



1. Johnson e Wichern - Exercício 8.03.
2. Johnson e Wichern - Exercício 8.06.
3. Johnson e Wichern - Exercício 8.07.
4. Johnson e Wichern - Exercício 8.11. - Dados no arquivo Census-tract (T8-5.DAT).

\mathbf{X}_1 : Total Population (thousands); \mathbf{X}_2 : Median School Years;
 \mathbf{X}_3 : Total Employment (thousands); \mathbf{X}_4 : Health Services Employment (hundreds);
 \mathbf{X}_5 : Median Value Home (\$10,000s).

5. **Johnson e Wichern - Exercício 8.12.** Dados no arquivo Air Pollution (T1-5.DAT). Os dados correspondem a 42 medidas de poluição do ar observadas na área de Los Angeles em um mesmo horário.

\mathbf{X}_1 : Wind; \mathbf{X}_2 : Solar Radiation; \mathbf{X}_3 : CO; \mathbf{X}_4 : NO;
 \mathbf{X}_5 : NO₂; \mathbf{X}_6 : O₃; \mathbf{X}_7 : HC.

6. Izenman (2008) apresenta um estudo sobre dados nutricionais que analisa 961 alimentos. Os componentes nutricionais observados foram: gordura (gramas), energia (calorias), carboidratos (gramas), proteínas (gramas), colesterol (miligramas), peso (gramas) e gordura saturada (gramas). Os alimentos são listados em porções variadas e por isso as porções são divididas pelo peso correspondente de cada item.

A matriz de variância dos dados é (colunas referentes a gordura, calorias, carboidratos, proteínas, colesterol e gordura saturada),

$$\mathbf{S} = \begin{pmatrix} 0.037 & 0.306 & -0.007 & 0.003 & 0.023 & 0.010 \\ 0.306 & 3.747 & 0.157 & 0.045 & 0.178 & 0.082 \\ -0.007 & 0.157 & 0.062 & -0.002 & -0.027 & -0.002 \\ 0.003 & 0.045 & -0.002 & 0.008 & 0.020 & 0.001 \\ 0.023 & 0.178 & -0.027 & 0.020 & 0.456 & 0.014 \\ 0.010 & 0.082 & -0.002 & 0.001 & 0.014 & 0.004 \end{pmatrix}$$

com autovalores $\lambda_1^v = 3.790$, $\lambda_2^v = 0.451$, $\lambda_3^v = 0.061$, $\lambda_4^v = 0.008$, $\lambda_5^v = 0.004$, $\lambda_6^v = 0.002$ e respectivos autovetores (v_i),

Item	v_1	v_2	v_3	v_4	v_5	v_6
Gord	-0.081	-0.019	0.364	-0.506	0.759	-0.167
Calo	-0.994	0.052	0.011	0.044	-0.083	-0.010
Carb	-0.041	0.092	-0.921	-0.186	0.325	0.037
Prot	-0.012	-0.040	0.028	0.827	0.554	0.080
Cole	-0.053	-0.993	-0.095	-0.035	-0.012	-0.019
GorS	-0.022	-0.022	0.093	-0.147	0.071	0.982



A matriz de correlação,

$$\mathbf{R} = \begin{pmatrix} 1.000 & 0.816 & -0.140 & 0.157 & 0.175 & 0.747 \\ 0.816 & 1.000 & 0.324 & 0.261 & 0.136 & 0.643 \\ -0.140 & 0.324 & 1.000 & -0.087 & -0.163 & -0.141 \\ 0.157 & 0.261 & -0.087 & 1.000 & 0.328 & 0.142 \\ 0.175 & 0.136 & -0.163 & 0.328 & 1.000 & 0.311 \\ 0.747 & 0.643 & -0.141 & 0.142 & 0.311 & 1.000 \end{pmatrix}$$

com autovalores $\lambda_1^c = 2.649$, $\lambda_2^c = 1.330$, $\lambda_3^c = 1.020$, $\lambda_4^c = 0.680$, $\lambda_5^c = 0.267$, $\lambda_6^c = 0.055$ e respectivos autovetores (c_i),

