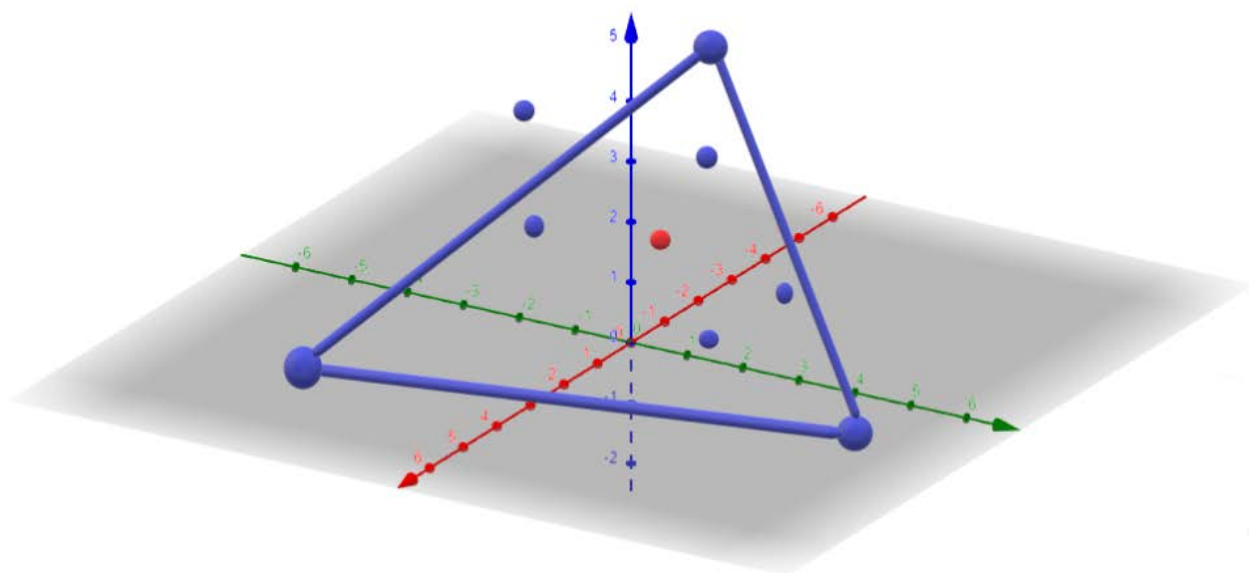


ANÁLISE MULTIVARIADA

APLICADA



Adair José Regazzi

Cosme Damião Cruz

Adair José Regazzi

D.S. em Estatística Experimental

Prof. Aposentado – Departamento de Estatística da UFV

adairreg@ufv.br

Cosme Damião Cruz

D.S. em Genética e Melhoramento

Prof. Aposentado – Departamento de Biologia Geral da UFV

cdcruz@ufv.br

ANÁLISE MULTIVARIADA APLICADA

Edição revista e ampliada

UNIVERSIDADE FEDERAL DE VIÇOSA – UFV

VIÇOSA – MG

2020

Prefácio

Existe uma infinidade de situações, nas quais o pesquisador está interessado no estudo de um número p de características ou variáveis sendo que, na maioria dos casos, interessa-nos o comportamento simultâneo de todas as variáveis e não o estudo isolado de cada uma delas. Neste sentido, a Análise Multivariada exerce papel fundamental dentro da Estatística Matemática, assim como em aplicações nas mais diversas áreas do conhecimento. Muito pouco era conhecido no campo da Análise Multivariada, em termos de teoria e aplicações, antes das investigações pioneiras realizadas por R. A. Fisher. Muitas contribuições de outros autores neste campo foram inspiradas no trabalho de Fisher, sendo este considerado “O Arquiteto da Análise Multivariada”.

O texto apresentado constitui notas de aulas de cursos ministrados pelos autores em várias instituições e em cursos de extensão. Versões anteriores (manuscritas ou digitadas) já foram disponibilizadas, mas não em sites oficiais. A ideia foi escrever um material introdutório que poderá ser utilizado em cursos de pós-graduação, mesmo para alunos que não tenham um curso de Álgebra de Matrizes e Modelos Lineares. Obviamente um conhecimento sobre Estatística Básica e Estatística Experimental é indispensável. Os aspectos didáticos e práticos, sem relaxar com a teoria, foram priorizados. Dos nove Capítulos apresentados, os Capítulos 1 e 2 fornecem base inicial sobre vetores e matrizes, inclusive sobre inversas generalizadas, o que facilita o entendimento dos demais Capítulos. Desse modo, serve de base para a leitura de textos mais avançados, tais como, Mardia et al. (1997), Johnson e Wichern (1998), Rencher (2002) e Ferreira (2008), entre outros.

Em nossa exposição, procuramos ser bastante didáticos, ilustrando com exemplos os principais conceitos. Ao final de todos os capítulos, apresentamos um conjunto de exercícios.

Escrever um livro ou mesmo uma apostila, por mais simples que seja, não é uma tarefa fácil, pois demanda bastante sacrifício e muito esforço. Nesse contexto, agradecemos a todos que contribuíram de alguma forma, principalmente nossos estudantes de pós-graduação.

Os autores

ÍNDICE		Página
CAPÍTULO 1		1
RESULTADOS PRELIMINARES		1
1.1.	INTRODUÇÃO	1
1.2.	MATRIZES E VETORES	1
1.2.1.	Matrizes, vetores e escalares	1
1.2.2.	Igualdade de matrizes	2
1.2.3.	Matriz quadrada	2
1.2.4.	Matriz transposta	2
1.2.5.	Alguns tipos especiais de matrizes	3
1.3.	OPERAÇÕES COM MATRIZES	4
1.3.1.	Adição e subtração de matrizes	4
1.3.2.	Produto de duas matrizes	5
1.3.3.	Soma direta	7
1.3.4.	Produto direto ou de Kronecker	7
1.3.5.	Matriz idempotente	8
1.4.	POSTO DE UMA MATRIZ	8
1.5.	DETERMINANTES	9
1.6.	INVERSA DE UMA MATRIZ	13
1.7.	AUTOVALORES E AUTOVETORES	16
1.8.	MATRIZ ORTOGONAL	20
1.9.	TRAÇO DE UMA MATRIZ	22
1.10.	FORMAS QUADRÁTICAS	23
1.10.1.	Introdução	23
1.10.2.	Classificação das formas quadráticas	25
1.11.	VETORES DE OPERADORES DIFERENCIAIS	28
1.12.	SISTEMAS DE EQUAÇÕES E INVERSAS GENERALIZADAS	31
1.12.1.	Sistemas de equações lineares com matrizes não singulares	31
1.12.2.	Sistemas de equações lineares com matrizes quaisquer	32
1.12.3.	Inversas generalizadas	33
1.13.	CONSISTÊNCIA EM SISTEMAS DE EQUAÇÕES LINEARES	42
1.14.	EXERCÍCIOS	45
1.15.	REFERÊNCIAS	47
CAPÍTULO 2		49
CONCEITOS INTRODUTÓRIOS		49
2.1.	VETORES E MATRIZES ALEATÓRIAS	49
2.2.	VETORES DE MÉDIAS, MATRIZES DE COVARIÂNCIAS E DE CORRELAÇÕES	49
2.3.	MÉDIA, VARIÂNCIA, COVARIÂNCIA E CORRELAÇÃO (Amostral)	54
2.4.	EXERCÍCIOS	59
2.5.	REFERÊNCIAS	60
CAPÍTULO 3		61
DISTRIBUIÇÕES MULTIVARIADAS		61
3.1.	DISTRIBUIÇÃO NORMAL MULTIVARIADA	61
3.1.1.	Definição	61
3.1.2.	Função geradora de momentos	65

ÍNDICE		Página
3.1.3.	Distribuição marginal e condicional	69
3.1.4.	Correlações parciais	71
3.2.	DISTRIBUIÇÃO DE WISHART	82
3.2.1.	Introdução	82
3.2.2.	Propriedades	84
3.3.	DISTRIBUIÇÃO DE HOTELLING	84
3.3.1.	Introdução	84
3.3.2.	Teste de um vetor de médias contra um padrão populacional	85
3.3.3.	Teste dos vetores de médias de duas populações	89
3.4.	EXERCÍCIOS	93
3.5.	REFERÊNCIAS	97
CAPÍTULO 4		98
ANÁLISE DE VARIÂNCIA MULTIVARIADA E VARIÁVEIS CANÔNICAS		98
4.1.	INTRODUÇÃO	98
4.2.	ANÁLISE DE VARIÂNCIA MULTIVARIADA EM BLOCOS CASUALIZADOS	98
4.2.1.	Modelo estatístico	98
4.2.2.	Matrizes de sistemas de equações normais	101
4.2.3.	Estimabilidade	102
4.2.4.	Decomposição das somas de quadrados e de produtos totais	103
4.3.	TESTES DE HIPÓTESES	110
4.3.1.	Teste de Wilks	111
4.3.2.	Teste de Pillai	113
4.3.3.	Teste de Hotelling-Lawley	114
4.3.4.	Teste de Roy	115
4.4.	COMPARAÇÕES MÚLTIPLAS	117
4.4.1.	Teste T^2 de Hotelling	117
4.4.2.	Teste de Roy e Bose	119
4.4.3.	Intervalo de confiança simultâneo de Bonferroni	120
4.4.4.	Função discriminante linear de Fisher	122
4.5.	EXEMPLO DE UMA ANÁLISE DE VARIÂNCIA MULTIVARIADA DE UM EXPERIMENTO EM BLOCOS COMPLETOS CASUALIZADOS	126
4.6.	RESULTADOS OBTIDOS ATRAVÉS DO PROGRAMA SAS – EXEMPLO 1	127
4.7.	ANÁLISE DE VARIÂNCIA MULTIVARIADA DE UM EXPERIMENTO INTEIRAMENTE CASUALIZADO COM DESENVOLVIMENTO MATRICIAL – EXEMPLO 2	136
4.8.	RESULTADOS OBTIDOS ATRAVÉS DO PROGRAMA SAS – EXEMPLO 2	145
4.9.	EXPERIMENTO FATORIAL MULTIVARIADO	153
4.9.1.	Resultados obtidos através do programa SAS	158
4.10.	VARIÁVEIS CANÔNICAS	163
4.10.1.	Sobre o descarte de variáveis	171
4.10.2.	Resultados obtidos através do programa SAS	176
4.11.	EXERCÍCIOS	186
4.12.	REFERÊNCIAS	188
CAPÍTULO 5		190
ANÁLISE DE COMPONENTES PRINCIPAIS		190

	ÍNDICE	Página
5.1.	INTRODUÇÃO	190
5.2.	COMPONENTES PRINCIPAIS	190
5.2.1.	Solução utilizando a matriz de correlação R	194
5.2.2.	Solução utilizando a matriz de covariância S	197
5.2.3.	Algumas considerações	206
5.2.4.	Exemplos de aplicação	208
5.2.4.1.	Exemplo 1 - Componentes principais a partir de uma matriz de covariância 2x2	208
5.2.4.2.	Exemplo 2 - Componentes principais a partir de uma matriz de correlação 2x2	213
5.2.4.3.	Exemplo 3 - Componentes principais a partir de uma matriz de correlação 3x3	217
5.2.5	Resultados obtidos através do programa SAS	229
5.2.5.1.	Exemplo 1	229
5.2.5.2.	Exemplo 2	231
5.2.5.3.	Exemplo 3	234
5.3.	EXERCÍCIOS	238
5.4.	REFERÊNCIAS	239
	CAPÍTULO 6	240
	ANÁLISE DE AGRUPAMENTO	240
6.1.	INTRODUÇÃO	240
6.2.	MEDIDAS DE SIMILARIDADE E DE DISSIMILARIDADE	241
6.2.1.	Medidas de dissimilaridade	242
6.2.1.1.	Distância euclidiana	243
6.2.1.2.	Distância euclidiana média	243
6.2.1.3.	Distância ponderada	244
6.2.1.4.	Distância de Gower	245
6.2.1.5.	Distância de Mahalanobis	245
6.2.2.	Medidas de similaridade	246
6.3.	MÉTODOS DE AGRUPAMENTO	247
6.3.1.	Métodos hierárquicos	247
6.3.1.1.	Método do vizinho mais próximo	248
6.3.1.2.	Método do vizinho mais distante	249
6.3.1.3.	Método da ligação média entre grupos	249
6.3.1.4.	Método da variância mínima ou método de Ward	259
6.3.2.	Métodos de otimização	272
6.4.	EXEMPLO UTILIZANDO O SAS	294
6.5.	EXERCÍCIO	296
6.6.	REFERÊNCIAS	296
	CAPÍTULO 7	298
	ANÁLISE DISCRIMINANTE	298
7.1.	INTRODUÇÃO	298
7.2.	CLASSIFICAÇÃO EM UMA DE DUAS POPULAÇÕES	299
7.2.1.	Função discriminante linear de Fisher	299
7.2.2.	Avaliação da função discriminante	304
7.2.3.	Métodos de estimação das probabilidades de má classificação	305

	ÍNDICE	Página
7.2.3.1.	Método devido a Okamoto (1963)	305
7.2.3.2.	Método devido a Smith (1947)	305
7.2.4.	Exemplo	306
7.3.	CLASSIFICAÇÃO EM UMA DE VÁRIAS POPULAÇÕES: FUNÇÕES DISCRIMINANTES DE ANDERSON (1958)	312
7.3.1.	Introdução	312
7.3.2.	Estimação das funções discriminantes e classificação	312
7.3.3.	Distância de Mahalanobis entre dois grupos	318
7.3.4.	Exemplo	318
7.4.	EXEMPLO UTILIZANDO O SAS (Dados da Tabela 7.1)	331
7.5.	Análise discriminante baseada em componentes principais	338
7.6.	EXERCÍCIOS	346
7.7.	REFERÊNCIAS	349
	CAPÍTULO 8	351
	CORRELAÇÕES CANÔNICAS	351
8.1.	INTRODUÇÃO	351
8.2.	ESTIMAÇÃO DAS CORRELAÇÕES CANÔNICAS E DOS PARES CANÔNICOS	352
8.2.1.	Autovalores e autovetores da matriz $R_{11}^{-1}R_{12}R_{22}^{-1}R_{21}$	361
8.3.	TESTES DE SIGNIFICÂNCIA	361
8.4.	EXEMPLOS	364
8.5.	EXERCÍCIOS	376
8.6.	REFERÊNCIAS	377
	CAPÍTULO 9	378
	ANÁLISE DE FATORES	378
9.1.	INTRODUÇÃO	378
9.2.	O MODELO DA ANÁLISE DE FATORES ORTOGONAL COM m FATORES COMUNS	378
9.3.	MÉTODOS DE ESTIMAÇÃO DAS CARGAS FATORIAIS	384
9.3.1.	Método do componente principal	384
9.3.2.	Método da máxima verossimilhança	385
9.4.	ROTAÇÃO DOS FATORES	386
9.5.	ESCORES FATORIAIS	386
9.6.	COMPARAÇÃO ENTRE AS TÉCNICAS ANÁLISE DE FATORES E ANÁLISE DE COMPONENTES PRINCIPAIS	387
9.7.	EXEMPLO	388
9.8.	EXERCÍCIO	400
9.9.	REFERÊNCIAS	401

CAPÍTULO 1

RESULTADOS PRELIMINARES

1.1. INTRODUÇÃO

Com o objetivo de facilitar o entendimento dos Métodos de Análise Estatística Multivariada apresentados no texto, são introduzidos aqui alguns conceitos básicos de uma maneira bastante simples, ou seja, sem maiores demonstrações algébricas. Obviamente que, para o leitor que possui conhecimentos sobre Álgebra de Matrizes e Modelos Lineares, este capítulo é quase que desnecessário.

1.2. MATRIZES E VETORES

1.2.1. Matrizes, vetores e escalares

Chama-se **matriz** de ordem m por n a um quadro de mxn elementos (números, letras etc.) dispostos em m linhas e n colunas:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

Cada elemento da matriz A tem dois índices, o primeiro de linha e o segundo de coluna. Assim, a_{12} é o elemento da 1ª linha e 2ª coluna.

A matriz A pode ser representada abreviadamente por $A = (a_{ij})$, com $i=1,2,\dots,m$ e $j=1,2,\dots,n$. A dimensão ou a ordem da matriz é especificada pelo número de linhas e colunas que ela contém. Por exemplo, a matriz B dada a seguir, possui dimensão 2x3, ou seja, possui duas linhas e três colunas. Para indicar a dimensão da matriz, escreve-se simplesmente ${}_2B_3$ ou $B_{(2 \times 3)}$.

$$B = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}$$

Um **vetor (vetor coluna)** é uma matriz de ordem $n \times 1$. Por exemplo, \underline{y} é um vetor 3×1 , sendo

$$\underline{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

Um **vetor linha** é uma matriz de ordem $1 \times n$ e aproveitando o exemplo anterior, será utilizada a seguinte notação:

$$\underline{y}' = [y_1, y_2, y_3] = [y_1 \quad y_2 \quad y_3]$$

No contexto de matrizes e vetores, um número real é chamado de um **escalar**. Assim, 2 e 10 são escalares.

1.2.2. Igualdade de matrizes

Dois matrizes são iguais quando têm as mesmas dimensões e os elementos de posições correspondentes são iguais. Por exemplo:

$$\begin{bmatrix} 1 & 3 & -2 \\ 2 & 4 & 6 \end{bmatrix} = \begin{bmatrix} 1 & 3 & -2 \\ 2 & 4 & 6 \end{bmatrix}$$

1.2.3. Matriz quadrada

Quando o número de linhas é igual ao número de colunas, tem-se uma **matriz quadrada**. A ordem da matriz quadrada é $n \times n$, ou simplesmente n . Por exemplo:

$$A_{(3)} = \begin{bmatrix} 4 & 2 & 2 \\ 2 & 2 & 0 \\ 2 & 0 & 2 \end{bmatrix}$$

1.2.4. Matriz transposta

Uma matriz A' é dita **transposta** de A , se todos os elementos das colunas de A' forem iguais aos elementos das linhas de A . Por exemplo:

$$\text{Se } A = \begin{bmatrix} 1 & 3 & -2 \\ 2 & 4 & 6 \end{bmatrix} \Rightarrow A' = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ -2 & 6 \end{bmatrix} \text{ é a sua transposta.}$$

Para A de ordem $m \times n \Rightarrow A'$ de ordem $n \times m$. Tem-se ainda que $(A')' = A$, isto é, a transposta da transposta de uma matriz é ela mesma.

1.2.5. Alguns tipos especiais de matrizes

Matriz simétrica: Chama-se matriz simétrica à matriz quadrada em que os elementos $a_{ij} = a_{ji}$ ou, equivalentemente, $A' = A$. Por exemplo,

$$A = \begin{bmatrix} 6 & 4 & 2 \\ 4 & 4 & 0 \\ 2 & 0 & 2 \end{bmatrix}$$

é simétrica.

Matriz diagonal: Uma matriz D quadrada, é dita diagonal se os elementos $d_{ij} = 0$ quando $i \neq j$, ou seja, se todos os elementos situados fora da diagonal principal são nulos. Por exemplo,

$$D1 = \begin{bmatrix} 2 & 0 \\ 0 & 8 \end{bmatrix} \quad \text{e} \quad D2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$

são matrizes diagonais, que também podem ser denotadas como $D1 = \text{diag}(2, 8)$ e $D2 = \text{diag}(1, 0, 4)$, respectivamente.

Pode-se usar a notação $\text{diag}(A)$ para indicar a matriz diagonal com os mesmos elementos da diagonal A , como por exemplo,

$$A = \begin{bmatrix} -1 & 3 & 0 \\ 3 & 4 & 1 \\ 0 & 1 & 5 \end{bmatrix} \Rightarrow \text{diag}(A) = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 5 \end{bmatrix}$$

Matriz identidade: Toda matriz quadrada onde os elementos da diagonal principal são iguais a 1 e os demais são nulos, é dita **matriz identidade**. Indica-se a matriz identidade por $I_{(n)}$, ou simplesmente I . Por exemplo,

$$I_{(2)} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{e} \quad I_{(3)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

são matrizes identidades.

Matriz triangular superior: Uma matriz triangular superior é uma matriz quadrada onde todos os elementos abaixo da diagonal principal são nulos, isto é, $m = n$ e $a_{ij} = 0$ para $i > j$, como por exemplo,

$$A = \begin{bmatrix} 1 & 5 & 7 \\ 0 & 4 & 2 \\ 0 & 0 & 9 \end{bmatrix}.$$

A notação para um vetor de zeros será $\underline{0}$ e para uma matriz de zeros será ϕ , como por exemplo,

$$\underline{0} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad \text{e} \quad \phi = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

1.3. OPERAÇÕES COM MATRIZES

1.3.1. Adição e subtração de matrizes

A soma de duas matrizes A e B de dimensão $m \times n$ é uma matriz ${}_m C_n$ tal que $c_{ij} = a_{ij} + b_{ij}$.

Assim, se

$$A = \begin{bmatrix} 1 & 4 & 2 \\ -1 & 0 & 5 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & -2 & 1 \\ 4 & 5 & 0 \end{bmatrix},$$

então,

$$C = A + B = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 5 & 5 \end{bmatrix}$$

A diferença $A - B$ de duas matrizes de dimensão $m \times n$ é uma matriz ${}_m C_n$ tal que $c_{ij} = a_{ij} - b_{ij}$. Assim, do exemplo anterior, segue que:

$$C = A - B = \begin{bmatrix} 1 & 6 & 1 \\ -5 & -5 & 5 \end{bmatrix}$$

Se A , B e C são matrizes de dimensão $m \times n$, então:

- (i) $A + B = B + A$
- (ii) $A + \phi = \phi + A = A$
- (iii) $-A + A = A - A = \phi$

$$(iv) \quad A + (B+C) = (A + B) + C$$

$$(v) \quad (A+B)' = A' + B'$$

1.3.2. Produto de duas matrizes

Sejam as matrizes ${}_m A_n = (a_{ij})$ e ${}_n B_p = (b_{kj})$. O produto AB é uma matriz

$${}_m C_p = AB = (c_{ij}) \text{ tal que } c_{ij} = \sum_{k=1}^n a_{ik} b_{kj}.$$

Note que o produto de duas matrizes só é possível quando o número de colunas da matriz que pré-multiplica for igual ao número de linhas da matriz que pós-multiplica. Por exemplo,

$${}_2 A_3 = \begin{bmatrix} 2 & 1 & 1 \\ 3 & 0 & 1 \end{bmatrix} \text{ e } {}_3 B_2 = \begin{bmatrix} 0 & 1 \\ 4 & 2 \\ 1 & 3 \end{bmatrix}.$$

$$\text{Assim, } AB = \begin{bmatrix} (2)(0) + (1)(4) + (1)(1) & (2)(1) + (1)(2) + (1)(3) \\ (3)(0) + (0)(4) + (1)(1) & (3)(1) + (0)(2) + (1)(3) \end{bmatrix} = \begin{bmatrix} 5 & 7 \\ 1 & 6 \end{bmatrix}$$

$$\text{e } BA = \begin{bmatrix} 3 & 0 & 1 \\ 14 & 4 & 6 \\ 11 & 1 & 4 \end{bmatrix}.$$

Mesmo nos casos onde os produtos AB e BA sejam possíveis e tenham a mesma dimensão, em geral, $AB \neq BA$. A multiplicação matricial não é, em geral, comutativa.

Se k é um escalar, o **produto** de uma matriz ${}_m A_n = (a_{ij})$ por esse escalar é uma matriz ${}_m B_n = k {}_m A_n = (b_{ij})$ tal que $b_{ij} = k a_{ij}$. Assim, se

$$k = 5, \quad A = \begin{bmatrix} 1 & -1 \\ 2 & 4 \end{bmatrix},$$

então

$$B = 5 \begin{bmatrix} 1 & -1 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} 5 & -5 \\ 10 & 20 \end{bmatrix}$$

A multiplicação de matrizes é distributiva em relação à soma e subtração:

$$(i) \quad A(B \pm C) = AB \pm AC$$

$$(ii) \quad (A \pm B)C = AC \pm BC$$

A multiplicação envolvendo vetores segue as mesmas regras das matrizes. Assim, se

$$\underline{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \underline{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$$

então,

$$\underline{y}'\underline{y} = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = y_1^2 + y_2^2 + \cdots + y_n^2 = \sum_{i=1}^n y_i^2$$

$$\underline{X}'\underline{y} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} y_1 + y_2 + \cdots + y_n \\ x_1 y_1 + x_2 y_2 + \cdots + x_n y_n \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \end{bmatrix}$$

Propriedades da matriz transposta:

- (i) $(\underline{A} + \underline{B})' = \underline{A}' + \underline{B}'$
- (ii) $(k\underline{A})' = k\underline{A}'$
- (iii) $(\underline{A}')' = \underline{A}$
- (iv) $(\underline{AB})' = \underline{B}'\underline{A}'$

As verificações destas propriedades são imediatas.

Norma de um vetor real: A norma ou comprimento de um vetor real

$$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix},$$

representada por $\|\underline{x}\|$, é o número real não-negativo dado por:

$$\begin{aligned} \|\underline{x}\| &= \sqrt{\underline{x}'\underline{x}} \\ &= \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} \\ &= \left(\sum_{i=1}^n x_i^2 \right)^{1/2}. \end{aligned}$$

Para $\underline{x} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$, sua norma é

$\|\underline{x}\| = \sqrt{2^2 + 3^2} = \sqrt{4+9} = \sqrt{13}$. Assim, $\sqrt{13}$ é o comprimento do vetor \underline{x} , que geometricamente é a distância entre um ponto $P = (2, 3)$ e a origem $O = (0, 0)$.

A partir de cada vetor $\underline{x} \neq \underline{0}$ é possível obter um vetor normalizado (ou unitário) \underline{y} fazendo

$$\underline{y} = \frac{\underline{x}}{\|\underline{x}\|}.$$

Assim, $\underline{y}'\underline{y} = 1$, ou seja, o comprimento do vetor normalizado \underline{y} é igual a 1. Logo, a forma normalizada de \underline{x} do exemplo anterior é a seguinte:

$$\underline{y} = \begin{bmatrix} 2/\sqrt{13} \\ 3/\sqrt{13} \end{bmatrix}.$$

1.3.3. Soma direta

Dadas as matrizes ${}_m A_n$ e ${}_r B_s$, define-se a **soma direta** entre elas, como a matriz

$${}_{m+r} C_{n+s} = A \oplus B = \begin{bmatrix} A & \phi \\ \phi & B \end{bmatrix}. \quad \phi \text{ pode ser um vetor ou uma matriz nula. Por exemplo:}$$

Sejam as matrizes

$$A = \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix} \quad \text{e} \quad B = \begin{bmatrix} 3 & 5 & 4 & 3 \\ 1 & 4 & 2 & 1 \\ 4 & 2 & 2 & 8 \end{bmatrix},$$

então

$$C = \begin{bmatrix} 2 & 3 & 0 & 0 & 0 & 0 \\ 1 & 4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 5 & 4 & 3 \\ 0 & 0 & 1 & 4 & 2 & 1 \\ 0 & 0 & 4 & 2 & 2 & 8 \end{bmatrix}$$

1.3.4. Produto direto ou de Kronecker

Dadas as matrizes ${}_m A_n$ e ${}_r B_s$, define-se o **produto direto** ou **produto de Kronecker** entre elas, como a matriz

$${}_{mr}C_{ns} = A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{bmatrix}.$$

Exemplo: Sejam as matrizes

$$A = \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix} \quad \text{e} \quad B = \begin{bmatrix} 1 & 2 & 3 \\ 5 & 0 & -2 \end{bmatrix},$$

então

$$C = \begin{bmatrix} 2 & 4 & 6 & 3 & 6 & 9 \\ 10 & 0 & -4 & 15 & 0 & -6 \\ 1 & 2 & 3 & 4 & 8 & 12 \\ 5 & 0 & -2 & 20 & 0 & -8 \end{bmatrix}$$

1.3.5. Matriz idempotente

Dada uma matriz P , tal que $P^2 = PP = P$, então P é uma matriz idempotente.

Exemplo: Seja a matriz

$$P = \begin{bmatrix} 2 & -1 & 1 \\ -3 & 4 & -3 \\ -5 & 5 & -4 \end{bmatrix},$$

então

$$P^2 = \begin{bmatrix} 2 & -1 & 1 \\ -3 & 4 & -3 \\ -5 & 5 & -4 \end{bmatrix} \begin{bmatrix} 2 & -1 & 1 \\ -3 & 4 & -3 \\ -5 & 5 & -4 \end{bmatrix} = \begin{bmatrix} 2 & -1 & 1 \\ -3 & 4 & -3 \\ -5 & 5 & -4 \end{bmatrix}$$

Logo, a matriz P é idempotente, uma vez que $P^2 = P$.

1.4. POSTO DE UMA MATRIZ

Antes de apresentar a definição de posto de uma matriz, são colocados alguns conceitos de uma maneira bastante simples, de modo a facilitar o entendimento.

Combinação linear: Sejam os vetores $\underline{v}_1, \underline{v}_2, \dots, \underline{v}_p$ de um espaço vetorial V e os escalares

c_1, c_2, \dots, c_p . Qualquer vetor $\underline{v} \in V$ da forma

$$\underline{v} = c_1 \underline{v}_1 + c_2 \underline{v}_2 + \cdots + c_p \underline{v}_p$$

é uma **combinação linear** dos vetores $\underline{v}_1, \underline{v}_2, \dots, \underline{v}_p$.

Vetores linearmente dependentes e linearmente independentes: Um conjunto de vetores $\underline{v}_1, \underline{v}_2, \dots, \underline{v}_p$ é dito **linearmente dependente** (l.d.), se existirem escalares c_1, c_2, \dots, c_p (nem todos nulos) de tal forma que

$$c_1 \underline{v}_1 + c_2 \underline{v}_2 + \cdots + c_p \underline{v}_p = \underline{0}.$$

Caso contrário, o conjunto de vetores é dito **linearmente independentes** (l. i.).

