

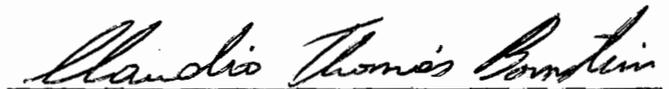
O MODELO DAS COMPONENTES PRINCIPAIS:

apresentação e interpretações

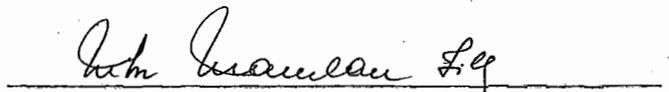
Luiz Cláudio Teixeira Herig

TESE SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS PROGRAMAS DE PÓS-GRADUAÇÃO DE ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS (M.Sc.).

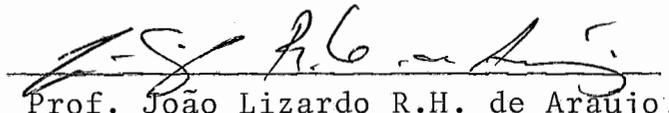
Aprovada por:



Prof. Claudio Thomás Bornstein
(Presidente)



Prof. Nelson Maculan Filho



Prof. João Lizardo R.H. de Araújo



Prof. Anibal Santana

RIO DE JANEIRO, RJ - BRASIL
DEZEMBRO DE 1979

HERIG, LUIZ CLÁUDIO TEIXEIRA

O Modelo Das Componentes Principais: Apresentação e Interpretações (Rio de Janeiro) 1979.

V, 222p. 29,7 cm (COPPE-UFRJ, M.Sc., Engenharia de Sistemas e Computação, 1979)

Tese - Univ. Fed. Rio de Janeiro. Fac. Engenharia

1. Análise Multivariável I. COPPE/UFRJ II. Título(série).

RESUMO

O objetivo desse trabalho é apresentar, de modo didático, um modelo de análise de dados no campo da Estatística Multivariada - o Modelo das Componentes Principais.

Dentro desse enfoque buscou-se transmitir, ao longo do texto, a filosofia geral do modelo, sem pretender com isso exaurir o assunto.

O trabalho encontra-se dividido em tres capítulos. O primeiro posiciona o modelo, dentro do âmbito da Estatística e o segundo se propõe a fornecer o embasamento necessário ao bom entendimento do último capítulo que faz a apresentação detalhada do modelo e descreve algumas de suas possíveis interpretações.

ABSTRACT

This work intends to present, in a didactical way, a data-analyzing model in the Multivariate Statistical field - the Principal Components Model.

In this way, we tried to transmit, in this work, the general idea of the model, without expecting to exhaust the matter.

The work is divided in three chapters. The first one locates the model inside the statistical field and the second tries to furnish the necessary basis for the understanding of the last chapter, which presents the model in a detailed way, and describes some of the possible interpretations.

ÍNDICEPáginasCAPÍTULO I - INTRODUÇÃO

1.1. - Análise Multivariada	1
1.2. - Análise Fatorial	2
1.3. - Objetivo do Texto	4
1.4. - Notação Empregada	4

CAPÍTULO II - NOÇÕES MATEMÁTICAS

2.1. - Introdução	11
2.2. - Somatórios	11
2.3. - Noções de Estatística	17
2.4. - Noções de Geometria no \mathbb{R}^n	47
2.5. - Noções de Álgebra Linear	60
2.6. - Resumo	81

CAPÍTULO III - COMPONENTES PRINCIPAIS

3.1. - Introdução	90
3.2. - O Modelos das Componentes Principais (Problema 1) . . .	91
3.3. - O Problema a Ser Resolvido	98
3.4. - Resolução do Problema 1.	103
3.5. - Revisão	123
3.6. - Um Exemplo Resolvido	125
3.7. - Transformação Ortogonal (Problema 2)	136
3.8. - Comentários Sobre o Exemplo	146
3.9. - A Conservação da Variância	149
3.10.- Comentários Sobre o Exemplo	152
3.11.- Enfoque Geométrico do Modelo (Problema 3) . . .	154
3.12.- Resolução do Problema 3	158

	<u>Páginas</u>
3.13. - Interpretação do Problema 3	183
3.14. - Comparação dos Problemas 1, 2 e 3	184
3.15. - Comentários Sobre o Exemplo	189
3.16. - Analogia Com a Mecânica	190
3.17. - Casos Particulares	194
3.18. - Escalas	217
 ANEXO	 220
 BIBLIOGRAFIA	 221

CAPÍTULO IINTRODUÇÃO1.1. Análise Multivariada

Um problema que frequentemente surge, em estatística, é que temos um conjunto (ou população) de N elementos nos quais observamos n características e, baseados nessas observações estamos interessados em tirar conclusões sobre a população, sobre os elementos dessa população, ou ainda sobre as características que estão sendo observadas.

Se for possível medir essas características poderemos associar a cada uma delas uma VARIÁVEL. Como esse problema geralmente envolve várias variáveis, vamos chama-lo de PROBLEMA MULTIVARIADO. Podemos então, sem muito rigor matemático, considerar a ANÁLISE ESTATÍSTICA MULTIVARIADA (ou, ANÁLISE MULTIVARIADA), como o ramo da ESTATÍSTICA que estuda os problemas multivariados.

Um exemplo típico de problema multivariado é o seguinte:

EXEMPLO 1.1.1:

Numa turma de 20 alunos observamos as notas que cada um deles tirou em matemática, física, geografia e história e, estamos interessados em ordenar esses alunos segundo um "coeficiente de aproveitamento" que irá ser definido. ▲

Nesse caso os elementos são os alunos ($N = 20$) e as variáveis, em número de quatro ($n = 4$), representarão cada uma das provas. Assim cada variável assumirá um valor para ca

da aluno. Esse valor será a nota que o aluno tirou nessa prova.

Nesse exemplo, podíamos ainda agrupar os alunos segundo áreas de interesse, ou testar a hipótese de que um aluno bom em matemática e física também é bom em história ou ainda verificar qual a prova mais difícil, etc... . Como podemos notar, a abordagem de um problema multivariado pode ser feita de várias formas, segundo diversos objetivos.

Para um melhor entendimento, podemos lembrar aqui a definição dada por KENDALL^[19], pág.1:

"A análise multivariada é um ramo da estatística que estuda o interrelacionamento entre conjuntos de variáveis dependentes e os elementos nos quais medimos essas variáveis".

Para finalizar, é bom lembrar que os problemas multivariados surgem nas mais diversas áreas do conhecimento humano. Poderíamos, por exemplo, desejar ordenar um conjunto de países baseados em indicadores econômicos (nesse caso os elementos seriam os países e as variáveis os indicadores). Poderíamos também desejar descobrir quais as variáveis que influenciam consideravelmente no rendimento de um atleta numa competição, ou ainda, se devemos ou não abrir as comportas de uma usina hidroelétrica baseado em fluxos hidrológicos locais.

Como vemos, a área abrangida é grande e não fica restrita aos exemplos aqui citados. Outros exemplos poderão ser encontrados em KENDALL^[19], págs.2-3.

1.2. Análise Fatorial

Os problemas multivariados frequentemente envolvem

um número muito grande de variáveis; o que dificulta a sua análise. Devido à essa dificuldade vários modelos matemáticos foram desenvolvidos com o objetivo de possibilitar uma simplificação na abordagem de tais problemas.

Entre os modelos que tem esse objetivo podemos destacar aqueles que procuram, uma vez de posse das n variáveis observadas, encontrar um novo conjunto de variáveis, tal que as "novas variáveis" (que chamaremos de COMPONENTES ou FATORES) possam "descrever", de uma forma conveniente, as variáveis inicialmente observadas (que chamaremos simplesmente de VARIÁVEIS), ou seja, que as variáveis sejam uma função dessas componentes (ou fatores).

A idéia fundamental desses modelos pode ser resumida na seguinte equação:

$$X_i = f_i(Y_1, Y_2, \dots, Y_p) \quad i = 1, 2, \dots, n \quad (1.2.1)$$

$$p \leq n$$

onde

X_i - é a variável, ou seja, a i -ésima característica, que observamos na população.

Y_j - ($j = 1, 2, \dots, p$) é a componente (ou fator).

f_i - é a função que relaciona as componentes Y_j à variável X_i

OBS.: Podemos notar que tanto X_i quanto Y_j são variáveis matemáticas. O fato de chamarmos X_i de variável e Y_j de componente (ou fator) deve-se somente à didática do texto.

Embora não havendo um consenso geral em torno de

uma definição, podemos dizer que a ANÁLISE FATORIAL estuda problemas multivariados, utilizando modelos que sejam descritos pela equação 1.2.1.

1.3. Objetivo do Texto

Como podemos concluir, o campo da Análise Fatorial é muito amplo não só do ponto de vista teórico como também em suas aplicações práticas. Seu estudo completo foge do objetivo desse texto. Aqui, a nossa meta será desenvolver um modelo que hoje é frequentemente utilizado naquele campo, o MODELO DAS COMPONENTES PRINCIPAIS (MCP).

Para a compreensão desse modelo é necessário algum conhecimento de Álgebra Linear, Estatística e Geometria no R^n , que pressupomos que o leitor possua. No Capítulo II, teremos um breve enfoque dessa matéria com o objetivo de permitir consultas que esclareçam o modelo em estudo. No Capítulo III, será então desenvolvido o modelo.

1.4. Notação Empregada

Esta seção contém um resumo da simbologia adotada no texto, e destina-se somente a consultas sobre a notação aqui empregada. A compreensão completa dessa simbologia virá com o decorrer do texto.

A notação aqui utilizada é uma reunião das que normalmente são usadas na Estatística, Álgebra Linear, Geometria do R^n e da notação de KENDALL [19], que pareceu-nos eficiente.

Vamos então dividir os símbolos em seis grupos:

- 1) Símbolos referentes às definições
- 2) Parâmetros
- 3) Índices
- 4) As variáveis X, Y e Z
- 5) Vetores e Matrizes Importantes
- 6) Outros.

1) Símbolos referentes às definições:

$$\left. \begin{array}{l} E[X_i] \\ \bar{X}_i \end{array} \right\} \text{ média ou valor esperado da variável } X_i$$

$$\left. \begin{array}{l} \sigma_{X_i}^2 \\ \text{VAR}[X_i] \\ c_{ii} \end{array} \right\} \text{ variância da variável } X_i$$

σ_{X_i} - desvio médio padrão da variável X_i

$\text{COR}[X_i; X_{i'}]$ - correlação entre as variáveis X_i e $X_{i'}$

$P(X_i = a)$ - probabilidade da variável X_i assumir o valor a

$P(X_i = a | X_{i'} = b)$ - probabilidade da variável X_i assumir o valor a dado que $X_{i'}$ assumiu o valor b

$\text{dist}(X(k); X(k'))$ - distância entre os pontos $X(k)$ e $X(k')$

$|\vec{X}(k)|$ - norma do vetor $\vec{X}(k)$

$\langle \vec{X}_i, \vec{X}_i \rangle$ - produto escalar entre os vetores \vec{X}_i e \vec{X}_i ,

A^T - matriz (ou vetor) transposta(o) da matriz (ou vetor) A

2) Parâmetros

N - número de elementos da população em estudo

n - número de características observadas em cada elemento da população

p - número de componentes ou fatores.

3) Índices

Como o problema é multivariado os índices assumem um papel importante. São eles:

i - se refere as variáveis que serão observadas na população. Exemplo: X_i . Da mesma forma utilizaremos, às vezes i' e i'' .

j - se refere às componentes. Exemplo: Y_j . Da mesma forma utilizaremos, às vezes j' e j'' .

k - se refere aos elementos da população. Exemplo: $X(k)$. Da mesma forma utilizaremos k' e k'' .

Notaremos que geralmente esses índices assumem os seguintes valores:

$$i = 1, 2, \dots, n$$

$$j = 1, 2, \dots, p$$

$$k = 1, 2, \dots, N$$

4) As variáveis X, Y e z

O Modelo das Componentes Principais (MCP) parte da hipótese de que medimos características em elementos de uma população. Essas características serão representadas pelas variáveis X_i e, dessa medição, decorrem várias notações. Se não vejamos:

X_i - [variável] característica i da população.

X - [$X = (X_1, X_2, \dots, X_n)$] conjunto de características da população.

$X_i(k)$ - [realização da variável X_i] valor da característica i apresentada pelo elemento k da população.

$X(k)$ - [$X(k) = (X_1(k), X_2(k), \dots, X_n(k))$] conjunto de valores das características apresentadas pelo elemento k da população. Ou seja $X(k)$ é uma n -upla ordenada dos valores $X_i(k)$ ($i = 1, 2, \dots, n$) e por isso poderá ser considerado como um ponto pertencente ao \mathbb{R}^n .

Se considerarmos essas características medidas em relação às suas médias, ou seja, usando desvios (ver definição de desvio na seção 2.3) ao invés do próprio valor da medição, usaremos as variáveis acima como letras minúsculas

$$x_i, x, x_i(k), x(k)$$

onde

$$x_i = X_i - \bar{X}_i$$

$$x = (x_1, x_2, \dots, x_n)$$

$$x_i(k) = X_i(k) - \bar{X}_i$$

$$x(k) = (x_1(k), x_2(k), \dots, x_n(k))$$

Se além disso utilizarmos essas características na forma padronizada teremos:

$$z_i, z, z_i(k), z(k)$$

onde

$$z_i = \frac{X_i - \bar{X}_i}{\sigma_{X_i}}$$

$$z = (z_1, z_2, \dots, z_n)$$

$$z_i(k) = \frac{X_i(k) - \bar{X}_i}{\sigma_{X_i}}$$

$$z(k) = (z_1(k), z_2(k), \dots, z_n(k))$$

No MCP as variáveis Y são utilizadas para denotar grandezas relacionadas com as componentes principais.

Y_j - [variável] é a j -ésima componente

Y - [$Y = (Y_1, Y_2, \dots, Y_n)$] conjunto de componentes

$Y_j(k)$ - [realização da variável Y_j] valor da componente j para o elemento k da população.

$Y(k)$ - [$Y(k) = (Y_1(k), Y_2(k), \dots, Y_n(k))$] valores assumidos pelas componentes para o elemento k da população. $Y(k)$ é uma n -upla ordenada dos valores $Y_j(k)$ ($j=1, 2, \dots, n$) e por isso poderá ser considerado como um ponto do R^n .

Da mesma forma que fizemos para as variáveis X ,

utilizaremos também as variáveis

$$y_j, y, y_j(k), y(k)$$

no caso de estarmos utilizando medidas em relação à média.

5) Vetores e Matrizes Importantes

C - Matriz de Covariâncias entre as variáveis X_i e $X_{i'}$, ($i, i' = 1, 2, \dots, n$) ou entre as variáveis x_i e $x_{i'}$, o que dá no mesmo (ver seção 2.3).

Como podemos observar é uma matriz $n \times n$ de elementos $c_{ii'}$.

Como veremos mais adiante c_{ii} é a variância da variável X_i , logo a diagonal principal dessa matriz conterá as variâncias das variáveis X_i

L - matriz de elementos l_{ij} que serão os coeficientes das componentes na composição das variáveis.

\hat{L} - matriz inversa da matriz L. Usualmente ela é representada por L^{-1} , mas nesse texto utilizaremos a outra notação. Ela possui elementos l_{ji} que serão os coeficientes das variáveis na composição das componentes.

l_i - vetor de componentes l_{ij} . Assim, podemos considerar também que L é um vetor de componentes l_i .

Senão vejamos:

$$L = \begin{pmatrix} l_{11} & l_{12} & \dots & l_{1n} \\ l_{21} & & & \vdots \\ \vdots & & & \vdots \\ l_{n1} & \dots & \dots & l_{nn} \end{pmatrix} = \begin{pmatrix} l_1 \\ l_2 \\ \vdots \\ l_n \end{pmatrix}$$

onde

$$l_i = (l_{i1}, l_{i2}, \dots, l_{in}) \quad i=1, 2, \dots, n$$

l_j - vetor de componentes l_{ji} . Da mesma forma podemos considerar que \hat{L} seja um vetor de componentes l_j .

D - matriz diagonal cujos componentes da diagonal principal são os $\lambda_j (j=1, 2, \dots, n)$.

6) Outros

g_{ji} - cosseno diretor do vetor (ou reta) j em relação ao vetor i da base do espaço vetorial.

g_j - [$g_j = (g_{j1}, g_{j2}, \dots, g_{jn})$] vetor de componentes g_{ji} , ou seja, conjunto de cossenos diretores de um vetor ou (reta).

λ_j - j -ésimo autovalor da matriz C e elemento da diagonal principal da matriz D. Representa a variância da j -ésima componente.

CAPÍTULO II
NOÇÕES MATEMÁTICAS

2.1. Introdução

O objetivo desse capítulo será reunir uma série de conceitos que serão necessários à compreensão do MODELO DAS COMPONENTES PRINCIPAIS (MCP). Sendo assim, na seção 2.2 daremos algumas propriedades dos somatórios. Na seção 2.3 enfocaremos a parte da Estatística que diz respeito à média, variância, covariância, correlação e padronização de variáveis aleatórias. Nesta seção, veremos as definições desses conceitos e algumas de suas propriedades, porém as demonstrações não serão feitas. Já na seção 2.4 o tema abordado será referente à geometria no \mathbb{R}^n , principalmente os conceitos de distância e norma. Outro assunto, a ser rapidamente abordado, será a Álgebra Linear. Isso acontecerá na seção 2.5 onde faremos um estudo sobre as Transformações Lineares. Finalmente, na seção 2.6, para simplificar consultas, teremos um resumo dos pontos importantes desse capítulo.

2.2. Somatórios

Várias são as operações matemáticas que envolvem somatórios. Gostaríamos aqui de lembrar algumas de suas características. Para isso vamos partir do seguinte sistema de equações:

$$\begin{aligned}
X_1 &= \ell_{11}Y_1 + \ell_{12}Y_2 + \dots + \ell_{1p}Y_p \\
X_2 &= \ell_{21}Y_1 + \ell_{22}Y_2 + \dots + \ell_{2p}Y_p \\
&\vdots \\
X_n &= \ell_{n1}Y_1 + \ell_{n2}Y_2 + \dots + \ell_{np}Y_p
\end{aligned}
\tag{2.2.1}$$

que pode ser representado por:

$$X_i = \sum_{j=1}^p \ell_{ij}Y_j \quad \forall i = 1, 2, \dots, n \tag{2.2.2}$$

onde X_i e Y_j serão variáveis e os ℓ_{ij} são os coeficientes.

Se desejarmos calcular o somatório das variáveis

X_i teremos:

$$\sum_{i=1}^n X_i = \sum_{i=1}^n \left[\sum_{j=1}^p \ell_{ij} Y_j \right] \tag{2.2.3}$$

Como a ordem das parcelas não altera a soma podemos então concluir que:

$$\boxed{\sum_{i=1}^n \sum_{j=1}^p \ell_{ij} Y_j = \sum_{j=1}^p \sum_{i=1}^n \ell_{ij} Y_j} \tag{2.2.4}$$

Poderíamos também estar interessados em calcular $(X_i X_{i'})$. Nesse caso teríamos

$$\begin{aligned}
X_i X_{i'} &= \left(\sum_{j=1}^p \ell_{ij} Y_j \right) \left(\sum_{j'=1}^p \ell_{i'j'} Y_{j'} \right) \\
&\forall i = 1, 2, \dots, n \\
&\forall i' = 1, 2, \dots, n
\end{aligned}
\tag{2.2.5}$$

(observe que o índice j' foi usado unicamente para distinguir-se do índice j).

$$\boxed{X_i^2 = \left(\sum_{j=1}^p \ell_{ij} Y_j \right)^2 = \sum_{j=1}^p \sum_{j'=1}^p \ell_{ij} \ell_{ij'} Y_j Y_{j'}} \quad \forall i=1,2,\dots,n \quad (2.2.10)$$

Suponhamos agora, que o objetivo seja calcular a derivada parcial de X_i em relação a um dado ℓ_{ij} . Partindo de (2.2.2) temos:

$$\frac{\delta X_i}{\delta \ell_{ij}} = \frac{\delta \sum_{j'=1}^p \ell_{ij'} Y_{j'}}{\delta \ell_{ij}} \quad \begin{array}{l} \forall i = 1, 2, \dots, n \\ \forall j = 1, 2, \dots, p \end{array} \quad (2.2.11)$$

Como, caso exista, a derivada de uma soma é igual à soma das derivadas das parcelas, temos:

$$\begin{aligned} \frac{\delta X_i}{\delta \ell_{ij}} &= \frac{\delta \ell_{i1} Y_1}{\delta \ell_{ij}} + \frac{\delta \ell_{i2} Y_2}{\delta \ell_{ij}} + \dots + \\ &+ \dots + \frac{\delta \ell_{ip} Y_j}{\delta \ell_{ij}} \quad \begin{array}{l} \forall i=1,2,\dots,p \\ \forall j=1,2,\dots,n \end{array} \end{aligned} \quad (2.2.12)$$

Observamos que todas as parcelas do segundo membro são nulas a menos daquela cujo numerador for igual a $\ell_{ij} Y_j$, logo

$$\boxed{\frac{\delta X_i}{\delta \ell_{ij}} = Y_j} \quad \begin{array}{l} \forall i=1,2,\dots,n \\ \forall j=1,2,\dots,p \end{array} \quad (2.2.13)$$

Poderíamos também estar interessados em calcular

$\frac{\delta X_i^2}{\delta l_{ij}}$. Nesse caso, utilizando a equação 2.2.10 teremos:

$$\frac{\delta X_i^2}{\delta l_{ij}} = \frac{\delta \left(\sum_{j'=1}^p \sum_{j''=1}^p l_{ij'} l_{ij''} Y_{j'} Y_{j''} \right)}{\delta l_{ij}} \quad \begin{matrix} \forall i=1,2,\dots,n \\ \forall j=1,2,\dots,p \end{matrix} \quad (2.2.14)$$

Se desenvolvermos o somatório como foi feito na equação 2.2.12 teremos p^2 parcelas pois $j'=1,2,\dots,p$ e $j''=1,2,\dots,p$. Chegaremos então à conclusão que essas parcelas podem ser agrupadas em quatro tipos, segundo os seus índices.

- a) $j' = j \neq j''$
- b) $j' \neq j = j''$
- c) $j' = j = j''$
- d) $j' \neq j \neq j''$

Assim, a equação 2.2.14 ficará:

$$\begin{aligned} \frac{\delta X_i^2}{\delta l_{ij}} &= \underbrace{\frac{\delta \left(l_{ij} \sum_{j''=1}^p l_{ij''} Y_{j''} Y_{j''} \right)}{\delta l_{ij}}}_{a} + \underbrace{\frac{\delta \left(l_{ij} \sum_{j'=1}^p l_{ij'} Y_{j'} Y_{j'} \right)}{\delta l_{ij}}}_{b} \\ &+ \underbrace{\frac{\delta \left(l_{ij}^2 Y_j^2 \right)}{\delta l_{ij}}}_{c} + \underbrace{\frac{\delta \left(\sum_{j'=1}^p \sum_{j''=1}^p l_{ij'} l_{ij''} Y_{j'} Y_{j''} \right)}{\delta l_{ij}}}_{d} \end{aligned}$$

$$\forall i=1,2,\dots,n$$

$$\forall j=1,2,\dots,p$$

(2.2.15)

Podemos então simplificar a equação 2.2.15 observando que a parcela a é igual à parcela b e que a parcela d é igual à zero. Assim, teremos:

$$\frac{\delta(X_i^2)}{\delta l_{ij}} = 2 \frac{\delta(\ell_{ij} \sum_{\substack{j'=1 \\ j' \neq j}}^p \ell_{ij', Y_j Y_{j'}})}{\delta l_{ij}} + \frac{\delta(\ell_{ij}^2 Y_j^2)}{\delta l_{ij}}$$

$$\forall i=1,2,\dots,n$$

$$\forall j=1,2,\dots,p$$

(2.2.16)

Calculando-se então as derivadas teremos:

$$\frac{\delta(X_i^2)}{\delta l_{ij}} = 2 \sum_{\substack{j'=1 \\ j' \neq j}}^p \ell_{ij', Y_j Y_{j'}} + 2 \ell_{ij} Y_j^2$$

$$\forall i=1,2,\dots,n$$

$$\forall j=1,2,\dots,p$$

(2.2.17)

E chegamos finalmente a

$$\frac{\delta(X_i^2)}{\delta l_{ij}} = 2 \sum_{j'=1}^p \ell_{ij', Y_j Y_{j'}}$$

$$\forall i=1,2,\dots,n$$

$$\forall j=1,2,\dots,p$$

(2.2.18)

Levando a equação 2.2.10 na equação 2.2.18 teremos:

$$\frac{\delta \left[\left(\sum_{j'=1}^p \ell_{ij'} Y_{j'} \right)^2 \right]}{\delta \ell_{ij}} = 2 \sum_{j'=1}^p \ell_{i \cdot j'} Y_j Y_{j'}$$

$\forall i=1, 2, \dots, n$
 $\forall j=1, 2, \dots, p$

(2.2.19)

2.3. Noções de Estatística

Como o nosso objetivo é analisar uma população de N elementos da qual observamos n características, direcionaremos essa seção para esse caso particular. Assim, X_i será para nós uma variável aleatória discreta que representará uma característica da população. Logo, X_i poderá assumir quaisquer valores $X_i(k)$. Uma generalização para o caso contínuo não é difícil e fica a cargo do leitor.

Consideraremos também que se pode medir as características em todos os elementos da população. Caso isso não seja possível, técnicas de estimativa poderão ser empregadas e os valores estimados numa amostra passarão a representar o parâmetro da população. Nesse caso, alguns cuidados devem ser tomados. Podemos ver isso detalhadamente em [2], [16], [21] e [25].

MÉDIA

Definição 2.3.1: A média, ou valor esperado, de uma variável aleatória discreta X_i , medida em uma população de N elementos

é dada por:

$$E[X_i] = \bar{X}_i = \frac{1}{N} \sum_{k=1}^N X_i(k) \quad (2.3.1)$$

onde k representa um elemento dessa população.

OBS.:-serão usadas indistintamente as notações $E[X_i]$ e \bar{X}_i .

Outra notação bastante comum é μ_{X_i} .

-podemos denominar também $E[X_i]$ de EXPECTÂNCIA ou ESPERANÇA MATEMÁTICA de X_i .

Como propriedades da média temos: (sejam a e b constantes e $i, i' = 1, 2, \dots, n$)

$$i) \text{ Se } X_i \text{ é constante então } E[X_i] = X_i = \text{cte} \quad (2.3.2)$$

$$ii) E[aX_i] = aE[X_i] \quad (2.3.3)$$

$$iii) E[X_i + X_{i'}] = E[X_i] + E[X_{i'}] \quad (2.3.4)$$

Podemos generalizar essa propriedade

$$E[X_1 + X_2 + \dots + X_n] = E[X_1] + E[X_2] + \dots + E[X_n] \quad (2.3.5)$$

o que pode ser escrito como:

$$E\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n E[X_i] \quad (2.3.6)$$

$$iv) \text{ Se } H(X_i) = aX_i + b \text{ então } E[H(X_i)] = a E[X_i] + b \quad (2.3.7)$$

Podemos generalizar essa propriedade. Seja então

$$H(X_1; X_2, \dots, X_n) = a_1 X_1 + a_2 X_2 + \dots + a_n X_n + b \quad (2.3.8)$$

(onde a_1, a_2, \dots, a_n e b são constantes) então

$$E[H(X_1, X_2, \dots, X_n)] = H(E[X_1]; E[X_2]; \dots, E[X_n]) \quad (2.3.9)$$

VARIÂNCIA

Definição 2.3.2: A variância de uma variável aleatória X_j é dada por:

$$\sigma_{X_i}^2 = E[(X_i - \bar{X}_i)^2] \quad (2.3.10)$$

A notação $\sigma_{X_i}^2$ é muito utilizada em estatística para representar a variância de uma variável X_i . Outra forma também usada é $\text{VAR}[X_i]$. Para facilitar o desenvolvimento do Capítulo III, usaremos indistintamente, além dessas duas, a notação c_{ii} utilizada por KENDALL [19].

Agora a equação 2.3.10 pode ser escrita da seguinte forma:

$$\sigma_{X_i}^2 = \text{VAR}[X_i] = c_{ii} = E[(X_i - \bar{X}_i)^2] \quad (2.3.11)$$

Essa mesma definição pode ser dada na forma de somatório. Substituindo a equação 2.3.1 na 2.3.11 teremos:

$$\sigma_{X_i}^2 = \text{VAR}[X_i] = c_{ii} = \frac{1}{N} \sum_{k=1}^N (X_i(k) - \bar{X}_i)^2 \quad (2.3.12)$$

Podemos notar que a variância é uma grandeza que mostra quanto uma variável aleatória se afasta de sua média. É portanto uma medida de dispersão.

Como propriedades da variância temos: (sejam a e b constantes e $i, i' = 1, 2, \dots, n$)

$$\text{i) Se } X_i \text{ é constante então } \text{VAR}[X_i] = 0 \quad (2.3.13)$$

$$\text{ii) } \text{VAR}[X_i + a] = \text{VAR}[X_i] \quad (2.3.14)$$

Podemos concluir que a origem usada na medição de X_i não influencia a sua variância.

$$\text{iii) } \text{VAR}[aX_i] = a^2 \text{VAR}[X_i] \quad (2.3.15)$$

Resumindo as propriedades ii e iii concluimos que:

$$\text{se } H(X_i) = aX_i + b \text{ então } \text{VAR}[H(X_i)] = a^2 \text{VAR}[X_i] \quad (2.3.16)$$

$$\text{iv) } \text{VAR}[X_i + X_{i'}] = \text{VAR}[X_i] + \text{VAR}[X_{i'}] + 2E[(X_i - \bar{X}_i)(X_{i'} - \bar{X}_{i'})] \quad (2.3.17)$$

Como será visto mais adiante, se as variáveis X_i e $X_{i'}$ ($i \neq i'$) não forem correlacionadas a expressão $E[(X_i - \bar{X}_i)(X_{i'} - \bar{X}_{i'})]$ se anulará e sob essa hipótese teremos:

$$\text{VAR}[X_i + X_{i'}] = \text{VAR}[X_i] + \text{VAR}[X_{i'}] \quad (2.3.18)$$

Ainda sob a hipótese de não correlação entre as variáveis X_i e $X_{i'}$ ($\forall i, i' = 1, 2, \dots, n; i \neq i'$) podemos generalizar

$$\text{VAR}[X_1 + X_2 + \dots + X_n] = \text{VAR}[X_1] + \text{VAR}[X_2] + \dots + \text{VAR}[X_n] \quad (2.3.19)$$

ou, sob a forma de somatório,

$$\text{VAR}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \text{VAR}[X_i] \quad (2.3.20)$$

Na prática, é mais fácil calcular a variância empregando o seguinte teorema:

TEOREMA 2.3.1: $\text{VAR}[X_i] = E[X_i^2] - (E[X_i])^2$

Demonstração:

$$\text{VAR}[X_i] = E[(X_i - \bar{X}_i)^2] \quad (2.3.21)$$

$$= E[X_i^2 - 2X_i\bar{X}_i + \bar{X}_i^2] \quad (2.3.22)$$

$$= E[X_i^2] - E[2X_i\bar{X}_i] + E[\bar{X}_i^2] \quad (2.3.23)$$

$$= E[X_i^2] - 2E[X_i]E[\bar{X}_i] + (E[X_i])^2 \quad (2.3.24)$$

$$= E[X_i^2] - (E[X_i])^2 \quad (2.3.21)$$

△

Assim, para calcular a variância de X_i basta achar a média dos quadrados de $X_i(k)$ ($k = 1, 2, \dots, N$), o quadrado da média de X_i e depois subtraí-los.

DESVIO MÉDIO PADRÃO

Definição 2.3.3: O desvio médio padrão de uma variável aleatória X_i é dado por:

$$\sigma_{X_i} = \sqrt{\sigma_{X_i}^2} \quad (2.3.26)$$

Observando a equação 2.3.12 podemos reescrever a equação 2.3.26 da seguinte forma:

$$\sigma_{X_i} = \sqrt{\frac{\sum_k^n (X_i(k) - \bar{X}_i)^2}{N}} \quad (2.3.27)$$

Essa definição é útil, entre outras coisas, devido ao fato de σ_{X_i} ser expresso na mesma unidade de X_i , o que não ocorre com a variância. Como propriedades do desvio médio

padrão temos: (sejam a e b constantes e $i, i' = 1, 2, \dots, n$)

$$i) \text{ Se } X_i \text{ é constante então } \sigma_{X_i} = 0 \quad (2.3.28)$$

$$ii) \sigma_{[X_i+a]} = \sigma_{X_i} \quad (2.3.29)$$

Mais uma vez a origem usada não interferirá no cálculo.

$$iii) \sigma_{[aX_i]} = |a| \sigma_{X_i} \quad (2.3.30)$$

Resumindo as propriedades ii e iii concluimos

$$\text{que se } H(X_i) = aX_i + b \text{ então } \sigma_{H(X_i)} = |a| \sigma_{X_i} \quad (2.3.31)$$

$$\text{Observe que se } a = \frac{1}{\sigma_{X_i}} \text{ então } \sigma_{H(X_i)} = \sigma_{H(X_i)}^2 = 1$$

COVARIÂNCIA

Definição 2.3.4: A covariância entre as variáveis aleatórias

X_i e $X_{i'}$, é dada por:

$$\text{COV}[X_i; X_{i'}] = E[(X_i - \bar{X}_i)(X_{i'} - \bar{X}_{i'})] \quad (2.3.32)$$

$$\forall i, i' = 1, 2, \dots, n$$

$$i \neq i'$$

Além da notação $\text{COV}[X_i; X_{i'}]$, utilizaremos também a notação $c_{ii'}$, adotada por KENDALL [19].

Essas notações foram adotadas visando facilitar o desenvolvimento do Capítulo III. Outra notação também encontrada na literatura é $\sigma_{X_i X_{i'}}$.

Agora a equação 2.3.32 pode ser escrita da seguinte forma:

$$\text{COV}[X_i; X_{i'}] = c_{ii'} = E[(X_i - \bar{X}_i)(X_{i'} - \bar{X}_{i'})] \quad (2.3.33)$$

$$\forall i, i' = 1, 2, \dots, n$$

$$i \neq i'$$

Essa definição pode ser dada também na forma de somatório. Substituindo a equação 2.3.1 na 2.3.33 teremos:

$$\text{COV}[X_i; X_{i'}] = c_{ii'} = \frac{1}{N} \sum_{k=1}^N [(X_i(k) - \bar{X}_i)(X_{i'}(k) - \bar{X}_{i'})] \quad (2.3.34)$$

$$\forall i, i' = 1, 2, \dots, n$$

$$i \neq i'$$

Observando a equação 2.3.32 notamos que $\text{COV}[X_i; X_{i'}]$ será positiva se, em média, $X_i(k) > \bar{X}_i$ acarretar $X_{i'}(k) > \bar{X}_{i'}$, ou $X_i(k) < \bar{X}_i$ acarretar $X_{i'}(k) < \bar{X}_{i'}$, (é bom notarmos que a $\text{COV}[X_i; X_{i'}]$ não depende de um dado par de valores $(X_i(k); X_{i'}(k))$ mas sim do conjunto de todos os pares $(X_i(k); X_{i'}(k))$ $k = 1, 2, \dots, N$. Por isso o emprego do termo "em média"). Nesse caso, em média, $(X_i(k) - \bar{X}_i)$ terá o mesmo sinal de $(X_{i'}(k) - \bar{X}_{i'})$ o que implicará em $\text{COV}[X_i; X_{i'}] > 0$. Por outro lado se, em média, $X_i(k) > \bar{X}_i$ acarretar $X_{i'}(k) < \bar{X}_{i'}$, ou $X_i(k) < \bar{X}_i$ acarretar $X_{i'}(k) > \bar{X}_{i'}$, teremos $\text{COV}[X_i; X_{i'}] < 0$. Observamos ainda que se nenhum desses dois casos ocorrer é de se esperar que $\text{COV}[X_i; X_{i'}]$ seja aproximadamente igual a zero. Assim, verificamos que essa grandeza nos mostra o quanto o fato de X_i se afastar de sua média implica em que $X_{i'}$ também o faça. Ou seja, qual a tendência de X_i e $X_{i'}$, variarem conjuntamente.

A covariância é de grande importância no Modelo das Componentes Principais. Porém, há um problema. Ela depende das unidades em que X_i e $X_{i,}$ são medidas, como veremos nos exemplos mais adiante. Assim, o conhecimento da $COV[X_i; X_{i,}]$, sem o conhecimento da $VAR[X_i]$ e $VAR[X_{i,}]$ não nos dá uma noção real do relacionamento de X_i com $X_{i,}$, já que mudando as unidades de medida ela se altera. Torna-se então necessário o conceito de correlação que irá ser abordado mais adiante.

Veremos agora um teorema que será de grande utilidade no Capítulo III.

Teorema 2.3.2: $COV[X_i; X_{i,}] = E[X_i X_{i,}] - E[X_i] E[X_{i,}]$

Demonstração:

$$COV[X_i; X_{i,}] = E[(X_i - \bar{X}_i)(X_{i,} - \bar{X}_{i,})] \quad (2.3.35)$$

$$= E[(X_i X_{i,}) - (\bar{X}_i X_{i,}) - (X_i \bar{X}_{i,}) + (\bar{X}_i \bar{X}_{i,})] \quad (2.3.36)$$

mas pela equação 2.3.5 podemos concluir que:

$$COV[X_i; X_{i,}] = E[(X_i X_{i,})] - E[(\bar{X}_i X_{i,})] - E[(X_i \bar{X}_{i,})] + E[(\bar{X}_i \bar{X}_{i,})] \quad (2.3.37)$$

como \bar{X}_i e $\bar{X}_{i,}$ são constantes temos:

$$COV[X_i; X_{i,}] = E[X_i X_{i,}] - \bar{X}_i E[X_{i,}] - \bar{X}_{i,} E[X_i] + \bar{X}_i \bar{X}_{i,} \quad (2.3.38)$$

calculando os valores esperados temos:

$$COV[X_i; X_{i,}] = E[X_i X_{i,}] - \bar{X}_i \bar{X}_{i,} - \bar{X}_{i,} \bar{X}_i + \bar{X}_i \bar{X}_{i,} \quad (2.3.39)$$

$$= E[X_i X_{i,}] - \bar{X}_i \bar{X}_{i,} \quad (2.3.40)$$

$$= E[X_i X_{i,}] - (E[X_i] E[X_{i,}]) \quad (2.3.41)$$

△

Uma observação que pode ser feita é que se nas equações 2.3.32-34 fizermos $i = i'$ teremos a definição de variância. Assim, podemos estender o conceito de covariância para esse caso e então a variância passará a ser um caso particular de $c_{ii'}$, onde $i = i'$, ou seja:

$$c_{ii} = \text{COV}[X_i; X_i] = \text{VAR}[X_i] \quad (2.3.42)$$

Podemos então observar que o teorema 2.3.1 torna-se um caso particular do teorema 2.3.2 para $i = i'$.

Além disso, podemos notar ainda que:

$$\text{COV}[X_i; X_{i'}] = \text{COV}[X_{i'}; X_i] \quad \text{ou} \quad c_{ii'} = c_{i'i} \quad (2.3.43)$$

CORRELAÇÃO

Definição 2.3.5: A correlação (ou coeficiente de correlação) entre as variáveis aleatórias X_i e $X_{i'}$, é dada por:

$$\text{COR}[X_i; X_{i'}] = \frac{\text{COV}[X_i; X_{i'}]}{\sigma_{X_i} \sigma_{X_{i'}}} \quad (2.3.44)$$

Outra notação também encontrada na literatura é

$\rho_{X_i X_{i'}}$

Na forma de somatório a equação 2.3.44 ficará:

$$\text{COR}[X_i; X_{i'}] = \frac{\frac{1}{N} \sum_{k=1}^N [(X_i(k) - \bar{X}_i)(X_{i'}(k) - \bar{X}_{i'})]}{\sqrt{\frac{1}{N} \sum_{k=1}^N (X_i(k) - \bar{X}_i)^2} \sqrt{\frac{1}{N} \sum_{k=1}^N (X_{i'}(k) - \bar{X}_{i'})^2}} \quad (2.3.45)$$

que pode ser simplificada para:

$$\text{COR}[X_i; X_{i'}] = \frac{\sum_{k=1}^N [(X_i(k) - \bar{X}_i)(X_{i'}(k) - \bar{X}_{i'})]}{\sqrt{\sum_{k=1}^N (X_i(k) - \bar{X}_i)^2} \sqrt{\sum_{k=1}^N (X_{i'}(k) - \bar{X}_{i'})^2}} \quad (2.3.46)$$

Essa definição corresponde a calcular a variância entre X_i e $X_{i'}$, tomando como unidade na medição de cada variável o seu desvio médio padrão, senão vejamos. Inicialmente, medimos as variáveis X_i numa unidade qualquer. Depois calculamos σ_{X_i} . Em seguida dividimos cada um dos valores $X_i(k)$ ($k = 1, 2, \dots, n$) por σ_{X_i} . Sabemos (ver equações 2.3.26-31) que as novas variáveis terão desvio padrão igual à unidade. Levando esse resultado na definição 2.3.5 podemos então notar que a correlação das novas variáveis será igual à covariância entre elas.

Notamos então que $\text{COR}[X_i; X_{i'}]$ é adimensional pois X_i e σ_{X_i} são medidos na mesma unidade (ver equação 2.3.27), o mesmo acontecendo para $X_{i'}$ e $\sigma_{X_{i'}}$.

Outra propriedade da correlação é que:

$$-1 \leq \text{COR}[X_i; X_{i'}] \leq 1 \quad (2.3.47)$$

(ver demonstração em [21], pág.183)

Podemos também concluir da definição de correlação que:

$$\text{COR}[X_i; X_{i'}] = \text{COR}[X_{i'}; X_i] \quad (2.3.48)$$

É bom lembrar que a correlação mede o grau de linearidade entre X_i e $X_{i'}$, isto é, o quanto X_i se aproxima de

ser uma função linear de $X_{i'}$, e vice-versa. Assim, se $|\text{COR}[X_i; X_{i'}]| \approx 1$ podemos afirmar que X_i é aproximadamente uma função linear de $X_{i'}$, ou seja, $X_i \approx a X_{i'} + b$ ($a, b = \text{ctes}$). Já se $|\text{COR}[X_i; X_{i'}]| \approx 0$ só podemos afirmar que X_i não é uma função linear de $X_{i'}$, o que não impede que X_i seja um outro tipo de função de $X_{i'}$, como veremos no exemplo 2.3.4.

Podemos então concluir o seguinte:

TEOREMA 2.3.3: Se $X_i = a X_{i'} + b$ (a e b constantes) então $|\text{COR}[X_i; X_{i'}]| = 1$

Demonstração:

Por definição (ver 2.3.44) temos:

$$\text{COR}[X_i; X_{i'}] = \frac{\text{COV}[X_i; X_{i'}]}{\sigma_{X_i} \sigma_{X_{i'}}} \quad (2.3.49)$$

mas pelo teorema 2.3.2

$$\text{COV}[X_i; X_{i'}] = E[X_i X_{i'}] - E[X_i] E[X_{i'}] \quad (2.3.50)$$

Como, por hipótese,

$$X_i = a X_{i'} + b \quad (2.3.51)$$

temos:

$$\text{COV}[X_i; X_{i'}] = E[aX_{i'}^2 + bX_{i'}] - E[X_{i'}] E[aX_{i'} + b] \quad (2.3.52)$$

$$= aE[X_{i'}^2] + bE[X_{i'}] - a\{E[X_{i'}]\}^2 - bE[X_{i'}] \quad (2.3.52A)$$

$$= a\{E[X_{i'}^2] - (E[X_{i'}])^2\} \quad (2.3.52B)$$

o que, pelo teorema 2.3.1, nos leva a

$$\text{COV}[X_i; X_{i'}] = a \sigma_{X_{i'}}^2 \quad (2.3.53)$$

Por outro lado, de acordo com as equações 2.3.29-

30

$$\sigma_{X_i} = \sigma_{aX_i + b} = |a| \sigma_{X_i}, \quad (2.3.54)$$

Substituindo as equações 2.3.53-4 na 2.3.49 teremos:

$$\text{COR}[X_i; X_i] = \frac{a \sigma_{X_i}^2}{|a| \sigma_{X_i}^2} \quad (2.3.55)$$

$$= \frac{a}{|a|} \quad (2.3.55A)$$

$$\text{logo } |\text{COR}[X_i; X_i]| = 1$$

PADRONIZAÇÃO DE VARIÁVEIS

Vimos, no conceito de covariância, que a escala em que as variáveis são medidas é importante. Daí surgiu o conceito de correlação, que, como vimos, é adimensional.

Se por outro lado, observarmos a equação 2.3.7, concluimos que a origem a partir da qual os valores foram tomados influenciam em $E[X_i]$.

Assim, surgiu a idéia de padronizar as variáveis a fim de que elas tenham:

- a) média nula
- b) variância unitária

Essa padronização ditará a partir de que origem faremos as medições e em que escala. Observe que ela nada tem a haver com as características apresentadas pela população mas sim com a forma de se efetuar a medição dessas características.

DESVIOS

Definição 2.3.6: O desvio x_i da variável X_i é dado pela equação

$$x_i = X_i - \bar{X}_i \quad (2.3.56)$$

Muitas vezes usaremos x_i ao invés de X_i já que os desvios apresentam propriedades interessantes. São elas:

a) x_i e X_i são expressos na mesma unidade.

$$b) E[x_i] = 0 \quad (2.3.57)$$

$$c) \text{VAR}[x_i] = \text{VAR}[X_i] = c_{ii} \quad (2.3.58)$$

essa propriedade já era de se esperar que se verificasse, já que a variância é uma medida de dispersão, não dependendo dos valores absolutos $X_i(k)$ ou $x_i(k)$.

$$d) \sigma_{x_i} = \sigma_{X_i} \quad (2.3.59)$$

$$e) \text{COV}[x_i; x_i] = E[x_i x_i] = c_{ii}, \quad (2.3.59A)$$

$$f) \text{COV}[x_i; x_i] = \text{COV}[X_i; X_i] \quad (2.3.60)$$

$$g) \text{COR}[x_i; x_i] = \text{COR}[X_i; X_i] \quad (2.3.61)$$

Podemos também notar que como, em geral, $\sigma_{x_i} \neq 1$; teremos:

$$\text{COR}[x_i; x_i] \neq \text{COV}[x_i; x_i] \quad (2.3.62)$$

Definimos os desvios, que são variáveis com média nula. Agora vamos definir outras variáveis de forma a que além da média nula, elas tenham variância unitária.

VARIÁVEIS PADRONIZADAS

Definição 2.3.7: Uma variável padronizada z_i é dada pela fórmula

$$z_i = \frac{x_i}{\sigma_{X_i}} \quad (2.3.63)$$

ou equivalentemente

$$z_i = \frac{X_i - \bar{X}_i}{\sigma_{X_i}} \quad (2.3.64)$$

Muitas vezes usaremos z_i ao invés de X_i , já que as variáveis padronizadas apresentam propriedades interessantes. São elas:

a) z_i é adimensional

b) $E[z_i] = 0$

c) $\text{VAR}[z_i] = 1$

d) $\sigma_{z_i} = 1$

e) $\text{COR}[z_i; z_{i'}] = \text{COV}[z_i; z_{i'}] = E[z_i z_{i'}] \quad (2.3.68)$

f) $\text{COR}[z_i; z_{i'}] = \text{COR}[X_i; X_{i'}] \quad (2.3.69)$

Com essas definições podemos notar que toda variável padronizada é um desvio, porém o contrário não é verdadeiro.

