

# Descrição de dados Multivariados

EFT

# índice

## 1 Introdução

- Pasos Gerais
- Análise Univariante
- Sistemas de Coordenadas

## 2 Medida de Centralização

- O vetor de médias
- A matriz de variâncias e covariâncias

# Introdução

- Tipos de variáveis - codificação.
- Suposição: Observa-se uma amostra de variáveis numa amostra de elementos de uma população.
- Métodos para resumir os valores das variáveis e descrever a estrutura de dependência.

## Matriz de dados

$$X = \{x_{ij}\}$$

onde  $i = 1, \dots, n$  (indivíduo),  $j = 1, \dots, p$  (variável)

A matriz  $\mathbf{X}$  de ordem  $m \times n$  (com entrada geral  $x_{ij}$ ) é uma arranjo retangular de números reais dispostos em  $n$  linhas e  $m$  colunas.

Casos especiais para a designação de matrizes:

- 1. Se  $n = m$  então  $\mathbf{X}$  é uma matriz quadrada.
- 2. Se  $x_{ij} = x_{ji}$  para todo  $i, j$ , então  $\mathbf{X}$  é uma matriz simétrica.
- 3. Se  $x_{ij} = 0$  para todo  $i \neq j$ , então  $\mathbf{X}$  é uma matriz diagonal.
- 4. Se  $x_{ii} = 1$  e  $x_{ij} = 0$  para todo  $i \neq j$ , então  $\mathbf{X}$  é chamada de matriz Identidade e é representada por  $\mathbf{1}$ .

# Introdução

Descrever dados multivariados supõe estudar cada variável isoladamente e as relações entre elas.

- O estudo univariante da variável  $x_j$  implica calcular a média:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_{ij},$$

- se definirmos os desvíos da média como  $d_{ij} = (x_{ij} - \bar{x}_j)^2$ , define-se o desvio como:

$$s_j = \sqrt{\frac{\sum_i d_{ij}}{n}} = \sqrt{\frac{\sum_i (x_{ij} - \bar{x}_j)^2}{n}} \quad (1)$$

- o Coeficiente de Variação compara a variabilidade das distintas variáveis, e é definida como:  $CV_j = \sqrt{\frac{s_j^2}{\bar{x}_j^2}}$  supomos aquí que  $\bar{x}_j$  é diferente de zero.

# Introdução

- Calculamos também o coeficiente de assimetria (medem a simetria dos dados respeito do centro):  $A_j = \frac{1}{n} \frac{d_{ij}^3}{s_j^3}$ , este coeficiente é zero para uma variável simétrica. Se o coeficiente é aproximadamente maior do que 1, concluímos que os dados tem claramente uma distribuição assimétrica.
- Uma característica é também a homogeneidade, se os desvios  $d_{ij}$  são muito distintos, sugere que há muitos dados que se separam da média (alta heterogeneidade). Uma medida da variabilidade dos  $d_{ij}$  é  $\frac{1}{n} \sum_i (d_{ij} - s_j^2)^2$ , já que  $\bar{d}_j = s_j^2$ .

# Introdução

- O coeficiente de Homogeneidade é definido como

$H_j = \frac{\frac{1}{n} \sum_i (d_{ij} - s_j^2)^2}{s_j^4}$  que é sempre maior do que zero. Este coeficiente pode também ser escrito como:  $H_j = \frac{1}{n} \frac{\sum_i (x_{ij} - \bar{x}_j)^4}{s_j^4} - 1 = K_j - 1$  onde  $K$  é definido como o coeficiente de Kurtosis. Como  $H_j \geq 0$  o valor de  $K$  será sempre maior ou igual que 1.

# Sistemas

Podemos atribuir a cada ponto no sistema um par ordenado de números :  
 $\mathbf{a}'$ :

$$\mathbf{a}' = (a_1, a_2)$$

que pode também ser escrito como

$$\mathbf{a}' = a_1(1, 0) + a_2(0, 1)$$

o Vetor  $\mathbf{a}'$  pode ser representado no espaço de coordenadas como um segmento de linha direcionado ou como um ponto. A escolha dependerá do contexto.