Exemplo: Seja a matriz

$$A = \begin{bmatrix} 2 & 1 & 4 \\ -1 & 2 & -7 \\ 3 & 4 & 1 \end{bmatrix},$$

então as colunas de A são l.d., pois

$$3 \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix} - 2 \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix} - \begin{bmatrix} 4 \\ -7 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

O **posto**, também chamado característica ou “rank” de uma matriz A, quadrada ou não, é igual ao número de linhas ou colunas linearmente independentes. Logo, a matriz A do exemplo anterior tem posto 2, isto é, $\text{posto}(A) = 2$, porque o número de colunas l.i. é 2. Pode-se mostrar que o número de colunas l.i. de qualquer matriz é igual ao número de linhas l.i. dessa matriz.

Dada uma matriz retangular ${}_n A_p$. Então $\text{posto}(A) \leq \min(n, p)$. Se $\text{posto}(A) = n$, diz-se que A é de posto linha completo e, se $\text{posto}(A) = p$, diz-se que A é de posto coluna completo.

1.5. DETERMINANTES

Não será dada a definição geral de determinante por ela não ser muito útil para o cálculo do determinante de uma matriz. É interessante ressaltar que só há determinante de matriz quadrada.

A representação do determinante de uma matriz A quadrada de ordem n, que será designado por $\det(A)$, faz-se de maneira análoga à da matriz, colocada entre dois traços verticais, como a seguir:

$$\det(A) = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}.$$

(i) Para $A = [a_{11}]$, $\det(A) = |a_{11}| = a_{11}$.

Exemplo: Se

$$A = [3] \text{ e } B = [-10],$$

então

$$|A| = 3 \text{ e } |B| = -10.$$

(ii) Cálculo do determinante de 2ª ordem

O determinante da matriz $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ pode ser obtido por:

$$\det(A) = |A| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}.$$

Exemplo: Se

$$A = \begin{bmatrix} 7 & 5 \\ 2 & 4 \end{bmatrix}$$

então

$$\det(A) = |A| = \begin{vmatrix} 7 & 5 \\ 2 & 4 \end{vmatrix} = 7(4) - 5(2) = 28 - 10 = 18.$$

(iii) Cálculo do determinante de 3ª ordem

O determinante da matriz

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

pode ser calculado pela regra de Sarrus, como indicado a seguir:

$$\det(A) = |A| = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \begin{array}{cc} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{array}$$

- - - + + +

$$=a_{11}a_{22}a_{33}+a_{12}a_{23}a_{31}+a_{13}a_{21}a_{32}-a_{13}a_{22}a_{31}-a_{11}a_{23}a_{32}-a_{12}a_{21}a_{33}$$

Exemplo: Se

$$A=\begin{bmatrix} 2 & 5 & 7 \\ 3 & 1 & 4 \\ 6 & 8 & 2 \end{bmatrix}$$

então

$$\begin{aligned} \det(A)=|A| &= \begin{vmatrix} 2 & 5 & 7 \\ 3 & 1 & 4 \\ 6 & 8 & 2 \end{vmatrix} \begin{vmatrix} 2 & 5 \\ 3 & 1 \\ 6 & 8 \end{vmatrix} \\ &= 2 \times 1 \times 2 + 5 \times 4 \times 6 + 7 \times 3 \times 8 - 7 \times 1 \times 6 - 2 \times 4 \times 8 - 5 \times 3 \times 2 \\ &= 4 + 120 + 168 - 42 - 64 - 30 \\ &= 292 - 136 \\ &= 156 \end{aligned}$$

(iv) Cálculo do determinante de matrizes de maiores dimensões

Seja A uma matriz quadrada de ordem n . O determinante de $A = (a_{ij})$ para $i, j = 1, 2, \dots, n$, pode ser expandido como:

$$\det(A)=|A|=\sum_{j=1}^n a_{ij} (-1)^{i+j} |M_{ij}|, \text{ para qualquer } i,$$

onde $|M_{ij}|$ é o menor de a_{ij} , que é o determinante da matriz obtida de A , suprimindo-se a i -ésima linha e a j -ésima coluna. Esta é a expansão de um determinante pelos elementos de uma linha.

Quando o determinante é expandido pelos elementos de uma coluna, temos:

$$\det(A)=|A|=\sum_{i=1}^n a_{ij} (-1)^{i+j} |M_{ij}|, \text{ para qualquer } j.$$

Para a matriz

$$A=\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix},$$

expandindo-se o determinante pela primeira linha, virá:

$$\begin{aligned} |A| &= \sum_{j=1}^3 a_{1j} (-1)^{1+j} |M_{1j}| \\ &= a_{11} (-1)^{1+1} |M_{11}| + a_{12} (-1)^{1+2} |M_{12}| + a_{13} (-1)^{1+3} |M_{13}| \end{aligned}$$

$$= a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

Exemplo: Se

$$A = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 2 & 3 \\ 1 & 3 & 0 \end{bmatrix}$$

então expandindo o determinante pela 1ª linha, virá:

$$\begin{aligned} |A| &= \sum_{j=1}^3 a_{1j} (-1)^{1+j} |M_{1j}| \\ &= 1 \begin{vmatrix} 2 & 3 \\ 3 & 0 \end{vmatrix} - 2 \begin{vmatrix} 0 & 3 \\ 1 & 0 \end{vmatrix} + 1 \begin{vmatrix} 0 & 2 \\ 1 & 3 \end{vmatrix} \\ &= 1(2 \times 0 - 3 \times 3) - 2(0 - 3 \times 1) + 1(0 - 2 \times 1) \\ &= 1(-9) - 2(-3) + 1(-2) \\ &= -9 + 6 - 2 \\ &= -5 \end{aligned}$$

(v) Alguns resultados:

- Toda matriz quadrada singular tem determinante nulo. Se A é não singular, $\det(A) \neq 0$.
- Se A é não singular, $\det(A^{-1}) = 1/\det(A)$, onde A^{-1} é a inversa de A.
- Se $D = \text{diag}(d_1, d_2, \dots, d_n)$ então $\det(D) = \prod_{i=1}^n d_i$.
- O determinante de uma matriz triangular é o produto dos elementos da diagonal.
- $\det(A) = \det(A')$.
- Se A e B são matrizes quadradas e de mesma ordem, então $|AB| = |A| |B|$.
- Se $|A_{(n)}| \neq 0$, então posto $(A_{(n)}) = n$. Neste caso, $A_{(n)}$ é de posto completo.

Uma matriz não nula A tem posto r se pelo menos um dos seus menores quadrados de ordem r é diferente de zero e todos os menores quadrados de ordem (r + 1), se existirem, são nulos. O posto de uma matriz nula é zero.

Exemplo: Se

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 5 & 7 \end{bmatrix},$$

então posto $(A) = r = 2$, pois $\begin{vmatrix} 1 & 2 \\ 2 & 3 \end{vmatrix} = -1 \neq 0$ e $|A| = 0$.

1.6. INVERSA DE UMA MATRIZ

Dada uma matriz quadrada A , sua inversa denotada por A^{-1} é uma matriz tal que:

$$AA^{-1} = A^{-1}A = I.$$

Nem toda matriz quadrada tem inversa. As que têm, chamam-se não singulares; as que não têm são singulares.

Seja uma matriz quadrada A não singular. Um algoritmo para obter a inversa A^{-1} consiste em justapor à matriz A uma matriz identidade I . Por meio de operações elementares nas linhas das duas matrizes, obtém-se a inversa A^{-1} , como indicado a seguir:

$$\left[A \mid I \right] \underset{\text{operações elementares}}{\sim} \left[I \mid A^{-1} \right]$$

Operações elementares: Denominam-se operações elementares de uma matriz as seguintes:

- (i) Permutação de duas linhas (ou de duas colunas);
- (ii) Multiplicação de todos os elementos de uma linha (ou coluna) por um número real diferente de zero;
- (iii) Substituição dos elementos de uma linha (coluna) pela soma deles com os elementos correspondentes de outra linha (coluna) previamente multiplicados por um número real diferente de zero.

Exemplo: Obter a inversa A^{-1} sendo

$$A = \begin{bmatrix} 2 & 1 & 3 \\ 4 & 2 & 2 \\ 2 & 5 & 3 \end{bmatrix}$$

Usando o algoritmo dado anteriormente, virá:

$$\left[\begin{array}{ccc|ccc} 2 & 1 & 3 & 1 & 0 & 0 \\ 4 & 2 & 2 & 0 & 1 & 0 \\ 2 & 5 & 3 & 0 & 0 & 1 \end{array} \right] \rightarrow L'_1 = L_1 (1/2)$$

$$\left[\begin{array}{ccc|ccc} 1 & 1/2 & 3/2 & 1/2 & 0 & 0 \\ 4 & 2 & 2 & 0 & 1 & 0 \\ 2 & 5 & 3 & 0 & 0 & 1 \end{array} \right] \rightarrow L'_2 = L_2 + L_1 (-4)$$

$$\left[\begin{array}{ccc|ccc} 1 & 1/2 & 3/2 & 1/2 & 0 & 0 \\ 0 & 0 & -4 & -2 & 1 & 0 \\ 2 & 5 & 3 & 0 & 0 & 1 \end{array} \right] \rightarrow L'_3 = L_3 + L_1 (-2)$$

$$\left[\begin{array}{ccc|ccc} 1 & 1/2 & 3/2 & 1/2 & 0 & 0 \\ 0 & 0 & -4 & -2 & 1 & 0 \\ 0 & 4 & 0 & -1 & 0 & 1 \end{array} \right] \rightarrow L'_{23}$$

$$\left[\begin{array}{ccc|ccc} 1 & 1/2 & 3/2 & 1/2 & 0 & 0 \\ 0 & 4 & 0 & -1 & 0 & 1 \\ 0 & 0 & -4 & -2 & 1 & 0 \end{array} \right] \rightarrow L'_2 = L_2 (1/4)$$

$$\rightarrow L'_3 = L_3 (-1/4)$$

$$\left[\begin{array}{ccc|ccc} 1 & 1/2 & 3/2 & 1/2 & 0 & 0 \\ 0 & 1 & 0 & -1/4 & 0 & 1/4 \\ 0 & 0 & 1 & 1/2 & -1/4 & 0 \end{array} \right] \rightarrow L'_1 = L_1 + L_2 (-1/2)$$

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 3/2 & 5/8 & 0 & -1/8 \\ 0 & 1 & 0 & -1/4 & 0 & 1/4 \\ 0 & 0 & 1 & 1/2 & -1/4 & 0 \end{array} \right] \rightarrow L'_1 = L_1 + L_3 (-3/2)$$

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -1/8 & 3/8 & -1/8 \\ 0 & 1 & 0 & -1/4 & 0 & 1/4 \\ 0 & 0 & 1 & 1/2 & -1/4 & 0 \end{array} \right]$$

Logo,

$$A^{-1} = \begin{bmatrix} -1/8 & 3/8 & -1/8 \\ -1/4 & 0 & 1/4 \\ 1/2 & -1/4 & 0 \end{bmatrix}$$

Outro método para determinar a inversa A^{-1} é dado a seguir: Se A é uma matriz de ordem n , não singular, então a inversa A^{-1} pode ser obtida por:

$$A^{-1} = \frac{1}{\det(A)} (\text{adjunta de } A).$$

Sendo adjunta de A igual à transposta da matriz dos cofatores de A , então

$$A^{-1} = \frac{1}{\det(A)} (\text{cofatores de } A)',$$

em que o cofator de a_{ij} , denotado por c_{ij} , é dado por:

$$c_{ij} = (-1)^{i+j} |M_{ij}|, \text{ onde } |M_{ij}| \text{ é o menor de } a_{ij}.$$

Exemplo: Se

$$A = \begin{bmatrix} 5 & 7 \\ 6 & 10 \end{bmatrix},$$

obter a inversa A^{-1} pelo método apresentado anteriormente.

$$\det(A) = \begin{vmatrix} 5 & 7 \\ 6 & 10 \end{vmatrix} = 5(10) - 7(6) = 50 - 42 = 8$$

$$c_{11} = (-1)^{1+1} |10| = 10$$

$$c_{12} = (-1)^{1+2} |6| = -6$$

$$c_{21} = (-1)^{2+1} |7| = -7$$

$$c_{22} = (-1)^{2+2} |5| = 5$$

Assim, cofatores de $A = \begin{bmatrix} 10 & -6 \\ -7 & 5 \end{bmatrix}$. Logo, a inversa de A é

$$A^{-1} = \frac{1}{8} \begin{bmatrix} 10 & -7 \\ -6 & 5 \end{bmatrix}.$$

Alguns resultados:

- A inversa de uma matriz diagonal não singular $D = \text{diag}(k_1, k_2, \dots, k_n)$ é a matriz diagonal $D^{-1} = \text{diag}(1/k_1, 1/k_2, \dots, 1/k_n)$.
- Se A é não singular, então sua inversa A^{-1} é única e $(A^{-1})^{-1} = A$.
- Se A é não singular, então sua transposta A' também é, e ainda $(A')^{-1} = (A^{-1})'$.
- Se A e B são matrizes não singulares de mesma ordem, então AB é não singular e $(AB)^{-1} = B^{-1}A^{-1}$.

1.7. AUTOVALORES E AUTOVETORES

Apresenta-se uma definição bastante simples e como são obtidos os autovalores e autovetores de uma matriz quadrada A . Neste texto consideram-se apenas autovalores reais.

Definição: Seja a equação

$$A \underline{x} = \lambda \underline{x}$$

em que, A é uma matriz quadrada de ordem n , \underline{x} é um vetor $n \times 1$ não-nulo e λ um escalar tal que $\lambda \in \mathbb{R}$. O número real λ tal que $A \underline{x} = \lambda \underline{x}$ é chamado autovalor de A associado ao autovetor \underline{x} . Os autovalores são também denominados valores próprios ou valores característicos, enquanto os autovetores são também denominados vetores próprios ou vetores característicos.

A determinação dos autovalores e autovetores pode ser feita como a seguir:

$$A \underline{x} = \lambda \underline{x}$$

$$A \underline{x} - \lambda I \underline{x} = \underline{0}$$

$$(A - \lambda I) \underline{x} = \underline{0} \quad (\text{Sistema homogêneo})$$

Para que esse sistema homogêneo admita soluções não nulas, isto é $\underline{x} \neq \underline{0}$, deve-se ter:

$$\det(A - \lambda I) = 0 \quad \text{ou} \quad |A - \lambda I| = 0$$

A equação $|A - \lambda I| = 0$ é denominada equação característica da matriz A , e suas raízes são os autovalores da matriz A . O determinante $|A - \lambda I|$ é um polinômio em λ denominado polinômio característico. A substituição de λ pelos seus valores no sistema homogêneo de equações lineares

permite determinar os autovetores associados. A equação característica de uma matriz real e simétrica tem apenas raízes reais.

Se A é real e simétrica de ordem n , a equação característica terá n raízes, isto é, A terá n autovalores $\lambda_1, \lambda_2, \dots, \lambda_n$. Os λ 's não serão necessariamente distintos ou todos diferentes de zero. Se pelo menos um $\lambda_i = 0$, então a matriz A é singular. Neste caso, o autovetor correspondente ao autovalor λ_i não é o vetor nulo, pois, de $(A - \lambda I)\tilde{x} = \tilde{0}$, tem-se $A\tilde{x} = \tilde{0}$ que tem solução $\tilde{x} \neq \tilde{0}$.

Exemplo: Considere a matriz

$$A = \begin{bmatrix} 3 & -\sqrt{2} \\ -\sqrt{2} & 2 \end{bmatrix}$$

(i) Determinação dos autovalores λ_1 e λ_2

A equação característica é:

$$|A - \lambda I| = 0$$

$$\begin{vmatrix} 3 - \lambda & -\sqrt{2} \\ -\sqrt{2} & 2 - \lambda \end{vmatrix} = 0$$

$$(3 - \lambda)(2 - \lambda) - (-\sqrt{2})(-\sqrt{2}) = 0$$

$$6 - 3\lambda - 2\lambda + \lambda^2 - 2 = 0$$

$$\lambda^2 - 5\lambda + 4 = 0$$

$$\lambda = \frac{5 \pm \sqrt{25 - 16}}{2} = \frac{5 \pm 3}{2}$$

$$\lambda_1 = 4 \quad \text{e} \quad \lambda_2 = 1$$

(ii) Determinação dos autovetores

Um autovetor $\tilde{x}_1 = \begin{bmatrix} x_{11} \\ x_{21} \end{bmatrix}$ associado a $\lambda_1 = 4$, é dado por:

$$(A - \lambda_1 I)\tilde{x}_1 = \tilde{0}$$

$$A\tilde{x}_1 = \lambda_1 \tilde{x}_1$$

$$\begin{bmatrix} 3 & -\sqrt{2} \\ -\sqrt{2} & 2 \end{bmatrix} \begin{bmatrix} x_{11} \\ x_{21} \end{bmatrix} = 4 \begin{bmatrix} x_{11} \\ x_{21} \end{bmatrix}$$

$$\begin{cases} 3x_{11} - \sqrt{2} x_{21} = 4x_{11} \\ -\sqrt{2}x_{11} + 2x_{21} = 4x_{21} \end{cases}$$

$$\begin{cases} -x_{11} - \sqrt{2} x_{21} = 0 \\ -\sqrt{2} x_{11} - 2x_{21} = 0 \end{cases}$$

$$\begin{cases} x_{11} + \sqrt{2} x_{21} = 0 \\ \sqrt{2} x_{11} + 2x_{21} = 0 \end{cases}$$

A matriz do sistema homogêneo de equações lineares é singular sendo o sistema consistente e indeterminado, isto é, admite infinitas soluções.

Para obter uma solução não nula, podemos abandonar uma das equações e dar a uma das incógnitas um valor arbitrário não nulo. Por exemplo, abandonar a segunda equação e fazer $x_{21} = -1$. Assim virá:

$$x_{11} - \sqrt{2} = 0 \quad \Rightarrow \quad x_{11} = \sqrt{2}$$

Logo, $\tilde{x}_1 = \begin{bmatrix} \sqrt{2} \\ -1 \end{bmatrix}$ é um autovetor associado a $\lambda_1 = 4$.

Cabe ressaltar que sempre haverá duas soluções, uma obtida da outra multiplicada por -1.

Um autovetor $\tilde{x}_2 = \begin{bmatrix} x_{12} \\ x_{22} \end{bmatrix}$ associado a $\lambda_2 = 1$, é dado por:

$$(A - \lambda_2 I) \tilde{x}_2 = 0$$

$$A \tilde{x}_2 = \lambda_2 \tilde{x}_2$$

$$\begin{bmatrix} 3 & -\sqrt{2} \\ -\sqrt{2} & 2 \end{bmatrix} \begin{bmatrix} x_{12} \\ x_{22} \end{bmatrix} = 1 \begin{bmatrix} x_{12} \\ x_{22} \end{bmatrix}$$

$$\begin{cases} 3x_{12} - \sqrt{2} x_{22} = x_{12} \\ -\sqrt{2} x_{12} + 2x_{22} = x_{22} \end{cases}$$

$$\begin{cases} 2x_{12} - \sqrt{2} x_{22} = 0 \\ -\sqrt{2} x_{12} + x_{22} = 0 \end{cases}$$

Fazendo $x_{22} = \sqrt{2}$ na primeira equação, virá:

$$2x_{12} - \sqrt{2}(\sqrt{2}) = 0$$

$$2x_{12} - 2 = 0$$

$$x_{12}=1$$

Logo, $\tilde{x}_2 = \begin{bmatrix} 1 \\ \sqrt{2} \end{bmatrix}$ é um autovetor associado a $\lambda_2 = 1$.

A forma normalizada dos vetores \tilde{x}_1 e \tilde{x}_2 são dadas por \tilde{v}_1 e \tilde{v}_2 , respectivamente, como a seguir:

$$\tilde{v}_1 = \frac{1}{\|\tilde{x}_1\|} \tilde{x}_1 = \frac{1}{\sqrt{x_{11}^2 + x_{21}^2}} \begin{bmatrix} x_{11} \\ x_{21} \end{bmatrix}$$

$$\tilde{v}_1 = \frac{1}{\sqrt{(\sqrt{2})^2 + (-1)^2}} \begin{bmatrix} \sqrt{2} \\ -1 \end{bmatrix} = \begin{bmatrix} \sqrt{2}/\sqrt{3} \\ -1/\sqrt{3} \end{bmatrix}$$

$$\tilde{v}_2 = \frac{1}{\|\tilde{x}_2\|} \tilde{x}_2 = \frac{1}{\sqrt{x_{12}^2 + x_{22}^2}} \begin{bmatrix} x_{12} \\ x_{22} \end{bmatrix}$$

$$\tilde{v}_2 = \frac{1}{\sqrt{1^2 + (\sqrt{2})^2}} \begin{bmatrix} 1 \\ \sqrt{2} \end{bmatrix} = \begin{bmatrix} 1/\sqrt{3} \\ \sqrt{2}/\sqrt{3} \end{bmatrix}$$

Note que $\tilde{v}_1' \tilde{v}_1 = 1$ e $\tilde{v}_2' \tilde{v}_2 = 1$, isto é, \tilde{v}_1 e \tilde{v}_2 são autovetores normalizados.

Para matrizes de ordem superior a dois, o cálculo dos autovalores é, em geral, bastante trabalhoso. Mas esse cálculo pode ser facilitado com o uso de computadores, que já dispõem para isso, de programas apropriados.

Alguns resultados:

- Os autovalores de $A_{(n)}$ e $A_{(n)}'$ são os mesmos.
- Se $\lambda_1, \lambda_2, \dots, \lambda_n$ são os autovalores de uma matriz quadrada A de ordem n e se k é um escalar não nulo, então $k\lambda_1, k\lambda_2, \dots, k\lambda_n$ são os autovalores de kA .
- Se λ é um autovalor da matriz não singular A , então $1/\lambda$ é um autovalor de A^{-1} e \tilde{x} é um autovetor tanto de A quanto de A^{-1} .
- Se A é uma matriz quadrada de ordem n com autovalores $\lambda_1, \lambda_2, \dots, \lambda_n$, então

$$\det(A) = |A| = \prod_{i=1}^n \lambda_i.$$

1.8. MATRIZ ORTOGONAL

Uma matriz quadrada A é ortogonal se

$$AA' = A'A = I,$$

isto é, se

$$A' = A^{-1}.$$

Assim, para

$$A = \begin{bmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ v_{21} & v_{22} & \cdots & v_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ v_{n1} & v_{n2} & \cdots & v_{nn} \end{bmatrix} = \begin{bmatrix} \underline{v}_1 & \underline{v}_2 & \cdots & \underline{v}_n \end{bmatrix}, \text{ ortogonal, tem-se que:}$$

$$\underline{v}_i' \underline{v}_i = 1 \quad (\underline{v}_i' \text{ s são vetores normalizados})$$

e

$$\underline{v}_i' \underline{v}_j = 0 \quad \text{para todo } i \neq j \text{ (dois vetores quaisquer, distintos, são ortogonais, isto é, são}$$

perpendiculares. O ângulo formado entre eles é de 90°).

É evidente que os vetores coluna (vetores linha) de uma matriz ortogonal A são vetores normalizados ortogonais dois a dois, e são ditos vetores ortonormais.

Exemplo:

$$A = \begin{bmatrix} 1/\sqrt{3} & 0 & -2/\sqrt{6} \\ 1/\sqrt{3} & -1/\sqrt{2} & 1/\sqrt{6} \\ 1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6} \end{bmatrix} \text{ é uma matriz ortogonal.}$$

Pode-se verificar que:

$$AA' = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$A'A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$A^{-1} = A' = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} \\ -2/\sqrt{6} & 1/\sqrt{6} & 1/\sqrt{6} \end{bmatrix}$$

Outro exemplo de matriz ortogonal é a matriz P dada a seguir:

$$P = \begin{bmatrix} \tilde{v}_1 & \tilde{v}_2 \end{bmatrix} = \begin{bmatrix} \sqrt{2}/\sqrt{3} & 1/\sqrt{3} \\ -1/\sqrt{3} & \sqrt{2}/\sqrt{3} \end{bmatrix}$$

Note que a matriz P foi obtida no exemplo de cálculo de autovalores e autovetores da matriz

$$A = \begin{bmatrix} 3 & -\sqrt{2} \\ -\sqrt{2} & 2 \end{bmatrix}.$$

Se A é uma matriz simétrica de ordem n, com autovalores $\lambda_1, \lambda_2, \dots, \lambda_n$ e autovetores normalizados $\tilde{v}_1, \tilde{v}_2, \dots, \tilde{v}_n$, respectivamente, então A pode ser expressa como:

$$A = P \Lambda P' \\ = \sum_{i=1}^n \lambda_i \tilde{v}_i \tilde{v}_i'$$

em que

$\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ e $P = \begin{bmatrix} \tilde{v}_1 & \tilde{v}_2 & \dots & \tilde{v}_n \end{bmatrix}$ é a matriz ortogonal. Este resultado é chamado

decomposição espectral da matriz A. Note que, neste caso, $P'AP = \Lambda$ e assim a matriz A é diagonalizável.

Exemplo: A decomposição espectral da matriz

$$A = \begin{bmatrix} 3 & -\sqrt{2} \\ -\sqrt{2} & 2 \end{bmatrix}$$

é dada por:

$$A = P \Lambda P' = \begin{bmatrix} \sqrt{2}/\sqrt{3} & 1/\sqrt{3} \\ -1/\sqrt{3} & \sqrt{2}/\sqrt{3} \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{2}/\sqrt{3} & -1/\sqrt{3} \\ 1/\sqrt{3} & \sqrt{2}/\sqrt{3} \end{bmatrix}$$

$$A = \begin{bmatrix} \sqrt{2}/\sqrt{3} & 1/\sqrt{3} \\ -1/\sqrt{3} & \sqrt{2}/\sqrt{3} \end{bmatrix} \begin{bmatrix} 4\sqrt{2}/\sqrt{3} & -4/\sqrt{3} \\ 1/\sqrt{3} & \sqrt{2}/\sqrt{3} \end{bmatrix}$$

$$A = \begin{bmatrix} 3 & -\sqrt{2} \\ -\sqrt{2} & 2 \end{bmatrix}$$

ou ainda,

$$A = \lambda_1 \tilde{v}_1 \tilde{v}_1' + \lambda_2 \tilde{v}_2 \tilde{v}_2'$$

$$A = 4 \begin{bmatrix} \sqrt{2}/\sqrt{3} \\ -1/\sqrt{3} \end{bmatrix} \begin{bmatrix} \sqrt{2}/\sqrt{3} & -1/\sqrt{3} \end{bmatrix} + 1 \begin{bmatrix} 1/\sqrt{3} \\ \sqrt{2}/\sqrt{3} \end{bmatrix} \begin{bmatrix} 1/\sqrt{3} & \sqrt{2}/\sqrt{3} \end{bmatrix}$$

$$A = 4 \begin{bmatrix} 2/3 & -\sqrt{2}/3 \\ -\sqrt{2}/3 & 1/3 \end{bmatrix} + 1 \begin{bmatrix} 1/3 & \sqrt{2}/3 \\ \sqrt{2}/3 & 2/3 \end{bmatrix}$$

$$A = \begin{bmatrix} 8/3 & -4\sqrt{2}/3 \\ -4\sqrt{2}/3 & 4/3 \end{bmatrix} + \begin{bmatrix} 1/3 & \sqrt{2}/3 \\ \sqrt{2}/3 & 2/3 \end{bmatrix}$$

$$A = \begin{bmatrix} 3 & -\sqrt{2} \\ -\sqrt{2} & 2 \end{bmatrix}$$

Alguns resultados:

- O determinante de uma matriz ortogonal é ± 1 .
- Se A é uma matriz simétrica com autovalores distintos, então os autovetores são ortogonais.
- Se A é uma matriz simétrica de ordem n , com autovalores $\lambda_1, \lambda_2, \dots, \lambda_n$, todos positivos, e considerando que a decomposição espectral da matriz A é dada por $A = P \Lambda P'$ como visto

anteriormente, tem-se que: $A^{-1} = P \Lambda^{-1} P' = \sum_{i=1}^n \frac{1}{\lambda_i} \tilde{v}_i \tilde{v}_i'$.

1.9. TRAÇO DE UMA MATRIZ

O traço de uma matriz quadrada $A = (a_{ij})$, com $i, j = 1, \dots, n$, denotado por $\text{tr}(A)$, é dado pela soma dos elementos da diagonal de A , isto é,

$$\text{tr}(A) = a_{11} + a_{22} + \dots + a_{nn} = \sum_{i=1}^n a_{ii}.$$

Exemplo: Se

$$A = \begin{bmatrix} 5 & 8 & 10 \\ 4 & 7 & 20 \\ 2 & 1 & -4 \end{bmatrix}$$

então,

$$\text{tr}(A) = 5 + 7 - 4 = 8$$

Alguns resultados:

- (i) Se A não é uma matriz quadrada, então o traço não é definido, isto é, ele não existe.
- (ii) Se k é um escalar, então $\text{tr}(k) = k$.
- (iii) Se A é uma matriz quadrada, então $\text{tr}(A') = \text{tr}(A)$.
- (iv) O traço da soma ou diferença de matrizes é igual à soma ou diferença dos traços, isto é, $\text{tr}(A \pm B) = \text{tr}(A) \pm \text{tr}(B)$, em que A e B são quadradas, de mesma ordem.
- (v) Se A é $n \times p$ e B é $p \times n$, então $\text{tr}(AB) = \text{tr}(BA)$.
- (vi) Se P é uma matriz simétrica e idempotente, então $\text{tr}(P) = \text{posto}(P)$.
- (vii) Se A é de ordem n e P é qualquer matriz não singular também de ordem n , então $\text{tr}(P^{-1}AP) = \text{tr}(A)$.

