Faculdade de Ciências e Tecnologia
boldsymbolniversidade Nova de Lisboa

Estatística Multivariada

Exercícios 2018/19

FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

1. Sejam $\boldsymbol{A} = \begin{bmatrix} 4 & 2 & 3 \\ 7 & 5 & 8 \end{bmatrix}$ e $\boldsymbol{B} = \begin{bmatrix} 3 & -2 & 4 \\ 6 & 9 & -5 \end{bmatrix}$.

   a) Determine $\boldsymbol{A}+\boldsymbol{B}$, $\boldsymbol{A}-\boldsymbol{B}$.

   b) Determine $\boldsymbol{A}'\boldsymbol{A}$ e $\boldsymbol{A}\boldsymbol{A}'$.

2. Sejam $\boldsymbol{A} = \begin{bmatrix} 1 & 3 \\ 2 & -1 \end{bmatrix}$ e $\boldsymbol{B} = \begin{bmatrix} 2 & 0 \\ 1 & 5 \end{bmatrix}$.

   a) Determine $\boldsymbol{AB}$, $\boldsymbol{BA}$.

   b) Determine $|\boldsymbol{AB}|$, $|\boldsymbol{A}|$ e $|\boldsymbol{B}|$.

3. Sejam $\boldsymbol{A} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 5 & 10 & 15 \end{bmatrix}$ e $\boldsymbol{B} = \begin{bmatrix} -1 & 1 & -2 \\ -1 & 1 & -2 \\ 1 & -1 & 2 \end{bmatrix}$.

   a) Mostre que $\boldsymbol{AB} = \mathbf{0}$.

   b) Encontre um vector $\boldsymbol{x} \neq \mathbf{0}$ tal que $\boldsymbol{Ax} = \mathbf{0}$.

   c) Determine $tr(\boldsymbol{A})$ e $tr(\boldsymbol{B})$.

   d) Mostre que $|\boldsymbol{A}| = 0$.

4. Seja $\boldsymbol{A} = \begin{bmatrix} 3 & 1 & 1 \\ 1 & 0 & 2 \\ 1 & 2 & 0 \end{bmatrix}$

   a) Determine os valores próprios de $\boldsymbol{A}$. A matriz $\boldsymbol{A}$ é definida positiva?

   b) Determine a matriz $\boldsymbol{P}$, constituída pelos vectores próprios normalizados associados a cada um dos valores próprios de $\boldsymbol{A}$.

   c) Calcule $tr(\boldsymbol{A})$ e $|\boldsymbol{A}|$ utilizando o resultado da alínea a).

   d) verifique que $\boldsymbol{P}$ é ortogonal. **P é ortogonal quando A' = A^(-1)**

   e) Determine $\boldsymbol{\Lambda} = \boldsymbol{P}'\boldsymbol{A}\boldsymbol{P}$.

   f) verifique que $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}'$.

5. Seja $\boldsymbol{A} = \begin{bmatrix} 3 & 6 & -1 \\ 6 & 9 & 4 \\ -1 & 4 & 3 \end{bmatrix}$.

   a) Determine a decomposição espectral de $\boldsymbol{A}$.

   b) A matriz $\boldsymbol{A}$ é definida positiva?

6. Sejam $\boldsymbol{x}$, $\boldsymbol{y}$ dois quaisquer vectores com a mesma dimensão. Mostre que $(\boldsymbol{x}-\boldsymbol{y})'(\boldsymbol{x}-\boldsymbol{y}) = \boldsymbol{x}'\boldsymbol{x} - 2\boldsymbol{x}'\boldsymbol{y} + \boldsymbol{x}'\boldsymbol{y}$.

7. Sejam $\boldsymbol{A}$ uma qualquer matriz do tipo $n \times p$. Mostre que $\boldsymbol{A}'\boldsymbol{A}$ é simétrica.

8. Seja $\boldsymbol{A}$ e $\boldsymbol{B}$ duas quaisquer matrizes quadradas invertíveis.

   a) Mostre que $(\boldsymbol{A}')^{-1} = \left(\boldsymbol{A}^{-1}\right)'$.

   b) $(\boldsymbol{A}\boldsymbol{B})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$.