## Exemplo

Temos as alturas e os pesos de 20 mulheres diferentes (dados reais em GB em 1951). Sejam  $X_1 = \textit{Altura}$  e  $X_2 = \textit{peso}$ . Temos que  $\bar{X}_1 = 62,85$  (1,60 m) e  $\bar{X}_2 = 123,6$  (60kg). Utilizamos  $X_{d1}$  e  $X_{d2}$  para representar os desvios da média, com médias iguais a zero e variâncias de 15,5 e 3,3 respectivamente. Os dados padronizados,  $(X_{s1}$  e  $X_{s2})$  apresentam médias iguais a zero e variâncias de 1,0.

## Exemplo

	$X_1$	$X_2$	$X_{d1}$	$X_{d2}$	$X_{s1}$	$X_{s2}$
1	57	93	-5,85	-30,6	-1,77	-1,96
2	58	110	-4,85	-13,6	-1,47	-0,87
3	60	99	-2,85	-24,6	-0,86	-1,58
4	59	111	-3,85	-12,6	-1,17	-0,81
5	61	115	-1,85	-8,6	-0,56	-0,55
6	60	122	-2,85	-1,6	-0,86	-0,10
7	62	110	-0,85	-13,6	-0,26	-0,87
8	61	116	-1,85	-7,6	-0,56	-0,49
9	62	122	-0,85	-1,6	-0,26	-0,10
10	63	128	0,15	4,4	0,04	0,28

	$X_1$	$X_2$	$X_{d1}$	$X_{d2}$	$X_{s1}$	$X_{s2}$
11	62	134	-0,85	10,4	-0,26	0,67
12	64	117	1,15	-6,6	0,35	-0,42
13	63	123	0,15	-0,6	0,05	-0,04
14	65	129	2,15	5,4	0,65	0,35
15	64	135	1,15	11,4	0,35	0,73
16	66	128	3,15	4,4	0,96	0,28
17	67	135	4,15	11,4	1,26	0,73
18	66	148	3,15	24,4	0,96	1,57
19	68	142	5,15	18,4	1,56	1,18
20	69	155	6,15	31,4	1,87	2,02

Algumas vezes é conveniente a representação do comprimento de um vetor particular. (Supondo que estamos no espaço euclidiano), no qual a distância entre dos pontos quaisquer é definida como:

$$\|\mathbf{a}-\mathbf{b}\| = [(a_1 - b_1)^2 + \dots + (a_m - b_m)^2]^{1/2}$$

onde  $m$  é a dimensão do espaço.

O comprimento de um vetor  $\mathbf{a}$  é a distância da origem ao ponto  $\mathbf{a}$ , ou

$$\|\mathbf{a}\| = \left[ \sum_i a_i^2 \right]^{1/2}$$

# Interpretação Geométrica das operações

- Multiplicação por um escalar:

$$r\mathbf{a}' = r(a_1, a_2) = (ra_1, ra_2)$$

o efeito é estender (Para  $r > 1$ ) ou encolher (para  $0 < r < 1$ ) o segmento de linha que representa o vetor  $\mathbf{a}$ .

- sabemos que o comprimento do vetor  $\mathbf{w}$  é  $\|\mathbf{w}\|$ . Para normalizar o vetor, realizamos uma multiplicação escalar, multiplicando o vetor  $\mathbf{w}$  pelo inverso do seu comprimento. (isto é,  $1/\|\mathbf{w}\|$ ). Se o vetor  $\mathbf{w}' = (1, 1)$ , temos que  $\|\mathbf{w}\| = (1^2 + 1^2)^{1/2} = \sqrt{2}$

$$\mathbf{w}'/\|\mathbf{w}\| = (1/\sqrt{2}, 1/\sqrt{2}) = (0,707; 0,707)$$

# Projeção

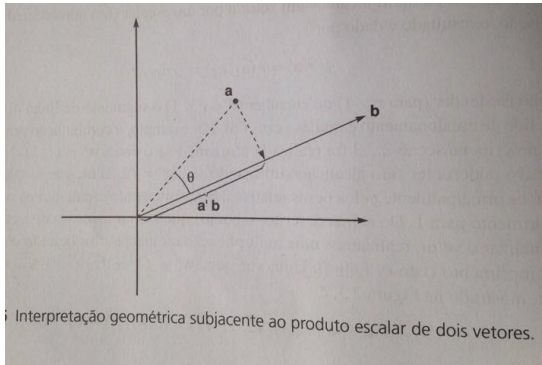
- produto escalar: Se  $a$  e  $b$  forem dois vetores coluna de ordem  $(2 \times 1)$ :

$$a' b = (a_1 \quad a_2) \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = a_1 b_1 + a_2 b_2$$

a ordem do produto é  $(1 \times 2)$  vezes  $2 \times 1 = (1 \times 1)$ .