1.10. FORMAS QUADRÁTICAS

1.10.1. Introdução

As formas quadráticas têm inúmeras aplicações na estatística, como por exemplo, no estudo de modelos lineares, na determinação de máximos e mínimos de equações de superfícies de resposta, dentre outras. A seguir são apresentadas algumas definições e outras considerações em geral.

Definição: Sejam x_1, x_2, \dots, x_n ; a_{ij} , $i, j = 1, 2, \dots, n$ números reais. A função homogênea $f : \mathbb{R}^n \rightarrow \mathbb{R}$ tal que

$$f(x) = \tilde{x}' A \tilde{x} = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j \quad (1)$$

é chamada forma quadrática, com $\tilde{x}' = [x_1 \ x_2 \ \dots \ x_n]$ e $A_{(n)} = (a_{ij})$ denominada matriz da forma quadrática. A equação (1) pode ainda ser escrita assim:

$$f(x) = \tilde{x}' A \tilde{x} = \sum_{i=1}^n a_{ii} x_i^2 + \sum_{i \neq j} \sum a_{ij} x_i x_j \quad (2)$$

Para uma forma quadrática $f(x) = \tilde{x}' A \tilde{x}$, existem infinitas matrizes $B \neq A$ tais que $f(x) = \tilde{x}' B \tilde{x}$, porém, existe uma única matriz simétrica C tal que, $f(x) = \tilde{x}' C \tilde{x}$. Por exemplo, no caso em que $n = 2$, seja $f(x) = ax_1^2 + bx_1x_2 + cx_2^2$. Na forma matricial, vem:

$$f(x) = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = a_{11}x_1^2 + (a_{12} + a_{21})x_1x_2 + a_{22}x_2^2.$$

Comparando os termos, temos:

$$a_{11}=a, \quad a_{12}+a_{21}=b \text{ e } a_{22}=c$$

Assim, existem infinitos valores a_{12} e a_{21} tais que $a_{12}+a_{21}=b$, porém, exigindo-se que A seja uma matriz simétrica, teremos $a_{12}=a_{21}$, conseqüentemente $a_{12}=a_{21}=\frac{b}{2}=\frac{a_{12}+a_{21}}{2}$, e a matriz da forma quadrática fica unicamente determinada. Desta forma, a matriz da forma quadrática será sempre considerada uma matriz simétrica, uma vez que existe uma correspondência biunívoca entre o conjunto das formas quadráticas e o conjunto das matrizes simétricas. Sendo $a_{ij}=a_{ji}$, a equação (2) pode ser escrita como:

$$f(\mathbf{x})=\mathbf{x}'\mathbf{A}\mathbf{x}=\sum_{i=1}^n a_{ii}x_i^2 + 2\sum_{i<j} a_{ij}x_i x_j$$

Exemplos:

(i) Seja $f: \mathbb{R}^3 \rightarrow \mathbb{R}$ definida por $f(\mathbf{x})=\mathbf{x}'\mathbf{A}\mathbf{x}$, onde $\mathbf{A}=\begin{bmatrix} 1 & 3 & 2 \\ 1 & -1 & 6 \\ 3 & 5 & 4 \end{bmatrix}$.

Obtenção da matriz simétrica para esta forma quadrática definida a partir de uma matriz não simétrica:

Seja

$$c_{11}=a_{11}=1$$

$$c_{12}=c_{21}=\frac{a_{12}+a_{21}}{2}=\frac{3+1}{2}=2$$

$$c_{13}=c_{31}=\frac{a_{13}+a_{31}}{2}=\frac{2+3}{2}=5/2$$

$$c_{22}=a_{22}=-1$$

$$c_{23}=c_{32}=\frac{a_{23}+a_{32}}{2}=\frac{6+5}{2}=11/2$$

$$c_{33}=a_{33}=4$$

Assim, a matriz C , simétrica, que substitui a matriz A é dada por $C=\begin{bmatrix} 1 & 2 & 5/2 \\ 2 & -1 & 11/2 \\ 5/2 & 11/2 & 4 \end{bmatrix}$.

(ii) Encontre a matriz A associada com a forma quadrática

$$f(\underline{x}) = \underline{x}' A \underline{x} = 2x_1^2 - 3x_1x_2 + \frac{7}{2}x_1x_3 + x_1x_4 + x_2^2 + 6x_2x_3 - 8x_3x_4 + 2x_4^2.$$

A matriz A simétrica é dada por:

$$A = \begin{bmatrix} 2 & -3/2 & 7/4 & 1/2 \\ -3/2 & 1 & 3 & 0 \\ 7/4 & 3 & 0 & -4 \\ 1/2 & 0 & -4 & 2 \end{bmatrix}.$$

(iii) Seja y_1, y_2, \dots, y_n , n observações de uma variável aleatória qualquer. Sabe-se que a soma de quadrado total corrigida pela média é dada por

$$SQTotal_{(corrigida \text{ pela média})} = \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n y_i^2 - \frac{(\sum_{i=1}^n y_i)^2}{n} \quad (3)$$

Expressar (3) em termos de forma quadrática.

Sendo $\underline{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_n$ e $\underline{u} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}_n$, pode-se escrever:

$$\begin{aligned} SQTotal_{(corrigida \text{ pela média})} &= \underline{y}' \underline{y} - \frac{1}{n} \underline{y}' \underline{u} \underline{u}' \underline{y} \\ &= \underline{y}' I_{(n)} \underline{y} - \underline{y}' \frac{\underline{u} \underline{u}'}{n} \underline{y} \\ &= \underline{y}' \left[\underbrace{I_{(n)} - \frac{\underline{u} \underline{u}'}{n}}_{=Q} \right] \underline{y} = \underline{y}' Q \underline{y} \end{aligned}$$

Q é a matriz núcleo da forma quadrática em y, é simétrica e idempotente $\Rightarrow \text{posto}(Q) = \text{tr}(Q) = n - 1$ = número de graus de liberdade.

1.10.2. Classificação das formas quadráticas

Dependendo do sinal de $f(\underline{x}) = \underline{x}' A \underline{x}$, $f(\underline{x})$ pode ter a seguinte classificação:

- (i) $f(\underline{x})$ é **positiva definida** se, e somente se, $f(\underline{x}) > 0$ para todo $\underline{x} \neq \underline{0}$ ($\underline{0}$ = vetor nulo);
- (ii) $f(\underline{x})$ é **positiva semidefinida** se, e somente se, $f(\underline{x}) \geq 0$ para todo \underline{x} e $f(\underline{x}) = 0$ para algum $\underline{x} \neq \underline{0}$;

- (iii) $f(x)$ é **negativa definida** se, e somente se, $f(x) < 0$ para todo $\underline{x} \neq \underline{0}$;
- (iv) $f(x)$ é **negativa semidefinida** se, e somente se, $f(x) \leq 0$ para todo \underline{x} e $f(x) = 0$ para algum $\underline{x} \neq \underline{0}$;
- (v) $f(x)$ é **não definida** se não se enquadra num dos casos anteriores, ou ainda, ela é positiva para alguns valores de \underline{x} e negativa para outros.

As mesmas definições se estendem para as matrizes das formas quadráticas devido a correspondência biunívoca anteriormente citada. Se uma forma quadrática $f(x) = \underline{x}' A \underline{x}$ satisfaz (i) ou (ii) costuma-se dizer que $f(x)$ é não negativa e se satisfaz (iii) ou (iv) é não positiva.

A classificação das formas quadráticas através da definição pode ser muito difícil ou pouco operacional. Para facilitar, uma alternativa é proceder à diagonalização da matriz núcleo, através de operações de congruência como apresentado a seguir.

Diagonalização de matrizes simétricas (por congruência): Duas matrizes D e A são chamadas congruentes se, $D = B'AB$ para alguma B não singular. A obtenção da matriz diagonal $D = \text{diag}(d_1, d_2, \dots, d_n)$ equivalente a A será feita, obviamente, por operações elementares nas linhas e nas colunas de A .

Método: As operações elementares feitas nas linhas de A são também efetuadas na matriz I justaposta à direita de A . Porém, as mesmas operações efetuadas passo a passo nas colunas de A não o são em I .

$$\begin{array}{c}
 \text{Op. congruência em } A \\
 [A \mid I] \quad \sim \quad \dots \quad \sim \quad \dots \quad \sim \quad [D \mid B'] \\
 \text{Op. elementares em } I
 \end{array}$$

Temos ainda que $B'AB = D$ e $\det(A) = \det(D)$. A classificação de $D = \text{diag}(d_1, d_2, \dots, d_n)$ é a mesma de A e também a mesma da forma quadrática $f(x) = \underline{x}' A \underline{x}$ como a seguir:

- (i) $f(x)$ é **positiva definida** $\Leftrightarrow d_i > 0$, para $i = 1, 2, \dots, n$;
- (ii) $f(x)$ é **positiva semidefinida** $\Leftrightarrow d_i \geq 0$, para $i = 1, 2, \dots, n$, com pelo menos um $d_i = 0$;
- (iii) $f(x)$ é **negativa definida** $\Leftrightarrow d_i < 0$, para $i = 1, 2, \dots, n$;
- (iv) $f(x)$ é **negativa semidefinida** $\Leftrightarrow d_i \leq 0$, para $i = 1, 2, \dots, n$, com pelo menos um $d_i = 0$;

(v) $f(x)$ é **não definida** se não se enquadra num dos casos anteriores.

Sabe-se da álgebra de matizes que se A é uma matriz real e simétrica de ordem n , existe uma matriz P ortogonal tal que $P'AP = \Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$, é a matriz diagonal que exhibe os autovalores de A . Assim, a classificação da forma quadrática $f(x) = \tilde{x}'A\tilde{x}$, pode ser feita através dos autovalores de A , bastando substituir d_i por λ_i na classificação dada anteriormente.

Exemplo: Seja a forma quadrática

$$f(x) = \tilde{x}'A\tilde{x} = 2x_1^2 - 2x_2^2 + 4x_3^2 + 8x_1x_2 + 6x_1x_3 + 2x_2x_3, \text{ ou na notação matricial,}$$

$$f(x) = \tilde{x}'A\tilde{x} = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} 2 & 4 & 3 \\ 4 & -2 & 1 \\ 3 & 1 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

A classificação desta forma quadrática diagonalizando a matriz A por meio de operações de congruência é apresentada a seguir:

$$\left[\begin{array}{ccc|ccc} 2 & 4 & 3 & 1 & 0 & 0 \\ 4 & -2 & 1 & 0 & 1 & 0 \\ 3 & 1 & 4 & 0 & 0 & 1 \end{array} \right] \begin{array}{l} L'_1 = L_1 \\ L'_2 = L_2 + L_1(-2) \\ L'_3 = L_3 \end{array} \quad \left[\begin{array}{ccc|ccc} 2 & 4 & 3 & 1 & 0 & 0 \\ 0 & -10 & -5 & -2 & 1 & 0 \\ 3 & 1 & 4 & 0 & 0 & 1 \end{array} \right] \begin{array}{l} C'_1 = C_1 \\ C'_2 = C_2 + C_1(-2) \\ C'_3 = C_3 \end{array}$$

$$\left[\begin{array}{ccc|ccc} 2 & 0 & 3 & 1 & 0 & 0 \\ 0 & -10 & -5 & -2 & 1 & 0 \\ 3 & -5 & 4 & 0 & 0 & 1 \end{array} \right] \begin{array}{l} L'_1 = L_1 \\ L'_2 = L_2 \\ L'_3 = L_3 + L_1(-3/2) \end{array}$$

$$\left[\begin{array}{ccc|ccc} 2 & 0 & 3 & 1 & 0 & 0 \\ 0 & -10 & -5 & -2 & 1 & 0 \\ 0 & -5 & -1/2 & -3/2 & 0 & 1 \end{array} \right] \begin{array}{l} C'_1 = C_1 \\ C'_2 = C_2 \\ C'_3 = C_3 + C_1(-3/2) \end{array}$$

$$\left[\begin{array}{ccc|ccc} 2 & 0 & 0 & 1 & 0 & 0 \\ 0 & -10 & -5 & -2 & 1 & 0 \\ 0 & -5 & -1/2 & -3/2 & 0 & 1 \end{array} \right] \begin{array}{l} L'_1 = L_1 \\ L'_2 = L_2 \\ L'_3 = L_3 + L_2(-1/2) \end{array}$$

$$\left[\begin{array}{ccc|ccc} 2 & 0 & 0 & 1 & 0 & 0 \\ 0 & -10 & -5 & -2 & 1 & 0 \\ 0 & 0 & 2 & -1/2 & -1/2 & 1 \end{array} \right] \begin{array}{l} C'_1 = C_1 \\ C'_2 = C_2 \\ C'_3 = C_3 + C_2(-1/2) \end{array}$$

$$\underbrace{\left[\begin{array}{ccc|ccc} 2 & 0 & 0 & 1 & 0 & 0 \\ 0 & -10 & 0 & -2 & 1 & 0 \\ 0 & 0 & 2 & -1/2 & -1/2 & 1 \end{array} \right]}_D \underbrace{\left[\begin{array}{ccc|ccc} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -1/2 & -1/2 & 1 \end{array} \right]}_{B'}$$

O apóstrofo nas letras L e C está significando apenas novas linhas e colunas, respectivamente. Note que, $D=B'AB$ e $\det(A)=\det(D)=-40$. Como d_1 muda de sinal, a forma quadrática é não definida, ou ainda, a matriz A é não definida.

Os autovalores da matriz A da forma quadrática do exemplo anterior são: $\lambda_1=7,343501$, $\lambda_2=1,199093$ e $\lambda_3=-4,542594$, o que nos leva obviamente, à mesma classificação dada anteriormente.

1.11. VETORES DE OPERADORES DIFERENCIAIS

Derivadas de funções lineares e formas quadráticas têm grande utilidade em certas demonstrações e aplicações na área de estatística. Por esta razão, são apresentados a seguir, de uma maneira bastante simples, alguns resultados importantes.

i) Escalares

Seja $\lambda=f(x)$ uma função das variáveis x_1, x_2, \dots, x_p em que $\tilde{x} = \begin{bmatrix} x_1 & x_2 & \dots & x_p \end{bmatrix}$ e

sejam $\partial\lambda/\partial x_1, \partial\lambda/\partial x_2, \dots, \partial\lambda/\partial x_p$ as derivadas parciais. Nós definimos $\partial\lambda/\partial \tilde{x}$ como:

$$\frac{\partial\lambda}{\partial \tilde{x}} = \begin{bmatrix} \frac{\partial\lambda}{\partial x_1} \\ \frac{\partial\lambda}{\partial x_2} \\ \vdots \\ \frac{\partial\lambda}{\partial x_p} \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial x_1} \\ \frac{\partial}{\partial x_2} \\ \vdots \\ \frac{\partial}{\partial x_p} \end{bmatrix} \lambda.$$

O vetor $\frac{\partial}{\partial \tilde{x}}$ é um vetor de operadores diferenciais.

Exemplo: Se $\lambda=4x_1-7x_2+10x_3$

em que λ, x_1, x_2 e x_3 são escalares, então as derivadas de λ em relação a x_1, x_2 e x_3 são

$$\frac{\partial\lambda}{\partial \tilde{x}} = \begin{bmatrix} 4 \\ -7 \\ 10 \end{bmatrix}.$$

Note que

$$\lambda = 4x_1 - 7x_2 + 10x_3 = \begin{bmatrix} 4 & -7 & 10 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \tilde{a}' \tilde{x},$$

e assim, sendo $\lambda = \tilde{a}' \tilde{x} = \tilde{x}' \tilde{a}$, onde $\tilde{a}' = [a_1 \quad a_2 \quad a_3]$ é um vetor de constantes, então

$$\frac{\partial}{\partial \tilde{x}} (\tilde{a}' \tilde{x}) = \frac{\partial}{\partial \tilde{x}} (\tilde{x}' \tilde{a}) = \tilde{a}.$$

ii) Vetores

Seja $\tilde{x}' A = \begin{bmatrix} \tilde{x}' a_1 & \tilde{x}' a_2 & \cdots & \tilde{x}' a_p \end{bmatrix}$, em que a_i é a i-ésima coluna da matriz A. Assim,

$$\frac{\partial}{\partial \tilde{x}} (\tilde{x}' A) = \begin{bmatrix} \frac{\partial \tilde{x}' a_1}{\partial \tilde{x}} & \frac{\partial \tilde{x}' a_2}{\partial \tilde{x}} & \cdots & \frac{\partial \tilde{x}' a_p}{\partial \tilde{x}} \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & \cdots & a_p \end{bmatrix} = A.$$

Temos ainda que

$$\frac{\partial}{\partial \tilde{x}} (A \tilde{x}) \equiv \frac{\partial}{\partial \tilde{x}} (A \tilde{x})' = \frac{\partial}{\partial \tilde{x}} (\tilde{x}' A') = A'.$$

Exemplo: Seja

$$A \tilde{x} = \begin{bmatrix} 5 & 2 & 4 \\ -2 & 7 & 9 \\ 3 & -8 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 5x_1 + 2x_2 + 4x_3 \\ -2x_1 + 7x_2 + 9x_3 \\ 3x_1 - 8x_2 + x_3 \end{bmatrix}. \text{ Então,}$$

$$\begin{aligned} \frac{\partial}{\partial \tilde{x}} (A \tilde{x}) &= \begin{bmatrix} \frac{\partial}{\partial \tilde{x}} (5x_1 + 2x_2 + 4x_3) & \frac{\partial}{\partial \tilde{x}} (-2x_1 + 7x_2 + 9x_3) & \frac{\partial}{\partial \tilde{x}} (3x_1 - 8x_2 + x_3) \end{bmatrix} \\ &= \begin{bmatrix} 5 & -2 & 3 \\ 2 & 7 & -8 \\ 4 & 9 & 1 \end{bmatrix} = A'. \end{aligned}$$

iii) Formas quadráticas

Seja $f(\tilde{x}) = \tilde{x}' A \tilde{x}$ uma forma quadrática.

Então,

$$\begin{aligned} \frac{\partial}{\partial \tilde{x}} (\tilde{x}' A \tilde{x}) &= \frac{\partial}{\partial \tilde{x}} (\tilde{x}' P) + \frac{\partial}{\partial \tilde{x}} (Q \tilde{x}) \text{ para } P = A \tilde{x} \text{ e } Q = \tilde{x}' A \\ &= P + Q' \end{aligned}$$

$$= A \tilde{x} + A' \tilde{x}$$

Para A simétrica, temos que:

$$\frac{\partial}{\partial \tilde{x}} (\tilde{x}' A \tilde{x}) = 2A \tilde{x} \quad (\text{Vetor coluna})$$

$$e \quad \frac{\partial}{\partial \tilde{x}'} (\tilde{x}' A \tilde{x}) = 2 \tilde{x}' A \quad (\text{Vetor linha})$$

Exemplo: Para

$$f(x) = \tilde{x}' A \tilde{x} = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} 2 & 4 & 3 \\ 4 & -2 & 1 \\ 3 & 1 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$= 2x_1^2 - 2x_2^2 + 4x_3^2 + 8x_1x_2 + 6x_1x_3 + 2x_2x_3,$$

$$\frac{\partial f(x)}{\partial \tilde{x}} = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \frac{\partial f(x)}{\partial x_2} \\ \frac{\partial f(x)}{\partial x_3} \end{bmatrix} = \begin{bmatrix} 4x_1 + 8x_2 + 6x_3 \\ 8x_1 - 4x_2 + 2x_3 \\ 6x_1 + 2x_2 + 8x_3 \end{bmatrix}$$

$$= 2 \begin{bmatrix} 2 & 4 & 3 \\ 4 & -2 & 1 \\ 3 & 1 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 2A \tilde{x}.$$

Temos ainda que

$$\begin{aligned} \frac{\partial^2 f(x)}{\partial \tilde{x} \partial \tilde{x}'} &= \begin{bmatrix} \frac{\partial}{\partial \tilde{x}} \end{bmatrix} \begin{bmatrix} \frac{\partial f(x)}{\partial \tilde{x}'} \end{bmatrix} \\ &= \begin{bmatrix} \frac{\partial}{\partial x_1} \\ \frac{\partial}{\partial x_2} \\ \vdots \\ \frac{\partial}{\partial x_p} \end{bmatrix} \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} & \frac{\partial f(x)}{\partial x_2} & \dots & \frac{\partial f(x)}{\partial x_p} \end{bmatrix} \end{aligned}$$

$$= \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_p} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x)}{\partial x_p \partial x_1} & \frac{\partial^2 f(x)}{\partial x_p \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_p \partial x_p} \end{bmatrix} = 2A$$

Esta é a chamada Matriz Hessiana. Uma aplicação é no estudo do ponto crítico em superfícies de resposta.

1.12. SISTEMAS DE EQUAÇÕES E INVERSAS GENERALIZADAS

1.12.1. Sistema de equações lineares com matrizes não singulares

Seja o sistema de equações lineares com n equações e n incógnitas apresentado a seguir:

$$\begin{aligned} a_{11} x_1 + a_{12} x_2 + \cdots + a_{1n} x_n &= g_1 \\ a_{21} x_1 + a_{22} x_2 + \cdots + a_{2n} x_n &= g_2 \\ \vdots & \quad \quad \quad \vdots \quad \quad \quad \vdots \\ a_{n1} x_1 + a_{n2} x_2 + \cdots + a_{nn} x_n &= g_n \end{aligned}$$

Este sistema pode ser escrito na forma matricial como:

$$A \tilde{x} = \tilde{g}$$

em que,

$A_{(n)}$, é uma matriz quadrada de ordem n , de elementos conhecidos e não singular, ${}_n \tilde{x}_1$ é um vetor coluna de n componentes desconhecidos e ${}_n \tilde{g}_1$ é um vetor coluna de n componentes conhecidos.

Logo, $\exists A^{-1}$, então $A \tilde{x} = \tilde{g}$ é um sistema consistente e determinado, isto é, possui uma única solução que pode ser obtida por:

$$A^{-1} A \tilde{x} = A^{-1} \tilde{g} \quad \Rightarrow \quad \tilde{x} = A^{-1} \tilde{g}$$

Exemplo: Seja $A \tilde{x} = \tilde{g}$ caracterizado por:

$$\begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 30 \\ 30 \end{bmatrix}$$

$$\det(A) = |A| = 3 \neq 0 \Rightarrow \exists A^{-1}$$

$$\text{Então, } \underline{\tilde{x}} = A^{-1} \underline{\tilde{g}} = \frac{1}{3} \begin{bmatrix} 1 & 1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 30 \\ 30 \end{bmatrix} = \begin{bmatrix} 20 \\ 10 \end{bmatrix}$$

Note que, se $\underline{\tilde{x}}$ é solução, então,

$$\begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 20 \\ 10 \end{bmatrix} = \begin{bmatrix} 30 \\ 30 \end{bmatrix} \text{ ou } 20 \begin{bmatrix} 2 \\ 1 \end{bmatrix} + 10 \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 30 \\ 30 \end{bmatrix},$$

isto é, $x_1 c_1 + x_2 c_2 = \underline{\tilde{g}}$, e assim, o vetor solução indica a combinação linear das colunas de A que reproduz $\underline{\tilde{g}}$. Em outras palavras, se $A \underline{\tilde{x}} = \underline{\tilde{g}}$ é consistente, então $\underline{\tilde{g}}$ pertence ao espaço vetorial gerado pelas colunas de A , fato que pode ser denotado por $\underline{\tilde{g}} \in C(A)$.

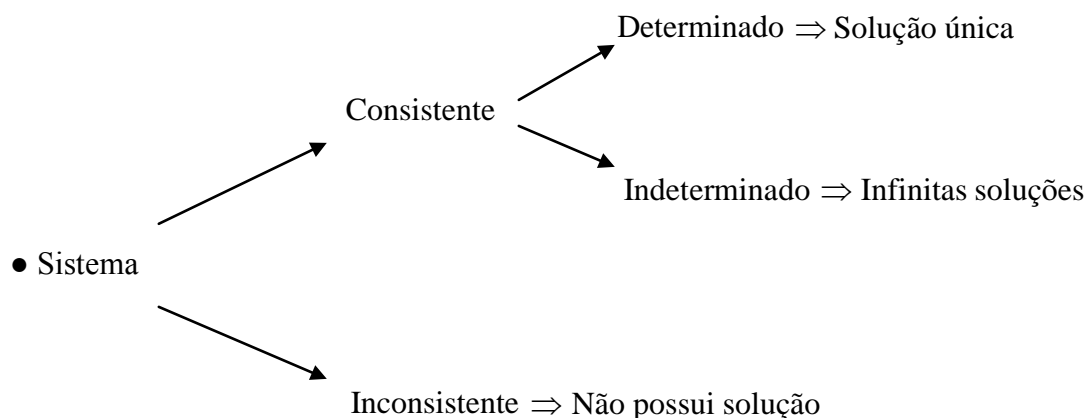
Os sistemas de equações lineares com matrizes não singulares são naturalmente desejáveis. No entanto, existem situações em que isto não ocorre. Outras vezes a matriz A nem é quadrada e não há sentido em falarmos em A^{-1} .

1.12.2. Sistema de equações lineares com matrizes quaisquer

Seja agora um sistema mais geral,

$$A \underline{\tilde{x}} = \underline{\tilde{g}}$$

onde A é $m \times n$ singular ou não singular, $\underline{\tilde{x}}$ é $n \times 1$ e $\underline{\tilde{g}}$ é $m \times 1$. Este sistema pode ser como apresentado a seguir:



Em Modelos de Delineamentos Experimentais, em geral, o sistema de equações é consistente e indeterminado. Neste caso, se o usuário optar por processos matriciais para a

resolução do sistema, então um conceito mais abrangente e menos restritivo de matriz inversa deve ser adotado: o conceito de inversa generalizada.

1.12.3. Inversas generalizadas

Uma solução para um sistema consistente de equações $Ax = g$ pode ser expresso em termos de uma inversa generalizada de A . Neste texto, dada a matriz ${}_m A_n$ então sua inversa generalizada ${}_n A_m^G$ terá os seguintes nomes e notações:

- Inversa condicional de $A = A^-$
- Inversa de Moore-Penrose de $A = A^+$
- Inversa de mínimos quadrados de $A = A^\ell$

As matrizes A^- , A^+ e A^ℓ são matrizes inversas generalizadas.

Inversa condicional de A (A^-)

Definição: Dada uma matriz ${}_m A_n$, então toda matriz ${}_n A_m^-$ que satisfaz a condição

$$A A^- A = A$$

é definida como uma inversa condicional de A .

Uma inversa condicional só é única quando A for não singular. Neste caso, $A^- = A^{-1}$.

Métodos para obtenção de A^-

A seguir são apresentados dois métodos para se obter inversas condicionais de uma matriz A . Antes, porém, considere a seguinte definição: Uma matriz H quadrada de ordem n é dita estar na forma de Hermite (superior) se e somente se ela satisfaz as seguintes condições:

- H é triangular superior;
- Tem somente 0 e 1 na diagonal;
- Se uma linha tem 0 (zero) na diagonal, então todo elemento daquela linha é zero;
- Se uma linha tem 1 na diagonal, então os outros elementos da coluna na qual o 1 aparece são nulos.

Método 1: Seja A uma matriz quadrada. Seja $A \sim H$ para indicar que H foi obtida de A por meio de operações elementares nas linhas de A . Assim, a obtenção de A^- é feita da seguinte forma:

$$[A \mid I] \sim \dots \sim [H \mid L]$$

Então, para H na forma de Hermite, tem-se que $L = A^-$ é uma inversa condicional de A .

Exemplo: Obter uma A^- para

$$A = \begin{bmatrix} 4 & 2 & 2 \\ 2 & 2 & 0 \\ 2 & 0 & 2 \end{bmatrix}$$

$$\left[\begin{array}{ccc|ccc} 4 & 2 & 2 & 1 & 0 & 0 \\ 2 & 2 & 0 & 0 & 1 & 0 \\ 2 & 0 & 2 & 0 & 0 & 1 \end{array} \right] \sim \left[\begin{array}{ccc|ccc} 1 & 1/2 & 1/2 & 1/4 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1/2 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1/2 \end{array} \right]$$

$$\left[\begin{array}{ccc|ccc} 1 & 1/2 & 1/2 & 1/4 & 0 & 0 \\ 0 & 1/2 & -1/2 & -1/4 & 1/2 & 0 \\ 0 & -1/2 & 1/2 & -1/4 & 0 & 1/2 \end{array} \right] \sim \left[\begin{array}{ccc|ccc} 1 & 0 & 1 & 1/2 & -1/2 & 0 \\ 0 & 1/2 & -1/2 & -1/4 & 1/2 & 0 \\ 0 & 0 & 0 & -1/2 & 1/2 & 1/2 \end{array} \right]$$

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 1 & 1/2 & -1/2 & 0 \\ 0 & 1 & -1 & -1/2 & 1 & 0 \\ 0 & 0 & 0 & -1/2 & 1/2 & 1/2 \end{array} \right] = [H \mid L]$$

$$\text{Logo, } A^- = \begin{bmatrix} 1/2 & -1/2 & 0 \\ -1/2 & 1 & 0 \\ -1/2 & 1/2 & 1/2 \end{bmatrix}.$$

É fácil ver que H é uma forma de Hermite, e ainda é imediato verificar que $AA^-A = A$ e H é idempotente.

Método 2: O método dado a seguir é uma algoritmo apresentado por Searle (1992) para encontrar inversas condicionais.

Dada a matriz ${}_n A_p$ de posto r :

- i) Tomar qualquer submatriz não singular, M , de ordem r ;
- ii) Obter $(M^{-1})'$;
- iii) Substituir em A , os elementos de M por seus correspondentes em $(M^{-1})'$;
- iv) Fazer todos os outros elementos iguais a zero;
- v) Obter a transposta da matriz resultante;
- vi) O resultado é uma inversa condicional de A .