9. Seja $\boldsymbol{A} = \begin{bmatrix} 9 & -2 \\ -2 & 9 \end{bmatrix}$.

   a) $\boldsymbol{A}$ é definida positiva?

   b) Determine a decomposição espectral de $\boldsymbol{A}$.

   c) Determine $\boldsymbol{A}^{-1}$.

   d) Determine os valores próprios de $\boldsymbol{A}^{-1}$.

   e) Determine $\boldsymbol{A}^{1/2}$ e $(\boldsymbol{A}^{1/2})^{-1}$.

10. Considere a seguinte matriz de dados

$$\boldsymbol{X} = \begin{bmatrix} 9 & 2 & 6 \\ 12 & 8 & 6 \\ 3 & 4 & 0 \\ 5 & 8 & 4 \\ 10 & 2 & 1 \end{bmatrix}.$$

    Determine $\bar{\boldsymbol{x}}$, $\boldsymbol{S}_n$ e $\boldsymbol{R}$.

11. Dados relativos 10 grandes corporações industriais americanas de três variáveis, 'activos (em milhões de dólares)', 'lucro liquido (em milhões de dólares)' e 'património líquido (em milhões de dólares)' estão disponíveis no ficheiro exe11.xlsx.

    a) Faça a representação dos diagramas de dispersão para os pares de variáveis (x_1, x_2), (x_1, x_3) e (x_2, x_3) e comente.

    b) Calcule $\bar{\boldsymbol{x}}$, $\boldsymbol{S}_n$ e $\boldsymbol{R}$.

12. O ficheiro exe12.xlsx contém dados relativos a 42 medições de 7 variáveis relacionadas com a poluição do ar, 'vento', 'Radiação Solar', 'CO' (monóxido de carbono), 'NO' (óxido nítrico), 'NO_2' (dióxido de nitrogénio), 'O_3'(ozono), 'HC' (hidrocarbonetos).

   a) Faça a representação dos diagramas de dispersão para os pares de variáveis (x_1, x_2), (x_2, x_3) e (x_3, x_5). Comente os resultados.

   b) Calcule $\bar{\boldsymbol{x}}$, $\boldsymbol{S}_n$ e $\boldsymbol{R}$.

13. Num estudo sobre diabetes considerou-se uma amostra de 46 pacientes e mediram-se 'y_1-peso relativo', 'y_2-glicemia em jejum', 'x_1-intolerância à glicose', 'x_2-resposta da insulina à glicose oral' e 'x_3-resistência à insulina'. Os dados estão no ficheiro exe13.xlsx. As três variáveis de maior interesse são x_1, x_2 e x_3, as duas variáveis adicionais de menor interesse são y_1 e y_2 (no ficheiro de dados as colunas têm a seguinte ordem x_1, x_2, x_3, y_1 e y_2).
   Calcule $\begin{bmatrix} \bar{\boldsymbol{y}} \\ -- \\ \bar{\boldsymbol{x}} \end{bmatrix}$ e $\boldsymbol{S}_n = \begin{bmatrix} \boldsymbol{S}_{11} & | & \boldsymbol{S}_{12} \\ -- & - & -- \\ \boldsymbol{S}_{21} & | & \boldsymbol{S}_{22} \end{bmatrix}$.

14. Considere os dados que constam no ficheiro exe13.xlsx relativos a 11 observações de 5 variáveis x_1, x_2, x_3, x_4 e x_5.

   a) Determine $\bar{\boldsymbol{x}}$ e $\boldsymbol{S}$.

   b) Determine a média e variância amostrais de:

$$z = 3x_1 - 2x_2 + 4x_3 - x_4 + x_5\,.$$

   c) Considere agora o vector aleatório $\boldsymbol{z}$ dado por:

$$\boldsymbol{z} = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 2 & -3 & 1 & -2 & -1 \\ -1 & -2 & 1 & -2 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \boldsymbol{AX}$$

   Determine a média e matriz de covariância amostrais de $\boldsymbol{z}$ e a respectiva matriz de correlações amostrais.

15. Usando os dados do ficheiro exe13.xlsx, calcule $|\boldsymbol{S}|$, $tr(\boldsymbol{S})$ e $|\boldsymbol{R}|$.

16. Considere os dados no ficheiro exe16.xlsx relativos medições em quatro idades diferentes do comprimento de um osso facial de 20 rapazes.

   a) Determine $\bar{\boldsymbol{x}}$, $\boldsymbol{S}$ e $\boldsymbol{R}$.

   b) Determine $|\boldsymbol{S}|$ e $tr(\boldsymbol{S})$.

   c) Considere agora as variáveis $z = x_1 + 2x_2 + x_3 - 3x_4$ e $w = -2x_1 + 3x_2 - x_3 + 2x_4$.

      i) Determine $\bar{z}$, $\bar{w}$, s_z^2, s_w^2.

      ii) Determine a covariância amostral entre z e w.