Veremos agora exemplos que ilustram os conceitos vistos acima. Como o objetivo é apenas didático os dados serão hipotéticos e a forma de resolver não será a mais prática mas sim aquele que permita melhor assimilar o que foi visto.

EXEMPLO 2.3.1:

Vamos admitir que, numa linha de fabricação, tenhamos medido o comprimento (X_1) e o peso (X_2) de lâminas de aço e encontramos os valores que constam do quadro 2.3.1. Representando graficamente esses dados encontramos a figura 2.3.1

X_1 (em metros)	X_2 (em Kg)
16	100
9	60
8	40
13	80
18	110
15	80
4	20
11	60
6	40
18	100

Quadro 2.3.1: Comprimento x peso de 10 lâminas de aço

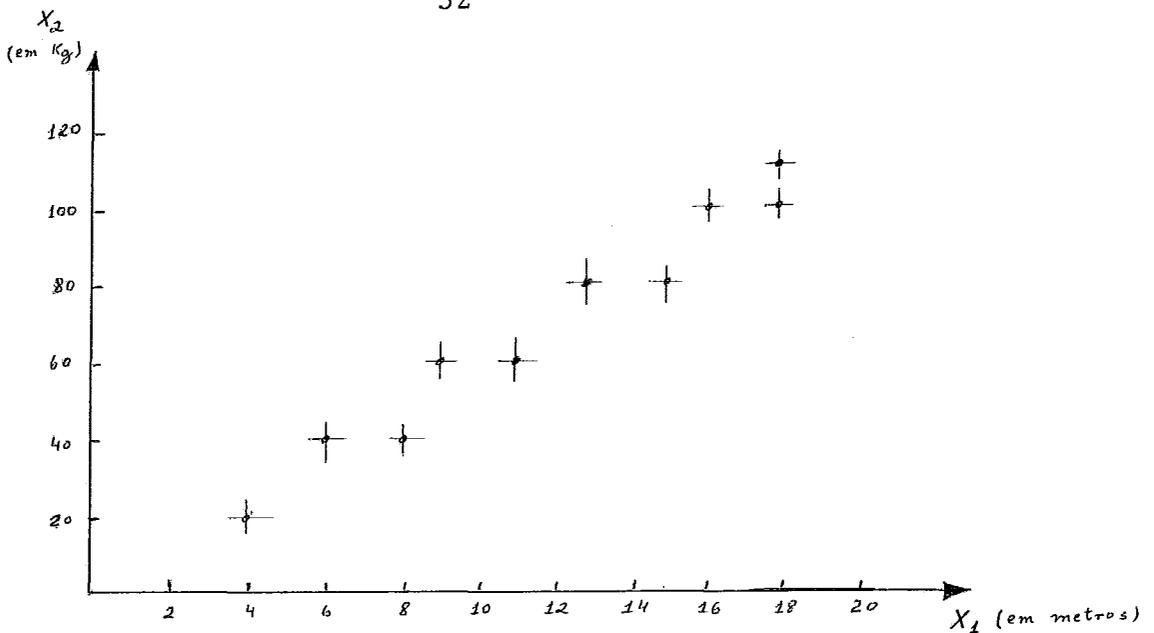


FIGURA 2.3.1: Representação do quadro 2.3.1

Observando essa figura notamos que parece haver influência do comprimento das lâminas e seu peso (e vice-versa) o que, lógicamente era de se esperar. Podemos notar que essa influência parece poder ser expressa aproximadamente por uma função linear, já que podemos passar uma reta por entre os pontos da figura 2.3.1 tal que os pontos não fiquem muito afastados dela. Isso nos leva a concluir que deverá haver uma forte correlação entre X_1 e X_2 .

Vamos agora fazer uma análise estatística dos dados. Efetuando os cálculos podemos chegar ao seguinte quadro:

$\bar{X}_1 = 11,8 \text{ m}$	$\bar{X}_2 = 69 \text{ Kg}$
$\text{VAR}[X_1] = 22,63 \text{ m}^2$	$\text{VAR}[X_2] = 809 \text{ Kg}^2$
$\sigma_{X_1} = 4,729 \text{ m}$	$\sigma_{X_2} = 28,443 \text{ Kg}$
$\text{COV}[X_1; X_2] = 131,8 \text{ mKg}$	
$\text{COR}[X_1; X_2] = 0,98 = 98\%$	

QUADRO 2.3.2: Valor de grandezas estatísticas correspondentes aos valores do quadro 2.3.1.

Dele podemos tirar algumas conclusões, que vem a seguir, e que confirmam o que foi visto anteriormente.

- a) de fato X_1 e X_2 são altamente correlacionados já que $COR[X_1;X_2] = 98\%$. Assim podemos dizer que $X_2 \approx aX_1+b$ (a e b constantes)
- b) a correlação é adimensional
- c) só a covariância não nos dá uma noção real de como X_1 e X_2 se relacionam. Podemos observar que o valor 131,8 mKg é um valor que só tem algum significado em presença das variâncias de X_1 e X_2 .
- d) o desvio médio padrão de uma variável é medido na mesma unidade em que o foi essa variável.

Vamos ver, ainda nesse exemplo, o que aconteceria se tivéssemos padronizado as variáveis. De posse dos dados dos quadros 2.3.1 e 2.3.2, poderíamos formar o quadro 2.3.3, que está representado na figura 2.3.2

z_1	z_2
0.888	1.090
-0.592	-0.316
-0.804	-1.020
0.254	0.387
1.311	1.441
0.676	0.387
-1.649	-1.723
-0.169	-0.316
-1.226	-1.020
1.311	1.090

QUADRO 2.3.3: Comprimento x peso (padronizados) de 10 lâminas de aço

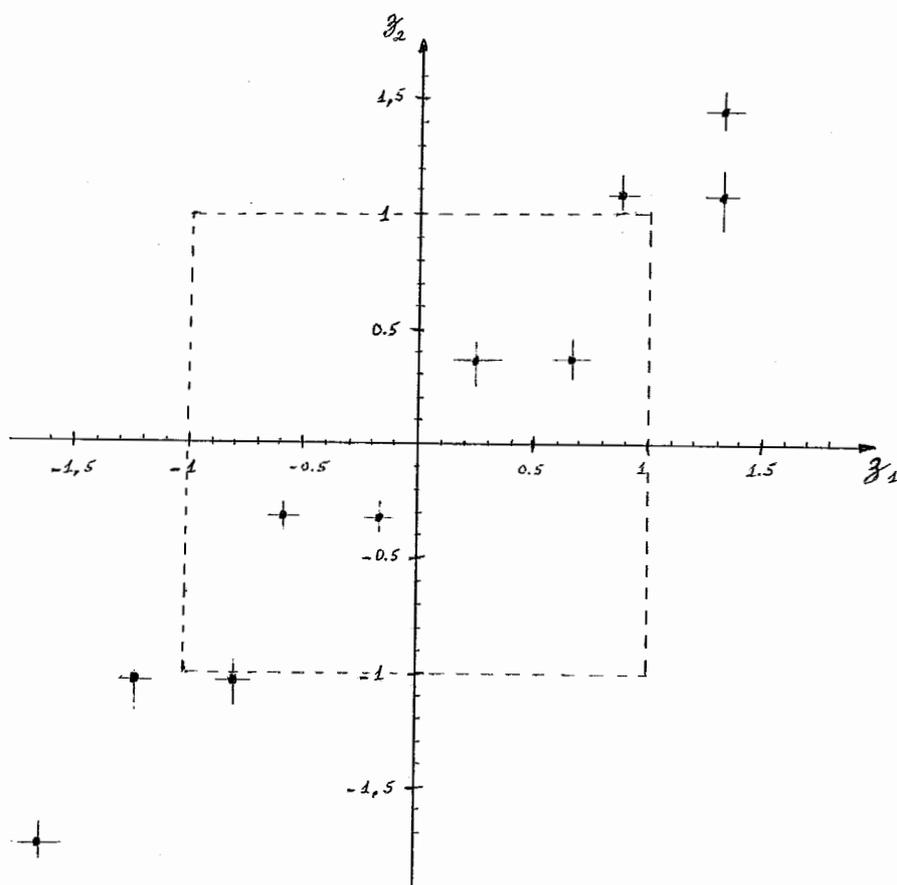


FIGURA 2.3.2: Representação do quadro 2.3.3

Efetuada, mais uma vez, uma análise estatística nos dados, encontraremos o seguinte quadro:

$\bar{z}_1 = 0$	$\bar{z}_2 = 0$
$\text{VAR}[z_1] = 1$	$\text{VAR}[z_2] = 1$
$\sigma_{z_1} = 1$	$\sigma_{z_2} = 1$
$\text{COV}[z_1; z_2] = 0,98$	
$\text{COR}[z_1; z_2] = 0,98$	

Quadro 2.3.4: Valor de grandezas estatísticas correspondentes aos valores do quadro 2.3.3

Desse quadro podemos tirar algumas conclusões que vem confirmar o que foi estudado. São elas:

- a) as variáveis z_1 e z_2 são adimensionais
- b) como z_1 e z_2 são padronizadas as suas médias são nulas e suas variâncias e desvio médio padrão são unitários
- c) $COR[z_1; z_2] = COV[z_1; z_2]$
- d) $COR[z_1; z_2] = COR[X_1; X_2]$

Finalmente é bom lembrar que como as médias de z_1 e z_2 são nulas, os pontos $z(k)$ ($k = 1, 2, \dots, 10$) se distribuem em torno da origem, e como suas variâncias são unitárias (o que significa dizer que a média dos quadrados das distâncias desses pontos a cada eixo coordenado é 1. Ver definição de distância na seção 2.4) esses pontos, em geral, não se afastam muito do quadrado definido pelos pontos $(1;1)(-1;1)(1;-1)(-1;-1)$.



EXEMPLO 2.3.2:

Fez-se um estudo, em uma cidade de clima frio, sobre o consumo de combustível empregado no aquecimento doméstico. Para isso, procurou-se analisar duas variáveis, a temperatura média diária (X_1) e o consumo diário de gás (X_2). Os valores obtidos na medição estão no quadro 2.3.5 e a sua representação se encontra na figura 2.3.3.

X_1 (em $^{\circ}\text{C}$)	X_2 (em 1000 m^3)
15	160
22	120
10	160
34	110
26	100
-10	260
8	180
-6	210
10	200
18	180
12	180
30	80
18	120
6	210
24	140

Quadro 2.3.5: Temperatura x consumo de gás (valores médios de 15 dias)

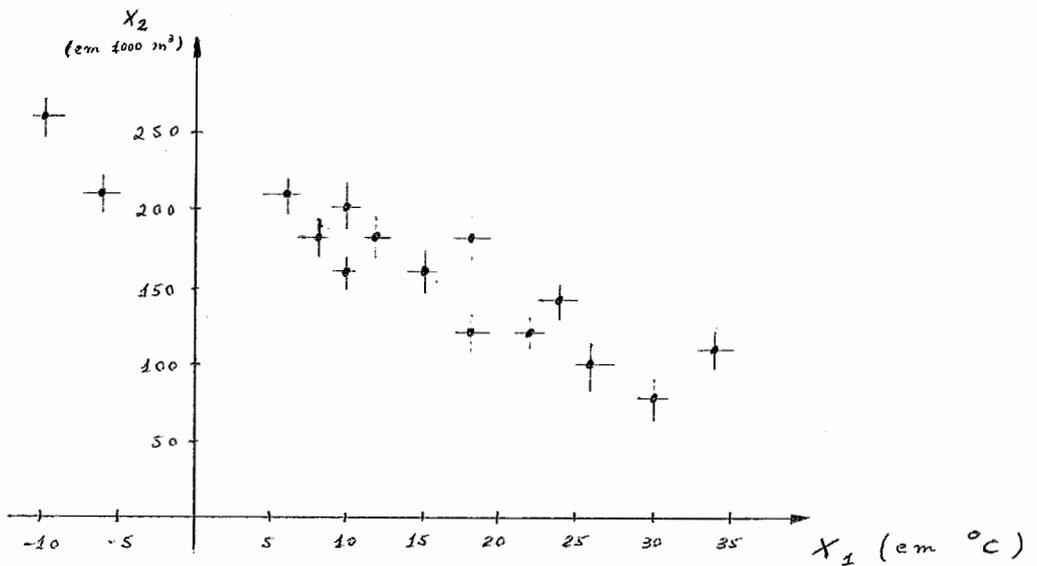


FIGURA 2.3.3: Representação do quadro 2.3.5

Podemos então observar que parece também haver uma forte correlação entre as variáveis X_1 e X_2 . Para a comprovação disso vamos efetuar os cálculos, criando o quadro 2.3.6.

$\bar{X}_1 = 14,467 \text{ } ^\circ\text{C}$	$\bar{X}_2 = 160,667 \times 1000 \text{ m}^3$
$\text{VAR}[X_1] = 140,373 (\text{ } ^\circ\text{C})^2$	$\text{VAR}[X_2] = 2256,235 \times (1000 \text{ m}^3)^2$
$\sigma_{X_1} = 11,848 \text{ } ^\circ\text{C}$	$\sigma_{X_2} = 47,499 \times 1000 \text{ m}^3$
$\text{COV}[X_1; X_2] = -512,977 \times 1000 \text{ m}^3 \text{ } ^\circ\text{C}$	
$\text{COR}[X_1; X_2] = -0,912 = -91,2\%$	

QUADRO 2.3.6: Valor de Grandezas Estatísticas referente aos valores do quadro 2.3.5.

Observando o quadro 2.3.6 poderíamos tirar conclusões análogas às do exemplo anterior. Observamos então que entre X_1 e X_2 há uma forte correlação já que $|\text{COR}[X_1; X_2]| \approx 1$. Devido à correlação ser negativa, notamos ainda que X_1 e X_2 variam inversamente, ou seja, $X_2 \approx a X_1 + b$ (a e b constantes) $a < 0$. Quando X_1 tende a aumentar, X_2 tende, em geral, a diminuir.

Para ilustrar o exemplo temos a seguir o quadro 2.3.7 com as variáveis X_1 e X_2 padronizadas, ou seja, z_1 e z_2 e, também, a figura 2.3.4 que representa os pontos $z(k)$ ($k = 1, 2, \dots, 15$).

z_1	z_2
0,005	-0,014
0,636	-0,856
-0,377	-0,014
1,649	-1,067
0,973	-1,277
-2,065	2,091
-0,546	0,407
-1,727	1,039
-0,377	0,828
0,298	0,407
-0,208	0,407
1,311	-1,698
0,298	-0,856
-0,715	1,039
0,805	-0,435

QUADRO 2.3.7: Temperatura x consumo de gás, padronizados (valores médios de 15 dias)

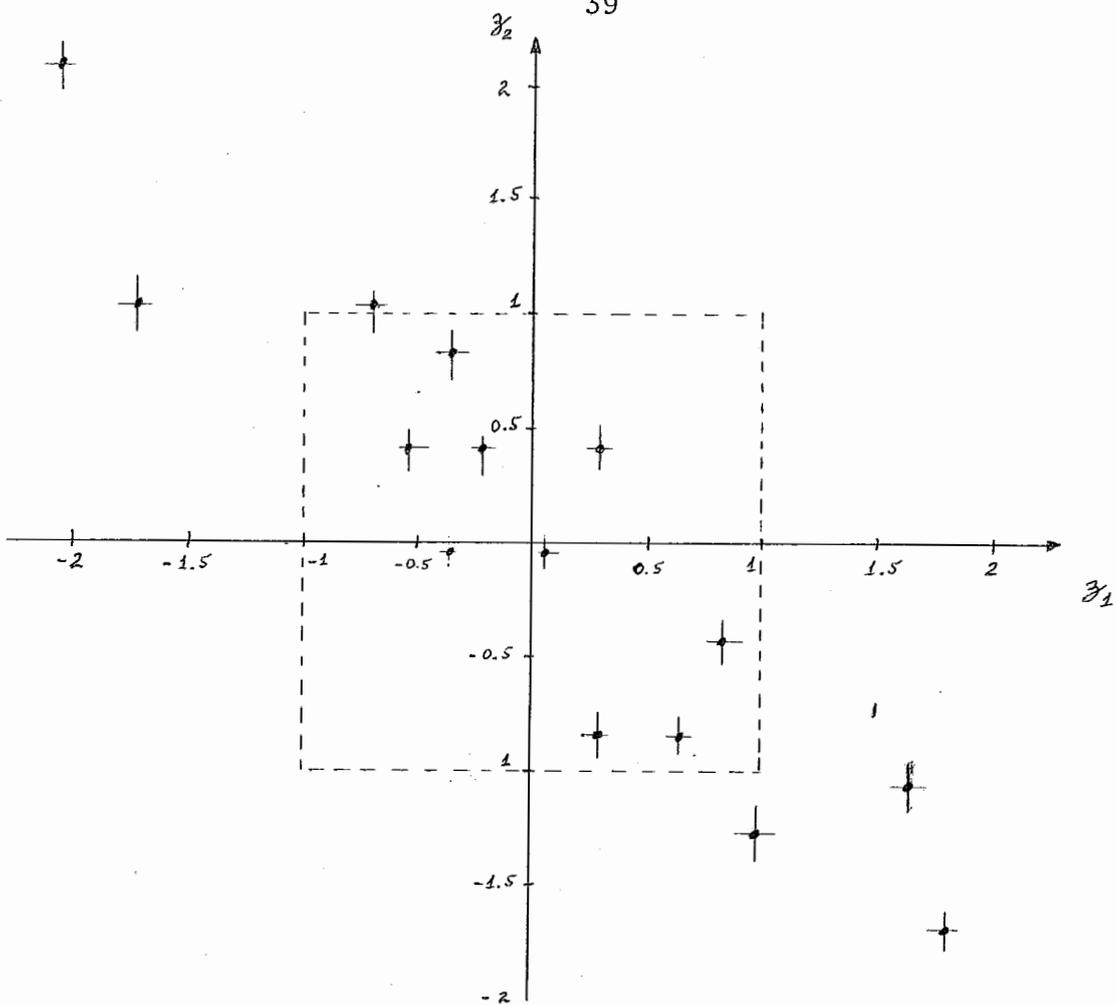


FIGURA 2.3.4: Representação do quadro 2.3.4

Mais uma vez poderíamos efetuar os cálculos com z_1 e z_2 e formar um quadro como 2.3.4. Isso não será feito pois os resultados são análogos.

Finalmente podemos lembrar que mais uma vez se verifica o que foi dito no exemplo anterior, ou seja, que os pontos $z(k)$ ($k = 1, 2, \dots, 15$) se distribuem em torno da origem e em geral, não se afastam muito do quadrado $(1;1)$ $(-1;1)$ $(1;-1)$ e $(-1;-1)$. ▲

EXEMPLO 2.3.3:

Medimos, num grupo de pessoas, de 25 a 50 anos as suas idades (X_1) e seus pesos (X_2) e tentamos encontrar alguma relação entre essas variáveis. Temos no quadro 2.3.8 os valores encontrados. A representação desses dados está na figura 2.3.5.

X_1 (em anos)	X_2 (em Kg)
47	60
40	50
25	60
30	84
49	60
42	56
33	64
36	52
28	92
43	92
48	50
48	95
44	64
29	68
35	100
26	80
45	76
41	76
38	75
32	80

QUADRO 2.3.8: Idade x peso de 20 pessoas

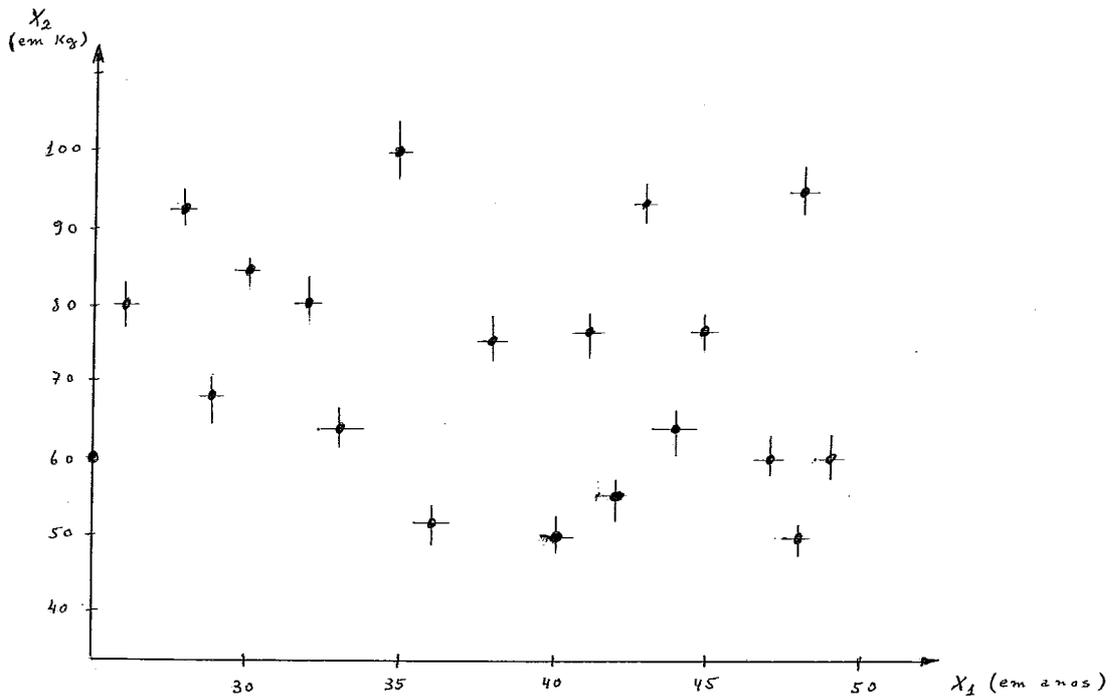


FIGURA 2.3.5: Representação do quadro 2.3.8

Efetuada os cálculos encontraremos que $COR[X_1; X_2] = -0,12$, o que demonstra que praticamente X_1 e X_2 não são correlacionados (é bom lembrar mais uma vez que o fato delas não serem correlacionadas indica apenas não haver uma função linear de X_1 em X_2) o que pode ser visto na figura 2.3.5.

A seguir, temos o quadro 2.3.9 e a figura 2.3.6 com as variáveis padronizadas z_1 e z_2 .

z_1	z_2
1,182	-0,769
0,268	-1,427
-1,691	-0,769
-1,038	0,809
1,443	-0,769
0,529	0,112
-0,646	0,506
-0,255	0,295
-1,299	1,335
0,659	1,335
1,312	-1,427
1,312	1,533
2,096	0,506
-1,169	0,243
-0,385	1,861
-1,560	0,546
0,920	0,283
0,398	0,283
0,006	0,217
-0,777	0,546

QUADRO 2.3.9: Idade x peso (padronizados) de 20 pessoas.

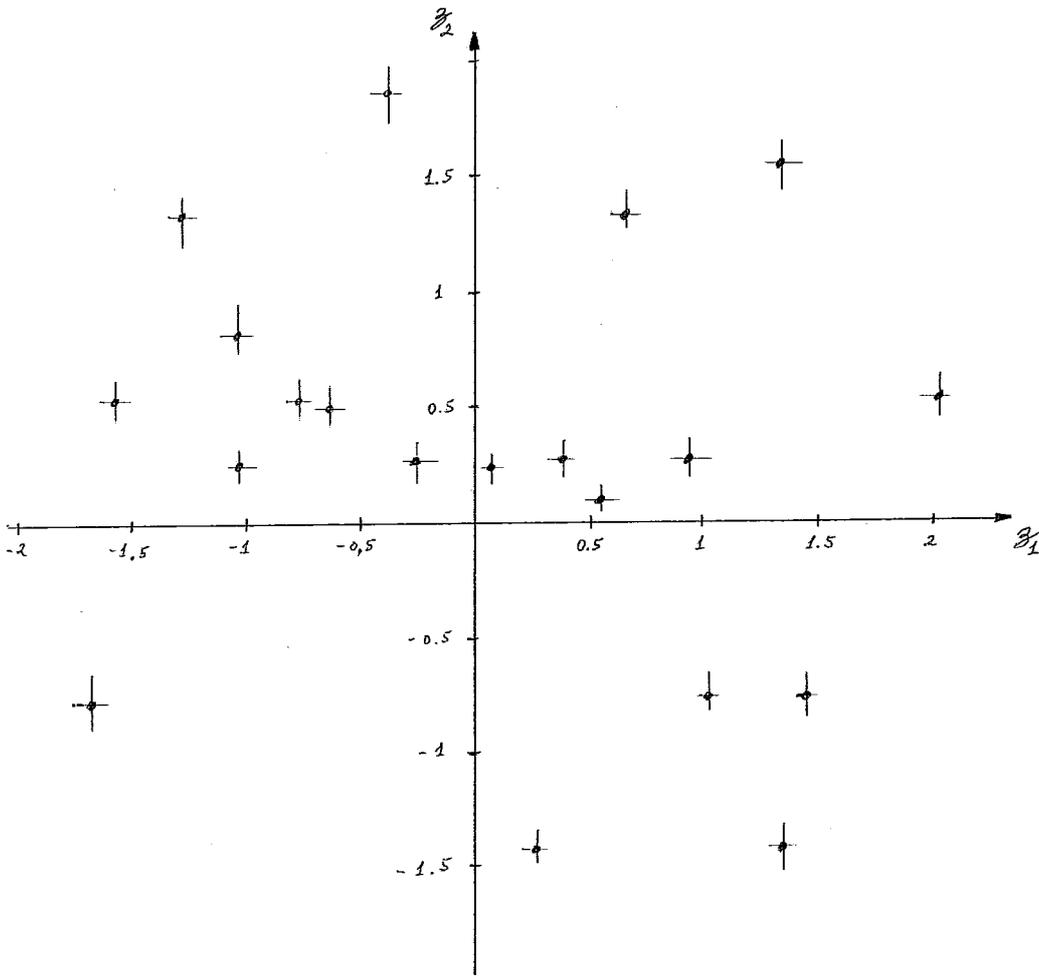


FIGURA 2.3.6: Representação do quadro 2.3.9



EXEMPLO 2.3.4:

Fotografou-se a tela de um radar durante 10 segundos em intervalos de 1 segundo, verificando-se que um móvel tomava as posições indicadas no quadro 2.3.10 e na figura 2.3.7.

X_1 (em cm)	X_2 (em cm)
0	0
3	4,5
6	8,0
9	10,5
12	12,0
15	12,5
18	12,0
21	10,5
24	8,0
27	4,5
30	0

QUADRO 2.3.10: Posição de um móvel numa tela de radar.

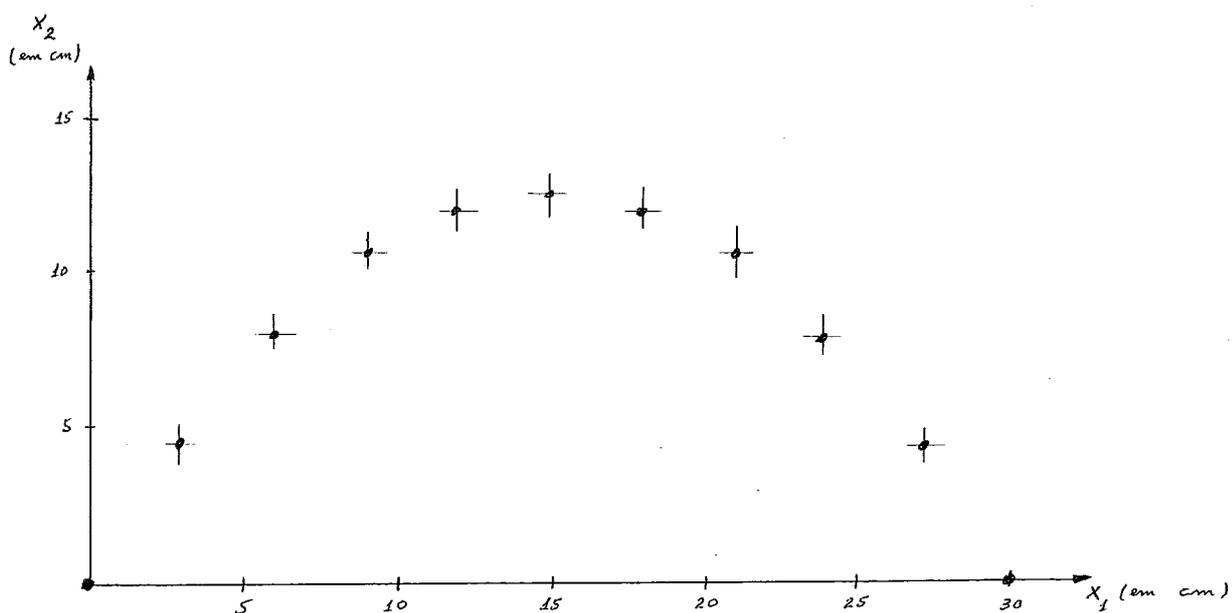


FIGURA 2.3.7: Representação do quadro 2.3.10

Calculando-se a correlação entre X_1 e X_2 encontraremos $COR[X_1; X_2] = 0$, ou seja, X_1 e X_2 não são correlacionados, o que significa meramente que X_2 não é função linear de X_1 (ou vice-versa). Porém, examinando a curva com mais cuidado, vemos que ela é uma parábola, podendo ser representada pela equação $18X_2 = 30X_1 - X_1^2$. Assim X_2 depende da variável X_1 embora não tenha nenhuma correlação com ela. ▲

Como já foi dito nesse texto, a correlação indica o grau de linearidade entre as variáveis, ou seja, o quanto uma variável se aproxima de ser uma função linear da outra. Mas o fato de não existir uma função linear não implica em que não exista outro tipo de função. Aliás, foi isso o que ocorreu no exemplo acima, onde a função é uma parábola. Notamos então que o conceito de correlação é um pouco limitado e a sua característica linear é a principal responsável pelo seu largo emprego.

Um conceito mais amplo e que não se limita a linearidade é o conceito de INDEPENDÊNCIA entre variáveis aleatórias. De uma forma intuitiva, duas variáveis X_1 e X_2 são independentes quando não é possível fazer afirmações acerca do comportamento de X_1 , uma vez conhecido o comportamento de X_2 , isto é, não é possível determinar valores de X_1 a partir dos valores de X_2 .

Podemos formalizar a noção de independência assim:

Definição 2.3.8: Duas variáveis aleatórias discretas X_1 e X_2 são INDEPENDENTES se e somente se a probabilidade de $X_1 = a$ dado que $X_2 = b$ é igual à probabilidade de $X_1 = a$ para todo a e $b \in \mathbb{R}$.

Em outras palavras, a probabilidade de X_1 assumir um valor a independe de que valor X_2 tenha assumido.

Podemos escrever essa definição da seguinte forma:

$$P(X_1 = a \mid X_2 = b) = P(X_1 = a) \quad (2.3.70)$$

Sabemos no entanto, pelo conceito de probabilidade condicional (ver [2¹], pág.53) que:

$$P(X_1 = a \mid X_2 = b) = \frac{P(X_1 = a \text{ e } X_2 = b)}{P(X_2 = b)} \quad (2.3.71)$$

Reunindo as equações 2.3.70 e 2.3.71 podemos concluir que X_1 e X_2 são independentes, se e somente se

$$\boxed{P(X_1 = a \text{ e } X_2 = b) = P(X_1 = a)P(X_2 = b)} \quad (2.3.72)$$

Geralmente a equação 2.3.72 é utilizada na literatura para definir o conceito de independência de variáveis aleatórias.

Dessa definição o leitor poderá concluir que se X_1 é independente de X_2 então X_2 é independente de X_1 e vice-versa.

Nesse texto o conceito de correlação será preferido ao de independência (como geralmente acontece), porém é importante que tenhamos em mente as restrições desse conceito.

2.4. Noções de Geometria no \mathbb{R}^n

Partimos, nessa seção, do pressuposto que o leitor tenha conhecimento de dependência e independência linear de vetores, espaços vetoriais, base e dimensão de um espaço vetorial, etc... . Essa matéria pode ser vista em detalhes em [6], [7], [15] e [23]. Vamos aqui abordar somente o que será diretamente utilizado no texto.

Como vimos anteriormente o nosso objetivo será analisar uma população de N elementos dos quais observamos n características. Vimos nos exemplos da seção 2.3 que em alguns casos (como são os daquela seção) podemos representar graficamente esse problema. Para isso utilizamos um sistema de coordenadas cartesianas onde cada um dos dois eixos perpendiculares representava uma característica observada (no caso X_1 ou X_2) e cada $X(k)$ era representado por um ponto de coordenadas $(X_1(k); X_2(k))$ *. Surge então a necessidade de se generalizar para o caso de $n > 2$. Podemos então dizer que quando observamos n características em cada elemento da população, cada $X(k)$ terá coordenadas $(X_1(k), X_2(k), \dots, X_n(k))$. Logicamente se $n > 3$ não poderemos fazer uma representação gráfica do problema e nesse caso teremos que lançar mão de técnicas menos intuitivas para a análise do problema.

Observamos que se associarmos vetores unitários

* Nesse texto utilizaremos somente o sistema de coordenadas cartesianas ortogonais. O sistema não ortogonal pode ser visto em [12].

\vec{X}_i a cada característica X_i e vetores $\vec{X}(k)$ a cada ponto $X(k)$ estaremos diante de um espaço vetorial (o \mathbb{R}^n no nosso caso, que é um espaço vetorial real n-dimensional) onde os vetores \vec{X}_i formam uma base para esse espaço e os $\vec{X}(k)$ são vetores nele contidos. Agora podemos estudar o nosso problema não só sob o aspecto geométrico mas também utilizando a teoria dos espaços vetoriais. No nosso estudo muitas vezes utilizaremos um ou outro enfoque sem fazer muita distinção entre eles.

Visto isso podemos passar a alguns conceito importantes.

Os três conceitos fundamentais são os de PONTO, RETA e PLANO. Já sabemos que os pontos (no caso $X(k)$) terão coordenadas $(X_1(k), X_2(k), \dots, X_n(k))$. As retas, como na geometria clássica, ficam determinadas por 2 pontos ou um ponto e uma direção, e os planos por 3 pontos não colineares ou uma reta e um ponto não pertencente a essa reta.

Além desses, se num espaço vetorial n-dimensional tivermos um sub-espaço de dimensão (n-1) diremos que esse sub-espaço é um HIPERPLANO desse espaço vetorial.

Tratando-se de pontos, surge então o conceito de DISTÂNCIA que vemos a seguir.

DISTÂNCIA ENTRE DOIS PONTOS NO \mathbb{R}^n

Definição 2.4.1: Sejam $X(k)$ e $X(k')$ dois pontos do \mathbb{R}^n . A DISTÂNCIA entre eles é definida por:

$$\text{dist}(X(k); X(k')) = \sqrt{\sum_{i=1}^n (X_i(k) - X_i(k'))^2} \quad (2.4.1)$$

Nota-se que essa definição é uma generalização para o \mathbb{R}^n do conceito habitual de distância no \mathbb{R}^2 .

Podemos verificar que as distâncias satisfazem as seguintes propriedades:

$$\text{i) } \text{dist}(X(k); X(k)) = 0 \quad (2.4.2)$$

$$\text{ii) se } X(k) \neq X(k') \text{ então } \text{dist}(X(k); X(k')) > 0 \quad (2.4.3)$$

$$\text{iii) } \text{dist}(X(k); X(k')) = \text{dist}(X(k'); X(k)) \quad (2.4.4)$$

$$\text{iv) } \text{dist}(X(k); X(k')) + \text{dist}(X(k'); X(k'')) \geq \text{dist}(X(k); X(k'')) \quad (2.4.5)$$

Como já dissemos a cada ponto $X(k)$ podemos associar um vetor $\vec{X}(k)$. Seja então $\vec{X}(k)$ um vetor que vá da origem do espaço vetorial a esse ponto. Sem fazer distinção entre ponto e vetor podemos nos referir simplesmente ao vetor $\vec{X}(k) = (X_1(k), X_2(k), \dots, X_n(k))$. Vejamos então a seguinte definição.

NORMA DE UM VETOR $\vec{X}(k)$

Definição 2.4.2: A NORMA (módulo ou comprimento) de um vetor $\vec{X}(k)$ (acima definido) é dada pela distância de $X(k)$ à origem, ou seja:

$$\text{Norma}(\vec{X}(k)) = |\vec{X}(k)| = \sqrt{\sum_{i=1}^n (X_i(k))^2} \quad (2.4.6)$$

Observe mais uma vez que nesse estudo o conceito de ponto e vetor estão intimamente ligados. Por isso só usaremos a notação $\vec{X}(k)$, ao invés de $X(k)$, quando o conceito de

vetor se fizer indispensável. Vamos agora estudar $\vec{X}(k)$.

Como todo vetor ele deverá ter norma (ou módulo), direção e sentido. A norma, nós já sabemos que é dada pela equação (2.4.6). A direção é a direção da reta que passa pela origem e pelo ponto $X(k)$ e o sentido é o da origem para $X(k)$. Agora podemos defini-lo através dessas tres grandezas, ou seja, se tivermos a norma, a direção e o sentido então $\vec{X}(k)$ estará definido. Para identificarmos o vetor dessa forma, vamos à seguinte definição.

COSSENO DIRETORES DE UM VETOR

Definição 2.4.3: Os COSSENO DIRETORES de um vetor no R^n são dados pela seguinte fórmula:

$$g_i = \cos \alpha_i = \frac{\langle \vec{X}(k) \cdot \vec{X}_i \rangle}{|\vec{X}(k)|} \quad (2.4.7)$$

$$i = 1, 2, \dots, n$$

onde g_i é o cosseno diretor de $\vec{X}(k)$ em relação a X_i , α_i é o ângulo entre $\vec{X}(k)$ e \vec{X}_i , \vec{X}_i é um vetor unitário da base do R^n (vetor unitário é aquele que tem norma igual a 1) e $\langle . \rangle$ significa produto escalar (ou produto interno) de vetores.

Para ilustrar a definição vamos a um exemplo no R^2 .

EXEMPLO 2.4.1

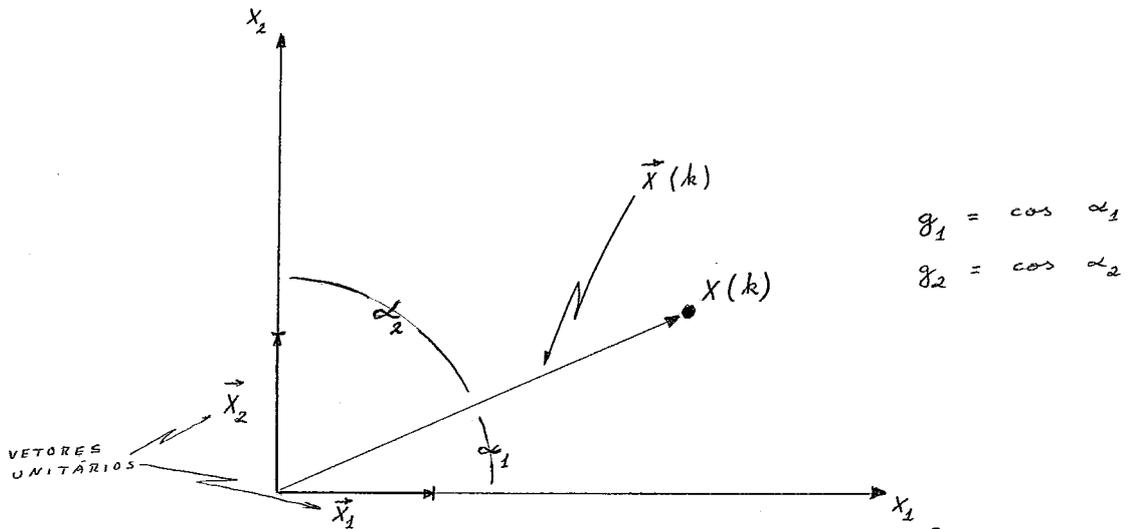


FIGURA 2.4.1: Cossenos diretores de um vetor no \mathbb{R}^2 .

O produto escalar entre os vetores $\vec{X}(k)$ e \vec{X}_1 é dado por:

$$\langle \vec{X}(k) \cdot \vec{X}_1 \rangle = |\vec{X}(k)| |\vec{X}_1| \cos \alpha_1 \quad (2.4.8)$$

mas como \vec{X}_1 é um vetor unitário

$$|\vec{X}_1| = 1 \quad (2.4.9)$$

o que nos leva a concluir que

$$\langle \vec{X}(k) \cdot \vec{X}_1 \rangle = |\vec{X}(k)| \cos \alpha_1 \quad (2.4.10)$$

logo

$$g_1 = \cos \alpha_1 = \frac{\langle \vec{X}(k) \cdot \vec{X}_1 \rangle}{|\vec{X}(k)|} \quad (2.4.11)$$

Analogamente

$$g_2 = \cos \alpha_2 = \frac{\langle \vec{X}(k) \cdot \vec{X}_2 \rangle}{|\vec{X}(k)|} \quad (2.4.12)$$

E teremos g_1 e g_2 como cossenos diretores do vetor (ou ponto) $X(k)$. Logo, de posse dos cossenos sabemos a direção do vetor. ▲

Os cossenos diretores tem a seguinte propriedade:

TEOREMA 2.4.1: Seja $\vec{X}(k) = (X_1(k), X_2(k), \dots, X_n(k))$ um vetor de um espaço vetorial n-dimensional real cuja base é composta pelos vetores $\vec{X}_1, \vec{X}_2, \dots, \vec{X}_n$ onde \vec{X}_i ($i = 1, 2, \dots, n$) é um vetor onde a i-ésima coordenada é 1 e as demais são zero. Seja g_i o cosseno do ângulo entre $\vec{X}(k)$ e \vec{X}_i . Então:

$$\sum_{i=1}^n g_i^2 = 1$$

Demonstração:

Pela definição de produto escalar (ou produto interno) temos:

$$\langle \vec{X}(k) \cdot \vec{X}_i \rangle = |\vec{X}(k)| \cdot |\vec{X}_i| \cdot g_i \quad (2.4.13)$$

Logo

$$g_i = \frac{\langle \vec{X}(k) \cdot \vec{X}_i \rangle}{|\vec{X}(k)| |\vec{X}_i|} \quad (2.4.14)$$

Mas

$$\langle \vec{X}_i(k) \cdot X_i \rangle = \langle (X_1(k), X_2(k), \dots, X_n(k)) \cdot \underbrace{(0, 0, 0, \dots, 0, 1, 0, \dots, 0)}_{(i-1) \text{ zeros}} \rangle \quad (2.4.15)$$

ou seja

$$\langle \vec{X}(k) \cdot X_i \rangle = X_i(k)$$

ou por outro lado

$$|\vec{X}(k)| = \sqrt{X_1(k)^2 + X_2(k)^2 + \dots + X_n(k)^2} \quad (2.4.16)$$

$$e \quad |\vec{X}_i| = 1 \quad (2.4.17)$$

Substituindo as equações 2.4.15-17 na 2.4.14 teremos:

$$g_i = \frac{X_i(k)}{\sqrt{X_1(k)^2 + X_2(k)^2 + \dots + X_n(k)^2}} \quad (2.4.18)$$

Desta forma teremos:

$$\sum_{i=1}^n g_i^2 = \sum_{i=1}^n \frac{X_i(k)^2}{X_1(k)^2 + X_2(k)^2 + \dots + X_n(k)^2} = 1 \quad (2.4.19)$$

△

Já sabemos que uma reta no \mathbb{R}^n fica definida por um ponto e uma direção. Assim podemos definir uma reta por um ponto e uma direção de um vetor desse espaço vetorial. Desta forma podemos estender o conceito de cossenos diretores de um vetor também para retas no \mathbb{R}^n e utilizamos a expressão "cosseno diretor de uma reta".

RETAS ORTOGONAIS

Na geometria clássica, para que duas retas sejam ortogonais (ou perpendiculares) o ângulo entre as suas direções deverá ser 90° , ou seja, o cosseno do ângulo formado por elas deverá ser zero. Vamos estender esse conceito para o \mathbb{R}^n .

Definição 2.4.4: Duas retas do \mathbb{R}^n são ortogonais se e somente se o cosseno do ângulo formado por suas direções for nulo.

Dessa definição podemos concluir o seguinte:

TEOREMA 2.4.2: Duas retas no \mathbb{R}^n de cossenos diretores $g_1 = (g_{11}, g_{12}, \dots, g_{1n})$ e $g_2 = (g_{21}, g_{22}, \dots, g_{2n})$ são ortogonais se e somente se

$$\sum_{i=1}^n g_{1i} g_{2i} = 0$$

Demonstração:

Como vimos anteriormente a direção dessas retas pode ser dada por:

$$\vec{X}(1) \rightarrow (g_{11}, g_{12}, \dots, g_{1n})$$

e

$$\vec{X}(2) \rightarrow (g_{21}, g_{22}, \dots, g_{2n})$$

Vamos então encontrar o ângulo α formado pelos vetores $\vec{X}(1)$ e $\vec{X}(2)$.

$$\cos \alpha = \frac{\langle \vec{X}(1) \cdot \vec{X}(2) \rangle}{|\vec{X}(1)| |\vec{X}(2)|} \quad (2.4.20)$$

mas

$$\langle \vec{X}(1) \cdot \vec{X}(2) \rangle = g_{11}g_{21} + g_{12}g_{22} + \dots + g_{1n}g_{2n} \quad (2.4.21)$$

$$= \sum_{i=1}^n g_{1i}g_{2i} \quad (2.4.22)$$

e

$$|\vec{X}(1)| = \sqrt{g_{11}^2 + g_{12}^2 + \dots + g_{1n}^2} \quad (2.4.23)$$

$$= \sqrt{\sum_{i=1}^n g_{1i}^2} \quad (2.4.24)$$

e analogamente

$$|\vec{X}(2)| = \sqrt{g_{21}^2 + g_{22}^2 + \dots + g_{2n}^2} \quad (2.4.25)$$

$$\sqrt{\sum_{i=1}^n g_{2i}^2} \quad (2.4.26)$$

Levando as equações 2.4.22, 2.4.24 e 2.4.26 na 2.4.20 teremos:

$$\frac{\sum_{i=1}^n g_{1i} g_{2i}}{\sqrt{\sum_{i=1}^n g_{1i}^2} \sqrt{\sum_{i=1}^n g_{2i}^2}} \quad (2.4.27)$$

Levando o resultado do teorema 2.4.1 na equação acima concluímos que:

$$\cos \alpha = \frac{\sum_{i=1}^n g_{1i} g_{2i}}{\sqrt{\sum_{i=1}^n g_{1i}^2} \sqrt{\sum_{i=1}^n g_{2i}^2}} \quad (2.4.28)$$

Mas pela definição 2.4.4 as retas serão ortogonais se e somente se $\cos \alpha = 0$, o que equivale a dizer que:

$$\sum_{i=1}^n g_{1i} g_{2i} = 0 \quad (2.4.29)$$

△

Visto isso podemos passar a um cálculo que será de muita utilidade no Capítulo III.