Exemplo: Obter uma A^- para

$$A = \begin{bmatrix} 1 & -1 & 1 \\ 2 & 3 & 0 \end{bmatrix}$$

$$\text{posto}(A) = 2$$

$$\text{Seja } M = \begin{bmatrix} 1 & 1 \\ 2 & 0 \end{bmatrix}, \text{ então } M^{-1} = \begin{bmatrix} 0 & 1/2 \\ 1 & -1/2 \end{bmatrix}$$

$$(M^{-1})' = \begin{bmatrix} 0 & 1 \\ 1/2 & -1/2 \end{bmatrix}$$

$$\begin{bmatrix} 0 & 0 & 1 \\ 1/2 & 0 & -1/2 \end{bmatrix}$$

$$\text{Logo, } A^- = \begin{bmatrix} 0 & 1/2 \\ 0 & 0 \\ 1 & -1/2 \end{bmatrix} \text{ é uma inversa condicional de } A.$$

Escolhendo-se outras submatrizes M , outras inversas condicionais para A poderão ser obtidas.

Se A é simétrica, então o algoritmo torna-se ainda mais simples:

- i) Tomar alguma submatriz principal não singular, M , de ordem r . Submatriz principal: uma submatriz de uma matriz A é dita principal se é obtida pela escolha de certas linhas de A e as colunas com a mesma numeração. Assim, os elementos diagonais de uma submatriz principal de A são elementos diagonais de A .
- ii) Obter M^{-1} ;
- iii) Substituir em A , os elementos de M por seus correspondentes em M^{-1} ;
- iv) Fazer todos os outros elementos iguais a zero;

v) O resultado é uma inversa condicional simétrica de A.

Toda matriz simétrica de posto r tem pelo menos um menor principal de ordem r diferente de zero, isto é, o determinante de pelo menos uma submatriz principal de ordem r é diferente de zero.

Exemplo: Seja

$$A = \begin{bmatrix} 4 & 2 & 2 \\ 2 & 2 & 0 \\ 2 & 0 & 2 \end{bmatrix} \text{ com } \text{posto}(A) = 2.$$

São inversas condicionais de A, dentre outras:

$$A_1^- = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/2 \end{bmatrix}; \quad A_2^- = \begin{bmatrix} 1/2 & 0 & -1/2 \\ 0 & 0 & 0 \\ -1/2 & 0 & 1 \end{bmatrix}; \quad A_3^- = \begin{bmatrix} 1/2 & -1/2 & 0 \\ -1/2 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

As propriedades apresentadas a seguir podem ser vistas em Graybill, 1969, Searle, 1971, dentre outros.

Dada a matriz ${}_n A_p$ e uma condicional ${}_p A_n^-$ então:

- i) A^- só é única se $n = p$ e $\text{posto}(A) = n$. Nesse caso, $A^- = A^{-1}$.
- ii) As matrizes A , A^- , AA^- , A^-A e $A(A'A)^-A'$ têm o mesmo posto;
- iii) $A(A'A)^-A'$ é invariante para qualquer inversa condicional de $A'A$ e, é simétrica e idempotente;
- iv) $\begin{bmatrix} I_{(n)} - A(A'A)^-A' \end{bmatrix}$ é simétrica e idempotente.

Inversa de Moore-Penrose de A (A^+)

Definição: Dada a matriz ${}_m A_n$ de posto k , então a matriz ${}_n A_m^+$ de posto k , que satisfaz às quatro condições seguintes, é definida como a inversa de Moore-Penrose de A.

- i) $AA^+A = A$
- ii) $A^+AA^+ = A^+$
- iii) $AA^+ = (AA^+)'$ (simétrica)
- iv) $A^+A = (A^+A)'$ (simétrica)

Teorema: Para cada matriz A existe uma e uma só A^+ .

i) Existência:

- i.1) Se ${}_m A_n = {}_m \phi_n \quad \exists \quad {}_n A_m^+ = {}_n \phi_m$ (trivial)
- i.2) Se ${}_m A_n \neq {}_m \phi_n$ e se $\text{posto}(A) = k > 0$, então $\exists \quad {}_m B_k$ e ${}_k C_n$, ambas de posto k , tais que ${}_m A_n = {}_m B_k {}_k C_n$.

Sendo B de posto coluna completo e C de posto linha completo, então $B'B$ e CC' são não singulares.

Seja $A^+ = C'(CC')^{-1}(B'B)^{-1}B'$, então

- a) $AA^+A = BCC'(CC')^{-1}(B'B)^{-1}B'BC = BIIC = BC = A$;
- b) $A^+AA^+ = C'(CC')^{-1}(B'B)^{-1}B'BCC'(CC')^{-1}(B'B)^{-1}B' = C'(CC')^{-1}(B'B)^{-1}B' = A^+$;
- c) $AA^+ = BCC'(CC')^{-1}(B'B)^{-1}B' = B(B'B)^{-1}B'$ (simétrica)
- d) $A^+A = C'(CC')^{-1}(B'B)^{-1}B'BC = C'(CC')^{-1}C$ (simétrica)

$$\therefore \exists A^+ = C'(CC')^{-1}(B'B)^{-1}B'$$

ii) Unicidade:

Sejam A_1^+ e A_2^+ duas inversas de Moore-Penrose para A. Então:

- a) $AA_1^+A = A$
- a') $AA_2^+A = A$
- b) $A_1^+AA_1^+ = A_1^+$
- b') $A_2^+AA_2^+ = A_2^+$
- c) $AA_1^+ = (AA_1^+)'$
- c') $AA_2^+ = (AA_2^+)'$
- d) $A_1^+A = (A_1^+A)'$
- d') $A_2^+A = (A_2^+A)'$

ii.1) Pós-multiplicando A por A_2^+ , vem:

$$AA_2^+ \stackrel{(a)}{=} AA_1^+AA_2^+ \stackrel{(c)}{=} (AA_1^+)'(AA_2^+)' \stackrel{(a')}{=} (AA_2^+AA_1^+)' \stackrel{(c)}{=} (AA_1^+)' = AA_1^+ \quad \therefore AA_2^+ = AA_1^+ \quad (e)$$

ii.2) Pré-multiplicando A por A_1^+ , fica:

$$A_1^+ A = A_1^+ A A_2^+ A \stackrel{(d)}{=} (A_1^+ A)' (A_2^+ A)' = (A_2^+ A A_1^+ A)' \stackrel{(a)}{=} (A_2^+ A)' \stackrel{(d')}{=} A_2^+ A \quad \therefore A_1^+ A = A_2^+ A \quad (f)$$

ii.3) Além disso,

$$A_1^+ = A_1^+ A A_1^+ \stackrel{(f)}{=} A_2^+ A A_1^+ \stackrel{(e)}{=} A_2^+ A A_2^+ = A_2^+ \stackrel{(b')}{=}$$

$$\therefore A_1^+ = A_2^+$$

Método para obtenção de A^+

Foi visto que, para obtenção de A^+ , fatora-se a matriz ${}_m A_n$ na forma ${}_m A_n = {}_m B_{kk} C_n$ e usa-se a expressão:

$$A^+ = C'(CC')^{-1}(B'B)^{-1}B'$$

Sem dúvida, a obtenção da inversa de Moore-Penrose é, em geral, bastante trabalhosa, mormente quando não se pode contar com pacotes estatísticos. Essa dificuldade é devida, em grande parte, à fatoração de A em $A = BC$. No entanto, a obtenção de B e C pode ser sensivelmente simplificada através do uso do excelente algoritmo de DWIVEDI (1975), exposto a seguir.

Dada a matriz ${}_m A_n = (a_{ij})$, $i = 1, \dots, p, \dots, m$; $j = 1, \dots, q, \dots, n$; tal que $\text{posto}[A] = k$:

i) Escolher algum elemento $a_{pq} \neq 0$;

ii) Obter $\underline{u}_1 \ \underline{v}'_1$, onde

$$\underline{u}_1 = \frac{1}{a_{pq}} \begin{bmatrix} a_{1q} \\ a_{2q} \\ \dots \\ a_{mq} \end{bmatrix}; \quad \underline{v}'_1 = [a_{p1}, \ a_{p2}, \ \dots, \ a_{pn}] ;$$

iii) Fazer $A_1 = A - \underline{u}_1 \underline{v}'_1$

iv) Se $A_1 = \phi$ então o processo está encerrado. Neste caso, $B = \underline{u}_1$ e $C = \underline{v}'_1$.

Se $A_1 \neq \phi$, repetir o processo para A_1 e assim por diante, até que se obtenha $A_k = \phi$.

Uma vez que em cada passo aumenta de pelo menos um o número de linhas e colunas nulas, é claro que o processo terminará depois de um número finito de passos.

Uma propriedade importante do algoritmo é a da convergência: o número de ciclos é igual ao posto de A .

Assim virá:

$$\begin{aligned}
 A - u_1 v_1' &= A_1 \\
 A_1 - u_2 v_2' &= A_2 \\
 &\dots \\
 A_{k-2} - u_{k-1} v_{k-1}' &= A_{k-1} \neq \phi \\
 A_{k-1} - u_k v_k' &= \phi
 \end{aligned}$$

Ao final do processo, tem-se:

$${}_m A_n = u_1 v_1' + u_2 v_2' + \dots + u_k v_k' = {}_m B_{k \times k} C_n, \text{ onde naturalmente,}$$

$$B = \begin{bmatrix} u_1 & u_2 & \dots & u_k \end{bmatrix}, \quad C = \begin{bmatrix} v_1' \\ v_2' \\ \dots \\ v_k' \end{bmatrix}$$

Em sequência, a inversa de Moore-Penrose pode ser facilmente obtida por:

$$A^+ = C'(CC')^{-1}(B'B)^{-1}B'$$

Exemplo: Dada a matriz

$$A = \begin{bmatrix} 2 & 6 & 4 & 2 \\ 4 & 15 & 14 & 7 \\ 2 & 9 & 10 & 5 \end{bmatrix}, \text{ obter a inversa de Moore-Penrose de } A.$$

i) Seja $a_{pq} = a_{11} = 2$

$$\begin{aligned}
 u_1 &= \frac{1}{a_{11}} \begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 2 \\ 4 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \quad v_1' = [a_{11} \quad a_{12} \quad a_{13} \quad a_{14}] \\
 &= [2 \quad 6 \quad 4 \quad 2]
 \end{aligned}$$

$$A_1 = A - u_1 v_1' = \begin{bmatrix} 2 & 6 & 4 & 2 \\ 4 & 15 & 14 & 7 \\ 2 & 9 & 10 & 5 \end{bmatrix} - \begin{bmatrix} 2 & 6 & 4 & 2 \\ 4 & 12 & 8 & 4 \\ 2 & 6 & 4 & 2 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 3 & 6 & 3 \\ 0 & 3 & 6 & 3 \end{bmatrix}$$

ii) Seja em A_1 , $a_{pq} = a_{22} = 3$, então:

$$u_2 = \frac{1}{3} \begin{bmatrix} 0 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \quad v_2' = [0 \quad 3 \quad 6 \quad 3]$$

$$A_2 = A_1 - \underline{u}_2 \underline{v}_2' = \phi \quad \Rightarrow \quad p[A] = 2$$

$$\therefore B = \begin{bmatrix} \underline{u}_1 & \underline{u}_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 2 & 1 \\ 1 & 1 \end{bmatrix} \text{ e } C = \begin{bmatrix} \underline{v}_1' \\ \underline{v}_2' \end{bmatrix} = \begin{bmatrix} 2 & 6 & 4 & 2 \\ 0 & 3 & 6 & 3 \end{bmatrix}$$

Assim, como $A^+ = C'(CC')^{-1}(B'B)^{-1}B'$, virá:

$$(CC') = \begin{bmatrix} 60 & 48 \\ 48 & 54 \end{bmatrix} \quad \Rightarrow \quad (CC')^{-1} = \begin{bmatrix} 3/52 & -2/39 \\ -2/39 & 5/78 \end{bmatrix}$$

$$(B'B) = \begin{bmatrix} 6 & 3 \\ 3 & 2 \end{bmatrix} \quad \Rightarrow \quad (B'B)^{-1} = \begin{bmatrix} 2/3 & -1 \\ -1 & 2 \end{bmatrix}$$

$$\therefore A^+ = \frac{1}{78} \begin{bmatrix} 14 & 3 & -11 \\ 19 & 5 & -14 \\ -18 & -2 & 16 \\ -9 & -1 & 8 \end{bmatrix},$$

que, sem dúvida, satisfaz às quatro condições da definição de A^+ .

FATO: A matriz A^+ é de grande valia na obtenção de resultados teóricos. Nesse contexto, as propriedades relacionadas a seguir são importantes. Outras podem ser vistas em GRAYBILL, 1969.

Propriedades:

P.1) $(A^+)' = (A')^+$ e $(A^+)^+ = A$;

P.2) $(A'A)^+ = A^+A'^+$ e $(AA^+)^+ = AA^+$;

P.3) AA^+ e $I - AA^+$ são simétricas e idempotentes;

P.4) $AA^+ = A(A'A)^-A'$;

P.5) Se $\text{posto}[A] = k$, então $\text{posto}[A^+] = \text{posto}[AA^+] = k$ e

$$\text{posto} \begin{bmatrix} I_{(n)} & -AA^+ \end{bmatrix} = n - k;$$

P.6) Se A é não singular, então $A^+ = A^- = A^{-1}$;

P.7) Se ${}_m A_n$ tem posto m , então $A^+ = A'(AA')^{-1}$ e $AA^+ = I$;

Se ${}_m A_n$ tem posto n , então $A^+ = (A'A)^{-1}A'$ e $A^+A = I$;

P.8) Se A é simétrica e idempotente, então $A^+ = A$;

P.9) Se $A = A_1 + A_2 + \dots + A_p$ e se $A_i A_{i'} = \phi$ e $A_i' A_i = \phi$, então

$$A^+ = A_1^+ + A_2^+ + \dots + A_p^+ \quad (i, i' = 1, \dots, p; i \neq i') ;$$

P.10) Se ${}_m A_n$ é uma matriz com $m \cdot n$ elementos iguais a 1, então $A^+ = \frac{1}{m \cdot n} A'$.

Inversa de mínimos quadrados de A (A^ℓ)

Definição: Dada a matriz ${}_m A_n$, então uma matriz ${}_n A_m^\ell$ que satisfaz às duas condições seguintes, é definida como inversa de mínimos quadrados de A .

- i) $AA^\ell A = A$
- ii) $(AA^\ell)' = AA^\ell$ (simétrica)

Obtenção de A^ℓ

Teorema: Se A é uma matriz de ordem $m \times n$, então $A^\ell = (A'A)^- A'$ é uma inversa de mínimos quadrados de A , onde $(A'A)^-$ é uma inversa condicional de $A'A$.

Exemplo: Seja $A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix}$, obter duas A^ℓ .

$$A'A = \begin{bmatrix} 4 & 2 & 2 \\ 2 & 2 & 0 \\ 2 & 0 & 2 \end{bmatrix}$$

Uma inversa condicional de $A'A$ é por exemplo:

$$(A'A)^- = \begin{bmatrix} 1/2 & -1/2 & 0 \\ -1/2 & 1 & 0 \\ -1 & 1 & 1 \end{bmatrix}$$

Assim virá:

$$A^\ell = \begin{bmatrix} 1/2 & -1/2 & 0 \\ -1/2 & 1 & 0 \\ -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1/2 & 1/2 \\ 1/2 & 1/2 & -1/2 & -1/2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Vamos obter uma outra inversa de mínimos quadrados de A .

$$\text{Sabemos que } (A'A)^- = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/2 \end{bmatrix} \text{ é também uma inversa condicional de } A'A.$$

Assim virá,

$$A^\ell = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 \end{bmatrix}$$

Sem dúvida, as duas matrizes A^ℓ obtidas, satisfazem às duas condições da definição.

1.13. CONSISTÊNCIA EM SISTEMAS DE EQUAÇÕES LINEARES

Foi visto que, se A é não singular, então o sistema de equações lineares $A \underline{x} = \underline{g}$ tem solução única, sendo, portanto consistente.

No entanto, como será visto adiante, é importante que se generalize o conceito de consistência para sistemas nos quais a matriz A não é necessariamente não singular.

A seguir são apresentadas algumas regras que nos permitem avaliar a consistência de sistemas mais gerais.

Teorema: (GRAYBILL, 1969, Cap. 7)

Uma condição necessária e suficiente para que o sistema $A \underline{x} = \underline{g}$ seja consistente é que o posto da matriz A seja igual ao posto da matriz A aumentada com \underline{g} .

$$\exists \underline{x} : A \underline{x} = \underline{g} \quad \Leftrightarrow \quad \text{posto}[A] = \text{posto}[A : \underline{g}]$$

Fato: Pode ser notado que essa condição é equivalente a exigir que \underline{g} pertença ao espaço coluna de A : $\underline{g} \in C(A)$.

Exemplo: Seja o sistema de equações lineares $A \underline{x} = \underline{g}$ dado por:

$$\begin{bmatrix} 6 & 3 & 3 \\ 3 & 3 & 0 \\ 3 & 0 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 10 \\ 6 \\ 4 \end{bmatrix}, \quad \text{posto}[A] = 2 \Rightarrow \text{não existe } A^{-1}$$

Para aplicação do Teorema anterior, basta operar nas linhas de A aumentada com \underline{g} .

$$\begin{bmatrix} 6 & 3 & 3 & : & 10 \\ 3 & 3 & 0 & : & 6 \\ 3 & 0 & 3 & : & 4 \end{bmatrix} \sim \dots \sim \begin{bmatrix} 1 & 0 & 1 & : & 4/3 \\ 0 & 1 & -1 & : & 2/3 \\ 0 & 0 & 0 & : & 0 \end{bmatrix}$$

$\therefore \text{posto}[A] = \text{posto}[A : \underline{g}] = 2$, e o sistema é consistente.

Vejam agora uma regra geral para obtenção de soluções de sistemas consistentes.

Teorema: Uma condição necessária e suficiente para que o sistema de equações $A \tilde{x} = \tilde{g}$ seja

consistente é que exista uma inversa condicional de A tal que $AA^{-} \tilde{g} = \tilde{g}$

$$a) A \tilde{x} = \tilde{g} \text{ consistente} \Rightarrow AA^{-} \tilde{g} = \tilde{g}$$

$$A \tilde{x} = \tilde{g} \text{ consistente} \Rightarrow \exists \tilde{x}^0 : A \tilde{x}^0 = \tilde{g} \quad (I)$$

Seja A^{-} alguma inversa condicional de A. Pré-multiplicando (I) por AA^{-} , vem

$$AA^{-} A \tilde{x}^0 = AA^{-} \tilde{g} \quad \therefore A \tilde{x}^0 = AA^{-} \tilde{g} \text{ e usando (I) temos } \tilde{g} = AA^{-} \tilde{g}$$

$$b) AA^{-} \tilde{g} = \tilde{g} \Rightarrow A \tilde{x} = \tilde{g} \text{ consistente}$$

Seja $AA^{-} \tilde{g} = \tilde{g}$. Basta então tomar $\tilde{x}^0 = A^{-} \tilde{g}$ e desse modo $A \tilde{x}^0 = \tilde{g}$ e o sistema é

consistente.

FATOS:

i) Note que sendo A^{+} também A^{-} , então, uma condição necessária e suficiente para a consistência de $A \tilde{x} = \tilde{g}$ é $AA^{+} \tilde{g} = \tilde{g}$. Nesse contexto $\tilde{x}^0 = A^{+} \tilde{g}$ é solução.

Para o sistema do exemplo anterior, virá:

$$AA^{-} \tilde{g} = \begin{bmatrix} 6 & 3 & 3 \\ 3 & 3 & 0 \\ 3 & 0 & 3 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1/3 & 0 \\ 0 & 0 & 1/3 \end{bmatrix} \begin{bmatrix} 10 \\ 6 \\ 4 \end{bmatrix} = \begin{bmatrix} 10 \\ 6 \\ 4 \end{bmatrix} = \tilde{g} \quad (\text{Consistente})$$

De modo análogo, tomando-se a A^{+} , virá:

$$AA^{+} \tilde{g} = \frac{1}{27} \begin{bmatrix} 6 & 3 & 3 \\ 3 & 3 & 0 \\ 3 & 0 & 3 \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 \\ 1 & 5 & -4 \\ 1 & -4 & 5 \end{bmatrix} \begin{bmatrix} 10 \\ 6 \\ 4 \end{bmatrix} = \begin{bmatrix} 10 \\ 6 \\ 4 \end{bmatrix} = \tilde{g}$$

E, dentre as possíveis soluções, duas delas são apresentadas a seguir:

$$\tilde{x}_1^0 = A^{-} \tilde{g} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1/3 & 0 \\ 0 & 0 & 1/3 \end{bmatrix} \begin{bmatrix} 10 \\ 6 \\ 4 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \\ 4/3 \end{bmatrix}$$

$$\tilde{x}_2^0 = A^{+} \tilde{g} = \frac{1}{27} \begin{bmatrix} 2 & 1 & 1 \\ 1 & 5 & -4 \\ 1 & -4 & 5 \end{bmatrix} \begin{bmatrix} 10 \\ 6 \\ 4 \end{bmatrix} = \begin{bmatrix} 10/9 \\ 8/9 \\ 2/9 \end{bmatrix}.$$

ii) Note que o sistema apresenta "várias" soluções, duas das quais foram aqui obtidas. Isto quer dizer que se o sistema é consistente e indeterminado, existem "várias" combinações lineares das colunas de A que reproduzem \underline{g} .

No exemplo em questão:

$$\underline{x}_{\tilde{1}}^0 \Rightarrow 0 \begin{bmatrix} 6 \\ 3 \\ 3 \end{bmatrix} + 2 \begin{bmatrix} 3 \\ 3 \\ 0 \end{bmatrix} + \frac{4}{3} \begin{bmatrix} 3 \\ 0 \\ 3 \end{bmatrix} = \begin{bmatrix} 10 \\ 6 \\ 4 \end{bmatrix}$$

$$\underline{x}_{\tilde{2}}^0 \Rightarrow \frac{10}{9} \begin{bmatrix} 6 \\ 3 \\ 3 \end{bmatrix} + \frac{8}{9} \begin{bmatrix} 3 \\ 3 \\ 0 \end{bmatrix} + \frac{2}{9} \begin{bmatrix} 3 \\ 0 \\ 3 \end{bmatrix} = \begin{bmatrix} 10 \\ 6 \\ 4 \end{bmatrix}$$

iii) Mesmo que a matriz do sistema não seja quadrada, pode-se ter solução única. Basta que ela tenha posto coluna completo. Assim, se ${}_m A_n$ tem posto n , então $A^+ = (A'A)^{-1} A'$. Logo, $\underline{x}^0 = A^+ \underline{g} = (A'A)^{-1} A' \underline{g}$.

1.14. EXERCÍCIOS

1. Dadas as matrizes:

$$A = \begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix}, \quad I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad D = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}, \quad B = \begin{bmatrix} 5 & 3 & 2 \\ 3 & 3 & 0 \\ 2 & 0 & 2 \end{bmatrix}, \quad C = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1/3 & 0 \\ 0 & 0 & 1/2 \end{bmatrix}$$

$$\underline{Y} = \begin{bmatrix} 18 \\ 12 \\ 6 \end{bmatrix}, \quad M = \begin{bmatrix} 5 & 0 & 0 \\ 3 & 3 & 0 \\ 2 & 0 & 2 \end{bmatrix} \text{ e } E = \begin{bmatrix} 2 & 5 \\ 6 & 4 \end{bmatrix},$$

pede-se calcular:

a) AI , IA , AA , AA' , $A'A$, AD e DA ;

b) CB e $C\underline{Y}$;

c) $W'CB$, onde $W' = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & -1 \end{bmatrix}$;

d) $\text{Det}(B)$ e $\text{Posto}(B)$;

e) $(AE)'$ e $E'A'$ (Note que: $(AE)' = E'A'$);

f) Dado o vetor:

$$\underline{X} = \begin{bmatrix} 3 \\ 1 \\ 4 \end{bmatrix}, \text{ calcular } \underline{X}'\underline{X} \text{ e } \underline{X}\underline{X}';$$

g) Verifique que a matriz A é ortogonal, em que:

$$A = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}.$$

2. Verifique que a matriz B é singular, em que:

$$B = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix}.$$

3. Determinar a e x de modo que sejam singulares as matrizes:

$$M = \begin{bmatrix} 1 & 8 \\ 3 & a \end{bmatrix} \text{ e } N = \begin{bmatrix} 1-x & 2 \\ 4 & 8-x \end{bmatrix}.$$

4. Seja o sistema de equações:

$$2x_1 + x_2 = 3$$

$$2x_1 + 3x_2 = 5$$

Pede-se:

- Escreva o sistema na forma matricial $\tilde{A} \tilde{x} = \tilde{g}$;
- Determine a solução do sistema sabendo-se que $\tilde{x} = \tilde{A}^{-1} \tilde{g}$;
- Pré-multiplique ambos os lados da igualdade de (a) por \tilde{A}' , obtendo $\tilde{A}'\tilde{A}x = \tilde{A}'\tilde{g}$;
- Determine a solução para o “novo” sistema através de $\tilde{x} = (\tilde{A}'\tilde{A})^{-1} \tilde{A}'\tilde{g}$;
- Compare os resultados obtidos em (b) e (d).

5. Seja o sistema de equações:

$$x_1 + x_2 + x_3 = 9$$

$$2x_1 + x_2 = 5$$

$$x_2 + x_3 = 8$$

Pede-se:

- Escreva o sistema na forma matricial $\tilde{A} \tilde{x} = \tilde{g}$;
- Determine a solução do sistema, sabendo-se que $\tilde{x} = \tilde{A}^{-1} \tilde{g}$.

6. Obter os autovalores e o autovetor normalizado associado ao maior autovalor das seguintes matrizes:

$$A = \begin{bmatrix} 2 & 3 \\ 3 & 10 \end{bmatrix} \text{ e } B = \begin{bmatrix} 2 & 3 & 3 \\ 2 & 7 & 6 \\ 3 & 9 & 10 \end{bmatrix}.$$

7. Dada a matriz abaixo:

$$A = \begin{bmatrix} 8 & 5 & 3 \\ 5 & 5 & 0 \\ 3 & 0 & 3 \end{bmatrix}.$$

Pede-se:

- Obter duas inversas condicionais para A (uma por cada método apresentado);
- Verifique numericamente que $AA^{-1}A = A$.

$$8. \text{ Determinar o posto das matrizes } A \text{ e } B, \text{ onde } A = \begin{bmatrix} 1 & 0 & -1 & 2 \\ 3 & 1 & 4 & 2 \\ 5 & 2 & 9 & 2 \end{bmatrix} \text{ e } B = \begin{bmatrix} 7 & 3 \\ 4 & 6 \end{bmatrix}.$$

9. A matriz $P = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$ é ortogonal? **Justifique sua resposta.**

10. Dada a forma quadrática $Q(\underline{x}) = \underline{x}' A \underline{x}$, onde $\underline{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ e $A = \begin{bmatrix} 1 & 3 \\ 3 & 2 \end{bmatrix}$, pede-se:

- a) Diagonalizar A por operações de congruência;
- b) Classifique a forma quadrática com base em (a).

6) Dada a forma quadrática

$$Q(\underline{x}) = \underline{x}' A \underline{x} = 6x_1^2 + 5x_2^2 + 9x_3^2 + 4x_1x_2 + 6x_1x_3 + 8x_2x_3, \text{ pede-se:}$$

- a) A matriz A ;
- b) Calcule o determinante de A ;
- c) Obter a inversa A^{-1} ;
- d) Obter os autovalores de A ;
- e) Classifique a matriz A com base em (d);
- f) Obter os autovetores normalizados associados aos autovalores;
- g) Obtenha a decomposição espectral de A ;
- h) Se ${}_p A_p$ for **positiva definida**, considere os autovetores normalizados as colunas de uma

outra matriz $P = \begin{bmatrix} \underline{v}_1 & \underline{v}_2 & \dots & \underline{v}_p \end{bmatrix}$. Neste caso, pede-se:

h₁) P é ortogonal? Por quê?

h₂) Verifique a relação $A = P\Lambda P'$, onde Λ é uma matriz diagonal constituída pelos autovalores de A , isto é,

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & \lambda_p \end{bmatrix}, \text{ com } \lambda_i > 0.$$

h₃) Obtenha A^{-1} tal que: $A^{-1} = P\Lambda^{-1}P' = \sum_{i=1}^p \frac{1}{\lambda_i} \underline{v}_i \underline{v}_i'$.

- i) Diagonalizar A por operações de congruência;
- j) Classifique a matriz A com base na matriz diagonal obtida em (i).

1.15. REFERÊNCIAS

DWIVEDI, T. D. A method to compute the rank factors of a matrix. Sankhyā: The Indian Journal of Statistics, v. 37, p. 463-464, 1975.

GRAYBILL, F. A. **Introduction to Matrices with Application in Statistic**. California: Wadsworth Publishing, 1969. 372p.

- GRAYBILL, F. A. **Theory and application of the linear model**. USA: Duxbury Press, 1976. 704p.
- JOHNSON, R. A.; WICHERN, D.W. **Applied multivariate statistical analysis**. 4th ed., USA, Upper Saddle River, New Jersey: Prentice Hall, Inc., 1998. 816p.
- MARDIA, K. V.; KENT, J. T.; BIBBY, J. M. **Multivariate analysis**. 6th ed., London: Academic Press, 1997. 518p.
- RENCER, A. C. **Methods of multivariate analysis**. 2nd ed., USA: John Wiley and Sons, 2002. 738p.
- RENCER, A.C.; SCHALJE, G.B. **Linear models in statistics**. 2nd ed., USA: John Wiley and Sons, 2008. 672p.
- SEARLE, S.R. **Linear Models**. New York: John Wiley and Sons, 1971. 532p.
- SEARLE, S.R. **Matrix Algebra Useful for Statistics**. 2nd ed., USA: John Wiley and Sons, 1992. 527p.
- STEINBRUCH, A.; WINTERLE, P. **Álgebra linear**. 2. ed., São Paulo: McGraw-Hill do Brasil, 1987. 583p.