17. Considere a matriz de dados

$$\boldsymbol{X} = \begin{bmatrix} -1 & 3 & -2 \\ 2 & 4 & 2 \\ 5 & 2 & 3 \end{bmatrix}.$$

   a) Calcule a matriz $\boldsymbol{X} - \mathbf{1}_n \bar{\boldsymbol{x}}'$.
   b) Determine $\boldsymbol{S}$ (através de operações com matrizes) e calcule $|\boldsymbol{S}|$.
   c) Calcule a variância amostral total.

18. Considere os dados no ficheiro exe18.xlsx, que contém medições referentes a 6 variáveis aleatórias $x_1, x_2, x_3, x_4, x_5, x_6$. Sejam $\boldsymbol{u}' = [u_1, u_2]$ e $\boldsymbol{v}' = [v_1, v_2, v_3]$ tais que

$$u_1 = x_1 + x_2 + x_3 \quad , \quad u_2 = x_1 - 2x_2 + 2x_3 \,,$$

$$v_1 = 3x_4 + x_5 - 2x_6 \quad , \quad v_2 = -x_4 - 2x_5 + x_6 \quad , \quad v_3 = 2x_4 - 3x_5 + x_6 \,.$$

   a) Determine a covariância amostral de $\boldsymbol{u}$ e de $\boldsymbol{v}$, isto é, $\boldsymbol{S}_{\boldsymbol{u}}$ e $\boldsymbol{S}_{\boldsymbol{v}}$.
   b) Determine a covariância amostral de cada variável de $\boldsymbol{u}$ com cada variável de $\boldsymbol{v}$, isto é determine a matriz $\boldsymbol{S}_{\boldsymbol{u},\boldsymbol{v}}$.

19. Considere o ficheiro exe19.xlsx com uma amostra de alturas (cm) e pesos (Kg) de duas populações, A e B.

   (a) Considere os dados relativos à população A.
      i. Determine as medidas amostrais; média, variância, desvio padrão, quartis, máximo e minimo para as alturas e os pesos.
      ii. Apresente os histogramas, com 20 classes, para cada uma das varáveis em estudo.
      iii. Apresente o diagrama de quartis para cada uma das varáveis em estudo; apresente a figuras com os diagramas isolados e uma figura com os dois diagramas.
      iv. Assumindo que a população tem distribuição multivariada normal, teste para um nível de significância de 0.05.
         i) $H_0 : \mu_{alturasA} = 170$ vs $\mu_{alturasA} \neq 170$.
         ii) $H_0 : \mu_{pesosA} = 70$ vs $\mu_{pesosA} \neq 70$.

         Apresente todos resultados e as conclusões.

   (b) Considerando agora os dados relativos às duas populações. Teste a igualdade dos valores médios, das alturas e dos pesos, das duas populações A e B. Apresente e discuta os resultados obtidos.

20. Seja $\boldsymbol{x} = [x_1, x_2, x_3]' \frown N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, com

$$\boldsymbol{\mu} = \begin{bmatrix} 3 \\ 1 \\ 4 \end{bmatrix} \quad , \quad \boldsymbol{\Sigma} = \begin{bmatrix} 6 & 1 & -2 \\ 1 & 13 & 4 \\ -2 & 4 & 4 \end{bmatrix}$$

   a) Indique a distribuição de $y = 2x_1 - x_2 + 3x_3$.
   b) Indique a distribuição conjunta de $y_1 = x_1 + x_2 + x_3$ e $y_2 = x_1 - x_2 + 2x_3$.

c) Indique a distribuição de x_2.

d) Indique a distribuição conjunta de x_1 e x_3.

e) Indique a distribuição conjunta de x_1 e $1/2(x_1 + x_2)$.

21. Considere uma população bivariada Normal com $\mu_1 = 0$, $\mu_2 = 2$, $\sigma_{11} = 2$, $\sigma_{22} = 1$ e $\rho_{12} = 0.5$. Indique a expressão da densidade desta distribuição bivariada Normal.