Uma definição mais geral do produto escalar entre dois vetores  $\mathbf{a}$  e  $\mathbf{b}$  envolve o ângulo entre os dois vetores (representado por  $\theta_{ab}$ ):

$$\mathbf{a}' \mathbf{b} = \|\mathbf{a}\| \|\mathbf{b}\| \cos \theta_{ab}$$

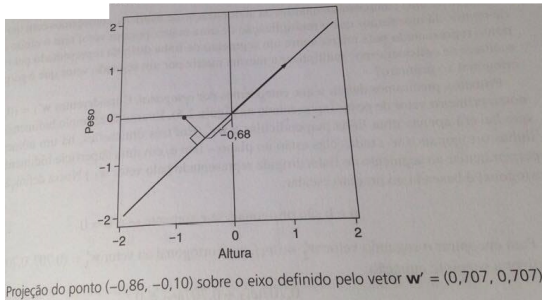


Com efeito, o que a multiplicação de vetor realiza é uma projecção: projeta uma linha perpendicular do ponto **a** para a linha dirigida **b**. A distância da origem à intersecção da linha perpendicular com a linha direta é dada por

$$\mathbf{a} \cdot \mathbf{b} = \|\mathbf{a}\| \|\mathbf{b}\| \cos \theta_{ab}$$



considere o ponto 6, isto é, uma mulher com 60 polegadas de altura e 122 libras de peso. Para esse ponto, os dados padronizados são  $(-0,86, -0,10)$ . O efeito de combinar essas duas informações usando os pesos  $\mathbf{w}' = (0,707; 0,707)$  resulta em  $-0,86 * 0,707 - 0,10 * 0,707 = -0,68$ . Na figura abaixo, podemos ver a projeção do ponto  $(-0,86; -0,10)$  sobre o eixo definido pelo vetor  $\mathbf{w}' = (0,707; 0,707)$ .



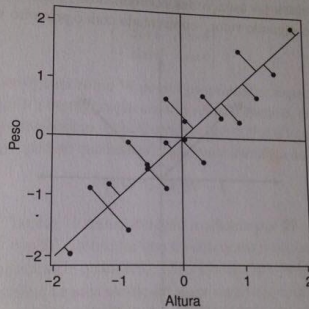
Seja  $\mathbf{X}_s$  uma matriz formada pelos vetores  $\mathbf{X}_{s1}$  e  $\mathbf{X}_{s2}$ , temos

$$\mathbf{X}_s = [\mathbf{X}_{s1}, \mathbf{X}_{s2}]$$

Podemos realizar uma multiplicação vetorial para cada observação na amostra:

$$\mathbf{z}_1 = \mathbf{X}_s \mathbf{w}$$

$\mathbf{z}_1$  é um vetor coluna (20x1), onde cada elemento é a projeção de cada ponto sobre o segmento de linha dirigida dada pelo vetor  $w$ . Veja figura abaixo.



Nova variável  $z_1 = X_s w$  (pontos projetados sobre o eixo definido por  $w$ ).

# Rotação

- Ortogonalidade: dois vetores **a** e **b** são ortogonais sse  $\mathbf{a}'\mathbf{b} = 0$ .
- Seja  $\mathbf{w}'_1 = (0,707; 0,707)$ , para encontrar um segundo vetor  $\mathbf{w}'_2 = (w_{12}, w_{22})$  ortogonal a  $\mathbf{w}'_1$  precisamos resolver

$$0,707 * w_{12} + 0,707 * w_{22} = 0$$

A solução é dada por quaisquer valores de  $w_{12}$  e  $w_{22}$  tais que

$$w_{12} = -w_{22}$$

temos para o caso que o vetor seja de comprimento unitário:

$$\mathbf{w}'_2 = (-0,707; 0,707)$$

ou

$$\mathbf{w}'_2 = (0,707, -0,707)$$

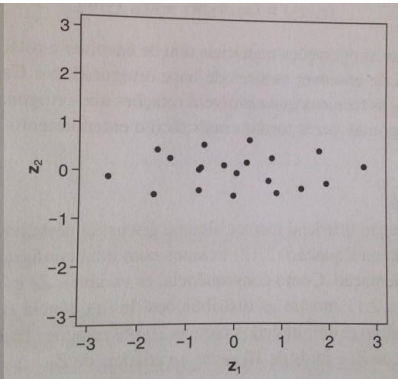
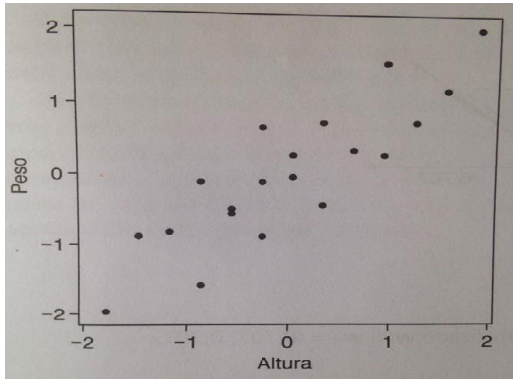
ambas as soluções apontam para direções opostas.

Se multiplicarmos  $\mathbf{X}_s$  pela matriz  $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2]$ , a operação resulta em:

$$[\mathbf{X}_{s1} \quad \mathbf{X}_{s2}][\mathbf{w}_1 \quad \mathbf{w}_2] = [\mathbf{z}_1 \quad \mathbf{z}_2]$$

Foram criadas duas combinações lineares, os vetores  $\mathbf{z}_1$  e  $\mathbf{z}_2$ , cada um dos quais é interpretável como o vetor cujas componentes são projeções dos pontos sobre um segmento de linha dirigida. Se plotarmos  $\mathbf{z}_1$  contra  $\mathbf{z}_2$  a configuração se parecerá com aquela original, só que o efeito da multiplicação foi o de rotar a configuração para uma nova orientação enquanto preserva seu tamanho e formato essenciais.

Chamamos a essa operação de rotação ortogonal ou rígida.



A matriz de rotação  $\mathbf{W}$  corresponde a uma rotação no sentido anti-horário dos eixos em  $45^\circ$

Em geral, podemos realizar uma rotação ortogonal em sentido horário através de qualquer ângulo  $\theta$  via a seguinte matriz:

$$\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

Uma matriz de rotação ortogonal como  $\mathbf{W}$  possui propriedades especiais, se  $\mathbf{W}$  gira a configuração em sentido horário, então  $\mathbf{W}'$  gira em sentido antihorário. O efeito líquido de se rotar por  $\mathbf{W}$  e de se voltar por  $\mathbf{W}'$  (ou vice-versa) garante que não haja qualquer mudança na configuração original:

$$\mathbf{W}\mathbf{W}' = \mathbf{W}'\mathbf{W} = \mathbf{I}$$

neste caso

$$\mathbf{W}' = \mathbf{W}^{-1}$$



Observe que as transformações acima ditas são verdadeiras somente para matrizes de rotação ortogonal e não para qualquer matriz quadrada.

outra maneira é imaginar uma mudança nos vetores de base do sistema de coordenadas. Para qualquer observação  $(x_1, x_2)$  podemos escrever:

$$(x_1, x_2) = x_1 \mathbf{e}'_1 + x_2 \mathbf{e}'_2$$

onde  $\mathbf{e}'_1 = (1, 0)$  e  $\mathbf{e}'_2 = (0, 1)$ .

Multiplicando  $(x_1, x_2)$  por  $\mathbf{W}$  temos a localização do ponto relativo ao novo conjunto de vetores de base  $\mathbf{w}_1$  e  $\mathbf{w}_2$

$$(z_1, z_2) = (x_1, x_2)\mathbf{w}_1 + (x_1, x_2)\mathbf{w}_2$$

Naturalmente, nem todas as operações matriciais têm de envolver a rotação ortogonal, e nem todas as mudanças de base têm de envolver vetores da base ortogonal.

## Estendendo e encolhendo

Como consequência da rotação ortogonal, ficamos com uma configuração que possui o mesmo formato, mas uma nova orientação. Como consequência, as variáveis  $Z_1$  e  $Z_2$  exibem variâncias muito diferentes. Dado que as variâncias de  $Z_1$  e  $Z_2$  são diferentes, podemos estar interessados em padroniza-las. Para isso, multiplicamos a matriz  $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2]$  por uma matriz diagonal que possui o escalar apropriado sobre a diagonal.