CAPÍTULO 2

CONCEITOS INTRODUTÓRIOS

2.1. VETORES E MATRIZES ALEATÓRIAS

Um **vetor aleatório** é um vetor cujos elementos são variáveis aleatórias. Similarmente, uma **matriz aleatória** é uma matriz cujos elementos são variáveis aleatórias.

Um vetor aleatório ($p \times 1$) pode ser escrito como:

$$\tilde{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}.$$

Uma matriz aleatória ($p \times n$) pode ser escrita como:

$$X = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ X_{p1} & X_{p2} & \cdots & X_{pn} \end{bmatrix}.$$

O valor esperado de uma matriz aleatória (ou vetor) é a matriz constituída pelos valores esperados de cada um de seus elementos. Especificamente, seja $X = \{X_{ij}\}$ uma matriz aleatória ($p \times n$). O valor esperado de X , denotado por $E(X)$ é a matriz de números ($p \times n$) (se eles existem). Logo,

$$E(\tilde{X}) = \begin{bmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_p) \end{bmatrix} \quad \text{e} \quad E(X) = \begin{bmatrix} E(X_{11}) & E(X_{12}) & \cdots & E(X_{1n}) \\ E(X_{21}) & E(X_{22}) & \cdots & E(X_{2n}) \\ \vdots & \vdots & \cdots & \vdots \\ E(X_{p1}) & E(X_{p2}) & \cdots & E(X_{pn}) \end{bmatrix}.$$

2.2. VETORES DE MÉDIAS, MATRIZES DE COVARIÂNCIAS E DE CORRELAÇÕES

Seja $\tilde{X} = \{X_i\}$ um vetor aleatório ($p \times 1$). Cada elemento de \tilde{X} é uma variável aleatória com sua própria distribuição de probabilidade marginal. No caso contínuo, o vetor aleatório \tilde{X} é descrito por uma função densidade de probabilidade conjunta $f_{X_1, X_2, \dots, X_p}(x_1, x_2, \dots, x_p)$. A média do vetor aleatório \tilde{X} é dada por:

$$E(\tilde{X}) = \begin{bmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} = \tilde{\mu}.$$

A matriz de covariância (ou matriz de dispersão) do vetor aleatório \tilde{X} é uma matriz simétrica dada por:

$$\begin{aligned} \Sigma &= D(\tilde{X}) = \text{Cov}(\tilde{X}) = E \left[\left(\tilde{X} - \tilde{\mu} \right) \left(\tilde{X} - \tilde{\mu} \right)' \right] = E(\tilde{X} \tilde{X}') - \tilde{\mu} \tilde{\mu}' \\ &= E \left\{ \begin{bmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \\ \vdots \\ X_p - \mu_p \end{bmatrix} \begin{bmatrix} X_1 - \mu_1 & X_2 - \mu_2 & \cdots & X_p - \mu_p \end{bmatrix} \right\} \\ &= E \begin{bmatrix} (X_1 - \mu_1)^2 & (X_1 - \mu_1)(X_2 - \mu_2) & \cdots & (X_1 - \mu_1)(X_p - \mu_p) \\ (X_2 - \mu_2)(X_1 - \mu_1) & (X_2 - \mu_2)^2 & \cdots & (X_2 - \mu_2)(X_p - \mu_p) \\ \vdots & \vdots & \cdots & \vdots \\ (X_p - \mu_p)(X_1 - \mu_1) & (X_p - \mu_p)(X_2 - \mu_2) & \cdots & (X_p - \mu_p)^2 \end{bmatrix} \\ &= \begin{bmatrix} E[(X_1 - \mu_1)^2] & E[(X_1 - \mu_1)(X_2 - \mu_2)] & \cdots & E[(X_1 - \mu_1)(X_p - \mu_p)] \\ E[(X_2 - \mu_2)(X_1 - \mu_1)] & E[(X_2 - \mu_2)^2] & \cdots & E[(X_2 - \mu_2)(X_p - \mu_p)] \\ \vdots & \vdots & \cdots & \vdots \\ E[(X_p - \mu_p)(X_1 - \mu_1)] & E[(X_p - \mu_p)(X_2 - \mu_2)] & \cdots & E[(X_p - \mu_p)^2] \end{bmatrix}, \end{aligned}$$

ou ainda,

$$\Sigma = \text{Cov}(\tilde{X}) = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix},$$

em que,

$\sigma_{ii} = \sigma_i^2 = E[X_i - E(X_i)]^2 = E(X_i^2) - [E(X_i)]^2$ é a variância da variável X_i .

$\sigma_{ik} = \sigma_{ki} = E\{[X_i - E(X_i)][X_k - E(X_k)]\} = E(X_i X_k) - E(X_i)E(X_k)$ é a covariância entre as variáveis aleatórias X_i e X_k

A matriz de correlação populacional (ρ) é dada por:

$$\rho = \begin{bmatrix} \frac{\sigma_{11}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{11}}} & \frac{\sigma_{12}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}} & \cdots & \frac{\sigma_{1p}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{pp}}} \\ \frac{\sigma_{12}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{22}}} & \frac{\sigma_{22}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{22}}} & \cdots & \frac{\sigma_{2p}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{pp}}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sigma_{1p}}{\sqrt{\sigma_{11}}\sqrt{\sigma_{pp}}} & \frac{\sigma_{2p}}{\sqrt{\sigma_{22}}\sqrt{\sigma_{pp}}} & \cdots & \frac{\sigma_{pp}}{\sqrt{\sigma_{pp}}\sqrt{\sigma_{pp}}} \end{bmatrix}$$

$$\rho = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{12} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ \rho_{1p} & \rho_{2p} & \cdots & 1 \end{bmatrix}, \quad -1 \leq \rho \leq 1$$

Assim, o coeficiente de correlação (ou coeficiente de correlação momento produto de Pearson) entre as variáveis aleatórias X_i e X_k é dado por:

$$\rho_{ik} = \frac{\text{Cov}(X_i, X_k)}{\sqrt{V(X_i)}\sqrt{V(X_k)}} = \frac{\sigma_{ik}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{kk}}}$$

ou ainda,

$$\rho_{ik} = \frac{\sigma_{ik}}{\sqrt{\sigma_{ii} \cdot \sigma_{kk}}} = \frac{\sigma_{ik}}{\sqrt{\sigma_i^2 \cdot \sigma_k^2}}.$$

Fato: Para

$$V^{1/2} = \begin{bmatrix} \sqrt{\sigma_{11}} & 0 & \cdots & 0 \\ 0 & \sqrt{\sigma_{22}} & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \sqrt{\sigma_{pp}} \end{bmatrix}$$

é fácil verificar que:

i) $V^{1/2} \rho V^{1/2} = \Sigma$;

$$\text{ii) } \rho = (V^{1/2})^{-1} \Sigma (V^{1/2})^{-1}.$$

Assim, Σ pode ser obtida de $V^{1/2}$ e ρ , enquanto ρ pode ser obtida de Σ .

Para X e Y variáveis aleatórias, é importante lembrar que:

- a) Se X e Y são independentes, então $\text{Cov}(X, Y) = 0$ e, portanto $\rho_{XY} = 0$.
- b) Se $\text{Cov}(X, Y) = 0$ e consequentemente $\rho_{XY} = 0$, não é possível concluir, em geral, que as variáveis são independentes.
- c) Se X e Y apresentam uma distribuição normal bivariada, então $\text{Cov}(X, Y) = \rho_{XY} = 0$ é condição suficiente para que as variáveis sejam independentes. A prova deste fato é apresentada no Capítulo 3.

Seja um vetor aleatório \tilde{X} com média $\mu_{\tilde{X}}$ e matriz de covariância $\Sigma_{\tilde{X}}$. Então para

$$\tilde{Z} = C' \tilde{X}, \text{ segue que } E(\tilde{Z}) = C' \mu_{\tilde{X}} = \mu_{\tilde{Z}} \text{ e } \text{Cov}(\tilde{Z}) = V(\tilde{Z}) = C' \Sigma_{\tilde{X}} C = \Sigma_{\tilde{Z}}.$$

Prova:

$$\text{i) } E(\tilde{Z}) = E(C' \tilde{X}) = C' E(\tilde{X}) = C' \mu_{\tilde{X}} = \mu_{\tilde{Z}}.$$

$$\begin{aligned} \text{ii) } \text{Cov}(\tilde{Z}) &= E \left\{ \begin{bmatrix} \tilde{Z} - E(\tilde{Z}) \end{bmatrix} \begin{bmatrix} \tilde{Z} - E(\tilde{Z}) \end{bmatrix}' \right\} \\ &= E \left\{ \begin{bmatrix} C' \tilde{X} - E(C' \tilde{X}) \end{bmatrix} \begin{bmatrix} C' \tilde{X} - E(C' \tilde{X}) \end{bmatrix}' \right\} \\ &= E \left\{ C' \begin{bmatrix} \tilde{X} - E(\tilde{X}) \end{bmatrix} \begin{bmatrix} \tilde{X} - E(\tilde{X}) \end{bmatrix}' C \right\} \\ &= E \left\{ C' \begin{bmatrix} \tilde{X} - E(\tilde{X}) \end{bmatrix} \begin{bmatrix} \tilde{X} - E(\tilde{X}) \end{bmatrix}' C \right\} \end{aligned}$$

$$\begin{aligned}
&= C' E \left\{ \underbrace{\begin{bmatrix} \tilde{X} - E(\tilde{X}) \\ \tilde{X} - E(\tilde{X}) \end{bmatrix}}_{=\sum_x} \right\} C \\
&= C' \sum_x C = \sum_z
\end{aligned}$$

Exemplo 1:

$$\text{Para } \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{bmatrix} = \begin{bmatrix} 4 & 1 & 2 \\ 1 & 9 & -3 \\ 2 & -3 & 25 \end{bmatrix},$$

pede-se:

a) Obter a matriz de correlação populacional ρ ;

$$\text{b) Obter } V^{1/2} = \begin{bmatrix} \sqrt{\sigma_{11}} & 0 & 0 \\ 0 & \sqrt{\sigma_{22}} & 0 \\ 0 & 0 & \sqrt{\sigma_{33}} \end{bmatrix};$$

c) Obter ρ onde $\rho = (V^{1/2})^{-1} \Sigma (V^{1/2})^{-1}$;

d) Verifique que $V^{1/2} \rho V^{1/2} = \Sigma$.

Solução:

$$\text{a) } \rho_{12} = \frac{1}{\sqrt{4 \times 9}} = 1/6; \rho_{13} = \frac{2}{\sqrt{4 \times 25}} = 1/5; \rho_{23} = \frac{-3}{\sqrt{9 \times 25}} = -1/5.$$

$$\text{Logo, } \rho = \begin{bmatrix} 1 & 1/6 & 1/5 \\ 1/6 & 1 & -1/5 \\ 1/5 & -1/5 & 1 \end{bmatrix}.$$

$$\text{b) } V^{1/2} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{bmatrix}.$$

$$\text{c) } \rho = \begin{bmatrix} 1/2 & 0 & 0 \\ 0 & 1/3 & 0 \\ 0 & 0 & 1/5 \end{bmatrix} \begin{bmatrix} 4 & 1 & 2 \\ 1 & 9 & -3 \\ 2 & -3 & 25 \end{bmatrix} \begin{bmatrix} 1/2 & 0 & 0 \\ 0 & 1/3 & 0 \\ 0 & 0 & 1/5 \end{bmatrix} = \begin{bmatrix} 1 & 1/6 & 1/5 \\ 1/6 & 1 & -1/5 \\ 1/5 & -1/5 & 1 \end{bmatrix}.$$

$$d) \sum = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{bmatrix} \begin{bmatrix} 1 & 1/6 & 1/5 \\ 1/6 & 1 & -1/5 \\ 1/5 & -1/5 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{bmatrix} = \begin{bmatrix} 4 & 1 & 2 \\ 1 & 9 & -3 \\ 2 & -3 & 25 \end{bmatrix}.$$

Exemplo 2:

Seja $\tilde{X}' = [X_1, X_2]$ um vetor com vetor de médias $\mu'_X = [\mu_1, \mu_2]$ e matriz de covariância $\sum_X = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}$. Encontrar o vetor de médias e a matriz de covariância de \tilde{Z} , sendo

$$\tilde{Z} = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} = \begin{bmatrix} X_1 - X_2 \\ X_1 + X_2 \end{bmatrix}.$$

Solução:

$$\tilde{Z} = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = C' \tilde{X}$$

$$\mu_Z = E(\tilde{Z}) = C' E(\tilde{X}) = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} \mu_1 - \mu_2 \\ \mu_1 + \mu_2 \end{bmatrix}.$$

$$\begin{aligned} \sum_Z = \text{Cov}(\tilde{Z}) &= C' \sum_X C = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} \sigma_{11} - \sigma_{12} & \sigma_{12} - \sigma_{22} \\ \sigma_{11} + \sigma_{12} & \sigma_{12} + \sigma_{22} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} \sigma_{11} - 2\sigma_{12} + \sigma_{22} & \sigma_{11} - \sigma_{22} \\ \sigma_{11} - \sigma_{22} & \sigma_{11} + 2\sigma_{12} + \sigma_{22} \end{bmatrix}. \end{aligned}$$

Note que, se $\sigma_{11} = \sigma_{22}$, isto é, se X_1 e X_2 têm variâncias iguais, então $\sigma_{11} - \sigma_{22} = 0$ na matriz \sum_Z . Isto demonstra o resultado bem conhecido que a soma e diferença de duas variáveis aleatórias com variâncias idênticas são não correlacionadas.

2.3. MÉDIA, VARIÂNCIA, COVARIÂNCIA E CORRELAÇÃO (Amostrai)

Seja a notação X_{ij} para indicar um particular valor da i -ésima variável que é observado no j -ésimo item, ou indivíduo ou repetição. Assim, X_{ij} é a medida da i -ésima variável no j -ésimo item. Consequentemente, n medidas em p variáveis podem ser dispostas como apresentado na Tabela 2.1.

Tabela 2.1 – Matriz de dados de n medidas em p variáveis

Variáveis	Itens					
	1	2	...	j	...	n
1	X_{11}	X_{12}	...	X_{1j}	...	X_{1n}
2	X_{21}	X_{22}	...	X_{2j}	...	X_{2n}
...
i	X_{i1}	X_{i2}	...	X_{ij}	...	X_{in}
...
p	X_{p1}	X_{p2}	...	X_{pj}	...	X_{pn}

- Média amostral para a variável X_i :

$$\bar{X}_i = \frac{1}{n} \sum_{j=1}^n X_{ij}$$

- Variância amostral para a variável X_i :

$$s_i^2 = s_{ii} = \hat{V}(X_i) = \frac{\sum_{j=1}^n (X_{ij} - \bar{X}_i)^2}{n-1} = \frac{\sum_{j=1}^n X_{ij}^2 - \frac{(\sum_{j=1}^n X_{ij})^2}{n}}{n-1}$$

- Covariância amostral entre as variáveis X_i e X_k :

$$s_{ik} = \hat{Cov}(X_i, X_k) = \frac{\sum_{j=1}^n (X_{ij} - \bar{X}_i)(X_{kj} - \bar{X}_k)}{n-1}$$

$$= \frac{\sum_{j=1}^n X_{ij} X_{kj} - \frac{(\sum_{j=1}^n X_{ij})(\sum_{j=1}^n X_{kj})}{n}}{n-1},$$

com $i = 1, 2, \dots, p$ e $k = 1, 2, \dots, p$.

- O coeficiente de correlação amostral entre a i -ésima e a k -ésima variáveis é definido como:

$$r_{ik} = \hat{\rho}_{ik} = \frac{s_{ik}}{\sqrt{s_{ii} \cdot s_{kk}}} = \frac{s_{ik}}{\sqrt{s_i^2 \cdot s_k^2}}$$

$$r_{ik} = \frac{\sum_{j=1}^n (X_{ij} - \bar{X}_i)(X_{kj} - \bar{X}_k)}{\sqrt{\left[\sum_{j=1}^n (X_{ij} - \bar{X}_i)^2 \right] \left[\sum_{j=1}^n (X_{kj} - \bar{X}_k)^2 \right]}} = \frac{SPD_{X_i X_k}}{\sqrt{SQD_{X_i} \cdot SQD_{X_k}}}.$$

É importante ressaltar que um coeficiente de correlação igual a zero não implica em ausência de relação entre as duas variáveis. Um coeficiente de correlação nulo somente implica na ausência de relação linear entre as duas variáveis. Correlação linear perfeita ($r = \pm 1$) aconteceria quando todos os pontos caíssem exatamente sobre uma linha reta.

Seja a matriz de dados $p \times n$ como a seguir:

$$X = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ X_{p1} & X_{p2} & \cdots & X_{pn} \end{bmatrix} = \begin{bmatrix} X_1 & X_2 & \cdots & X_n \end{bmatrix}.$$

As estatísticas descritivas computadas das n medidas em p variáveis podem ser organizadas do seguinte modo:

- Vetor de médias amostrais: $\bar{X} = \frac{1}{n} \sum_{j=1}^n X_j = \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_p \end{bmatrix}$

- Matriz de covariância amostral:

$$S = \sum = \frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X})(X_j - \bar{X})' = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1p} \\ s_{21} & s_{22} & \cdots & s_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_{pp} \end{bmatrix} \quad (\text{simétrica: } s_{ik} = s_{ki})$$

- Matriz de correlação amostral:

$$R = \hat{\rho} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{bmatrix} \quad (\text{simétrica: } r_{ik} = r_{ki})$$

Exemplo 3:

Considerando duas variáveis X_1 e X_2 , com 4 medidas em cada uma delas, obteve-se os seguintes resultados:

$$X_{11} = 42 \quad X_{12} = 52 \quad X_{13} = 48 \quad X_{14} = 58$$

$$X_{21} = 4 \quad X_{22} = 5 \quad X_{23} = 4 \quad X_{24} = 3$$

$$\text{Assim, } \mathbf{X} = \begin{bmatrix} 42 & 52 & 48 & 58 \\ 4 & 5 & 4 & 3 \end{bmatrix}.$$

Pede-se calcular:

a) $\bar{\mathbf{X}}$;

b) S com $n - 1$ graus de liberdade;

c) $R = \hat{\rho}$.

Solução:

a) As médias amostrais são:

$$\bar{X}_1 = \frac{1}{4} \sum_{j=1}^4 X_{1j} = \frac{1}{4}(42 + 52 + 48 + 58) = 50$$

$$\bar{X}_2 = \frac{1}{4} \sum_{j=1}^4 X_{2j} = \frac{1}{4}(4 + 5 + 4 + 3) = 4$$

$$\text{Logo, } \bar{\mathbf{X}} = \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \end{bmatrix} = \begin{bmatrix} 50 \\ 4 \end{bmatrix}.$$

Pelo cálculo matricial, virá:

$$\bar{\mathbf{X}} = \frac{1}{4} \left\{ \begin{bmatrix} 42 \\ 4 \end{bmatrix} + \begin{bmatrix} 52 \\ 5 \end{bmatrix} + \begin{bmatrix} 48 \\ 4 \end{bmatrix} + \begin{bmatrix} 58 \\ 3 \end{bmatrix} \right\} = \frac{1}{4} \begin{bmatrix} 200 \\ 16 \end{bmatrix} = \begin{bmatrix} 50 \\ 4 \end{bmatrix}.$$

b) As variâncias e covariâncias amostrais são:

$$\begin{aligned} s_{11} = s_1^2 &= \frac{1}{4-1} \sum_{j=1}^4 (X_{1j} - \bar{X}_1)^2 = \frac{1}{3} [(42-50)^2 + (52-50)^2 + (48-50)^2 + (58-50)^2] \\ &= \frac{136}{3} \cong 45,33 \end{aligned}$$

$$s_{22} = s_2^2 = \frac{1}{4-1} \sum_{j=1}^4 (X_{2j} - \bar{X}_2)^2 = \frac{1}{3} [(4-4)^2 + (5-4)^2 + (4-4)^2 + (3-4)^2] = \frac{2}{3} \cong 0,67$$

$$\begin{aligned} s_{12} &= \frac{1}{4-1} \sum_{j=1}^4 (X_{1j} - \bar{X}_1)(X_{2j} - \bar{X}_2) = \frac{1}{3} [(42-50)(4-4) + (52-50)(5-4) \\ &\quad + (48-50)(4-4) + (58-50)(3-4)] = \frac{-6}{3} = -2 \end{aligned}$$

Logo, $S = \begin{bmatrix} 45,33 & -2 \\ -2 & 0,67 \end{bmatrix}$.

Pelo cálculo matricial, virá:

$$\begin{aligned} S &= \frac{1}{4-1} \sum_{j=1}^4 (\tilde{X}_j - \bar{\tilde{X}})(\tilde{X}_j - \bar{\tilde{X}})' \\ &= \frac{1}{3} \left[(\tilde{X}_1 - \bar{\tilde{X}})(\tilde{X}_1 - \bar{\tilde{X}})' + (\tilde{X}_2 - \bar{\tilde{X}})(\tilde{X}_2 - \bar{\tilde{X}})' + (\tilde{X}_3 - \bar{\tilde{X}})(\tilde{X}_3 - \bar{\tilde{X}})' + (\tilde{X}_4 - \bar{\tilde{X}})(\tilde{X}_4 - \bar{\tilde{X}})' \right] \\ &= \frac{1}{3} \left\{ \begin{bmatrix} -8 \\ 0 \end{bmatrix} \begin{bmatrix} -8 & 0 \end{bmatrix} + \begin{bmatrix} 2 \\ 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \end{bmatrix} + \begin{bmatrix} -2 \\ 0 \end{bmatrix} \begin{bmatrix} -2 & 0 \end{bmatrix} + \begin{bmatrix} 8 \\ -1 \end{bmatrix} \begin{bmatrix} 8 & -1 \end{bmatrix} \right\} \\ &= \frac{1}{3} \left\{ \begin{bmatrix} 64 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix} + \begin{bmatrix} 4 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 64 & -8 \\ -8 & 1 \end{bmatrix} \right\} \\ &= \frac{1}{3} \begin{bmatrix} 136 & -6 \\ -6 & 2 \end{bmatrix} \cong \begin{bmatrix} 45,33 & -2 \\ -2 & 0,67 \end{bmatrix} \end{aligned}$$

c) $r_{11} = r_{22} = 1$, $r_{12} = \frac{-2}{\sqrt{(45,33)(0,67)}} = -0,36$

Logo, $R = \hat{\rho} = \begin{bmatrix} 1 & -0,36 \\ -0,36 & 1 \end{bmatrix}$.

Observação: Suponha que a matriz de dados fosse disposta da seguinte maneira:

$${}_n X_p = \begin{bmatrix} \tilde{X}_1 & \tilde{X}_2 & \cdots & \tilde{X}_p \end{bmatrix}.$$

No exemplo 3, virá: $X = \begin{bmatrix} \tilde{X}_1 & \tilde{X}_2 \end{bmatrix} = \begin{bmatrix} 42 & 4 \\ 52 & 5 \\ 48 & 4 \\ 58 & 3 \end{bmatrix}$.

Assim, $\bar{\tilde{X}} = \frac{1}{n} X' \tilde{u}$, em que \tilde{u} é um vetor de 1's de ordem n.

$$\bar{\tilde{X}} = \frac{1}{4} \begin{bmatrix} 42 & 52 & 48 & 58 \\ 4 & 5 & 4 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 200 \\ 16 \end{bmatrix} = \begin{bmatrix} 50 \\ 4 \end{bmatrix}.$$

$$S = \frac{1}{n-1} X' \begin{bmatrix} I_{(n)} & -\frac{1}{n} \tilde{u} \tilde{u}' \end{bmatrix} X = \frac{1}{n-1} \left(X'X - \frac{1}{n} X' \tilde{u} \tilde{u}' X \right)$$

$$\begin{aligned}
&= \frac{1}{4-1} \left\{ \begin{bmatrix} 42 & 52 & 48 & 58 \\ 4 & 5 & 4 & 3 \end{bmatrix} \begin{bmatrix} 42 & 4 \\ 52 & 5 \\ 48 & 4 \\ 58 & 3 \end{bmatrix} - \frac{1}{4} \begin{bmatrix} 200 \\ 16 \end{bmatrix} \begin{bmatrix} 200 & 16 \end{bmatrix} \right\} \\
&= \frac{1}{3} \left\{ \begin{bmatrix} 10136 & 794 \\ 794 & 66 \end{bmatrix} - \begin{bmatrix} 10000 & 800 \\ 800 & 64 \end{bmatrix} \right\} = \frac{1}{3} \begin{bmatrix} 136 & -6 \\ -6 & 2 \end{bmatrix} \cong \begin{bmatrix} 45,33 & -2 \\ -2 & 0,67 \end{bmatrix}
\end{aligned}$$

2.4. EXERCÍCIOS

1. Um estudo envolvendo as variáveis aleatórias X_1, X_2, X_3 e X_4 , para $n=9$ observações por variável, obteve-se os resultados apresentados na Tabela 2.2.

Tabela 2.2 – Resultados experimentais

Observação nº	Variáveis			
	X_1	X_2	X_3	X_4
1	46,2	11,2	32,9	157,1
2	48,6	12,6	13,2	174,4
3	42,6	10,6	28,7	160,8
4	39,0	10,4	26,1	162,0
5	38,7	9,3	30,1	140,8
6	44,5	10,8	18,5	164,6
7	39,1	12,7	24,3	163,7
8	40,1	10,0	18,6	174,5
9	45,9	12,0	20,4	185,7

Pede-se obter:

- O vetor de médias \bar{X} ;
- A matriz de covariância S ;
- A matriz de correlação R .

2. Considerando a matriz de covariância amostral $S = \begin{bmatrix} 25 & -2 \\ -2 & 4 \end{bmatrix}$, pede-se:

a) Sendo $V^{1/2} = \begin{bmatrix} \sqrt{s_{11}} & 0 & \cdots & 0 \\ 0 & \sqrt{s_{22}} & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \sqrt{s_{pp}} \end{bmatrix}$, obtenha a matriz de correlação amostral que é dada

pela seguinte expressão: $R = (V^{1/2})^{-1} S (V^{1/2})^{-1}$;

b) Verifique a relação $S = V^{1/2} R V^{1/2}$.

2.5. REFERÊNCIAS

GRAYBILL, F. A. **Theory and application of the linear model**. USA: Duxbury Press, 1976. 704p.

HOFFMANN, R.; VIEIRA, S. **Análise de regressão. Uma introdução à econometria**. 4. ed., São Paulo: Editora Ucitec, 2006. 378p.

JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. 4th ed., USA, Upper Saddle River, New Jersey: Prentice Hall, Inc., 1998. 816p.

MARDIA, K. V.; KENT, J. T.; BIBBY, J. M. **Multivariate analysis**. 6th ed., London: Academic Press, 1997. 518p.

MONTGOMERY, D. C.; PECK, E. A. **Introduction to linear regression analysis**. 2nd ed., USA: John Wiley and Sons, 1992. 527p.

RENCHE, A. C. **Methods of multivariate analysis**. 2nd ed., USA: John Wiley and Sons, 2002. 738p.

SAS INSTITUTE INC. **Statistical Analysis System**. Version 9.1, USA: Cary, NC, 2004.

CAPÍTULO 3

DISTRIBUIÇÕES MULTIVARIADAS

3.1. DISTRIBUIÇÃO NORMAL MULTIVARIADA

3.1.1. Definição

Seja a seguinte notação: $\exp(u) = e^u$

$$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_p \end{bmatrix} \quad \text{e} \quad \underline{x}' = [x_1, x_2, \dots, x_p]$$

Sejam X_1, X_2, \dots, X_p , p variáveis aleatórias do tipo contínuo.

O vetor aleatório $\underline{X}' = [X_1, X_2, \dots, X_p]$ tem distribuição normal multivariada ou também chamada distribuição multinormal de probabilidade se a função densidade de probabilidade (f.d.p.) conjunta de X_1, X_2, \dots, X_p for do tipo:

$$f_{X_1, X_2, \dots, X_p}(x_1, x_2, \dots, x_p) = K \exp \left[-\frac{1}{2} \left(\underline{x} - \underline{\mu} \right)' M \left(\underline{x} - \underline{\mu} \right) \right] \quad (1)$$

em que, K é uma constante, $\underline{\mu}' = [\mu_1, \mu_2, \dots, \mu_p]$ é um vetor de constantes e ${}_p M_p$ é uma matriz real e simétrica positiva definida e, $-\infty < x_i < \infty$ para $i = 1, 2, \dots, p$.

• Determinação de K e M

Se (1) é uma f.d.p., então:

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_{X_1, X_2, \dots, X_p}(x_1, x_2, \dots, x_p) dx_1 dx_2 \dots dx_p = 1$$

ou

$$K \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \exp \left[-\frac{1}{2} \left(\underline{x} - \underline{\mu} \right)' M \left(\underline{x} - \underline{\mu} \right) \right] dx_1 dx_2 \dots dx_p = 1 \quad (2)$$

Sendo M uma matriz real e simétrica positiva definida, existe uma matriz B não singular tal que $B'MB = I_p$.

Seja então a transformação não singular (biunívoca):

$$\underline{x} = \underline{\mu} + B \underline{z}$$

O Jacobiano da transformação é $J = \text{módulo } |B|$.