22. Seja $\boldsymbol{X} = [x_1, x_2, x_3]' \frown N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, com

$$\boldsymbol{\mu} = \begin{bmatrix} -3 \\ 1 \\ 4 \end{bmatrix} \quad , \quad \boldsymbol{\Sigma} = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

a) Indique, justificando, quais das seguintes variáveis(vectores) aleatórias(os) são independentes:

i) x_1 e x_2.

ii) x_2 e x_3.

iii) $[x_1, x_2]'$ e x_3.

iv) $\dfrac{x_1 + x_2}{2}$ e x_3.

v) x_2 e $x_2 - \dfrac{5}{2}x_1 - x_3$.

b) Indique a distribuição de $(\boldsymbol{X} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{X} - \boldsymbol{\mu})$.

23. Seja $\boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}_1 \\ -- \\ \boldsymbol{x}_2 \end{bmatrix} \frown N_4\left(\begin{bmatrix} 2 \\ -1 \\ -- \\ 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 7 & 3 & | & -3 & 2 \\ 3 & 6 & | & 0 & 4 \\ - & - & | & - & - \\ -3 & 0 & | & 5 & -2 \\ 2 & 4 & | & -2 & 4 \end{bmatrix} \right).$

a) Determine $E(\boldsymbol{x}_1|\boldsymbol{x}_2)$.

b) Determine $Cov(\boldsymbol{x}_1|\boldsymbol{x}_2)$.

24. Utilizando os estimadores de máxima verosimilhança, determine estimativas para $\boldsymbol{\mu}$ e $\boldsymbol{\Sigma}$ com base na amostra aleatória

$$\boldsymbol{X} = \begin{bmatrix} 3 & 6 \\ 4 & 4 \\ 5 & 7 \\ 4 & 7 \end{bmatrix}$$

extraída de uma população bivariada Normal.

25. Considere a amostra aleatória $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{20}$ extraída de uma população $N_6(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, determine:

a) a distribuição de $(\boldsymbol{x}_1 - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_1 - \boldsymbol{\mu})$.

b) a distribuição de $\bar{\boldsymbol{x}}$ e $\sqrt{n}\,(\bar{\boldsymbol{x}} - \boldsymbol{\mu})$.

c) a distribuição de $(n-1)\boldsymbol{S}$.

26. Considere a amostra aleatória $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{25}$ extraída de uma população $N_8(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, determine:

   a) a distribuição de $\bar{\boldsymbol{x}}$.

   b) a distribuição de $24\boldsymbol{S}$.

27. (a) Seja $\boldsymbol{x}' = [x_1, x_2, x_3, x_4] \frown N_4(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, com

$$\boldsymbol{\mu} = \begin{bmatrix} -2 \\ 1 \\ -1 \\ 3 \end{bmatrix} \quad , \quad \boldsymbol{\Sigma} = \begin{bmatrix} 11 & 3 & 0 & 0 \\ 3 & 9 & 0 & 0 \\ 0 & 0 & 2 & 2 \\ 0 & 0 & 2 & 4 \end{bmatrix} .$$

   i. Indique a distribuição de x_2 .

   ii. Indique a distribuição de $\boldsymbol{y} = [y_1, y_2, y_3]'$, sendo $y_1 = x_1 + x_2 + x_3 - x_4$, $y_2 = x_1 - x_2 + 2x_3 + x_4$ e $y_3 = x_1 + x_2$

   iii. Indique, justificando, se $[x_1, x_2]'$ e $[x_3, x_4]'$ são independentes.

   iv. Seja $\boldsymbol{x}^{(1)} = [x_1, x_2, x_3]'$, determine $E(\boldsymbol{x}^{(1)}|x_4 = x_4)$ e $Cov(\boldsymbol{x}^{(1)}|x_4 = x_4)$.

   (b) Seja $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ uma amostra aleatória extraída da população com distribuição $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

   i. Indique a distribuição de $\boldsymbol{x}$.

   ii. Mostre que $E(\sqrt{n}(\boldsymbol{x} - \boldsymbol{\mu})) = \boldsymbol{0}$ e que $Cov(\sqrt{n}(\boldsymbol{x} - \boldsymbol{\mu})) = \boldsymbol{\Sigma}$.

   iii. Indique a distribuição de $n(\boldsymbol{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})$ e a distribuição de $(n-1)\boldsymbol{S}$.

