

Universidade Federal de Alfenas - UNIFAL-MG - *campus* Varginha
Bacharelado Interdisciplinar em Ciência e Economia
Disciplina: Análise multivariada - Profa. Patrícia de Siqueira Ramos
Lista 2 - Amostras aleatórias

Essa lista está dividida em duas partes.

- Manual: deverá ser resolvida e entregue escrita à mão de forma individual ou em dupla mostrando os cálculos utilizados
- Python: quando aparecer (Python), tal questão será resolvida usando o programa Python e deverá ser usado o mesmo **notebook** onde a lista 1 foi resolvida.

1 Quatro recibos de uma livraria foram selecionados para avaliar a natureza das vendas de livros. Há informações sobre o número de livros vendidos e o valor total, dentre outras. Os valores das vendas (US\$) foram 42, 52, 48, 58, enquanto o número de livros vendidos em cada foi 4, 5, 4, 3.

- a) Escreva a matriz de dados \mathbf{X} correspondente.
- b) Obtenha $\bar{\mathbf{X}}$, \mathbf{S} , \mathbf{R} e \mathbf{W} .
- c) (Python!) Resolva o item b no Python usando as funções `mean`, `cov` e `corrcoef` da forma vista na aula prática.

2 (Python!) Obtenha as matrizes de covariâncias amostrais e de correlações amostrais do seguinte conjunto de dados:

indivíduo	sexo	idade	QI	depressão	indivíduo	sexo	idade	QI	depressão
1	F	19	119	N	13	F	15	122	S
2	M	34	82	S	14	M	51	99	S
3	F	32	106	N	15	M	55	85	N
4	F	37	134	N	16	M	32	106	S
5	M	39	100	S	17	F	27	93	N
6	F	30	87	S	18	M	59	109	N
7	F	14	97	N	19	M	13	97	N
8	M	54	84	N	20	F	39	89	N
9	M	22	95	S	21	F	35	80	N
10	M	37	96	N	22	F	38	127	S
11	M	42	101	S	23	F	42	107	N
12	M	16	110	S	24	F	21	85	N

Dica: o conjunto de dados acima já está salvo num arquivo ('depressao.csv'). Pra carregá-lo no Python use:

```
import pandas as pd
# carregar o conjunto de dados usando as formas vistas na aula prática
# pode ser necessário corrigir as aspas

dados = pd.read_csv('https://patriciasiqueira.github.io/arquivos/depressao.csv')
dados = dados.iloc[:, 2:4] # selecionar apenas variáveis numéricas
dados.head() # visualizar as primeiras observações
```

3 Considere a matriz de covariâncias abaixo:

$$\mathbf{S} = \begin{bmatrix} 3,8778 & 2,8110 & 3,1480 & 3,5062 \\ 2,8110 & 2,1210 & 2,2669 & 2,5690 \\ 3,1480 & 2,2669 & 2,6550 & 2,8341 \\ 3,5062 & 2,5690 & 2,8341 & 3,2352 \end{bmatrix}.$$

- a) (Python!) Obtenha a matriz \mathbf{R} a partir de \mathbf{S} utilizando o fato: $\mathbf{R} = \mathbf{D}^{-1/2} \mathbf{S} \mathbf{D}^{-1/2}$.
 b) Como você avalia as correlações entre os pares de variáveis?

4 Considerando os vetores $\mathbf{X}_{1\bullet}$, $\mathbf{X}_{2\bullet}$, $\mathbf{X}_{3\bullet}$, $\mathbf{X}_{4\bullet}$ e $\mathbf{X}_{5\bullet}$:

$$\mathbf{X}_{1\bullet} = \begin{bmatrix} 9 \\ 12 \\ 3 \end{bmatrix}, \mathbf{X}_{2\bullet} = \begin{bmatrix} 2 \\ 8 \\ 4 \end{bmatrix}, \mathbf{X}_{3\bullet} = \begin{bmatrix} 6 \\ 6 \\ 0 \end{bmatrix}, \mathbf{X}_{4\bullet} = \begin{bmatrix} 5 \\ 4 \\ 2 \end{bmatrix}, \mathbf{X}_{5\bullet} = \begin{bmatrix} 8 \\ 10 \\ 1 \end{bmatrix}.$$

Dica: para criar uma matriz no Python a partir de vetores prontos, por exemplo, uma matriz de dimensão 3×2 a partir de 3 vetores 2×1 (a , b e c), pode-se usar `np.vstack((a, b, c)).reshape(3, 2)`.

Obtenha:

- a) $\bar{\mathbf{X}}$, \mathbf{S} e \mathbf{R} (no Python e manualmente).
 b) Como você avalia as correlações entre as variáveis?

5 Considere a matriz de dados \mathbf{X} abaixo (que já se encontra no arquivo 'matrizX.csv' no meu site):

$$\begin{pmatrix} 3 & 6 & 4 & 0 & 7 \\ 4 & 2 & 7 & 4 & 6 \\ 4 & 0 & 3 & 1 & 5 \\ 6 & 2 & 6 & 1 & 1 \\ 1 & 6 & 2 & 1 & 4 \\ 5 & 1 & 2 & 0 & 2 \\ 1 & 1 & 2 & 6 & 1 \\ 1 & 1 & 5 & 4 & 4 \\ 7 & 0 & 1 & 3 & 3 \\ 3 & 3 & 0 & 5 & 1 \end{pmatrix}.$$

Antes:

```
# pode ser necessário corrigir as aspas do endereço
matriz = pd.read_csv('https://patriciasiqueira.github.io/arquivos/matrizX.csv', sep=';')
X = matriz.values[:, :] # valores numéricos
X = np.asarray(X, dtype=float) # transformar em matriz de reais
```

- a) (Python!) Obtenha as matrizes distância euclidiana, euclidiana padronizada e Mahalanobis entre as observações.
 b) Se apenas houvesse as quatro primeiras observações e as três primeiras variáveis, calcule manualmente todas as matrizes distância (use o Python para obter a matriz de covariâncias amostrais \mathbf{S} e sua inversa \mathbf{S}^{-1}).