DISTÂNCIA DE UM PONTO À UMA RETA NO \mathbb{R}^n

Seja uma reta no \mathbb{R}^n definida pelo ponto $A = (a_1, a_2, \dots, a_n)$ e pelos cosenos diretores g_1, g_2, \dots, g_n . Seja $Q = (Q_1, Q_2, \dots, Q_n)$ um ponto qualquer dessa reta que passaremos a chamar de reta AQ . Seja também $X(k) = (X_1(k), X_2(k), \dots, X_n(k))$ um ponto qualquer do \mathbb{R}^n , cuja projeção ortogonal so-

bre a reta AQ é o ponto $P(k) = (P_1(k), P_2(k), \dots, P_n(k))$. Assim podemos dizer que:

$$[\text{dist}(X(k); A)]^2 = [\text{dist}(X(k); P(k))]^2 + [\text{dist}(P(k); A)]^2 \quad (2.4.30)$$

(grosseiramente falando é a equação de Pitágoras aplicada ao \mathbb{R}^n). Para uma visualização geométrica veja a figura 2.4.2.

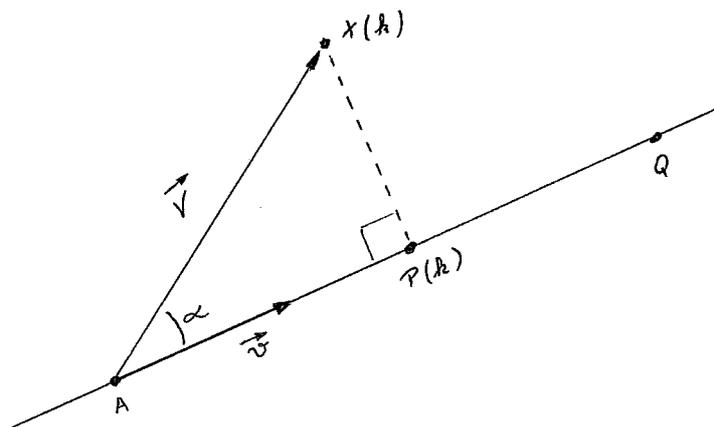


FIGURA 2.4.2: Representação geométrica da equação 2.4.30.

Usando o conceito de distância (ver equação 2.4.1) sabemos que:

$$\text{dist}(X(k); A) = \sqrt{\sum_{i=1}^n (X_i(k) - a_i)^2} \quad (2.4.31)$$

Por outro lado a reta AQ tem como equação:

$$Q_i - a_i = g_i t \quad \forall i = 1, 2, \dots, n \quad (2.4.32)$$

onde t é um valor que independe de i .

Seja então:

$$\vec{v} = (g_1, g_2, \dots, g_n) \quad (2.4.33)$$

Podemos notar que \vec{v} é unitário (ver teorema 2.4.1) e que \vec{v} tem a direção da reta AQ. Assim a projeção AP(k) do vetor $\vec{AX}(k)$ (vetor \vec{V} na figura 2.4.2) na reta AQ é dada por:

$$\text{dist}(A;P(k)) = \langle \vec{V}, \vec{V} \rangle \quad (2.4.34)$$

Mas

$$\begin{aligned} \langle \vec{V}, \vec{V} \rangle &= (X_1(k) - a_1, X_2(k) - a_2, \dots, X_n(k) - a_n) \cdot \\ &\quad \cdot (g_1, g_2, \dots, g_n) \end{aligned} \quad (2.4.35)$$

ou seja

$$\langle \vec{V}, \vec{V} \rangle = \sum_{i=1}^n g_i (X_i(k) - a_i) \quad (2.4.36)$$

Levando a equação 2.4.36 na 2.4.34 concluímos que:

$$\text{dist}(A;P(k)) = \sum_{i=1}^n g_i (X_i(k) - a_i) \quad (2.4.37)$$

Finalmente, levando as equações 2.4.31 e 2.4.37 na 2.4.30 concluímos que:

$$[\text{dist}(X(k);P(k))]^2 = \sum_{i=1}^n (X_i(k) - a_i)^2 - \left[\sum_{i=1}^n g_i (X_i(k) - a_i) \right]^2 \quad (2.4.38)$$

Logo a distância entre $X(k)$ e a reta AQ será:

$$\text{dist}(X(k); \text{reta AQ}) = \sqrt{\sum_{i=1}^n (X_i(k) - a_i)^2 - \left[\sum_{i=1}^n g_i (X_i(k) - a_i) \right]^2} \quad (2.4.39)$$

ou simplesmente

$$\boxed{[\text{dist}(X(k); A)]^2 = \sum_{i=1}^2 (X_i(k) - a_i)^2} \quad (2.4.42)$$

Seja agora

$$\vec{v} = (g_1, g_2, \dots, g_n)$$

mais uma vez \vec{v} será unitário e terá a direção da reta AQ.

Além disso AP(k) será o produto escalar \vec{V} por \vec{v} , ou seja

$$\text{dist}(P(k); A) = \langle \vec{V}, \vec{v} \rangle = (X_1(k) - a_1, X_2(k) - a_2) \times (g_1, g_2) \quad (2.4.43)$$

logo

$$\text{dist}(P(k); A) = g_1(X_1(k) - a_1) + g_2(X_2(k) - a_2) \quad (2.4.44)$$

ou

$$\boxed{\text{dist}(P(k); A) = \sum_{i=1}^2 g_i (X_i(k) - a_i)} \quad (2.4.45)$$

Levando as equações 2.4.42 e 2.4.45 na 2.4.40 temos:

$$\boxed{\text{dist}(X(k); P(k)) = \sqrt{\sum_{i=1}^2 (X_i(k) - a_i)^2 - \left[\sum_{i=1}^2 g_i (X_i(k) - a_i) \right]^2}} \quad (2.4.46)$$

Podemos então observar que a equação 2.4.46 é um caso particular de 2.4.39 para $n = 2$



2.5. Noções de Álgebra Linear

Partimos nessa seção, do pressuposto que o leitor tenha conhecimento de funções, vetores, matrizes, determinantes e sistemas de equações lineares. Essa matéria pode ser vista em detalhes em [6], [7], [15] e [23]. Vamos aqui abordar somente o que será utilizado nesse texto.

SISTEMAS DE EQUAÇÕES E TRANSFORMAÇÕES LINEARES

Vamos partir do seguinte sistema de equações

$$\begin{aligned} \ell_{11}X_1 + \ell_{12}X_2 + \dots + \ell_{1n}X_n + \ell_{1(n+1)}X_{(n+1)} &= b_1 \\ \ell_{21}X_1 + \ell_{22}X_2 + \dots + \ell_{2n}X_n + \ell_{2(n+1)}X_{(n+1)} &= b_2 \\ &\vdots \\ \ell_{p1}X_1 + \ell_{p2}X_2 + \dots + \ell_{pn}X_n + \ell_{p(n+1)}X_{(n+1)} &= b_p \end{aligned} \tag{2.5.1}$$

onde os X_i são variáveis e os ℓ_{ji} e b_j são constantes.

Na forma de somatório teríamos:

$$\sum_{i=1}^{n+1} \ell_{ji}X_i = b_j \quad j = 1, 2, \dots, p \tag{2.5.2}$$

Seja

$$Y_j = b_j - X_{(n+1)} \ell_{j(n+1)} \quad j = 1, 2, \dots, p \tag{2.5.3}$$

O sistema de equações pode então ser escrito na seguinte forma:

$$\begin{aligned}
 Y_1 &= \ell_{11}X_1 + \ell_{12}X_2 + \dots + \ell_{1n}X_n \\
 Y_2 &= \ell_{21}X_1 + \ell_{22}X_2 + \dots + \ell_{2n}X_n \\
 &\vdots \\
 &\vdots \\
 Y_p &= \ell_{p1}X_1 + \ell_{p2}X_2 + \dots + \ell_{pn}X_n
 \end{aligned}
 \tag{2.5.4}$$

ou simplesmente

$$Y_j = \sum_{i=1}^n \ell_{ji} X_i \quad j = 1, 2, \dots, p
 \tag{2.5.5}$$

Representando esse sistema de equações na forma ma
tricial temos:

$$Y = \hat{L}X \quad \text{ou} \quad \vec{Y} = \hat{L}\vec{X}
 \tag{2.5.6}$$

ou seja

$$\vec{Y} = f(\vec{X})
 \tag{2.5.7}$$

onde

$$Y = \vec{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad \hat{L} = \begin{pmatrix} \ell_{11} & \ell_{12} & \dots & \ell_{1n} \\ \ell_{21} & & & \vdots \\ \vdots & & & \vdots \\ \ell_{p1} & \dots & \dots & \ell_{pn} \end{pmatrix} \quad \text{e} \quad X = \vec{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}$$

Podemos agora considerar a equação 2.5.6 não só co
mo um sistema de equações lineares, como também uma função
 $Y = f(X)$. Dizemos então, que essa função é uma TRANSFORMAÇÃO
LINEAR de X em Y e que \hat{L} é o OPERADOR (ou MATRIZ TRANSFORMA-
ÇÃO) que representa, de forma matricial, essa função. (Essa ma
téria pode ser vista com maiores detalhes em [7], seção 2.4).

Vejamos um exemplo bem simples.

EXEMPLO 2.5.1:

Seja a função:

$$f(X) = X_1 \quad \text{onde} \quad X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \quad (2.5.8)$$

Colocando na forma das equações 2.5.4 temos:

$$Y = f(X) = 1X_1 + 0X_2 \quad (2.5.9)$$

Para passarmos para a forma matricial é necessário notarmos que $X \in \mathbb{R}^2$ e $Y \in \mathbb{R}^1$, logo o operador \hat{L} terá uma linha e duas colunas.

Seja então:

$$\hat{L} = (\mathcal{L}_{11}; \mathcal{L}_{12}) \quad (2.5.10)$$

Nesse caso teríamos:

$$Y = (\mathcal{L}_{11}; \mathcal{L}_{12}) \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \mathcal{L}_{11}X_1 + \mathcal{L}_{12}X_2 \quad (2.5.11)$$

e pelas equações 2.5.9 e 2.5.11 facilmente concluimos que $\mathcal{L}_{11} = 1$ e $\mathcal{L}_{12} = 0$.

Assim teremos:

$$Y = (1; 0)X \quad (2.5.12)$$

A figura 2.5.1 representa graficamente a aplicação dessa transformação aos pontos $X(1)$, $X(2)$, $X(3)$ e $X(4)$. Nota mos que essa transformação é uma projeção ortogonal dos pontos $X(k) = (X_1(k); X_2(k))$ ($k = 1, 2, 3$) no eixo X_1 . Logo $f(X_1; X_2) = X_1$ é uma transformação linear de X em Y e $\hat{L} = (1; 0)$ é o operador que representa essa função.

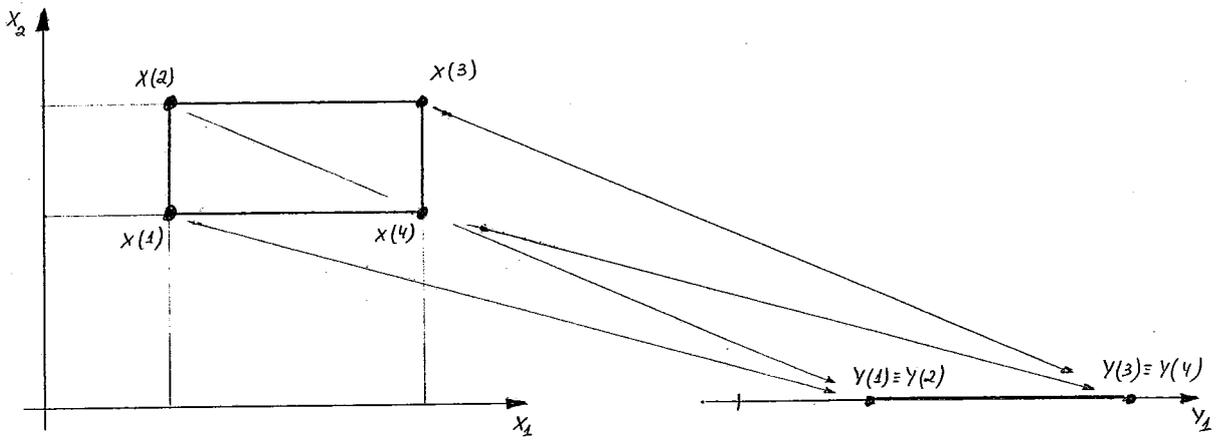


FIGURA 2.5.1: Projeção ortogonal de X em X_1

É bom observar que $Y \in \mathbb{R}^p$ e $X \in \mathbb{R}^n$. Assim a matriz \hat{L} representa uma função cujo domínio é um espaço vetorial n -dimensional (o \mathbb{R}^n) e seu contra-domínio é um espaço vetorial p -dimensional (o \mathbb{R}^p). (Ver definição de espaço vetorial em [7], seção 2.1 ou [23], cap.IV). No exemplo acima $n = 2$ e $p = 1$.

São de particular interesse as transformações lineares nas quais $n = p$. Vejamos um exemplo.

EXEMPLO 2.5.2:

Seja uma função $Y = f(X)$ tal que:

$$\begin{aligned} Y_1 &= a X_1 + 0 X_2 \\ Y_2 &= 0 X_1 + a X_2 \end{aligned} \quad a \neq 0 \quad (2.5.13)$$

Essa também é uma transformação linear que na forma matricial ficará:

$$Y = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} X \quad (2.5.14)$$

A figura 2.5.2 nos mostra a transformação aplicada aos pontos $X(1)$, $X(2)$, $X(3)$ e $X(4)$. Notamos que ela tem domi

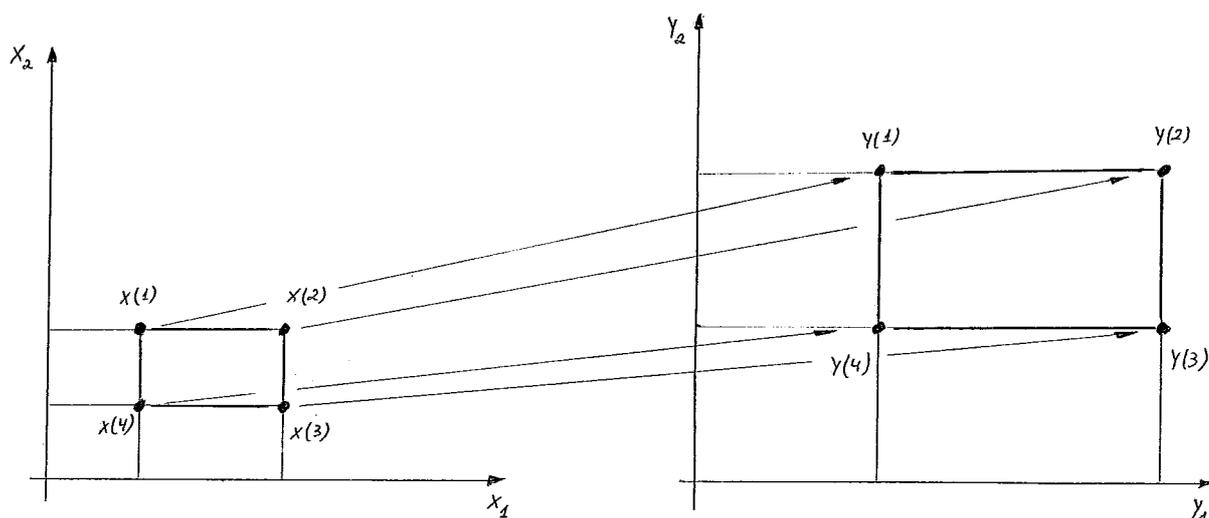


FIGURA 2.5.2: Exemplo de transformação linear.

nio e contra-domínio no mesmo espaço vetorial, ou seja $X \in \mathbb{R}^2$ e $Y \in \mathbb{R}^2$.

Observando os exemplos 2.5.1 e 2.5.2 surge uma pergunta. Dado que tenhamos o vetor Y , podemos ou não encontrar novamente o vetor X ? Nota-se que no exemplo 2.5.1 isso é impossível, o que já não acontece no exemplo 2.5.2, senão vejamos:

Da equação 2.5.13 temos:

$$Y_1 = a X_1 + 0 X_2 \quad (2.5.15)$$

$$Y_2 = 0 X_1 + a X_2$$

logo

$$X_1 = \frac{Y_1}{a} \quad \text{e} \quad X_2 = \frac{Y_2}{a} \quad (2.5.16)$$

ou

$$X_1 = \frac{1}{a} Y_1 + 0 Y_2$$

$$X_2 = 0 Y_1 + \frac{1}{a} Y_2 \quad (2.5.17)$$

assim

$$X = \begin{pmatrix} \frac{1}{a} & 0 \\ 0 & \frac{1}{a} \end{pmatrix} Y \quad (2.5.18)$$

ou seja, dado Y podemos novamente encontrar X . Assim o conceito de transformação inversa se torna importante na análise.

TRANSFORMAÇÕES LINEARES INVERSÍVEIS

Definição 2.5.1: Uma transformação linear $Y = \hat{L}X$ (ou $Y = f(X)$) é inversível se e somente se existe uma transformação $X = LY$ (ou $X = f^{-1}(Y)$).

Nesse caso dizemos que L é a matriz inversa de \hat{L} ou, equivalentemente, que \hat{L} é a matriz inversa de L .

Em outras palavras estamos dizendo que uma transformação linear é inversível, se e somente se existe uma transformação linear inversa, o que é o mesmo que dizer que existe uma matriz L , inversa a \hat{L} .

É importante observar que para que isso aconteça X e Y precisam ter a mesma dimensão ou seja $X \in \mathbb{R}^n$ e $Y \in \mathbb{R}^n$, senão vejamos. Vamos imaginar que $X \in \mathbb{R}^n$ e $Y \in \mathbb{R}^p$ $p < n$. Neste caso, Y não traz nenhuma informação sobre as $(n - p)$ dimensões que X possui e Y não, logo não poderíamos a partir de Y recuperar X através de uma transformação linear inversa já que a transformação não cria informação.

Isso nos leva a concluir que L e \hat{L} são matrizes

quadradas*.

Porém só o fato de \hat{L} ser uma matriz quadrada não nos garante a existência de L . Lembrando a teoria da Álgebra Linear (ver [6] [7] [15] e [23]) sabemos que para L existir é necessário e suficiente que:

- a) \hat{L} seja uma matriz quadrada
- b) determinante de \hat{L} seja diferente de zero**

Sob essas condições então podemos garantir a existência de L e dizer que L é a inversa de \hat{L} ou seja:

$$L = (\hat{L})^{-1} \quad (2.5.19)$$

o que equivale a

$$L^{-1} = \hat{L} \quad (2.5.20)$$

ou ainda a

$$L\hat{L} = \hat{L}L = I \quad (2.5.21)$$

onde I é a matriz identidade***

Essas afirmativas podem ser verificadas no exemplo 2.5.2. Nele tínhamos

$$Y = \hat{L}X \quad \text{onde} \quad \hat{L} = \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} \quad (2.5.22)$$

* Matriz quadrada é aquela cujo número de linhas é igual ao número de colunas.

** Essa condição equivale a dizer que \hat{L} é não-singular ou que o núcleo da transformação representada por \hat{L} tem dimensão nula. Para maiores esclarecimentos consultar [7], cap.II e [23], cap.VI.

*** Matriz identidade é uma matriz quadrada tal que os elementos da diagonal principal são iguais a 1 e os demais são nulos.

depois verificamos que:

$$X = LY \quad \text{onde} \quad L = \begin{pmatrix} \frac{1}{a} & 0 \\ 0 & \frac{1}{a} \end{pmatrix} \quad (2.5.23)$$

Podemos então notar que as equações 2.5.19, 2.5.20 e 2.5.21 se verificam.

Inicialmente falamos sobre transformações lineares. Depois focalizamos dentre essas transformações aquelas que são inversíveis. Vamos agora estudar dentre as transformações lineares inversíveis um tipo de transformação que nos será muito útil.

TRANSFORMAÇÕES ORTOGONAIS

Definição 2.5.2: Uma transformação linear representada pela matriz L é ortogonal se e somente se

$$\hat{L} = L^T$$

onde \hat{L} é a matriz inversa de L e L^T a sua transposta.

(as transformações ortogonais também são chamadas de AUTO-ORTOGONAIS, UNITÁRIAS ou AUTO-ADJUNTAS)

Observe que essa definição pressupõe várias condições, senão vejamos:

- a) L é uma matriz que representa uma transformação linear
- b) essa transformação é inversível (pois se não fosse L^{-1} não existiria)
- c) e finalmente $\hat{L} = L^T$

A primeira condição caracteriza a existência de

uma transformação linear. A segunda condição seleciona entre as transformações lineares aquelas que são inversíveis e finalmente a terceira seleciona aquelas que além de inversíveis, são ortogonais.

Dessa definição, juntamente com as equações 2.5.19, 2.5.20 e 2.5.21, decorrem uma série de conclusões que nos serão úteis. Senão vejamos.

Seja L a representação de uma transformação ortogonal (simplicadamente chamada de Matriz Ortogonal) logo

$$\hat{L} = L^T \quad (2.5.24)$$

transpondo as matrizes acima e observando que a transposta de uma matriz transposta é a própria matriz, temos:

$$\hat{L}^T = L \quad (2.5.25)$$

Levando a equação 2.5.24 na 2.5.21 temos:

$$L L^T = I \quad (2.5.26)$$

$$e \quad L^T L = I \quad (2.5.26A)$$

e levando a equação 2.5.25 na 2.5.21 temos:

$$\hat{L}^T \hat{L} = I \quad (2.5.27)$$

$$e \quad \hat{L} \hat{L}^T = I \quad (2.5.27A)$$

É bom lembrar que o fato de termos $\hat{L} = L^T$ não acarreta que $L^T = L$, ou seja, L não é necessariamente uma matriz simétrica.

Vamos agora ver como poderíamos representar a definição 2.5.2 em termos algébricos.

Seja:

$$L = \begin{pmatrix} l_{11} & \cdots & l_{1n} \\ \vdots & & \vdots \\ l_{n1} & \cdots & l_{nn} \end{pmatrix} \quad (2.5.28)$$

Pela equação 2.5.26 se L representa uma transformação ortogonal ela satisfaz a

$$\begin{pmatrix} l_{11} & \cdots & l_{1n} \\ \vdots & & \vdots \\ l_{n1} & \cdots & l_{nn} \end{pmatrix} \begin{pmatrix} l_{11} & \cdots & l_{n1} \\ \vdots & & \vdots \\ l_{1n} & \cdots & l_{nn} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & & \\ \vdots & & & \\ 0 & \cdots & & 1 \end{pmatrix} \quad (2.5.29)$$

Desmembrando essa equação teremos:

$$\left. \begin{array}{l} l_{11}l_{11} + l_{12}l_{12} + \cdots + l_{1n}l_{1n} = 1 \\ l_{11}l_{21} + l_{12}l_{22} + \cdots + l_{1n}l_{2n} = 0 \\ \vdots \\ l_{11}l_{n1} + l_{12}l_{n2} + \cdots + l_{1n}l_{nn} = 0 \end{array} \right\} \text{ n equações}$$

$$\left. \begin{array}{l} l_{21}l_{11} + l_{22}l_{12} + \cdots + l_{2n}l_{1n} = 0 \\ l_{21}l_{21} + l_{22}l_{22} + \cdots + l_{2n}l_{2n} = 1 \\ \vdots \\ l_{21}l_{n1} + l_{22}l_{n2} + \cdots + l_{2n}l_{nn} = 0 \end{array} \right\} \text{ n equações} \quad (2.5.30)$$

$$\left. \begin{array}{l} l_{n1}l_{11} + l_{n2}l_{12} + \cdots + l_{nn}l_{1n} = 0 \\ \vdots \\ l_{n1}l_{(n-1)1} + l_{n2}l_{(n-1)2} + \cdots + l_{nn}l_{(n-1)n} = 0 \\ l_{n1}l_{n1} + l_{n2}l_{n2} + \cdots + l_{nn}l_{nn} = 1 \end{array} \right\} \text{ n equações}$$

ou simplesmente

$$\sum_{j=1}^n l_{ij} l_{i'j} = \begin{cases} 1 & i = i' \\ 0 & i \neq i' \end{cases} \quad \begin{matrix} i = 1, 2, \dots, n \\ i' = 1, 2, \dots, n \end{matrix} \quad (2.5.31)$$

Analogamente se desenvolvêssemos as equações

2.5.26A, 2.5.27 e 2.5.27A chegaríamos a:

$$\sum_{i=1}^n \ell_{ij} \ell_{ij'} = 1 \quad j = j' \quad j = 1, 2, \dots, n \quad (2.5.32)$$

$$= 0 \quad j \neq j' \quad j' = 1, 2, \dots, n$$

$$\sum_{i=1}^n \ell_{ji} \ell_{j'i} = 1 \quad j = j' \quad j = 1, 2, \dots, n \quad (2.5.33)$$

$$= 0 \quad j \neq j' \quad j' = 1, 2, \dots, n$$

$$\sum_{j=1}^n \ell_{ji} \ell_{ji'} = 1 \quad i = i' \quad i = 1, 2, \dots, n \quad (2.5.34)$$

$$= 0 \quad i \neq i' \quad i' = 1, 2, \dots, n$$

Podemos então concluir que dizer que L representa uma transformação ortogonal é o mesmo que afirmar que cada uma das equações 2.5.24-27A e 2.5.31-34 ocorre.

É importante observar que o sistema de equações 2.5.31 (e analogamente 2.5.32-34) não é linearmente independente, o que é lógico, já que se ele o fosse os ℓ_{ij} estariam definidos, o que é absurdo.

Vamos agora estudar algumas propriedades das transformações ortogonais.

TEOREMA 2.5.1: O rank da matriz L é igual a n .

Demonstração:

Pela inequação de Sylvester (ver [7], pág.33) temos:

$$\text{rank}(L L^T) \leq \min(\text{rank}(L), \text{rank}(L^T)) \quad (2.5.35)$$

mas da equação 2.5.26 temos que:

$$\text{rank}(L L^T) = \text{rank}(I) = n \quad (2.5.36)$$

e por outro lado sabemos que:

$$\text{rank}(L) = \text{rank}(L^T) \quad (2.5.37)$$

Levando esses resultados na equação 2.5.35 chegamos a:

$$n \leq \text{rank}(L) \quad (2.5.38)$$

Mas, por definição, $\text{rank}(L) \leq n$ o que acarreta que:

$$\text{rank}(L) = n$$

△

Esse teorema nos leva a concluir que o núcleo de L tem dimensão zero, (ver definição de núcleo, ou "null space", em [7], seção 2.4) o que acarreta que em $Y = LX$ podemos dizer, sem rigor matemático, que "Y contém a mesma quantidade de informação que X", aliás, característica de toda transformação inversível.

TEOREMA 2.5.2: A transformação inversa a uma transformação ortogonal também é ortogonal.

Demonstração:

Seja

$$X = LY \quad (2.5.39)$$

onde L representa uma transformação ortogonal. Pela definição 2.5.2 temos:

$$L^{-1} = L^T \quad (2.5.40)$$

e como toda transformação ortogonal é inversível

$$Y = L^{-1} X \quad (2.5.41)$$

Pela equação 2.5.40 concluímos que:

$$(L^{-1})^T = (L^T)^T = L \quad (2.5.42)$$

mas

$$(L^{-1})^{-1} = L \quad (2.5.43)$$

Reunindo as equações 2.5.42 e 2.5.43 temos:

$$(L^{-1})^{-1} = L(-1)^T \quad (2.5.44)$$

O que nos leva a concluir que L^{-1} também representa uma transformação ortogonal. \triangle

Com esse resultado, se estivermos de posse de X podemos obter Y e obter novamente X , utilizando apenas transformações ortogonais.

Outra característica desse tipo de transformação é que ela não altera a distância entre dois pontos quaisquer, senão vejamos.

TEOREMA 2.5.3: Seja $Y = \hat{L} X$ onde \hat{L} representa uma transformação ortogonal. Seja $Y(k) = \hat{L} X(k)$ e $Y(k') = \hat{L} X(k')$ onde $X(k)$ e $X(k')$ são pontos pertencentes ao R^n . Então $\text{dist}(X(k), X(k')) = \text{dist}(Y(k), Y(k'))$

Demonstração:

Por definição (ver equação 2.4.1)

$$\text{dist}(X(k); X(k')) = \sqrt{\sum_{i=1}^n (X_i(k) - X_i(k'))^2} \quad (2.5.45)$$

$$\text{e } \text{dist}(Y(k); Y(k')) = \sqrt{\sum_{j=1}^n (Y_j(k) - Y_j(k'))^2} \quad (2.5.46)$$

Mas como $Y(k) = \hat{L} X(k)$ e $Y(k') = \hat{L} X(k')$ temos:

$$Y(k) = \begin{pmatrix} Y_1(k) \\ Y_2(k) \\ \vdots \\ Y_n(k) \end{pmatrix} = \begin{pmatrix} \ell_{11} & \cdots & \ell_{1n} \\ \vdots & & \vdots \\ \ell_{n1} & \cdots & \ell_{nn} \end{pmatrix} \begin{pmatrix} X_1(k) \\ X_2(k) \\ \vdots \\ X_n(k) \end{pmatrix} \quad (2.5.47)$$

$$\text{ou seja } Y_j(k) = \sum_{i=1}^n \ell_{ji} X_i(k) \quad (2.5.48)$$

e analogamente

$$Y_j(k') = \sum_{i=1}^n \ell_{ji} X_i(k') \quad (2.5.49)$$

Substituindo as equações 2.5.48 e 2.5.49 na 2.5.46

teremos:

$$\text{dist}(Y(k); Y(k')) = \sqrt{\sum_{j=1}^n \left(\sum_{i=1}^n \ell_{ji} X_i(k) - \sum_{i=1}^n \ell_{ji} X_i(k') \right)^2} \quad (2.5.50)$$

Utilizando as propriedades dos somatórios podemos escrever que

$$\text{dist}(Y(k); Y(k')) = \sqrt{\sum_{j=1}^n \left(\sum_{i=1}^n (\ell_{ji} X_i(k) - \ell_{ji} X_i(k')) \right)^2} \quad (2.5.51)$$

ou

$$\text{dist}(Y(k); Y(k')) = \sqrt{\sum_{j=1}^n \left(\sum_{i=1}^n \ell_{ji} (X_i(k) - X_i(k')) \right)^2} \quad (2.5.52)$$

Usando a equação 2.2.10 podemos escrever que:

$$\text{dist}(Y(k); Y(k')) = \sqrt{\sum_{j=1}^n \left\{ \sum_{i=1}^n \sum_{i'=1}^n \ell_{ji} \ell_{ji'} [(X_i(k) - X_i(k')) (X_{i'}(k) - X_{i'}(k'))] \right\}} \quad (2.5.53)$$

$$= \sqrt{\sum_{i=1}^n \sum_{i'=1}^n \left\{ \sum_{j=1}^n \ell_{ji} \ell_{ji'} [(X_i(k) - X_i(k')) (X_{i'}(k) - X_{i'}(k'))] \right\}} \quad (2.5.54)$$

$$= \sqrt{\sum_{i=1}^n \sum_{i'=1}^n \left\{ [(X_i(k) - X_i(k')) (X_{i'}(k) - X_{i'}(k'))] \sum_{j=1}^n \ell_{ji} \ell_{ji'} \right\}} \quad (2.5.55)$$

Mas pela equação 2.5.34 temos:

$$\sum_{j=1}^n \ell_{ji} \ell_{ji'} = 1 \quad i = i' \quad (2.5.56)$$

$$= 0 \quad i \neq i'$$

o que nos leva a

$$\text{dist}(Y(k); Y(k')) = \sqrt{\sum_{i=1}^n [X_i(k) - X_i(k')]^2} \quad (2.5.57)$$

Comparando as equações 2.5.45 e 2.5.57 concluímos que:

$$\text{dist}(Y(k); Y(k')) = \text{dist}(X(k); X(k')) \quad (2.5.58)$$

△

Como acabamos de ver, uma transformação ortogonal não altera a distância entre dois pontos quaisquer. É portanto uma transformação ISOMÉTRICA. Se isso é verdade, em particular a distância de cada ponto à origem se mantém constante.

Podemos notar também que a origem se mantém inalterada sob essa transformação, ou seja, se

$$Y = \hat{L} X, \quad \hat{L}^T = L^{-1} \quad \text{e} \quad X = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

então

$$Y = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Logo como a origem não se altera com uma transformação ortogonal e as distâncias de pontos a essa origem também não se alteram, surge a idéia da seguinte definição.

Definição 2.5.3: Denominamos de ROTAÇÃO de um vetor $X \in \mathbb{R}^n$ em torno da origem, uma transformação ortogonal $Y = \hat{L} X$ (ver

[³], pág.173).

Para ilustrar essa definição vamos ver um exemplo no \mathbb{R}^2 .

EXEMPLO 2.5.3:

Seja uma transformação ortogonal representada pela matriz

$$\hat{L} = \begin{pmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix} \quad (2.5.59)$$

Podemos notar que de fato \hat{L} representa uma transformação ortogonal já que a equação 2.5.26 se verifica, ou seja:

$$\begin{pmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix} \begin{pmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad (2.5.60)$$

Seja:

$$X(1) = \begin{pmatrix} 3 \\ 1 \end{pmatrix} ; \quad X(2) = \begin{pmatrix} 7 \\ 1 \end{pmatrix} \quad e \quad X(3) = \begin{pmatrix} 7 \\ 3 \end{pmatrix} \quad (2.5.61)$$

Assim teremos:

$$Y(1) = \begin{pmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix} \begin{pmatrix} 3 \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{3\sqrt{3}+1}{2} \\ \frac{-3+\sqrt{3}}{2} \end{pmatrix} \approx \begin{pmatrix} 3,098 \\ -0,634 \end{pmatrix}$$

$$\begin{aligned}
 Y(2) &= \begin{pmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix} \begin{pmatrix} 7 \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{7\sqrt{3}+1}{2} \\ \frac{-7+\sqrt{3}}{2} \end{pmatrix} \approx \begin{pmatrix} 6,562 \\ -2,634 \end{pmatrix} \\
 Y(3) &= \begin{pmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix} \begin{pmatrix} 7 \\ 3 \end{pmatrix} = \begin{pmatrix} \frac{7\sqrt{3}+3}{2} \\ \frac{-7+3\sqrt{3}}{2} \end{pmatrix} \approx \begin{pmatrix} 7,562 \\ -0,901 \end{pmatrix}
 \end{aligned}$$

(2.5.62)

Essa transformação pode ser vista na figura 2.5.3.

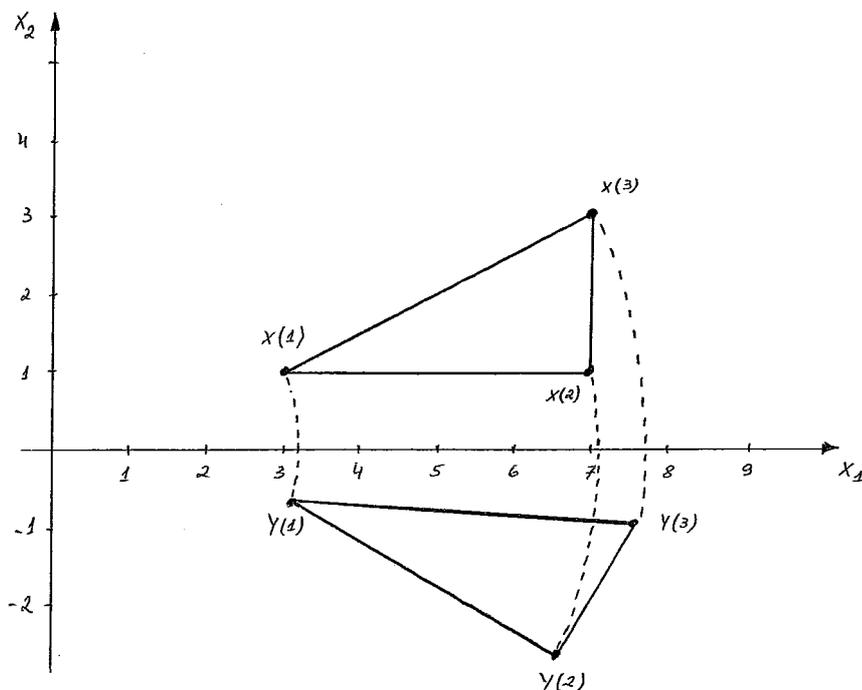


FIGURA 2.5.3: Representação gráfica da transformação ortogonal representada por \hat{L} .

Observe que as distâncias dos pontos $X(1)$, $X(2)$ e $X(3)$ à origem são as mesmas de $Y(1)$, $Y(2)$ e $Y(3)$ respectivamente.

Quando efetuamos uma rotação podemos focaliza-la de duas formas:

ENFOQUE A: Mantemos fixa a base do sistema de coordenadas (no caso X_1, X_2, \dots, X_n) e rotacionamos os pontos do R^n (no caso os $X(k)$).

ENFOQUE B: Mantemos fixos os pontos do R^n (no caso os $X(k)$) e rotacionamos a base do sistema de coordenadas (no caso X_1, X_2, \dots, X_n).

Assim, no exemplo 2.5.3 poderíamos ter a figura 2.5.4, ao invés da 2.5.3

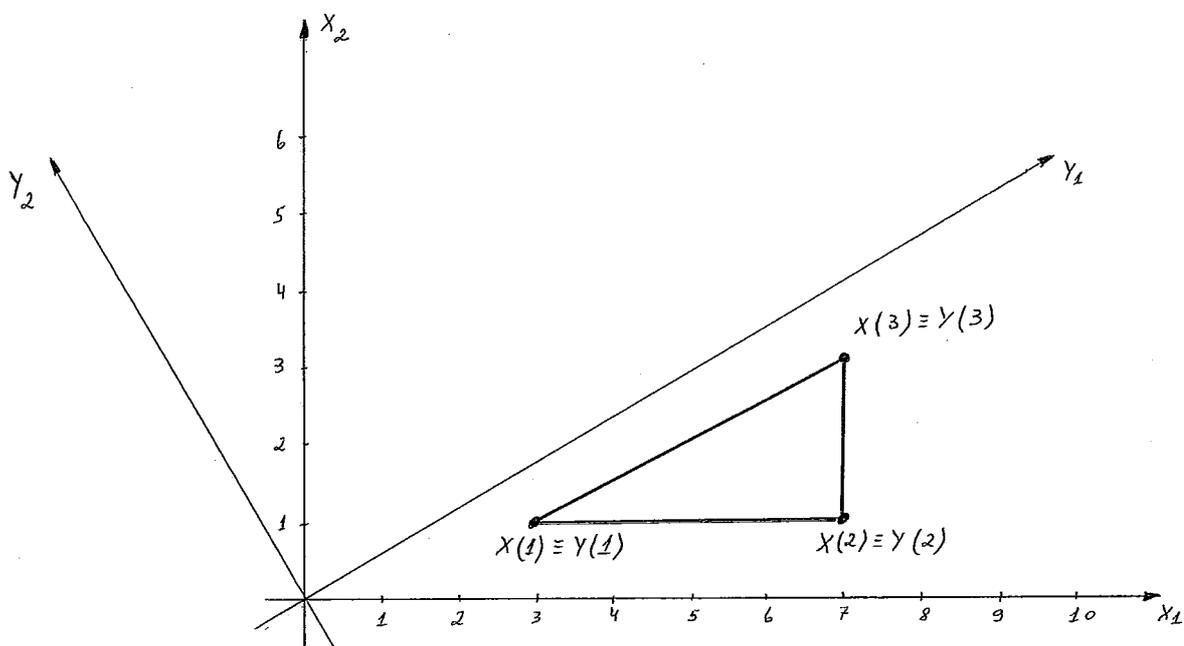


FIGURA 2.5.4: Representação gráfica da transformação ortogonal representada por \hat{L} .

Observe que a posição dos pontos $X(k) \equiv Y(k)$ ($k = 1, 2, 3$) em relação aos eixos Y_1 e Y_2 na figura 2.5.4 é a mesma que as dos pontos $Y(k)$ ($k = 1, 2, 3$) em relação aos eixos

X_1 e X_2 na figura 2.5.3.

Se utilizamos o ENFOQUE B, poderemos concluir que as transformações ortogonais nos permitem observar os pontos $X(k)$ segundo um outro ângulo, sem alterar a distância entre eles. Isso é muito importante pois assim podemos analisar as características dos elementos da população segundo um novo ângulo, sem alterar essas características medidas. Escolhendo uma rotação apropriada, poderemos facilitar a análise que se deseja fazer com os pontos $X(k)$.

É importante observar que uma rotação no R^n nos casos de $n = 2$ e 3 pode não ser exatamente o que chamamos de "rotação" intuitivamente. Senão vejamos um exemplo:

EXEMPLO 2.5.4:

Seja uma transformação ortogonal representada pela matriz

$$\hat{L} = \begin{pmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{\sqrt{3}}{2} \end{pmatrix}$$

Procedendo de forma análoga a do exemplo 2.5.3 podemos comprovar que de fato \hat{L} representa uma transformação ortogonal.

Seja:

$$X(1) = \begin{pmatrix} 3 \\ 1 \end{pmatrix} ; \quad X(2) = \begin{pmatrix} 7 \\ 1 \end{pmatrix} \quad \text{e} \quad X(3) = \begin{pmatrix} 7 \\ 3 \end{pmatrix} \quad (2.5.63)$$

Assim teremos:

$$Y(1) = \begin{pmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{\sqrt{3}}{2} \end{pmatrix} \begin{pmatrix} 3 \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{3\sqrt{3}+1}{2} \\ \frac{3-\sqrt{3}}{2} \end{pmatrix} \approx \begin{pmatrix} 3,098 \\ 0,634 \end{pmatrix}$$

$$Y(2) = \begin{pmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{\sqrt{3}}{2} \end{pmatrix} \begin{pmatrix} 7 \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{7\sqrt{3}+1}{2} \\ \frac{7-\sqrt{3}}{2} \end{pmatrix} \approx \begin{pmatrix} 6,562 \\ 2,634 \end{pmatrix}$$

$$Y(3) = \begin{pmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{\sqrt{3}}{2} \end{pmatrix} \begin{pmatrix} 7 \\ 3 \end{pmatrix} = \begin{pmatrix} \frac{7\sqrt{3}+3}{2} \\ \frac{7-3\sqrt{3}}{2} \end{pmatrix} \approx \begin{pmatrix} 7,562 \\ 0,901 \end{pmatrix}$$

(2.5.64) ▲

Comparando $Y(1)$, $Y(2)$ e $Y(3)$ da equação 2.5.64 com os da equação 2.5.62 observamos que as primeiras coordenadas de cada ponto são iguais, o que não acontece com as segundas coordenadas por causa do sinal. Assim, a menos do sinal, uma rotação no R^n em torno da origem para $n = 2$ é o que chamamos intuitivamente de "rotação". Por outro lado, essa troca de sinal verificada nesse exemplo nada mais é do que uma reflexão em relação ao eixo X_1 , como podemos ver na figura 2.5.5. Logo uma rotação no R^n em torno da origem para $n=2$, intuitivamente falando, poderá ser uma "rotação" (exemplo 2.5.3) ou uma "rotação" seguida de uma reflexão (exemplo 2.5.4) em torno de um eixo.

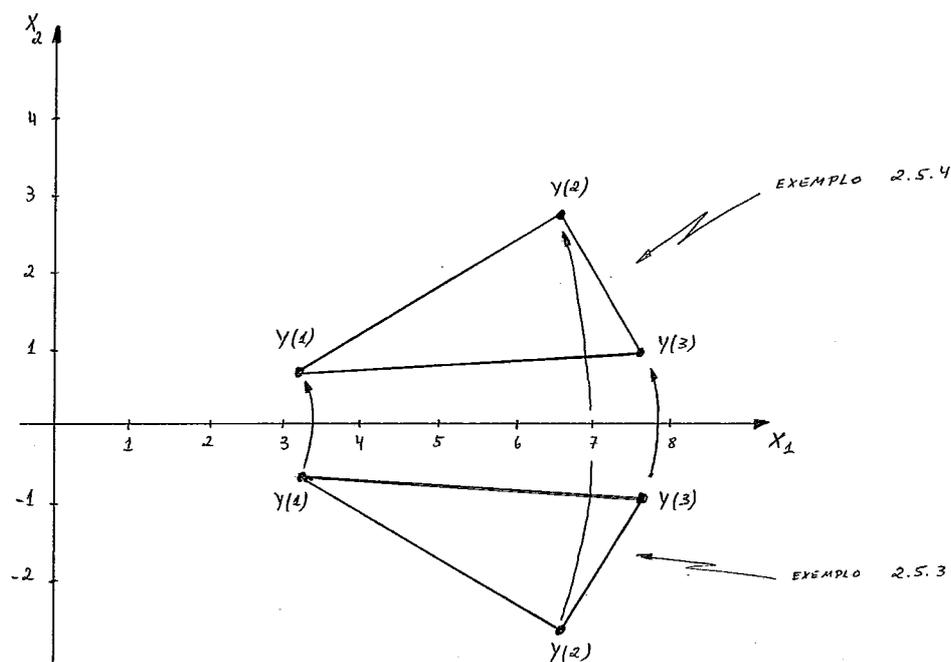


FIGURA 2.5.5: Reflexão em torno de X_1

Apresentamos a seguir outra característica das transformações ortogonais.

TEOREMA 2.5.4: Se $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n$ formam uma base ortonormal* para o \mathbb{R}^n e \hat{L} uma matriz que representa uma transformação ortogonal tal que $\vec{y}_j = \hat{L} \vec{x}_i$ ($i = j = 1, 2, \dots, n$), então $\vec{y}_1, \vec{y}_2, \dots, \vec{y}_n$ também formam uma base ortonormal para o \mathbb{R}^n .

Demonstração:

Consultar ref. [6], pág.160.



Esse teorema nos mostra que se $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n$ formam uma base ortonormal para o \mathbb{R}^n e efetuamos uma rotação, representada pela matriz \hat{L} , obteremos uma nova base $\vec{y}_1, \vec{y}_2, \dots, \vec{y}_n$ que também será ortonormal.

* base ortonormal é aquela constituída por vetores unitários e ortogonais ([15], seção 8.2)

Sob o ponto de vista geométrico ele nos mostra que se utilizarmos um sistema de coordenadas ortogonais cartesianas para representar o R^n , quando efetuarmos uma rotação nesse sistema de coordenadas o novo sistema que surgirá também será de coordenadas ortogonais cartesianas.

Observe que estamos utilizando o enfoque \underline{B} , já que o objetivo é observar os pontos $X(k)$ de uma outra forma, e não alterá-los. Logicamente, quando alterarmos o sistema de coordenadas, os pontos passam a ter novas coordenadas em relação ao novo sistema. Assim, cada ponto $X(k)$ de coordenada $X_1(k), X_2(k), \dots, X_n(k)$ em um sistema, terá coordenadas $Y_1(k), Y_2(k), \dots, Y_n(k)$ no outro sistema. Para identificar o sistema em uso utilizaremos $X(k)$ para um e $Y(k)$ para outro, não se esquecendo que, segundo o enfoque \underline{B} , $X(k)$ e $Y(k)$ representam o mesmo ponto só que com coordenadas diferentes.

2.6. Resumo

Com o objetivo de facilitar as consultas vamos aqui reunir os resultados das seções 2.2, 2.3, 2.4 e 2.5 .