28. Seja $\boldsymbol{x}' = [x_1, x_2, x_3] \frown N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, com

$$\boldsymbol{\mu} = \begin{bmatrix} -2 \\ 3 \\ -1 \end{bmatrix} \quad , \quad \boldsymbol{\Sigma} = \begin{bmatrix} 11 & 0 & 0 \\ 0 & 9 & -3 \\ 0 & -3 & 2 \end{bmatrix} .$$

   (a) i. Indique a distribuição de x_3 .

   ii. Indique a distribuição conjunta de $y_1 = x_1 + x_2 + x_3$ e $y_2 = x_1 - x_2 + 2x_3$.

   iii. Indique, justificando, se x_1 e $[x_2, x_3]'$ são independentes.

   iv. Seja $\boldsymbol{x}^{(1)} = [x_1, x_2]'$, determine $E(\boldsymbol{x}^{(1)}|x_3 = x_3)$ e $Cov(\boldsymbol{x}^{(1)}|x_3 = x_3)$.

   (b) Seja $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ uma amostra aleatória extraída da população $\boldsymbol{x}$ com a distribuição $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Indique a distribuição de $\boldsymbol{x}$, $\sqrt{n}(\boldsymbol{x} - \boldsymbol{\mu})$, $n(\boldsymbol{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})$ e a distribuição de $(n-1)\boldsymbol{S}$.

29. Considere a seguinte matriz de dados

$$\boldsymbol{X} = \begin{bmatrix} 2 & 12 \\ 8 & 9 \\ 6 & 9 \\ 8 & 10 \end{bmatrix} .$$

Com base nesta amostra pretendemos testar

$$H_0 : \boldsymbol{\mu}' = [7, 11] \quad \text{vs} \quad H_1 : \boldsymbol{\mu}' \neq [7, 11] .$$

a) Indique qual a distribuição da estatística T^2 de Hotelling.

b) Determine o valor observado da estatística T^2.

c) Teste H_0 para $\alpha = 0.05$.

30. Com base numa amostra de dimensão 3 extraída de um população bivariada Normal obteve-se a seguinte matriz dos dados:

$$\boldsymbol{X} = \begin{bmatrix} 6 & 9 \\ 10 & 6 \\ 8 & 3 \end{bmatrix}.$$

Mostre que, num teste $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ vs $H_1 : H_0 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$, o valor observado da estatística T^2, para $\boldsymbol{\mu}'_0 = [9, 5]$, não se altera se cada observação $\boldsymbol{x}_j$, $j = 1, 2, 3$ for substituída por $\boldsymbol{C}\boldsymbol{x}_j$ sendo

$$\boldsymbol{C} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}.$$

31. Com base numa amostra de dimensão 42 extraída de um população bivariada Normal obtiveram-se as seguintes matrizes amostrais:

$$\bar{\boldsymbol{x}} = \begin{bmatrix} .564 \\ .603 \end{bmatrix}, \boldsymbol{S} = \begin{bmatrix} 0.0144 & 0.0117 \\ 0.0117 & 0.0146 \end{bmatrix} \text{ e } \boldsymbol{S}^{-1} = \begin{bmatrix} 203.018 & -163.391 \\ -163.391 & 200.228 \end{bmatrix}.$$

a) Teste, utlizando a estatística T^2, ao nível de 5%

$$H_0 : \boldsymbol{\mu} = [0.55, 0.60]' \quad \text{vs} \quad H_1 : \boldsymbol{\mu} \neq [0.55, 0.60]'.$$

b) Determine a expressão que define a região 95% confiança para $\boldsymbol{\mu}$. Faça um esboço gráfico da elipse.

c) verifique se o vector $\boldsymbol{\mu}' = [0.562, 0.589]$ pertence à região de confiança determinada na alínea anterior.

d) Determine, para $1 - \alpha = 0.95$, os intervalos de confiança simultâneos para μ_1, μ_2 utilizando dois métodos diferentes. Compare os resultados.