De $B'MB = I \Rightarrow |B'MB| = 1 \Rightarrow |B'| |M| |B| = 1$. Como $|B'| = |B|$, então

$$|B| = \frac{1}{\sqrt{|M|}} \quad (\text{raiz quadrada positiva})$$

Logo, $J = \frac{1}{\sqrt{|M|}}$.

$$\begin{aligned} \text{Sendo } \left(\tilde{x} - \tilde{\mu} \right)' M \left(\tilde{x} - \tilde{\mu} \right) &= \left(B \tilde{z} \right)' M \left(B \tilde{z} \right) = \tilde{z}' B' M B \tilde{z} \\ &= \tilde{z}' \tilde{z} = \sum_{i=1}^p z_i^2, \end{aligned}$$

a expressão (2) pode ser escrita como:

$$\begin{aligned} K \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp \left(-\frac{1}{2} \tilde{z}' \tilde{z} \right) \frac{1}{\sqrt{|M|}} dz_1 dz_2 \dots dz_p &= 1 \\ \frac{K}{\sqrt{|M|}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{-\frac{1}{2} \sum_{i=1}^p z_i^2} \prod_{i=1}^p dz_i &= \frac{K}{\sqrt{|M|}} \prod_{i=1}^p \int_{-\infty}^{\infty} e^{-\frac{1}{2} z_i^2} dz_i = 1 \end{aligned} \quad (3)$$

$$\text{Mas } \int_{-\infty}^{\infty} e^{-\frac{1}{2} t^2} dt = 2 \int_0^{\infty} e^{-\frac{1}{2} t^2} dt.$$

$$\text{Fazendo } v = \frac{t^2}{2} \Rightarrow t = (2v)^{1/2} \therefore dt = \frac{1}{2} (2v)^{-1/2} 2dv$$

$$dt = \frac{\sqrt{2}}{2} v^{-1/2} dv.$$

$$\begin{aligned} \text{Logo, } \int_{-\infty}^{\infty} e^{-\frac{1}{2} t^2} dt &= 2 \int_0^{\infty} \frac{\sqrt{2}}{2} e^{-v} v^{-1/2} dv = \sqrt{2} \int_0^{\infty} e^{-v} v^{-1/2} dv \\ &= \sqrt{2} \Gamma \left(\frac{1}{2} \right) = \sqrt{2} \sqrt{\pi} = \sqrt{2\pi}. \end{aligned}$$

Voltando em (3), virá:

$$\begin{aligned} \frac{K}{\sqrt{|M|}} \prod_{i=1}^p (2\pi)^{1/2} &= 1 \Rightarrow \frac{K}{\sqrt{|M|}} (2\pi)^{p/2} = 1 \\ K &= \frac{\sqrt{|M|}}{(2\pi)^{p/2}}. \end{aligned}$$

Assim, a f.d.p. da distribuição multinormal (p- variada) pode ser escrita como:

$$f_{X_1, X_2, \dots, X_p}(x_1, x_2, \dots, x_p) = \frac{\sqrt{|M|}}{(2\pi)^{p/2}} \exp \left[-\frac{1}{2} \left(\underline{x} - \underline{\mu} \right)' M \left(\underline{x} - \underline{\mu} \right) \right].$$

Por outro lado denotando por Σ a matriz de dispersão do vetor \underline{X} dada por:

$$\Sigma = E \left[\left(\underline{X} - \underline{\mu} \right) \left(\underline{X} - \underline{\mu} \right)' \right] = E \begin{bmatrix} (X_1 - \mu_1)^2 & (X_1 - \mu_1)(X_2 - \mu_2) & \cdots & (X_1 - \mu_1)(X_p - \mu_p) \\ (X_2 - \mu_2)(X_1 - \mu_1) & (X_2 - \mu_2)^2 & \cdots & (X_2 - \mu_2)(X_p - \mu_p) \\ \vdots & \vdots & \cdots & \vdots \\ (X_p - \mu_p)(X_1 - \mu_1) & (X_p - \mu_p)(X_2 - \mu_2) & \cdots & (X_p - \mu_p)^2 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} E[(X_1 - \mu_1)^2] & E[(X_1 - \mu_1)(X_2 - \mu_2)] & \cdots & E[(X_1 - \mu_1)(X_p - \mu_p)] \\ E[(X_2 - \mu_2)(X_1 - \mu_1)] & E[(X_2 - \mu_2)^2] & \cdots & E[(X_2 - \mu_2)(X_p - \mu_p)] \\ \vdots & \vdots & \cdots & \vdots \\ E[(X_p - \mu_p)(X_1 - \mu_1)] & E[(X_p - \mu_p)(X_2 - \mu_2)] & \cdots & E[(X_p - \mu_p)^2] \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) & \cdots & \text{Cov}(X_1, X_p) \\ \text{Cov}(X_2, X_1) & \text{Var}(X_2) & \cdots & \text{Cov}(X_2, X_p) \\ \vdots & \vdots & \cdots & \vdots \\ \text{Cov}(X_p, X_1) & \text{Cov}(X_p, X_2) & \cdots & \text{Var}(X_p) \end{bmatrix},$$

$$\text{virá: } \Sigma = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\underline{x} - \underline{\mu} \right) \left(\underline{x} - \underline{\mu} \right)' \frac{\sqrt{|M|}}{(2\pi)^{p/2}} \exp \left[-\frac{1}{2} \left(\underline{x} - \underline{\mu} \right)' M \left(\underline{x} - \underline{\mu} \right) \right] dx_1 dx_2 \cdots dx_p$$

Com a transformação anterior $\left(\underline{x} - \underline{\mu} \right) = B \underline{z}$, com B tal que $B' M B = I$ e

$$J = \text{mod}|B| = \frac{1}{\sqrt{|M|}}, \text{ virá:}$$

$$\Sigma = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} B \underline{z} \underline{z}' B' \frac{\sqrt{|M|}}{(2\pi)^{p/2}} \exp \left(-\frac{1}{2} \underline{z}' B' M B \underline{z} \right) \frac{1}{\sqrt{|M|}} dz_1 \cdots dz_p$$

$$\Sigma = \frac{1}{(2\pi)^{p/2}} B \left[\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \underline{z} \underline{z}' \exp \left(-\frac{1}{2} \underline{z}' \underline{z} \right) dz_1 \cdots dz_p \right] B'$$

$$\Sigma = \frac{1}{(2\pi)^{p/2}} B \left\{ \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \begin{bmatrix} z_1^2 & \cdots & z_1 z_p \\ \vdots & \cdots & \vdots \\ z_1 z_p & \cdots & z_p^2 \end{bmatrix} e^{-\frac{1}{2} \sum_{i=1}^p z_i^2} dz_1 \cdots dz_p \right\} B'$$

Sendo a integral de uma matriz a matriz das integrais de seus elementos, podem-se considerar os dois casos possíveis:

a) Elementos da diagonal principal da matriz $\underline{z} \underline{z}'$

$$I_1 = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} z_i^2 e^{-\frac{1}{2}(z_1^2 + z_2^2 + \cdots + z_i^2 + \cdots + z_p^2)} dz_1 \cdots dz_i \cdots dz_p$$

$$I_1 = \int_{-\infty}^{\infty} e^{-\frac{1}{2}z_1^2} dz_1 \cdots \int_{-\infty}^{\infty} z_i^2 e^{-\frac{1}{2}z_i^2} dz_i \cdots \int_{-\infty}^{\infty} e^{-\frac{1}{2}z_p^2} dz_p \quad (4)$$

Sabe-se que $\int_{-\infty}^{\infty} e^{-\frac{1}{2}t^2} dt = \sqrt{2\pi}$.

Por outro lado $\int_{-\infty}^{\infty} t^2 e^{-\frac{1}{2}t^2} dt = 2 \int_0^{\infty} t^2 e^{-\frac{1}{2}t^2} dt$

Fazendo, $v = \frac{t^2}{2} \Rightarrow t = \sqrt{2v} \Rightarrow dt = \frac{\sqrt{2}}{2} \cdot \frac{1}{\sqrt{v}} dv$, virá:

$$\int_{-\infty}^{\infty} t^2 e^{-\frac{1}{2}t^2} dt = 2 \int_0^{\infty} 2v \cdot e^{-v} \frac{\sqrt{2}}{2} \cdot \frac{1}{\sqrt{v}} dv$$

$$= 2\sqrt{2} \int_0^{\infty} e^{-v} v^{1/2} dv = 2\sqrt{2} \Gamma\left(\frac{3}{2}\right) = 2\sqrt{2} \cdot \frac{\sqrt{\pi}}{2} = \sqrt{2\pi}$$

Voltando em (4), virá:

$$I_1 = \prod_{i=1}^p (2\pi)^{1/2} = (2\pi)^{p/2}$$

b) Elementos fora da diagonal principal da matriz $\mathbf{z} \mathbf{z}'$

$$I_2 = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} z_i z_j e^{-\frac{1}{2}(z_1^2 + \cdots + z_i^2 + \cdots + z_j^2 + \cdots + z_p^2)} dz_1 \cdots dz_i \cdots dz_j \cdots dz_p$$

$$I_2 = \int_{-\infty}^{\infty} e^{-\frac{1}{2}z_1^2} dz_1 \cdots \int_{-\infty}^{\infty} z_i e^{-\frac{1}{2}z_i^2} dz_i \cdots \int_{-\infty}^{\infty} z_j e^{-\frac{1}{2}z_j^2} dz_j \cdots \int_{-\infty}^{\infty} e^{-\frac{1}{2}z_p^2} dz_p$$

Mas $\int_{-\infty}^{\infty} e^{-\frac{1}{2}t^2} dt = \sqrt{2\pi}$

$$e \int_{-\infty}^{\infty} t e^{-\frac{1}{2}t^2} dt = \int_{-\infty}^0 t e^{-\frac{1}{2}t^2} dt + \int_0^{\infty} t e^{-\frac{1}{2}t^2} dt$$

$$= -\int_0^{\infty} t e^{-\frac{1}{2}t^2} dt + \int_0^{\infty} t e^{-\frac{1}{2}t^2} dt = 0$$

Logo, $I_2 = 0$

Assim, \sum pode ser escrita como:

$$\sum = \frac{1}{(2\pi)^{p/2}} \mathbf{B} (2\pi)^{p/2} \mathbf{I}_p \mathbf{B}' = \mathbf{B} \mathbf{B}' \therefore \sum = \mathbf{B} \mathbf{B}'$$

De $\mathbf{B}' \mathbf{M} \mathbf{B} = \mathbf{I} \xRightarrow{\exists \mathbf{B}^{-1}} \mathbf{M} = (\mathbf{B}')^{-1} \mathbf{B}^{-1} = (\mathbf{B} \mathbf{B}')^{-1} \therefore \mathbf{M}^{-1} = \mathbf{B} \mathbf{B}'$

Logo, $\sum = \mathbf{M}^{-1} \Leftrightarrow \mathbf{M} = \sum^{-1}$

$$|M| = |\Sigma^{-1}| = \frac{1}{|\Sigma|}$$

Assim, a matriz da forma quadrática $\left(\underline{\tilde{x}} - \underline{\tilde{\mu}}\right)' M \left(\underline{\tilde{x}} - \underline{\tilde{\mu}}\right)$ corresponde à recíproca da matriz de dispersão do vetor $\underline{\tilde{X}}$. Pode-se então escrever:

$$f_{X_1, X_2, \dots, X_p}(x_1, x_2, \dots, x_p) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp \left[-\frac{1}{2} \left(\underline{\tilde{x}} - \underline{\tilde{\mu}}\right)' \Sigma^{-1} \left(\underline{\tilde{x}} - \underline{\tilde{\mu}}\right) \right]$$

3.1.2. Função geradora de momentos

Por definição, a função geradora de momentos (f.g.m.) do vetor aleatório $\underline{\tilde{X}}$ pode ser escrita como:

$$M_{X_1, X_2, \dots, X_p}(t_1, t_2, \dots, t_p) = E\left(e^{t_1 X_1 + t_2 X_2 + \dots + t_p X_p}\right)$$

$$\text{com } \underline{t}' = \begin{bmatrix} t_1 & t_2 & \dots & t_p \end{bmatrix} \text{ e } -h_i < t_i < h_i.$$

$$\text{Ou ainda, } M_{\underline{\tilde{X}}}(\underline{t}') = E\left(e^{\underline{t}' \underline{\tilde{X}}}\right).$$

Logo,

$$M_{\underline{\tilde{X}}}(\underline{t}') = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{\underline{t}' \underline{\tilde{x}}} \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} e^{-\frac{1}{2} \left(\underline{\tilde{x}} - \underline{\tilde{\mu}}\right)' \Sigma^{-1} \left(\underline{\tilde{x}} - \underline{\tilde{\mu}}\right)} dx_1 dx_2 \dots dx_p$$

$$M_{\underline{\tilde{X}}}(\underline{t}') = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{\underline{t}' \underline{\tilde{x}} - \frac{1}{2} \left(\underline{\tilde{x}} - \underline{\tilde{\mu}}\right)' \Sigma^{-1} \left(\underline{\tilde{x}} - \underline{\tilde{\mu}}\right)} dx_1 dx_2 \dots dx_p$$

Seja a transformação: $\underline{\tilde{x}} = \underline{y} + \underline{\tilde{x}}_0$ com $\underline{\tilde{x}}_0$ tal que anule $d\phi(\underline{\tilde{x}})$, em que:

$$\phi(\underline{\tilde{x}}) = \underline{\tilde{x}}' \underline{t} - \frac{1}{2} \left(\underline{\tilde{x}} - \underline{\tilde{\mu}}\right)' \Sigma^{-1} \left(\underline{\tilde{x}} - \underline{\tilde{\mu}}\right)$$

$$d\phi(\underline{\tilde{x}}) = d\underline{\tilde{x}}' \underline{t} - \frac{1}{2} d\left(\underline{\tilde{x}}' \Sigma^{-1} \underline{\tilde{x}} - \underline{\tilde{x}}' \Sigma^{-1} \underline{\tilde{\mu}} - \underline{\tilde{\mu}}' \Sigma^{-1} \underline{\tilde{x}} + \underline{\tilde{\mu}}' \Sigma^{-1} \underline{\tilde{\mu}}\right)$$

$$d\phi(\underline{\tilde{x}}) = d\underline{\tilde{x}}' \underline{t} - \frac{1}{2} \left(d\underline{\tilde{x}}' \Sigma^{-1} \underline{\tilde{x}} + \underline{\tilde{x}}' \Sigma^{-1} d\underline{\tilde{x}} - d\underline{\tilde{x}}' \Sigma^{-1} \underline{\tilde{\mu}} - \underline{\tilde{\mu}}' \Sigma^{-1} d\underline{\tilde{x}}\right)$$

$$d\phi(\underline{\tilde{x}}) = d\underline{\tilde{x}}' \left(\underline{t} - \Sigma^{-1} \underline{\tilde{x}} + \Sigma^{-1} \underline{\tilde{\mu}}\right)$$

$$d\phi(\underline{\tilde{x}}) = d\underline{\tilde{x}}' \left[\underline{t} - \Sigma^{-1} \left(\underline{\tilde{x}} - \underline{\tilde{\mu}}\right)\right]$$

$$d\varphi(\underline{x}) = 0 \Rightarrow \underline{t} - \sum^{-1} (\underline{x}_0 - \underline{\mu}) = 0 \Rightarrow \underline{x}_0 = \sum \underline{t} + \underline{\mu}$$

Desenvolvendo $\varphi(\underline{x})$ em série de Taylor em torno do ponto \underline{x}_0 , virá:

$$\varphi(\underline{x}) = \varphi(\underline{x}_0) + (\underline{x} - \underline{x}_0)' \varphi'(\underline{x}_0) + \frac{1}{2!} (\underline{x} - \underline{x}_0)' \varphi''(\underline{x}_0) (\underline{x} - \underline{x}_0) \quad (5)$$

$$\varphi(\underline{x}_0) = \left(\underline{t}' \sum + \underline{\mu}' \right) \underline{t} - \frac{1}{2} (\underline{t}' \sum \sum^{-1} \sum \underline{t})$$

$$\varphi(\underline{x}_0) = \underline{t}' \sum \underline{t} + \underline{\mu}' \underline{t} - \frac{1}{2} \underline{t}' \sum \underline{t}$$

$$\varphi(\underline{x}_0) = \underline{t}' \underline{\mu} + \frac{1}{2} \underline{t}' \sum \underline{t}$$

Por outro lado:

$$\varphi''(\underline{x}) = -\sum^{-1} \therefore \varphi''(\underline{x}_0) = -\sum^{-1}$$

Logo, sendo $\underline{x} - \underline{x}_0 = \underline{y}$ e $d\underline{x} = d\underline{y}$, $J = 1$, de (5) virá:

$$\varphi(\underline{x}) = \underline{t}' \underline{\mu} + \frac{1}{2} \underline{t}' \sum \underline{t} - \frac{1}{2} \underline{y}' \sum^{-1} \underline{y}$$

e

$$M_{\underline{x}'}(\underline{t}') = \frac{1}{(2\pi)^{p/2} |\sum|^{1/2}} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{\underline{t}' \underline{\mu} + \frac{1}{2} \underline{t}' \sum \underline{t} - \frac{1}{2} \underline{y}' \sum^{-1} \underline{y}} dy_1 \dots dy_p$$

$$M_{\underline{x}'}(\underline{t}') = e^{\underline{t}' \underline{\mu} + \frac{1}{2} \underline{t}' \sum \underline{t}} \underbrace{\frac{1}{(2\pi)^{p/2} |\sum|^{1/2}} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{-\frac{1}{2} \underline{y}' \sum^{-1} \underline{y}} dy_1 \dots dy_p}_{=1, \text{ pois é a f.d.p. de } \underline{Y} \sim N_p\left(\underline{0}; \sum\right)}$$

$$\text{Logo, } M_{\underline{x}'}(\underline{t}') = e^{\underline{t}' \underline{\mu} + \frac{1}{2} \underline{t}' \sum \underline{t}}.$$

• Determinação do primeiro e segundo momento e matriz de variâncias e covariâncias

$$\text{Primeiro momento: } E(\underline{X}) = \left. \frac{dM_{\underline{x}'}(\underline{t}')}{d\underline{t}} \right|_{\underline{t}=\underline{0}}$$

$$E(\tilde{X}) = e^{\tilde{t}'\tilde{\mu} + \frac{1}{2}\tilde{t}'\Sigma\tilde{t}} \cdot \left(\tilde{\mu} + \frac{1}{2}2\sum\tilde{t} \right) \Bigg|_{\tilde{t}=0}$$

$E(\tilde{X}) = \tilde{\mu}$, isto é, o vetor $\tilde{\mu}$ é a média do vetor aleatório \tilde{X} .

Segundo momento:
$$E(\tilde{X}\tilde{X}') = \frac{d}{d\tilde{t}} \left[\frac{dM_{\tilde{X}}(\tilde{t})}{d\tilde{t}'} \right] \Bigg|_{\tilde{t}=0}$$

$$E(\tilde{X}\tilde{X}') = \frac{d}{d\tilde{t}} \left[e^{\tilde{t}'\tilde{\mu} + \frac{1}{2}\tilde{t}'\Sigma\tilde{t}} \left(\tilde{\mu}' + \frac{1}{2} \cdot 2\tilde{t}'\Sigma \right) \right] \Bigg|_{\tilde{t}=0}$$

$$E(\tilde{X}\tilde{X}') = e^{\tilde{t}'\tilde{\mu} + \frac{1}{2}\tilde{t}'\Sigma\tilde{t}} \cdot \Sigma + e^{\tilde{t}'\tilde{\mu} + \frac{1}{2}\tilde{t}'\Sigma\tilde{t}} \left(\tilde{\mu} + \sum\tilde{t} \right) \left(\tilde{\mu} + \sum\tilde{t} \right)' \Bigg|_{\tilde{t}=0}$$

$$E(\tilde{X}\tilde{X}') = \Sigma + \tilde{\mu}\tilde{\mu}'$$

Matriz de covariância (variâncias e covariâncias $[D(\tilde{X})]$)

$$D(\tilde{X}) = E(\tilde{X}\tilde{X}') - \tilde{\mu}\tilde{\mu}'$$

$$D(\tilde{X}) = \Sigma + \tilde{\mu}\tilde{\mu}' - \tilde{\mu}\tilde{\mu}'$$

$$D(\tilde{X}) = \Sigma$$

• Dois teoremas importantes sobre a distribuição normal multivariada

Teorema 1: Se $\tilde{X} \sim N_p(\tilde{\mu}, \Sigma)$ e se $\tilde{Y} = {}_q C' {}_p \tilde{X}$

$$\text{Com Posto}(C') = q \leq p, \text{ então } \tilde{Y} \sim N_q(C'\tilde{\mu}; C'\Sigma C)$$

Prova: Em geral (não só para a multinormal) segue que:

$$E(\tilde{Y}) = E(C'\tilde{X}) = C'E(\tilde{X}) = C'\tilde{\mu}$$

$$D(\tilde{Y}) = D(C'\tilde{X}) = C'D(\tilde{X})C = C'\Sigma C$$

Por outro lado, pela f.g.m., virá:

$$M_{\tilde{Y}'=(C'\tilde{X})'}(\tilde{t}') = E\left(e^{\tilde{t}'\tilde{Y}}\right) = E\left(e^{\tilde{t}'C'\tilde{X}}\right) = E\left[e^{\tilde{\tau}'\tilde{X}}\right] = M_{\tilde{X}}(\tilde{\tau}')$$

Logo,

$$M_{\tilde{Y}'}(\tilde{t}') = e^{\tilde{t}'\tilde{\mu} + \frac{1}{2}\tilde{t}'\tilde{\Sigma}\tilde{t}'} = e^{\tilde{t}'(C'\tilde{\mu}) + \frac{1}{2}\tilde{t}'(C'\tilde{\Sigma}C)\tilde{t}'}, \text{ e desde que a f.g.m. caracteriza a distribuição,}$$

segue que $\tilde{Y} = C'\tilde{X} \sim N_q(C'\tilde{\mu}; C'\tilde{\Sigma}C)$.

Teorema 2: Seja $\tilde{X} \sim N_p(\tilde{\mu}; \tilde{\Sigma})$ e a partição

$$\tilde{X} = \begin{bmatrix} \tilde{X}_1 \\ \tilde{X}_2 \end{bmatrix}; \tilde{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}; \tilde{\Sigma} = \begin{bmatrix} \sum_{11} & \sum_{12} \\ \sum_{12}' & \sum_{22} \end{bmatrix}; q^{(x_1)}_1; p-q^{(x_2)}_1$$

A condição necessária e suficiente para que \tilde{X}_1 e \tilde{X}_2 sejam independentes e normalmente distribuídos é que $\sum_{12} = \phi$.

Prova:

(\Rightarrow) Seja $\sum_{12} = \phi$ e a f.g.m.

$$M_{\tilde{X}'}(\tilde{t}') = e^{\tilde{t}'\tilde{\mu} + \frac{1}{2}\tilde{t}'\tilde{\Sigma}\tilde{t}'}, \text{ com } \tilde{t}' = \begin{bmatrix} t_1 \\ t_2 \end{bmatrix} \text{ e } \tilde{\Sigma} = \begin{bmatrix} \sum_{11} & \phi \\ \phi & \sum_{22} \end{bmatrix}$$

Assim,

$$\begin{aligned} M_{\tilde{X}'}(\tilde{t}') &= \exp \left[\tilde{t}'_1 \mu_1 + \tilde{t}'_2 \mu_2 + \frac{1}{2} \left(\tilde{t}'_1 \sum_{11} \tilde{t}_1 + \tilde{t}'_2 \sum_{22} \tilde{t}_2 \right) \right] \\ M_{\tilde{X}'}(\tilde{t}') &= \exp \left(\tilde{t}'_1 \mu_1 + \frac{1}{2} \tilde{t}'_1 \sum_{11} \tilde{t}_1 \right) \exp \left(\tilde{t}'_2 \mu_2 + \frac{1}{2} \tilde{t}'_2 \sum_{22} \tilde{t}_2 \right) \\ \therefore M_{\tilde{X}'}(\tilde{t}') &= M_{\tilde{X}'_1}(\tilde{t}'_1) \cdot M_{\tilde{X}'_2}(\tilde{t}'_2) \end{aligned}$$

Desde que a f.g.m. quando existe, caracteriza a distribuição, segue que:

$$\tilde{X}_i \sim N_{q_i}(\mu_i; \sum_{ii}); i = 1, 2, \text{ com } q_1 = q \text{ e } q_2 = p - q$$

Por outro lado, desde que a distribuição conjunta de \tilde{X}_1 e \tilde{X}_2 pode ser fatorada no produto das marginais, temos que \tilde{X}_1 e \tilde{X}_2 são independentes.

(\Leftarrow) Se \tilde{X}_1 e \tilde{X}_2 são independentes é imediato que $\sum_{12} = \phi$.

3.1.3 Distribuição marginal e condicional

Pelo Teorema 1, foi visto que a distribuição marginal de qualquer subconjunto de variáveis X_1, \dots, X_r ($r \leq p$) tem distribuição normal r-variada, com vetor média $(\mu_1, \dots, \mu_r)'$ e matriz de covariância obtida tomando-se de Σ os elementos correspondentes.

Considere inicialmente um Teorema de grande utilidade nesse ponto.

Teorema 3: Seja Σ simétrica, positiva definida e a partição $\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$.

Seja $\Sigma^{-1} = M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}$. Então, $\Sigma_{11}^{-1} = M_{11} - M_{12}M_{22}^{-1}M_{21}$.

Prova:

$$\Sigma M = I \Rightarrow \begin{cases} \Sigma_{11}M_{11} + \Sigma_{12}M_{21} = I & (a) \\ \Sigma_{11}M_{12} + \Sigma_{12}M_{22} = 0 & (b) \end{cases}$$

De (b), vem:

$\Sigma_{12} = -\Sigma_{11}M_{12}M_{22}^{-1}$, que substituindo em (a), virá:

$$\Sigma_{11}M_{11} - \Sigma_{11}M_{12}M_{22}^{-1}M_{21} = I$$

$$\Sigma_{11}(M_{11} - M_{12}M_{22}^{-1}M_{21}) = I$$

$$\Sigma_{11}^{-1} = M_{11} - M_{12}M_{22}^{-1}M_{21}$$

Da mesma forma demonstra-se que $\Sigma_{22}^{-1} = M_{22} - M_{21}M_{11}^{-1}M_{12}$.

Seja agora, $\tilde{X} \sim N_p(\mu, \Sigma)$ e a seguinte partição:

$$\tilde{X} = \begin{bmatrix} \tilde{X}_r \\ \tilde{X}_s \end{bmatrix} \text{ (com } r+s=p); \mu = \begin{bmatrix} \mu_r \\ \mu_s \end{bmatrix}; \Sigma = \begin{bmatrix} \Sigma_{rr} & \Sigma_{rs} \\ \Sigma_{sr} & \Sigma_{ss} \end{bmatrix} \text{ e } \Sigma^{-1} = M = \begin{bmatrix} M_{rr} & M_{rs} \\ M_{sr} & M_{ss} \end{bmatrix}$$

A distribuição marginal do vetor \tilde{X}_r dada pelo Teorema 1, será:

$$\tilde{X}_r \sim N_r(\mu_r; \Sigma_{rr}),$$

$$\text{com f.d.p. } f_{\tilde{X}_r}(\tilde{x}_r) = \frac{\sqrt{|\Sigma_{rr}^{-1}|}}{(2\pi)^{r/2}} \exp \left[-\frac{1}{2}(\tilde{x}_r - \mu_r)' \Sigma_{rr}^{-1}(\tilde{x}_r - \mu_r) \right].$$

Pelo Teorema 3, foi visto que: $\Sigma_{rr}^{-1} = M_{rr} - M_{rs}M_{ss}^{-1}M_{sr}$.

Logo,

$$f_{\tilde{X}_r}(\tilde{x}_r) = \frac{\sqrt{|\mathbf{M}_{rr} - \mathbf{M}_{rs}\mathbf{M}_{ss}^{-1}\mathbf{M}_{sr}|}}{(2\pi)^{r/2}} \exp \left[-\frac{1}{2} \left(\tilde{x}_r - \tilde{\mu}_r \right)' \left(\mathbf{M}_{rr} - \mathbf{M}_{rs}\mathbf{M}_{ss}^{-1}\mathbf{M}_{sr} \right) \left(\tilde{x}_r - \tilde{\mu}_r \right) \right]$$

Para obtenção da distribuição condicional de \tilde{X}_r dado $\tilde{X}_s = \tilde{x}_s$, considere a transformação a seguir:

$$\begin{cases} \tilde{Y}_r = \tilde{X}_r + \mathbf{M}\tilde{X}_s \\ \tilde{Y}_s = \tilde{X}_s \end{cases} \Rightarrow \begin{bmatrix} \tilde{Y}_r \\ \tilde{Y}_s \end{bmatrix} = \begin{bmatrix} \mathbf{I}_r & \mathbf{M} \\ \phi & \mathbf{I}_s \end{bmatrix} \begin{bmatrix} \tilde{X}_r \\ \tilde{X}_s \end{bmatrix},$$

com \mathbf{M} tal que \tilde{Y}_r e \tilde{Y}_s sejam independentes.

