

Universidade Federal de Alfenas - UNIFAL-MG - *campus* Varginha
Bacharelado Interdisciplinar em Ciência e Economia
Disciplina: Análise multivariada - Profa. Patrícia de Siqueira Ramos
Lista 3 - Componentes principais

- Resolva as questões 1, 2 e 3 da lista manualmente (pode ser em dupla).
- Faça no `python` a questão 4 (as outras também podem ser feitas no `python` para conferir).

1 Seja a matriz de covariâncias amostral:

$$\mathbf{S} = \begin{bmatrix} 5 & 2 \\ 2 & 2 \end{bmatrix},$$

- a) Determine os componentes principais Y_1 e Y_2 a partir de \mathbf{S} ;
- b) Obtenha a proporção da variância total explicada pelo primeiro componente principal;
- c) Comprove que $tr(\mathbf{S}) = \sum_{i=1}^p \lambda_i$.

2 Converta a matriz \mathbf{S} do exercício 1 na matriz \mathbf{R} .

- a) Determine os componentes principais Y_1 e Y_2 a partir de \mathbf{R} e calcule a proporção da variância total explicada pelo primeiro componente principal;
- b) Compare os componentes obtidos na letra a dos exercícios 1 e 2 e informe: Eles são iguais? Deveriam ser?
- c) Calcule as correlações r_{Y_1, Z_1} , r_{Y_1, Z_2} , r_{Y_2, Z_1} e r_{Y_2, Z_2} . O que essas correlações nos informam?

3 Seja

$$\mathbf{\Sigma} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix}.$$

- a) Determine os componentes principais a partir de $\mathbf{\Sigma}$;
- b) Quais os valores de λ_i e o que eles significam?
- c) O que é possível dizer sobre os componentes principais obtidos?

4 (no `Python`!) Considere as informações sobre produção de origem animal dos estados do Brasil em 2016 (fonte: <https://sidra.ibge.gov.br>). As variáveis são: leite (milhares de litros), ovos de galinha (milhares de dúzias) e mel de abelha (kg). Os dados estão disponibilizados no arquivo ‘prod-animal-2016’ (versão xlsx e csv). Efetue toda a análise de componentes principais (a partir da matriz \mathbf{R}) dos dados usando o `python` e responda o que se pede:

- a) Expresse os componentes principais obtidos como uma combinação linear das variáveis.
- b) Obtenha o *scree plot* (gráfico com a variação explicada por cada CP e a acumulada) e o *biplot* (gráfico com os escores dos dois primeiros CPs e as correlações dos CPs com as variáveis).
- c) Qual a proporção da variância total é contabilizada pelos dois primeiros componentes principais?
- d) Quantos componentes principais deveriam ser usados? Informe o critério utilizado.
- e) Interprete os componentes principais que foram considerados por você (analisando os valores dos coeficientes).
- f) Analise quais os estados apresentam os maiores e menores escores do CP1 e interprete.