shipout/backgroundshipout/foreground

Departamento de Matemática

Fichas de exercícios
Estatística Multivariada

Universidade de Aveiro

Adelaide Freitas
(adelaide@ua.pt)

Conteúdo

Capítulo 1

Introdução à Análise Multivariada

- Modelo multivariado
- Parâmetros de um modelo multivariado: Esperança e matriz de covariância (definição e propriedades)
- Amostra aleatória multivariada
- Medidas amostrais multivariadas.

1.1 Exercícios

Ex. 1.1 — Seja $\mathbf{X} = [X_1\ X_2]'$ um v.a. tal que $E(X_1) = E(X_2) = 2$, $Var(X_1) = 1$, $Var(X_2) = 5$ e $E(X_1X_2) = 2$. Seja $\mathbf{a} = [1\ 0]'$ e $\mathbf{b} = [2\ \ -1]'$
Encontre:

1. $E(\mathbf{a} - \mathbf{X})$;
2. a matriz de covariâncias de $\mathbf{X}$ e verifique que ela é definida positiva;
3. $Var(2 - \mathbf{b}'\mathbf{X})$;
4. a esperança e a matriz de covariâncias de

$$\mathbf{Z} = \begin{bmatrix} X_1 \\ X_1 - X_2 \\ X_2 \end{bmatrix}$$

Ex. 1.2 — Seja $\mathbf{X} = [X_1\ X_2\ X_3]'$ um v.a. tal que $E(X_1 + X_2 + X_3) = 6$, $E(X_1) > 0$, $E(X_1^2) = 15$, $E(2X_3) = 6$, $E(X_1X_2) = -4$, $Var(X_1) = 11$, $Var(X_2 + X_3) = 8$, $Var(3X_3) = 54$, $Cov(X_1 + X_2, X_3) = -2$, $Cov(X_3, 2X_1) = 4$.
Determine:

1. a matriz de covariâncias de $\mathbf{X}$.
2. $E(\mathbf{a}'\mathbf{X})$, onde $\mathbf{a} = [1\ 2\ 3]'$;

3. $Var(3\mathbf{X})$;
4. a esperança e a matriz de covariâncias de
   (a) $Z = \frac{X_1+X_2+X_3}{3}$
   (b) $\mathbf{Z} = \begin{bmatrix} X_1 - X_2 \\ X_2 - X_3 \end{bmatrix}$

Ex. 1.3 — Seja $\mathbf{X} = [X_1 \ X_2 \ X_3]'$ um v.a. com matriz de covariâncias

$$\mathbf{\Sigma} = \begin{bmatrix} 25 & -2 & 4 \\ -2 & 4 & 1 \\ 4 & 1 & 9 \end{bmatrix}$$

1. Determine a matriz de correlações de $\mathbf{X}$;
2. Calcule a correlação entre X_1 e $\frac{1}{2}X_2 + \frac{1}{2}X_3$;
3. Determine a variância da combinação linear $3X_1 - X_2 + 2X_3$
4. Determine a matriz de covariâncias 2×2 do vetor de componentes:

$$\begin{aligned} Z_1 &= X_1 - X_2 \\ Z_2 &= X_1 + X_2 \end{aligned}$$

Ex. 1.4 — (R) Considere o ficheiro iris do R. Trata-se de uma base de dados originalmente construída para quantificar a variação geográfica das flores íris na Península de Gaspé (Quebec, Canadá) e estudada por Sir R. Fisher (1936). O ficheiro contém o comprimento e largura da sépala e o comprimento e largura da pétala de 150 flores de 3 espécies distintas: Iris Setosa, Iris Versicolor e Iris Virgínica. Para cada espécie existem 50 observações.

Verifique a estrutura dos dados usando o comando `str`. Seja $\underset{\sim}{\mathbf{x}}$ a matriz de dados numéricos (150×4).

1. Obtenha o vetor de médias e as matrizes de covariâncias e de correlações de $\underset{\sim}{\mathbf{x}}$.
2. Obtenha uma sequência de comandos que lhe permita obter a variância (empírica) de $\underset{\sim}{\mathbf{y}} = \underset{\sim}{\mathbf{x}}\,[1 \ 0 \ 1 \ 0]'$. Interprete a nova matriz de dados $\underset{\sim}{\mathbf{y}}$.
3. Obtenha uma sequência de comandos que lhe permita obter a matriz de correlações (empírica) de $\underset{\sim}{\mathbf{z}} = \underset{\sim}{\mathbf{x}} \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}'$. Interprete a nova matriz de dados $\underset{\sim}{\mathbf{z}}$.

1.2 Soluções

Resposta do ex. 1.1 — 1. $[-1 \quad -2]'$; 2. $\begin{bmatrix} 1 & -2 \\ -2 & 5 \end{bmatrix}$; 3. 17; 4. $[2 \ \ 0 \ \ 2]'$, $\begin{bmatrix} 1 & 3 & -2 \\ 3 & 10 & -7 \\ -2 & -7 & 5 \end{bmatrix}$

Resposta do ex. 1.2 — 1. $\begin{bmatrix} 11 & -6 & 2 \\ -6 & 10 & -4 \\ 2 & -4 & 6 \end{bmatrix}$; 2. 13; 3. $\begin{bmatrix} 99 & -54 & 18 \\ -54 & 90 & -36 \\ 18 & -36 & 54 \end{bmatrix}$; 4. a) 2 , 11/9; b) $[1 \quad -2]'$, $\begin{bmatrix} 33 & -22 \\ -22 & 24 \end{bmatrix}$

Resposta do ex. 1.3 — 1. $\begin{bmatrix} 1 & -1/5 & 4/15 \\ -1/5 & 1 & 1/6 \\ 4/15 & 1/6 & 1 \end{bmatrix}$; 2. $2/5\sqrt{15}$; 3. 321; 4. $\begin{bmatrix} 33 & 21 \\ 21 & 25 \end{bmatrix}$

Resposta do ex. 1.4 — Comandos do R:

```
> str(iris)
```

1.
```
> mean(iris[,1:4]); cov(iris[,1:4]); cor(iris[,1:4])
```

2.
```
> Y = iris[,1:4] %*% t(c(1,0,1,0))
> cov(Y)
```
Dados univariados com a soma dos comprimentos das pétalas e sépalas.

3.
```
> a=matrix(c(1,0,1,0,0,1,0,1),nrows=4)
> Z = iris[,1:4] %*% a
> cor(Z)
```
Dados bivariados com a soma dos comprimentos das pétalas e sépalas numa dimensão e com a soma das larguras das pétalas e das sépalas na segunda dimensão.