32. Considere uma amostra aleatória $\boldsymbol{x}_1, \ldots,$ $\boldsymbol{x}_{35}$ (ficheiro exe32.xlsx) extraída de uma população $N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Teste, utilizando a estatística T^2 de Hotelling e a estatística de razão de verosimilhanças, Λ, ao nível de 5%, a hipótese nula

$$H_0 : \boldsymbol{\mu} = [15, 6, 2.85]' \quad \text{vs} \quad H_1 : \boldsymbol{\mu} \neq [15, 6, 2.85]'.$$

33. Considere os dados (ficheiro exe33.xlsx) relativos medições em quatro idades diferentes do comprimento de um osso facial de 20 rapazes.

a) Teste, para $\alpha = 0.05$, $H_0 :\ \boldsymbol{\mu} = [48, 49, 50, 51]'$.

b) Determine, para $1 - \alpha = 0.95$, intervalos de confiança simultâneos para μ_1, μ_2, μ_3 e μ_4.

c) Utilizando o método de Bonferroni, determine, para $1 - \alpha = 0.95$, intervalos de confiança simultâneos para μ_1, μ_2, μ_3 e μ_4.

34. Com base numa amostra (ficheiro exe34.xlsx) extraída de uma população multivariada Normal, $N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$:

   (a) Teste, utilizando a estatística de razão de verosimilhanças, Λ, para $\alpha = 0.1$,

   $$H_0 : \boldsymbol{\mu} = [1, 1.5, 4]' \text{ vs } H_1 : \boldsymbol{\mu} \neq [1, 1.5, 4]'.$$

   (b) Determine a expressão que define a região de confiança a 95% para $\boldsymbol{\mu}$ e indique o centro e os vectores que a definem.

   (c) Determine, para $1 - \alpha = 0.95$, e com base na estatística T^2 de Hotelling, intervalos de confiança simultâneos para μ_1, μ_2 e μ_3.

35. Considere os dados (ficheiro exe35.xlsx) relativos medições relativas a cinco variáveis.

   a) Teste utilizando a estatística T^2 de Hotelling, para $\alpha = 0.05$, $H_0 : \boldsymbol{\mu} = [47, 46, 48, 50, 52]'$.

   b) Teste, utilizando a estatística de razão de verosimilhanças, Λ, ao nível de 5%, $H_0 : \boldsymbol{\mu} = [47, 46, 48, 50, 52]'$.

   c) utilizando o método de Bonferroni, determine, para $1 - \alpha = 0.95$, intervalos de confiança simultâneos para μ_1, μ_2, μ_3, μ_4 e μ_5.

36. Considere os dados no ficheiro exe36.xlsx, relativos a medições relativas a 4 variáveis x_1, x_2, x_3 e x_4. Nas alíneas seguintes, sempre que possível, faça a interpretação dos resultados obtidos.

   (a) Crie uma amostra de dimensão 1000 (sem reposição) a partir do ficheiro de dados fornecidos e com base nessa amostra responda às seguintes alíneas.

      i. Determine, considerando os dados amostrais do ficheiro, estimativas para o vetor de médias, a matriz de covariância e a matriz de correlações.

      ii. Apresente diagramas de dispersão para as variáveis x_1 e x_2 e também com as variáveis x_3 e x_4.

      iii. Apresente uma figura com os diagramas de quartis das quatro variáveis.

   (b) Considerando agora o ficheiro inicial, teste as hipóteses

   $$H_0 : \boldsymbol{\mu} = [10, 15, 1, 100]' \text{ vs } H_1 : \boldsymbol{\mu} \neq [10, 15, 1, 100]'$$

   apresentando todos os resultados obtidos e conclusões.

   (c) Estude o impacto do tamanho da amostra na resposta à questão anterior, considerando 100 réplicas de amostras de dimensões 100,500,1000 e 5000.

37. Suponhamos que foram recolhidas duas amostras aleatórias (independentes), uma de dimensão $n_1 = 7$ e outra de dimensão $n_2 = 5$ (ficheiro exe37.xlsx), em dois povoamentos distintos de pinheiro bravo, referentes às variáveis x_1- altura da ávore (em m), x_2- diâmetro da árvore (em cm) e x_3- idade da árvore (em anos). Teste, ao nível de 5%, se é plausível considerar que os dois povoamentos apresentam altura, diâmetro e idade médias iguais. (Nota: assuma que as matrizes de covariância populacionais são iguais e que as populações são multivariadas Normais)

38. Com base em duas amostras independentes de dimensão 50 extraídas de duas populações multivariadas $N_2(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$ e $N_2(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$ obtiveram-se os seguintes valores:

$$\bar{\boldsymbol{x}}_1 = \begin{bmatrix} 8.3 \\ 4.1 \end{bmatrix}, \quad \boldsymbol{S}_1 = \begin{bmatrix} 2 & 1 \\ 1 & 6 \end{bmatrix}$$

$$\bar{\boldsymbol{x}}_2 = \begin{bmatrix} 10.2 \\ 3.9 \end{bmatrix}, \quad \boldsymbol{S}_2 = \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix}.$$

   a) Verifique se $\mathbf{0}$ pertence a região 95% confiança para $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$.

   b) Identifique o centro e os vectores que determinam os eixos da elipse de confiança para $1 - \alpha = 0.95$.

   c) Teste, ao nível de 1%, a hipótese nula $H_0 : \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \mathbf{0}$.