A) SOMATÓRIOS (Seção 2.2)

$$i) \sum_{i=1}^n \sum_{j=1}^p \ell_{ij} Y_j = \sum_{j=1}^p \sum_{i=1}^n \ell_{ij} Y_j \quad (2.2.4)$$

$$ii) \left(\sum_{j=1}^p \ell_{ij} Y_j \right) \left(\sum_{j'=1}^p \ell_{i'j'} Y_{j'} \right) = \sum_{j=1}^p \sum_{j'=1}^p \ell_{ij} \ell_{i'j'} Y_j Y_{j'} \quad (2.2.9)$$

$$V_i = 1, 2, \dots, n$$

$$V_{i'} = 1, 2, \dots, n$$

$$\text{iii) } \left(\sum_{j=1}^p \ell_{ij} Y_j \right)^2 = \sum_{j=1}^p \sum_{j'=1}^p \ell_{ij} \ell_{ij'} Y_j Y_{j'} \quad (2.2.10)$$

$$\forall i = 1, 2, \dots, n$$

$$\text{iv) Dado } X_i = \sum_{j=1}^p \ell_{ij} Y_j \text{ então } \frac{\delta X_i}{\delta \ell_{ij}} = Y_j \quad \begin{matrix} \forall i=1, 2, \dots, n \\ \forall j=1, 2, \dots, p \end{matrix} \quad (2.2.13)$$

$$\text{v) } \frac{\delta \left[\left(\sum_{j'=1}^p \ell_{ij'} Y_{j'} \right)^2 \right]}{\delta \ell_{ij}} = 2 \sum_{j'=1}^p \ell_{ij'} Y_j Y_{j'} \quad (2.2.19)$$

$$\begin{matrix} \forall i=1, 2, \dots, n \\ \forall j=1, 2, \dots, p \end{matrix}$$

B) ESTATÍSTICA (Seção 2.3)

i) Média ou Valor Esperado

$$\text{Definição: } E[X_i] = \bar{X}_i = \frac{1}{N} \sum_{k=1}^N X_i(k) \quad (2.3.1)$$

$$\text{Propriedades: } \rightarrow \text{ se } X_i = \text{cte} \text{ então } E[X_i] = X_i = \text{cte} \quad (2.3.2)$$

$$\rightarrow E[aX_i] = aE[X_i] \quad (a=\text{cte}) \quad (2.3.3)$$

$$\rightarrow E\left[\sum_{i=1}^n X_i \right] = \sum_{i=1}^n E[X_i] \quad (2.3.6)$$

$$\rightarrow \text{ Se } H(X_1; X_2, \dots; X_n) = a_1 X_1 + a_2 X_2 + \dots$$

$$\dots + a_n X_n + b \text{ onde } a_1; a_2; \dots; a_n \text{ e } b$$

são constantes então

$$E[H(X_1; X_2; \dots; X_n)] = H(E[X_1]; E[X_2]; \dots; E[X_n]) \quad (2.3.9)$$

ii) Variância

Definição: $\sigma_{X_i}^2 = \text{VAR}[X_i] = c_{ii} = E[(X_i - \bar{X}_i)^2] =$

$$= \frac{1}{N} \sum_{k=1}^N (X_i(k) - \bar{X}_i)^2 \quad (2.3.12)$$

Propriedades: \rightarrow se $X_i = \text{cte}$ então $\text{VAR}[X_i] = 0 \quad (2.3.13)$

$$\rightarrow \text{VAR}[X_i + a] = \text{VAR}[X_i] \quad (2.3.14)$$

$$\rightarrow \text{VAR}[aX_i] = a^2 \text{VAR}[X_i] \quad (2.3.15)$$

$$\rightarrow \text{VAR}[X_i + X_{i'}] = \text{VAR}[X_i] + \text{VAR}[X_{i'}] +$$

$$+ 2E[(X_i - \bar{X}_i)(X_{i'} - \bar{X}_{i'})] \quad (2.3.17)$$

\rightarrow se as variáveis X_i não forem correlacionadas entre si então

$$\text{VAR}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \text{VAR}[X_i] \quad (2.3.20)$$

Teorema 2.3.1: $\text{VAR}[X_i] = E[X_i^2] - (E[X_i])^2$

iii) Desvio Médio Padrão

Definição: $\sigma_{X_i} = \sqrt{\sigma_{X_i}^2} = \sqrt{\frac{\sum_{k=1}^N (X_i(k) - \bar{X}_i)^2}{N}} \quad (2.3.27)$

Propriedades: \rightarrow se $X_i = \text{cte}$ então $\sigma_{X_i} = 0 \quad (2.3.28)$

$$\rightarrow \sigma[X_i + a] = \sigma_{X_i} \quad (2.3.29)$$

$$\rightarrow \sigma[aX_i] = |a| \sigma_{X_i} \quad (2.3.30)$$

iv) Covariância

Definição: $COV[X_i; X_{i'}] = c_{ii'} = E[(X_i - \bar{X}_i)(X_{i'} - \bar{X}_{i'})] =$

$$= \frac{1}{N} \sum_{k=1}^N [(X_i(k) - \bar{X}_i)(X_{i'}(k) - \bar{X}_{i'})] \quad (2.3.34)$$

$i \neq i'$

Teorema 2.3.2: $COV[X_i; X_{i'}] = E[X_i X_{i'}] - E[X_i]E[X_{i'}]$

Propriedades: $\rightarrow c_{ii} = COV[X_i; X_i] = VAR[X_i]$ (2.3.42)

$\rightarrow c_{ii'} = c_{i'i}$ (2.3.43)

v) Correlação

Definição: $COR[X_i; X_{i'}] = \frac{COV[X_i; X_{i'}]}{\sigma_{X_i} \sigma_{X_{i'}}} =$

$$= \frac{\sum_{k=1}^N [(X_i(k) - \bar{X}_i)(X_{i'}(k) - \bar{X}_{i'})]}{\sqrt{\sum_{k=1}^N (X_i(k) - \bar{X}_i)^2} \sqrt{\sum_{k=1}^N (X_{i'}(k) - \bar{X}_{i'})^2}}$$

(2.3.46)

Propriedades: $\rightarrow -1 \leq COR[X_i; X_{i'}] \leq 1$ (2.3.47)

$\rightarrow COR[X_i; X_{i'}] = COR[X_{i'}; X_i]$ (2.3.48)

Teorema 2.3.3: Se $X_i = aX_{i'} + b$ (a e b constantes)
então

$$|COR[X_i; X_{i'}]| = 1$$

vi) Desvios

$$\text{Definição: } x_i = X_i - \bar{X}_i \quad (2.3.56)$$

$$\text{Propriedades: } \rightarrow E[x_i] = 0 \quad (2.3.57)$$

$$\rightarrow \text{VAR}[x_i] = \text{VAR}[X_i] = c_{ii} \quad (2.3.58)$$

$$\rightarrow \sigma_{x_i} = \sigma_{X_i} \quad (2.3.59)$$

$$\rightarrow \text{COV}[x_i; x_{i'}] = E[x_i x_{i'}] \quad (2.3.59A)$$

$$\rightarrow \text{COV}[x_i; x_{i'}] = \text{COV}[X_i; X_{i'}] \quad (2.3.60)$$

$$\rightarrow \text{COR}[x_i; x_{i'}] = \text{COR}[X_i; X_{i'}] \quad (2.3.61)$$

vii) Variáveis Padronizadas

$$\text{Definição: } z_i = \frac{x_i}{\sigma_{X_i}} = \frac{X_i - \bar{X}_i}{\sigma_{X_i}} \quad (2.3.64)$$

$$\text{Propriedades: } \rightarrow E[z_i] = 0 \quad (2.3.65)$$

$$\rightarrow \text{VAR}[z_i] = 1 \quad (2.3.66)$$

$$\rightarrow \sigma_{z_i} = 1 \quad (2.3.67)$$

$$\begin{aligned} \rightarrow \text{COR}[z_i; z_{i'}] &= \text{COV}[z_i; z_{i'}] = \\ &= E[z_i z_{i'}] \end{aligned} \quad (2.3.68)$$

$$\rightarrow \text{COR}[z_i; z_{i'}] = \text{COR}[X_i; X_{i'}] \quad (2.3.69)$$

viii) Independência de Variáveis Aleatórias

As seguintes afirmativas são equivalentes:

a) X_1 e X_2 são variáveis aleatórias discretas independentes.

$$b) P(X_1 = a | X_2 = b) = P(X_1 = a) \quad (2.3.70)$$

$$c) P(X_1 = a \text{ e } X_2 = b) = P(X_1 = a)P(X_2 = b) \quad (2.3.72)$$

C) GEOMETRIA NO \mathbb{R}^n (Seção 2.4)

i) Distância entre dois pontos no \mathbb{R}^n

$$\text{Definição: } \text{dist}(X(k); X(k')) = \sqrt{\sum_{i=1}^n (X_i(k) - X_i(k'))^2} \quad (2.4.1)$$

$$\text{Propriedades: } \rightarrow \text{dist}(X(k); X(k)) = 0 \quad (2.4.2)$$

\rightarrow se $X(k) \neq X(k')$ então

$$\text{dist}(X(k); X(k')) > 0 \quad (2.4.3)$$

$\rightarrow \text{dist}(X(k); X(k')) = \text{dist}(X(k'); X(k))$

$$(2.4.4)$$

$\rightarrow \text{dist}(X(k); X(k')) + \text{dist}(X(k'); X(k'')) \geq$

$$\geq \text{dist}(X(k); X(k'')) \quad (2.4.5)$$

ii) Norma de um vetor $\vec{X}(k)$

Definição: $|\vec{X}(k)| = \text{dist}(X(k); \text{origem}) =$

$$= \sqrt{\sum_{i=1}^n X_i(k)^2} \quad (2.4.6)$$

iii) Cossenos diretores de um vetor $\vec{X}(k)$

$$\text{Definição: } g_i = \cos \alpha_i = \frac{\langle \vec{X}(k), \vec{X}_i \rangle}{|\vec{X}(k)|} \quad (2.4.7)$$

$i=1, 2, \dots, n$

$$\text{Teorema 2.4.1: } \sum_{i=1}^n g_i^2 = 1$$

iv) Retas Ortogonais

Teorema 2.4.2: Duas retas no \mathbb{R}^n de cossenos diretores

$g_1 = (g_{11}, g_{12}, \dots, g_{1n})$ e $g_2 = (g_{21}, g_{22}, \dots, g_{2n})$ são ortogonais se e somente

$$\text{se } \sum_{i=1}^n g_{1i} g_{2i} = 0$$

v) Distância de um ponto à uma reta no \mathbb{R}^n

$$\text{dist}(X(k); \text{reta } AQ) = \sqrt{\sum_{i=1}^n (X_i(k) - a_i)^2 - \left[\sum_{i=1}^n g_i (X_i(k) - a_i) \right]^2}$$

(2.4.39)

D) ÁLGEBRA LINEAR (Seção 2.5)i) Transformações Lineares

$$Y = \hat{L} X \quad \text{ou} \quad \vec{Y} = \hat{L} \vec{X} \quad (2.5.6)$$

onde a matriz \hat{L} é um operador linear.

ii) Transformações Lineares Inversíveis

Definição: Uma transformação linear $Y = \hat{L} X$ (ou $Y = f(X)$) é inversível se e somente se existe uma transformação $X = L Y$ (ou $X = f^{-1}(Y)$).

Condições necessárias e suficientes para a existência de L:

- a) \hat{L} ser uma matriz quadrada.
- b) determinante de \hat{L} ser diferente de zero.

iii) Transformações Ortogonais

Definição: Uma transformação linear representada pela matriz L é ortogonal se e somente se

$$\hat{L} = L^T$$

onde \hat{L} é a matriz inversa de L e L^T a sua transposta.

Afirmações equivalentes:

→ L representa uma transformação ortogonal

→ \hat{L} representa uma transformação ortogonal

$$\rightarrow \hat{L} = L^T \quad (2.5.24)$$

$$\rightarrow \hat{L}^T = L \quad (2.5.25)$$

$$\rightarrow L L^T = I \quad (2.5.26)$$

$$\rightarrow L^T L = I \quad (2.5.26A)$$

$$\rightarrow \hat{L}^T \hat{L} = I \quad (2.5.27)$$

$$\rightarrow \hat{L} \hat{L}^T = I \quad (2.5.27A)$$

$$\rightarrow \sum_{j=1}^n \ell_{ij} \ell_{i'j} = \begin{cases} 1 & i = i' \\ 0 & i \neq i' \end{cases} \quad \begin{matrix} i=1,2,\dots,n \\ i'=1,2,\dots,n \end{matrix} \quad (2.5.31)$$

$$\rightarrow \sum_{i=1}^n \ell_{ij} \ell_{ij'} = \begin{cases} 1 & j = j' \\ 0 & j \neq j' \end{cases} \quad \begin{matrix} j=1,2,\dots,n \\ j'=1,2,\dots,n \end{matrix} \quad (2.5.32)$$

$$\rightarrow \sum_{i=1}^n \hat{\ell}_{ji} \hat{\ell}_{j'i} = \begin{cases} 1 & j = j' \\ 0 & j \neq j' \end{cases} \quad \begin{matrix} j=1,2,\dots,n \\ j'=1,2,\dots,n \end{matrix} \quad (2.5.33)$$

$$\rightarrow \sum_{j=1}^n \hat{\ell}_{ji} \hat{\ell}_{ji'} = \begin{cases} 1 & i = i' \\ 0 & i \neq i' \end{cases} \quad \begin{matrix} i=1,2,\dots,n \\ i'=1,2,\dots,n \end{matrix} \quad (2.5.34)$$

Teorema 2.5.1: O rank da matriz L é igual a n .

Teorema 2.5.2: A transformação inversa a uma transformação ortogonal também é ortogonal.

Teorema 2.5.3: Seja $Y = \hat{L} X$ onde \hat{L} representa uma transformação ortogonal. Seja $Y(k) = \hat{L} X(k)$ e $Y(k') = \hat{L} X(k')$ onde $X(k)$ e $X(k')$ são pontos pertencentes ao \mathbb{R}^n . Então $\text{dist}(X(k); X(k')) = \text{dist}(Y(k); Y(k'))$.

Enfoques dados à uma transformação ortogonal:

ENFOQUE A: mantem-se fixa a base do sistema de coordenadas e rotaciona-se os pontos do \mathbb{R}^n .

ENFOQUE B: mantem-se fixos os pontos do \mathbb{R}^n e rotaciona-se a base do sistema de coordenadas.

Teorema 2.5.4: Se $\vec{X}_1, \vec{X}_2, \dots, \vec{X}_n$ formam uma base ortonormal para o \mathbb{R}^n e \hat{L} é uma matriz que representa uma transformação ortogonal tal que $\vec{Y}_j = \hat{L} \vec{X}_i$ ($i=j=1, 2, \dots, n$), então $\vec{Y}_1, \vec{Y}_2, \dots, \vec{Y}_n$ também formam uma base ortonormal para o \mathbb{R}^n .

CAPÍTULO III
COMPONENTES PRINCIPAIS

3.1. Introdução

O Modelo das Componentes Principais (MCP), utilizado na Análise Fatorial, trabalha com uma transformação do tipo representado pela equação 1.2.1 fazendo, no entanto, uma série de restrições adicionais acerca desta transformação. No decorrer deste capítulo examinaremos com detalhe o significado destas restrições.

Sejam X_1, X_2, \dots, X_n as variáveis medidas nos elementos de uma população. Basicamente o que o MCP se propõe a fazer é aplicar uma transformação a essas variáveis, tal que as novas variáveis obtidas Y_1, Y_2, \dots, Y_n permitam uma melhor discriminação dos elementos da população. Esta melhor discriminação é de extrema utilidade quando se trata de criar uma tipologia para a população, fazer uma classificação dos elementos segundo critérios dados, etc...

Nesse capítulo vamos apresentar a estrutura matemática do modelo. Inicialmente, nas seções 3.2, 3.3 e 3.4 vamos estudar o enfoque estatístico do MCP, após o que, na seção 3.5, faremos uma revisão do que foi visto. Em seguida, na seção 3.6, apresentamos um exemplo numérico com o objetivo de fixar mais os conceitos estudados. Já na seção 3.7 abordamos o MCP sobre outro enfoque, o da Álgebra Linear, mostrando a equivalência desse tipo de abordagem com a anterior e fazendo, na seção 3.8, comentários sobre o exemplo já visto. Na seção 3.9

é mostrada a característica do MCP, de conservação da variância, e novos comentários sobre o exemplo da seção 3.6 serão feitos na 3.10. Em seguida, abordamos o MCP sobre o enfoque geométrico nas seções 3.11, 3.12 e 3.13 após o que, (seção 3.14), fazemos uma comparação entre as três abordagens (estatística, Álgebra Linear e geométrica) e novos comentários (seção 3.15) sobre o exemplo da seção 3.6. Para finalizar faremos uma analogia do MCP com a Mecânica (seção 3.16), estudaremos alguns casos particulares (seção 3.17) e o problema das escalas empregadas na medição das variáveis (seção 3.18).

3.2. O Modelo das Componentes Principais

Inicialmente vamos relembrar o problema a que nos propusemos a estudar.

Temos um conjunto (ou população) de N elementos nos quais observamos n características e associamos a cada característica observada uma variável X_i ($i=1,2,\dots,n$). Essa associação será de tal forma que $E[X_i] = 0$, ou seja, X_i será um desvio (ver definição de desvio na seção 2.3) e por isso passaremos a representa-lo por x_i . Assim $x_i(k)$ será a representação da i -ésima característica do elemento k da população e cada elemento dessa população apresentará n características expressas pelo vetor $x(k) = (x_1(k), x_2(k), \dots, x_n(k))$.

Vamos agora estudar essa população utilizando o enfoque do MCP. Portanto, nos basearemos na equação 1.2.1 que, utilizando os desvios x_i , tomará a seguinte forma:

$$x_i = f_i(y_1, y_2, y_3, \dots, y_p) \quad i = 1, 2, \dots, n \quad (3.2.1)$$

onde

x_i - representa a i -ésima característica que observamos na população

y_j - ($j=1,2,\dots,p$) é a componente

f_i - é a função que relaciona as componentes y_j ao x_i

OBS.: Mostraremos mais adiante (teorema 3.3.1) que no MCP utilizando desvios x_i obtemos como resultado desvios y_j . Esta é a razão porque, desde já, utilizamos a notação referente a desvios.

Podemos observar que esse enfoque é bem geral pois não faz nenhuma restrição às funções f_i , às componente y_j e nem em relação ao valor de p que poderá ser maior, menor ou igual ao número de características observadas (n). Por sua generalidade, a sua identificação torna-se impraticável.

O MCP faz, como dissemos, na seção 3.1, uma série de restrições à equação 3.2.1. Passamos a descreve-las lembrando que os seus significados ficarão mais claros à medida que o modelo for sendo desenvolvido.

1.^a RESTRIÇÃO

Inicialmente, nesse modelo, f_i é uma função LINEAR. Essa restrição se deve ao fato de os modelos lineares terem uma série de propriedades que os não lineares não tem. Essas propriedades simplificam o problema e, uma vez resolvido o caso linear, poderemos inferir sobre os não lineares. Com essa restrição podemos escrever a equação 3.2.1 da seguinte forma:

$$x_i = \ell_{i1}y_1 + \ell_{i2}y_2 + \dots + \ell_{ip}y_p \quad i=1,2,\dots,n \quad (3.2.2)$$

2ª RESTRIÇÃO

Uma segunda restrição é que p seja igual a n . Essa restrição define o número de componentes que estamos interessados em encontrar. Ao estabelecer que o número de componentes p seja igual ao número de variáveis observadas (n), o modelo fica matematicamente simplificado e esta simplificação será explorada no desenvolvimento do modelo. Agora podemos escrever a equação 3.2.2 assim:

$$x_i = \ell_{i1}y_1 + \ell_{i2}y_2 + \dots + \ell_{in}y_n \quad i = 1,2,\dots,n \quad (3.2.3)$$

ou simplesmente

$$x_i = \sum_{j=1}^n \ell_{ij}y_j \quad i=1,2,\dots,n \quad (3.2.4)$$

3ª RESTRIÇÃO

Outra restrição é que as componentes y_j ($j=1,2,\dots,n$) não sejam correlacionadas entre si. Isso significa impor que não haja nenhuma dependência funcional linear entre as componentes (ver conceito de correlação na seção 2.3). Essa dependência é indesejável já que estamos interessados nas componentes para que possamos observar, de uma forma mais conveniente, as características e os elementos da população em estudo, e não as relações entre as componentes. Assim podemos impor que:

$$\text{COR}[y_j; y_{j'}] = 0 \quad j \neq j' \quad (3.2.7)$$

$$j, j' = 1, 2, \dots, n$$

4.^a RESTRIÇÃO

Antes de definirmos essa restrição vamos ver o seguinte.

Escrevendo a equação 3.2.4 na forma matricial teremos.

$$\underline{x} = L \underline{y} \quad (3.2.8)$$

onde

$$\underline{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad L = \begin{pmatrix} l_{11} & l_{12} & \dots & l_{1n} \\ l_{21} & & & \vdots \\ \vdots & & & \vdots \\ l_{n1} & \dots & \dots & l_{nn} \end{pmatrix} \quad \text{e} \quad \underline{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

Assim, uma vez determinada a matriz L , dado um vetor \underline{y} , temos um vetor \underline{x} . Porém a observação da população em estudo nos dá somente informações sobre as variáveis x_i e o nosso modelo procura justamente a determinação das componentes y_j . Ou seja, dado \underline{x} , o nosso problema consiste em encontrar \hat{L} tal que:

$$\underline{y} = \hat{L} \underline{x} \quad (3.2.9)$$

isto é

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} l_{11} & l_{12} & \dots & l_{1n} \\ l_{21} & & & \vdots \\ \vdots & & & \vdots \\ l_{n1} & \dots & \dots & l_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad (3.2.10)$$

Realmente, como veremos mais adiante, no modelo que desenvolvemos aparecem os coeficientes l_{ji} e não os coeficiente l_{ij} .

Acerca dos vetores $\ell_j = (\ell_{j1}, \ell_{j2}, \dots, \ell_{jn})$ no entanto impomos uma restrição adicional, que é justamente o que caracterizamos como a 4.^a restrição. Impomos a condição de que os ℓ_j sejam unitários, isto é, que

$$\sum_{i=1}^n \ell_{ji}^2 = 1 \quad (3.2.11)$$

O leitor poderia se perguntar como é possível garantir a existência de \hat{L} . Isto é, como é possível garantir que existe a inversa de L ? Nas seções posteriores será mostrada que o atendimento às condições do modelo garante na verdade esta existência.

OBJETIVO

Dentro desse conjunto de restrições o objetivo inicial do MCP será encontrar a primeira componente y_1 tal que ela discrimine ao máximo os elementos da população, ou seja, que os $y_1(k)$ (valores assumidos pela componente y_1 para o elemento k da população) sejam os mais variados possível, o que significa que a variância de y_1 seja máxima. Em outras palavras o objetivo inicial será:

$$\text{MAXIMIZAR VAR}[y_1]$$

O objetivo seguinte será fazer o mesmo para a segunda componente (y_2), lembrando que pela 3.^a restrição a correlação entre y_2 e y_1 deverá ser zero.

Depois iremos repetir esse procedimento mais uma vez para a terceira componente (y_3), lembrando agora que a correlação entre ela e as componentes y_1 e y_2 deverá ser zero.

E assim sucessivamente até a n -ésima componente que deverá ter correlação nula com todas as demais componentes.

É importante lembrar que se cada componente y_j for de tal forma que

$$\text{COR}[y_j; y_{j'}] = 0 \quad \begin{array}{l} j = 1, 2, \dots, n \\ j' = 1, 2, \dots, (j-1) \end{array} \quad (3.2.12)$$

então, segundo a equação 2.3.48, qualquer par de componentes distintas terá correlação nula, coerentemente com a 3.^a restrição.

Podemos agora formular o MCP como um problema de otimização dividido em \underline{n} etapas. Senão vejamos:

PROBLEMA 1:

Dados x_1, x_2, \dots, x_n , encontrar y_1, y_2, \dots, y_n

(onde $y_j = \sum_{i=1}^n \ell_{ji} x_i$) de forma a

1.^a ETAPA

MAXIMIZAR VAR $[y_1]$

sujeito \tilde{a}

$$\sum_{i=1}^n \ell_{1i}^2 = 1$$

2.^a ETAPA

MAXIMIZAR VAR $[y_2]$

sujeito \tilde{a}

$$\sum_{i=1}^n \ell_{2i}^2 = 1$$

$$\text{COR}[y_2; y_1] = 0$$

3.^a ETAPAMAXIMIZAR $\text{VAR}[y_3]$

sujeito a

$$\sum_{i=1}^n \ell_{3i}^2 = 1$$

$$\text{COR}[y_3; y_1] = 0$$

$$\text{COR}[y_3; y_2] = 0$$

$$\vdots$$
n.^a ETAPAMAXIMIZAR $\text{VAR}[y_n]$

sujeito a

$$\sum_{i=1}^n \ell_{ni}^2 = 1$$

$$\text{COR}[y_n; y_1] = 0$$

$$\text{COR}[y_n; y_2] = 0$$

$$\text{COR}[y_n; y_3] = 0$$

$$\vdots$$

$$\text{COR}[y_n; y_{n-1}] = 0$$

Podemos então enunciar o problema da seguinte forma:

PROBLEMA 1:

Dado que numa população de N elementos foram medidas n características em cada um deles, e que essas características são representadas pelas variáveis x_i ($i=1, 2, \dots, n$), en-

contrar n componentes $y_j = \sum_{i=1}^n \ell_{ji} x_i$ ($j=1,2,\dots,n$) sujeitas à restrição $\sum_{i=1}^n \ell_{ji}^2 = 1$ ($j=1,2,\dots,n$) de tal forma que:

- 1) y_1 tenha variância máxima.
 - 2) y_2 tenha variância máxima mas não seja correlacionada com y_1 .
 - 3) y_3 tenha variância máxima mas não seja correlacionada nem com y_1 nem com y_2 .
- ⋮
- n) y_n tenha variância máxima mas não seja correlacionada nem com y_1 , nem com y_2 , nem com y_3, \dots , nem com $y_{(n-1)}$

3.3. O Problema a Ser Resolvido

Como salientamos na seção 3.2, a população em estudo só nos fornece informações diretas sobre as variáveis x_i . Assim, surge a necessidade de escrevermos o Problema 1 em termos das variáveis x_i e não das componentes y_j . Vamos então ver como isso será feito.

Inicialmente, partindo da equação 3.2.9, que pode ser escrita assim:

$$y_j = \sum_{i=1}^n \ell_{ji} x_i \quad j=1,2,\dots,n \quad (3.3.1)$$

vamos verificar uma característica das componentes.

TEOREMA 3.3.1: Se x_i for um desvio então y_j também será.

Demonstração:

Pela definição de desvio (ver seção 2.3)

$$x_i = X_i - \bar{X}_i \quad (3.3.2)$$

Substituindo x_i na equação 3.3.1 vem:

$$y_j = \sum_{i=1}^n \ell_{ji} (X_i - \bar{X}_i) \quad (3.3.3)$$

$$= \sum_{i=1}^n (\ell_{ji} X_i - \ell_{ji} \bar{X}_i) \quad (3.3.4)$$

Usando a propriedade associativa dos somatórios te
mos:

$$y_j = \left(\sum_{i=1}^n \ell_{ji} X_i \right) - \left(\sum_{i=1}^n \ell_{ji} \bar{X}_i \right) \quad (3.3.5)$$

$$= \left(\sum_{i=1}^n \ell_{ji} X_i \right) - \left(\sum_{i=1}^n \ell_{ji} E[X_i] \right) \quad (3.3.6)$$

Como os ℓ_{ij} são constantes para todo elemento da população (ver propriedades das médias na seção 2.3) então

$$\sum_{i=1}^n \ell_{ji} E[X_i] = E \left[\sum_{i=1}^n \ell_{ji} X_i \right] \quad (3.3.7)$$

Substituindo esse resultado na equação 3.3.6 temos:

$$y_j = \left(\sum_{i=1}^n \ell_{ji} X_i \right) - E \left[\sum_{i=1}^n \ell_{ji} X_i \right] \quad (3.3.8)$$

Levando em conta que $Y_j = \sum_{i=1}^n \ell_{ji} X_i$ a equação

3.3.8 ficará

$$y_j = Y_j - E[Y_j] \quad (3.3.9)$$

$$= Y_j - \bar{Y}_j \quad (3.3.10)$$

Assim concluímos que y_j é um desvio. △

É importante lembrar que tanto os x_i como os y_j são desvios mas não necessariamente são variáveis padronizadas.

Podemos supor que nem todos os valores $y_j(1), y_j(2), \dots, y_j(N)$ sejam iguais pois medimos esses valores numa população. Nesse caso teremos:

$$\text{VAR}[y_j] \neq 0 \quad (3.3.13)$$

E sob essa hipótese podemos afirmar que

$$\text{COR}[y_j; y_{j'}] = 0 \quad \text{o que equivale a} \quad \text{COV}[y_j; y_{j'}] = 0$$

Por outro lado, como y_j é um desvio, temos

$$\text{COV}[y_j; y_{j'}] = E[y_j \cdot y_{j'}] \quad (3.3.14)$$

Substituindo a equação 3.3.1 na 3.3.14 chegamos a

$$\text{COV}[y_j; y_{j'}] = E\left[\left(\sum_{i=1}^n \ell_{ji} x_i\right) \left(\sum_{i'=1}^n \ell_{j'i'} x_{i'}\right)\right] \quad (3.3.15)$$

Usando as propriedades dos somatórios a equação acima ficará

$$\text{COV}[y_j; y_{j'}] = E\left[\sum_{i=1}^n \sum_{i'=1}^n \ell_{ji} x_i \ell_{j'i'} x_{i'}\right] \quad (3.3.16)$$

$$= E\left[\sum_{i=1}^n \sum_{i'=1}^n \ell_{ji} \ell_{j'i'} x_i x_{i'}\right] \quad (3.3.17)$$

Como os ℓ_{ji} são constantes para todos os elementos da população encontramos

$$\text{COV}[y_j; y_{j'}] = \sum_{i=1}^n \sum_{i'=1}^n (\ell_{ji} \ell_{j'i'} E[x_i x_{i'}]) \quad (3.3.18)$$

Mas pela equação 2.3.59A

$$\text{COV}[x_i; x_{i'}] = E[x_i x_{i'}] \quad (3.3.19)$$

Substituindo a equação 3.3.19 na 3.3.18 encontramos

$$\text{COV}[y_j; y_{j'}] = \sum_{i=1}^n \sum_{i'=1}^n \ell_{ji} \ell_{j'i'} \text{COV}[x_i; x_{i'}] \quad (3.3.20)$$

ou

$$\text{COV}[y_j; y_{j'}] = \sum_{i=1}^n \sum_{i'=1}^n \ell_{ji} \ell_{j'i'} c_{ii'} \quad (3.3.21)$$

onde

$$c_{ii'} = \text{COV}[x_i; x_{i'}] \quad (3.3.22)$$

Desta forma concluímos que a 3ª restrição do Problema 1 pode ser expressa pela seguinte equação

$$\boxed{\sum_{i=1}^n \sum_{i'=1}^n \ell_{ji} \ell_{j'i'} c_{ii'} = 0} \quad \begin{array}{l} j = 1, 2, \dots, n \\ j' = 1, 2, \dots, (j-1) \end{array} \quad (3.3.23)$$

Por outro lado, sabemos da seção 2.3 que

$$\text{VAR}[Y_j] = \text{VAR}[y_j] = \text{COV}[y_j; y_j] \quad (3.3.24)$$

logo com base na equação 3.3.21, chegamos a

$$\boxed{\text{VAR}[Y_j] = \text{VAR}[y_j] = \sum_{i=1}^n \sum_{i'=1}^n \ell_{ji} \ell_{ji'} c_{ii'}} \quad (3.3.25)$$

Visto isso podemos escrever o Problema 1 da seguinte forma:

PROBLEMA 1:

Dados x_1, x_2, \dots, x_n , encontrar y_1, y_2, \dots, y_n

(onde $y_j = \sum_{i=1}^n \ell_{ji} x_i$) de forma a

1.^a ETAPA

$$\text{MAXIMIZAR } \sum_{i=1}^n \sum_{i'=1}^n \ell_{1i} \ell_{1i'}, c_{ii'} \quad (3.3.26)$$

sujeito a

$$\left\{ \sum_{i=1}^n \ell_{1i}^2 = 1 \right. \quad (3.3.27)$$

2.^a ETAPA

$$\text{MAXIMIZAR } \sum_{i=1}^n \sum_{i'=1}^n \ell_{2i} \ell_{2i'}, c_{ii'} \quad (3.3.28)$$

sujeito a

$$\left\{ \sum_{i=1}^n \ell_{2i}^2 = 1 \right. \quad (3.3.29)$$

$$\left\{ \sum_{i=1}^n \sum_{i'=1}^n \ell_{2i} \ell_{1i'}, c_{ii'} = 0 \right. \quad (3.3.30)$$

3.^a ETAPA

$$\text{MAXIMIZAR } \sum_{i=1}^n \sum_{i'=1}^n \ell_{3i} \ell_{3i'}, c_{ii'} \quad (3.3.31)$$

sujeito a

$$\left\{ \sum_{i=1}^n \ell_{3i}^2 = 1 \right. \quad (3.3.32)$$

$$\left\{ \begin{array}{l} \sum_{i=1}^n \sum_{i'=1}^n \ell_{3i} \ell_{1i'} c_{ii'} = 0 \\ \sum_{i=1}^n \sum_{i'=1}^n \ell_{3i} \ell_{2i'} c_{ii'} = 0 \\ \vdots \end{array} \right. \quad (3.3.33)$$

$$\left\{ \begin{array}{l} \sum_{i=1}^n \sum_{i'=1}^n \ell_{3i} \ell_{2i'} c_{ii'} = 0 \\ \vdots \end{array} \right. \quad (3.3.34)$$

n.^a ETAPA

$$\text{MAXIMIZAR} \quad \sum_{i=1}^n \sum_{i'=1}^n \ell_{ni} \ell_{ni'} c_{ii'} \quad (3.3.35)$$

sujeito a

$$\left\{ \begin{array}{l} \sum_{i=1}^n \ell_{ni}^2 = 1 \\ \sum_{i=1}^n \sum_{i'=1}^n \ell_{ni} \ell_{1i'} c_{ii'} = 0 \\ \sum_{i=1}^n \sum_{i'=1}^n \ell_{ni} \ell_{2i'} c_{ii'} = 0 \\ \sum_{i=1}^n \sum_{i'=1}^n \ell_{ni} \ell_{3i'} c_{ii'} = 0 \\ \vdots \\ \sum_{i=1}^n \sum_{i'=1}^n \ell_{ni} \ell_{(n-1)i'} c_{ii'} = 0 \end{array} \right. \quad (3.3.36)$$

$$\sum_{i=1}^n \sum_{i'=1}^n \ell_{ni} \ell_{1i'} c_{ii'} = 0 \quad (3.3.37)$$

$$\sum_{i=1}^n \sum_{i'=1}^n \ell_{ni} \ell_{2i'} c_{ii'} = 0 \quad (3.3.38)$$

$$\sum_{i=1}^n \sum_{i'=1}^n \ell_{ni} \ell_{3i'} c_{ii'} = 0 \quad (3.3.39)$$

$$\left. \begin{array}{l} \vdots \\ \sum_{i=1}^n \sum_{i'=1}^n \ell_{ni} \ell_{(n-1)i'} c_{ii'} = 0 \end{array} \right\} \quad (3.3.40)$$

3.4. Resolução do Problema 1

Para resolver o Problema 1 basta encontrarmos os ℓ_{ji} , já que as covariâncias $c_{ii'}$ serão obtidas através de medições efetuadas no conjunto de elementos (ou população) em estudo.

Essas covariâncias podem ser agrupadas numa matriz da seguinte forma

$$C = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & \vdots \\ \vdots & \vdots & \cdots & \vdots \\ c_{n1} & \cdots & \cdots & c_{nn} \end{pmatrix}$$

Podemos notar que essa é uma matriz quadrada de \underline{n} linhas por \underline{n} colunas.

Como sabemos da Álgebra Linear (ver [7], seção 2.6), toda matriz quadrada $n \times n$ possui \underline{n} autovalores. Nessa seção vamos nos preocupar somente com o caso mais geral no qual esses \underline{n} autovalores sejam todos diferentes entre si e que nenhum seja nulo. Os casos em que isso não acontecer serão abordados posteriormente na seção 3.17.

Dividiremos o Problema 1 em \underline{n} etapas:

1.^a ETAPA

$$\text{MAXIMIZAR} \quad \sum_{i=1}^n \sum_{i'=1}^n \ell_{1i} \ell_{1i'} c_{ii'} \quad (3.4.1)$$

sujeito a

$$\left\{ \sum_{i=1}^n \ell_{1i}^2 = 1 \right. \quad (3.4.2)$$

Solução:

Utilizaremos o Método dos Multiplicadores de Lagrange (ver [1^o], [1⁴] e [2⁴]) na resolução desse problema. Assim o problema passará a ser

$$\text{MAXIMIZAR } T_1 = \left[\sum_{i'=1}^n \sum_{i''=1}^n \ell_{1i'} \ell_{1i''} c_{i'i''} \right] - \lambda_1 \left[\left(\sum_{i'=1}^n \ell_{1i'}^2 \right) - 1 \right] \quad (3.4.3)$$

onde λ_1 é o Multiplicador de Lagrange. (os índices foram re-arrumados de forma mais conveniente já que iremos utilizar ℓ_{1i} , mais adiante, para outro fim). Agora T_1 é função dos ℓ_{1i} e de λ_1 , e o problema não possui restrições.

Diferenciando a equação 3.4.3 em relação a cada uma dessas variáveis e, além disso, observando que $c_{i'i''} = c_{i''i'}$, teremos

$$\left\{ \begin{array}{l} \frac{\delta T_1}{\delta \ell_{1i}} = 2 \sum_{i'=1}^n \ell_{1i'} c_{ii'} - 2\lambda_1 \ell_{1i} \\ \hspace{20em} i=1,2,\dots,n \end{array} \right. \quad (3.4.4)$$

$$\left\{ \begin{array}{l} \frac{\delta T_1}{\delta \lambda_1} = \left(\sum_{i'=1}^n \ell_{1i'}^2 \right) - 1 \end{array} \right. \quad (3.4.5)$$

Mas para T_1 ser máximo suas derivadas parciais em relação aos ℓ_{1i} e λ_1 deverão ser nulas o que acarreta

$$\left\{ \begin{array}{l} \sum_{i'=1}^n \ell_{1i'} c_{ii'} - \lambda_1 \ell_{1i} = 0 \\ \hspace{15em} i=1,2,\dots,n \end{array} \right. \quad (3.4.6)$$

$$\left\{ \begin{array}{l} \sum_{i'=1}^n \ell_{1i'}^2 = 1 \end{array} \right. \quad (3.4.7)$$

Dessas equações podemos tirar duas conclusões.

1.^a) λ_1 é a variância de y_1

2.^a) λ_1 é um autovalor da matriz C.

senão vejamos.

TEOREMA 3.4.1: λ_1 é a variância de y_1 .

Demonstração:

Multiplicando a equação 3.4.6 por ℓ_{1i} e somando para todo $i=1,2,\dots,n$ teremos

$$\sum_{i=1}^n \sum_{i'=1}^n \ell_{1i} \ell_{1i'} c_{ii'} - \lambda_1 \sum_{i=1}^n \ell_{1i}^2 = 0 \quad (3.4.8)$$

Substituindo a equação 3.4.7 na 3.4.8 temos:

$$\sum_{i=1}^n \sum_{i'=1}^n \ell_{1i} \ell_{1i'} c_{ii'} = \lambda_1 \quad (3.4.9)$$

Por outro lado, fazendo $j = 1$ na equação 3.3.25, temos:

$$\text{VAR}[y_1] = \sum_{i=1}^n \sum_{i'=1}^n \ell_{1i} \ell_{1i'} c_{ii'} \quad (3.4.10)$$

Reunindo as equações 3.4.9 e 3.4.10 teremos finalmente que:

$$\lambda_1 = \text{VAR}[y_1] \quad (3.4.11)$$

Com relação à segunda conclusão temos o seguinte teorema. △

TEOREMA 3.4.2: λ_1 é um autovalor da matriz C.

Demonstração:

Seja

$$\ell_1 = (\ell_{11}, \ell_{12}, \dots, \ell_{1n})$$

e

$$C = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & & & \vdots \\ \vdots & & & \vdots \\ c_{n1} & \dots\dots\dots & & c_{nn} \end{pmatrix}$$

Podemos então escrever a equação 3.4.6 assim:

$$C \mathcal{L}_1^T = \lambda_1 \mathcal{L}_1^T \quad (3.4.12)$$

que nada mais é do que a definição de autovalor. Logo λ_1 é um autovalor da matriz C e \mathcal{L}_1^T é um autovetor da matriz C associado ao autovalor λ_1 . △

Voltando às duas conclusões, observamos que o valor máximo da $\text{VAR}[y_1]$ será o maior autovalor da matriz C e os \mathcal{L}_{1i} serão as componentes de um autovetor dessa matriz, associado a esse autovalor.

Sabemos, no entanto, da Álgebra Linear que associado a um autovalor existe uma infinidade de autovetores. Porém lembrando a 4ª restrição observamos que:

$$\sum_{i=1}^n \mathcal{L}_{1i}^2 = 1 \quad (3.4.13)$$

logo o autovetor escolhido será aquele que satisfizer a equação 3.4.13. Assim os \mathcal{L}_{1i} serão as componentes do autovetor unitário da matriz C associado ao autovalor λ_1 (Obs.: Autovetor, ou vetor, unitário é aquele cujo somatório dos quadrados de suas componentes é igual a 1).

Fim da 1ª etapa □

2.^a ETAPA

$$\text{MAXIMIZAR } \sum_{i=1}^n \sum_{i'=1}^n \ell_{2i} \ell_{2i'} c_{ii'} \quad (3.4.14)$$

sujeito a

$$\left\{ \begin{array}{l} \sum_{i=1}^n \ell_{2i}^2 = 1 \end{array} \right. \quad (3.4.15)$$

$$\left\{ \begin{array}{l} \sum_{i=1}^n \sum_{i'=1}^n \ell_{2i} \ell_{1i'} c_{ii'} = 0 \end{array} \right. \quad (3.4.16)$$

Solução:

Mais uma vez, utilizando o Método dos Multiplicadores de Lagrange, o problema ficará da seguinte forma:

$$\begin{aligned} \text{MAXIMIZAR } T_2 = & \left[\sum_{i'=1}^n \sum_{i''=1}^n \ell_{2i'} \ell_{2i''} c_{ii''} \right] - \lambda_2 \left[\left(\sum_{i'=1}^n \ell_{2i'}^2 \right) - 1 \right] - \\ & - \alpha_1 \left(\sum_{i'=1}^n \sum_{i''=1}^n \ell_{2i'} \ell_{1i''} c_{i'i''} \right) \end{aligned} \quad (3.4.17)$$

onde λ_2 e α_1 são os Multiplicadores de Lagrange. (os índices, mais uma vez, foram re-arrumados de forma mais conveniente já que iremos utilizar ℓ_{2i} , mais adiante para outro fim). Agora T_2 é função dos ℓ_{2i} , de λ_2 e também de α_1 e o problema não possui mais restrições.

Diferenciando a equação 3.4.17 em relação a cada uma dessas variáveis e, além disso, observando que $c_{i'i''} = c_{i''i'}$, teremos:

$$\left\{ \begin{array}{l} \frac{\delta T_2}{\delta \ell_{2i}} = 2 \sum_{i'=1}^n \ell_{2i'} c_{ii'} - 2\lambda_2 \ell_{2i} - \alpha_1 \sum_{i'=1}^n \ell_{1i'} c_{ii'} \\ \hspace{15em} i = 1, 2, \dots, n \end{array} \right. \quad (3.4.18)$$

$$\left\{ \begin{array}{l} \frac{\delta T_2}{\delta \lambda_2} = \sum_{i'=1}^n \ell_{2i'}^2 - 1 \end{array} \right. \quad (3.4.19)$$

$$\left\{ \begin{array}{l} \frac{\delta T_2}{\delta \alpha_1} = \sum_{i'=1}^n \sum_{i''=1}^n \ell_{2i'} \ell_{1i''} c_{i'i''} \end{array} \right. \quad (3.4.20)$$

Mas para T_2 ser máximo suas derivadas parciais em relação aos ℓ_{2j} , a λ_2 e α_1 deverão ser nulas o que acarreta

$$\left\{ \begin{array}{l} 2 \sum_{i'=1}^n \ell_{2i'} c_{ii'} - 2\lambda_2 \ell_{2i} - \alpha_1 \sum_{i'=1}^n \ell_{1i'} c_{ii'} = 0 \\ \hspace{15em} i = 1, 2, \dots, n \end{array} \right. \quad (3.4.21)$$

$$\left\{ \begin{array}{l} \sum_{i'=1}^n \ell_{2i'}^2 = 1 \end{array} \right. \quad (3.4.22)$$

$$\left\{ \begin{array}{l} \sum_{i'=1}^n \sum_{i''=1}^n \ell_{2i'} \ell_{1i''} c_{i'i''} = 0 \end{array} \right. \quad (3.4.23)$$

Dessas equações podemos tirar duas conclusões:

- 1.^a) λ_2 é a variância de y_2
- 2.^a) λ_2 é um autovalor da matriz c .

senão vejamos

TEOREMA 3.4.3: λ_2 é a variância de y_2

Demonstração:

Multiplicando a equação 3.4.21 por ℓ_{2i} e somando para todo $i = 1, 2, \dots, n$ teremos:

$$2 \sum_{i=1}^n \sum_{i'=1}^n \ell_{2i} \ell_{2i'} c_{ii'} - 2\lambda_2 \sum_{i=1}^n \ell_{2i}^2 - \alpha_1 \sum_{i=1}^n \sum_{i'=1}^n \ell_{2i} \ell_{1i'} c_{ii'} = 0 \quad (3.2.24)$$

$i=1, 2, \dots, n$

Substituindo as equações 3.4.22-23 na 3.4.24 chegaremos a:

$$\sum_{i=1}^n \sum_{i'=1}^n \ell_{2i} \ell_{2i'} c_{ii'} = \lambda_2 \quad (3.4.25)$$

Comparando as equações 3.4.25 e 3.3.25 chegamos a

$$\lambda_2 = \text{VAR}[y_2] \quad (3.4.27)$$

△

Com relação à segunda conclusão temos o seguinte teorema.