Seja  $\mathbf{D}^{-1}$  uma matriz diagonal com os elementos  $1/s$  na diagonal, onde  $s_i$  é o desvio padrão da amostra da variável  $Z_i$ .

$$\mathbf{D}^{-1} = \begin{bmatrix} 1/s_1 & 0 \\ 0 & 1/s_2 \end{bmatrix}$$

onde  $\mathbf{D}^{-1}$  é o inverso de  $\mathbf{D}$  e  $s_1$  e  $s_2$  são os desvios padrões das variáveis  $Z_1$  e  $Z_2$ . Desta forma,

$$\mathbf{Z}_s = \mathbf{Z}\mathbf{D}^{-1}$$

O efeito final da multiplicação de  $\mathbf{Z}$  por  $\mathbf{D}^{-1}$  é encolher  $Z_1$  (porque  $1/s_1 = 0,73$ ) e estender  $Z_2$  (porque  $1/s_2 = 2,78$ ), de modo que ambas exibem uma variância de 1.

O formato muda do elíptico para o agora circular como na figura abaixo:

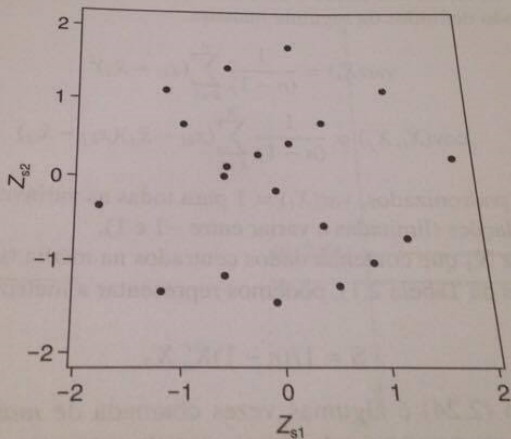


Diagrama de dispersão das variáveis padronizadas  $Z_{s1}$  e  $Z_{s2}$ .

# Decomposição em valores singulares

vimos que  $X$  é uma matriz com  $X_1$  e  $X_2$  altamente correlacionados. Podemos rotacionar via uma operação matricial estendendo-a ou encolhendo-a ao longo dos eixos. (45 graus no primeiro caso e depois uma configuração circular). As transformações realizadas podem ser escritas como:

$$\mathbf{Z}_s = \mathbf{XWD}^{-1}$$

onde  $\mathbf{W}$  é a rotação ortogonal e  $\mathbf{D}^{-1}$  a transformação diagonal que se estende ou encolhe.

Se multiplicarmos ambos os lados por  $\mathbf{D}$ , temos:

$$\mathbf{Z}_s \mathbf{D} = \mathbf{XW}$$

Podemos desfazer a rotação original imposta por  $\mathbf{W}$  pela rotação da configuração de outra maneira, isto é, multiplicando por  $\mathbf{W}'$ :

$$\mathbf{X} = \mathbf{Z}_s \mathbf{D} \mathbf{W}'$$

Isto é, qualquer matriz de dados  $\mathbf{X}$  pode ser decomposta em três partes componentes: uma matriz de variáveis  $\mathbf{Z}_s$ , que não estão correlacionadas e que têm uma variância unitária, uma transformação de extensão ou encolhimento ( $\mathbf{D}$ ) e uma rotação ortogonal  $\mathbf{W}'$ . Este processo é chamado de decomposição em valores singulares (Singular Value Decomposition)

## Medidas de Centralização

$$\bar{\mathbf{x}} = \frac{1}{n} \sum_i \mathbf{x}_i = \begin{bmatrix} \bar{x}_1 \\ \cdot \\ \cdot \\ \bar{x}_p \end{bmatrix} \quad (2)$$

que a partir da matriz de dados pode-se expressar como

$$\bar{\mathbf{x}} = \frac{1}{n} \mathbf{X}^t \mathbf{1} \quad (3)$$



# Variâncias e covariâncias

a covariância entre variáveis  $x_j$  e  $x_k$  se calcula por:

$$s_{jk} = \frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k) \quad (4)$$

para uma variável multivariante, define-se a matriz de variâncias e covariâncias como:

$$\mathbf{S} = \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})' \quad (5)$$

é uma matriz quadrada e simétrica que contém na diagonal as variâncias e fora da diagonal as covariâncias entre as variáveis.