Capítulo 2

Distribuição Normal Multivariada

5.1. Normal univariada (revisão)

5.2. Normal multivariada

a. Definição e propriedades

b. Estimação de máxima verosimilhança

c. Distribuições de Amostragem. TLC multivariado. Transformações de Estatísticas.

d. Inferência sobre um vetor de médias

5.3. Validação da normalidade

2.1 Exercícios

Ex. 2.1 — Seja $\mathbf{X} \frown N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ com $\boldsymbol{\mu} = \mathbf{0}$ e $\boldsymbol{\Sigma} = \begin{bmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{bmatrix}$.

Identifique a distribuição de

1. cada uma das marginais do vetor aleatório $\mathbf{X}$;
2. $\mathbf{cX}$, onde $\mathbf{c} = \begin{bmatrix} 1 & 0 & -1 \\ 1 & -2 & 1 \end{bmatrix}$;
3. $(\mathbf{X} - \boldsymbol{\mu})' \, \boldsymbol{\Sigma}^{-1} \, (\mathbf{X} - \boldsymbol{\mu})$.

Ex. 2.2 — Seja $X_1 \frown N(0, 1)$ e seja

$$X_2 = \begin{cases} -X_1 & , \text{se} -1 \leq X_1 \leq 1 \\ X_1 & , \text{outros casos} \end{cases}$$

Mostre que

1. $X_2 \frown N(0, 1)$;

2. (X_1, X_2) não segue uma distribuição normal bivariada;
3. A seguinte afirmação é falsa: " *Um v.a. p-dimensional tem distribuição normal se e só se as suas p marginais seguem uma distribuição normal unidimensional*"

Ex. 2.3 — Sejam $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{X}_3$ e $\mathbf{X}_4$ quatro v.a's independentes com distribuição normal p-dimensional $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

1. Encontre as distribuições marginais dos v.a's:

$$\mathbf{V}_1 = \frac{1}{4}\mathbf{X}_1 - \frac{1}{4}\mathbf{X}_2 + \frac{1}{4}\mathbf{X}_3 - \frac{1}{4}\mathbf{X}_4 \ \text{ e } \ \mathbf{V}_2 = \frac{1}{4}\mathbf{X}_1 + \frac{1}{4}\mathbf{X}_2 - \frac{1}{4}\mathbf{X}_3 - \frac{1}{4}\mathbf{X}_4$$

2. Identifique a distribuição de $\begin{bmatrix} \mathbf{V}_1 \\ \mathbf{V}_2 \end{bmatrix}$.

Sugestão: Trabalhe com bloco de matrizes.

Ex. 2.4 — Seja $\mathbf{X} \frown N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ com $\boldsymbol{\mu}' = [1\ 2]$ e

$$\boldsymbol{\Sigma} = \begin{bmatrix} 4 & -2 \\ -2 & 2 \end{bmatrix}$$

1. R Obtenha uma sequência de comandos que lhe permita construir um gráfico com elipses correspondentes a contornos da superfície onde a função densidade de probabilidade de $\mathbf{X}$ se mantêm constante.
2. Identifique o centro e os comprimentos dos semi-eixos da elipse onde **a função densidade** de $\mathbf{X}$ toma o valor constante igual a 0.03.

Ex. 2.5 — Seja $\mathbf{X} = [X_1\ X_2\ X_3]' \frown N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ com $\boldsymbol{\mu}' = [-3\ 1\ 4]$ e

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

Qual das seguintes v.a's são independentes

1. X_1 e X_2 ?
2. X_2 e X_3 ?
3. (X_1, X_2) e X_3 ?
4. $\frac{X_1+X_2}{2}$ e X_3 ?

Ex. 2.6 — Seja $\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \frown N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, com $\mathbf{X}_1$ $(q \times 1)$, $\mathbf{X}_2$ $((p-q) \times 1)$ e

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$$

Prove que $\boldsymbol{\Sigma}_{12} = \mathbf{0}$ sse $\mathbf{X}_1$ é independente de $\mathbf{X}_2$.

Ex. 2.7 — Seja $\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \frown N_2(\mathbf{0}, \boldsymbol{\Sigma})$, com $\boldsymbol{\Sigma} = \begin{bmatrix} 1 & -0.8 \\ -0.8 & 2 \end{bmatrix}$.
Nestas condições,

1. identifique a f.d.p. das marginais X_1 e X_2;
2. identifique a distribuição condicional de X_2 dado X_1

Ex. 2.8 — Seja $X_1 \frown N(0,1)$ e $\mathbf{X} = \begin{bmatrix} X_1 \\ \mathbf{X}_2 \end{bmatrix} \frown N_3 \left(\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 & 2 & 1 \\ 2 & 5 & 2 \\ 1 & 2 & 2 \end{bmatrix} \right)$.

Nestas condições,

1. identifique a distribuição de $\mathbf{X}_2$;
2. identifique a distribuição condicional de $\mathbf{X}_2$ dado X_1.

Ex. 2.9 — Considere uma população normal bivariada de média $\boldsymbol{\mu}$ e matriz de covariâncias $\boldsymbol{\Sigma}$.

1. Mostre que a media amostral é um estimador centrado de $\boldsymbol{\mu}$ mas a matriz de covariâncias amostral não é um estimador centrado de $\boldsymbol{\Sigma}$.
2. Com base na seguinte amostra

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

extraída daquela população, obtenha as estimativas de máxima verosimilhança de $\boldsymbol{\mu}$ e de $\boldsymbol{\Sigma}$.

Ex. 2.10 — Considere a matriz de dados $\underset{\sim}{\mathbf{x}} = \begin{bmatrix} 0 & 1 & 2 \\ 0 & 4 & 2 \end{bmatrix}'$ extraída de uma população normal bivariada.

1. Calcule o vetor de médias amostrais $\bar{\mathbf{x}}$ e a matriz de covariâncias amostrais $\mathbf{s}_c$.
2. Qual a distribuição da estatística T^2 de Hotelling para testar a média populacional com base na amostra dada?
3. Teste, ao nível de significância de 10%,

$$H_0 \;:\; \boldsymbol{\mu} = [0\ 2]' \quad \text{vs} \quad H_1 \;:\; \boldsymbol{\mu} \neq [0\ 2]'$$

(Nota: o objetivo desta questão é que sejam efectuados todos os cálculos sem recurso ao computador).

Ex. 2.11 — Considere uma amostra aleatória $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$ extraída de uma população bidimensional $\mathbf{X}$ discreta tal que

$$P(\mathbf{X} = (0,0)) = P(\mathbf{X} = (0,1)) = P(\mathbf{X} = (1,0)) = P(\mathbf{X} = (1,1)) = \frac{1}{4}$$

1. Usando o TLC multivariado, o que pode concluir quanto ao comportamento assintótico da média amostral $\bar{\mathbf{X}}$?
2. Obtenha uma sequência de comandos que lhe permita visualizar uma parte da superfície da função densidade de probabilidade limite identificada na alínea anterior.

3. Identifique o comportamento assintótico da transformação $g(\bar{\mathbf{x}}) = \bar{\mathbf{x}}'\mathbf{A}\bar{\mathbf{x}}$, onde $\mathbf{A}$ é uma matriz simétrica de constantes reais, 2×2, definida positiva.