TEOREMA 3.4.4: λ_2 é um autovalor da matriz C

Demonstração:

Seja:

$$\ell_1 = (\ell_{11}, \ell_{12}, \dots, \ell_{1n})$$

$$\ell_2 = (\ell_{21}, \ell_{22}, \dots, \ell_{2n})$$

e

$$C = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & & & \vdots \\ \vdots & & & \vdots \\ c_{n1} & \cdots & \cdots & c_{nn} \end{pmatrix}$$

então podemos escrever as equações 3.4.6-7 e 3.4.21-23 da seguinte forma:

$$C \ell_1^T = \lambda_1 \ell_1^T \quad (3.4.28)$$

$$\ell_1 \ell_1^T = 1 \quad (3.4.29)$$

$$2 C \ell_2^T - 2\lambda_2 \ell_2^T = \alpha_1 C \ell_1^T \quad (3.4.30)$$

$$\ell_2 \ell_2^T = 1 \quad (3.4.31)$$

$$\ell_2 C \ell_1^T = \ell_1 C \ell_2^T = 0 \quad (3.4.32)$$

Pré-multiplicando a equação 3.4.30 pelo vetor ℓ_1 temos:

$$2\ell_1 C \ell_2^T - 2\lambda_2 \ell_1 \ell_2^T = \alpha_1 \ell_1 C \ell_1^T \quad (3.4.33)$$

Por outro lado, reunindo as equações 3.4.28 e 3.4.32 concluímos que:

$$\lambda_1 \ell_2 \ell_1^T = 0 \quad (3.4.34)$$

Como por hipótese (ver início desta seção) $\lambda_1 \neq 0$ então

$$\ell_2 \ell_1^T = \ell_1 \ell_2^T = 0 \quad (3.4.35)$$

Levando as equações 3.4.32 e 3.4.35 na 3.4.33 concluímos que:

$$\alpha_1 \mathcal{L}_1 C \mathcal{L}_1^T = 0 \quad (3.4.36)$$

Substituindo a equação 3.4.28 na 3.4.36 temos:

$$\alpha_1 \mathcal{L}_1 \lambda_1 \mathcal{L}_1^T = \alpha_1 \lambda_1 \mathcal{L}_1 \mathcal{L}_1^T = 0 \quad (3.4.37)$$

Mas pela equação 3.4.29, $\mathcal{L}_1 \mathcal{L}_1^T = 1$. Logo para que a equação 3.4.37 se verifique temos $\alpha_1 \lambda_1 = 0$. Mas como por hipótese (ver início desta seção) $\lambda_1 \neq 0$ segue que $\alpha_1 = 0$.

Fazendo $\alpha_1 = 0$ na equação 3.4.30 temos:

$$C \mathcal{L}_2^T = \lambda_2 \mathcal{L}_2^T \quad (3.4.38)$$

que nada mais é do que a definição de autovalor. Logo λ_2 é um autovalor da matriz C e \mathcal{L}_2^T é um autovetor da matriz C associado ao autovalor λ_2 . △

Voltando às duas conclusões, observamos que o valor máximo da $\text{VAR}[y_2]$ será o segundo maior autovalor da matriz C (já que pela 3a. restrição a $\text{COR}[y_2; y_1] = 0$ o que acarreta $y_2 \neq y_1$) e os \mathcal{L}_{2i} serão as componentes de um autovetor dessa matriz, associado a esse autovalor.

Como fizemos na 1.^a Etapa, o autovetor \mathcal{L}_{2i} será unitário.

Fim da 2.^a etapa □

Vamos agora resolver uma etapa genérica do Problema 1.

j.^a ETAPA
(j = 3, 4, ..., n)

$$\text{MAXIMIZAR } \sum_{i=1}^n \sum_{i'=1}^n \ell_{ji} \ell_{ji'} c_{ii'} \quad (3.4.39)$$

sujeito a

$$\sum_{i=1}^n \ell_{ji}^2 = 1 \quad (3.4.40)$$

$$\sum_{i=1}^n \sum_{i'=1}^n \ell_{ji} \ell_{1i'} c_{ii'} = 0 \quad (3.4.41)$$

$$\sum_{i=1}^n \sum_{i'=1}^n \ell_{ji} \ell_{2i'} c_{ii'} = 0 \quad (3.4.42)$$

' ' ' ' '
' ' ' ' '
' ' ' ' '

$$\sum_{i=1}^n \sum_{i'=1}^n \ell_{ji} \ell_{(j-1)i'} c_{ii'} = 0 \quad (3.4.43)$$

Solução:

Utilizando o Método dos Multiplicadores de Lagrange, como nas outras etapas, o problema ficará:

$$\begin{aligned} \text{MAXIMIZAR } T_j = & \left[\sum_{i'=1}^n \sum_{i''=1}^n \ell_{ji'} \ell_{ji''} c_{i'i''} \right] - \\ & - \lambda_j \left[\left(\sum_{i'=1}^n \ell_{ji'}^2 \right) - 1 \right] - \\ & - \sum_{j'=1}^{(j-1)} \alpha_{j'} \left(\sum_{i'=1}^n \sum_{i''=1}^n \ell_{ji'} \ell_{j'i''} c_{i'i''} \right) \end{aligned} \quad (3.4.44)$$

onde λ_j e os $\alpha_{j'}$ são os Multiplicadores de Lagrange (como das outras vezes, os índices foram re-arrumados). Agora T_j é função dos ℓ_{ji} , de λ_j e também dos $\alpha_{j'}$, e o problema não possui restrições.

Substituindo as equações 3.4.46-47 na 3.4.49 chegamos a:

$$\sum_{i=1}^n \sum_{i'=1}^n \ell_{ji} \ell_{ji'} c_{ii'} = \lambda_j \quad (3.4.49)$$

Finalmente, reunindo a equação acima com a 3.3.25, chegamos a:

$$\lambda_j = \text{VAR}[y_j] \quad (3.4.50)$$

Com relação à segunda conclusão temos o seguinte teorema. (ver anexo) \triangle

TEOREMA 3.4.6: λ_j é um autovalor da matriz C.

Demonstração:

Vamos partir da hipótese que já calculamos $y_{j'}$ ($j'=1,2,\dots,(j-1)$) e concluímos que os $\lambda_{j'}$ são autovalores da matriz C e os $\ell_{j'}^T$ são os autovetores unitários associados aos $\lambda_{j'}$.

Sendo assim, podemos escrever que:

$$C \ell_{j'}^T = \lambda_{j'} \ell_{j'}^T \quad j' = 1, 2, \dots, (j-1) \quad (3.4.51)$$

$$\ell_{j'} \ell_{j'}^T = 1 \quad j' = 1, 2, \dots, (j-1) \quad (3.4.52)$$

onde

$$\ell_{j'} = (\ell_{j',1}, \ell_{j',2}, \dots, \ell_{j',n})$$

Por outro lado, podemos escrever as equações 3.4.45-47 na forma matricial, assim:

$$2 C \mathcal{L}_j^T - 2\lambda_j \mathcal{L}_j^T = \sum_{j'=1}^{(j-1)} \alpha_{j'} [C \mathcal{L}_{j'}^T] \quad (3.4.53)$$

$$\mathcal{L}_j \mathcal{L}_j^T = \mathcal{L}_j^T \mathcal{L}_j = 1 \quad (3.4.54)$$

$$\mathcal{L}_j C \mathcal{L}_{j'}^T = \mathcal{L}_{j'} C \mathcal{L}_j^T = 0 \quad (3.4.55)$$

$$j'=1, 2, \dots, (j-1)$$

Pré-multiplicando a equação 3.4.53 pelo vetor \mathcal{L}_j , ($\forall j'=1, 2, \dots, (j-1)$) teremos:

$$2\mathcal{L}_j C \mathcal{L}_j^T - 2\lambda_j \mathcal{L}_j \mathcal{L}_j^T = \sum_{j'=1}^{(j-1)} \alpha_{j'} \mathcal{L}_j [C \mathcal{L}_{j'}^T] \quad (3.4.56)$$

Por outro lado, reunindo as equações 3.4.51 e 3.4.55 concluímos que:

$$\lambda_{j'} \mathcal{L}_j \mathcal{L}_{j'}^T = 0 \quad j'=1, 2, \dots, (j-1) \quad (3.4.57)$$

Como por hipótese (ver início da seção) $\lambda_{j'} \neq 0$ então:

$$\mathcal{L}_j \mathcal{L}_{j'}^T = \mathcal{L}_{j'} \mathcal{L}_j^T = 0 \quad j'=1, 2, \dots, (j-1) \quad (3.4.58)$$

Levando as equações 3.4.55 e 3.4.58 na 3.4.56 concluímos que:

$$\sum_{j'=1}^{(j-1)} \alpha_{j'} \mathcal{L}_j [C \mathcal{L}_{j'}^T] = 0 \quad (3.4.59)$$

Substituindo a equação 3.4.51 na 3.4.59 temos:

$$\sum_{j'=1}^{(j-1)} \alpha_{j'} \mathcal{Z}_{j'} \lambda_{j'} \mathcal{Z}_{j'}^T = \sum_{j'=1}^{(j-1)} \alpha_{j'} \lambda_{j'} \mathcal{Z}_{j'} \mathcal{Z}_{j'}^T = 0 \quad (3.4.60)$$

Mas pela equação 3.4.52, $\mathcal{Z}_{j'} \mathcal{Z}_{j'}^T = 1$ para $j'=1, 2, \dots, (j-1)$. Logo para que a equação 3.4.60 se verifique temos:

$$\sum_{j'=1}^{(j-1)} \alpha_{j'} \lambda_{j'} = 0 \quad (3.4.60A)$$

Mas como por hipótese (ver início da seção) $\lambda_{j'} \neq 0$ ($\forall j'=1, 2, \dots, (j-1)$) e supondo que nas $2, \dots, j-1$ etapas mostramos que $\alpha_{j', -1} = 0$ segue que $\alpha_{j'} = 0$ ($\forall j' = 1, 2, \dots, (j-1)$).

Levando esse resultado na equação 3.4.53 temos:

$$C \mathcal{Z}_j^T = \lambda_j \mathcal{Z}_j^T \quad (3.4.61)$$

que nada mais é do que a definição de autovalor. Logo λ_j é um autovalor da matriz C e \mathcal{Z}_j^T é um autovetor da matriz C associado ao autovalor λ_j

△

Voltando, mais uma vez, às duas conclusões observa

mos que o valor máximo da $\text{VAR}[y_j]$ será o j -ésimo maior autovalor da matriz C (já que se $j \neq j'$ então $y_j \neq y_{j'}$, pois $\text{COR}[y_j; y_{j'}] = 0$) e os ℓ_{ji} serão as componentes do autovetor unitário (por causa da equação 3.4.40) dessa matriz, associado a esse autovalor.

Fim da j -ésima etapa □

Com isso, concluímos que para resolver o Problema 1 precisamos inicialmente calcular os autovalores da matriz C (matriz de covariâncias) e depois encontrar os autovetores unitários associados a esses autovalores. Em outras palavras, precisamos resolver o seguinte sistema de equações:

$$(C - \lambda_j I) \ell_j^T = \emptyset \quad j=1, 2, \dots, n \quad (3.4.62)$$

$$\sum_{i=1}^n \ell_{ji}^2 = 1 \quad j=1, 2, \dots, n \quad (3.4.63)$$

onde I é a Matriz Identidade, λ_j é o j -ésimo maior autovalor da matriz C , ℓ_j^T o autovetor unitário da matriz C associado ao autovalor λ_j e \emptyset é um vetor coluna de n componentes nulas. (para maiores esclarecimentos sobre a equação 3.4.62 consultar [7], seção 2.6 ou [23], capítulo IX).

É importante que notemos que sob a hipótese feita no início dessa seção de que a matriz C possui n autovalores distintos e não nulos, o sistema de equações 3.4.62-3 fica determinado a menos dos sinais dos ℓ_{ji} . Senão vejamos:

TEOREMA 3.4.7: Seja o sistema de equações:

$$(C - \lambda_j I) \ell_j^T = \emptyset \quad j=1, 2, \dots, n \quad (3.4.64)$$

$$\sum_{i=1}^n \ell_{ji}^2 = 1 \quad j=1,2,\dots,n \quad (3.4.65)$$

onde:

C é a Matriz de Covariâncias;

I é a Matriz Identidade;

λ_j é o j-ésimo maior autovalor da matriz C;

ℓ_j^T é o autovetor unitário da matriz C, associado ao autovalor λ_j , de componentes ℓ_{ji} ;

\emptyset é um vetor coluna de n componentes nulas.

Então, para cada j, o sistema possui 2 soluções que diferem apenas no sinal.

Demonstração:

Vamos iniciar essa demonstração usando como lema uma afirmativa extraída da referência [2], pág.338-9 que escrevemos a seguir:

LEMA 1

"Dada qualquer matriz simétrica G, existe uma transformação ortogonal representada pela matriz L tal que:

$$L^T G L = D = \begin{pmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ 0 & 0 & \lambda_3 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & \lambda_n \end{pmatrix} \quad (3.4.66)$$

onde $\lambda_1, \lambda_2, \dots, \lambda_n$ são autovalores da matriz G."

Como a Matriz de Covariâncias C é simétrica então existe uma matriz D tal que:

$$L^T C L = D = \begin{pmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ 0 & 0 & \lambda_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \lambda_n \end{pmatrix} \quad (3.4.67)$$

onde $\lambda_1, \lambda_2, \dots, \lambda_n$ são os autovalores da matriz C .

Por outro lado podemos escrever que:

$$L^T (C - \lambda_j I) L = (L^T C L) - (\lambda_j L^T L) \quad (3.4.68)$$

$$j=1, 2, \dots, n$$

Mas como L representa uma transformação ortogonal temos:

$$L^T L = I \quad (3.4.69)$$

logo, reunindo as equações 3.4.67-69 teremos:

$$L^T (C - \lambda_j I) L = D - \lambda_j I \quad (3.4.70)$$

$$j=1, 2, \dots, n$$

ou seja:

$$L^T (C - \lambda_j I) L = \begin{pmatrix} \lambda_1 - \lambda_j & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 - \lambda_j & 0 & \dots & 0 \\ 0 & 0 & \lambda_3 - \lambda_j & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \lambda_n - \lambda_j \end{pmatrix}$$

$$j=1, 2, \dots, n \quad (3.4.71)$$

Podemos então notar que sendo os autovalores λ_j ($j=1,2,\dots,n$) todos diferentes e não nulos, a matriz $(D - \lambda_j I)$ possui $(n-1)$ colunas linearmente independentes, ou seja, seu rank é $(n-1)$ (ver definição de rank, ou posto, de uma matriz em [6],[7],[15] ou [23]). Logo teremos:

$$\text{rank}[L^T(C - \lambda_j I)L] = n - 1 \quad (3.4.72)$$

Seja então:

$$\text{rank}[C - \lambda_j I] = v \quad (3.4.73)$$

Por outro lado, pela equação 3.4.69, nota-se que, como $\text{rank}(I) = n$ então:

$$\text{rank}[L] = \text{rank}[L^T] = n \quad (3.4.74)$$

Porém, sabemos da Álgebra Linear (ver [7],pág.33, inequação de Sylvester) que:

$$\begin{aligned} \text{rank}[L^T] + \text{rank}[(C - \lambda_j I)L] - n &\leq \text{rank}[L^T(C - \lambda_j I)L] \leq \\ &\leq \text{mínimo}\{\text{rank}[L^T], \text{rank}[(C - \lambda_j I)L]\} \end{aligned} \quad (3.4.75)$$

Levando as equações 3.4.72 e 3.4.74 na 3.4.75 teremos:

$$\begin{aligned} \cancel{n} + \text{rank}[(C - \lambda_j I)L] - \cancel{n} &\leq n - 1 \leq \\ &\leq \text{mínimo}\{n, \text{rank}[(C - \lambda_j I)L]\} \end{aligned} \quad (3.4.76)$$

logo:

$$\text{rank}[(C - \lambda_j I)L] = n - 1 \quad (3.4.77)$$

Aplicando novamente a inequação de Sylvester teremos:

$$\begin{aligned} \text{rank}[C - \lambda_j I] + \text{rank}[L] - n &\leq \text{rank}[(C - \lambda_j I)L] \leq \\ &\leq \text{mínimo}\{\text{rank}[C - \lambda_j I], \text{rank}[L]\} \end{aligned} \quad (3.4.78)$$

Levando as equações 3.4.73-4 e 3.4.77 na 3.4.78 teremos:

$$v + \cancel{n} - \cancel{n} \leq n - 1 \leq \underline{\text{mínimo}}\{v, n\} \quad (3.4.79)$$

Logo:

$$v = \text{rank}[C - \lambda_j I] = n - 1 \quad (3.4.80)$$

Seja então $U_1, U_2, U_3, \dots, U_p$ soluções da equação:

$$(C - \lambda_j I)U_j = \emptyset \quad (3.4.81)$$

(onde $U_j^T = (U_{j1}, U_{j2}, \dots, U_{jn})$ e \emptyset é um vetor de n componentes nulas). Seja $U = (U_1, U_2, \dots, U_p)$. Usando agora a matriz U ao invés dos vetores U_j temos:

$$(C - \lambda_j I)U = \phi \quad (3.4.82)$$

(onde ϕ é uma matriz $n \times p$ com n colunas iguais a \emptyset)

Aplicando a inequação de Sylvester à equação 3.4.82 temos:

$$\text{rank}[C - \lambda_j I] + \text{rank}[U] - n \leq \text{rank}[\phi] \quad (3.4.83)$$

ou seja:

$$(n - 1) + \text{rank}[U] - n \leq 0 \quad (3.4.84)$$

logo:

$$\text{rank } U \leq 1 \quad (3.4.85)$$

o que nos leva a concluir que se \mathcal{L}_j^T é solução da equação 3.4.64, então qualquer outra solução daquela equação será da forma $\beta \mathcal{L}_j^T$ onde $\beta \in \mathbb{R}$.

Por outro lado seja \mathcal{L}_j^T e $\beta \mathcal{L}_j^T$ soluções do sistema de equações formado pelas equações 3.4.64-65 logo:

$$\sum_{i=1}^n \ell_{ji}^2 = 1 \quad (3.4.86)$$

e

$$\sum_{i=1}^n (\beta \ell_{ji})^2 = 1 \quad (3.4.87)$$

mas

$$\sum_{i=1}^n (\beta \ell_{ji})^2 = \beta^2 \sum_{i=1}^n \ell_{ji}^2 = \beta^2 = 1 \quad (3.4.88)$$

ou seja:

$$\beta = 1 \quad \text{ou} \quad \beta = -1 \quad (3.4.90)$$

Logo como queríamos provar, o sistema de equações 3.8.64-65, possui duas soluções que diferem unicamente no sinal.

△

3.5. Revisão

Com o objetivo de reunir o que foi visto nesse capítulo, vamos fazer uma revisão.

Inicialmente vimos na seção 3.1 que o MCP se propõe a aplicar uma transformação às variáveis observadas numa população (x_i), encontrando um novo conjunto de variáveis (y_j), de modo a permitir uma melhor discriminação dos elementos dessa população.

Na seção 3.2 vimos que essa transformação pode ser resumida pela equação:

$$x_i = f_i(y_1, y_2, \dots, y_p) \quad i=1, 2, \dots, n \quad (3.5.1)$$

e que o MCP se limita a estudar os casos em que:

- a) f_i é uma função LINEAR;
- b) $p = n$;
- c) As componentes y_j ($j=1,2,\dots,n$) não sejam correlacionadas entre si;
- d) $\sum_{i=1}^n \lambda_{ji}^2 = 1$ (Obs.: Essa restrição como vimos na seção 3.4 corresponde a exigir que o autovetor λ_j^T seja unitário. Seu significado, como já dissemos, será visto mais adiante).

Vimos, também, que dentro desse conjunto de restrições o MCP irá encontrar as componentes y_j ($j=1,2,\dots,n$) de tal forma que:

- a) A variância de y_1 seja máxima;
- b) A variância de y_2 seja máxima mas y_2 não seja correlacionado com y_1 ;
- c) A variância de y_3 seja máxima mas y_3 não seja correlacionado nem com y_2 nem y_1 ;
- : : : : :
- : : : : :
- : : : : :
- n) A variância de y_n seja máxima mas y_n não seja correlacionado com qualquer y_j ($j=1,2,\dots,(n-1)$).

Este problema denominamos de Problema 1.

Já na seção 3.3 colocamos as restrições e as funções objetivo do Problema 1 de uma forma que possibilitasse a sua resolução de uma forma mais simples. Concluimos então que a 3.^a restrição poderia ser expressa pela equação:

$$\sum_{i=1}^n \sum_{i'=1}^n \ell_{ji} \ell_{j'i'} c_{ii'} = 0 \quad \begin{array}{l} j = 1, 2, \dots, n \\ j' = 1, 2, \dots, (j-1) \end{array} \quad (3.5.2)$$

e que a variância das componentes poderia ser representada por:

$$\text{VAR}[Y_j] = \text{VAR}[y_j] = \sum_{i=1}^n \sum_{i'=1}^n \ell_{ji} \ell_{ji'} c_{ii'} \quad (3.5.3)$$

E finalmente, na seção 3.4, desenvolvemos a resolução do Problema 1 sob a hipótese de a matriz C possuir n autovalores distintos e não nulos e chegamos a conclusão que para resolve-lo bastaria resolver o seguinte sistema de equações:

$$(C - \lambda_j I) \ell_j^T = \emptyset \quad j=1, 2, \dots, n \quad (3.5.4)$$

$$\sum_{i=1}^n \ell_{ji}^2 = 1 \quad j=1, 2, \dots, n \quad (3.5.5)$$

Resolvendo esse sistema teremos então os ℓ_{ji} e consequentemente estaremos de posse da matriz \hat{L} . Nas seções posteriores mostraremos que a inversa de \hat{L} existe de modo que podemos chegar aos ℓ_{ij} , que inicialmente estávamos interessados em encontrar.

3.6. Um Exemplo Resolvido

Temos, a seguir, o exemplo 3.6.1. Os dados utilizados nele são hipotéticos já que visamos somente uma apresentação didática do problema.

Também por esse motivo a forma de resolver o problema não será a mais usual mas sim aquela que permita ver claramente os passos desenvolvidos nas seções anteriores.

Neste exemplo não estaremos preocupados com uma possível interpretação do significado de cada uma das componentes, mas somente com as suas características matemáticas.

EXEMPLO 3.6.1:

PROBLEMA:

Com o objetivo de analisar um conjunto de 10 países obtivemos os valores de três variáveis (que chamaremos de X_1, X_2 e X_3) para cada um desses países. Gostaríamos de através da utilização do MCP, encontrar uma componente que englobando as informações contidas nas três variáveis, discriminasse ao máximo esses 10 países, permitindo ordená-los, segundo essa componente.

Os valores das variáveis encontram-se no quadro

VARIÁVEL \ PAÍS	PAÍS									
	1	2	3	4	5	6	7	8	9	10
X_1	80	30	90	70	35	50	40	80	30	90
X_2	70	40	90	75	40	65	45	90	30	85
X_3	80	40	90	80	30	60	40	80	35	90

Quadro 3.6.1: Valores de 3 variáveis observadas em 10 países.

3.6.1 e sua representação gráfica na figura 3.6.1.

SOLUÇÃO:

Inicialmente vamos encontrar os desvios x_1, x_2, x_3 . Para isso calculamos as médias das variáveis:

$$\bar{X}_1 = 59,5 \quad \bar{X}_2 = 63,0 \quad \bar{X}_3 = 62,5 \quad (3.6.1)$$

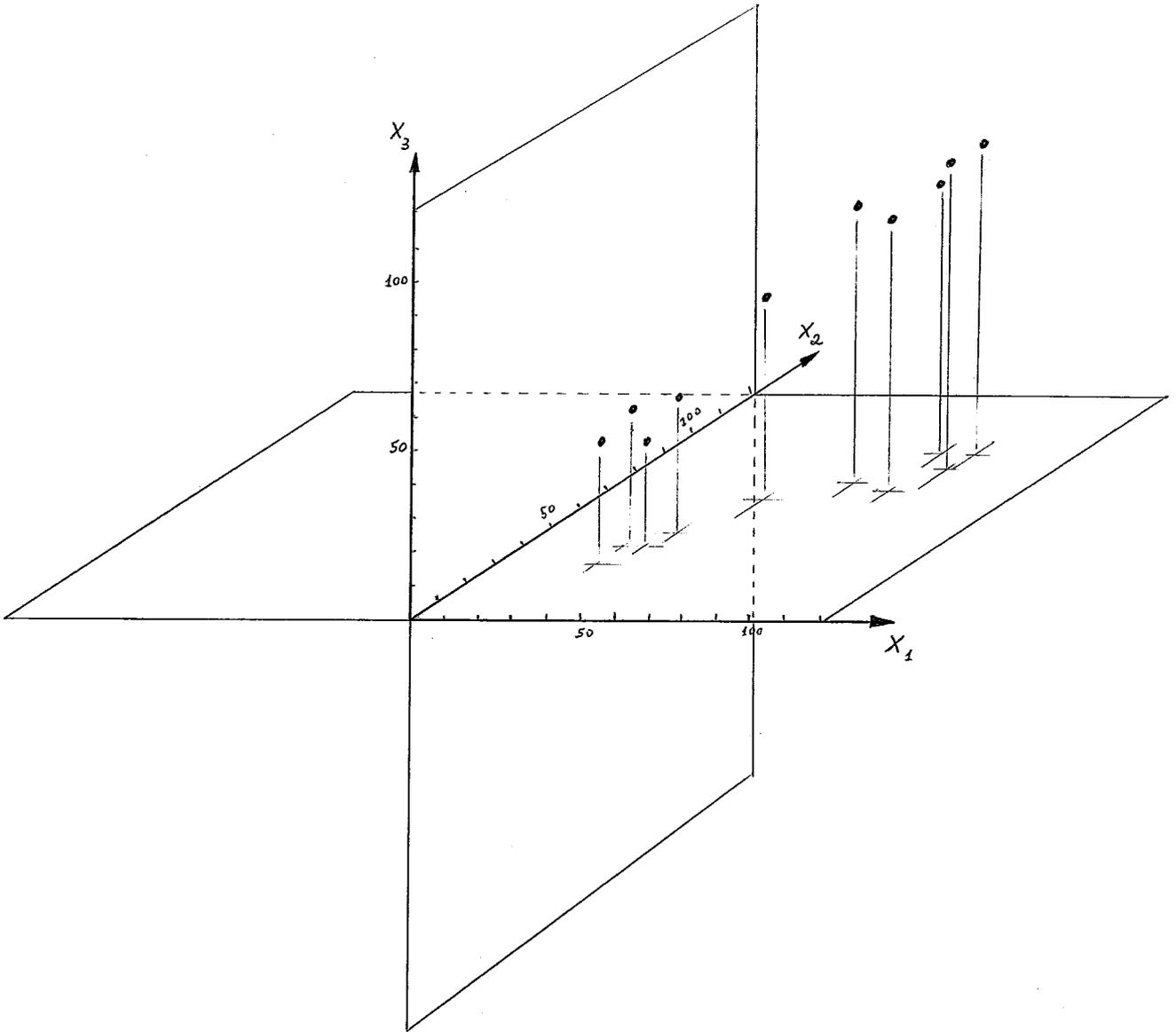


FIGURA 3.6.1: Representação gráfica do quadro 3.6.1

Em seguida, subtraindo de cada $X_i(k)$ ($i=1,2,3$; $k=1,2,\dots,10$) a média \bar{X}_i , encontraremos os $x_i(k)$ que podem ser vistos no quadro 3.6.2 e na figura 3.6.2

PAÍS \ DESVIO	1	2	3	4	5	6	7	8	9	10
x_1	20,5	-29,5	30,5	10,5	-24,5	-9,5	-19,5	20,5	-29,5	30,5
x_2	7,0	-23,0	27,0	12,0	-23,0	2,0	-18,0	27,0	-33,0	22,0
x_3	17,5	-22,5	27,5	17,5	-32,5	-2,5	-22,0	17,5	-27,5	27,5

Quadro 3.6.2: Valores de 3 desvios observados em 10 países.

O próximo passo será encontrar a matriz C. Para isso é necessário encontrar as variâncias e covariâncias de x_1, x_2 e x_3 . Efetutando os cálculos teremos:

$$\text{VAR}[x_1] = 562,25$$

$$\text{COV}[x_1; x_2] = 486,50$$

$$\text{VAR}[x_2] = 461,00$$

$$\text{COV}[x_1; x_3] = 531,25 \quad (3.6.2)$$

$$\text{VAR}[x_3] = 526,25$$

$$\text{COV}[x_2; x_3] = 472,50$$

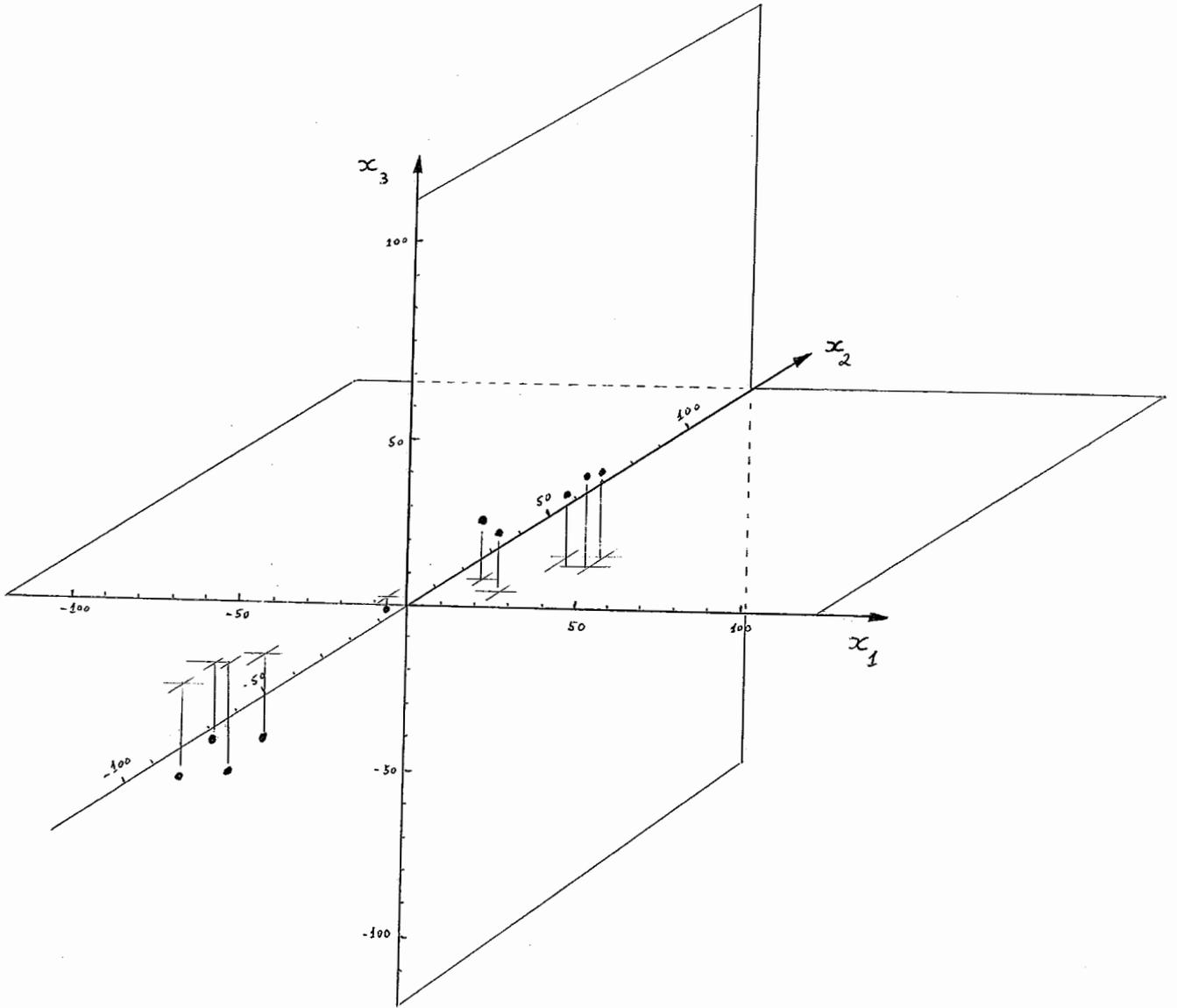


FIGURA 3.6.2: Representação gráfica do quadro 3.6.2

e a Matriz de Covariâncias ficará:

$$C = \begin{pmatrix} 562,25 & 486,50 & 531,25 \\ 486,50 & 461,00 & 472,50 \\ 531,25 & 472,50 & 526,25 \end{pmatrix} \quad (3.6.3)$$

Agora precisamos encontrar os autovalores dessa matriz. (para maior detalhamento dos cálculos que seguem, consultar [7], seção 2.6; [2³], capítulo 9 e [1³], seção 6.3.4).

Para encontrá-los vamos calcular o resultado da seguinte equação:

$$\det(\lambda I - C) = 0 \quad (3.6.4)$$

onde \det simboliza o determinante; λ é o autovalor que se deseja encontrar; I é a Matriz Identidade ($n \times n$) e C a Matriz de Covariâncias.

Desenvolvendo a equação 3.6.4. chegaremos à seguinte equação:

$$\lambda^3 - 1549,5 \lambda^2 + 55.517,5 \lambda - 454.451,1 = 0 \quad (3,6.5)$$

que apresenta as seguintes soluções:

$$\begin{aligned} \lambda_1 &\cong 1.513,01 && (1^\circ \text{ autovalor}) \\ \lambda_2 &\cong 23,95 && (2^\circ \text{ autovalor}) \\ \lambda_3 &\cong 12,54 && (3^\circ \text{ autovalor}) \end{aligned} \quad (3.6.6)$$

De posse dos autovalores, vamos então resolver o sistema de equações 3.5.4 e 3.5.5. Assim teremos:

$$\begin{pmatrix} (562,25-\lambda_j) & 486,50 & 531,25 \\ 486,50 & (461,00-\lambda_j) & 472,50 \\ 531,25 & 472,50 & (526,25-\lambda_j) \end{pmatrix} \begin{pmatrix} \ell_{j1} \\ \ell_{j2} \\ \ell_{j3} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad j=1,2,3 \quad (3.6.7)$$

$$\sum_{i=1}^3 \ell_{ji}^2 = 1 \quad j=1,2,3 \quad (3.6.8)$$

Substituindo o valor de cada autovalor (equações 3.6.6) nas equações 3.6.7 e 3.6.8 teremos:

$$\left. \begin{aligned} - 950,76 \ell_{11} + 486,50 \ell_{12} + 531,25 \ell_{13} &= 0 \\ 486,50 \ell_{11} - 1052,01 \ell_{12} + 472,50 \ell_{13} &= 0 \\ 531,25 \ell_{11} + 472,50 \ell_{12} - 987,76 \ell_{13} &= 0 \\ \ell_{11}^2 + \ell_{12}^2 + \ell_{13}^2 &= 1 \end{aligned} \right\} \text{1º grupo}$$

$$\left. \begin{aligned} 538,30 \ell_{21} + 486,50 \ell_{22} + 531,25 \ell_{23} &= 0 \\ 486,50 \ell_{21} + 437,05 \ell_{22} + 472,50 \ell_{23} &= 0 \\ 531,25 \ell_{21} + 472,50 \ell_{22} + 502,30 \ell_{23} &= 0 \\ \ell_{21}^2 + \ell_{22}^2 + \ell_{23}^2 &= 1 \end{aligned} \right\} \text{2º grupo}$$

$$\left. \begin{aligned} 549,71 \ell_{31} + 486,50 \ell_{32} + 531,25 \ell_{33} &= 0 \\ 486,50 \ell_{31} + 448,46 \ell_{32} + 472,50 \ell_{33} &= 0 \\ 531,25 \ell_{31} + 472,50 \ell_{32} + 513,71 \ell_{33} &= 0 \\ \ell_{31}^2 + \ell_{32}^2 + \ell_{33}^2 &= 1 \end{aligned} \right\} \text{3º grupo}$$

Observamos então que cada grupo de equações está associado a um autovalor, sendo o 1º grupo ao maior autovalor (λ_1), o 2º grupo ao 2º maior autovalor (λ_2) e finalmente o 3º grupo ao 3º maior autovalor (λ_3).

Resolvendo cada grupo de equações encontraremos:

$$\begin{aligned} \ell_{11} &= 0,604 & \ell_{12} &= 0,542 & \ell_{13} &= 0,586 \\ \ell_{21} &= -0,469 & \ell_{22} &= 0,835 & \ell_{23} &= -0,289 \\ \ell_{31} &= -0,646 & \ell_{32} &= -0,098 & \ell_{33} &= 0,758 \end{aligned} \quad (3.6.10)$$

Logo a matriz \hat{L} ficará:

$$\hat{L} = \begin{pmatrix} 0,604 & 0,542 & 0,586 \\ -0,469 & 0,835 & -0,289 \\ -0,646 & -0,098 & 0,758 \end{pmatrix} \quad (3.6.11)$$

Substituindo os valores dos ℓ_{ji} na equação 3.3.1 teremos:

$$\begin{aligned} y_1 &= 0,604x_1 + 0,542x_2 + 0,586x_3 \\ y_2 &= -0,469x_1 + 0,835x_2 - 0,289x_3 \\ y_3 &= -0,646x_1 - 0,098x_2 + 0,758x_3 \end{aligned} \quad (3.6.12)$$

onde y_1 é a 1ª Componente (pois está associada ao maior autovalor (λ_1)), y_2 é a 2ª Componente (pois está associada a λ_2) e y_3 a 3ª Componente (pois está associada a λ_3).

Vamos agora determinar o valor de cada $y_j(k)$ ($j=1, 2, 3$ e $k=1, 2, \dots, 10$), ou seja, o valor de cada componente y_j para cada um dos 10 países.

Efetuada os cálculos com os valores do quadro 3.6.2 e com as equações 3.6.12 encontraremos os valores

que estão no quadro 3.6.3 e cuja representação gráfica está na figura 3.6.3

COMPONENTE \ PAIS	PAIS									
	1	2	3	4	5	6	7	8	9	10
Y_1	26,431	-43,469	49,211	23,101	-46,309	-6,119	-34,719	37,271	-51,819	46,461
Y_2	-8,827	1,133	0,293	0,038	1,678	6,848	0,618	7,873	-5,772	-3,882
Y_3	-0,664	4,256	-1,504	5,306	-6,554	4,046	-2,694	-2,624	1,446	-1,014

Quadro 3.6.3: Valores das 3 componentes para cada um dos 10 países.

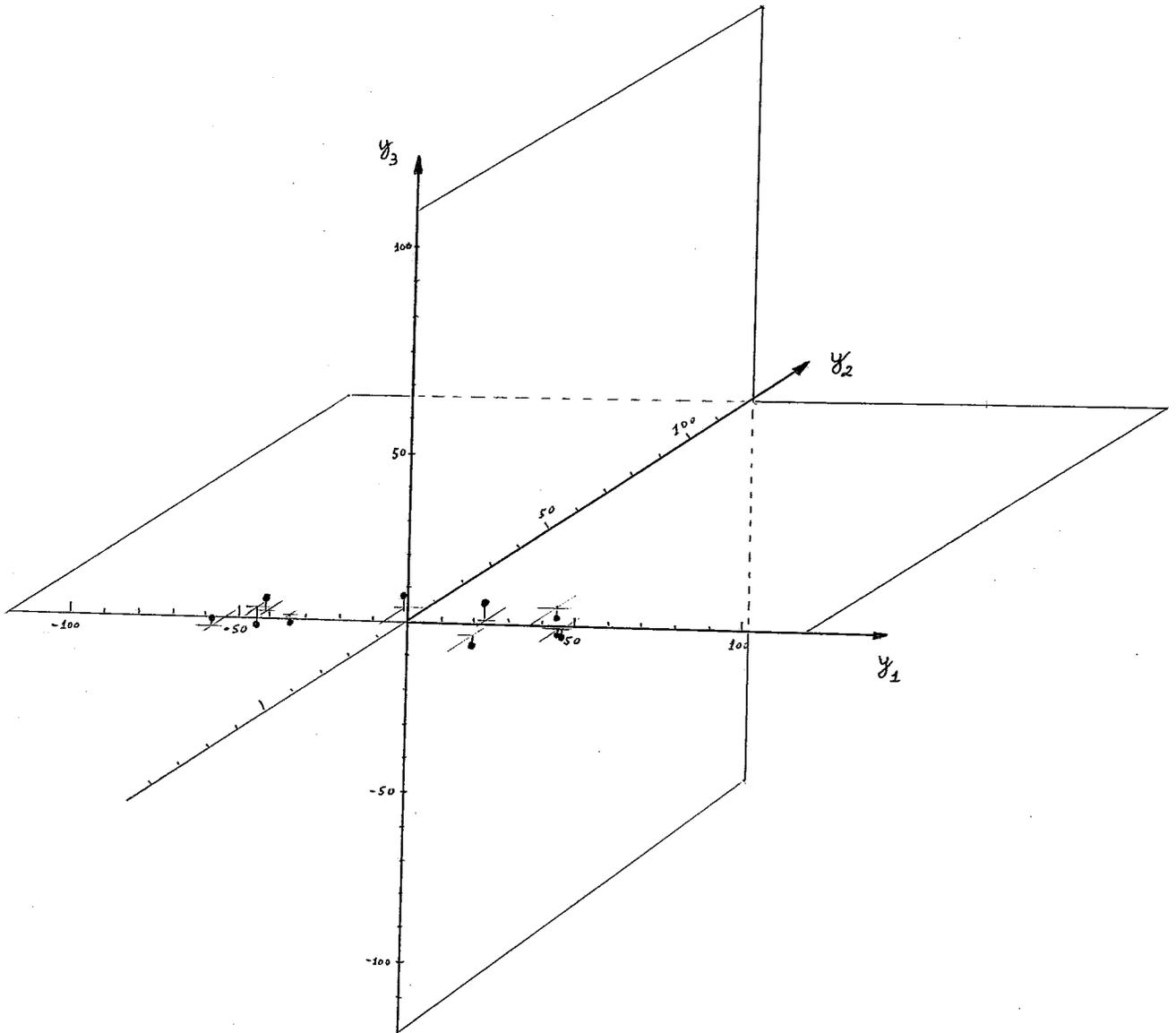


FIGURA 3.63.: Representação gráfica do quadro 3.6.3

Observando o quadro 3.6.3 e a figura 3.6.3 podemos notar (como já era de se esperar pela definição do Problema 1) que a 1.^a Componente, y_1 , é a que mais discrimina os países. Assim a utilizaremos para ordenar os países, como mostra a figura 3.6.4.

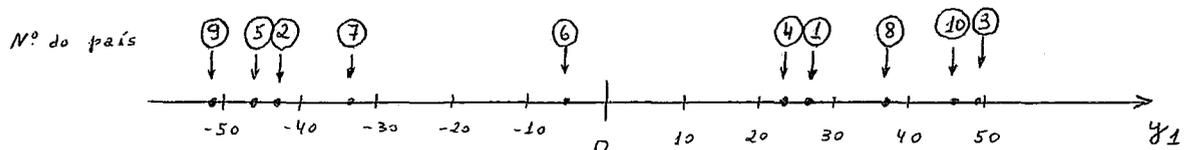


FIGURA 3.6.4: Representação gráfica dos valores $y_1(k)$.

Se este exemplo fosse um caso real poderíamos ainda tentar encontrar algum significado para essa componente ou ainda analisar porque os países 4,1,8,10 e 3 estão no intervalo $20 \rightarrow 50$ e os países 9,5,2 e 3 estão no intervalo $(-52) \rightarrow (-34)$.

É importante lembrar que neste exemplo poderíamos ter outras matrizes \hat{L} que satisfizessem às equações 3.6.9. Como podemos ver o Teorema 3.4.7 nos afirma que para cada j temos duas soluções para y_j que diferem apenas no sinal.

Por exemplo poderíamos ter:

$$\hat{L} = \begin{pmatrix} -0,604 & -0,542 & -0,586 \\ 0,469 & -0,835 & 0,289 \\ -0,646 & -0,098 & 0,758 \end{pmatrix} \quad (3.6.13)$$

ou ainda:

$$\hat{L} = \begin{pmatrix} -0,604 & -0,542 & -0,586 \\ 0,469 & -0,835 & 0,289 \\ 0,646 & 0,098 & -0,758 \end{pmatrix} \quad (3.6.14)$$

Como para cada j temos 2 soluções, ao todo poderíamos ter 2^3 (já que são 3 as componentes) matrizes \hat{L} como solução desse problema. Porém, o fato de utilizarmos uma ou outra solução não é relevante para o problema pois o objetivo do MCP é discriminar os elementos da população e isso independe dos sinais adotados.

3.7. Transformação Ortogonal

Voltando ao modelo vamos agora tirar mais conclusões sobre ele. Vamos, a partir do Problema 1, ver que o que o MCP faz na verdade é uma transformação ortogonal.

Vamos então partir da condição de não correlação dos fatores dada pela equação 3.3.23.

$$\sum_{i=1}^n \sum_{i'=1}^n \lambda_{ji} \lambda_{j'i'} c_{ii'} = 0 \quad \begin{matrix} j=1,2,\dots,n \\ j'=1,2,\dots,(j-1) \end{matrix} \quad (3.7.1)$$

Vamos também lembrar que ao resolver o Problema 1, na seção 3.4, concluimos que o vetor

$$\lambda_j^T = \begin{pmatrix} \lambda_{j1} \\ \lambda_{j2} \\ \vdots \\ \lambda_{jn} \end{pmatrix}$$

é um autovetor da Matriz de Covariâncias o que nos permite, pela definição de autovetor escrever que:

$$C \mathcal{L}_j^T = \lambda_j \mathcal{L}_j^T \quad (3.7.2)$$

onde λ_j é o autovalor ao qual \mathcal{L}_j^T está associado.

Podemos então escrever a equação 3.7.2 da seguinte forma:

$$\begin{pmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & & & \vdots \\ \vdots & & & \vdots \\ c_{n1} & \dots & \dots & c_{nn} \end{pmatrix} \begin{pmatrix} \mathcal{L}_{j1} \\ \mathcal{L}_{j2} \\ \vdots \\ \mathcal{L}_{jn} \end{pmatrix} = \lambda_j \begin{pmatrix} \mathcal{L}_{j1} \\ \mathcal{L}_{j2} \\ \vdots \\ \mathcal{L}_{jn} \end{pmatrix} \quad (3.7.3)$$

ou ainda:

$$\sum_{i'=1}^n \mathcal{L}_{ji'} c_{ii'} = \lambda_j \mathcal{L}_{ji} \quad \begin{matrix} i=1,2,\dots,n \\ j=1,2,\dots,n \end{matrix} \quad (3.7.4)$$

Essa equação em nada se altera se for escrita assim:

$$\sum_{i'=1}^n \mathcal{L}_{j'i'} c_{ii'} = \lambda_{j'} \mathcal{L}_{j'i} \quad \begin{matrix} i=1,2,\dots,n \\ j'=1,2,\dots,n \end{matrix} \quad (3.7.5)$$

Por outro lado podemos escrever a equação 3.7.1 assim:

$$\sum_{i=1}^n \mathcal{L}_{ji} \left[\sum_{i'=1}^n \mathcal{L}_{j'i'} c_{ii'} \right] = 0 \quad \begin{matrix} j=1,2,\dots,n \\ j'=1,2,\dots,(j-1) \end{matrix} \quad (3.7.6)$$

Levando a equação 3.7.5 na 3.7.6 teremos finalmente:

$$\lambda_{j'} \sum_{i=1}^n \mathcal{L}_{ji} \mathcal{L}_{j'i} = 0 \quad \begin{matrix} j=1,2,\dots,n \\ j'=1,2,\dots,(j-1) \end{matrix} \quad (3.7.7)$$

Como pela hipótese da seção 3.4 os autovalores são não nulos (o que acarreta que $\lambda_j \neq 0$) concluímos que:

$$\sum_{i=1}^n \ell_{ji} \ell_{j'i} = 0 \quad \begin{array}{l} j = 1, 2, \dots, n \\ j' = 1, 2, \dots, (j-1) \end{array} \quad (3.7.8)$$

que é equivalente a:

$$\boxed{\sum_{i=1}^n \ell_{ji} \ell_{j'i} = 0} \quad \begin{array}{l} j, j' = 1, 2, \dots, n \\ j \neq j' \end{array} \quad (3.7.9)$$

Sabíamos ainda da seção 3.4 que:

$$\sum_{i=1}^n \ell_{ji}^2 = 1 \quad j = 1, 2, \dots, n \quad (3.7.10)$$

Reunindo as equações 3.7.9 e 3.7.10 teremos:

$$\sum_{i=1}^n \ell_{ji} \ell_{j'i} = \begin{cases} 1 & \text{se } j=j' \quad j = 1, 2, \dots, n \\ 0 & \text{se } j \neq j' \quad j' = 1, 2, \dots, n \end{cases} \quad (3.7.11)$$

Vemos finalmente que as equações 2.5.33 e 3.7.11 são idênticas o que nos leva a concluir que a matriz \hat{L} representa uma transformação ortogonal e logicamente

$$\hat{L} = L^T \quad (3.7.12)$$

ou

$$L^T = \hat{L} \quad (3.7.13)$$

A partir desta idéia de transformação ortogonal, definiremos a seguir um novo problema que denominaremos de Problema 2. Como veremos na formulação a seguir não se trata, neste caso, de um problema de otimização. O que o Problema 2 se propõe a fazer é meramente uma transformação ortogonal, isto é, uma rotação dos eixos de forma que a correlação entre as novas variáveis que obteremos após a rotação (componentes) seja nula.