39. Para a realização de testes fisiológicos, relativos a quatro variáveis (x_1, x_2, x_3, x_4), foram considerados dois grupos, um com 32 homens (1 - homem) e outro com 32 mulheres (2- mulheres) (ficheiro exe39.xlsx).

   a) Ao nível de 1%, teste se podemos considerar que os resultados obtidos nos dois grupos são diferentes, ou seja teste

$$H_0 : \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \mathbf{0} \quad \text{vs} \quad H_1 : \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 \neq \mathbf{0}.$$

   b) Determine intervalos de confiança simultâneos para a diferença de valores médios. Considere um coeficiente de confiança global de 95%.

40. Considere os dados que constam no ficheiro exe40.xlsx relativos a medições de três variáveis x_1, x_2 e x_3. Teste, ao nível de 5%, a hipótese nula

$$\mu_1 = \mu_2 = \mu_3.$$

41. Considere uma amostra aleatória (ficheiro exe41.xlsx) extraída de uma população $N_5(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Teste, ao nível de 5%, a hipótese $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$.

42. Considere os ficheiros exe42a.xlsx e exe42b.xlsx que contém dados relativos a medições de quatro variáveis x_1, x_2, x_3 e x_4 em duas espécies de besouros:

    - x_1: distância do sulco transversal à borda posterior do pró-torax (μm)
    - x_2: comprimento dos élitros $(0,01mm)$
    - x_3: comprimento do segundo segmento da antena (μm)
    - x_3: comprimento do terceiro segmento da antena (μm).

    Para um nível de confiança 95%, determine intervalos de confiança simultâneos para $\mu_{1j} - \mu_{2j}$, $j = 1, 2, 3, 4$.

43. Foi pedido a 15 alunos para escreverem um texto informal e outro formal. As variáveis registadas foram o número de palavras e o número de verbos, mais precisamente (ficheiro exe43.xlsx):

    y_1 = número de palavras no ensaio informal,

    y_2 = número de verbos no ensaio informal,

x_1 = número de palavras no ensaio formal,

x_2 = número de verbos no ensaio formal.

a) Teste se existem diferenças nos resultados obtidos nos diferentes tipos de texto.
b) Teste individualmente para cada tipo de texto se existiram diferenças nos resultados obtidos.

44. Realizaram-se testes psicotécnicos em dois grupos, um com 32 homens (1 - homem) e o outro com 32 mulheres (2 - mulher). As variáveis em estudo foram x_1-raciocínio lógico, x_2-raciocínio numérico, x_3-inconsistências pictóricas e x_4- vocabulário. Os dados encontram-se no ficheiro exe44.xlsx.

   a) Teste, ao nível de significância de 5%,

   $$H_0: \quad \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \quad \text{vs} \quad H_1: \quad \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2 \,.$$

   Indique quais os pressupostos que teve que assumir.

   b) Teste a significância das variáveis x_3 e x_4 além das variáveis x_1 e x_2 na distinção dos dois grupos.
   c) Teste agora a significância de cada variável além das restantes três na distinção dos dois grupos.

45. Considere os dados no ficheiro exe45.xlsx, relativos a medições relativas a 5 variáveis x_1, x_2, x_3, x_4 e x_5 em três populações. Nas alíneas seguintes, sempre que possível, faça a interpretação dos resultados obtidos.

   (a) Crie amostras, para cada uma das 3 populações (sem reposição), com 1% dos dados originais a partir do ficheiro de dados fornecidos. Com base nessas amostras responda às seguintes alíneas.

      i. Determine estimativas para o vetor de médias, a matriz de covariância e a matriz de correlações para cada uma das populações.
      ii. Apresente diagramas de dispersão para as variáveis x_1 e x_2 e também com as variáveis x_3 e x_5, em cada uma das populações.
      iii. Apresente uma figura com os diagramas de quartis das cinco variáveis, para cada uma das populações.
      iv. Apresente histogramas comparativos, para a variável x_4, considerando as populações duas a duas.