Ex. 2.12 — Considere uma amostra aleatória $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$ extraída de uma população bidimensional $\mathbf{X}$ com vetor média $\boldsymbol{\mu} = [0\ 0]'$ e matriz de covariâncias

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$$

Encontre a distribuição limite, quando $n \to +\infty$, do vetor aleatório

$$\begin{bmatrix} \bar{X}_1^2 - \bar{X}_2 \\ \bar{X}_1 + 3\bar{X}_2 \end{bmatrix}$$

Ex. 2.13 — Considere a matriz de dados $\underset{\sim}{\mathbf{x}} = \begin{bmatrix} 2 & 8 & 6 & 8 \\ 12 & 9 & 9 & 10 \end{bmatrix}'$ para a qual se tem a matriz de covariâncias amostral

$$\mathbf{s}_c = \begin{bmatrix} 8 & -\frac{10}{3} \\ -\frac{10}{3} & 2 \end{bmatrix}, \text{ sendo } \mathbf{s}_c^{-1} = \begin{bmatrix} 0.409 & 0.682 \\ 0.682 & 1.636 \end{bmatrix}$$

Assumindo que os dados provêm de uma população normal,

1. determine a elipse fronteira da região que contém $\boldsymbol{\mu}$ com uma confiança de 95%;
2. teste, ao nível de significância de 5%,

$$H_0 : \boldsymbol{\mu} = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} \quad \text{vs} \quad H_1 : \boldsymbol{\mu} \neq \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}.$$

Ex. 2.14 — Considere uma amostra com $n = 42$ observações de uma população normal bivariada de média $\boldsymbol{\mu}$. Da amostra obteve-se:

$$\bar{\mathbf{x}} = \begin{bmatrix} 0.564 \\ 0.603 \end{bmatrix} \quad \text{e} \quad \mathbf{s}_c = \begin{bmatrix} 0.0144 & 0.0017 \\ & 0.0146 \end{bmatrix}, \quad \text{sendo que} \quad \mathbf{s}_c^{-1} = \begin{bmatrix} 70.41 & -8.20 \\ & 69.45 \end{bmatrix}.$$

Teste, ao nível de significância de 1%,

$$H_0 \ : \quad \boldsymbol{\mu} = \begin{bmatrix} 7 \\ 11 \end{bmatrix} \quad \text{vs} \quad H_1 \ : \quad \boldsymbol{\mu} \neq \begin{bmatrix} 7 \\ 11 \end{bmatrix}$$

Ex. 2.15 — As pontuações obtida por $n = 87$ estudantes a três disciplinas A, B e C, foram registadas sendo que para a amostra dada se obteve:

$$\bar{\mathbf{x}} = \begin{bmatrix} 527.74 \\ 54.69 \\ 25.13 \end{bmatrix} \quad \text{e} \quad \mathbf{s}_c = \begin{bmatrix} 5691.34 & & \\ 600.51 & 126.05 & \\ 217.25 & 23.34 & 23.11 \end{bmatrix}, \quad \text{sendo que}$$

$$\mathbf{s}_c^{-1} = 10^{-3} \begin{bmatrix} 0.448 & & \\ -1.666 & 15.955 & \\ -2.529 & -0.449 & 67.502 \end{bmatrix}.$$

1. Encontre os intervalos simultâneos de Roy para μ_1, μ_2, μ_3 e $\frac{1}{3}(\mu_1 + \mu_2 + \mu_3)$ a 95% de confiança.
2. Encontre os intervalos simultâneos segundo Bonferroni para μ_1, μ_2, μ_3 e $\frac{1}{3}(\mu_1 + \mu_2 + \mu_3)$ a 95% de confiança.

3. Compare a amplitude dos intervalos correspondentes calculados acima e comente.
4. Obtenha a região a 95% de confiança para $\boldsymbol{\mu}$.
5. Com base nas alíneas anteriores, há evidência estatisticamente significativa para concluir que $\boldsymbol{\mu} = [500\ 50\ 50]'$?

Ex. 2.16 — Sejam $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{10}$ dez observações 3-dimensionais independentes, extraídas de uma população $\mathbf{X} \frown N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, de vetor média amostral e matriz de covariâncias amostral dados por:

$$\bar{\mathbf{x}} = \begin{bmatrix} 194.5 \\ 136.9 \\ 185.9 \end{bmatrix} \qquad \mathbf{s}_c = \begin{bmatrix} 187.6 & 45.9 & 113.6 \\ & 69.2 & 15.3 \\ & & 239.9 \end{bmatrix}$$

sendo

$$\mathbf{s}_c^{-1} = \begin{bmatrix} 0.0089 & & \\ -0.0051 & 0.1754 & \\ -0.0039 & 0.0013 & 0.0059 \end{bmatrix}$$

1. Obtenha os intervalos de confiança simultâneos de Roy para μ_i, $i = 1, 2, 3$, com grau de confiança de 95%; recorrendo ao construa uma representação gráfica desses intervalos;
2. Obtenha intervalos de confiança simultâneos para μ_i, $i = 1, 2, 3$, com grau de confiança de 95% e usando o método de Bonferroni; recorrendo ao construa uma representação gráfica desses intervalos;
3. Obtenha a região a 95% de confiança para $\boldsymbol{\mu}$; recorrendo ao construa uma representação gráfica dessa região;
4. Faça uma análise comparativa dos três gráficos obtidos;
5. Com base nas alíneas anteriores, conclua se $\boldsymbol{\mu}$ é significativamente diferente de $[195\ 135\ 185]'$, a um nível de significância a sua escolha.

Ex. 2.17 — Oito pacientes receberam 2.0 mg/kg de uma certa droga. As alterações nos níveis de açúcar no sangue (glicose) e a pressão arterial (sistólica e diastólica) foram registadas na seguinte tabela:

	Paciente							
Variáveis	1	2	3	4	5	6	7	8
Glicose (x_1)	30	90	-10	10	30	60	0	40
Pressão sistólica (x_2)	-8	7	-2	0	-2	0	-2	1
Pressão diastólica (x_3)	-1	6	4	2	5	3	4	2

1. Determine o vetor média amostral, a matriz de covariâncias amostral (corrigida) e a sua inversa.
2. Determine as distâncias de Mahalanobis observadas para os 8 pacientes e determine o número de pacientes que caem fora da linha de contorno de probabilidade 50% e 75% da normal multivariada 3-dimensional, de média e matriz de covariâncias estimadas da amostra.
3. construa um gráfico de qui-quadrado (*chi-square plot*) das distâncias ordenadas calculadas em questão anterior 2. Analise o gráfico comentando sobre a validade da normalidade do v.a. (X_1, X_2, X_3) de onde foram extraídas as observações (x_1, x_2, x_3) dos 8 pacientes.