Mostraremos que o Problema 2 é equivalente ao Problema 1, isto é, ambos tem o mesmo conjunto de soluções. Para isto, primeiramente formularemos o Problema 2 passando depois à sua resolução.

PROBLEMA 2:

Dado que numa população de N elementos foram medidas n características em cada um deles, e que essas características são representadas pelas variáveis x_i ($i=1,2,\dots,n$) encontrar n componentes $y_j = \sum_{i=1}^n \ell_{ji} x_i$ ($j=1,2,\dots,n$) tal que as componentes y_j não sejam correlacionadas entre si e que a matriz \hat{L} (dos ℓ_{ji}) represente uma Transformação Ortogonal.

Vamos escrever o problema da seguinte forma:

PROBLEMA 2:

$$y_j = \sum_{i=1}^n \ell_{ji} x_i \quad j=1,2,\dots,n \quad (3.7.14)$$

$$\text{COR}[y_j; y_{j'}] = 0 \quad j \neq j' \quad (3.7.15)$$

$$\begin{aligned} \sum_{i=1}^n \ell_{ji} \ell_{j'i} &= 1 && \text{se } j = j' \\ &= 0 && \text{se } j \neq j' \end{aligned} \quad (3.7.16)$$

Podemos observar que agora a equação 3.7.15 impõe que as componentes não sejam correlacionadas entre si e a 3.7.16 que a matriz \hat{L} represente uma Transformação Ortogonal.

Vamos então estudar a não correlação das componentes.

Usando o raciocínio da seção 3.3 podemos dizer que:

$$\text{COR}[y_j; y_{j'}] = 0 \quad \text{equivale a} \quad \text{COV}[y_j; y_{j'}] = 0 \quad (3.7.17)$$

$j \neq j'$

Pelas equações 2.3.42 e 3.4.50 temos:

$$\text{COV}[y_j; y_j] = \text{VAR}[y_j] = \lambda_j \quad (3.7.18)$$

Reunindo as equações 3.7.17 e 3.7.18 teremos:

$$\begin{aligned} \text{COV}[y_j; y_{j'}] &= 0 & j \neq j' \\ &= \lambda_j & j = j' \end{aligned} \quad (3.7.19)$$

Vamos então escrever a equação 3.7.19 de uma outra forma:

$$\begin{pmatrix} \text{COV}[y_1; y_1] & \text{COV}[y_1; y_2] & \dots & \text{COV}[y_1; y_n] \\ \text{COV}[y_2; y_1] & \text{COV}[y_2; y_2] & \dots & \dots \\ \vdots & \vdots & \vdots & \vdots \\ \text{COV}[y_n; y_1] & \dots & \dots & \text{COV}[y_n; y_n] \end{pmatrix} = D \quad (3.7.20)$$

onde

$$D = \begin{pmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ 0 & 0 & \lambda_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \lambda_n \end{pmatrix} \quad (3.7.21)$$

(como podemos observar, enquanto C é a Matriz Covariância das Variáveis (x_i) , D é a Matriz Covariância das componentes (y_j)).

Mas, como y_j é um desvio, pela equação 2.3.59A.

$$\text{COV}[y_j; y_{j'}] = E[y_j y_{j'}] \quad (3.7.22)$$

Logo a equação 3.7.20 pode ser escrita assim:

$$\begin{pmatrix} E[y_1 y_1] & E[y_1 y_2] & \dots & E[y_1 y_n] \\ E[y_2 y_1] & E[y_2 y_2] & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ E[y_n y_1] & \dots & \dots & E[y_n y_n] \end{pmatrix} = D \quad (3.7.28)$$

Seja então:

$$\begin{pmatrix} E[y_1 y_1] & E[y_1 y_2] & \dots & E[y_1 y_n] \\ E[y_2 y_1] & E[y_2 y_2] & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ E[y_n y_1] & \dots & \dots & E[y_n y_n] \end{pmatrix} = E \left[\begin{pmatrix} (y_1 y_1) & (y_1 y_2) & \dots & (y_1 y_n) \\ (y_2 y_1) & (y_2 y_2) & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ (y_n y_1) & \dots & \dots & (y_n y_n) \end{pmatrix} \right] \quad (3.7.24)$$

e

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

Agora a equação 3.7.23 pode ser descrita assim:

$$E = \left[\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} (y_1 y_2 \dots y_n) \right] = D \quad (3.7.26)$$

ou simplesmente:

$$E[y y^T] = D \quad (3.7.27)$$

Por outro lado, sendo:

$$\hat{L} = \begin{pmatrix} \ell_{11} & \ell_{12} & \dots & \ell_{1n} \\ \ell_{21} & \ell_{22} & \dots & \vdots \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ \ell_{n1} & \dots & \dots & \ell_{nn} \end{pmatrix} \quad (3.7.28)$$

podemos escrever a equação 3.7.14 assim:

$$y = \hat{L} x \quad (3.7.29)$$

onde

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_n \end{pmatrix} \quad (3.7.30)$$

Substituindo a equação 3.7.29 na 3.7.27 teremos:

$$E[\hat{L} x (\hat{L} x)^T] = D \quad (3.7.31)$$

Mas sabemos da teoria das matrizes (ver [23], teorema 3.3) que:

$$(\hat{L} x)^T = x^T \hat{L}^T \quad (3.7.32)$$

o que nos leva a concluir que a equação 3.7.31 pode ser escrita como:

$$E[\hat{L} x x^T \hat{L}^T] = D \quad (3.7.33)$$

Como a matriz \hat{L} é constante para todos os elementos da população encontramos:

$$\hat{L} E [x x^T] \hat{L}^T = D \tag{3.7.34}$$

Por outro lado,

$$E [x x^T] = E \left[\begin{pmatrix} (x_1 x_1) & (x_1 x_2) & \dots & (x_1 x_n) \\ (x_2 x_1) & \dots & \dots & \vdots \\ \vdots & \dots & \dots & \vdots \\ (x_n x_1) & \dots & \dots & (x_n x_n) \end{pmatrix} \right] = \begin{pmatrix} E[x_1 x_1] & E[x_1 x_2] & \dots & E[x_1 x_n] \\ E[x_2 x_1] & \dots & \dots & \vdots \\ \vdots & \dots & \dots & \vdots \\ E[x_n x_1] & \dots & \dots & E[x_n x_n] \end{pmatrix} \tag{3.7.35}$$

Mas pela equação 2.3.59A temos que:

$$E [x_i x_i] = COV [x_i; x_i] \tag{3.7.35A}$$

logo, pela equação:

$$E [x x^T] = \begin{pmatrix} COV [x_1; x_1] & COV [x_1; x_2] & \dots & COV [x_1; x_n] \\ COV [x_2; x_1] & \dots & \dots & \vdots \\ \vdots & \dots & \dots & \vdots \\ COV [x_n; x_1] & \dots & \dots & COV [x_n; x_n] \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & \dots & \dots & \vdots \\ \vdots & \dots & \dots & \vdots \\ c_{n1} & \dots & \dots & c_{nn} \end{pmatrix} = C \tag{3.7.35b}$$

Reunindo as equações 3.7.34 e 3.7.35b concluímos que a condição de não correlação dos fatores pode ser expressa

pela equação:

$$\boxed{\hat{L} C \hat{L}^T = D} \quad (3.7.36)$$

A condição de ortogonalidade também pode ser colocada na forma matricial. Comparando a equação 3.7.16 com as 2.5.33 e 2.5.27 vemos que ela pode ser expressa por:

$$\boxed{\hat{L}^T \hat{L} = I} \quad (3.7.37)$$

onde I é a Matriz Identidade.

Agora podemos escrever o problema da seguinte forma:

PROBLEMA 2:

$$y = \hat{L} x \quad (3.7.38)$$

$$\hat{L} C \hat{L}^T = D \quad (3.7.39)$$

$$\hat{L}^T \hat{L} = I \quad (3.7.40)$$

Para resolver o problema vamos pré-multiplicar a equação 3.7.39 pela matriz \hat{L}^T . Assim teremos:

$$\hat{L}^T \hat{L} C \hat{L}^T = \hat{L}^T D \quad (3.7.41)$$

Substituindo a equação 3.7.40 na 3.7.41 teremos:

$$C \hat{L}^T = \hat{L}^T D \quad (3.7.43)$$

Seja então:

$$\ell_j = (\ell_{j1}, \ell_{j2}, \dots, \ell_{jn}) \quad (3.7.44)$$

logo podemos escrever a equação 3.7.43 assim:

$$C \ell_j^T = \lambda_j \ell_j^T \quad j=1, 2, \dots, n \quad (3.7.45)$$

o que mais uma vez demonstra que λ_j é um autovalor da matriz C e \mathcal{L}_j^T é um autovetor da matriz C associado a λ_j . Logo o Problema 2 se resume em resolver a equação:

$$(C - \lambda_j I) \mathcal{L}_j^T = 0 \quad (3.7.46)$$

onde o autovetor \mathcal{L}_j^T deverá ser unitário (por causa da equação 3.7.16 ou 3.7.40). Assim, as soluções do Problema 2 deverão atender a:

$$(C - \lambda_j I) \mathcal{L}_j^T = 0 \quad j=1,2,\dots,n \quad (3.7.47)$$

$$\sum_{i=1}^n \mathcal{L}_{ji}^2 = 1 \quad j=1,2,\dots,n \quad (3.7.48)$$

E, finalmente, se compararmos as equações acima com as 3.4.62 e 3.4.63 observamos que os problemas 1 e 2 possuem o mesmo conjunto de soluções.

Acabamos de ver que os problemas 1 e 2 se "equivalem", já que possuem o mesmo conjunto de soluções. Sendo assim uma série de características das Transformações Ortogonais podem ser estendidas ao MCP. * Senão vejamos:

i) A matriz \hat{L} na equação

$$y = \hat{L} x \quad (3.7.49)$$

representa uma Transformação Ortogonal. Logo ela possui inversa, o que confirma a suposição feita na seção 3.2.

ii) Podemos ainda lembrar a equação 2.5.24

$$\hat{L} = L^T \quad (3.7.50)$$

ou seja:

$$\begin{pmatrix} l_{11} & l_{12} & \dots & l_{1n} \\ l_{21} & & & \vdots \\ \vdots & & & \vdots \\ l_{n1} & \dots & \dots & l_{nn} \end{pmatrix} = \begin{pmatrix} l_{11} & l_{21} & \dots & l_{n1} \\ l_{12} & & & \vdots \\ \vdots & & & \vdots \\ l_{1n} & \dots & \dots & l_{nn} \end{pmatrix} \quad (3.7.51)$$

o que nos leva a concluir que:

$$\boxed{l_{ji} = l_{ij}} \quad (3.7.52)$$

iii) Pelo Teorema 2.5.3 a distância entre $x(k)$ e $x(k')$ será a mesma que entre $y(k)$ e $y(k')$.

iv) Como \hat{L} representa uma Transformação Ortogonal, estamos diante de uma Rotação no \mathbb{R}^n (observe que as distâncias entre os $x(k)$ não se alteram (ver iii) e para $x(k) = (0,0,\dots,0)$ teremos $y(k) = (0,0,\dots,0)$).

v) Pelo Teorema 2.5.4 se os vetores x_i formarem uma base ortonormal então os y_j também formarão.

Essas conclusões podem ser observadas no exemplo 3.6.1 como veremos a seguir.

3.8. Comentários Sobre o Exemplo

Vamos então verificar se as características das Transformações Ortogonais aparecem no Exemplo 3.6.1.

i) Naquele exemplo temos:

$$\hat{L} = \begin{pmatrix} 0,604 & 0,542 & 0,586 \\ -0,469 & 0,835 & -0,289 \\ -0,646 & -0,098 & 0,758 \end{pmatrix} \quad (3.8.1)$$

sabemos também que:

$$\hat{L} = L^T \quad (3.8.2)$$

logo:

$$L = \begin{pmatrix} 0,604 & -0,469 & -0,646 \\ 0,542 & 0,835 & -0,098 \\ 0,586 & -0,289 & 0,758 \end{pmatrix} \quad (3.8.3)$$

Efetuando o produto de L por \hat{L} teremos:

$$\begin{pmatrix} 0,604 & -0,469 & -0,646 \\ 0,542 & 0,835 & -0,098 \\ 0,586 & -0,289 & 0,758 \end{pmatrix} \begin{pmatrix} 0,604 & 0,542 & 0,586 \\ -0,469 & 0,835 & -0,289 \\ -0,646 & -0,098 & 0,758 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (3.8.4)$$

o que nos confirma não só que a matriz \hat{L} representa uma Transformação Ortogonal, possuindo portanto uma inversa, como também que a equação 3.7.52 se verifica.

ii) Tomando-se dois $x(k)$ quaisquer, como por exemplo $x(3)$ e $x(10)$ teremos:

$$\begin{aligned} \text{dist}(x(3); x(10)) &= \sqrt{(30,5-30,5)^2 + (27,0-22,0)^2 + (27,5-27,5)^2} \\ &= 5 \end{aligned} \quad (3.8.5)$$

Por outro lado teremos:

$$\begin{aligned} \text{dist}(y(3);y(10)) &= \sqrt{(49,211-46,461)^2 + (0,293+3,882)^2 + (-1,504+1,014)^2} \\ &\cong 5 \end{aligned} \quad (3.8.6)$$

Assim vemos que não se alterou a distância entre os pontos 3 e 10.

iii) E finalmente, na figura 3.8.1 podemos ver a Rotação no \mathbb{R}^3 e que como x_1, x_2, x_3 são ortogonais entre si, os y_1, y_2 e y_3 também são.

Usando o enfoque b dado ao exemplo 2.5.3 (ver figura 2.5.4) podemos dizer que sem alterar os pontos $x(k)$, rotacionamos os eixos x_1, x_2 e x_3 fazendo com que eles tomassem a posição y_1, y_2 e y_3 .

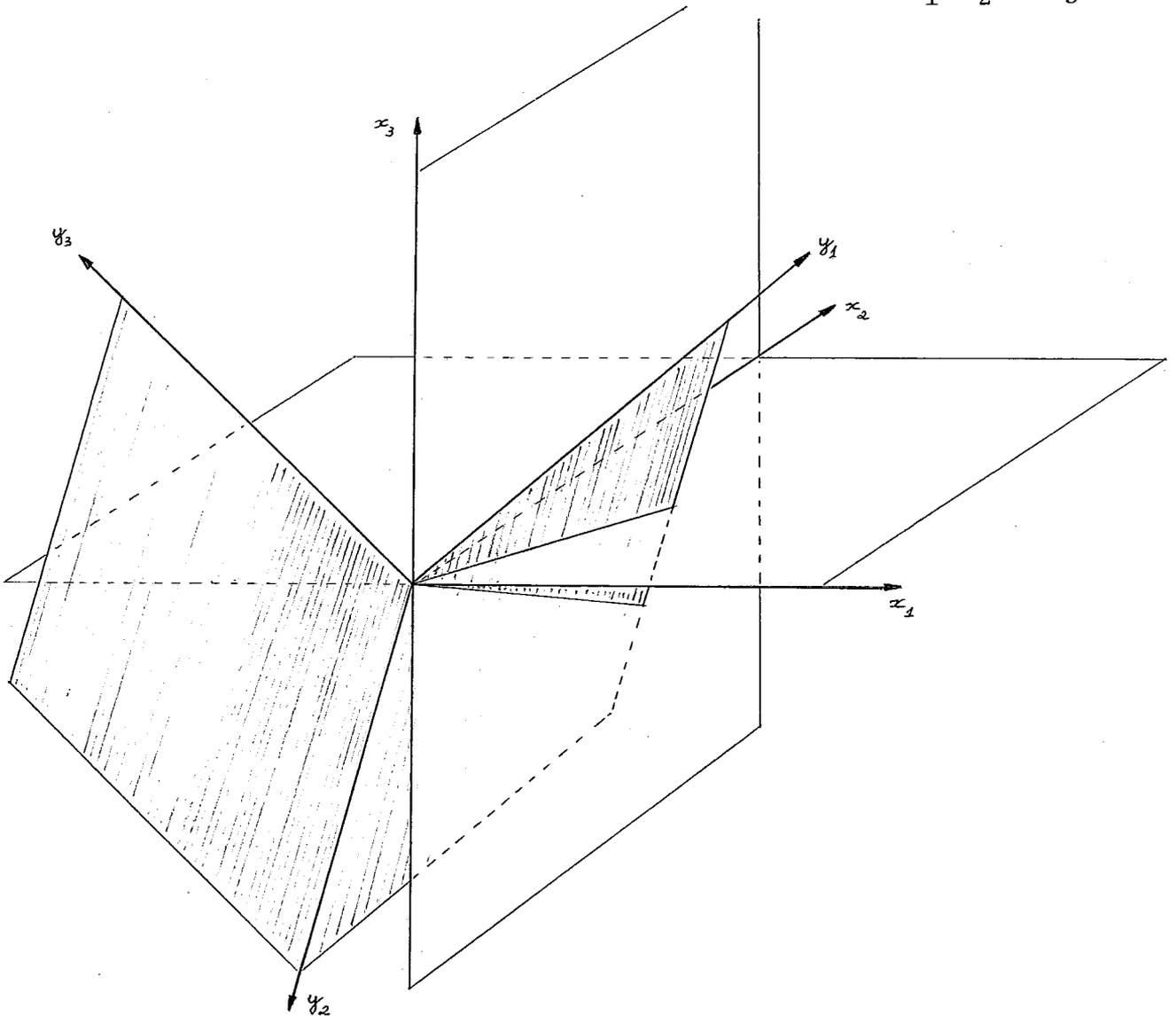


FIGURA 3.8.1: Representação gráfica das componentes y_1, y_2 e y_3 do exemplo 3.6.1.

3.9. A Conservação da Variância

Vamos agora à outra propriedade do MCP.

TEOREMA 3.9.1:
$$\sum_{j=1}^n \text{VAR}[y_j] = \sum_{i=1}^n \text{VAR}[x_i]$$

Demonstração:

Pela equação 3.3.25

$$\text{VAR}[y_j] = \sum_{i=1}^n \sum_{i'=1}^n \ell_{ji} \ell_{ji'} c_{ii'} \quad (3.9.1)$$

$$= \sum_{i=1}^n \sum_{i'=1}^n \ell_{ji} \ell_{ji'} \text{COV}[x_i; x_{i'}] \quad (3.9.2)$$

logo

$$\sum_{j=1}^n \text{VAR}[y_j] = \sum_{j=1}^n \sum_{i=1}^n \sum_{i'=1}^n \ell_{ji} \ell_{ji'} \text{COV}[x_i; x_{i'}] \quad (3.9.3)$$

ou, usando as propriedades dos somatórios:

$$\sum_{j=1}^n \text{VAR}[y_j] = \sum_{i=1}^n \sum_{i'=1}^n \left[\left(\sum_{j=1}^n \ell_{ji} \ell_{ji'} \right) \text{COV}[x_i; x_{i'}] \right] \quad (3.9.4)$$

Mas pela equação 3.7.16.

$$\begin{aligned} \sum_{i=1}^n \ell_{ji} \ell_{j'i} &= 1 & \text{se } j = j' & \quad j=1,2,\dots,n \\ &= 0 & \text{se } j \neq j' & \quad j'=1,2,\dots,n \end{aligned} \quad (3.9.5)$$

Como vimos na seção 2.5, as equações 2.5.33 e 2.5.34 se equivalem logo, podemos escrever que:

$$\begin{aligned} \sum_{j=1}^n \tilde{\ell}_{ji} \tilde{\ell}_{j'i'} &= 1 & \text{se } i = i' & \quad i=1,2,\dots,n \\ &= 0 & \text{se } i \neq i' & \quad i'=1,2,\dots,n \end{aligned} \quad (3.9.6)$$

Substituindo a equação 3.9.6 na 3.9.4 teremos:

$$\sum_{j=1}^n \text{VAR}[y_j] = \sum_{i=1}^n \text{COV}[x_i; x_i] \quad (3.9.7)$$

Mas pela definição de variância e covariância sabemos que:

$$\text{COV}[x_i; x_i] = \text{VAR}[x_i] \quad (3.9.8)$$

Levando então a equação 3.9.8 na 3.9.7 concluímos finalmente que:

$$\sum_{j=1}^n \text{VAR}[y_j] = \sum_{i=1}^n \text{VAR}[x_i] \quad (3.9.9)$$

△

Esse teorema equivale a afirmar que o TRAÇO (soma dos elementos da diagonal principal) da Matriz Covariância das Variáveis (Matriz C) é igual ao TRAÇO da Matriz Covariância das Componentes (Matriz D). Chamaremos esses TRAÇOS de VARIÂNCIA TOTAL DA POPULAÇÃO.

Com isso fica provado que a transformação descrita na seção 3.7 não altera a variância total da população ou seja, a variância total das componentes (y_j) é igual à variância total das variáveis (x_i).

Vimos na seção 2.3 que a variância de uma variável é um indicador de dispersão dessa variável, ou seja, um indicador de quanto, em média, cada valor assumido por essa variável se afasta da média de todos os valores que essa variável assumiu para cada elemento da população em estudo. Sendo assim podemos dizer que a variância é um indicador das diferenças entre as características medidas nos elementos da população em estudo. Por exemplo se $\text{VAR}[x_5] = 0$ podemos ter certeza que todos os $x_5(k)$ ($k=1,2,\dots,N$) são iguais, ou seja, x_5 é uma variável que, por si só, não nos permite distinguir elementos nes-

sa população. Por outro lado se $\text{VAR}[x_7]$ é grande em relação aos valores $x_7(k)$ então, x_7 é uma variável que nos permite, em geral, fazer uma boa distinção entre os elementos em estudo.

Por outro lado, na seção 3.4 (teorema 3.4.5), concluímos que:

$$\lambda_j = \text{VAR}[y_j] \quad j=1, 2, \dots, n \quad (3.9.10)$$

sendo assim, reunindo esse resultado com o teorema 3.9.1 podemos dizer que y_1 engloba $\frac{\lambda_1}{\sum_{j=1}^n \lambda_j}$ da variância total da população,

y_2 engloba $\frac{\lambda_2}{\sum_{j=1}^n \lambda_j}$ e, em geral, y_j engloba $\frac{\lambda_j}{\sum_{j=1}^n \lambda_j}$ da

variância total da população. Como por hipótese, $\lambda_1 > \lambda_2 > \lambda_3 > \dots > \lambda_n > 0$, podemos dizer que as componentes estarão ordenadas segundo a sua "importância", entendendo-se como importância a parte da variância total da população englobada por cada componente. Por isso é comum denominarmos y_1 de 1.^a Componente Principal (ou Componente Principal), y_2 de 2.^a Componente Principal (ou 2.^a Componente) e assim por diante até a n-ésima Componente.

É importante observar que se as variáveis forem padronizadas (ver seção 2.3) teremos:

$$\sum_{j=1}^n \lambda_j = n \quad (3.9.11)$$

já que as variáveis terão variância unitária. Logo, nesse caso, cada componente y_j englobará

$$\frac{\lambda_j}{n}$$

da variância total da população.

3.10. Comentários Sobre o Exemplo

Na seção 3.6, ao resolvermos o problema ali proposto, chegamos ao seguinte resultado:

$$\begin{aligned} \text{VAR}[x_1] &= 562,25 \\ \text{VAR}[x_2] &= 461,00 \\ \text{VAR}[x_3] &= 526,25 \end{aligned} \tag{3.10.1}$$

o que nos leva a concluir que a variância total da população é:

$$\sum_{i=1}^3 \text{VAR}[x_i] = 1549,5 \tag{3.10.2}$$

Por outro lado, a seção 3.6 também nos dá o seguinte:

$$\begin{aligned} \text{VAR}[y_1] &= \lambda_1 \approx 1513,01 \\ \text{VAR}[y_2] &= \lambda_2 \approx 23,95 \\ \text{VAR}[y_3] &= \lambda_3 \approx 12,54 \end{aligned} \tag{3.10.3}$$

o que nos leva a

$$\sum_{j=1}^3 \lambda_j = 1549,5 \tag{3.10.4}$$

Como podemos notar, o teorema 3.9.1 se verifica.

Além disso vamos calcular o quanto da variância total da população cabe a cada componente.

$$y_1 \rightarrow \frac{1513,01}{1549,50} = 97,65\%$$

$$y_2 \rightarrow \frac{23,95}{1549,50} = 1,55\%$$

$$y_3 \rightarrow \frac{12,54}{1549,50} = 0,80\%$$

Isso nos leva a concluir que em presença de y_1 as demais componentes não são "importantes", ou seja, se desejamos discriminar os elementos da população basta utilizarmos y_1 .

É importante observar que, logicamente, essa conclusão não é geral. Certamente existirão casos em que mais de uma componente serão "importantes". Porém nos casos em que uma componente for suficiente para exprimir quase toda a variância total da população, como no exemplo 3.6.1, essa componente poderá ser empregada como um índice segundo o qual os elementos da população possam ser ordenados.

A título de ilustração, temos na figura 3.10.1 a representação de cada valor assumido pelas componentes y_1 , y_2 e y_3 para cada um dos dez países.

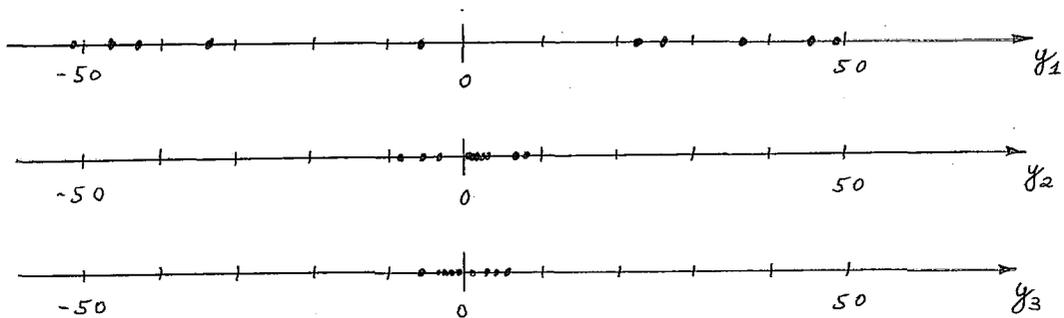


FIGURA 3.10.1: Representação gráfica dos $y(k)$ do exemplo 3.6.1.

Como podemos observar, de fato é a 1.^a Componente, aquela que mais discrimina os países.

3.11. Enfoque Geométrico do Modelo

Voltando ao modelo vamos agora enfoca-lo de uma nova maneira.

Vamos inicialmente considerar os valores $x(k)$ ($k=1,2,\dots,N$) como pontos de um espaço vetorial n -dimensional e definir o seguinte problema.

PROBLEMA 3:

Dado que numa população de N elementos foram medidas n características em cada um deles, e que essas características são representadas pelas variáveis x_i ($i=1,2,\dots,n$) e considerando-se os valores $x(k)$ ($k=1,2,\dots,N$), que são os valores assumidos pelas variáveis x_1, x_2, \dots, x_n para cada elemento k da população, como pontos do espaço vetorial \mathbb{R}^n ; encontrar um conjunto de n retas nesse espaço de tal forma que:

- 1) a soma dos quadrados das distâncias dos pontos $x(k)$ à primeira reta seja mínima.
- 2) a soma dos quadrados das distâncias dos pontos $x(k)$ à segunda reta seja mínima, sendo essa reta ortogonal à primeira.
- 3) a soma dos quadrados das distâncias dos pontos $x(k)$ à terceira reta seja mínima, sendo essa reta ortogonal às anteriores.

n) a soma dos quadrados das distâncias dos pontos $x(k)$ à n -ésima reta seja mínima, sendo essa reta ortogonal às anteriores.

Vamos colocar esse problema em termos matemáticos.

Sejam n retas no \mathbb{R}^n definidas pelos pontos $A_j = (a_{j1}, a_{j2}, \dots, a_{jn})$ ($j = 1, 2, \dots, n$) e pelos cossenos diretores $g_{j1}, g_{j2}, \dots, g_{jn}$ ($j=1, 2, \dots, n$) respectivamente. Seja $Q_j = (Q_{j1}, Q_{j2}, \dots, Q_{jn})$ ($j=1, 2, \dots, n$) um ponto qualquer da j -ésima reta que passaremos a chamar de reta $A_j Q_j$.

Assim, essas retas podem ser expressas pela seguinte equação:

$$Q_{ji} - a_{ji} = g_{ji} t_j \quad \forall i=1, 2, \dots, n \quad (3.11.1)$$

onde t_j é um valor que independe de i .

Utilizando a equação 2.4.39, temos:

$$\text{dist}(x(k); \text{reta } A_j Q_j) = \sqrt{\sum_{i=1}^n (x_i(k) - a_{ji})^2 - \left[\sum_{i=1}^n g_{ji} (x_i(k) - a_{ji}) \right]^2} \quad (3.11.2)$$

Logo a soma dos quadrados das distâncias dos pontos $x(k)$ às retas $A_j Q_j$ será expressa por:

$$S_j = \sum_{k=1}^N \left\{ \sum_{i=1}^n (x_i(k) - a_{ji})^2 - \left[\sum_{i=1}^n g_{ji} (x_i(k) - a_{ji}) \right]^2 \right\}$$

(3.11.3)

Por outro lado a condição de ortogonalidade entre as retas (ver teorema 2.4.2) pode ser expressa por:

$$\boxed{\sum_{i=1}^n g_{ji} g_{j'i} = 0} \quad \begin{array}{l} j = 1, 2, \dots, n \\ j' = 1, 2, \dots, n \\ j \neq j' \end{array} \quad (3.11.4)$$

Se lembrarmos que pelo teorema 2.4.1

$$\boxed{\sum_{i=1}^n g_{ji}^2 = 1} \quad j = 1, 2, \dots, n \quad (3.11.5)$$

podemos escrever o Problema 3 como um problema de otimização dividido em n etapas (como fizemos com o Problema 1).

PROBLEMA 3:

1.^a ETAPA:

$$\text{MINIMIZAR } S_1 \quad (3.11.6)$$

sujeito a

$$\left\{ \sum_{i=1}^n g_{ji} g_{ji} = 1 \right. \quad (3.11.7)$$

2.^a ETAPA:

$$\text{MINIMIZAR } S_2 \quad (3.11.8)$$

sujeito a

$$\left\{ \sum_{i=1}^n g_{2i} g_{1i} = 0 \right. \quad (3.11.9)$$

$$\left\{ \sum_{i=1}^n g_{2i} g_{2i} = 1 \right. \quad (3.11.10)$$

3.^a ETAPA:

$$\text{MINIMIZAR } S_3 \quad (3.11.11)$$

sujeito a

$$\left\{ \begin{array}{l} \sum_{i=1}^n g_{3i} g_{1i} = 0 \end{array} \right. \quad (3.11.12)$$

$$\left\{ \begin{array}{l} \sum_{i=1}^n g_{3i} g_{2i} = 0 \end{array} \right. \quad (3.11.13)$$

$$\left\{ \begin{array}{l} \sum_{i=1}^n g_{3i} g_{3i} = 1 \end{array} \right. \quad (3.11.14)$$

n.^a ETAPA:

$$\text{MINIMIZAR } S_n \quad (3.11.15)$$

sujeito a

$$\left\{ \begin{array}{l} \sum_{i=1}^n g_{ni} g_{1i} = 0 \end{array} \right. \quad (3.11.16)$$

$$\left\{ \begin{array}{l} \sum_{i=1}^n g_{ni} g_{2i} = 0 \end{array} \right. \quad (3.11.17)$$

$$\left\{ \begin{array}{l} \sum_{i=1}^n g_{ni} g_{3i} = 0 \end{array} \right. \quad (3.11.18)$$

⋮

$$\left\{ \begin{array}{l} \sum_{i=1}^n g_{ni} g_{(n-1)i} = 0 \end{array} \right. \quad (3.11.19)$$

$$\left\{ \begin{array}{l} \sum_{i=1}^n g_{ni} g_{ni} = 1 \end{array} \right. \quad (3.11.20)$$

Vamos agora resolver cada uma dessas etapas.

3.12. Resolução do Problema 31.^a ETAPA:

$$\text{MINIMIZAR } S_1 = \sum_{k=1}^N \left\{ \sum_{i=1}^n (x_i(k) - a_{1i})^2 - \left[\sum_{i=1}^n g_{1i} (x_i(k) - a_{1i}) \right]^2 \right\} \quad (3.12.1)$$

sujeito a

$$\sum_{i=1}^n g_{1i} g_{1i} = 1 \quad (3.12.2)$$

Solução:

Utilizando o Método dos Multiplicadores de Lagrange (ver [1^o], [1⁴] e [2⁴]) concluímos que minimizar S_1 equivale a:

$$\begin{aligned} \text{MINIMIZAR } S'_1 = & \sum_{k=1}^N \left\{ \sum_{i=1}^n (x_i(k) - a_{1i})^2 - \left[\sum_{i=1}^n g_{1i} (x_i(k) - a_{1i}) \right]^2 \right\} + \\ & + N\lambda_1 \left[\left(\sum_{i=1}^n g_{1i}^2 \right) - 1 \right] \end{aligned} \quad (3.12.3)$$

onde $N\lambda_1$ é o Multiplicador de Lagrange.

Porém, utilizando propriedades dos somatórios (ver equação 2.2.10) teremos:

$$\left[\sum_{i=1}^n g_{1i} (x_i(k) - a_{1i}) \right]^2 = \sum_{i=1}^n \sum_{i'=1}^n g_{1i} g_{1i'} (x_i(k) - a_{1i}) (x_{i'}(k) - a_{1i'}) \quad (3.12.4)$$

Substituindo a equação 3.12.4 na 3.12.3 e colocando os índices numa forma conveniente teremos:

$$\text{MINIMIZAR } S'_1 = \sum_{k=1}^N \left\{ \sum_{i'=1}^n (x_{i',(k)} - a_{1i'})^2 - \sum_{i'=1}^n \sum_{i''=1}^n g_{1i'} g_{1i''} (x_{i',(k)} - a_{1i'}) (x_{i'',(k)} - a_{1i''}) \right\} + N\lambda_1 \left[\left(\sum_{i'=1}^n g_{1i'}^2 \right) - 1 \right] \quad (3.12.5)$$

Para que S'_1 seja mínima, as suas derivadas parciais em relação aos a_{1i} , g_{1i} e λ_1 deverão ser nulas para todo $i=1,2,\dots$ Vejamos inicialmente em relação a a_{1i} .

$$\frac{\delta S'_1}{\delta a_{1i}} = \sum_{k=1}^N \left\{ -2(x_{i,(k)} - a_{1i}) + 2 \sum_{i'=1}^n g_{1i} g_{1i'} (x_{i',(k)} - a_{1i'}) \right\} \quad i=1,2,\dots,n \quad (3.12.6)$$

ou seja

$$\begin{aligned} \frac{\delta S'_1}{\delta a_{1i}} = & -2 \sum_{k=1}^N x_{i,(k)} + 2 \sum_{k=1}^N a_{1i} + 2 \sum_{k=1}^N \sum_{i'=1}^n g_{1i} g_{1i'} x_{i',(k)} - \\ & - 2 \sum_{k=1}^N \sum_{i'=1}^n g_{1i} g_{1i'} a_{1i'} \quad i=1,2,\dots,n \end{aligned} \quad (3.12.7)$$

Utilizando as propriedades dos somatórios teremos:

$$\begin{aligned} \frac{\delta S'_1}{\delta a_{1i}} = & -2 \sum_{k=1}^N x_{i,(k)} + 2 \sum_{k=1}^N a_{1i} + 2 \sum_{i'=1}^n g_{1i} g_{1i'} \left[\sum_{k=1}^N x_{i',(k)} \right] - \\ & - 2 \sum_{k=1}^N \sum_{i'=1}^n g_{1i} g_{1i'} a_{1i'} \quad i=1,2,\dots,n \end{aligned} \quad (3.12.8)$$

Mas como $x_{i,(k)}$ é um desvio (ver seção 2.3), teremos:

$$E[x_i] = \frac{1}{N} \sum_{k=1}^N x_i(k) = 0 \quad i=1,2,\dots,n \quad (3.12.9)$$

logo

$$\frac{\delta S'_1}{\delta a_{1i}} = 2 \sum_{k=1}^N a_{1i} - 2 \sum_{k=1}^N \sum_{i'=1}^n g_{1i} g_{1i'} a_{1i'} \quad i=1,2,\dots,n \quad (3.12.10)$$

ou

$$\frac{\delta S'_1}{\delta a_{1i}} = 2 N a_{1i} - 2 N g_{1i} \sum_{i'=1}^n g_{1i'} a_{1i'} \quad i=1,2,\dots,n \quad (3.12.11)$$

Anulando $\frac{\delta S'_1}{\delta a_{1i}}$ $i=1,2,\dots,n$ temos:

$$a_{1i} = g_{1i} \sum_{i'=1}^n g_{1i'} a_{1i'} \quad i=1,2,\dots,n \quad (3.12.12)$$

Fazendo $t^* = \sum_{i=1}^n g_{1i} a_{1i}$, e observando que

este somatório independe de i obtemos $a_{1i} = g_{1i} t^*$. Substituindo este valor na equação 3.11.1 para $j = 1$ obtemos:

$$Q_{1i} = g_{1i}(t - t^*) \quad i=1,2,\dots,n \quad (3.12.13)$$

Fazendo $t = t^*$ na equação 3.12.13 obtemos o ponto $Q_1^* = (0,0,\dots,0)$, ou seja, a reta A_1Q_1 passa pelo ponto de coordenadas $(0,0,\dots,0)$.

Esse resultado é importante pois nos mostra que a reta A_1Q_1 passa pela origem do espaço vetorial (R^n) ao qual os $x_i(k)$ pertencem. Como estamos utilizando $x_i(k)$, e não $X_i(k)$, a origem nada mais é do que a média dos $X_i(k)$, ou seja \bar{X}_i .

Podemos agora redefinir a função objetivo dessa 1.^a etapa (equação 3.12.1) tomando o ponto Q_1^* como referência para o cálculo das distâncias. Evidentemente isso seria o mesmo que simplesmente redefinir o ponto A_1 que passaria a ser a origem $(0,0,\dots,0)$. Fazendo diretamente essa mudança na equação 3.12.5, isto é, fazendo $a_{1i} = 0$ para $i=1,2,\dots,n$ temos:

$$\begin{aligned} \text{MINIMIZAR } S_1'' = & \sum_{k=1}^N \left\{ \sum_{i'=1}^n (x_{i'}(k))^2 - \sum_{i'=1}^n \sum_{i''=1}^n g_{1i'} g_{1i''} x_{i'}(k) x_{i''}(k) \right\} + \\ & + N\lambda_1 \left[\left(\sum_{i'=1}^n g_{1i'}^2 \right) - 1 \right] \end{aligned} \quad (3.12.14)$$

Vamos então calcular a derivada parcial de S_1'' em relação aos g_{1i} .

$$\frac{\delta S_1''}{\delta g_{1i}} = -2 \sum_{k=1}^N \sum_{i'=1}^n g_{1i'} x_{i'}(k) x_i(k) + 2N\lambda_1 g_{1i} \quad (3.12.15)$$

$i=1,2,\dots,n$

ou ainda

$$\frac{\delta S_1''}{\delta g_{1i}} = -2 \left[\sum_{i'=1}^n g_{1i'} \left(\sum_{k=1}^N x_{i'}(k) x_i(k) \right) \right] + 2N\lambda_1 g_{1i} \quad (3.12.16)$$

$i=1,2,\dots,n$

Mas como os $x_i(k)$ são desvios teremos: (ver seção 2.3)

$$\sum_{k=1}^N (x_i(k) x_{i'}(k)) = N \cdot \text{COV}[x_i; x_{i'}] = N c_{ii'} \quad (3.12.17)$$

logo

$$\frac{\delta S_1''}{\delta g_{1i}} = -2N \sum_{i'=1}^n g_{1i'} c_{ii'} + 2N\lambda_1 g_{1i} \quad (3.12.18)$$

$i=1,2,\dots,n$

Assim, para que $\frac{\delta S_1''}{\delta g_{1i}}$ seja nulo, teremos:

$$\boxed{\sum_{i'=1}^n g_{1i'} c_{ii'} - \lambda_1 g_{1j} = 0} \quad i=1,2,\dots,n \quad (3.12.19)$$

Finalmente, encontrando a derivada parcial de S_1 em relação a λ_1 teremos:

$$\frac{\delta S_1'}{\delta \lambda_1} = N \left(\sum_{i'=1}^n g_{1i'}^2 - 1 \right) \quad (3.12.20)$$

Logicamente, para $\frac{\delta S_1'}{\delta \lambda_1}$ se anular é necessário

que:

$$\boxed{\sum_{i'=1}^n g_{1i'}^2 = 1} \quad (3.12.21)$$

Para resolver então o problema vamos reunir as equações 3.12.19 e 3.12.21. Assim a solução do problema será a solução do seguinte sistema de equações:

$$\sum_{i'=1}^n g_{1i'} c_{ii'} = \lambda_1 g_{1i} \quad i=1,2,\dots,n \quad (3.12.23)$$

$$\sum_{i'=1}^n g_{1i'}^2 = 1 \quad (3.12.24)$$

OBSERVAÇÃO:

Podemos notar que no sistema de equações acima temos:

n equações independentes em 3.12.23

1 equação independente em 3.12.24

e também

n incógnitas g_{1i}

1 incógnita λ_1

sendo assim temos $n+1$ equação e $n+1$ incógnitas, o que nos leva a um sistema determinado, onde podemos calcular diretamente os valores g_{1i} e λ_1 . Uma vez determinados estes valores, temos a equação da reta correspondente pois sabemos que:

- 1) a reta passa pela origem, o que é consequência da equação 3.12.11
- 2) temos os cossenos diretores g_{1i} .

Utilizando equações 3.12.23-4 podemos tirar duas conclusões:

- 1.^a) λ_1 é um autovalor da matriz C
- 2.^a) λ_1 é o maior autovalor da matriz C .

Senão vejamos:

TEOREMA 3.12.1: λ_1 é um autovalor da matriz C .

Demonstração:

Seja $g_1 = (g_{11}, g_{12}, \dots, g_{1n})$

e

$$C = \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & & & \vdots \\ \vdots & & & \vdots \\ c_{n1} & \dots & \dots & c_{nn} \end{pmatrix}$$

Assim podemos escrever a equação 3.12.23 assim:

$$C g_1^T = \lambda_1 g_1^T \quad (3.12.25)$$

que nada mais é do que a definição de autovalor. Logo λ_1 é um autovalor da matriz C e g_1^T é um autovetor da matriz C associado ao autovalor λ_1 . △

Com relação à segunda conclusão teremos o seguinte teorema:

TEOREMA 3.12.2: λ_1 é o maior autovalor da matriz C.

Demonstração:

Multiplicando-se a equação 3.12.23 por g_{1i} e somando para todo $i=1,2,\dots,n$ teremos:

$$\sum_{i=1}^n \sum_{i'=1}^n g_{1i} g_{1i'} c_{ii'} = \lambda_1 \sum_{i=1}^n g_{1i}^2 \quad (3.12.26)$$

Mas pela equação 3.12.24

$$\sum_{i=1}^n g_{1i}^2 = 1 \quad (3.12.27)$$

logo

$$\sum_{i=1}^n \sum_{i'=1}^n g_{1i} g_{1i'} c_{ii'} = \lambda_1 \quad (3.12.28)$$

Levando a equação 3.12.17 na 3.12.28 teremos:

$$\sum_{i=1}^n \sum_{i'=1}^n g_{1i} g_{1i'} \left[\sum_{k=1}^n (x_i(k)x_{i'}(k)) \right] = N\lambda_1 \quad (3.12.29)$$

ou

$$\sum_{k=1}^N \sum_{i=1}^n \sum_{i'=1}^n g_{1i} g_{1i'} x_i(k) x_{i'}(k) = N\lambda_1 \quad (3.12.30)$$

Levando as equações 3.12.27 e 2.12.30 na 3.12.14

teremos:

$$\text{MINIMIZAR } S''_i = \left[\sum_{k=1}^N \sum_{i'=1}^n (x_{i'}(k))^2 \right] - N\lambda_1 \quad (3.12.31)$$

Logo, como os $x_{i'}(k)$ são parâmetros da população em estudo, minimizar S''_i é o mesmo que maximizar λ_1 . Como pelo teorema 3.11.1, λ_1 é um autovalor da matriz C, então ele será o maior autovalor dessa matriz. △

Voltando novamente às duas conclusões, verificamos que para S_1 (soma dos quadrados das distâncias dos pontos $x(k)$ à reta A_1Q_1) ser mínima, é necessário que λ_1 seja o maior autovalor da matriz C e que g_1^T seja o autovetor unitário associado a λ_1 .

Assim a reta A_1Q_1 será definida por passar na origem e ter cossenos diretores que sejam componentes do autovetor unitário da matriz C associado ao maior autovalor dessa matriz.

Fim da 1.^a etapa □

2ª ETAPA:

$$\text{MINIMIZAR } S_2 = \sum_{k=1}^N \left\{ \sum_{i=1}^n (x_i(k) - a_{2i})^2 - \left[\sum_{i=1}^n g_{2i} (x_i(k) - a_{2i}) \right]^2 \right\} \quad (3.12.32)$$

sujeito a

$$\sum_{i=1}^n g_{2i} g_{1i} = 0 \quad (3.12.33)$$

$$\sum_{i=1}^n g_{2i} g_{2i} = 1 \quad (3.12.34)$$

Solução:

Utilizando mais uma vez o Método dos Multiplicadores de Lagrange concluímos que minimizar S_2 equivale a:

$$\begin{aligned} \text{MINIMIZAR } S'_2 = & \sum_{k=1}^N \left\{ \sum_{i=1}^n (x_i(k) - a_{2i})^2 - \left[\sum_{i=1}^n g_{2i} (x_i(k) - a_{2i}) \right]^2 \right\} + \\ & + N\lambda_2 \left[\left(\sum_{i=1}^n g_{2i}^2 \right) - 1 \right] + \alpha_2 \left[\sum_{i=1}^n g_{2i} g_{1i} \right] \end{aligned} \quad (3.12.35)$$

onde $N\lambda_2$ e α_2 são os Multiplicadores de Lagrange.