Pelo Teorema 1, pode-se escrever:

$$D(\tilde{Y}) = \begin{bmatrix} \mathbf{I}_r & \mathbf{M} \\ \phi & \mathbf{I}_s \end{bmatrix} \begin{bmatrix} \sum_{rr} & \sum_{rs} \\ \sum_{sr} & \sum_{ss} \end{bmatrix} \begin{bmatrix} \mathbf{I}_r & \phi \\ \mathbf{M}' & \mathbf{I}_s \end{bmatrix}.$$

Pelo Teorema 2, segue que:

$$\sum_{rs} + \mathbf{M}\sum_{ss} = \phi \Rightarrow \mathbf{M} = -\sum_{rs} \sum_{ss}^{-1}.$$

Assim,

$$E(\tilde{Y}) = E \begin{bmatrix} \tilde{Y}_r \\ \tilde{Y}_s \end{bmatrix} = \begin{bmatrix} \tilde{\mu}_r - \sum_{rs} \sum_{ss}^{-1} \tilde{\mu}_s \\ \tilde{\mu}_s \end{bmatrix} = \begin{bmatrix} \tilde{\mu}_r^* \\ \tilde{\mu}_s \end{bmatrix}$$

$$D(\tilde{Y}) = \begin{bmatrix} \sum_{rr} - \sum_{rs} \sum_{ss}^{-1} \sum_{sr} & \phi \\ \phi & \sum_{ss} \end{bmatrix} = \begin{bmatrix} \sum_{rr-s} & \phi \\ \phi & \sum_{ss} \end{bmatrix}.$$

Logo,

$$g_{\tilde{Y}_r, \tilde{Y}_s}(\tilde{y}_r, \tilde{y}_s) = N_r(\tilde{y}_r; \tilde{\mu}_r^*; \sum_{rr-s}) \cdot N_s(\tilde{y}_s; \tilde{\mu}_s; \sum_{ss})$$

e

$$f_{\tilde{X}_r, \tilde{X}_s}(\tilde{x}_r, \tilde{x}_s) = g_{\tilde{Y}_r, \tilde{Y}_s}(\tilde{x}_r - \sum_{rs} \sum_{ss}^{-1} \tilde{x}_s; \tilde{x}_s) |J|, \text{ em que, } |J| = 1.$$

Logo,

$$f_{\underline{x}_r, \underline{x}_s} = N_r \left(\underline{x}_r - \sum_{rs} \sum_{ss}^{-1} \underline{x}_s : \underline{\mu}_r^* ; \sum_{rr \cdot s} \right) \cdot N_s \left(\underline{x}_s : \underline{\mu}_s ; \sum_{ss} \right)$$

$$f_{\underline{x}_r, \underline{x}_s} = f_{\underline{x}_r / \underline{x}_s}(\underline{x}_r, \underline{x}_s) \cdot f_{\underline{x}_s}(\underline{x}_s).$$

Assim,

$$f_{\underline{x}_r / \underline{x}_s} \left(\underline{x}_r / \underline{x}_s \right) = \frac{f_{\underline{x}_r, \underline{x}_s}(\underline{x}_r, \underline{x}_s)}{f_{\underline{x}_r}(\underline{x}_s)} = N_r \left(\underline{x}_r - \sum_{rs} \sum_{ss}^{-1} \underline{x}_s : \underline{\mu}_r^* ; \sum_{rr \cdot s} \right)$$

Note que:

$$\begin{aligned} \underline{x}_r - \sum_{rs} \sum_{ss}^{-1} \underline{x}_s - \underline{\mu}_r^* &= \underline{x}_r - \sum_{rs} \sum_{ss}^{-1} \underline{x}_s - \underline{\mu}_r + \sum_{rs} \sum_{ss}^{-1} \underline{\mu}_s \\ &= \underline{x}_r - \left[\underline{\mu}_r + \sum_{rs} \sum_{ss}^{-1} \left(\underline{x}_s - \underline{\mu}_s \right) \right] \\ &= \underline{x}_r - \underline{\mu}_{r \cdot s} \end{aligned}$$

Logo,

$$\underline{x}_r / \underline{x}_s \sim N_r \left[\underline{x}_r : \underline{\mu}_{r \cdot s} ; \sum_{rr \cdot s} \right],$$

$$\text{com f.d.p. } f_{\underline{x}_r / \underline{x}_s} \left(\underline{x}_r / \underline{x}_s \right) = \frac{1}{(2\pi)^{r/2} \left| \sum_{rr \cdot s} \right|^{1/2}} \cdot \exp \left[-\frac{1}{2} \left(\underline{x}_r - \underline{\mu}_{r \cdot s} \right)' \sum_{rr \cdot s}^{-1} \left(\underline{x}_r - \underline{\mu}_{r \cdot s} \right) \right]$$

3.1.4. Correlações parciais

Sejam p variáveis aleatórias X_1, X_2, \dots, X_p . A matriz de correlação simples entre essas variáveis é dada por:

$$\rho = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{21} & 1 & \cdots & \rho_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ \rho_{p1} & \rho_{p2} & \cdots & 1 \end{bmatrix},$$

em que $\rho_{ij} = \rho_{ji}$, isto é ρ é simétrica.

O termo correlação parcial designa a correlação entre duas variáveis quaisquer, quando os efeitos de outras variáveis forem mantidos fixos. Em certos casos o coeficiente de correlação simples pode produzir impressionantes equívocos a respeito da relação que há entre duas variáveis. Isto porque, correlação não implica em relação de causa e efeito. Assim, um alto (ou

baixo) coeficiente de correlação entre duas variáveis pode ser o resultado do efeito que sobre essas duas variáveis tem uma terceira variável, ou um grupo de outras variáveis.

Sejam $r + s$ variáveis aleatórias

$$\underbrace{X_1 \ X_2 \ \cdots \ X_r}_{1^\circ \text{ conjunto}} \quad \underbrace{X_{r+1} \ X_{r+2} \ \cdots \ X_s}_{2^\circ \text{ conjunto}} \Rightarrow \Sigma = \begin{bmatrix} \sum_{rr} & \cdots & \sum_{rs} \\ \vdots & & \vdots \\ \sum_{sr} & \cdots & \sum_{ss} \end{bmatrix}.$$

A matriz de variâncias e covariâncias condicionais das variáveis do primeiro conjunto dado as variáveis do segundo conjunto é dada por $\sum_{rr \cdot s} = \sum_{rr} - \sum_{rs} \sum_{ss}^{-1} \sum_{sr}$, sendo uma matriz simétrica de ordem r . O ij -ésimo elemento de $\sum_{rr \cdot s}$ pode ser denotado por $\sigma_{ij \cdot r+1, r+2, \dots, r+s}$.

A correlação parcial entre a i -ésima e j -ésima variáveis do primeiro conjunto, com todas as variáveis do segundo conjunto mantidas constantes é dada por:

$$\rho_{ij \cdot r+1, r+2, \dots, r+s} = \frac{\sigma_{ij \cdot r+1, r+2, \dots, r+s}}{\sqrt{\sigma_{ii \cdot r+1, r+2, \dots, r+s} \cdot \sigma_{jj \cdot r+1, r+2, \dots, r+s}}}.$$

É importante observar que:

(i) No cálculo de cada correlação parcial individualmente, temos que:

- Para obter $\rho_{12 \cdot 3} \Rightarrow \Sigma$ é de ordem 3 ($r + s = 3$).
- Para obter $\rho_{13 \cdot 46} \Rightarrow \Sigma$ é de ordem 4 ($r + s = 4$).

(ii) Para obtenção simultânea de $\rho_{12 \cdot 45}$, $\rho_{13 \cdot 45}$ e $\rho_{23 \cdot 45}$, por exemplo, pode-se usar Σ de ordem 5, com as variáveis X_1 , X_2 e X_3 no primeiro conjunto ($r = 3$) e, X_4 e X_5 no segundo conjunto ($s = 2$).

(iii) Para $D = \text{diag}(\sum_{rr \cdot s})$, tem-se que a matriz de correlações parciais pode ser escrita como:

$\rho_{\text{parc}} = D^{-1/2} (\sum_{rr \cdot s}) D^{-1/2}$, em que a potência $-1/2$ de D indica que a matriz consiste da raiz quadrada do inverso dos elementos de D .

(iv) **Notações:**

Parâmetros	Σ	σ	ρ	ρ_{ij}	D
Estimadores	S	$\hat{\sigma} = s$	$\hat{\rho} = R$	$\hat{\rho}_{ij} = r_{ij}$	\hat{D}

Exemplo: Considerando os dados experimentais apresentados na Tabela 3.1, pede-se calcular:

- A matriz de variâncias e covariâncias com 8 graus de liberdade;
- A matriz de correlação;
- As correlações parciais: $r_{12 \cdot 3}$, $r_{12 \cdot 34}$, $r_{12 \cdot 4}$, $r_{13 \cdot 4}$ e $r_{23 \cdot 4}$.

Tabela 3.1 – Resultados experimentais obtidos para quatro variáveis aleatórias e $n = 9$ observações

Observações	Variáveis			
	X_1	X_2	X_3	X_4
1	42,2	11,2	31,9	167,1
2	48,6	10,6	13,2	174,4
3	42,6	10,6	28,7	160,8
4	39,0	10,4	26,1	162,0
5	34,7	9,3	30,1	140,8
6	44,5	10,8	8,5	174,6
7	39,1	10,7	24,3	163,7
8	40,1	10,0	18,6	174,5
9	45,9	12,0	20,4	185,7

Solução:

$$a) \quad S = \begin{bmatrix} 17,442778 & 2,131111 & -20,395139 & 42,342083 \\ 2,131111 & 0,556944 & -1,020556 & 7,247083 \\ -20,395139 & -1,020556 & 62,851944 & -63,031667 \\ 42,342083 & 7,247083 & -63,031667 & 159,900000 \end{bmatrix}$$

$$b) \quad \hat{\rho} = R = \begin{bmatrix} 1 & 0,68374 & -0,61597 & 0,80175 \\ 0,68374 & 1 & -0,17249 & 0,76795 \\ -0,61597 & -0,17249 & 1 & -0,62875 \\ 0,80175 & 0,76795 & -0,62875 & 1 \end{bmatrix}$$

c) Cálculo de $r_{12.3} \rightarrow$ Variáveis: X_1 e X_2 ($r = 2$); X_3 ($s = 1$).

$$S = \begin{bmatrix} S_{rr} & S_{rs} \\ S_{sr} & S_{ss} \end{bmatrix} = \begin{bmatrix} 17,442778 & 2,131111 & -20,395139 \\ 2,131111 & 0,556944 & -1,020556 \\ -20,395139 & -1,020556 & 62,851944 \end{bmatrix}$$

$$S_{rr \cdot s} = S_{rr} - S_{rs} S_{ss}^{-1} S_{sr}$$

$$S_{rr \cdot s} = \begin{bmatrix} 17,442778 & 2,131111 \\ 2,131111 & 0,556944 \end{bmatrix} - \begin{bmatrix} -20,395139 \\ -1,020556 \end{bmatrix} \frac{1}{62,851944} \begin{bmatrix} -20,395139 & -1,020556 \end{bmatrix}$$

$$S_{rr \cdot s} = \begin{bmatrix} 17,442778 & 2,131111 \\ 2,131111 & 0,556944 \end{bmatrix} - \begin{bmatrix} 6,618120 & 0,331165 \\ 0,331165 & 0,016571 \end{bmatrix} = \begin{bmatrix} 10,824658 & 1,799946 \\ 1,799946 & 0,540373 \end{bmatrix}$$

Assim,

$$r_{12.3} = \hat{p}_{12.3} = \frac{s_{12.3}}{\sqrt{s_{11.3} \cdot s_{22.3}}} = \frac{1,799946}{\sqrt{10,824658 \cdot 0,540373}} = 0,7442 ,$$

ou ainda,

$$\hat{D} = \text{diag} (S_{rr \cdot s}) = \begin{bmatrix} 10,824658 & 0 \\ 0 & 0,540373 \end{bmatrix}$$

$$\hat{\rho}_{\text{Parc}} = \hat{D}^{-1/2} (S_{rr \cdot s}) \hat{D}^{-1/2}$$

$$\begin{aligned} \hat{\rho}_{\text{Parc}} &= \begin{bmatrix} \sqrt{1/10,824658} & 0 \\ 0 & \sqrt{1/0,540373} \end{bmatrix} \begin{bmatrix} 10,824658 & 1,799946 \\ 1,799946 & 0,540373 \end{bmatrix} \begin{bmatrix} \sqrt{1/10,824658} & 0 \\ 0 & \sqrt{1/0,540373} \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0,7442 \\ 0,7442 & 1 \end{bmatrix} \end{aligned}$$

Logo,

$$r_{12.3} = \hat{\rho}_{12.3} = 0,7442 .$$

Cálculo de $r_{12.34}$

Com procedimento análogo ao anterior, obtém-se:

$$r_{12.34} = \hat{\rho}_{12.34} = \frac{s_{12.34}}{\sqrt{s_{11.34} \cdot s_{22.34}}} = 0,4317.$$

Cálculo de $r_{12.4}$, $r_{13.4}$ e $r_{23.4}$

O cálculo de cada correlação pode ser feito separadamente ou simultaneamente como a seguir:

$$\underbrace{X_1 \ X_2 \ X_3}_{r=3} \ \underbrace{X_4}_{s=1}. \text{ Daí obtém-se: } \hat{D} = \text{diag} (S_{r \cdot s}).$$

$$\hat{\rho}_{\text{Parc}} = \hat{D}^{-1/2} (S_{r \cdot s}) \hat{D}^{-1/2} = \begin{bmatrix} 1 & 0,1777 & -0,2407 \\ 0,1777 & 1 & 0,6231 \\ -0,2407 & 0,6231 & 1 \end{bmatrix}.$$

Logo,

$$r_{12.4} = 0,1777, \ r_{13.4} = -0,2407 \text{ e } r_{23.4} = 0,6231.$$

Outros Exercícios

Exemplo 1: Obter a função geradora de momentos da distribuição normal univariada

Seja $X \sim N(x : \mu, \sigma^2)$. A função densidade de probabilidade de uma variável aleatória X com distribuição normal de probabilidade com média μ e variância σ^2 é dada por:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty.$$

A f.g.m. da variável aleatória X é dada por:

$$M_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f_X(x) dx$$

$$M_X(t) = E(e^{tX}) = E(e^{tX} e^{-t\mu} e^{t\mu}) = e^{t\mu} E[e^{t(X-\mu)}]$$

$$= e^{t\mu} \cdot \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{t(x-\mu) - \frac{(x-\mu)^2}{2\sigma^2}} dx$$

$$= e^{t\mu} \cdot \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-[\frac{(x-\mu)^2}{2\sigma^2} - t(x-\mu)]} dx \quad (1)$$

Somando-se e subtraindo-se ao expoente do integrando em (1) a quantidade $\frac{\sigma^2 t^2}{2}$, virá:

$$\begin{aligned}
M_X(t) &= e^{t\mu} \cdot \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\left(\frac{x-\mu}{\sigma\sqrt{2}} - \frac{\sigma t}{\sqrt{2}}\right)^2 + \frac{\sigma^2 t^2}{2}} dx \\
&= e^{t\mu} \cdot e^{\frac{t^2 \sigma^2}{2}} \cdot \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\left(\frac{x-\mu - \sigma^2 t}{\sigma\sqrt{2}}\right)^2} dx \\
&= e^{t\mu + \frac{t^2 \sigma^2}{2}} \cdot \underbrace{\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{[x-(\mu + \sigma^2 t)]^2}{2\sigma^2}} dx}_{=I} \quad (2)
\end{aligned}$$

Vamos mostrar que $I = 1$, pois I é uma f.d.p. de uma variável aleatória X normalmente distribuída com média $\mu + \sigma^2 t$ e variância σ^2 .

Em (2), seja $\frac{x - (\mu + \sigma^2 t)}{\sigma} = z$. Logo, $x = \sigma z + \mu + \sigma^2 t$ e $dx = \sigma dz$.

$$I = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z^2/2} \sigma dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z^2/2} dz = \frac{2}{\sqrt{2\pi}} \int_0^{\infty} e^{-z^2/2} dz \quad (3)$$

Fazendo em (3), $\frac{z^2}{2} = y$, virá:

$$z^2 = 2y$$

$$z = \sqrt{2} y^{1/2}$$

$$dz = \sqrt{2} \frac{1}{2} y^{-1/2} dy$$

Assim,

$$I = \frac{2}{\sqrt{2\pi}} \int_0^{\infty} e^{-y} \sqrt{2} \frac{1}{2} y^{-1/2} dy = \frac{1}{\sqrt{\pi}} \int_0^{\infty} e^{-y} y^{-1/2} dy \quad (4)$$

$$\text{Sabemos que } \Gamma(\alpha) = \int_0^{\infty} e^{-y} y^{\alpha-1} dy \text{ e } \Gamma\left(\frac{1}{2}\right) = \int_0^{\infty} e^{-y} y^{-1/2} dy = \sqrt{\pi} \quad (5)$$

Substituindo (5) em (4), virá:

$$I = \frac{1}{\sqrt{\pi}} \cdot \sqrt{\pi} = 1.$$

Assim,

$$M_X(t) = e^{t\mu + \frac{t^2 \sigma^2}{2}}.$$

A partir da f.g.m. é fácil obter a média e a variância de X , em que:

(a) Média de X : $E(X) = \mu$

(b) Variância de X : $V(X) = \sigma^2$.

Exemplo 2: Seja o vetor aleatório $\tilde{X}' = [X_1 \quad X_2]$, com distribuição normal bivariada em que,

$E(\tilde{X}') = [0 \quad 0]$, $D(\tilde{X}) = \Sigma$, isto é, $\tilde{X} \sim N_2(\tilde{x}; 0; \Sigma)$. Logo, $V(X_1) = \sigma_1^2$, $V(X_2) = \sigma_2^2$, $Cov(X_1, X_2) = \rho\sigma_1\sigma_2$. Obter a expressão da f.d.p. de \tilde{X} .

Solução:

Pela teoria vista anteriormente, tem-se que:

$$f_{X_1, X_2}(x_1, x_2) = \frac{1}{(2\pi)^{2/2} \sqrt{|\Sigma|}} \cdot e^{-\frac{1}{2} \tilde{x}' \Sigma^{-1} \tilde{x}}$$

$$\text{em que, } \tilde{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \text{ e } \Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}.$$

$$|\Sigma| = \sigma_1^2\sigma_2^2 - \rho^2\sigma_1^2\sigma_2^2 = (1 - \rho^2)\sigma_1^2\sigma_2^2$$

$$\Sigma^{-1} = \frac{1}{(1 - \rho^2)\sigma_1^2\sigma_2^2} \begin{bmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix} = \frac{1}{(1 - \rho^2)} \begin{bmatrix} \frac{1}{\sigma_1^2} & \frac{-\rho}{\sigma_1\sigma_2} \\ \frac{-\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{bmatrix}$$

Assim,

$$f_{X_1, X_2}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1 - \rho^2}} \cdot e^{-\frac{1}{2(1 - \rho^2)} \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma_1^2} & \frac{-\rho}{\sigma_1\sigma_2} \\ \frac{-\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}$$

$$f_{X_1, X_2}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1 - \rho^2}} \cdot e^{-\frac{1}{2(1 - \rho^2)} \left(\frac{x_1^2}{\sigma_1^2} - \frac{2\rho x_1 x_2}{\sigma_1\sigma_2} + \frac{x_2^2}{\sigma_2^2} \right)}$$

Esta é uma f.d.p. de uma normal bivariada, centrada no ponto zero.

Exemplo 3: Dada a densidade binormal (ou normal bivariada):

$$f_{X,Y}(x, y) = \frac{\sqrt{3}}{4\pi} \cdot e^{-\frac{1}{2}(x^2 - xy + y^2)}, I_{(-\infty, \infty)}(x) I_{(-\infty, \infty)}(y).$$

Pede-se calcular $P(Y \leq 2/X = 1)$.

Solução:

Pode-se determinar inicialmente $f_{Y/X}(y/x) = \frac{f_{X,Y}(x, y)}{f_X(x)}$, como a seguir:

$$\Sigma^{-1} = \mathbf{M} = \begin{bmatrix} 1 & -1/2 \\ -1/2 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{M}_{rr} & \mathbf{M}_{rs} \\ \mathbf{M}_{sr} & \mathbf{M}_{ss} \end{bmatrix}$$

$$\Sigma_{rr}^{-1} = \mathbf{M}_{rr} - \mathbf{M}_{rs} \mathbf{M}_{ss}^{-1} \mathbf{M}_{sr} = 1 - \left(-\frac{1}{2}\right) \left(1\right) \left(-\frac{1}{2}\right) = 1 - \frac{1}{4} = \frac{3}{4}$$

$$|\Sigma_{rr}^{-1}| = 3/4$$

$$f_x(x) = \frac{\sqrt{3/4}}{(2\pi)^{1/2}} \cdot e^{-\frac{1}{2}x\left(\frac{3}{4}\right)x}$$

$$f_x(x) = \frac{\sqrt{3}}{2\sqrt{2\pi}} \cdot e^{-\frac{3}{8}x^2}$$

$$f_{Y/X}(y/x) = \frac{\frac{\sqrt{3}}{4\pi} \cdot e^{-\frac{1}{2}(x^2 - xy + y^2)}}{\frac{\sqrt{3}}{2\sqrt{2\pi}} \cdot e^{-\frac{3}{8}x^2}}$$

$$f_{Y/X}(y/x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{1}{4}x^2 - xy + y^2\right)}$$

$$f_{Y/X=1}(y/x=1) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(y^2 - y + \frac{1}{4}\right)}$$

$$f_{Y/X=1}(y/x=1) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(y - \frac{1}{2}\right)^2}.$$

Logo,

$$P(Y \leq 2/X=1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^2 e^{-\frac{1}{2}\left(y - \frac{1}{2}\right)^2} dy$$

Fazendo $y - \frac{1}{2} = z \Rightarrow y = z + \frac{1}{2} \Rightarrow dy = dz$

y	z
$-\infty$	$-\infty$
2	3/2

$$\therefore P(Y \leq 2/X=1) = P\left(Z \leq \frac{3}{2}/X=1\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{3/2} e^{-\frac{z^2}{2}} dz$$

Pela Tabela da Distribuição Normal Padrão $[Z \sim N(0, 1)]$, virá:

$$P(Y \leq 2/X=1) = 0,9332.$$

Observação: A seguir, como ilustração, será obtida a função $f_{Y/X}(y/x)$, utilizando a fórmula deduzida anteriormente.

$$f_{X_r/X_s}(x_r/x_s) = \frac{1}{(2\pi)^{r/2} \left| \sum_{rr,ss} \right|^{1/2}} \cdot \exp \left[-\frac{1}{2} \left(x_r - \mu_{r,s} \right)' \sum_{rr,ss}^{-1} \left(x_r - \mu_{r,s} \right) \right]$$

em que,
$$\begin{cases} \mu_{r,s} = \mu_r + \sum_{rs} \sum_{ss}^{-1} \left(x_s - \mu_s \right) \\ \sum_{rr,ss} = \sum_{rr} - \sum_{rs} \sum_{ss}^{-1} \sum_{sr} \\ r + s = p \end{cases}$$

No exemplo:

$$\sum^{-1} = \begin{bmatrix} 1 & -1/2 \\ -1/2 & 1 \end{bmatrix} e \sum = \begin{bmatrix} \sum_{rr} & \sum_{rs} \\ \sum_{sr} & \sum_{ss} \end{bmatrix} = \begin{bmatrix} 4/3 & 2/3 \\ 2/3 & 4/3 \end{bmatrix}$$

$$r = s = 1; X_r = Y; X_s = X; \mu_r = \mu_Y = 0; \mu_s = \mu_X = 0.$$

Assim,

$$\mu_{r,s} = 0 + \left(\frac{2}{3} \right) \left(\frac{3}{4} \right) (x - 0) = \frac{1}{2} x$$

$$\sum_{rr,ss} = \frac{4}{3} - \left(\frac{2}{3} \right) \left(\frac{3}{4} \right) \left(\frac{2}{3} \right) = \frac{4}{3} - \frac{1}{3} = 1$$

$$f_{Y/X}(y/x) = \frac{1}{(2\pi)^{r/2} \left| \sum_{rr,ss} \right|^{1/2}} \cdot e^{-\frac{1}{2} \left(y - \frac{1}{2} x \right) \sum_{rr,ss}^{-1} \left(y - \frac{1}{2} x \right)}$$

$$f_{Y/X}(y/x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left(y - \frac{1}{2} x \right)^2}$$

$$f_{Y/X=1}(y/x=1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left(y - \frac{1}{2} \right)^2}.$$

Exemplo 4:

Dada a seguinte função densidade binormal:

$$f_{X,Y}(x,y) = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2 - 2xy + 2y^2)}.$$

Pede-se obter:

- a) $E(Y/X)$;
 b) Função geradora de momentos;
 c) O coeficiente de correlação ρ_{XY} .

Solução:

a) $E(Y/X) = \int_{-\infty}^{\infty} y f_{Y/X}(y/x) dy$. Para obtenção da função $f_{Y/X}(y/x)$, será usado o fato da observação anterior.

$$\Sigma^{-1} = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}, \Sigma = \begin{bmatrix} V(Y) & \text{Cov}(Y, X) \\ \text{Cov}(X, Y) & V(X) \end{bmatrix} = \begin{bmatrix} \sum_{rr} & \sum_{rs} \\ \sum_{sr} & \sum_{ss} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$$

$$\mu_{r.s} = \mu_r + \sum_{rs} \Sigma_{ss}^{-1} \left(x_s - \mu_s \right) = 0 + 1 \left(\frac{1}{2} \right) (x - 0) = \frac{x}{2}$$

$$\sum_{rr-s} = \sum_{rr} - \sum_{rs} \Sigma_{ss}^{-1} \sum_{sr} = 1 - (1) \left(\frac{1}{2} \right) (1) = \frac{1}{2}$$

$$f_{Y/X}(y/x) = \frac{1}{(2\pi)^{1/2} |1/2|^{1/2}} \cdot e^{-\frac{1}{2} \left(y - \frac{x}{2} \right) (2) \left(y - \frac{x}{2} \right)}$$

$$f_{Y/X}(y/x) = \frac{\sqrt{2}}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2} 2 \left(y - \frac{x}{2} \right)^2}$$

$$f_{Y/X}(y/x) = \frac{\sqrt{2}}{\sqrt{2\pi}} \cdot e^{-\frac{1}{2} \left(\sqrt{2} y - \frac{x}{\sqrt{2}} \right)^2}$$

$$\text{Assim virá: } E(Y/X) = \int_{-\infty}^{\infty} \frac{\sqrt{2}}{\sqrt{2\pi}} \cdot y e^{-\frac{1}{2} \left(\sqrt{2} y - \frac{x}{\sqrt{2}} \right)^2} dy$$

$$\text{Seja a transformação } \sqrt{2} y - \frac{x}{\sqrt{2}} = z$$

$$y = \frac{z}{\sqrt{2}} + \frac{x}{2}$$

$$dy = \frac{1}{\sqrt{2}} dz$$

$$E(Y/X) = \frac{\sqrt{2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left(\frac{z}{\sqrt{2}} + \frac{x}{2} \right) e^{-\frac{z^2}{2}} \frac{1}{\sqrt{2}} dz$$

$$E(Y/X) = \frac{1}{\sqrt{2}\sqrt{2\pi}} \int_{-\infty}^{\infty} z e^{-\frac{z^2}{2}} dz + \frac{x}{2} \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz$$

Sendo $f_Z(z) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{z^2}{2}}$, $-\infty < z < \infty$, sabe-se que $\int_{-\infty}^{\infty} f_Z(z) dz = 1$ e $Z \sim N(0, 1)$.

Logo,

$$E(Y/X) = \frac{1}{\sqrt{2}}(0) + \frac{x}{2}(1)$$

$E(Y/X) = \frac{x}{2}$. Note que este resultado é exatamente igual a $\mu_{r,s}$, que é a média condicional de

Y dado X.

b) Função geradora de momentos

$$\text{Sabe-se que: } M_{\underline{X}'}(\underline{t}') = e^{\underline{t}'\underline{\mu}' + \frac{1}{2}\underline{t}'\Sigma\underline{t}'}.$$

Seja, $\underline{X}' = [Y \quad X]$; $\underline{t}' = [t_1 \quad t_2]$; $\underline{\mu}' = [0 \quad 0]$.

$$\Sigma = \begin{bmatrix} V(Y) & \text{Cov}(Y, X) \\ \text{Cov}(X, Y) & V(X) \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$$

$$M_{Y,X}(t_1, t_2) = e^{\begin{bmatrix} t_1 & t_2 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} t_1 & t_2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \end{bmatrix}}$$

$$M_{Y,X}(t_1, t_2) = e^{\frac{1}{2}(t_1^2 + 2t_1t_2 + 2t_2^2)}.$$

$$\text{c) } \rho_{XY} = \frac{\text{Cov}(X, Y)}{\sqrt{V(X)V(Y)}} = \frac{1}{\sqrt{(2)(1)}} = \frac{1}{\sqrt{2}}.$$

Exemplo 5:

Mostrar que a função dada abaixo é uma binormal, a qual se identifica com a fórmula teórica.

$$f_{X,Y}(x, y) = \frac{\sqrt{15}}{4\pi} \cdot e^{-\frac{1}{2}(x^2 - xy + 4y^2)}, \quad \begin{matrix} -\infty < x < \infty \\ -\infty < y < \infty \end{matrix}$$

Solução:

$$\underline{\mu} = \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$f_{X,Y}(x, y) = \frac{\sqrt{|\Sigma^{-1}|}}{(2\pi)^{2/2}} \cdot e^{-\frac{1}{2}[\underline{x} \quad \underline{y}] \Sigma^{-1} \begin{bmatrix} x \\ y \end{bmatrix}}$$

$$\Sigma^{-1} = \begin{bmatrix} 1 & -1/2 \\ -1/2 & 4 \end{bmatrix}, \quad \det(\Sigma^{-1}) = |\Sigma^{-1}| = 4 - \frac{1}{4} = \frac{15}{4}.$$

Diagonalizando \sum^{-1} por congruência, virá:

$$\begin{bmatrix} 1 & -1/2 & 1 & 0 \\ -1/2 & 4 & 0 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & -1/2 & 1 & 0 \\ 0 & 15/4 & 1/2 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 15/4 & 1/2 & 1 \end{bmatrix}$$

Logo, \sum^{-1} é simétrica e positiva definida.

$$K = \frac{\sqrt{|\sum^{-1}|}}{(2\pi)^{p/2}} = \frac{\sqrt{\frac{15}{4}}}{(2\pi)^{2/2}} = \frac{\sqrt{15}}{4\pi}.$$

Assim, a função dada é uma binormal e se identifica com a fórmula teórica.