4. Assumindo a normalidade dos dados
   (a) Teste, ao nível de significância de 5%, $H_0 : \ \boldsymbol{\mu} = \mathbf{0} \quad \text{vs} \quad H_1 \ : \quad \boldsymbol{\mu} \neq \mathbf{0}$.
   (b) Obtenha a região a 95% de confiança para $\boldsymbol{\mu}$.
   (c) Obtenha intervalos de confiança simultâneos de Roy para μ_i, $i = 1, 2, 3$, com grau de confiança de 95%.
   (d) Obtenha intervalos de confiança simultâneos de Roy para $\mu_1 - \mu_2$ e $\frac{\mu_1+\mu_2}{2}$, com grau de confiança de 95%.
   (e) Obtenha intervalos de confiança simultâneos para μ_i, $i = 1, 2, 3$, com grau de confiança de 95% e usando o método de Bonferroni.

2.2 Soluções

Resposta do ex. 2.1 — 1. $N(0,1)$,$\forall$ marginal; 2. $N_2\left(\mathbf{0}, \begin{bmatrix} 2-2\rho & 0 \\ 0 & 6-6\rho \end{bmatrix}\right)$; 3. $\chi^2_{(3)}$

Resposta do ex. 2.2 — —

Resposta do ex. 2.6 — —

Capítulo 3

Análise Discriminante

3.1 Exercícios

Ex. 3.1 — Considere dois conjuntos de dados

$$\underset{\sim}{\mathbf{x}}_1 = \begin{bmatrix} 3 & 2 & 4 \\ 7 & 4 & 7 \end{bmatrix} \quad \text{e} \quad \underset{\sim}{\mathbf{x}}_2 = \begin{bmatrix} 6 & 5 & 4 \\ 9 & 7 & 8 \end{bmatrix}$$

para os quais

$$\bar{\mathbf{x}}_1 = \begin{bmatrix} 3 \\ 6 \end{bmatrix}, \qquad \bar{\mathbf{x}}_2 = \begin{bmatrix} 5 \\ 8 \end{bmatrix} \quad \text{e} \quad \mathbf{S}_p = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}.$$

Assuma a normalidade das populações Π_1 e Π_2, com matrizes de covariâncias iguais, onde foram extraidos os dados $\underset{\sim}{\mathbf{x}}_1$ e $\underset{\sim}{\mathbf{x}}_2$, respetivamente.

1. Determine a função discriminante linear associada à Regra de custo esperado de má classificação (*expected cost of misclassification, ECM*) mínimo, com iguais probabilidades à priori e iguais custos de má classificação;
2. Classifique a observação $\mathbf{x} = [2\ 7]'$ para uma das populações Π_1 e Π_2, usando a regra de classificação mencionada na alínea anterior;
3. Faça uma representação gráfica das observações $\underset{\sim}{\mathbf{x}}_1$ e $\underset{\sim}{\mathbf{x}}_2$ e da função discriminante considerada na alínea 1. Interprete a regra de classificação obtida na alínea anterior.

Ex. 3.2 — Um investigador pretende obter um procedimento para discriminar entre duas populações multivariadas. O investigador tem dados suficientes para estimar as funções densidade $f_1(\mathbf{x})$ e $f_2(\mathbf{x})$ associadas às populações Π_1 e Π_2, respectivamente. Seja $C(\Pi_2|\Pi_1) = 50$ e $C(\Pi_1|\Pi_2) = 100$. Além disso, é sabido que 20% de todos os indivíduos para os quais as medidas $\mathbf{x}$ podem ser registadas pertencem à população Π_2.

1. Encontre a regra de custo esperado de má classificação mínimo para alocar um novo individuo a uma das duas populações.
2. Medidas registadas de um novo indivíduo produziu valores para as densidades: $f_1(\mathbf{x}) = 0.3$ e $f_2(\mathbf{x}) = 0.5$. Com base na alínea anterior, aloque esse novo indivíduo em Π_1 ou Π_2.

Ex. 3.3 — Seja $f_1(x) = \frac{1}{2}(1 - |x|)$, para $|x| \leq 1$, e seja $f_2(x) = \frac{1}{2}(1 - |x - 0.5|)$, para $-0.5 \leq x \leq 1.5$.

1. Obtenha a Regra de classificação de custo esperado de má classificação mínimo quando $p_1 = p_2$ e $C(\Pi_2|\Pi_1) = C(\Pi_1|\Pi_2)$.
2. Obtenha a Regra de classificação de custo esperado de má classificação mínimo quando $p_1 = 0.2$ e $C(\Pi_2|\Pi_1) = C(\Pi_1|\Pi_2)$.

Ex. 3.4 — Suponha que duas amostras, com $n_1 = 11$ e $n_2 = 12$ observações, são extraídas de populações normais multivariadas $\mathbf{X}_1$ e $\mathbf{X}_2$, com iguais matrizes de covariâncias, respetivamente. Os vetores de média amostrais e a matriz de covariâncias combinada (*pooled*) são:

$$\bar{\mathbf{x}}_1 = \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \qquad \bar{\mathbf{x}}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \quad \text{e} \quad \mathbf{s}_p = \begin{bmatrix} 7.3 & -1.1 \\ -1.1 & 4.8 \end{bmatrix} \text{ com } \mathbf{s}_p^{-1} = \begin{bmatrix} 0.14 & 0.03 \\ 0.03 & 0.22 \end{bmatrix}.$$

1. Construa a função discriminante linear (amostral) de Fisher.
2. Classifique a observação $\underset{\sim}{\mathbf{x}} = [0\ 1]'$ para uma das populações $\mathbf{X}_1$ e $\mathbf{X}_2$, usando a regra de classificação baseada no discriminante linear de Fisher obtido na alínea anterior.

3.2 Soluções

Resposta do ex. 3.1 — 1) $\mathcal{F}(\mathbf{x}) = -2x_1$; 2) Regra: $x \in \Pi_1$ se $x_1 \leq 4$; 3) —

Resposta do ex. 3.2 — 1) $\mathcal{R}_1 = \left\{x : \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \geq \frac{1}{2}\right\}$; 2)Π_1.

Resposta do ex. 3.3 — 1) $\mathcal{R}_1 = [-1, 0.25]$ e $\mathcal{R}_2 =]0.25, 1.5]$; 2) $\mathcal{R}_1 = [-1, -1/3]$ e $\mathcal{R}_2 =]-1/3, 1.5]$

Resposta do ex. 3.4 — 1) $\mathcal{F}(\mathbf{x}) = -0.4907x_1 - 0.5291x_2$; 2) Π_2.