   (b) Considerando agora o ficheiro inicial.

      i. Teste as hipóteses

      $$H_0: \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \ \text{ vs } \ H_1: \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$$

      apresentando todos os resultados obtidos e conclusões.

      ii. Teste as hipóteses

      $$H_0: \boldsymbol{\mu}_2 = \boldsymbol{\mu}_3 \ \text{ vs } \ H_1: \boldsymbol{\mu}_2 \neq \boldsymbol{\mu}_3$$

      apresentando todos os resultados obtidos e conclusões.

   (c) Estude o impacto do tamanho da amostra na resposta à alínea (b) i., considerando 100 réplicas de amostras de dimensões 100,500,1000,1500,2500 e 5000.

46. Suponhamos que temos as seguintes três amostras independentes extraídas de populações univariadas Normais:

Pop. 1	Pop. 2	Pop. 3
9	0	3
6	2	1
9		2

Teste, ao nível de significância de 1%, $H_0 : \mu_1 = \mu_2 = \mu_3$.

47. Suponhamos que temos as seguintes três amostras independentes extraídas de populações bivariadas Normais:

Pop. 1		Pop. 2		Pop. 3	
9	3	0	4	3	8
6	2	2	0	1	9
9	7			2	7

Teste, ao nível de significância de 1%, $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \boldsymbol{\mu}_3$.

48. Numa experiência procurou-se estudar os efeitos da enxertia de Macieiras em diferentes tipos de porta-enxertos. Os dados relativos a oito árvores usadas em cada um dos três porta-enxertos estão no ficheiro exe48.xlsx. As variáveis em estudo são:

    - y_1 diâmetro do tronco, aos 4 anos ($mm \times 100$),
    - y_2 crescimento aos 4 anos (m),
    - y_3 diâmetro do troco aos 15 anos ($mm \times 100$),
    - y_4 peso da árvore acima do solo aos 15 anos ($kg \times 1000$).

    Teste, ao nível de 5%, $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \boldsymbol{\mu}_3$.

49. Numa competição de vinhos os juízes avaliaram vinhos produzidos em três regiões diferentes relativamente a quatro variáveis: x_1 - Cor, x_2 - Claridade x_3 - Aroma e x_4 - Acidez. Foram considerados doze vinhos de cada região e cada juíz deu a sua classificação. Os valores apresentados, no ficheiro exe49.xlsx, são as pontuações médias dos juízes para cada garrafa. verifique se existem diferenças significativas nas avaliações dos vinhos das três regiões.

50. No ficheiro exe50.xlsx, temos os resultados de análises ao sangue relativos a medições de 3 variáveis (x_1 - contagem de glóbulos brancos, x_2 - contagem de glóbulos vermelhos, e x_3 - contagem de hemoglobina) obtidas através da utilização de três reagentes diferentes. O primeiro reagente é actualmente o mais usado em análises clínicas contudo, com vista a uma redução de custos, pretende-se ter alternativas, mais baratas, a este reagente. Com o objectivo de verificar se os outros dois reagentes também podem ser utilizados, teste se os vectores de médias dos diferentes reagentes são iguais. Indique quais os pressupostos que teve que assumir.

51. Amostras de um determinado tipo de cerâmica foram consideradas em quatro locais distintos designados por L, C, I, A (ficheiro exe51.xlsx). usou-se um produto químico para medir a percentagem de quatro óxidos metálicos presentes em cada amostra: x_1- percentagem de óxido de alumínio, x_2- percentagem de óxido de ferro, x_3- percentagem de óxido de cálcio, x_4- percentagem de óxido de sódio. Teste, para um nível de significância de 5%, se existem diferenças significativas na composição química das cerâmicas produzidas nos diferentes locais.

52. Pretende-se analisar se existiram alterações no tamanho de crânios Egípcios de três períodos: 4000 AC (1º período), 3300 AC (2º período) e 1.850 AC (3º período). Com este objectivo foram feitas medições de crânios Egípcios masculinos dos três períodos. Os dados estão no ficheiro exe52.xlsx. As variáveis consideradas são:

    x_1 - largura máxima (mm)

    x_2 - altura basibregmatica (mm)

    x_3 - comprimento basialveolar (mm)

    x_4 - altura nasal (mm)

    O que pode concluir? Indique os pressuposto que teve que assumir e considere $\alpha = 0.01$.