Análise Multivariada

Mestrado em Matemática e Aplicações

1º semestre – 2014/2015
13/01/2015 – 11:30

Duração: 3 horas **1º Exame**

Justifique convenientemente todas as respostas!

Grupo I — 7.5 valores

1. Seja $\boldsymbol{X}$ um vetor aleatório de dimensão p, com distribuição normal multivariada de vector de valor esperado $\boldsymbol{\mu}$ e matriz de covariâncias $\boldsymbol{\Sigma}$. Dadas duas matrizes de coeficientes $\boldsymbol{A}$ e $\boldsymbol{B}$ de dimensão $q \times p$ $(q \leq p)$ prove que $\boldsymbol{AX}$ e $\boldsymbol{BX}$ são vetores aleatórios independentes se e somente se $\boldsymbol{A\Sigma B}^t = \boldsymbol{0}$. (1.5)

2. Seja $\boldsymbol{X} = (X_1, X_2, X_3)^t$ um vetor aleatório com distribuição normal multivariada e $\boldsymbol{X}_2 = (X_2, X_3)^t$, tais que:

$$X_1 \sim \mathcal{N}(0,1) \quad \text{e} \quad (\boldsymbol{X}_2|X_1 = x_1) \sim \mathcal{N}_2\left(\begin{pmatrix} 2x_1 \\ x_1+1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$$

(a) Determine a distribuição de $\boldsymbol{X}$. (2.0)

(b) Determine a distribuição de $\boldsymbol{X}_2$. (1.0)

3. Um campo de trigo foi dividido em 30 parcelas iguais, às quais foram aplicadas um de 3 possíveis fertilizantes, escolhidos aleatoriamente e representados por A, B e C. O volume de trigo colhido em cada parcela conduziu aos seguintes resultados: (3.0)

Parcela	Tipo de fertilizante		
	A	B	C
1	4	6	2
2	3	7	1
3	2	7	1
4	5	5	1
5	4	5	3
6	4	5	4
7	3	8	3
8	3	9	3
9	3	9	2
10	1	6	2
Média	3.8	4.2	4.1

$$S_{n-1} = \begin{pmatrix} 498.848 & 141.613 & 74.344 \\ 141.613 & 352.777 & -1.163 \\ 74.344 & -1.163 & 187.054 \end{pmatrix},$$

Admita as condições que achar necessárias para testar a hipótese de os três fertilizantes, em média, conduzirem a iguais volumes de produção de trigo.

Grupo II — 2.5 valores

1. Seja U uma variável aleatória com distribuição uniforme no intervalo $[0,1]$, $\boldsymbol{a} \in \mathcal{R}^2 \backslash \{\boldsymbol{0}\}$ um vetor de constantes e $\boldsymbol{X} = U\boldsymbol{a}$. Determina as componentes principais associadas a $\boldsymbol{X}$. Comente os resultados obtidos. (2.5)

Grupo III	5.0 valores

1. Considere que tem os seguintes 5 objetos caracterizados por duas variáveis:

Objeto	x_1	x_2
1	1	1
2	2	1
3	4	5
4	7	7
5	5	7

(a) Calcule a matriz de distâncias *City-Block* entre os objectos. (1.0)

(b) Mostre que a distância $d_{k(ij)}$ (entre *clusters*) usada pelos métodos da ligação média (*group average*) satisfaz a relação: (1.5)

$$d_{k(ij)} = \frac{n_i}{n_i + n_j} d_{ki} + \frac{n_j}{n_i + n_j} d_{kj},$$

onde n_i representa o número de observações do *cluster* i.

(c) Faça a análise de *clusters* usando o método da ligação média, recorrendo à relação anterior e à matriz de distâncias obtida na alínea 1(a). Desenhe o respectivo dendrograma e diga em quantos *clusters* sugere dividir os dados. (2.5)

Sugestão: Se não resolveu 1(a), use a matriz de distâncias Euclideanas para resolver esta questão.

Grupo IV	5.0 valores

1. Uma variável aleatória tem distribuição normal de valor esperado zero e variância unitária se é observada num item dito regular e tem distribuição normal de valor esperado 7 e variância também unitária se for observada num item anómalo. Sabe-se que 10% dos item produzidos são anómalos.

(a) Apresente a regra de classificação baseada no critério que consiste em minimizar a probabilidade total de má classificação e calcule essa probabilidade. (2.0)

(b) Mostre que esta regra é equivalente a maximizar a probabilidade à posteriori. (1.0)

(c) Admita que desconhece os parâmetros da distribuição dos item regulares e anómalos, mas que continua a saber que as populações são normais com iguais variâncias. Sabe-se ainda que o custo de classificar erradamente um item anómalo como regular é 5 vezes mais elevado do que classificar erradamente um item regular como anómalo. Considerando que dispoe de 80 observações de itens regulares e 20 anómalos que conduziram aos seguintes valores amostrais: (2.0)

Classe	Média	Desvio-Padrão
Regulares	-0.11	1.11
Anómalos	6.87	1.22

Calcule a regra de classificação estimada associada a estes dados. Classifique um item caracterizado pelo valor 2.1.

FORMULÁRIO:

- Se $X \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ então $E(X_1|X_2 = x_2) = \boldsymbol{\mu}_1 - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(x_2 - \boldsymbol{\mu}_2)$ e $Var(X_1|X_2 = x_2) = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$.
- Se $U \sim Unif(a,b)$ então $E(U) = (a+b)/2$ e $Var(U) = (b-a)^2/12$.
- $T^2(p,n) = \frac{np}{n-p+1} F(p, n-p+1)$.
- $(n-1)(\bar{X} - \boldsymbol{\mu})^t S_n^{-1}(\bar{X} - \boldsymbol{\mu}) \sim T^2(p, n-1)$.