Testes de Avaliação

Compilação de provas de avaliação sobre os Capítulos 1, 5 e 6 de Estatística Multivariada

Prova: 1º Teste, 2013/14

Ex. 0.1 — Seja $\mathbf{X} = [X_1\ X_2\ X_3\ X_4]' \frown N_4(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ com

$$\boldsymbol{\mu} = \begin{bmatrix} 5 \\ 6 \\ 7 \\ 8 \end{bmatrix} \quad \text{e} \quad \boldsymbol{\Sigma} = \begin{bmatrix} 2 & 0 & 1 & 0 \\ 0 & 3 & 2 & 0 \\ 1 & 2 & 4 & 0 \\ 0 & 0 & 0 & 9 \end{bmatrix}.$$

(a)Calcule $E\,(\,\mathbf{X} - \mathbf{a}\,)$, com $\mathbf{a} = [8\ 8\ 8\ 8]'$.

(b)Identifique a distribuição de $[X_2\ X_4]'$.

(c)Determine a média e a variância de $Z = X_1 + 2X_3 - X_4 + 2$

Ex. 0.2 — Sejam $Y_1, Y_2, \cdots, Y_m$ m v.a.'s i.i.d. com distribuição $N(\mu = 2, \sigma^2 = 1)$. Definem-se as v.a.'s $Z_i = Y_{i+1} - Y_i,\ \ i = 1, 2, \cdots, m-1$.
Encontre a distribuição do v.a. $\mathbf{Z} = [Z_1\ Z_2\ \cdots\ Z_{m-1}]'$.

Ex. 0.3 — Um tratamento foi administrado a 8 pacientes e respostas a duas medidas, x_1, x_2, foram registadas. Na tabela abaixo são ainda apresentados os valores das distâncias quadradas $d_j^2 = (\mathbf{x}_j - \bar{\mathbf{x}})'\mathbf{s}_c^{-1}(\mathbf{x}_j - \bar{\mathbf{x}}),\ j = 1, 2, \cdots, 8$, onde $\bar{\mathbf{x}}$ é a média amostral e $\mathbf{s}_c$ é a matriz de covariâncias amostral.

	paciente j							
Medida	1	2	3	4	5	6	7	8
x_1	3	4	1	2	3	1	3	2
x_2	5	5	2	3	3	3	5	3
d_j^2	1.7	2.4	2.0	0.3	3.2	2.5	1.7	0.3

sendo

$$\bar{\mathbf{x}} = \begin{bmatrix} 2.38 \\ 3.63 \end{bmatrix} \quad \mathbf{s}_c = \begin{bmatrix} 1.13 & 1.02 \\ 1.02 & 1.41 \end{bmatrix} \quad \text{e} \quad \mathbf{s}_c^{-1} = \begin{bmatrix} 2.56 & -1.85 \\ -1.85 & 2.04 \end{bmatrix}$$

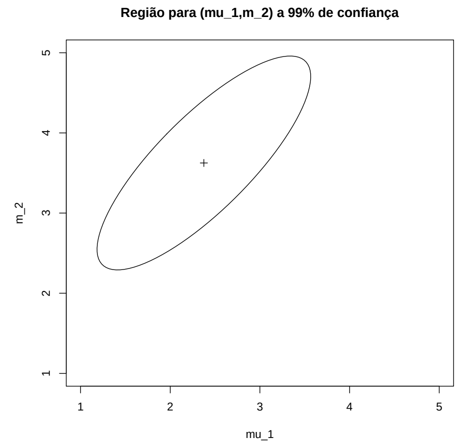
1. Determine a proporção de observações que caem dentro do contorno da distribuição normal bivariada, de média estimada por $\bar{\mathbf{x}}$ e matriz de covariâncias estimada por $\mathbf{s}_c$, com probabilidade de 75%.

2. Assuma a normalidade dos dados.

   (a) Obtenha intervalos de confiança simultâneos para μ_i, $i = 1, 2$, com grau de confiança de 99% e usando o método de Bonferroni.

   (b) Obtenha intervalos de confiança simultâneos de Roy para μ_i, $i = 1, 2$, com grau de confiança de 99%.

   (c) Com base na amostra, obteve-se a região a 99% de confiança para $\boldsymbol{\mu}$ dada geometricamente pela região interior delimitada pela elipse graficamente representada na seguinte figura.



   Conclua quanto ao teste $H_0 : \boldsymbol{\mu} = \begin{bmatrix} 4 \\ 3 \end{bmatrix}$, a um nível de significância a indicar.

Ex. 0.4 — Suponha que duas amostras, com $n_1 = n_2 = 20$ observações, são extraídas de populações bivariadas $\mathbf{X}_1$ e $\mathbf{X}_2$, respetivamente. Calculado o vetor de médias amostrais e a inversa da matriz de covariâncias amostral combinada (*pooled*) obteve-se:

$$\bar{\mathbf{x}}_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \qquad \bar{\mathbf{x}}_2 = \begin{bmatrix} 0 \\ 2 \end{bmatrix} \quad \text{e} \quad \mathbf{s}_p^{-1} = \begin{bmatrix} 2 & 1 \\ 1 & 5 \end{bmatrix}.$$

1. Obtenha a regra de classificação baseada no discriminante linear (amostral) de Fisher.

2. Classifique uma observação $\mathbf{x}_0 = [0\ 1]'$ para uma das populações $\mathbf{X}_1$ e $\mathbf{X}_2$, usando a regra obtida na alínea anterior.

Prova: 1º Teste, 2014/15

Ex. 0.5 — Seja $\mathbf{X} = [X_1\ X_2\ X_3]' \frown N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ com

$$\boldsymbol{\mu} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad \text{e} \quad \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & a & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

para qualquer $a \in \mathbb{R}$ tal que $\mathbf{\Sigma}$ satisfaça as propriedades necessárias para ser uma matriz de covariâncias.

(a)Calcule $E\,(1 \;-\; \mathbf{c}'\mathbf{X}\,)$, com $\mathbf{c} = [-1\ 0\ 1]'$.

(b)Considere-se $a = 3$. Identifique a distribuição de $\mathbf{Y} = [Y_1\ Y_2]'$ com

$$Y_1 = X_1 - X_2 \quad \text{e} \quad Y_2 = \frac{X_1 + X_2 + X_3}{3}$$

(c)Determine o conjunto de valor(es) de $a \in \mathbb{R}$ tal que $\mathbf{Z} = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}' = \begin{bmatrix} X_1 + aX_2 \\ X_2 + aX_3 \end{bmatrix} \frown N_2(\boldsymbol{\mu}_{\mathbf{Z}}, \mathbf{\Sigma}_{\mathbf{Z}})$, com $Cov(Z_1, Z_2) = 2$.

(d)Seja $\mathbf{X}_1, \mathbf{X}_2, \cdot, \mathbf{X}_{20}$ uma amostra aleatória de tamanho $n = 20$ extraída de $\mathbf{X}$.

(a) Identifique a distribuição de probabilidade de $(\mathbf{X}_1 - \boldsymbol{\mu})'\ \mathbf{\Sigma}^{-1}\ (\mathbf{X}_1 - \boldsymbol{\mu})$;

(b) Calcule $P\left((\bar{\mathbf{X}} - \boldsymbol{\mu})'\ \mathbf{\Sigma}^{-1}\ (\bar{\mathbf{X}} - \boldsymbol{\mu}) > 1\right)$.