Porém, utilizando propriedades dos somatórios (ver equação 2.2.10) teremos:

$$\left[\sum_{i=1}^n g_{2i} (x_i(k) - a_{2i}) \right]^2 = \sum_{i=1}^n \sum_{i'=1}^n g_{2i} g_{2i'} (x_i(k) - a_{2i})(x_{i'}(k) - a_{2i'}) \quad (3.12.36)$$

Substituindo a equação 3.12.36 na 3.12.35 e colocando os índices numa forma mais conveniente teremos:

$$\begin{aligned}
 \text{MINIMIZAR } S'_2 = & \sum_{k=1}^N \left\{ \sum_{i'=1}^n (x_{i',(k)} - a_{2i'})^2 - \sum_{i'=1}^n \sum_{i''=1}^n g_{2i'} g_{2i''} (x_{i',(k)} - \right. \\
 & \left. - a_{2i'}) (x_{i'',(k)} - a_{2i''}) \right\} + N\lambda_2 \left[\left(\sum_{i'=1}^n g_{2i'}^2 \right) - 1 \right] + \\
 & + \alpha_2 \left[\sum_{i'=1}^n g_{2i'} g_{1i'} \right] \quad (3.12.37)
 \end{aligned}$$

Para que S'_2 seja mínima, as suas derivadas parciais em relação aos $a_{2i}, g_{2i}, \lambda_2$ e α_2 terão que ser nulas para todo $i=1, 2, \dots, n$.

Em relação aos a_{2i} o raciocínio é análogo ao desenvolvido na etapa anterior e nos levará a concluir que a reta A_2Q_2 também passa pela origem do espaço vetorial ao qual os $x(k)$ pertencem.

Utilizando esse fato e sendo A_2 um ponto genérico da reta A_2Q_2 , vamos fazer $A_2 = (0, 0, 0, \dots, 0)$. Agora nosso objetivo será:

$$\begin{aligned}
 \text{MINIMIZAR } S''_2 = & \sum_{k=1}^N \left\{ \sum_{i'=1}^n (x_{i',(k)})^2 - \sum_{i'=1}^n \sum_{i''=1}^n g_{2i'} g_{2i''} x_{i',(k)} x_{i'',(k)} \right\} + \\
 & + N\lambda_2 \left[\left(\sum_{i'=1}^n g_{2i'}^2 \right) - 1 \right] + \alpha_2 \left[\sum_{i'=1}^n g_{1i'} g_{2i'} \right] \quad (3.12.39)
 \end{aligned}$$

Encontrando as demais derivadas parciais como fizemos na primeira etapa, teremos:

$$\left\{ \begin{array}{l} \frac{\delta S''_2}{\delta g_{2i}} = -2N \sum_{i'=1}^n g_{2i'} c_{ii'} + 2N\lambda_2 g_{2i} + \alpha_2 g_{1i} \\ \hspace{20em} i=1,2,\dots,n \end{array} \right. \quad (3.12.40)$$

$$\left\{ \begin{array}{l} \frac{\delta S''_2}{\delta \lambda_2} = \sum_{i'=1}^n (g_{2i'})^2 - 1 \end{array} \right. \quad (3.12.41)$$

$$\left\{ \begin{array}{l} \frac{\delta S''_2}{\delta \alpha_2} = \sum_{i'=1}^n g_{2i'} g_{1i'} \end{array} \right. \quad (3.12.42)$$

Anulando as derivadas parciais chegamos então a:

$$\left\{ \begin{array}{l} 2N \sum_{i'=1}^n g_{2i'} c_{ii'} - 2N\lambda_2 g_{2i} - \alpha_2 g_{1i} = 0 \\ \hspace{20em} i=1,2,\dots,n \end{array} \right. \quad (3.12.43)$$

$$\left\{ \begin{array}{l} \sum_{i'=1}^n g_{2i'}^2 = 1 \end{array} \right. \quad (3.12.44)$$

$$\left\{ \begin{array}{l} \sum_{i'=1}^n g_{2i'} g_{1i'} = 0 \end{array} \right. \quad (3.12.45)$$

Dessas equações, mais uma vez, podemos tirar duas conclusões:

- 1.^a) λ_2 é um autovalor da matriz C.
- 2.^a) λ_2 é o segundo maior autovalor da matriz C.

OBSERVAÇÃO:

É importante relembrar que por hipótese (ver seção 3.4) a matriz C possui n autovalores diferentes e não nulos. Os casos em que isso não ocorrer serão estudados na seção 3.17.

Passamos agora à confirmação das conclusões.

TEOREMA 3.12.3: λ_2 é um autovalor da matriz C

Demonstração:

Multiplicando a equação 3.12.43 por g_{1i} e somando para todo $i=1,2,\dots,n$ teremos:

$$2N \sum_{i=1}^n \sum_{i'=1}^n g_{2i'} g_{1i} c_{ii'} - 2N\lambda_2 \sum_{i=1}^n g_{2i} g_{1i} - \alpha_2 \sum_{i=1}^n g_{1i}^2 = 0 \quad (3.12.46)$$

Substituindo as equações 3.12.24 e 3.12.45 na 3.12.46 teremos:

$$2N \sum_{i=1}^n \sum_{i'=1}^n g_{2i'} g_{1i} c_{ii'} = \alpha_2 \quad (3.12.47)$$

ou, o que é equivalente

$$2N \sum_{i=1}^n g_{2i} \left[\sum_{i'=1}^n g_{1i'} c_{ii'} \right] = \alpha_2 \quad (3.12.48)$$

Mas pela equação 3.12.23 temos:

$$\sum_{i'=1}^n g_{1i'} c_{ii'} = \lambda_1 g_{1i} \quad (3.12.49)$$

Levando esse resultado na equação 3.12.48 teremos:

$$2N\lambda_1 \sum_{i=1}^n g_{2i} g_{1i} = \alpha_2 \quad (3.12.50)$$

Substituindo então a equação 3.12.45 na 3.12.50 concluímos que:

$$\boxed{\alpha_2 = 0} \quad (3.12.5)$$

Agora podemos escrever a equação 3.12.43 da seguinte forma:

$$\sum_{i'=1}^n g_{2i'} c_{ii'} = \lambda_2 g_{2i} \quad i=1,2,\dots,n \quad (3.12.52)$$

Seja:

$$g_2 = (g_{21}, g_{22}, \dots, g_{2n})$$

e

$$C = \begin{pmatrix} c_{11} & c_{12} & c_{13} & \dots & c_{1n} \\ c_{21} & & & & \vdots \\ \vdots & & & & \vdots \\ \vdots & & & & \vdots \\ c_{n1} & \dots & \dots & \dots & c_{nn} \end{pmatrix}$$

Podemos então escrever a equação 3.12.52 assim:

$$C g_2^T = \lambda_2 g_2^T \quad (3.12.53)$$

que nada mais é do que a definição de autovalor. Logo λ_2 é um autovalor da matriz C e g_2^T é um autovetor da matriz C associado ao autovalor λ_2 . \triangle

Com relação à segunda conclusão teremos o seguinte teorema.

TEOREMA 3.12.4: λ_2 é o segundo maior autovalor da matriz C .

Demonstração:

Multiplicando-se a equação 3.12.52 por g_{2i} e somando para todo $i=1,2,\dots,n$ teremos:

$$\sum_{i=1}^n \sum_{i'=1}^n g_{2i} g_{2i'} c_{ii'} = \lambda_2 \sum_{i=1}^n g_{2i}^2 \quad (3.12.54)$$

mas pela equação 3.12.44

$$\sum_{i=1}^n g_{2i}^2 = 1 \quad (3.12.55)$$

logo

$$\sum_{i=1}^n \sum_{i'=1}^n g_{2i} g_{2i'} c_{ii'} = \lambda_2 \quad (3.12.56)$$

Levando a equação 3.12.17 na 3.12.56 teremos:

$$\sum_{i=1}^n \sum_{i'=1}^n g_{2i} g_{2i'} \left[\sum_{k=1}^N (x_i(k) x_{i'}(k)) \right] = N\lambda_2 \quad (3.12.57)$$

ou

$$\sum_{k=1}^N \sum_{i=1}^n \sum_{i'=1}^n g_{2i} g_{2i'} x_i(k) x_{i'}(k) = N\lambda_2 \quad (3.12.58)$$

Levando as equações 3.12.44-5 e 3.12.58 na 3.12.39 chegamos a:

$$\text{MINIMIZAR } S_2'' = \left[\sum_{k=1}^n \sum_{i'=1}^n x_{i'}(k)^2 \right] - N\lambda_2 \quad (3.12.59)$$

Logo, como os $x_{i'}(k)$ e N são parâmetros da população em estudo, minimizar S_2'' é o mesmo que maximizar λ_2 . Como pelo teorema 3.12.3 λ_2 é um autovalor da matriz C , teremos que escolher λ_2 entre os autovalores da matriz C . A primeira vista poderia parecer que a escolha recairia sobre o

maior autovalor da matriz C . Porém se isso ocorresse λ_2 seria igual a λ_1 e conseqüentemente os autovetores g_2^T e g_1^T , associados respectivamente a λ_2 e λ_1 seriam iguais, o que implicaria numa violação da equação 3.12.45. Sendo assim o maior autovalor da matriz C que poderá ser escolhido, neste caso, será o segundo maior autovalor da matriz C . △

Voltando novamente às duas conclusões, verificamos que para S_2 (soma dos quadrados das distâncias dos pontos $x(k)$ à reta A_2Q_2) ser mínima sendo essa reta ortogonal a A_1Q_1 é necessário que λ_2 seja o segundo maior autovalor da matriz C e que g_2^T seja o autovetor unitário associado a λ_2 .

Assim a reta A_2Q_2 será definida por passar pela origem e ter cossenos diretores que sejam componentes do autovetor unitário da matriz C associado ao segundo maior autovalor dessa matriz.

Fim da 2ª etapa □

Vamos agora resolver uma etapa genérica do Problema 3.

j.^a ETAPA (j=3, 4, ..., n)

$$\text{MINIMIZAR } S_j = \sum_{k=1}^N \left\{ \sum_{i=1}^n (x_i(k) - a_{ji})^2 - \left[\sum_{i=1}^n g_{ji} (x_i(k) - a_{ji}) \right]^2 \right\} \quad (3.12.60)$$

sujeito a

$$\sum_{i=1}^n g_{ji} g_{1i} = 0 \quad (3.12.61)$$

$$\sum_{i=1}^n g_{ji} g_{2i} = 0 \quad (3.12.62)$$

⋮

$$\sum_{i=1}^n g_{ji} g_{(j-1)i} = 0 \quad (3.12.63)$$

$$\sum_{i=1}^n g_{ji} g_{ji} = 1 \quad (3.12.64)$$

Solução:

Utilizando o Método dos Multiplicadores de Lagrange, como nas outras etapas, o problema ficará:

$$\begin{aligned} \text{MINIMIZAR } S'_j = & \sum_{k=1}^N \left\{ \sum_{i=1}^n (x_i(k) - a_{ji})^2 - \left[\sum_{i=1}^n g_{ji} (x_i(k) - a_{ji}) \right]^2 \right\} + \\ & + N\lambda_j \left[\left(\sum_{i=1}^n g_{ji}^2 \right) - 1 \right] + \alpha_{j1} \left[\sum_{i=1}^n g_{ji} g_{1i} \right] + \\ & + \alpha_{j2} \left[\sum_{i=1}^n g_{ji} g_{2i} \right] + \dots + \alpha_{j(j-1)} \left[\sum_{i=1}^n g_{ji} g_{(j-1)i} \right] \end{aligned} \quad (3.12.65)$$

onde $N\lambda_j, \alpha_{j1}, \alpha_{j2}, \dots, \alpha_{j(j-1)}$ são os Multiplicadores de Lagrange.

Mais uma vez, utilizando propriedades dos somatórios (ver equação 2.2.10) teremos:

$$\left[\sum_{i=1}^n g_{ji} (x_i(k) - a_{ji}) \right]^2 = \sum_{i=1}^n \sum_{i'=1}^n g_{ji} g_{ji'} (x_i(k) - a_{ji}) (x_{i'}(k) - a_{ji'}) \quad (3.12.66)$$

Substituindo a equação 3.12.66 na 3.12.65 e colocando os índices numa forma mais conveniente teremos:

$$\begin{aligned} \text{MINIMIZAR } S_j^! &= \sum_{k=1}^N \left\{ \sum_{i=1}^n (x_{i'}(k) - a_{ji'})^2 - \sum_{i=1}^n \sum_{i''=1}^n g_{ji'} g_{ji''} (x_{i'}(k) - a_{ji'}) (x_{i''}(k) - a_{ji''}) \right\} + \\ &+ N\lambda_j \left[\left(\sum_{i=1}^n g_{ji}^2 \right) - 1 \right] + \alpha_{j1} \left[\sum_{i=1}^n g_{ji} g_{1i} \right] + \\ &+ \alpha_{j2} \left[\sum_{i=1}^n g_{ji} g_{2i} \right] + \dots + \alpha_{j(j-1)} \left[\sum_{i=1}^n g_{ji} g_{(j-1)i} \right] \end{aligned} \quad (3.12.67)$$

Para que $S_j^!$ seja mínima, as suas derivadas parciais em relação aos $a_{ji}, g_{ji}, \lambda_2, \alpha_{j1}, \alpha_{j2}, \dots, \alpha_{j(j-1)}$ terão que ser nulas para todo $i=1, 2, \dots, n$.

Em relação aos a_{ji} o raciocínio é análogo ao desenvolvido na primeira etapa e nos levará a concluir que a reta $A_j Q_j$ também passa pela origem do espaço vetorial ao qual os $x_i(k)$ pertencem.

Utilizando esse fato e sendo A_j um ponto genérico da reta $A_j Q_j$ vamos fazer $A_j = (0, 0, \dots, 0)$. Agora nosso objetivo será:

$$\begin{aligned}
 \text{MINIMIZAR } S''_j = & \sum_{k=1}^N \left\{ \sum_{i'=1}^n (x_{i'}(k))^2 - \sum_{i'=1}^n \sum_{i''=1}^n g_{ji'} g_{ji''} x_{i'}(k) x_{i''}(k) \right\} + \\
 & + N\lambda_j \left[\left(\sum_{i'=1}^n g_{ji'}^2 \right) - 1 \right] + \alpha_{j1} \left[\sum_{i=1}^n g_{ji} g_{1i} \right] + \\
 & + \alpha_{j2} \left[\sum_{i=1}^n g_{ji} g_{2i} \right] + \dots + \alpha_{j(j-1)} \left[\sum_{i=1}^n g_{ji} g_{(j-1)i} \right]
 \end{aligned}
 \tag{3.12.69}$$

Encontrando as demais derivadas parciais, como fizemos nas etapas anteriores, teremos:

$$\left. \begin{aligned}
 \frac{\delta S''_j}{\delta g_{ji}} = & -2N \sum_{i'=1}^n g_{ji'} c_{ii'} + 2N\lambda_j g_{ji} + \alpha_{j1} g_{1i} + \\
 & + \alpha_{j2} g_{2i} + \dots + \alpha_{j(j-1)} g_{(j-1)i}
 \end{aligned} \right\} \tag{3.12.70}$$

$$\left. \frac{\delta S''_j}{\delta \lambda_j} = \sum_{i'=1}^n (g_{ji'})^2 - 1 \right\} \tag{3.12.71}$$

$$\left. \frac{\delta S''_j}{\delta \alpha_{j1}} = \sum_{i'=1}^n g_{ji'} g_{1i'} \right\} \tag{3.12.72}$$

$$\frac{\delta S''_j}{\delta \alpha_{j2}} = \sum_{i'=1}^n g_{ji'} g_{2i'} \quad (3.12.73)$$

$$\frac{\delta S''_j}{\delta \alpha_{j(j-1)}} = \sum_{i'=1}^n g_{ji'} g_{(j-1)i'} \quad (3.12.74)$$

Anulando as derivadas parciais chegamos então a:

$$2N \sum_{i'=1}^n g_{ji'} c_{ii'} - 2N\lambda_j g_{ji} - \alpha_{j1} g_{1i} - \alpha_{j2} g_{2i} - \dots - \alpha_{j(j-1)} g_{(j-1)i} = 0 \quad (3.12.75)$$

$$\sum_{i'=1}^n g_{ji'}^2 = 1 \quad (3.12.76)$$

$$\sum_{i'=1}^n g_{ji'} g_{1i'} = 0 \quad (3.12.77)$$

$$\sum_{i'=1}^n g_{ji'} g_{2i'} = 0 \quad (3.12.78)$$

$$\sum_{i'=1}^n g_{ji'} g_{(j-1)i'} = 0 \quad (3.12.79)$$

Dessas equações, mais uma vez, podemos tirar duas conclusões:

1.^a) λ_j é um autovalor da matriz C.

2.^a) λ_j é o j-ésimo maior autovalor da matriz C.

OBSERVAÇÃO:

Lembramos mais uma vez que por hipótese (ver seção 3.4) a matriz C possui n autovalores diferentes e não nulos. Os casos em que isso não ocorrer serão estudados na seção 3.17.

Passamos agora à confirmação das conclusões.

TEOREMA 3.12.5: λ_j é um autovalor da matriz C .

Demonstração:

Vamos partir da hipótese que já resolvemos o Problema 3 até a etapa $j-1$. Sendo assim já sabemos que todos os $\lambda_{j'}$, ($j'=1,2,\dots,j-1$) são autovalores da matriz C e os $g_{j'}^T$ são os autovetores unitários associados aos $\lambda_{j'}$. Isso nos permite escrever que:

$$\sum_{i'=1}^n g_{j'i'} c_{ii'} = \lambda_{j'} g_{j'i} \quad \begin{array}{l} j'=1,2,\dots,(j-1) \\ i=1,2,\dots,n \end{array} \quad (3.12.80)$$

(que é a generalização, para j' , das equações 3.12.23 e 3.12.52)

Por outro lado, se já resolvemos até a etapa $j-1$, as restrições dessas etapas devem ter sido respeitadas o que nos leva a:

$$\sum_{i=1}^n g_{j'i} g_{j''i} = \begin{array}{ll} 0 & \text{se } j' \neq j'' \\ 1 & \text{se } j' = j'' \end{array} \quad (3.12.81)$$

$$j'=1,2,\dots,(j-1)$$

$$j'' \leq j'$$

(ver equações 3.11.7-20)

Multiplicando-se então a equação 3.12.75 por $g_{j'i}$ ($j'=1,2,\dots,(j-1)$) e somando para todo $i=1,2,\dots,n$ teremos:

$$\begin{aligned}
& 2N \sum_{i=1}^n \sum_{i'=1}^n g_{ji'} g_{j'i} c_{ii'} - 2N\lambda_j \sum_{i=1}^n g_{ji} g_{j'i} - \\
& - \alpha_{j1} \sum_{i=1}^n g_{j'i} g_{1i} - \alpha_{j2} \sum_{i=1}^n g_{j'i} g_{2i} - \dots - \\
& - \alpha_{j(j-1)} \sum_{i=1}^n g_{j'i} g_{(j-1)i} = 0 \quad j'=1,2,\dots,(j-1) \quad (3.12.82)
\end{aligned}$$

Substituindo a equação 3.12.81 na 3.12.82 teremos:

$$2N \sum_{i=1}^n \sum_{i'=1}^n g_{ji'} g_{j'i} c_{ii'} = \alpha_j j', \quad j'=1,2,\dots,(j-1) \quad (3.12.83)$$

ou, o que é equivalente

$$2N \sum_{i'=1}^n g_{ji'} \left[\sum_{i=1}^n g_{j'i} c_{ii'} \right] = \alpha_j j', \quad j'=1,2,\dots,(j-1) \quad (3.12.84)$$

Por outro lado podemos escrever a equação 3.12.80 da seguinte forma:

$$\begin{aligned}
\sum_{i=1}^n g_{j'i} c_{ii'} &= \lambda_j g_{j'i}, & j'=1,2,\dots,(j-1) & \quad (3.12.85) \\
& & i'=1,2,\dots,n &
\end{aligned}$$

Levando a equação 3.12.85 na 3.12.84 chegamos a:

$$2N\lambda_j \sum_{i=1}^n g_{ji'} g_{j'i} = \alpha_j j', \quad j'=1,2,\dots,(j-1) \quad (3.12.86)$$

Substituindo a equação 3.12.81 na 3.12.86 concluimos que:

$$\boxed{\alpha_j j' = 0} \quad j'=1,2,\dots,(j-1) \quad (3.12.87)$$

Agora podemos escrever a equação 3.12.75 da seguinte forma:

$$\sum_{i'=1}^n g_{ji'} c_{ii'} = \lambda_j g_{ji} \quad i=1,2,\dots,n \quad (3.12.88)$$

Seja

$$g_j = (g_{j1}, g_{j2}, \dots, g_{jn})$$

e

$$C = \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & & & \vdots \\ \vdots & & & \vdots \\ c_{n1} & \dots & \dots & c_{nn} \end{pmatrix}$$

Podemos então escrever a equação 3.12.88 assim:

$$C g_j^T = \lambda_j g_j^T \quad (3.12.89)$$

que nada mais é do que a definição de autovalor. Logo λ_j é um autovalor da matriz C e g_j^T é um autovetor da matriz C associado ao autovalor λ_j . △

Com relação à segunda conclusão teremos o seguinte teorema: (ver anexo)

TEOREMA 3.12.6: λ_j é o j -ésimo maior autovalor da matriz C

Demonstração:

Multiplicando-se a equação 3.12.88 por g_{ji} e somando para todo $i=1,2,\dots,n$ teremos:

$$\sum_{i=1}^n \sum_{i'=1}^n g_{ji} g_{ji'} c_{ii'} = \lambda_j \sum_{i=1}^n g_{ji}^2 \quad (3.12.90)$$

mas pela equação 3.12.76

$$\sum_{i=1}^n g_{ji}^2 = 1 \quad (3.12.91)$$

logo

$$\sum_{i=1}^n \sum_{i'=1}^n g_{ji} g_{ji'} c_{ii'} = \lambda_j \quad (3.12.92)$$

Levando a equação 3.12.17 na 3.12.92 teremos:

$$\sum_{i=1}^n \sum_{i'=1}^n g_{ji} g_{ji'} \left[\sum_{k=1}^N (x_{i_i}(k) x_{i_i'}(k)) \right] = N \lambda_j \quad (3.12.93)$$

ou

$$\sum_{k=1}^N \sum_{i=1}^n \sum_{i'=1}^n g_{ji} g_{ji'} x_{i_i}(k) x_{i_i'}(k) = N \lambda_j \quad (3.12.94)$$

Levando as equações 3.12.76-9 e 3.12.94 na 3.12.69 chegamos a:

$$\text{MINIMIZAR } S_j'' = \left[\sum_{k=1}^N \sum_{i=1}^n x_{i_i}(k)^2 \right] - N \lambda_j \quad (3.12.95)$$

Logo, como os $x_{i_i}(k)$ e N são parâmetros da população em estudo, minimizar S_j'' é o mesmo que maximizar λ_j . Como pelo teorema 3.12.5 λ_j é um autovalor da matriz C , teremos que escolher λ_j entre os autovalores da matriz C . A primeira vista, como no teorema 3.12.4, poderia parecer que a escolha recairia sobre o maior autovalor da matriz C . Porém pelos motivos expostos naquele teorema essa solução não obedeceria à equação 3.12.77. O segundo maior autovalor também não satisfaria, pois não satisfaria à equação 3.12.78. Observan-

do então as equações 3.12.77-9 concluimos que o maior autovalor λ_j que satisfaz a esse conjunto de equações é o j -ésimo maior autovalor da matriz C . △

Voltando, mais uma vez às duas conclusões, verificamos que para S_j (soma dos quadrados das distâncias dos pontos $x(k)$ à reta $A_j Q_j$) ser mínima, sendo essa reta ortogonal a todas as retas $A_{j'}, Q_{j'}$, ($j'=1, 2, \dots, (j-1)$) é necessário que λ_j seja o j -ésimo maior autovalor da matriz C e que g_j^T seja o autovetor unitário associado a λ_j .

Assim a reta $A_j Q_j$ será definida por passar pela origem e ter cossenos diretores que sejam componentes do autovetor unitário da matriz C associado ao j -ésimo maior autovalor dessa matriz.

Fim da j -ésima etapa □

Com isso, concluimos que para resolver o Problema 3 precisamos inicialmente calcular os autovalores da matriz C (matriz de covariâncias) e depois encontrar os autovetores unitários associados a esses autovalores. Em outras palavras, precisamos resolver o seguinte sistema de equações:

$$(C - \lambda_j I) g_j^T = \emptyset \quad j=1, 2, \dots, n \quad (3.12.96)$$

$$\sum_{i=1}^n g_{ji}^2 = 1 \quad j=1, 2, \dots, n \quad (3.12.97)$$

onde I é a Matriz Identidade, λ_j é o j -ésimo maior autovalor da matriz C , g_j^T é o autovetor unitário da matriz C associado ao autovalor λ_j e \emptyset é um vetor coluna de n componentes nulas. (para maiores esclarecimentos sobre a equação 3.12.96 con

sultar [7], seção 2.6 ou [23], capítulo 9).

3.13. Interpretação do Problema 3

Depois de resolvido o Problema 3 vamos enfoca-lo de uma forma intuitiva-geométrica.

Inicialmente vamos ver o caso de um problema que envolve somente tres variáveis x_1, x_2 e x_3 . (observe que essas variáveis são desvios). Nesse caso o problema se restringe ao espaço vetorial \mathbb{R}^3 . Para resolver o problema o nosso objetivo inicial é encontrar a reta $(A_1 Q_1)$ que mais se "adapta" aos pontos $x(k)$ (adaptar, aqui, no sentido de a soma dos quadrados das distâncias desses pontos a ela ser mínima). Depois o nosso objetivo passa a ser, escolher dentre as retas que são ortogonais à primeira, aquela que mais se "adapta" aos pontos $x(k)$. E finalmente, o nosso terceiro objetivo será escolher dentre as retas que são ortogonais às duas primeiras, aquela que mais se "adapta" aos pontos $x(k)$.

Para o caso de problemas que envolvam n variáveis o raciocínio é análogo. Sempre o objetivo será encontrar uma reta que seja ortogonal as anteriores e que mais se "adapte" aos $x(k)$.

É importante lembrar que, como vimos na seção 3.12, as retas que satisfazem as condições acima, passam pela origem do espaço vetorial ao qual os pontos $x(k)$ pertencem pois estamos utilizando as variáveis x_1, x_2, \dots, x_n (que são desvios) e não as variáveis X_1, X_2, \dots, X_n . Se transportássemos o problema para o espaço vetorial ao qual os pontos $X(k)$ pertencem,

então poderíamos garantir que as n retas ortogonais calculadas acima, passariam pelo ponto de coordenadas $(\bar{X}_1, \bar{X}_2, \dots, \bar{X}_n)$.

3.14. Comparação dos Problemas 1, 2 e 3

O objetivo inicial dessa seção será mostrar que os Problemas 1, 2 e 3 possuem o mesmo conjunto de soluções.

Na seção 3.7 já mostramos que isso ocorre para os Problemas 1 e 2. Vamos então mostrar que também o Problema 3 possui aquele mesmo conjunto de soluções.

Na seção 3.4 chegamos à conclusão de que para resolver o Problema 1 bastaria resolver o sistema de equações 3.4.62-3 que repetimos a seguir:

$$(C - \lambda_j I) \mathcal{L}_j^T = \emptyset \quad j=1, 2, \dots, n \quad (3.14.1)$$

$$\sum_{i=1}^n \mathcal{L}_{ji}^2 = 1 \quad j=1, 2, \dots, n \quad (3.14.2)$$

onde os \mathcal{L}_{ji} são as componentes do autovetor unitário \mathcal{L}_j^T da matriz C , associado ao autovalor λ_j . Por outro lado, na seção 3.12, chegamos a conclusão de que para resolver o Problema 3 bastaria resolver o sistema das equações 3.12.96-97 que repetimos a seguir:

$$(C - \lambda_j I) g_j^T = \emptyset \quad j=1, 2, \dots, n \quad (3.14.3)$$

$$\sum_{i=1}^n g_{ji}^2 = 1 \quad j=1, 2, \dots, n \quad (3.14.4)$$

onde os g_{ji} são as componentes do autovetor unitário g_j^T da matriz C , associado ao autovalor λ_j . Comparando a equação

3.14.1 com a 3.14.3 e a 3.14.2 com a 3.14.4 concluímos que os ℓ_{ji} do Problema 1 correspondem aos g_{ji} do Problema 3, mostrando assim que os Problemas 1 e 3 possuem o mesmo conjunto de soluções. Consequentemente os tres problemas possuem o mesmo conjunto de soluções.

Vamos aqui recapitular os tres problemas que nada mais são do que tres formas de abordar o MCP. O Problema 1, abordando o MCP de uma forma ESTATÍSTICA. O Problema 2 dando uma abordagem segundo a ÁLGEBRA LINEAR e o Problema 3 abordando-o segundo o enfoque da GEOMETRIA NO \mathbb{R}^n .

A título de recordação, repetimos aqui os enunciados dos problemas.

PROBLEMAS 1, 2 e 3

Dado que numa população de N elementos foram medidas n características em cada um deles, e que essas características são representadas pelas variáveis x_i ($i=1,2,\dots,n$).

PROBLEMA 1

Encontrar n componentes $y_j = \sum_{i=1}^n \ell_{ji} x_i$ ($j=1,2,\dots,n$) sujeitas à restrição $\sum_{i=1}^n \ell_{ji}^2 = 1$ ($j=1,2,\dots,n$) de tal forma que:

- 1) y_1 tenha variância máxima;
- 2) y_2 tenha variância máxima mas não seja correlacionada com y_1 ;
- 3) y_3 tenha variância máxima mas não seja correlacionada nem com y_1 nem com y_2 ;

⋮
⋮
⋮

n) y_n tenha variância máxima mas não seja correlacionada nem com y_1 , nem com y_2 , nem com y_3, \dots , nem com $y_{(n-1)}$.

PROBLEMA 2

Encontrar \underline{n} componentes $y_j = \sum_{i=1}^n \hat{\ell}_{ji} x_i$ ($j=1, 2,$

\dots, n) tal que as componentes y_j não sejam correlacionadas entre si e que a matriz \hat{L} (dos $\hat{\ell}_{ji}$) represente uma Transformação Ortogonal.

PROBLEMA 3

E considerando-se os valores $x(k)$ ($k=1, 2, \dots, N$), que são os valores assumidos pelas variáveis x_1, x_2, \dots, x_n para cada elemento k da população, como pontos do espaço vetorial \mathbb{R}^n , encontrar um conjunto de \underline{n} retas nesse espaço de tal forma que:

- 1) a soma dos quadrados das distâncias dos pontos $x(k)$ à primeira reta seja mínima;
- 2) a soma dos quadrados das distâncias dos pontos $x(k)$ à segunda reta seja mínima, sendo essa reta ortogonal à primeira;

,

,

,

- n) a soma dos quadrados das distâncias dos pontos $x(k)$ à n -ésima reta seja mínima, sendo essa reta ortogonal às anteriores.

Também para relembrar vamos rever a formulação matemática de cada um dos problemas.

PROBLEMA 1

j.^a ETAPA: (j=1,2,...,n)

$$\text{MAXIMIZAR } \sum_{i=1}^n \sum_{i'=1}^n \ell_{ji} \ell_{ji'} c_{ii'} \quad (3.14.5)$$

sujeito a

$$\sum_{i=1}^n \ell_{ji}^2 = 1 \quad (3.14.6)$$

$$\sum_{i=1}^n \sum_{i'=1}^n \ell_{ji} \ell_{j'i'} c_{ii'} = 0 \quad (3.14.7)$$

j'=1,2,...,(j-1)

PROBLEMA 2

$$y_j = \sum_{i=1}^n \ell_{ji} x_i \quad j=1,2,\dots,n \quad (3.14.8)$$

$$\text{COR}[y_j; y_{j'}] = 0 \quad j \neq j' \quad (3.14.9)$$

$$\sum_{i=1}^n \ell_{ji} \ell_{j'i} = 1 \quad \text{se } j = j' \quad (3.14.10)$$

$$= 0 \quad \text{se } j \neq j' \quad (3.14.11)$$

PROBLEMA 3

j.^a ETAPA : (j=1,2,...,n)

$$\text{MINIMIZAR } S_j \quad (3.14.12)$$

sujeito a

$$\sum_{i=1}^n g_{ji} g_{j'i} = 1 \quad \text{se } j = j' \quad (3.14.13)$$

$$= 0 \quad \text{se } j \neq j' \quad (3.14.14)$$

$$j'=1,2,\dots,j$$

onde S_j representa o somatório dos quadrados das distâncias dos pontos $x(k)$ à j -ésima reta.

Comparando essas formulações podemos notar que a restrição que impõe que o autovetor da matriz C (λ_j^T ou g_j^T) seja unitário está presente nos três problemas e é representada pelas equações 3.14.6, 3.14.10 e 3.14.13. Por outro lado a condição de não correlação das componentes aparece somente nos Problemas 1 e 2 e é representada pelas equações 3.14.7 e 3.14.9. Por sua vez a condição de ortogonalidade das componentes, ou das retas, aparece nos Problemas 2 e 3 e é representada pelas equações 3.14.11 e 3.14.14.

Esquemáticamente, e sem nenhum rigor matemático, poderíamos resumir os tres problemas na seguinte figura, que engloba objetivos e restrições:

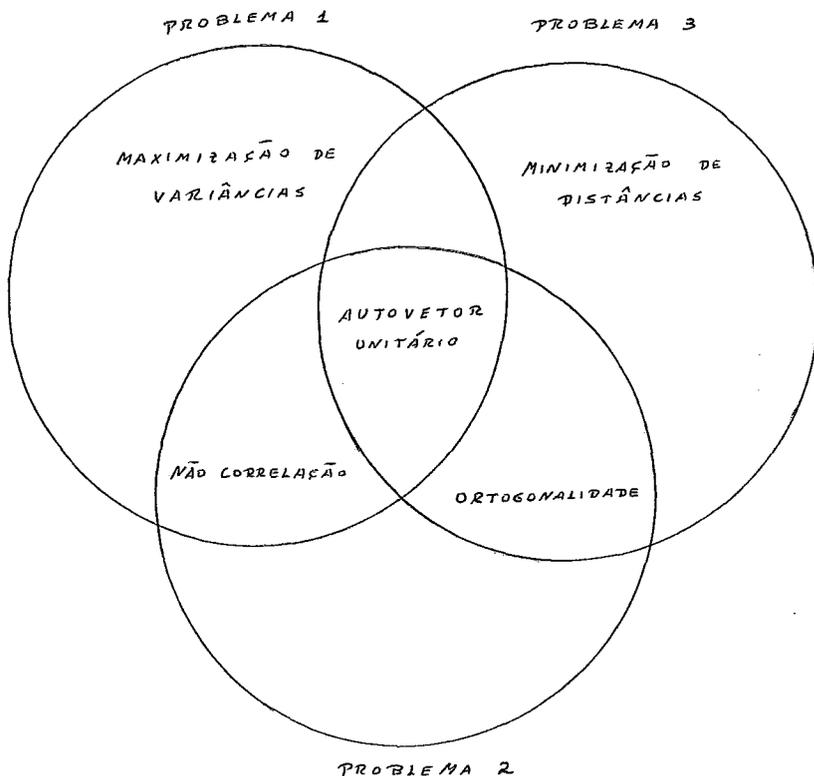


FIGURA 3.14.1: Representação esquemática dos Problemas 1, 2 e 3.

Já vimos que os tres problemas nada mais são que enfoques diferentes do MCP. Sendo assim uma série de características geométricas estudadas no modelo nada mais são que características do MCP. Vamos então relacioná-las:

- i) Podemos associar cada reta $A_j Q_j$ ($j=1,2,\dots,n$) à j -ésima componente e intuitivamente considerá-la como sua "representação geométrica";
- ii) Pelo teorema 2.5.4 se representarmos geometricamente as variáveis x_1, x_2, \dots, x_n por eixos ortogonais, as retas $A_1 Q_1, A_2 Q_2, \dots, A_n Q_n$ serão ortogonais entre si;
- iii) Com relação à primeira componente, a maximização da sua variância, corresponde geometricamente a uma minimização do quadrado das distâncias dos pontos $x(k)$ à reta $A_1 Q_1$.

3.15. Comentários Sobre o Exemplo

Nessa seção vamos verificar se de fato no Exemplo 3.6.1 as componentes y_j eram ortogonais entre si.

Fazendo os $\ell_{ji} = g_{ji}$ poderíamos dizer, sem muito rigor, que os "cossenos diretores das Componentes" (a rigor seriam os cossenos diretores da reta a qual podemos associar a Componente) naquele exemplo são:

$$g_1 = (0,604 ; 0,542 ; 0,586) \quad (3.15.1)$$

$$g_2 = (-0,469 ; 0,835 ; -0,289) \quad (3.15.2)$$

$$g_3 = (-0,646 ; -0,098 ; 0,758) \quad (3.15.3)$$

Para verificar se as Componentes (ou melhor, as retas que associamos as Componentes) são ortogonais, basta encontrar o produto escalar dos vetores g_1, g_2 e g_3 dois a dois e verificar se eles são nulos. Assim teremos:

$$\begin{aligned}
 \langle g_1 \cdot g_2 \rangle &= (0,604 \times -0,469) + (0,542 \times 0,835) + \\
 &+ (0,586 \times -0,289) \\
 &= -0,283 + 0,452 - 0,169 \\
 &= 0
 \end{aligned}
 \tag{3.15.4}$$

$$\begin{aligned}
 \langle g_1 \cdot g_3 \rangle &= (0,604 \times -0,646) + (0,542 \times -0,098) + \\
 &+ (0,586 \times 0,758) \\
 &= -0,390 - 0,053 + 0,444 \\
 &\approx 0
 \end{aligned}$$

$$\begin{aligned}
 \langle g_2 \cdot g_3 \rangle &= (-0,469 \times -0,646) + (0,835 \times -0,098) + \\
 &+ (-0,289 \times 0,758) \\
 &= 0,303 - 0,083 - 0,219 \\
 &\approx 0
 \end{aligned}$$

Assim fica comprovado que as Componentes (ou melhor, as retas que a elas associamos) são de fato ortogonais o que pode ser visto na figura 3.8.1.

3.16. Analogia Com a Mecânica

Uma outra forma intuitiva de percebermos o MCP é comparando o seu enfoque geométrico com a FÍSICA (mecânica). O que passamos a fazer.

Seja um conjunto composto por N partículas de massa unitária, que representaremos por suas posições $x(k)$ ($k=1, 2, \dots, N$) em relação ao sistema de coordenadas x_1, x_2, x_3 . (Lembremos que $x(k) = (x_1(k), x_2(k), x_3(k))$).

CENTRO DE MASSA

Uma das características físicas desse conjunto de partículas (que chamaremos de corpo) é o seu CENTRO DE MASSA (a grosso modo, chamado de CENTRO DE GRAVIDADE).

Para este corpo a fórmula para cálculo da posição de seu centro de massa é a seguinte:

$$CM = (CM_1, CM_2, CM_3) \quad (3.16.1)$$

e

$$CM_i = \frac{1}{N} \sum_{k=1}^N x_i(k) \quad i=1, 2, 3 \quad (3.16.2)$$

onde CM é a posição do CENTRO DE MASSA do corpo no sistema de coordenadas x_1, x_2, x_3 .

Logo,

$$CM_i = E[x_i] \quad i=1, 2, 3 \quad (3.16.3)$$

mas como x_i é um desvio

$$E[x_i] = 0 \quad (3.16.4)$$

e

$$CM = (0, 0, 0) \quad (3.16.5)$$

ou seja, o centro de massa desse corpo está na origem do sistema de coordenadas. Assim, concluir que as retas $A_j Q_j$ ($j=1, 2, 3$) (ver Problema 3) passam pela origem do espaço vetorial ao

qual os $x(k)$ pertencem, equivale a dizer que essas retas pas sam pelo Centro de Massa das partículas.

Para melhor entendimento do que seja este centro de massa, vamos transcrever um trecho de HALLIDAY & RESNICK ([¹¹], pág.187):

"mesmo quando o corpo roda ou vibra, enquanto se desloca, há um ponto no corpo, chamado CENTRO DE MASSA, que se desloca da mesma maneira que se deslocaria uma única partícula, sujeita ao mesmo sistema de forças externas".

Em outras palavras, para determinados tipos de estudos, poderíamos considerar apenas o centro de massa do corpo e não todas as suas partículas. Por exemplo, um avião lan ça uma bomba. A grosso modo não interessa se ela está rodando ou vibrando, interessa somente qual a trajetória que será percorrida pelo seu centro de massa, pois essa trajetória é que indicará onde a bomba cairá.

MOMENTO DE INÉRCIA

Outra característica desse corpo é o seu MOMENTO DE INÉRCIA. Ele é uma grandeza física que mede a inércia de rotação de um corpo em relação a um eixo.

Seja então o eixo definido por uma reta qualquer $A_j Q_j$ ($j=1,2,3$). O momento de inércia das partículas $x(k)$ em relação ao eixo $A_j Q_j$, é dado por:

$$MI_j = \sum_{k=1}^N [\text{dist}(x(k); A_j Q_j)]^2 \quad (3.16.6)$$

onde MI_j é o momento de inércia do corpo em relação ao eixo

$A_j Q_j$.

Se observarmos o Problema 3 (seção 3.11), veremos que MI_j nada mais é que S_j . Assim ao impor que S_j seja mínima estávamos definindo que a reta $A_j Q_j$ seria aquela, em relação a qual o MOMENTO DE INÉRCIA das partículas, que estão nas posições $x(k)$, fosse mínimo.

Da mesma forma que fizemos para $j=1,2,3$ (já que na mecânica estamos limitados a três dimensões), poderíamos estender esse conceito para $j=1,2,\dots,n$ e definir um outro problema análogo aos três anteriores.

PROBLEMA 4

Dado que numa população de N elementos foram medidas n características em cada um deles, e que essas características são representadas pelas variáveis x_i ($i=1,2,\dots,n$), e considerando os valores $x(k)$ ($k=1,2,\dots,N$), que são valores assumidos pelas variáveis x_1, x_2, \dots, x_n para cada elemento k da população, como a posição de partículas de massa unitária no espaço vetorial \mathbb{R}^n , encontrar um conjunto de n retas, nesse espaço, passando pela origem, de tal forma que:

- 1) o momento de inércia das partículas $x(k)$ em relação à primeira reta seja mínimo;
- 2) o momento de inércia das partículas $x(k)$ em relação à segunda reta seja mínimo, sendo essa reta ortogonal à primeira;

|
|
|

|
|
|

⋮
⋮
⋮
⋮

n) o momento de inércia das partículas $x(k)$ em relação à n -ésima reta seja mínimo, sendo essa reta ortogonal às anteriores.

3.17. Casos Particulares

Nas seções anteriores, ao desenvolver o MCP, partimos do pressuposto que a Matriz de Covariâncias (C) possuía n autovalores distintos e não nulos. Nesta seção vamos comparar esses casos com aqueles nos quais isso não ocorre.

Observamos, na seção 3.14, que o MCP se baseia na resolução do sistema de equações 3.14.1-2 ou equivalentemente 3.14.3-4 que repetimos a seguir:

$$(C - \lambda_j I) \mathcal{Z}_j^T = \emptyset \quad j=1,2,\dots,n \quad (3.17.1)$$

$$\sum_{i=1}^n \mathcal{Z}_{ji}^2 = 1 \quad j=1,2,\dots,n \quad (3.17.2)$$

Para nosso estudo vamos então utilizar o seguinte sistema de equações:

$$(C - \lambda_j I) U_j = \emptyset \quad j=1,2,\dots,n \quad (3.17.3)$$

$$\sum_{i=1}^n \mathcal{Z}_{ji}^2 = 1 \quad j=1,2,\dots,n \quad (3.17.4)$$

onde U_j é uma matriz composta por p colunas ($p \leq n$) que serão vetores \mathcal{L}_j^T , linearmente independentes e ortogonais entre si, soluções da equação 3.17.1.

Para a estudo desses casos vamos aos seguintes teoremas:

TEOREMA 3.17.1: Se a matriz L representa uma transformação ortogonal, C é a Matriz de Covariâncias e D é definida segundo a equação 3.4.67 então:

$$\text{rank}|C| = \text{rank } D$$

$$\text{e } \text{rank}|C - \lambda_j I| = \text{rank}|D - \lambda_j I|$$

Demonstração:

Seja A uma matriz $n \times n$ simétrica e B uma matriz $n \times n$ diagonal com os autovalores da matriz A , em sua diagonal principal, em ordem decrescente.

Pelo Lema 1 do teorema 3.4.7 temos:

$$L^T A L = B \quad (3.17.5)$$

onde L e L^T representam transformações ortogonais.

Aplicando a Equação de Sylvester (ver [7], pág.33) à equação 3.17.5 temos:

$$\text{rank}[L^T] + \text{rank}[AL] - n \leq \text{rank}[B] \leq \text{mínimo}\{\text{rank}[L^T], \text{rank}[AL]\} \quad (3.17.6)$$

porém como L representa uma transformação ortogonal temos, pela equação 3.4.74, que:

$$\text{rank}[L] = \text{rank}[L^T] = n \quad (3.17.7)$$

Logo, levando esse resultado na equação 3.17.6 temos:

$$\cancel{n} + \text{rank}[AL] - \cancel{n} \leq \text{rank}[B] \leq \text{mínimo}\{n, \text{rank}[AL]\} \quad (3.17.8)$$

$$\text{Logo, } \text{rank}[AL] = \text{rank}[B] \quad (3.17.9)$$

Aplicando novamente a Equação de Sylvester temos:

$$\text{rank}[A] + \text{rank}[L] - n \leq \text{rank}[B] \leq \text{mínimo}\{\text{rank}[A], \text{rank}[L]\} \quad (3.17.10)$$

Substituindo a equação 3.17.7 na 3.17.10 teremos:

$$\text{rank}[A] + \cancel{n} - \cancel{n} \leq \text{rank}[B] \leq \text{mínimo}\{\text{rank}[A], n\} \quad (3.17.11)$$

logo

$$\text{rank}[A] = \text{rank}[B] \quad (3.17.12)$$

Se observarmos a equação 3.4.67 observamos que podemos ter as matrizes A e B como sendo as matrizes C e D, ou seja:

$$A = C \quad \text{e} \quad B = D \quad (3.17.13)$$

o que nos leva a concluir que:

$$\text{rank}[C] = \text{rank}[D] \quad (3.17.14)$$

E se observarmos a equação 3.4.70 observamos que podemos ter:

$$A = (C - \lambda_j I) \quad \text{e} \quad B = (D - \lambda_j I) \quad (3.17.15)$$

o que nos leva a concluir que:

$$\text{rank}[C - \lambda_j I] = \text{rank}[D - \lambda_j I] \quad (3.17.16)$$



TEOREMA 3.17.2: Na equação:

$$(C - \lambda_j I)U_j = \emptyset$$

$$\text{o rank } [U_j] = n - \text{rank}[C - \lambda_j I]$$

Demonstração:

Ver [7] página 33, teorema 2.5 e corolário 2.5.

△

TEOREMA 3.17.3: Se $\text{rank}[C] < n$ então as variáveis x_1, x_2, \dots, x_n são linearmente dependentes.

