

Exercício 1. Seja \underline{X} um vetor aleatório $p \times 1$ com $E(\underline{X}) = \underline{\mu}$ e $\text{Var}(\underline{X}) = \Sigma$.

- (a) Determine as componentes principais Y_1 e Y_2 para a matriz de covariância

$$\Sigma = \begin{bmatrix} 5 & 2 \\ 2 & 2 \end{bmatrix}.$$

Calcule a proporção da variância total populacional explicada pela primeira componente principal.

- (b) Converta a matriz Σ para a matriz de correlação P . Determine as componentes principais Y_1 e Y_2 de P e calcule a proporção total da variância explicada por Y_1 .

Exercício 2. Obtenha as componentes principais e as proporções de variação total explicadas por cada componente de um vetor aleatório X com variância

$$\Sigma = \begin{bmatrix} \sigma^2 & \sigma^2\rho & 0 \\ \sigma^2\rho & \sigma^2 & \sigma^2\rho \\ 0 & \sigma^2\rho & \sigma^2 \end{bmatrix}, \quad \frac{-1}{\sqrt{2}} < \rho < \frac{1}{\sqrt{2}}.$$

Exercício 3. Os dados da Tabela 8.6 de Johnson e Wichern (2007) descrevem recordes nacionais de provas masculinas de corrida em 54 países: 100m (s), 200m (s), 400m (s), 800m (min), 1500m (min), 5000m (min), 10000m (min), maratona (min).

- (a) Realize uma análise de componentes principais considerando a matriz de variâncias e covariâncias amostrais. Escolha o número de componentes principais e justifique sua escolha. Interprete a(s) componente(s) obtida(s).
- (b) Repita o item (a) considerando a matriz de correlação amostral.

Exercício 4. O conjunto de dados **data (decathlon)** disponível na **library (FactoMineR)** do R contém dados de 41 participações em competições de decatlo, com o nome do atleta na primeira coluna e a competição (Decastar 2004 e Jogos Olímpicos 2004) na última. As provas são apresentadas na seguinte ordem: salto em distância, arremesso de peso, salto em altura, 400 metros rasos, 110 metros com barreiras, lançamento de disco, salto com vara, lançamento de dardo e 1500 metros.

- (a) Apresente uma breve análise exploratória dos dados.
- (b) Realize uma análise de componentes principais. Justifique a escolha do número de componentes finais e interprete-as.
- (c) Proponha um ranking para os atletas.

Exercício 5. Estudos mostram que grande parte de adultos e adolescentes norte-americanos usam regularmente substâncias psicoativas. Em um destes estudos (Huba et al. 1981, J. of Personality and Social Psychology), dados foram coletados de 1634 estudantes na área metropolitana de Los Angeles. Cada participante completou um questionário informando o número de vezes que cada item foi usado. Os itens são os seguintes: cigarro, cerveja, vinho, licor, cocaína, tranquilizantes, medicamentos, heroína, maconha, haxixe, inalantes, alucinógenos e anfetaminas. As respostas foram registradas em uma escala de cinco pontos: 1. nunca experimentei, 2. apenas uma vez, 3. poucas vezes, 4. muitas vezes e 5. regularmente.

O arquivo com a matriz de correlações das respostas encontra-se disponível na Coteia WIKI.

- (a) Analise a matriz de correlações dos dados.
- (b) Proponha um “índice de utilização de substâncias psicoativas”.

Referência

Johnson, R. A. and Wichern, D. W. (2007) Applied Multivariate Statistical Analysis. 5th edition. Prentice-Hall.