Ex. 0.6 — A transpiração de 20 fêmeas saudáveis foram analisadas. Duas componentes:

$$\begin{aligned} X_1 &= \text{taxa de sudorese} \\ X_2 &= \text{teor de potásio} \end{aligned}$$

foram medidas. Dos resultados observados obteve-se a média amostral $\bar{\mathbf{x}}$ e a matriz de covariâncias amostral $\mathbf{s}_c$ dadas por:

$$\bar{\mathbf{x}} = \begin{bmatrix} 4.6 \\ 10.0 \end{bmatrix} \qquad \mathbf{s}_c = \begin{bmatrix} 2.9 & -1.8 \\ -1.8 & 3.6 \end{bmatrix} \quad \text{sendo} \quad \mathbf{s}_c^{-1} = \begin{bmatrix} 0.50 & 0.25 \\ 0.25 & 0.40 \end{bmatrix}$$

1.Obtenha a matriz de correlações amostral.

2.Assuma a normalidade dos dados.

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

$$H_0 : \boldsymbol{\mu} = \begin{bmatrix} 6 \\ 9 \end{bmatrix} \quad \text{vs} \quad H_1 : \boldsymbol{\mu} \neq \begin{bmatrix} 6 \\ 9 \end{bmatrix}$$

(b)Obtenha intervalos de confiança simultâneos de Roy para μ_1, μ_2 e $\frac{\mu_1+\mu_2}{2}$ com grau de confiança de 95%.

(c)Explique, justificando, o que aconteceria à amplitude dos intervalos obtidos na alínea anterior se tivesse considerado o procedimento segundo Bonferroni.

Ex. 0.7 — Considere-se duas populações normais multivariadas, $\mathbf{X}_1$ e $\mathbf{X}_2$, sendo

$$\begin{aligned} \mathbf{X}_1 &\frown N_3(\boldsymbol{\mu}, \mathbf{\Sigma}) \\ \mathbf{X}_2 &\frown N_3(a\,\boldsymbol{\mu}, a\,\mathbf{\Sigma}) \quad \text{com} \quad a > 1. \end{aligned}$$

1. Mostre que a regra de Máxima Verosimilhança de custo esperado de má classificação mínimo, em função do valor de a, quando o custo de má classificação dentro da população $\mathbf{X}_1$ é o dobro ao custo de má classificação dentro da população $\mathbf{X}_2$ e é duas vezes mais provável que um indivíduo pertencem à população $\mathbf{X}_1$, é dada por:

$$\begin{aligned}\mathcal{R}_1 &= \left\{\mathbf{x} \in \mathbb{R}^3 \,:\, \mathbf{x}'\boldsymbol{\Sigma}^{-1}\mathbf{x} - 2a\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} \leq \frac{a\ln(16a)}{a-1}\right\} \\ \mathcal{R}_2 &= \mathbb{R}^3 \setminus \mathcal{R}_1\end{aligned}$$

2. Tome $a = 2$ e suponha que duas amostras, com $n_1 = n_2 = 10$ observações, são extraídas das populações $\mathbf{X}_1$ e $\mathbf{X}_2$, respetivamente. Calculado o vetor de médias amostrais e a matriz de covariâncias amostral combinada (*pooled*) obteve-se:

$$\bar{\mathbf{x}}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad \bar{\mathbf{x}}_2 = \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} \quad \text{e} \quad \mathbf{s}_p = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 10 & 3 \\ 0 & 3 & 1 \end{bmatrix} \quad \text{sendo} \quad \mathbf{s}_p^{-1} = \begin{bmatrix} 0.5 & 0 & 0 \\ 0 & 1 & -3 \\ 0 & -3 & 10 \end{bmatrix}.$$

Obtenha a regra discriminante de Fisher e classifique uma observação $\mathbf{x}_0 = [0\ 1\ 0]'$ para uma das populações, $\mathbf{X}_1$ ou $\mathbf{X}_2$, usando essa regra.

Prova: 1º Teste, 2015/16

Ex. 0.8 — Seja $\mathbf{X} = [X_1\ X_2\ X_3]' \frown N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ com

$$\boldsymbol{\mu} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \quad \text{e} \quad \boldsymbol{\Sigma} = \begin{bmatrix} 11 & -6 & 2 \\ -6 & 10 & -4 \\ 2 & -4 & 6 \end{bmatrix}.$$

(a) Encontre $E(X_1)$, $E(X_1^2)$ e $E(X_1X_2)$.

(b) Identifique a distribuição de $\mathbf{Y} = [Y_1\ Y_2]'$ com

$$Y_1 = X_1 + 2 \quad \text{e} \quad Y_2 = X_1 + X_2 + X_3$$

(c) Identifique a distribuição de X_3 condicionada a $[X_1\ X_2]'$.

Ex. 0.9 — Considere o seguinte Teorema que permite reconstruir a distribuição conjunta sabendo uma distribuição marginal $\mathbf{X}_1$ e a distribuição condicional de $\mathbf{X}_2|\mathbf{X}_1$.

Teorema: Se $\mathbf{X}_1 \frown N_r(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})$ e $\mathbf{X}_2|\mathbf{X}_1 = \mathbf{x}_1 \frown N_{p-r}(\mathbf{A}\mathbf{x}_1 + \mathbf{b}, \boldsymbol{\Sigma}^*)$, onde $\boldsymbol{\Sigma}^*$ não depende de $\mathbf{x}_1$, então:

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \frown N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

com

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \mathbf{A}\boldsymbol{\mu}_1 + \mathbf{b} \end{bmatrix} \quad \text{e} \quad \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{11}\mathbf{A}' \\ \mathbf{A}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}^* + \mathbf{A}\boldsymbol{\Sigma}_{11}\mathbf{A}' \end{bmatrix}$$

Usando este Teorema, encontre a distribuição de $[X\ Y\ Z]'$ sabendo que

$$Z \frown N_1(0,1), \qquad Y|Z = z \frown N_1(1+z, 1), \qquad \text{e} \qquad X|Y = y, Z = z \frown N_1(1-Y, 1)$$

Sugestão: Comece por obter a distribuição de $[Y\ Z]'$.