Demonstração:

Pela equação 3.7.35b temos:

$$C = \begin{pmatrix} E[x_1 x_1] & E[x_1 x_2] & \dots & E[x_1 x_n] \\ E[x_2 x_1] & & & \vdots \\ \vdots & & & \vdots \\ E[x_n x_1] & \dots & \dots & E[x_n x_n] \end{pmatrix} \quad (3.17.17)$$

se o $\text{rank } [C] < n$ então as colunas da matriz C são linearmente dependentes o que nos permite escrever que:

$$\exists \alpha_{j'} \neq 0 \quad j'=1,2,\dots,n$$

tal que:

$$\sum_{j'=1}^n \alpha_{j'} E[x_j x_{j'}] = 0 \quad j=1,2,\dots,n \quad (3.17.18)$$

ou, utilizando a definição de média que,

$$\sum_{j'=1}^n \alpha_{j'} \cdot \frac{1}{N} \sum_{k=1}^n x_{j'}(k) x_{j'}(k) = 0 \quad j=1,2,\dots,n \quad (3.17.18A)$$

Utilizando as propriedades dos somatórios podemos escrever a equação 3.17.18A assim:

$$\frac{1}{N} \sum_{k=1}^N x_j(k) \sum_{j'=1}^n \alpha_{j'} x_{j'}(k) = 0 \quad j=1,2,\dots,n \quad (3.17.19)$$

ou ainda

$$\begin{aligned} & \frac{1}{N} \{x_j(1) [\alpha_1 x_1(1) + \alpha_2 x_2(1) + \dots + \alpha_n x_n(1)] + x_j(2) [\alpha_1 x_1(2) + \\ & + \alpha_2 x_2(2) + \dots + \alpha_n x_n(2)] + \dots + x_j(N) [\alpha_1 x_1(N) + \alpha_2 x_2(N) + \dots + \alpha_n x_n(N)]\} = \\ & = 0 \quad j=1,2,\dots,n \quad (3.17.19A) \end{aligned}$$

Utilizando o conceito de produto escalar podemos escrever a equação 3.17.19A assim:

$$\frac{1}{N} \langle (x_j(1), x_j(2), \dots, x_j(N)) \cdot \sum_{j'=1}^n \alpha_{j'} (x_{j'}(1), x_{j'}(2), \dots, x_{j'}(N)) \rangle = 0 \quad j=1,2,\dots,n \quad (3.17.20)$$

Seja então o vetor

$$u = \sum_{j=1}^n \alpha_j (x_j(1), x_j(2), \dots, x_j(N)) \quad (3.17.20A)$$

logo

$$\begin{aligned} \langle u \cdot u \rangle = & \left\langle \sum_{j=1}^n \alpha_j (x_j(1), x_j(2), \dots, x_j(N)) \cdot \sum_{j'=1}^n \alpha_{j'} (x_{j'}(1), x_{j'}(2), \dots \right. \\ & \left. \dots, x_{j'}(N)) \right\rangle \quad (3.17.20B) \end{aligned}$$

Levando a equação 3.17.20 na 3.17.20B teremos:

$$\langle u \cdot u \rangle = 0 \quad (3.17.21)$$

Levando esse resultado na equação 3.17.20A concluímos que:

$$\sum_{j=1}^n \alpha_j x_j(k) = 0 \quad \forall k=1, 2, \dots, N \quad (3.17.21A)$$

o que indica que o conjunto das variáveis x_1, x_2, \dots, x_n é linearmente dependente para essa população de N elementos.

△

Passamos agora ao estudo de cada um dos casos.

1º CASO: n autovalores distintos e não nulos.

Este é o caso que foi abordado nas seções anteriores.

Para este caso teremos:

$$D = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & & \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \dots & \lambda_n \end{pmatrix} \quad (3.17.22)$$

onde podemos observar que temos n colunas linearmente independentes ou seja:

$$\text{rank}[D] = n \quad (3.17.23)$$

e conseqüentemente:

$$D - \lambda_j I = \begin{pmatrix} \lambda_1 - \lambda_j & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 - \lambda_j & 0 & \dots & 0 \\ 0 & 0 & \lambda_j - \lambda_j & \dots & 0 \\ 0 & 0 & 0 & \dots & \lambda_n - \lambda_j \end{pmatrix} \quad (3.17.24)$$

Agora podemos observar que $D - \lambda_j I$ tem somente $n-1$ colunas linearmente independentes ou seja:

$$\text{rank}[D - \lambda_j I] = n - 1 \quad (3.17.25)$$

Se reunirmos esse resultado com os teoremas 3.17.1 e 3.17.2 concluimos que:

$$\text{rank}[U_j] = 1 \quad (3.17.26)$$

o que significa dizer que não será possível encontrar 2 ou mais autovetores \mathcal{L}_j^T linearmente independentes e ortogonais entre si, para um mesmo j , que sejam solução da equação 3.17.3. Assim sendo se \mathcal{L}_j^T é solução daquela equação então qualquer outra solução será da forma $\alpha \mathcal{L}_j^T$ onde $\alpha \in \mathbb{R}$.

Por isso, se impusermos a restrição de autovetor unitário (equação 3.17.4), a menos do sinal (ver teorema 3.4.7), a solução será única.

Fim do 1º CASO \square

2º CASO: n autovalores distintos sendo um deles nulo

Para este caso teremos:

$$D = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 & 0 \\ 0 & \lambda_2 & \dots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & \lambda_{n-1} & 0 \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix} \quad (3.17.27)$$

onde podemos observar que temos $n-1$ colunas linearmente independentes, ou seja:

$$\text{rank}[D] = n - 1 \quad (3.17.28)$$

Neste caso, pelo teorema 3.17.1, teremos:

$$\text{rank}[C] = \text{rank}[D] = n - 1 \quad (3.17.29)$$

o que, pelo teorema 3.17.3, nos leva a concluir que as variáveis x_1, x_2, \dots, x_n são linearmente dependentes e que logicamente, pelo menos uma delas é consequência das demais e portanto pode ser desprezada na análise do MCP.

Para melhor entendimento desse caso vamos partir de um exemplo. (esse exemplo é apenas didático).

EXEMPLO 3.17.1:

PROBLEMA:

Foram medidas, em um conjunto de 7 famílias, tres variáveis, que denominaremos de X_1, X_2 e X_3 e que representam o seguinte:

- X_1 → percentual da renda familiar gasto em moradia;
- X_2 → percentual da renda familiar gasto em alimentação;
- X_3 → percentual da renda familiar que resta para as demais despesas e uma possível economia.

Queremos analisar esses dados para estudar a possibilidade de uma só componente do MCP ser a responsável pela variância dos dados.

O resultado da medição encontra-se no quadro 3.17.1

VARIÁVEIS	FAMÍLIA						
	1	2	3	4	5	6	7
X_1	20	40	10	30	25	50	15
X_2	30	30	15	20	10	40	15
X_3	50	30	75	50	65	10	70

QUADRO 3.17.1: Valores de 3 variáveis observadas em 7 famílias.

e na figura 3.17.1.

SOLUÇÃO:

Inicialmente vamos encontrar os desvios x_1, x_2 e x_3 . Para isso calculamos as médias das variáveis:

$$\bar{X}_1 \approx 27,14 \quad \bar{X}_2 \approx 22,86 \quad \bar{X}_3 \approx 50,00 \quad (3.17.30)$$

Em seguida, subtraindo de cada $X_i(k)$ ($i=1,2,3$; $k=1,2,\dots,10$) a média \bar{X}_i , encontraremos os $x_i(k)$ que podem ser vistos no quadro 3.17.2 e na figura 3.17.2.

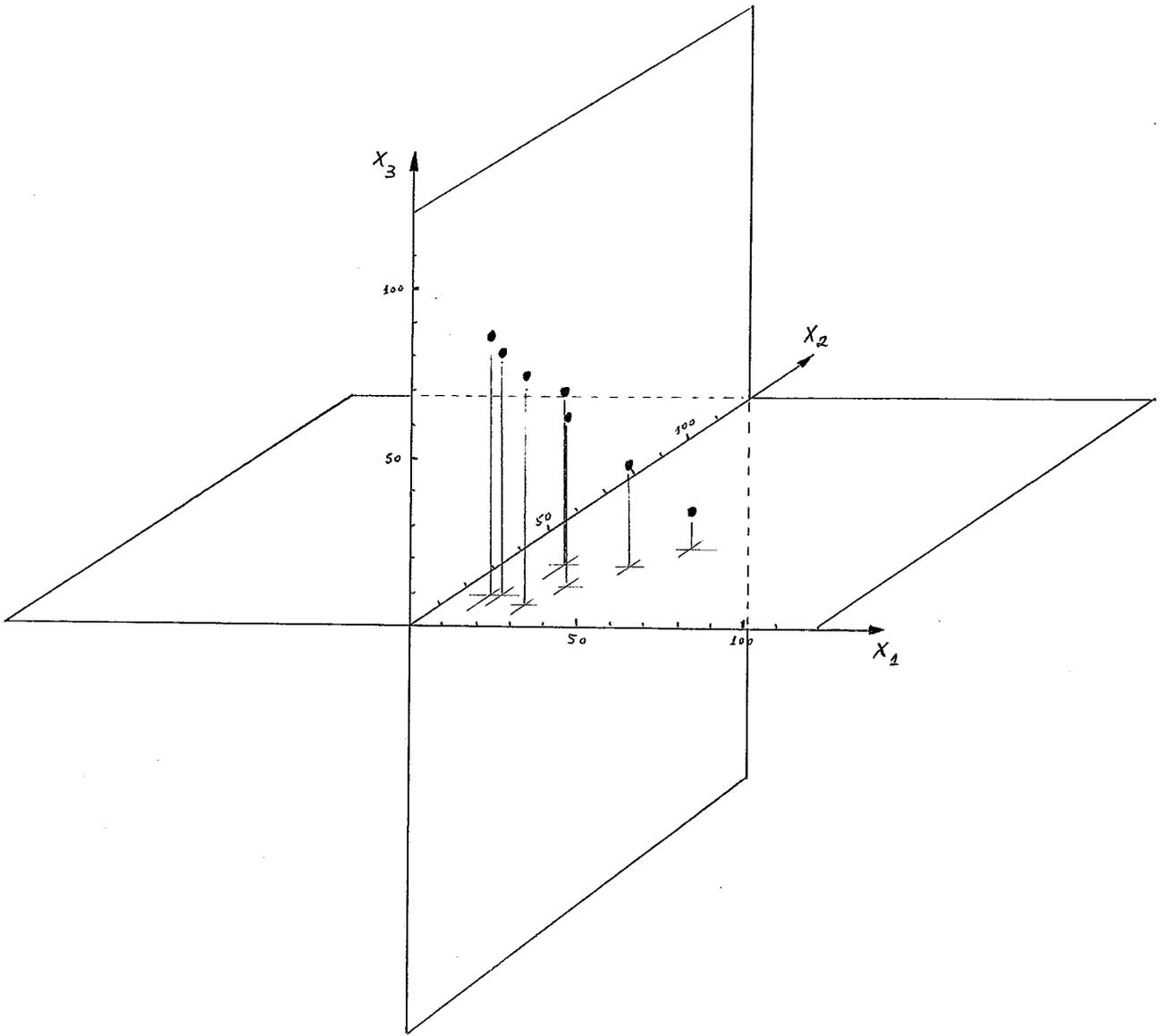


FIGURA 3.17.1: Representação gráfica do quadro 3.17.1.

DESvio	FAMÍLIA						
	1	2	3	4	5	6	7
x_1	-7,14	12,86	-17,14	2,86	- 2,14	22,86	-12,14
x_2	7,14	7,14	- 7,86	-2,86	-12,86	17,14	- 7,86
x_3	0	-20,00	25,00	0	15,00	-40,00	20,00

QUADRO 3.17.2: Valores de 3 desvios observados em 7 famílias.

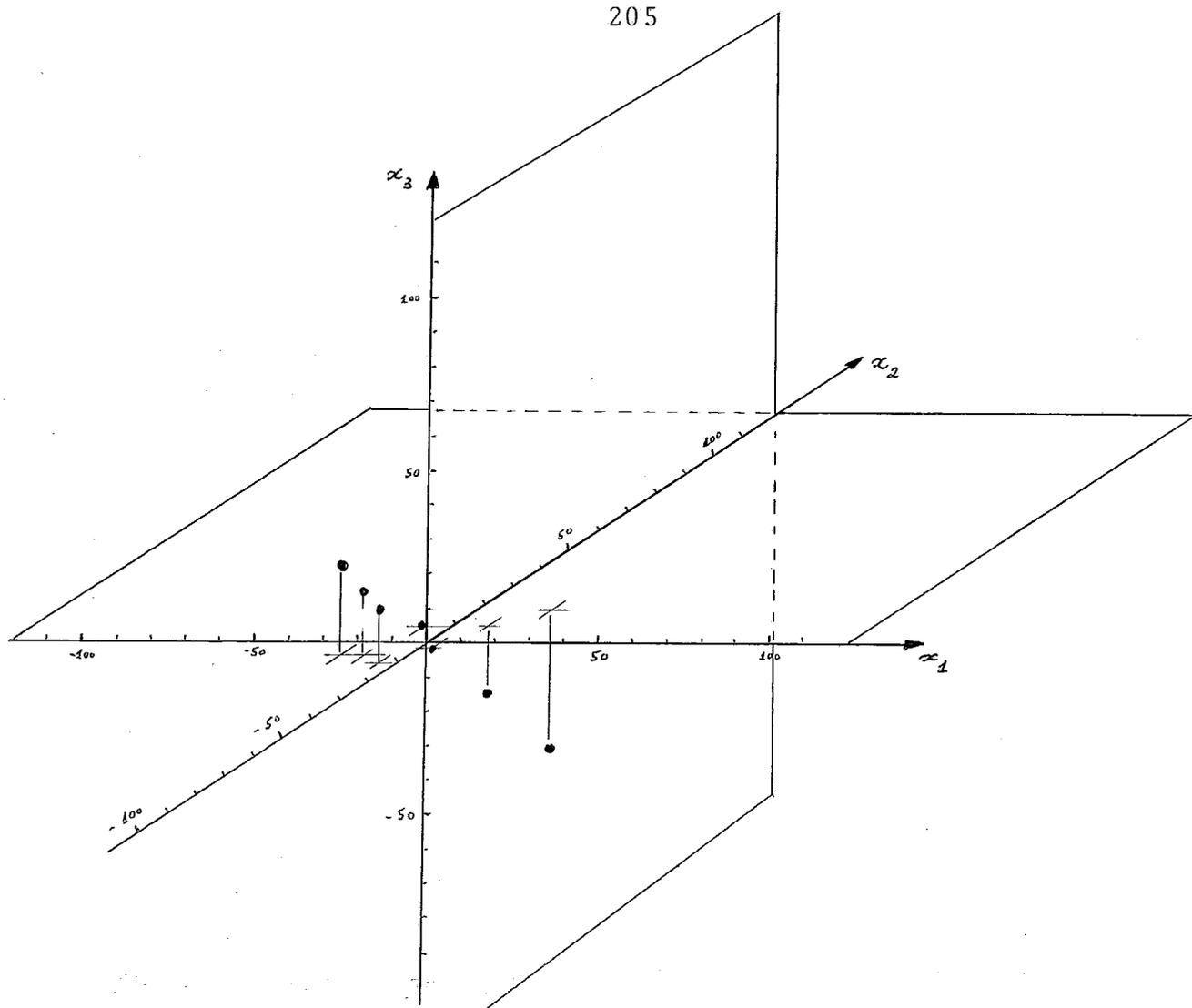


FIGURA 3.17.2: Representação gráfica do quadro 3.17.2.
Calculando então as variâncias e as covariâncias,
teremos:

$$\begin{array}{ll}
 \text{VAR}[x_1] \approx 170,41 & \text{COV}[x_1;x_2] \approx 97,45 \\
 \text{VAR}[x_2] \approx 98,98 & \text{COV}[x_1;x_3] \approx -267,86 \\
 \text{VAR}[x_3] \approx 464,26 & \text{COV}[x_2;x_3] \approx -196,43
 \end{array} \quad (3.17.2)$$

e a Matriz de Covariâncias ficará:

$$C = \begin{pmatrix} 170,41 & 97,45 & -267,86 \\ 97,45 & 98,98 & -196,43 \\ -267,86 & -196,43 & 464,29 \end{pmatrix} \quad (3.17.31)$$

Para encontrar os autovalores da matriz C vamos desenvolver o seu determinante e iguala-lo a zero. Assim teremos:

$$\lambda^3 - 733,68 \lambda^2 + 22112,04 \lambda = 0 \quad (3.17.32)$$

Logo os autovalores serão:

$$\begin{aligned} \lambda_1 &\approx 702,18 \\ \lambda_2 &\approx 31,50 \\ \lambda_3 &= 0 \end{aligned} \quad (3.17.33)$$

Neste caso teremos:

$$D = \begin{pmatrix} 702,18 & 0 & 0 \\ 0 & 31,50 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (3.17.34)$$

e, pela equação 3.17.29, teremos:

$$\text{rank}[C] = \text{rank}[D] = 2 \quad (3.17.35)$$

o que nos leva a concluir que as variáveis x_1, x_2 e x_3 (ou equivalentemente, X_1, X_2, X_3) são linearmente dependentes.

Para o nosso exemplo esse resultado já era esperado pois as variáveis foram definidas de tal forma que:

$$X_1(k) + X_2(k) + X_3(k) = 100 \quad \forall k=1,2,\dots,7 \quad (3.17.36)$$

Ou seja, uma das variáveis é linearmente dependente das outras duas. Sendo assim, somente duas são necessárias. Logo poderíamos escolher por exemplo X_1 e X_2 e resolver o MCP encontrando as duas componentes principais.

Como o nosso objetivo foi somente mostrar o que

acontece para o caso de autovalores nulos, não iremos até o fim do exercício, o que poderá ser feito pelo leitor. ▲

Geomêtricamente esse caso (existência de um autovalor nulo) pode ser comparado com uma distribuição de pontos sobre um plano (\mathbb{R}^2) num espaço de 3-dimensões (\mathbb{R}^3). Nesse caso poderemos eliminar uma variável já que ela seria função das duas outras.

OBS.: Um caso também que é interessante considerarmos é aquele em que um autovalor embora não seja nulo ele é aproximadamente zero. Este caso é geomêtricamente comparado por KENDALL ([¹⁹], pág.18) como uma distribuição de pontos na forma de um "bólo achatado" onde uma dimensão é desprezível em relação às demais. Então uma componente principal (a n-ésima) não será muito "importante" o que não implica dizer que uma das variáveis não é importante.

Observe que o caso do autovalor nulo e o caso do autovalor aproximadamente zero são conceitualmente diferentes. No primeiro há uma dependência linear entre as variáveis o que não ocorre no segundo caso.

OBS.: Para o caso de mais de um autovalor nulo, por exemplo q autovalores nulos, poderia ser demonstrado que q variáveis x_j poderiam ser desprezadas, já que elas poderiam ser representadas por combinações lineares das n-q restantes. A demonstração de tal fato é simples e fica a cargo do leitor.

Fim do 2º CASO □

3º CASO: entre os n autovalores não nulos há um e somente um par de autovalores iguais

Para este caso teremos:

$$D = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \lambda_r & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & \lambda_s & \dots & 0 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & \lambda_n \end{pmatrix} \quad \lambda_r = \lambda_s \quad (3.17.37)$$

onde:

$$\text{rank}[D] = n \quad (3.17.38)$$

e conseqüentemente:

$$D - \lambda_s I = D - \lambda_r I = \begin{pmatrix} \lambda_1 - \lambda_r & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 - \lambda_r & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & \lambda_n - \lambda_r \end{pmatrix} \quad (3.17.39)$$

onde:

$$\text{rank}[D - \lambda_r I] = n - 2 \quad (3.17.40)$$

Reunindo a equação 3.14.40 e os teoremas 3.17.1 e 3.17.2 teremos:

$$\text{rank}[U_r] = 2 \quad (3.17.41)$$

o que significa dizer que poderemos encontrar até 2 autoveto-

res ℓ_r^T e ℓ_s^T , linearmente independentes e ortogonais que sejam solução da equação 3.17.3.

Levando em conta a equação 3.17.4 para cada um desses autovetores, podemos escolhê-los de forma que eles sejam unitários.

Assim L representará, de fato, uma matriz ortogonal já que mesmo os ℓ_r^T e ℓ_s^T que são autovetores associados a autovalores iguais são ortogonais entre si e unitários.

Para melhor entendimento, vamos a um exemplo.

EXEMPLO 3.17.2:

PROBLEMA:

Numa população de 12 elementos foram medidas 3 variáveis X_1, X_2 e X_3 e os resultados obtidos encontram-se no

VARIÁVEL \ ELEMENTO	ELEMENTO											
	1	2	3	4	5	6	7	8	9	10	11	12
X_1	0	0	0	0	3	3	3	3	6	6	6	6
X_2	0	1	1	0	0	1	1	0	0	1	1	0
X_3	0	0	1	1	0	0	1	1	0	0	1	1

QUADRO 3.17.3: Valores de 3 variáveis observadas em 12 elementos.

quadro 3.17.3 e na figura 3.17.3.

Deseja-se encontrar as Componentes Principais.

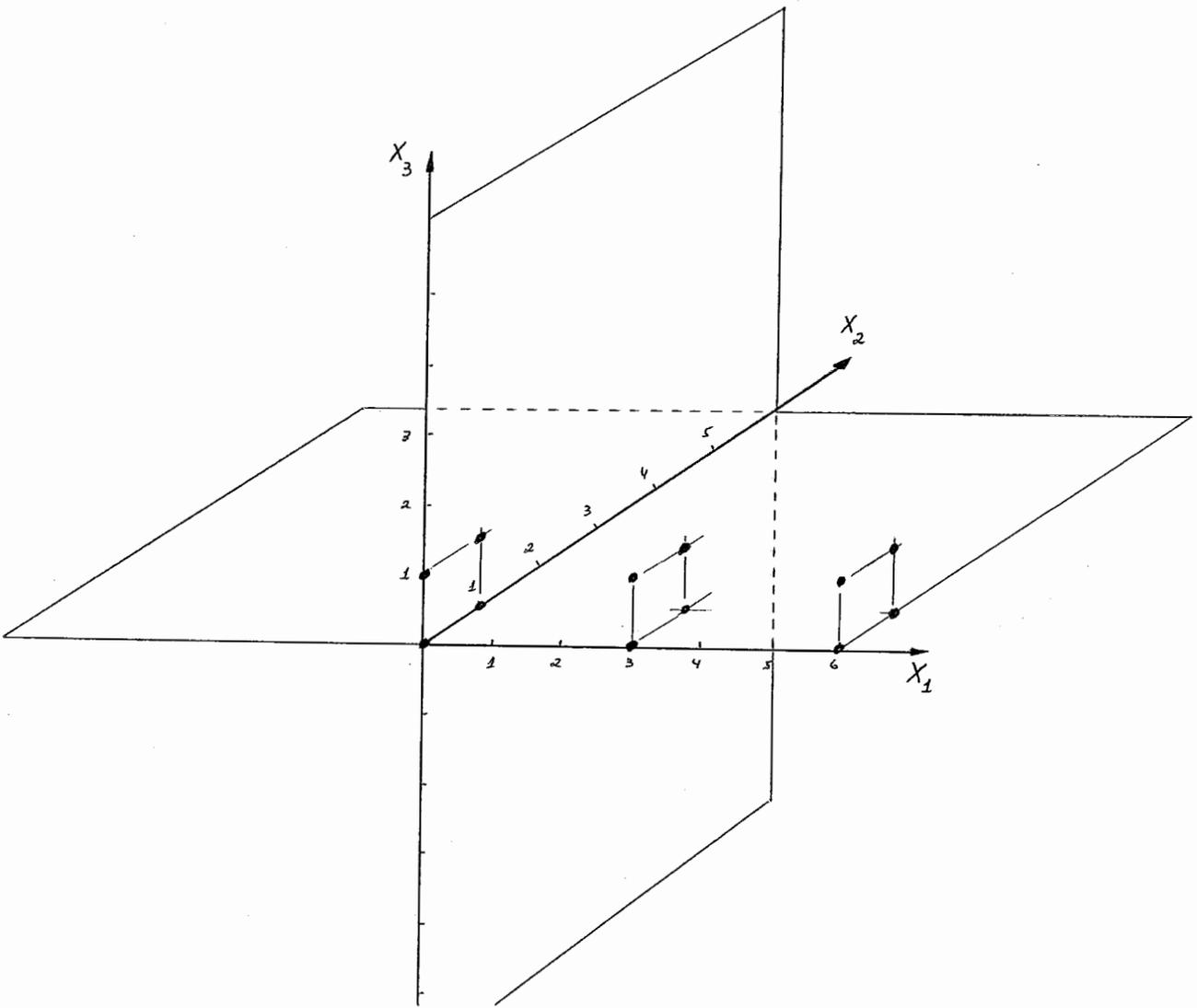


FIGURA 3.17.3: Representação gráfica do quadro 3.17.3.

SOLUÇÃO:

Inicialmente vamos encontrar os desvios x_1, x_2 e x_3 . Para isso calculamos as médias das variáveis

$$\bar{X}_1 \approx 3 \qquad \bar{X}_2 \approx 0,5 \qquad \bar{X}_3 \approx 0,5 \qquad (3.17.42)$$

Em seguida, subtraindo de cada $X_i(k)$ ($i=1,2,3$; $k=1,2,\dots,12$) a média \bar{X}_i , encontraremos os $x_i(k)$ que podem ser vistos no quadro 3.17.4 e na figura 3.17.4.

ELEMENTOS \ DESVIOS	1	2	3	4	5	6	7	8	9	10	11	12
x_1	-3	-3	-3	-3	0	0	0	0	-3	-3	-3	-3
x_2	-0,5	0,5	0,5	-0,5	-0,5	0,5	0,5	-0,5	-0,5	0,5	0,5	-0,5
x_3	-0,5	-0,5	0,5	0,5	-0,5	-0,5	0,5	0,5	-0,5	-0,5	0,5	0,5

QUADRO 3.17.4: Valores de 3 desvios observados em 12 elementos.

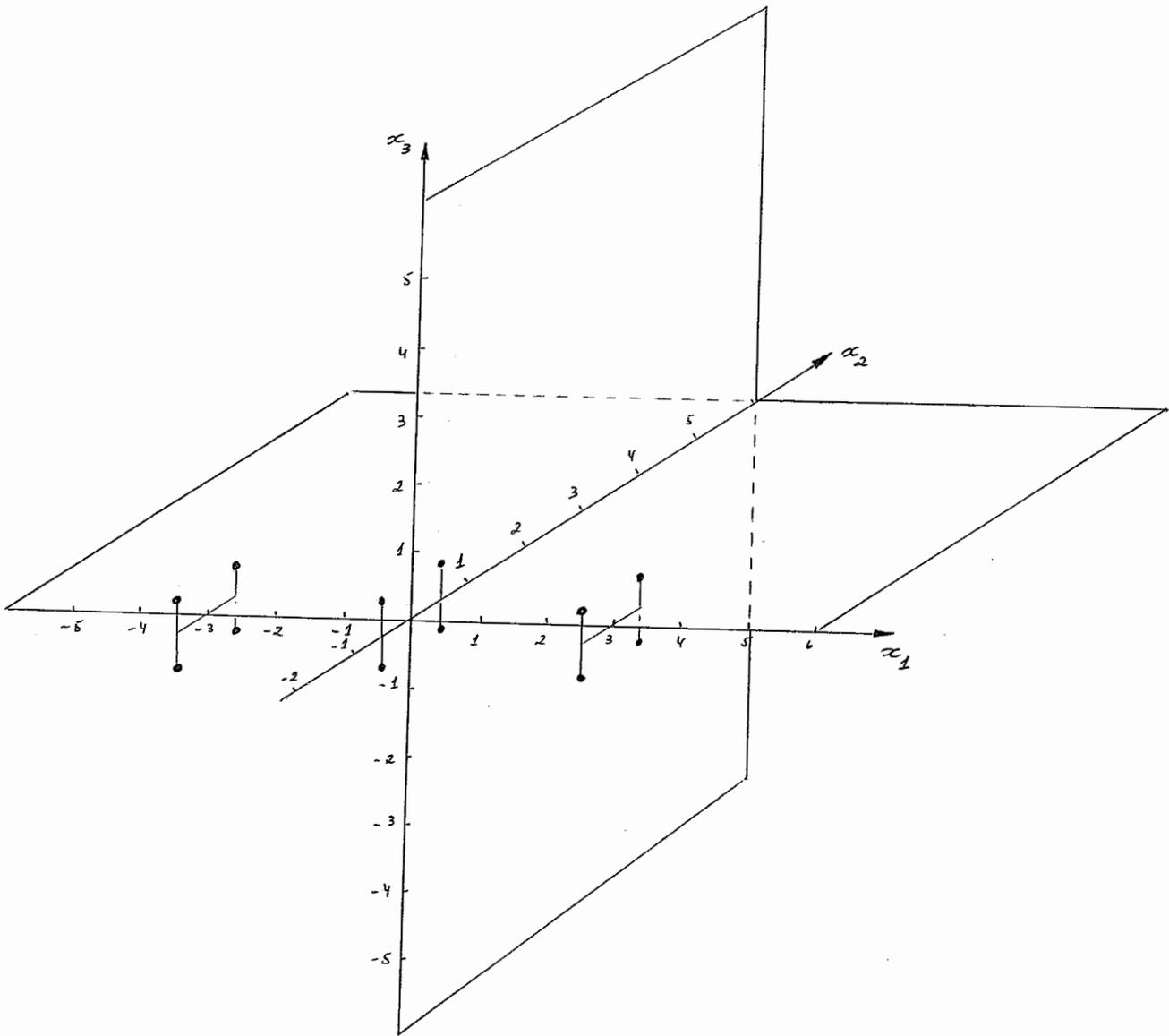


FIGURA 3.17.4: Valores de 3 desvios observados em 12 elementos.

Calculando então as variáveis e as covariâncias,

teremos:

$$\begin{aligned}
 \text{VAR}[x_1] &= 6 & \text{COV}[x_1;x_2] &= 0 \\
 \text{VAR}[x_2] &= 0,25 & \text{COV}[x_1;x_3] &= 0 \\
 \text{VAR}[x_3] &= 0,25 & \text{COV}[x_2;x_3] &= 0
 \end{aligned}
 \tag{3.17.43}$$

e a Matriz de Covariâncias ficará:

$$C = \begin{pmatrix} 6 & 0 & 0 \\ 0 & 0,25 & 0 \\ 0 & 0 & 0,25 \end{pmatrix}
 \tag{3.17.44}$$

Como C é uma matriz diagonal os seus autovalores são facilmente determinados:

$$\begin{aligned}
 \lambda_1 &= 6 \\
 \lambda_2 &= 0,25 \\
 \lambda_3 &= 0,25
 \end{aligned}
 \tag{3.17.45}$$

De posse dos autovalores, vamos encontrar os autovetores. Para isso vamos utilizar as equações 3.17.1 e 3.17.2 que transcrevemos abaixo.

$$(C - \lambda_j I) \mathcal{L}_j^T = \emptyset \quad j=1,2,\dots,n
 \tag{3.17.46}$$

$$\sum_{i=1}^n \mathcal{L}_{ji}^2 = 1 \quad j=1,2,\dots,n
 \tag{3.17.47}$$

No caso da Matriz de Covariâncias da equação 3.17.44 teremos:

$$\begin{pmatrix} 6-\lambda_j & 0 & 0 \\ 0 & 0,25-\lambda_j & 0 \\ 0 & 0 & 0,25-\lambda_j \end{pmatrix} \begin{pmatrix} \ell_{j1} \\ \ell_{j2} \\ \ell_{j3} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad (3.17.48)$$

$j=1,2,3$

$$\sum_{i=1}^3 \ell_{ji}^2 = 1 \quad j=1,2,3 \quad (3.17.49)$$

Substituindo o valor de cada autovalor (equações 3.17.45) teremos:

Para $\lambda_1 = 6$

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & -5,75 & 0 \\ 0 & 0 & -5,75 \end{pmatrix} \begin{pmatrix} \ell_{11} \\ \ell_{12} \\ \ell_{13} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad (3.17.50)$$

$$\ell_{11}^2 + \ell_{12}^2 + \ell_{13}^2 = 1 \quad (3.17.51)$$

logo $\ell_{11} = 1$

$$\ell_{12} = 0 \quad (3.17.52)$$

$$\ell_{13} = 0$$

que indica que:

$$y_1 = x_1 + 0 x_2 + 0 x_3 \quad (3.17.53)$$

ou seja a 1ª Componente é a própria variável x_1 .

Para $\lambda_2 = 0,25$

$$\begin{pmatrix} 5,75 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \ell_{21} \\ \ell_{22} \\ \ell_{23} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad (3.17.54)$$

$$\ell_{21}^2 + \ell_{22}^2 + \ell_{23}^2 = 1 \quad (3.17.55)$$

Neste caso o vetor $\ell_2 = (\ell_{21}; \ell_{22}; \ell_{23})$ não fica determinado pois qualquer vetor a forma

$$\ell_2 = (0, \ell_{22}, \ell_{23})$$

tal que:

$$\ell_{22}^2 + \ell_{23}^2 = 1 \quad (3.17.56)$$

é solução do sistema de equações 3.17.46-7, o mesmo acontecendo para o caso de λ_3 .

Podemos então encontrar 2 vetores ortogonais e unitários que satisfaçam o sistema de equações. Senão vejamos:

Seja:

$$\ell_2 = \left(0, \frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right) \quad (3.17.57)$$

e

$$\ell_3 = \left(0, \frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}\right) \quad (3.17.58)$$

Seu produto interno é nulo o que indica serem ℓ_2 e ℓ_3 ortogonais e sua norma igual a 1 o que indica que o autovetor é unitário. Assim definimos 3 componentes principais ortogonais entre si.

$$y_1 = x_1$$

$$y_2 = \frac{\sqrt{2}}{2} x_2 + \frac{\sqrt{2}}{2} x_3 \quad (3.17.59)$$

$$y_3 = \frac{\sqrt{2}}{2} x_2 - \frac{\sqrt{2}}{2} x_3 \quad \triangle$$

Porém poderíamos ter escolhido outros valores para ℓ_2 e ℓ_3 como por exemplo:

$$\ell_2 = \left(0, \frac{\sqrt{3}}{2}, \frac{1}{2}\right) \quad (3.17.60)$$

$$\ell_3 = \left(0, -\frac{1}{2}, \frac{\sqrt{3}}{2}\right) \quad (3.17.61)$$

e neste caso teríamos:

$$y_1 = x_1$$

$$y_2 = \frac{\sqrt{3}}{2} x_2 + \frac{1}{2} x_3 \quad (3.17.62)$$

$$y_3 = \frac{1}{2} x_2 + \frac{\sqrt{3}}{2} x_3$$

Essas soluções estão representadas na figura 3.17.5 juntamente com outras soluções possíveis.

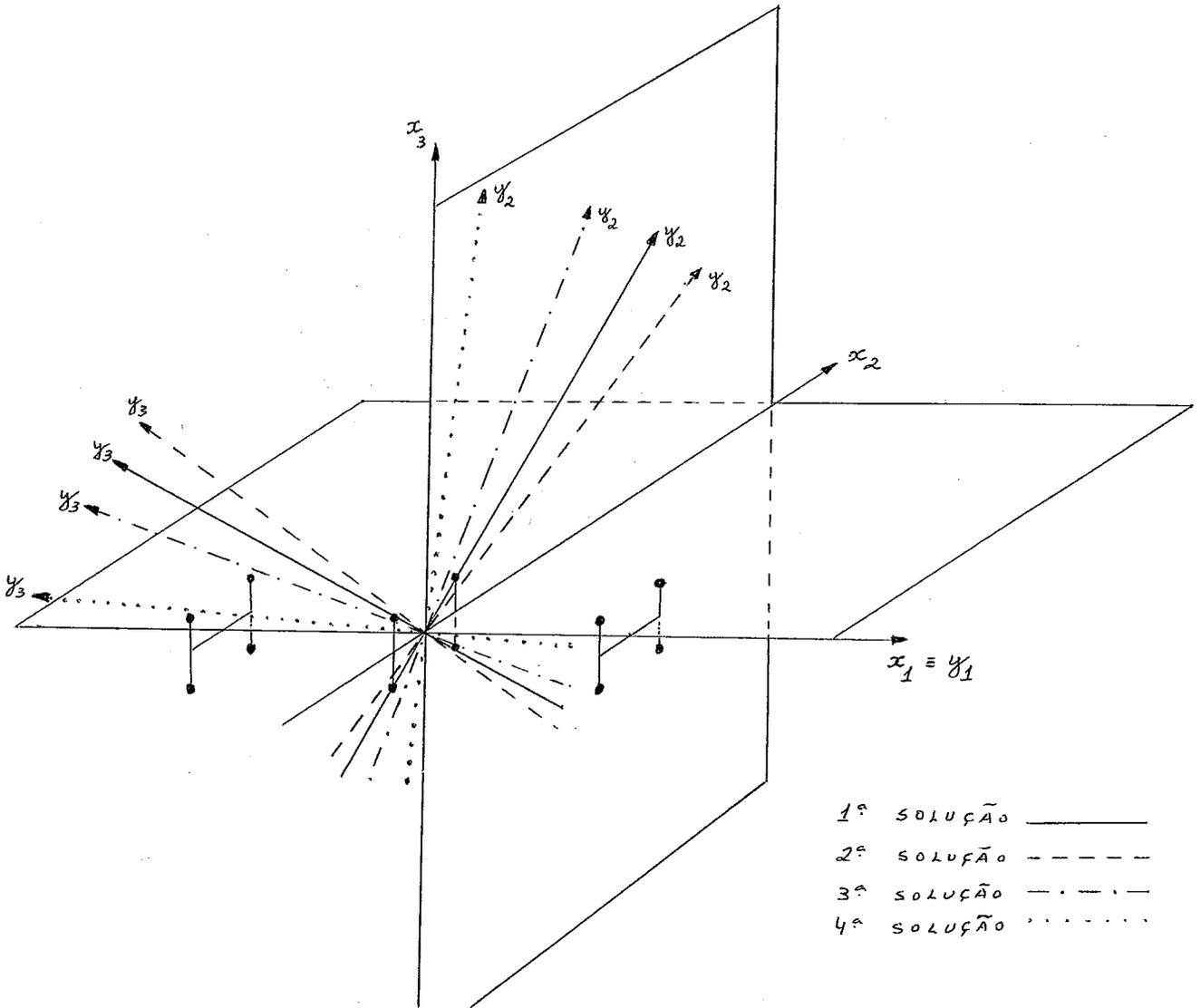


FIGURA 3.14.5: Representação gráfica das soluções do exemplo 3.17.2

Δ

No exemplo que acabamos de ver, podemos notar que há uma infinidade de y_2 e y_3 que são soluções do problema.

Geomêtricamente, KENDALL ([¹⁹], pág.20) compara estes casos com uma distribuição de pontos no formato de um "charuto" onde existe uma direção na qual há uma grande variância, a do comprimento do "charuto", e uma infinidade de maneiras de se fixar as duas direções restantes, cada uma delas

levando à mesma variância (Problema 1), a uma transformação ortogonal (Problema 2), e a minimização dos quadrados das distâncias dos pontos às retas que representam as componentes (Problema 3).

OBS.: Para o caso de igualdade de mais de 2 autovalores, por exemplo $\lambda_r = \lambda_s = \lambda_t$, poderia ser mostrado que seria possível encontrar componentes ortogonais entre si, satisfazendo as restrições do problema. Esses casos não serão aqui abordados, porém a forma de demonstração é semelhante ao caso aqui exposto e por isso deixamos a cargo do leitor.

Fim do 3º CASO \square

Existirão ainda casos em que poderá acontecer uma combinação dos casos aqui descritos. Por exemplo, um problema envolvendo seis variáveis, x_1, x_2, \dots, x_6 , onde $\lambda_1 \neq \lambda_2 = \lambda_3 = \lambda_4 \neq \lambda_5 = \lambda_6 = 0$. Nesses casos, inicialmente eliminamos as variáveis linearmente dependentes (2º caso. Nesse exemplo eliminaríamos 2 variáveis) e depois cairíamos no 3º caso.

3.18. Escalas

Se observamos a definição de COVARIÂNCIA, notaremos que a escala em que as variáveis são medidas tem influência sobre o valor das suas covariâncias. Senão vejamos.

Seja por exemplo uma população de N elementos onde medimos duas características X_1 e X_2 , ambas em quilôme-

tros. Seja essa mesma população e essas mesmas características, agora medidas em metros e denominadas X_3 e X_4 . Facilmente podemos notar que:

$$X_3(k) = 1000 X_1(k) \quad \forall k=1,2,\dots,n \quad (3.18.1)$$

$$X_4(k) = 1000 X_2(k) \quad \forall k=1,2,\dots,n \quad (3.18.2)$$

Assim a covariância entre X_1 e X_2 e entre X_3 e X_4 será:

$$\text{COV}[X_1;X_2] = \frac{1}{N} \sum_{k=1}^n [(X_1(k) - \bar{X}_1)(X_2(k) - \bar{X}_2)] \quad (3.18.3)$$

e

$$\text{COV}[X_3;X_4] = \frac{1}{N} \sum_{k=1}^n [(X_3(k) - \bar{X}_3)(X_4(k) - \bar{X}_4)] \quad (3.18.4)$$

Substituindo as equações 3.18.1-2 na 3.18.4 teremos:

$$\text{COV}[X_3;X_4] = \frac{1}{N} \sum_{k=1}^n [(1000X_1(k) - 1000\bar{X}_1)(1000X_2(k) - 1000\bar{X}_2)] \quad (3.18.5)$$

$$= \frac{1.000.000}{N} \sum_{k=1}^n [(X_1(k) - \bar{X}_1)(X_2(k) - \bar{X}_2)] \quad (3.18.6)$$

$$= 1.000.000 \text{ COV}[X_1;X_2] \quad (3.18.7)$$

Por outro lado como vimos nas seções anteriores, o MCP utiliza a Matriz de Covariâncias (C) como uma matriz que contém informações sobre as características da população que está sendo analisada. Sendo assim, mudando a escala da variável observada, muda a matriz C, e conseqüentemente, mudam as Componentes dessa análise.

Uma forma de evitar esse problema é utilizar uma medida adimensional de cada característica. Assim, ao invés de X_i ou x_i geralmente utilizamos a variável padronizada z_i (ver definição de variável padronizada na seção 2.3). Isto em outras palavras significa dizer que ao invés de utilizarmos a Matriz de Covariâncias estamos utilizando a Matriz de Correlações (ver equações 2.3.68-9 de onde se conclui que $COV[z_j; z_j,'] = COR[X_j; X_j,']$).

Um bom exemplo onde essa situação pode ser melhor compreendida é o exemplo 2.2. de KENDALL([¹⁹], pág.20).

ANEXOCOMENTÁRIO SOBRE O MÉTODO DOS MULTI-
PLICADORES DE LAGRANGE

Este método se baseia na anulação de derivadas. Sendo assim o máximo ou mínimo por ele fornecido não é necessariamente um máximo ou mínimo global. Ele fornece também máximos e mínimos locais.

Por isso, tanto no Problema 1 como no Problema 3 somente se estabelece, por esse método, que as soluções de cada uma das n etapas são autovalres. Para, no entanto, se verificar que a solução é o maior, o segundo maior, ..., o j -ésimo maior autovalor é necessário voltar ao problema de otimização.

Esta volta aparece no Problema 1 pelo teorema 3.4.5 que mostra que $\lambda_j = \text{VAR}[y_j]$. Com esse resultado e sabendo que $\text{COR}[y_j; y_{j'}] = 0 \quad \forall j = 1, 2, \dots, n$ e $\forall j' = 1, 2, \dots, (j-1)$ voltamos ao problema de otimização, cuja função objetiva é:

Maximizar $\text{VAR}[y_j]$

ou seja

Maximizar λ_j

concluindo portanto que λ_j é o j -ésimo maior autovalor.

Exatamente o mesmo acontece no Problema 3. Lá o teorema 3.12.6 garante que λ_j é o j -ésimo maior autovalor. Veja que para demonstrar este teorema tivemos que, mais uma vez, voltar ao problema de otimização, utilizando a função objetivo originalmente dada.

BIBLIOGRAFIA

- [¹] - AFFIFI, A.A. e AZEN, S.P. - "Statistical Analysis A Computer Oriented Approach", Academic Press, 1972.
- [²] - ANDERSON, T.W. - "An Introduction to Multivariate Statistical Analysis", John Wiley, 1958.
- [³] - BANACH, S. - "Théorie des Opérations Linéaires", Chelsea Publishing Company, 1932.
- [⁴] - BENZECRY, J.P. - "L'Analyse des Donnees. Vol.1. La Taxinomie", Dunod, 1973.
- [⁵] - BENZECRY, J.P. - "L'Analyse des Donnes. Vol.2. L'Analyse des Correspondences", Dunod, 1973.
- [⁶] - CARVALHO, J.P. - "Introdução à Álgebra Linear", IMPA, 1971.
- [⁷] - CHEN, C.T. - "Introduction to Linear System Theory", Holt, Rinehart and Winston, Inc., 1970.
- [⁸] - CLARKE, A.B. e R.L. DISNEY - "Probability and Random Processes for Engineers and Scientists", John Wiley, 1970.
- [⁹] - COLEY, W.W. e LOHNES, P.R. - "Multivariate Data Analises", John Wiley, 1971.
- [¹⁰] - DANTZIG, G.B. - "Linear Programming and Extensions", Princeton University Press, 1963.
- [¹¹] - HALLIDAY, D. e RESNICK, R. - "FÍSICA I", Livros Técnicos e Científicos Editora S.A., 1973 (tradução).
- [¹²] - HARMAN, H.H. - "Modern Factor Analysis", The University of Chicago Press, 1960.

- [¹³] - HARRIS, R.J. - "A Primer of Multivariate Statistics", Academic Press, Inc., 1975.
- [¹⁴] - HILLIER, F.S. e LIEBERMAN, G.J. - "Operations Research", Holden Days, Inc., 1967.
- [¹⁵] - HOFFMAN, K. e KUNZE, R. - "Algebra Linear", Editora Polígono, 1971 (tradução).
- [¹⁶] - HORST, P. - "Factor Analysis of Data Matrices", H.R. and Winston Inc., 1965.
- [¹⁷] - KENDALL, M.G. - "A Course in Multivariate Analysis", Charles Griffin & Company, 1965.
- [¹⁸] - KENDALL, M.G. e STUART, A. - "The Advanced Theory of Statistics", Charles Griffin & Company, 1966.
- [¹⁹] - KENDALL, M.G. - "Multivariate Analysis", Charles Griffin & Company, 1975.
- [²⁰] - KRISHNAIAH, P.R. - "Multivariate Analysis", Academic Press, 1966.
- [²¹] - LARSON, H.J. - "Introduction to Probability Theory and Statistical Inference", John Wiley, 1969.
- [²²] - LAWLEY, D.N. e MAXWELL, A.E. - "Factor Analysis as a Statistical Method", Butterworth & Co., 1963.
- [²³] - LIPSCHUTZ, S. - "Álgebra Linear", McGraw-Hill do Brasil, 1972.
- [²⁴] - McMILLAN, C.Jr. - "Mathematical Programming An Introduction to the Design and Application of Optimal Decision Machines", John Wiley, 1970.
- [²⁵] - MEYER, P.L. - "Probabilidade Aplicações à Estatística", Ao Livro Técnico, 1969 (tradução).
- [²⁶] - THRALL, R.M. e TORNHEIM, L. - "Vector Spaces and Matrices", John Wiley, 1957.