões } q \times 1\ (\boldsymbol{Y}) \text{ e } (p-q) \times 1\ (\boldsymbol{Z}),\ \ p > 1 \text{ e } 1 \le q < p.$$

   O vetor de médias e a matriz de covariâncias são divididos como

$$\boldsymbol{\mu} = \begin{pmatrix}\boldsymbol{\mu}_Y\\ \boldsymbol{\mu}_Z\end{pmatrix} \quad \text{e} \quad \boldsymbol{\Sigma} = \begin{bmatrix}\boldsymbol{\Sigma}_{YY} & \boldsymbol{\Sigma}_{YZ}\\ \boldsymbol{\Sigma}'_{YZ} & \boldsymbol{\Sigma}_{ZZ}\end{bmatrix}.$$

   $\boldsymbol{A}$ e $\boldsymbol{B}$ são matrizes $q \times q$ e $q \times (p-q)$, respectivamente. Determine a distribuição de $\boldsymbol{AY} + \boldsymbol{BZ}$.
   Solução. Escrevemos

$$\boldsymbol{AY} + \boldsymbol{BZ} = \begin{bmatrix}\boldsymbol{A}, & \boldsymbol{B}\end{bmatrix}\begin{pmatrix}\boldsymbol{Y}\\ \boldsymbol{Z}\end{pmatrix} = \boldsymbol{MX},$$

   em que $\boldsymbol{M} = [\boldsymbol{A}, \boldsymbol{B}]$ é uma matriz $q \times p$. Portanto, $\boldsymbol{AY} + \boldsymbol{BZ}$ é um vetor de q combinações lineares de $\boldsymbol{X}$. Logo, a distribuição de $\boldsymbol{AY} + \boldsymbol{BZ}$ é normal q-variada com vetor de médias

$$\boldsymbol{M\mu} = \begin{bmatrix}\boldsymbol{A}, & \boldsymbol{B}\end{bmatrix}\begin{pmatrix}\boldsymbol{\mu}_Y\\ \boldsymbol{\mu}_Z\end{pmatrix} = \boldsymbol{A\mu}_Y + \boldsymbol{B\mu}_Z$$

   e matriz de covariâncias

$$\begin{aligned}\boldsymbol{M\Sigma M}' &= \begin{bmatrix}\boldsymbol{A}, & \boldsymbol{B}\end{bmatrix}\begin{bmatrix}\boldsymbol{\Sigma}_{YY} & \boldsymbol{\Sigma}_{YZ}\\ \boldsymbol{\Sigma}'_{YZ} & \boldsymbol{\Sigma}_{ZZ}\end{bmatrix}\begin{bmatrix}\boldsymbol{A}'\\ \boldsymbol{B}'\end{bmatrix}\\ &= \begin{bmatrix}\boldsymbol{A\Sigma}_{YY} + \boldsymbol{B\Sigma}'_{YZ}, & \boldsymbol{A\Sigma}_{YZ} + \boldsymbol{B\Sigma}_{ZZ}\end{bmatrix}\begin{bmatrix}\boldsymbol{A}'\\ \boldsymbol{B}'\end{bmatrix}\\ &= \boldsymbol{A\Sigma}_{YY}\boldsymbol{A}' + \boldsymbol{B\Sigma}'_{YZ}\boldsymbol{A}' + \boldsymbol{A\Sigma}_{YZ}\boldsymbol{B}' + \boldsymbol{B\Sigma}_{ZZ}\boldsymbol{B}'.\end{aligned}$$

3. (a) O vetor $\boldsymbol{X} = (X_1, X_2)'$ tem função densidade $f(x_1, x_2) = C$, se $x_1^2 + x_2^2 \le k^2$, e $f(x_1, x_2) = 0$, caso contrário. Determine o valor de C.
   Solução. Definimos $A = \{(x_1, x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 \le k^2\}$. Como $f(x_1, x_2)$ é uma função densidade, devemos ter

$$\begin{aligned}1 &= \iint_{\mathbb{R}^2} f(x_1, x_2)dx_1dx_2 = \iint_A Cdx_1dx_2 = C\iint_A dx_1dx_2\\ &= C \times \text{área(A)} = C\pi k^2.\end{aligned}$$

   Logo, $C = \dfrac{1}{\pi k^2}$.

(b) Afirma-se que

$$f(\boldsymbol{x}) = \frac{1}{4\sqrt{3\pi^3}} \exp\left\{ -\left[\frac{(x_1-1)^2}{6} + \frac{(x_2-2)^2}{4} + \frac{x_3^2}{2} \right] \right\}, \ \boldsymbol{x} \in \mathbb{R}^3,$$

representa a função densidade de um vetor aleatório $\boldsymbol{X} = (X_1, X_2, X_3)'$ com distribuição $\mathrm{N}_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Determine $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$. As variáveis são independentes?
Solução. Como $p = 3$, $f(\boldsymbol{x})$ deve ser escrita como

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{3/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) \right].$$

Iniciamos simplificando as parcelas entre colchetes na expressão de $f(\boldsymbol{x})$ do enunciado, obtendo

$$f(\boldsymbol{x}) = \frac{1}{4\sqrt{3\pi^3}} \exp\left\{ -\frac{1}{2}\left[\frac{(x_1-1)^2}{3} + \frac{(x_2-2)^2}{2} + x_3^2 \right] \right\}, \ \boldsymbol{x} \in \mathbb{R}^3. \tag{2}$$

A expressão entre colchetes na eq. (2) corresponde a $(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$. Notando que na eq. (2) os coeficientes de $x_j x_l$, para $j \neq l$ são nulos, temos que $\boldsymbol{\mu} = (1,2,0)'$, $\boldsymbol{\Sigma}^{-1} = \mathrm{diag}(1/3, 1/2, 1)$, de modo que $\boldsymbol{\Sigma} = \mathrm{diag}(3,2,1)$, $|\boldsymbol{\Sigma}| = 6$ e $(2\pi)^{3/2}|\boldsymbol{\Sigma}|^{1/2} = 4\sqrt{3\pi^3}$. As variáveis são independentes, pois a distribuição é normal e a matriz de covariâncias $\boldsymbol{\Sigma}$ é diagonal.

4. Em uma aplicação de análise de componentes principais foram obtidas as estimativas dos autovalores da matriz de covariâncias amostral dadas abaixo.

2,53 2,12 1,18 0,54 0,42 0,20 0,01

Apresente o número de componentes principais que você sugere que sejam utilizados na análise.
Solução. Observando as proporções da variância total acumuladas, apresentadas abaixo, percebemos que os três primeiros componentes correspondem a mais de 80% da variância total. O gráfico de escarpa abaixo apresenta um ponto de “cotovelo” no quarto componente. Por simplicidade, e lembrando que o número de variáveis é 7, três componentes poderiam ser utilizados.

	Componente principal						
	1	2	3	4	5	6	7
Proporção da variância	0.361	0.303	0.169	0.077	0.060	0.029	0.001
Proporção acumulada	0.361	0.664	0.833	0.910	0.970	0.999	1.000