Ex. 0.10 — Um tratamento foi administrado a 5 pacientes e respostas a duas medidas, x_1, x_2, foram registadas. Na tabela abaixo são ainda apresentados os valores das distâncias quadradas $d_j^2 = (\mathbf{x}_j - \bar{\mathbf{x}})'\mathbf{s}_c^{-1}(\mathbf{x}_j - \bar{\mathbf{x}})$, $j = 1, 2, \cdots, 5$, onde $\bar{\mathbf{x}}$ é a média amostral e $\mathbf{s}_c$ é a matriz de covariâncias amostral (corrigida).

	paciente j				
Medida	1	2	3	4	5
x_1	3	4	1	2	3
x_2	5	5	2	3	3
d_j^2	2.1	1.5	2.0	0.3	2.1

sendo $\mathbf{s}_c = \begin{bmatrix} 1.3 & 1.3 \\ 1.3 & 1.8 \end{bmatrix}$ e $\mathbf{s}_c^{-1} = \begin{bmatrix} 2.8 & -2 \\ -2 & 2 \end{bmatrix}$

1. Obtenha o vector média amostral, $\bar{\mathbf{x}}$.
2. Determine a proporção de observações que caem dentro do contorno da distribuição normal bivariada, de média estimada por $\bar{\mathbf{x}}$ e matriz de covariâncias estimada por $\mathbf{s}_c$, com probabilidade de 25%.
3. Assuma a normalidade dos dados.
   (a) Teste, ao nível de significância de 5%,
   $$H_0 : \boldsymbol{\mu} = \begin{bmatrix} 2 \\ 4 \end{bmatrix} \quad \text{vs} \quad H_1 : \boldsymbol{\mu} \neq \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$
   (b) Obtenha intervalos de confiança simultâneos para μ_i, $i = 1, 2$, com grau de confiança de 95% e usando o método de Roy.
   (c) Com base nos intervalos calculados na alínea anterior, qual a sua decisão no teste da alínea 3a? Compare com a decisão tomada nessa alínea 3a e comente os resultados.

Ex. 0.11 — Suponha que duas amostras, com $n_1 = 8$ e $n_2 = 12$ observações, são extraídas das populações $\mathbf{X}_1$ e $\mathbf{X}_2$, respetivamente. Calculado o vetor de médias amostrais e as matrizes de covariâncias amostrais (corrigidas) correspondentes às duas amostras obteve-se:

$$\bar{\mathbf{x}}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \bar{\mathbf{x}}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad \text{e} \quad \mathbf{s}_1 = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix} \quad 1 \quad \mathbf{s}_2 = \begin{bmatrix} 0.5 & 0 \\ 0 & 1 \end{bmatrix}.$$

1. Obtenha a matriz de covariâncias amostral combinada (corrigida), $\mathbf{s}_p$, a custa de $\mathbf{s}_1$ e $\mathbf{s}_2$.
2. Sabendo que
   $$\mathbf{s}_p^{-1} = \begin{bmatrix} 1.00 & -0.22 \\ -0.22 & 0.61 \end{bmatrix},$$
   obtenha a regra discriminante de Fisher e classifique uma observação $\mathbf{x}_0 = [0.5\ 1]'$ para uma das populações, $\mathbf{X}_1$ ou $\mathbf{X}_2$, usando essa regra.

Prova: 1º Teste, 2016/17

Ex. 0.12 — Seja o v.a. $\mathbf{X} = [X_1\ \ X_2]' \frown N_2\left(\boldsymbol{\mu} = \begin{bmatrix} -4 \\ 2 \end{bmatrix}, \boldsymbol{\Sigma} = \begin{bmatrix} 6 & 1 \\ 1 & 3 \end{bmatrix}\right).$

1. Considere a v.a. Y tal que $[X_1 \;\; X_2 \;\; Y]'$ tem distribuição normal sendo $E(Y^2) = 5E(Y) = 10$, $E(X_1Y) = -3$ e $E(X_2Y) = 5$.
   (a) Obtenha a distribuição de $[X_1 \;\; X_2 \;\; Y]'$.
   (b) Obtenha a distribuição de $\mathbf{X}$ condicionada a Y .
2. Considere agora o v.a. $\mathbf{Z} = [X_1 \;\; aX_2 \;\; X_1]'$, para alguma constante a real.
   (a) Justifique que $\mathbf{Z}$ não tem distribuição normal qualquer que seja $a \neq 0$.
   (b) Encontre a esperança e a matriz de covariâncias de $\mathbf{Z}$.
   (c) Identifique a matriz de correlações de $\mathbf{Z}$ (com $a \neq 0$).
3. Seja $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_{50}$ uma amostra aleatória extraída de uma população $\mathbf{X}$ e seja $\bar{\mathbf{X}}$ o v.a. de médias amostrais.
   (a) Indique a distribuição de $(\bar{\mathbf{X}} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\bar{\mathbf{X}} - \boldsymbol{\mu})$.
   (b) Partindo do TLC (multivariado), mostre que

$$\frac{\bar{X}_1^2 + \bar{X}_2^2}{2} \frown_{aprox.} N(\mu = 10, \sigma^2 = 1.84)$$

Ex. 0.13 — Suponha que três questionários que permitem avaliar as atitudes de estudantes universitários de um curso de Engenharia face à Estatística (x_1), à Matemática (x_2) e ao Inglês (x_3) foram respondidos por 40 jovens universitários (25 do 1.º ano do curso e 15 do último ano do curso). Assuma que os dados provêm de uma população com distribuição normal 3-dimensional. Observou-se que os vetores de médias amostrais e as matrizes de covariâncias amostrais (corrigidas) correspondentes a cada amostra (1.º e último ano) de pontuações médias são:

- Amostra de 25 jovens universitários (do <u>1.º ano</u> do curso de Engenharia):

$$\bar{\mathbf{x}}_A = \begin{bmatrix} 14 \\ 16 \\ 14 \end{bmatrix}, \quad \mathbf{s}_{c_A} = \begin{bmatrix} 21 & -1.5 & 1 \\ & 3.0 & -0.5 \\ & & 10 \end{bmatrix} \quad \text{com } \mathbf{s}_{c_A}^{-1} = \begin{bmatrix} 0.05 & 0.02 & 0 \\ & 0.35 & 0.01 \\ & & 0.10 \end{bmatrix}.$$

- Amostra de 15 jovens (do <u>último ano</u> do curso de Engenharia):

$$\bar{\mathbf{x}}_B = \begin{bmatrix} 17 \\ 16 \\ 16 \end{bmatrix}$$

.

1. Sabendo que a inversa da matriz de covariâncias amostral (corrigida) combinada $\mathbf{s}_p$ é igual a:

$$\mathbf{s}_p^{-1} = \begin{bmatrix} 0.05 & 0.02 & 0 \\ & 0.34 & 0.01 \\ & & 0.06 \end{bmatrix},$$

   obtenha a regra discriminante de Fisher e, com base nela, diga qual o ano de curso (primeiro ou último) que suspeitaria de um estudante que, ao responder aos três questionários, obtivesse uma pontuação de 17 para Estatística, 15 para Matemática e 14 para Inglês.
2. Obtenha o vetor de médias amostrais associado ao total dos 40 jovens observados (sem discriminação de género).