# Variâncias e covariâncias

A matriz de variâncias e covariâncias, chamadas comumente de matriz de covariâncias é a matriz do tipo:

$$\mathbf{S} = \begin{bmatrix} s_1^2 & \dots & s_{1p} \\ \vdots & \vdots & \vdots \\ s_{p1} & \dots & s_p^2 \end{bmatrix} \quad (6)$$

A matriz  $\mathbf{S}$  pode-se obter da matriz de dados centrada  $\tilde{\mathbf{X}}$ , que é definido como  $\tilde{\mathbf{X}} = \mathbf{X} - \mathbf{1}\bar{\mathbf{x}}'$ . Se substituimos o vetor de médias pela expressão:  $\tilde{\mathbf{X}} = \mathbf{X} - \frac{1}{n}\mathbf{1}\mathbf{1}'\mathbf{X} = \mathbf{P}\mathbf{X}$ . Onde  $\mathbf{P} = \mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}'$  é simétrica e idempotente ( $\mathbf{P}\mathbf{P} = \mathbf{P}$ ), tem posto  $n - 1$ , é ortogonal ao espaço definido por  $\mathbf{1}$ , pois  $\mathbf{P}\mathbf{1} = \mathbf{0}$ .

# Variâncias e covariâncias

Desta forma, a matriz **S** pode-se escrever como :

$$\mathbf{S} = \frac{1}{n} \tilde{\mathbf{X}}' \tilde{\mathbf{X}} = \frac{1}{n} \mathbf{X}' \mathbf{P} \mathbf{X} \quad (7)$$

A matriz de variâncias corregida  $\hat{\mathbf{S}}$  é dada (pois só existirão  $n - 1$  desvios independentes):

$$\hat{\mathbf{S}} = \frac{1}{n - 1} \tilde{\mathbf{X}}' \tilde{\mathbf{X}}$$

## EXERCÍCIOS

- 2.1 Para cada par dos seguintes vetores, encontre o comprimento de **a**, o comprimento de **b**, o ângulo  $\theta$  formado pelos vetores **a** e **b** e seu produto escalar:
- a.  $\mathbf{a} = (1, 1)$  e  $\mathbf{b} = (-1, 1)$       c.  $\mathbf{a} = (4, 3)$  e  $\mathbf{b} = (-4, -3)$   
b.  $\mathbf{a} = (1, 0)$  e  $\mathbf{b} = (1, 1)$       d.  $\mathbf{a} = (1, 2, 3)$  e  $\mathbf{b} = (1, 1, 2)$
- 2.2 Um pesquisador do campo de pesquisa educacional coletou algumas pontuações de testes de um grupo de oito estudantes. Esses dados são mostrados a seguir e estão disponíveis no arquivo *EDUC\_SCORES*.

Estudante	Linguagem Atitude ( $x_1$ )	Raciocínio Analogico ( $x_2$ )	Raciocínio Geométrico ( $x_3$ )	Sexo do Estudante (Masculino = 1) ( $x_4$ )
A	2	3	15	1
B	6	8	9	1
C	5	2	7	0
D	9	4	3	1
E	11	10	2	0
F	12	15	1	0
G	1	4	12	1
H	7	3	4	0

Usando os dados da linha da matriz  $\mathbf{X} = [\mathbf{x}_1 \ \mathbf{x}_2 \ \mathbf{x}_3]$ , calcule as seguintes matrizes:

- o vetor centroide  $\bar{\mathbf{x}}' = (\bar{x}_1, \bar{x}_2, \bar{x}_3)$
- a matriz de média diferenciada  $\mathbf{X}_d$
- a matriz  $\mathbf{X}_d$  de dados padronizados
- a soma da matriz quadrada  $\mathbf{X}'_d \mathbf{X}_d$
- a matriz de covariância  $\mathbf{S} = 1/(n-1) \mathbf{X}'_d \mathbf{X}_d$
- a matriz de correlação  $\mathbf{R} = 1/(n-1) \mathbf{X}' \mathbf{X}$