Item	c_1	c_2	c_3	c_4	c_5	c_6
Gord	-0.557	0.099	0.275	0.130	-0.455	0.617
Calo	-0.536	0.357	-0.137	0.075	-0.273	-0.697
Carb	0.025	0.672	-0.568	-0.286	0.157	0.344
Prot	-0.235	-0.374	-0.639	0.599	0.154	0.119
Cole	-0.253	-0.521	-0.326	-0.717	-0.210	-0.003
GorS	-0.531	-0.019	0.261	-0.150	0.791	0.022

Ainda são fornecidas as seguintes medidas:

- Variância amostral generalizada: $VAG = 5.011 \times 10^{-9}$
- Variância amostral total: $VAT = 4.314$

Utilizando os resultados apresentados, responda as seguintes questões:

- Sobre a matriz \mathbf{S} :
 - Indique se esta matriz é positiva definida. Justifique.
 - Indique como você faria para obter a decomposição espectral da matriz \mathbf{S} ?
 - Indique um procedimento para obter \mathbf{S}^{-1} e $\sqrt{\mathbf{S}}$ (se existir).
- Indique como obter a variância generalizada e interprete o resultado.
- Qual o motivo da grande diferença entre VAG e VAT? Isto fato é refletido em alguma outra medida apresentada para \mathbf{S} ?
- Qual matriz, variância-covariância ou correlações você usaria para reduzir a dimensionalidade dos dados? Justifique.
- Com base em sua escolha para o item (d), obtenha as duas primeiras componentes principais para os dados e faça o gráfico *Scree Plot*.
- Quantas componentes principais você reteria no estudo? Justifique. Interprete (se possível) as duas primeiras componentes principais.



7. O pacote `kohonen` do R possui dados sobre análise química de 177 vinhos de três tipos de cultivares de uvas (Nebbiolo, Barberas e Grignolino) da região de Piemonte na Itália. O vinho da uva Nebbiolo é chamado de Barolo. A identificação dos vinhos está no objeto `vintages`. O artigo de origem destes dados é

Forina, M.; Armanino, C.; Castino, M. e Ubigli, M. (1986) Multivariate data analysis as a discriminating method of the origin of wines. *Vitis*, 189-201. O arquivo pode ser obtido com os seguintes comandos R:

```
install.packages("kohonen")
library(kohonen)
data("wines")
```

Desconsidere a variável `OD ratio` e faça uma análise de componentes principais deste conjunto de dados.

Para as questões 8 e 9, considere o seguinte enunciado e indique o item correto.

Uma análise de componentes principais foi realizada e os seguintes autovalores foram obtidos da matriz de variância-covariância dos dados: 3.731, 2.218, 0.442, 0.341, 0.183, 0.0085.

8. O número de componentes que retem pelo menos 90% da variação total dos dados é igual a:
- ☐ 1
 - ☐ 2
 - ☐ 3
 - ☐ 4
 - ☐ 5
 - ☐ Não é possível determinar com a informação fornecida.
9. (4 points) Sobre o conjunto de dados original, podemos afirmar que
- ☐ O número total de variáveis do estudo não pode ser determinado.
 - ☐ A variância generalizada é maior que 6.
 - ☐ As variáveis são independentes.
 - ☐ A variância total é menor que 5.
 - ☐ A variância total é igual a $\text{tr}(\text{diag}(3.731, 2.218, 0.442, 0.341, 0.183, 0.0085, 0))$.



10. (3 points) Suponha que uma matriz \mathbf{L} tem dimensão (4×25) . Seja $\mathbf{D} = \text{diag}(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$, em que $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \lambda_4$ são os autovalores de $\mathbf{L}^T \mathbf{L}$. Queremos obter a decomposição em valores singulares de \mathbf{L} , isto é, $\mathbf{L} = \mathbf{U} \mathbf{\Delta} \mathbf{V}^T$. Então, $\mathbf{\Delta}$ será,

☐ $\mathbf{\Delta} = \mathbf{D}$.

☐ $\mathbf{\Delta} = \mathbf{D}^{1/2}$

☐ $\mathbf{\Delta} = \mathbf{I}$

☐ $\mathbf{\Delta} = \mathbf{D}^2$.

☐ $\mathbf{\Delta}$ não pode ser determinado conhecendo \mathbf{D} .