3.Considere agora apenas a amostra dos estudantes do primeiro ano ($n = \mathbf{25}$)

(a)Teste, ao nível de significância de 1%,

$$H_0 : \boldsymbol{\mu} = \begin{bmatrix} 15 \\ 15 \\ 15 \end{bmatrix} \quad \text{vs} \quad H_1 : \boldsymbol{\mu} \neq \begin{bmatrix} 15 \\ 15 \\ 15 \end{bmatrix}$$

(b)Obtenha intervalos a 99% de confiança simultâneos para μ_i, $i = 1, 2, 3$, usando o método de Roy.

(c)Verifique que, tomando os intervalos de Roy calculados na alínea anterior, a sua decisão relativamente ao teste de hipóteses realizado acima não mudaria. Tal seria expectável? Justifique a sua resposta.

(d)Tendo em conta o resultado obtido do (figura ao lado) sobre os valores próprios (por ordem decrescente) e os correspondentes vetores próprios da matriz de covariâncias amostral (corrigida), determine o comprimento do eixo maior que define o elipsóide de fronteira da região a 99% de confiança de (μ_1, μ_2, μ_3) e identifique o ângulo que esse eixo faz com o plano XOY.
Nota. Sobre o ângulo, a resposta pode ser dada em termos da função arctan(.)

```
> print(eigen(scA),2)
$values
[1] 21.2  9.9  2.9

$vectors
       [,1]   [,2]   [,3]
[1,]  0.992 -0.097 -0.079
[2,] -0.084 -0.051 -0.995
[3,]  0.092  0.994 -0.059
```

(e)Pretende-se testar:

$$H_0 : 0.9\mu_1 = 0.9\mu_2 = \mu_3 \quad \text{vs} \quad H_1 : \mu_1 \neq \mu_2 \text{ ou } 0.9\mu_2 \neq \mu_3 \text{ ou } 0.9\mu_1 \neq \mu_3$$

Partindo de uma estatística com distribuição T^2 de Hotelling, mostre que, com base na amostra dos 25 estudantes, rejeitaria H_0, ao nível de significância de 1%, se:

$$[-2 \quad 0.4] \begin{bmatrix} 27 & -5.55 \\ -5.55 & 13.33 \end{bmatrix}^{-1} \begin{bmatrix} -2 \\ 0.4 \end{bmatrix} > 0.47$$

Sugestão: Considere o facto do teste pretendido ser equivalente a testar

$$H_0 : \begin{bmatrix} \mu_1 - \mu_2 \\ 0.9\mu_2 - \mu_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{vs} \quad H_1 : \begin{bmatrix} \mu_1 - \mu_2 \\ 0.9\mu_2 - \mu_3 \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

e aborde o problema inferencial sobre a população $\begin{bmatrix} X_1^* \\ X_2^* \end{bmatrix} = \begin{bmatrix} X_1 - X_2 \\ 0.9X_2 - X_3 \end{bmatrix}$.

Soluções

Resposta do ex. 0.1 — a) $[-3 \ -2 \ -1 \ 0]'$; $N_4\left([6 \ 8]', \begin{bmatrix} 3 & 0 \\ 0 & 9 \end{bmatrix}\right)$; c) média 13; Variância 31.

Resposta do ex. 0.2 — $N_{m-1}\left(\mathbf{0}, \begin{bmatrix} 2 & -1 & 0 & \cdots & 0 & 0 \\ -1 & 2 & -1 & \cdots & 0 & 0 \\ 0 & -1 & 2 & \cdots & 0 & 0 \\ \vdots & & & & & \\ 0 & 0 & 0 & \cdots & -1 & 0 \end{bmatrix}\right)$

Resposta do ex. 0.3 — 1) 7/8; 2) (a) $IC_{99\%}(\mu_1) =]0.866, 3.894[$, $IC_{99\%}(\mu_2) =]1.938, 5.322[$; (b) $IC_{99\%}(\mu_1) =]0.482, 4.278[$, $IC_{99\%}(\mu_2) =]1.510, 5.750[$; (c) Com base na região dada na figura, rejeita-se H_0, ao nível de significância de 1%.

Resposta do ex. 0.4 — a) $\mathcal{R}_1 = \{\mathbf{x} \in \mathbb{R}^2 : \ 2x_1 - 2x_2 \geq -1\}$ e $\mathcal{R}_2 = \mathbb{R}^2 \setminus \mathcal{R}_1$; b) $\mathbf{X}_2$.

Resposta do ex. 0.5 — FALTA

Resposta do ex. 0.6 — FALTA

Resposta do ex. 0.7 — FALTA

Resposta do ex. 0.8 — a) $E(X_1) = 1$, $E(X_1^2) = 12$ e $E(X_1X_2) = -6$; b) $\mathbf{Y} \frown N_2\left([3 \ 0]', \begin{bmatrix} 11 & 7 \\ 7 & 11 \end{bmatrix}\right)$;
c) $X_3|X_1, X_2 \frown N_1\left(-\frac{1}{74}(70 + 4X_1 + 32X_2)\frac{324}{74}\right)$.

Resposta do ex. 0.9 — $[X \ Y \ Z]' \frown N_3\left([0 \ 1 \ 0]', \begin{bmatrix} 3 & -2 & -1 \\ -2 & 2 & 1 \\ -1 & 1 & 1 \end{bmatrix}\right)$

Resposta do ex. 0.10 — a) $\bar{\mathbf{x}} = [2.6 \ 3.6]'$; b) 0.2 ; c) i) Com base na amostra, não existem evidências estatísticas para rejeitar H_0, ao nível de significância de 5%; ii)$IC_{95\%}(\mu_1)$=]0.03 , 5.17[, $IC_{95\%}(\mu_2)$=]0.57 , 6.63[iii) Usando os intervalos de Roy para μ_1 e μ_2 concluiríamos novamente pela não rejeição de H_0. A região de confiança é definida pela intersecção dos intervalos de Roy de todas as combinações lineares de μ_1 e μ_2. Ao basear a decisão em apenas 2 intervalos aumentamos o risco de tomar uma decisão errada (neste caso, como os dois intervalos definem uma região de confiança maior, então aumentamos a probabilidade de não rejeitar H_0).

Resposta do ex. 0.11 — a) $\mathbf{s}_p = \begin{bmatrix} 1.08 & 0.39 \\ 0.39 & 1.78 \end{bmatrix}$; b) Regra discriminante de Fisher: $\mathcal{R}_1 = \{(x_1, x_2) \in \mathbb{R}^2 \,:\, x_1 - 0.22x_2 \geq 0.5\}$ e $\mathcal{R}_2 = \mathbb{R}^2 \setminus \mathcal{R}_1$; $\mathbf{x}_0$ pertence à fronteira que separa as duas populações de acordo com a regra de Fisher. De acordo com a definição de $\mathcal{R}_1$ e $\mathcal{R}_2$, decide-se que $x_0 \in X_1$.

Resposta do ex. 0.12 — FALTA

Resposta do ex. 0.13 — FALTA