- 2.3 O pesquisador está interessado em encontrar uma combinação linear para compilar o desempenho no teste de cada um dos oito estudantes. Ele propõe dar um peso de 25% à pontuação em aptidão para a linguagem ( $\mathbf{x}_1$ ), um peso de 25% para a pontuação em raciocínio analógico ( $\mathbf{x}_2$ ) e um peso de 50% à pontuação em raciocínio geométrico ( $\mathbf{x}_3$ ).
- Encontre o vetor  $\mathbf{w}' = (w_1, w_2, w_3)$  de comprimento unitário que realize esse esquema relativo de ponderação.
  - Encontre a combinação linear usando o dado bruto ( $\mathbf{z}_1 = \mathbf{X} \mathbf{w}$ ) e o dado padronizado ( $\mathbf{z}_2 = \mathbf{X}_d \mathbf{w}$ ).
  - Compare as pontuações dos estudantes C e D. Qual estudante é melhor em  $\mathbf{z}_1$ ? Qual estudante é melhor em  $\mathbf{z}_2$ ?
- 2.4 Divida os dados no arquivo *EDUC\_SCORES* (descrito anteriormente) em dois grupos: homens e mulheres. Para cada grupo, forme a matriz  $\mathbf{X} = [\mathbf{x}_1 \ \mathbf{x}_2 \ \mathbf{x}_3]$  e compute o seguinte:
- o vetor centroide  $\bar{\mathbf{x}}' = (\bar{x}_1, \bar{x}_2, \bar{x}_3)$
  - a matriz de covariância  $\mathbf{S} = 1/(n-1) \mathbf{X}'_d \mathbf{X}_d$
  - Parece haver diferenças nos dois grupos tanto no nível ou na dispersão de suas pontuações no teste?
- 2.5 O arquivo *RANDOM1* contém três variáveis ( $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ ) que foram criadas utilizando-se um gerador de números randômicos. Cada uma consiste em  $n = 100$  observações retiradas independentemente de uma unidade de distribuição normal, e cada uma foi padronizada. Usando os dados em *RANDOM1*, forme as seguintes combinações lineares:

$$\mathbf{x}_4 = 0,80\mathbf{x}_1 + 0,60\mathbf{x}_2$$

$$\mathbf{x}_5 = 0,80\mathbf{x}_1 + 0,60\mathbf{x}_3$$

Forme a matriz  $\mathbf{X} = [\mathbf{x}_4 \ \mathbf{x}_5]$  e realize a seguinte multiplicação matricial:  $\mathbf{Z} = \mathbf{X} \mathbf{W}$ , onde

$$\mathbf{W} = \begin{bmatrix} 0,866 & -0,500 \\ 0,500 & 0,866 \end{bmatrix}$$

- Crie um gráfico de dispersão das novas variáveis  $\mathbf{z}_1$  e  $\mathbf{z}_2$ .
  - Calcule a matriz de covariância  $1/(n-1) \mathbf{Z}' \mathbf{Z}$  (observe que  $\mathbf{z}_1$  e  $\mathbf{z}_2$  já são centradas na média).
  - Qual é o determinante da matriz de covariância de  $\mathbf{Z}$ ? Como ele se compara ao determinante da matriz de covariância de  $\mathbf{X} = [\mathbf{x}_4 \ \mathbf{x}_5]$ ?
- 2.6 Usando os dados do arquivo *RANDOM1* do problema anterior, forme a matriz  $\mathbf{X} = [\mathbf{x}_1 \ \mathbf{x}_2]$  e realize a multiplicação matricial  $\mathbf{Z} = \mathbf{X} \mathbf{W}$ , onde

$$\mathbf{W} = \begin{bmatrix} 0,866 & 0,500 \\ 0,500 & 0,866 \end{bmatrix}$$

- Crie um gráfico de dispersão das novas variáveis  $\mathbf{z}_1$  e  $\mathbf{z}_2$ .
- Calcule a matriz de covariância  $1/(n-1) \mathbf{Z}' \mathbf{Z}$  (note que  $\mathbf{z}_1$  e  $\mathbf{z}_2$  já são centradas na média).
- Qual é o determinante da matriz de covariância de  $\mathbf{Z}$ ? Como ele se compara com o determinante da matriz de covariância de  $\mathbf{X} = [\mathbf{x}_1 \ \mathbf{x}_2]$ ?

# Exercícios

- Exercícios 2.1 à 2.6
- Páginas 28-29
- Livro Análise de Dados Multivariados

Grupos	Núm.
1-6	Ex. 2.1,2.3 e 2.5
7-12	Ex. 2.2,2.4 e 2.6