3.2. DISTRIBUIÇÃO DE WISHART

3.2.1. Introdução

A distribuição de Wishart é uma generalização multivariada da distribuição qui-quadrado (χ^2). No caso univariado, a distribuição qui-quadrado é definida como a soma de n variáveis X independentes e identicamente distribuídas como $N(0, 1)$ elevadas ao quadrado.

Teorema: Suponha que os vetores com p componentes $\tilde{Z}_1, \dots, \tilde{Z}_n$ ($n \geq p$) são independentes,

cada um distribuído de acordo com $N_p\left(0, \sum\right)$. Então, a densidade de $A = \sum_{\alpha=1}^n \tilde{Z}_{\alpha} \tilde{Z}_{\alpha}'$ é

$$w(A/\sum, n) = \frac{|\sum|^{-\frac{1}{2}(n-p-1)} \cdot e^{-\frac{1}{2}\text{tr}(\sum^{-1}A)}}{2^{\frac{1}{2}np} \cdot \pi^{p(p-1)/4} \cdot \left| \sum \right|^{\frac{1}{2}n} \cdot \prod_{i=1}^p \Gamma\left[\frac{1}{2}(n+1-i)\right]} \quad (1)$$

para A positiva definida e zero em caso contrário.

Corolário: Suponha que os vetores $\tilde{X}_1, \dots, \tilde{X}_N$ ($N > p$) são independentes, cada qual com

distribuição $N_p\left(\mu, \sum\right)$. Então, a densidade de $A = \sum_{\alpha=1}^N \left(\tilde{X}_{\alpha} - \bar{\tilde{X}}\right)\left(\tilde{X}_{\alpha} - \bar{\tilde{X}}\right)'$ é (1) para $n = N - 1$.

- A densidade (1) pode ser denotada por $w(A/\sum, n)$ e a distribuição associada por $W(\sum, n)$.
- Para $S = \frac{A}{N-1}$, tem-se que $(N-1)S = A \sim W_p(\sum, N-1)$.

$$w\left(\mathbf{A} / \sum, N-1\right)=\frac{\left|\mathbf{A}\right|^{\frac{1}{2}(N-p-2)} \cdot e^{-\frac{1}{2} \operatorname{tr}\left(\sum^{-1} \mathbf{A}\right)}}{2^{\frac{1}{2}(N-1) p} \cdot \pi^{p(p-1) / 4} \cdot\left|\sum\right|^{\frac{1}{2}(N-1)} \cdot \prod_{i=1}^p \Gamma\left[\frac{1}{2}(N-i)\right]} .$$

Nota: $\operatorname{tr}\left(\sum^{-1} \mathbf{A}\right)=\sum_{i=1}^p \sum_{j=1}^p \sigma^{ij} a_{ij}$

Exemplo: $p=1$ e $n=1$

A partir de (1) temos que:

$$\begin{aligned} w\left(\mathbf{A} / \sum, n\right) &= w\left(x_1^2 / \sigma_{11}, 1\right) \\ &= \frac{x_1^{-1} \cdot e^{-\frac{1}{2} \frac{x_1^2}{\sigma_{11}}}}{\sqrt{2}(1) \sigma_{11}^{1/2} \Gamma\left(\frac{1}{2}\right)} \\ &= \frac{1}{\sqrt{2 \pi} \sigma_{11}} x_1^{-1} \cdot e^{-\frac{1}{2} \frac{x_1^2}{\sigma_{11}}} \end{aligned} \quad (2)$$

A expressão (2) é a função densidade de probabilidade de uma variável aleatória X_1^2 que tem distribuição qui-quadrado com 1 grau de liberdade, isto é, $X_1^2 \sim \chi_{(1)}^2$.

Para $\sigma_{11}=1$ em (2), segue que:

$$X_1^2 \sim \chi_{(1)}^2 \text{ com f.d.p. } f_{x_1^2}\left(x_1^2\right)=\frac{1}{\sqrt{2 \pi}} x_1^{-1} \cdot e^{-\frac{x_1^2}{2}} .$$

Uma variável aleatória V_n com distribuição qui-quadrado com n graus de liberdade $\left(V_n \sim \chi_{(n)}^2\right)$, tem a seguinte função densidade de probabilidade:

$$f_{v_n}\left(v_n\right)=\begin{cases} \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} \cdot v_n^{\frac{n}{2}-1} \cdot e^{-\frac{v_n}{2}}, & \text{ para } v_n > 0 \\ 0 & , \text{ para outros valores de } v_n \end{cases}$$

Definição:

$$\text{Seja } {}_p \mathbf{X}_n=\left[\begin{array}{cccc} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ X_{p1} & X_{p2} & \cdots & X_{pn} \end{array}\right]=\left[\tilde{X}_1, \tilde{X}_2, \cdots, \tilde{X}_n\right],$$

n medidas em cada uma das p variáveis, com $n > p$. Sejam $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n$ uma amostra aleatória de tamanho n independentes e identicamente distribuídas como $N_p(0, \Sigma)$, isto é

$\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n \text{ IID } \sim N_p(0, \Sigma)$. Então a matriz aleatória

$${}_pA_p = \tilde{X}_1 \tilde{X}_1' + \tilde{X}_2 \tilde{X}_2' + \dots + \tilde{X}_n \tilde{X}_n'$$

tem distribuição de Wishart com parâmetros n e Σ , isto é, $A \sim W_p(n, \Sigma)$, em que n é o número de graus de liberdade.

3.2.2. Propriedades

Algumas propriedades associadas a esta distribuição são:

- i) Seja $A_1 \sim W_p(n_1, \Sigma)$ e $A_2 \sim W_p(n_2, \Sigma)$. Se A_1 e A_2 forem independentes, então $A_1 + A_2 \sim W_p(n_1 + n_2, \Sigma)$.
- ii) Se $A \sim W_p(n, \Sigma)$ e $\text{Posto}({}_qC'_p) = q \leq p$, então $C'AC \sim W_q(n, C'\Sigma C)$.
- iii) Sejam $\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_n$ uma amostra aleatória de tamanho n de uma distribuição normal p -variada com média μ e matriz de covariância Σ . Então,

- $\bar{\tilde{X}} = \frac{1}{n} \sum_{j=1}^n \tilde{X}_j$ é distribuído como $N_p(\mu, \frac{1}{n} \Sigma)$.

- Seja $S = \frac{A}{n-1} = \frac{1}{n-1} \sum_{j=1}^n \left(\tilde{X}_j - \bar{\tilde{X}} \right) \left(\tilde{X}_j - \bar{\tilde{X}} \right)'$. Então,

$$(n-1)S = \sum_{j=1}^n \left(\tilde{X}_j - \bar{\tilde{X}} \right) \left(\tilde{X}_j - \bar{\tilde{X}} \right)' \sim W_p(n-1, \Sigma).$$

- $\bar{\tilde{X}}$ e S são independentes.

3.3. DISTRIBUIÇÃO DE HOTELLING

3.3.1. Introdução

A distribuição de Hotelling foi inicialmente proposta por Harold Hotelling tendo como objetivo generalizar a distribuição t de Student no campo multivariado.

Definição:

Seja $\tilde{Y} \sim N_p(\mu, \frac{1}{a} \Sigma)$ e $A = bS \sim W_p(b, \Sigma)$ independentes. Para

$T^2 = a \left(\bar{Y} - \underline{\mu} \right)' S^{-1} \left(\bar{Y} - \underline{\mu} \right)$, tem-se que: $\frac{b-p+1}{bp} \cdot T^2 \sim F_{p, b-p+1}$. (Prova: ANDERSON, 1984).

Observação: \bar{Y} é um vetor $p \times 1$ e a é um escalar constante, cujo valor depende da situação.

3.3.2. Teste de um vetor de médias contra um padrão populacional

$$\text{Seja } {}_pX_n = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ X_{p1} & X_{p2} & \cdots & X_{pn} \end{bmatrix} = [X_1, X_2, \cdots, X_n],$$

n medidas em cada uma das p variáveis, com $n > p$.

Sejam X_1, X_2, \cdots, X_n uma amostra aleatória de tamanho n de uma distribuição normal

p -variada com média $\underline{\mu}$ e matriz de covariância $\underline{\Sigma}$. Assim,

$$\bar{X} = \frac{1}{n} \sum_{j=1}^n X_j \text{ é o vetor } p \times 1 \text{ de médias amostrais;}$$

$$S = \frac{1}{n-1} \sum_{j=1}^n \left(X_j - \bar{X} \right) \left(X_j - \bar{X} \right)' \text{ é a matriz } p \times p \text{ de covariância amostral.}$$

A hipótese pode ser formulada como: $H_0 : \underline{\mu} = \underline{\mu}_0$ vs. $H_a : \underline{\mu} \neq \underline{\mu}_0$, isto é,

$$H_0 : \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} = \begin{bmatrix} \mu_{01} \\ \mu_{02} \\ \vdots \\ \mu_{0p} \end{bmatrix} \text{ vs. } H_a : \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} \neq \begin{bmatrix} \mu_{01} \\ \mu_{02} \\ \vdots \\ \mu_{0p} \end{bmatrix}.$$

O problema é então construir um teste estatístico para a hipótese H_0 , isto é, construir um teste estatístico que permita verificar se o vetor de médias da população em estudo difere de um vetor padrão de médias.

Sendo $X_1, X_2, \cdots, X_n \text{ IID} \sim N_p(\underline{\mu}, \underline{\Sigma})$, então,

$$\bar{X} \sim N_p\left(\underline{\mu}, \frac{1}{n} \underline{\Sigma}\right)$$

e

$$A = (n-1)S \sim W_p(n-1, \underline{\Sigma}).$$

Para $a = n$ e $b = n-1$, virá:

$$T^2 = n \left(\bar{\underline{X}} - \underline{\mu}_0 \right)' S^{-1} \left(\bar{\underline{X}} - \underline{\mu}_0 \right)$$

e

$$\frac{n-p}{(n-1)p} \cdot T^2 \sim F_{p, n-p}.$$

Assim, a estatística usada para o teste do vetor de médias é a T^2 de Hotelling com p e $n-p$ graus de liberdade, que pode ser transformada em uma estatística F com p e $n-p$ graus de liberdade, o que torna mais acessível a verificação da significância.

Logo, para o teste da hipótese $H_0 : \underline{\mu} = \underline{\mu}_0$ vs. $H_a : \underline{\mu} \neq \underline{\mu}_0$, a um nível de significância

α , rejeita-se H_0 em favor de H_a se:

$$T^2 = n \left(\bar{\underline{X}} - \underline{\mu}_0 \right)' S^{-1} \left(\bar{\underline{X}} - \underline{\mu}_0 \right) > \frac{(n-1)p}{n-p} \cdot F_{\alpha, p, n-p}$$

ou ainda, se

$$F = \frac{n-p}{(n-1)p} \cdot T^2 > F_{\alpha, p, n-p}.$$

Observação: Pode-se escrever:

$$T^2 = \sqrt{n} \left(\bar{\underline{X}} - \underline{\mu}_0 \right)' \left(\frac{\sum_{j=1}^n \left(\underline{X}_j - \bar{\underline{X}} \right) \left(\underline{X}_j - \bar{\underline{X}} \right)'}{n-1} \right)^{-1} \sqrt{n} \left(\bar{\underline{X}} - \underline{\mu}_0 \right),$$

a qual é da forma

$$\left(\begin{array}{c} \text{vetor aleatório} \\ \text{normal multivariado} \end{array} \right)' \left(\frac{\text{matriz aleatória Wishart}}{\text{graus de liberdade}} \right)^{-1} \left(\begin{array}{c} \text{vetor aleatório} \\ \text{normal multivariado} \end{array} \right)$$

De modo análogo, para o caso univariado, sabe-se que:

$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \Rightarrow t^2 = \frac{(\bar{X} - \mu_0)^2}{s^2/n}, \text{ que pode ser escrita como,}$$

$$t^2 = \sqrt{n} (\bar{X} - \mu_0) (s^2)^{-1} \sqrt{n} (\bar{X} - \mu_0)$$

ou

$$\left(\begin{array}{c} \text{variável aleatória} \\ \text{normal} \end{array} \right) \left(\frac{\begin{array}{c} \text{(escalar) variável} \\ \text{aleatória qui-quadrado} \end{array}}{\text{graus de liberdade}} \right)^{-1} \left(\begin{array}{c} \text{variável aleatória} \\ \text{normal} \end{array} \right)$$

Exemplo:

Seja a matriz de dados de uma amostra aleatória de tamanho $n=3$ de uma população normal bivariada

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

Testar a hipótese $H_0 : \underline{\mu} = \begin{bmatrix} 9 \\ 5 \end{bmatrix}$ vs. $H_a : \underline{\mu} \neq \begin{bmatrix} 9 \\ 5 \end{bmatrix}$, através do teste T^2 de Hotelling ao

nível de significância $\alpha = 5\%$.

Solução:

- $H_0 : \underline{\mu} = \begin{bmatrix} 9 \\ 5 \end{bmatrix}$ vs. $H_a : \underline{\mu} \neq \begin{bmatrix} 9 \\ 5 \end{bmatrix}$

- Cálculo do vetor de médias amostrais: $\bar{\underline{X}}$

$$\bar{X}_i = \frac{1}{n} \sum_{j=1}^n X_{ij}, \quad i = 1, 2, \dots, p \quad (p = 2 \text{ no exemplo})$$

$$\bar{\underline{X}} = \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \end{bmatrix} = \begin{bmatrix} \frac{6+10+8}{3} \\ \frac{9+6+3}{3} \end{bmatrix} = \begin{bmatrix} 8 \\ 6 \end{bmatrix}$$

- Cálculo da matriz de covariância amostral: S

$$s_{ii} = s_i^2 = \hat{V}(X_i) = \frac{1}{n-1} \sum_{j=1}^n (X_{ij} - \bar{X}_i)^2 = \frac{\sum_{j=1}^n X_{ij}^2 - \frac{\left(\sum_{j=1}^n X_{ij}\right)^2}{n}}{n-1}$$

e

$$s_{ik} = \text{Cov}(X_i, X_k) = \frac{1}{n-1} \sum_{j=1}^n (X_{ij} - \bar{X}_i)(X_{kj} - \bar{X}_k) = \frac{\sum_{j=1}^n X_{ij}X_{kj} - \frac{\left(\sum_{j=1}^n X_{ij}\right)\left(\sum_{j=1}^n X_{kj}\right)}{n}}{n-1},$$

com $i = 1, 2, \dots, p$; $k = 1, 2, \dots, p$.

$$s_{11} = \frac{(6-8)^2 + (10-8)^2 + (8-8)^2}{2} = 4$$

$$s_{12} = \frac{(6-8)(9-6) + (10-8)(6-6) + (8-8)(3-6)}{2} = -3$$

$$s_{22} = \frac{(9-6)^2 + (6-6)^2 + (3-6)^2}{2} = 9$$

Assim,

$$S = \begin{bmatrix} 4 & -3 \\ -3 & 9 \end{bmatrix}$$

$$S^{-1} = \frac{1}{(4)(9) - (-3)(-3)} \begin{bmatrix} 9 & 3 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 1/3 & 1/9 \\ 1/9 & 4/27 \end{bmatrix}$$

- Cálculo das estatísticas T^2 e F

$$T^2 = n \left(\bar{\mathbf{X}} - \underline{\mu}_0 \right)' S^{-1} \left(\bar{\mathbf{X}} - \underline{\mu}_0 \right)$$

$$F = \frac{n-p}{(n-1)p} \cdot T^2, \text{ com } p \text{ e } n-p \text{ graus de liberdade.}$$

$$T^2 = 3 \begin{bmatrix} 8-9 & 6-5 \end{bmatrix} \begin{bmatrix} 1/3 & 1/9 \\ 1/9 & 4/27 \end{bmatrix} \begin{bmatrix} 8-9 \\ 6-5 \end{bmatrix} = 3 \begin{bmatrix} -1 & 1 \end{bmatrix} \begin{bmatrix} -2/9 \\ 1/27 \end{bmatrix} = \frac{7}{9} \cong 0,7778$$

$$F = \frac{3-2}{(3-1)2} \cdot \frac{7}{9} = \frac{7}{36} = 0,194^{\text{n.s.}} \quad \text{n.s. (P>0,05)}$$

$$F_{5\%}(2,1) = 199,5$$

Dado que $F = 0,194 < 199,5$, não se rejeita H_0 ao nível de 5% de probabilidade.

FATOS:

- i) Sabe-se que T^2 é distribuída como $\frac{(n-1)p}{n-p} \cdot F_{p, n-p}$ onde $F_{p, n-p}$ denota uma variável aleatória com distribuição F com p e n-p graus de liberdade. Assim, pode-se comparar o valor observado $T^2 = \frac{7}{9} = 0,7778$ com o valor crítico

$$\frac{(n-1)p}{(n-p)} F_{\alpha, p, n-p} = \frac{(3-1)2}{3-2} (199,5) = 798.$$

Dado que $T^2 = 0,7778 < 798$, não se rejeita H_0 ao nível de 5% de probabilidade pelo teste T^2 de Hotelling.

- ii) Foi visto que a estatística usada para o teste de um vetor de médias contra um padrão populacional é a T^2 de Hotelling com p e n-p graus de liberdade. Como as Tabelas para esta estatística são menos habituais, o valor calculado de T^2 pode ser transformado em uma estatística F com p e n-p graus de liberdade, como no exemplo ($F=0,194$), o que torna mais acessível a verificação da significância.

3.3.3. Teste dos vetores de médias de duas populações

Para testar a hipótese de igualdade de vetores de médias de duas populações normais multivariadas pode-se utilizar a estatística T^2 de Hotelling, como apresentada a seguir.

Considere uma amostra aleatória de tamanho n_1 da população 1 e outra de tamanho n_2 da população 2. As observações em p variáveis podem ser dispostas como na Tabela 3.2.

Tabela 3.2 – Estatísticas para duas populações normais multivariadas

Amostras	Estatísticas
(População 1) $X_{11}, X_{12}, \dots, X_{1n_1}$	$\bar{\tilde{X}}_1 = \frac{1}{n_1} \sum_{j=1}^{n_1} \tilde{X}_{1j}, \quad S_1 = \frac{A_1}{n_1 - 1} = \frac{1}{n_1 - 1} \sum_{j=1}^{n_1} \left(\tilde{X}_{1j} - \bar{\tilde{X}}_1 \right) \left(\tilde{X}_{1j} - \bar{\tilde{X}}_1 \right)'$
(População 2) $X_{21}, X_{22}, \dots, X_{2n_2}$	$\bar{\tilde{X}}_2 = \frac{1}{n_2} \sum_{j=1}^{n_2} \tilde{X}_{2j}, \quad S_2 = \frac{A_2}{n_2 - 1} = \frac{1}{n_2 - 1} \sum_{j=1}^{n_2} \left(\tilde{X}_{2j} - \bar{\tilde{X}}_2 \right) \left(\tilde{X}_{2j} - \bar{\tilde{X}}_2 \right)'$

- Nesta notação, o primeiro índice refere-se à população.
- $\bar{\tilde{X}}_1$ e $\bar{\tilde{X}}_2$ são vetores de médias amostrais $p \times 1$
- S_1 e S_2 são matrizes de covariância $p \times p$

Deseja-se testar se os vetores de médias das duas populações são iguais, isto é, se $\mu_1 = \mu_2$ (ou equivalentemente, $\mu_1 - \mu_2 = \underline{0}$).

A hipótese pode ser formulada como: $H_0 : \mu_1 = \mu_2$ vs. $H_a : \mu_1 \neq \mu_2$, isto é,

$$H_0 : \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \vdots \\ \mu_{1p} \end{bmatrix} = \begin{bmatrix} \mu_{21} \\ \mu_{22} \\ \vdots \\ \mu_{2p} \end{bmatrix} \text{ vs. } H_a : \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \vdots \\ \mu_{1p} \end{bmatrix} \neq \begin{bmatrix} \mu_{21} \\ \mu_{22} \\ \vdots \\ \mu_{2p} \end{bmatrix}.$$

Suposições:

Da população 1: $\tilde{X}_{11}, \tilde{X}_{12}, \dots, \tilde{X}_{1n_1} \text{ IID} \sim N_p(\mu_1, \Sigma)$

$$A_1 = (n_1 - 1)S_1 \sim W_p(n_1 - 1, \Sigma)$$

Da população 2: $\tilde{X}_{21}, \tilde{X}_{22}, \dots, \tilde{X}_{2n_2} \text{ IID} \sim N_p(\mu_2, \Sigma)$

$$A_2 = (n_2 - 1)S_2 \sim W_p(n_2 - 1, \Sigma)$$

Também, $\tilde{X}_{11}, \tilde{X}_{12}, \dots, \tilde{X}_{1n_1}$, são independente de $\tilde{X}_{21}, \tilde{X}_{22}, \dots, \tilde{X}_{2n_2}$.

Supondo, $\sum_1 = \sum_2 = \sum$, $\sum_{j=1}^{n_1} \left(\tilde{X}_{1j} - \bar{\tilde{X}}_1 \right) \left(\tilde{X}_{1j} - \bar{\tilde{X}}_1 \right)'$ é um estimador de $(n_1 - 1)\sum$ e

$\sum_{j=1}^{n_2} \left(\tilde{X}_{2j} - \bar{\tilde{X}}_2 \right) \left(\tilde{X}_{2j} - \bar{\tilde{X}}_2 \right)'$ é um estimador de $(n_2 - 1)\sum$. Combinando estas informações,

obtém-se a matriz comum de covariância amostral, que é dada por:

$$S_c = \frac{\sum_{j=1}^{n_1} \left(\tilde{X}_{1j} - \bar{\tilde{X}}_1 \right) \left(\tilde{X}_{1j} - \bar{\tilde{X}}_1 \right)' + \sum_{j=1}^{n_2} \left(\tilde{X}_{2j} - \bar{\tilde{X}}_2 \right) \left(\tilde{X}_{2j} - \bar{\tilde{X}}_2 \right)'}{n_1 + n_2 - 2},$$

ou ainda,

$$S_c = \frac{A_1 + A_2}{n_1 + n_2 - 2} = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{n_1 + n_2 - 2}.$$

$$E\left(\bar{\tilde{X}}_1 - \bar{\tilde{X}}_2\right) = E\left(\bar{\tilde{X}}_1\right) - E\left(\bar{\tilde{X}}_2\right) = \mu_1 - \mu_2.$$

$\bar{\tilde{X}}_1$ e $\bar{\tilde{X}}_2$ são independentes e assim $\text{Cov}\left(\bar{\tilde{X}}_1, \bar{\tilde{X}}_2\right) = \phi$.

$$\begin{aligned} \text{Cov}\left(\bar{\tilde{X}}_1 - \bar{\tilde{X}}_2\right) &= \text{Cov}\left(\bar{\tilde{X}}_1\right) + \text{Cov}\left(\bar{\tilde{X}}_2\right) \\ &= \frac{1}{n_1}\sum + \frac{1}{n_2}\sum = \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\sum. \end{aligned}$$

Assim, $\text{Cov}\left(\bar{\tilde{X}}_1 - \bar{\tilde{X}}_2\right) = \left(\frac{1}{n_1} + \frac{1}{n_2}\right)S_c$ é um estimador de $\text{Cov}\left(\bar{\tilde{X}}_1 - \bar{\tilde{X}}_2\right)$.

Sob $H_0 : \mu_1 = \mu_2$, $\bar{\tilde{X}}_1 - \bar{\tilde{X}}_2 \sim N_p\left(0, \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\sum\right)$.

Segue que, $A_1 + A_2 = (n_1 + n_2 - 2)S_c \sim W_p(n_1 + n_2 - 2, \sum)$.

Fazendo $b = n_1 + n_2 - 2$ e $a = \frac{n_1 \cdot n_2}{n_1 + n_2}$, e sob $H_0 : \mu_1 = \mu_2$,

$$T^2 = \frac{n_1 \cdot n_2}{n_1 + n_2} \left[\bar{\tilde{X}}_1 - \bar{\tilde{X}}_2 \right]' S_c^{-1} \left[\bar{\tilde{X}}_1 - \bar{\tilde{X}}_2 \right]$$

e $F = \frac{b - p + 1}{bp} \cdot T^2 = \frac{n_1 + n_2 - p - 1}{(n_1 + n_2 - 2)p} \cdot T^2$ com p e $n_1 + n_2 - p - 1$ graus de liberdade.

Logo, T^2 é distribuída como $\frac{(n_1 + n_2 - 2)p}{(n_1 + n_2 - p - 1)} F_{p, n_1 + n_2 - p - 1}$.

Assim, a estatística usada para o teste da igualdade dos dois vetores de médias é a T^2 de Hotelling com p e $n_1 + n_2 - p - 1$ graus de liberdade, que pode ser transformada em uma estatística F com p e $n_1 + n_2 - p - 1$ graus de liberdade, o que torna mais acessível a verificação da significância.

Logo, para o teste da hipótese $H_0 : \mu_1 = \mu_2$ vs. $H_a : \mu_1 \neq \mu_2$, a um nível de significância α , rejeita-se H_0 em favor de H_a se

$$T^2 = \frac{n_1 \cdot n_2}{n_1 + n_2} \left[\begin{matrix} \bar{X}_1 - \bar{X}_2 \\ \bar{X}_1 - \bar{X}_2 \end{matrix} \right]' S_c^{-1} \left[\begin{matrix} \bar{X}_1 - \bar{X}_2 \\ \bar{X}_1 - \bar{X}_2 \end{matrix} \right] > \frac{(n_1 + n_2 - 2)p}{(n_1 + n_2 - p - 1)} F_{\alpha, p, n_1 + n_2 - p - 1},$$

ou ainda, se

$$F = \frac{n_1 + n_2 - p - 1}{(n_1 + n_2 - 2)p} \cdot T^2 > F_{\alpha, p, n_1 + n_2 - p - 1}.$$

Exemplo:

De uma amostra de folhas de duas subespécies de orquídeas, mediu-se o comprimento (X_1) e a largura (X_2) das folhas, e obtiveram-se os seguintes resultados:

	SUBESPÉCIE 1		SUBESPÉCIE 2	
	X_1	X_2	X_1	X_2
	3	6	2	11
	9	6	7	14
	16	8	13	18
	19	13	19	18
	24	12	23	20
Totais:	71	45	64	81
Médias:	14,2	9,0	12,8	16,2

Testar a hipótese de igualdade dos vetores de médias das duas populações, pelo teste T^2 de Hotelling, para $\alpha = 5\%$.

Solução:

$$\bullet \quad \mu_1 = \begin{bmatrix} \mu_{11} \\ \mu_{12} \end{bmatrix} \quad \text{e} \quad \mu_2 = \begin{bmatrix} \mu_{21} \\ \mu_{22} \end{bmatrix}$$

$$H_0 : \mu_1 = \mu_2 \text{ vs. } H_a : \mu_1 \neq \mu_2$$

- Cálculo dos vetores de médias amostrais

$$\bar{\tilde{X}}_1 = \begin{bmatrix} \bar{X}_{11} \\ \bar{X}_{12} \end{bmatrix} = \begin{bmatrix} 14,2 \\ 9,0 \end{bmatrix} \quad \bar{\tilde{X}}_2 = \begin{bmatrix} \bar{X}_{21} \\ \bar{X}_{22} \end{bmatrix} = \begin{bmatrix} 12,8 \\ 16,2 \end{bmatrix}$$

- Cálculo da matriz comum de covariância amostral

$$S_c = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{n_1 + n_2 - 2}$$

Matriz de covariância amostral dentro da subespécie 1:

$$s_{11} = \frac{3^2 + 9^2 + \dots + 24^2 - \frac{(71)^2}{5}}{4} = \frac{1283 - 1008,2}{4} = 68,7$$

$$s_{12} = \frac{(3)(6) + (9)(6) + \dots + (24)(12) - \frac{(71)(45)}{5}}{4} = \frac{735 - 639}{4} = 24$$

$$s_{22} = \frac{6^2 + 6^2 + \dots + 12^2 - \frac{(45)^2}{5}}{4} = \frac{449 - 405}{4} = 11$$

$$S_1 = \begin{bmatrix} 68,7 & 24 \\ 24 & 11 \end{bmatrix}$$

Matriz de covariância amostral dentro da subespécie 2:

$$s_{11} = \frac{2^2 + 7^2 + \dots + 23^2 - \frac{(64)^2}{5}}{4} = \frac{1112 - 819,2}{4} = 73,2$$

$$s_{12} = \frac{(2)(11) + (7)(14) + \dots + (23)(20) - \frac{(64)(81)}{5}}{4} = \frac{1156 - 1036,8}{4} = 29,8$$

$$s_{22} = \frac{11^2 + 14^2 + \dots + 20^2 - \frac{(81)^2}{5}}{4} = \frac{1365 - 1312,2}{4} = 13,2$$

$$S_2 = \begin{bmatrix} 73,2 & 29,8 \\ 29,8 & 13,2 \end{bmatrix}$$

$$S_c = \frac{(5 - 1)S_1 + (5 - 1)S_2}{5 + 5 - 2} = \begin{bmatrix} 70,95 & 26,9 \\ 26,9 & 12,1 \end{bmatrix}$$

$$S_c^{-1} = \frac{1}{134,8850} \begin{bmatrix} 12,1 & -26,9 \\ -26,9 & 70,95 \end{bmatrix}$$

- Cálculo das estatísticas T^2 e F

$$T^2 = \frac{n_1 \cdot n_2}{n_1 + n_2} \left[\bar{\tilde{X}}_1 - \bar{\tilde{X}}_2 \right]' S_c^{-1} \left[\bar{\tilde{X}}_1 - \bar{\tilde{X}}_2 \right]$$

$$F = \frac{n_1 + n_2 - p - 1}{(n_1 + n_2 - 2)p} \cdot T^2, \text{ com } p \text{ e } n_1 + n_2 - p - 1 \text{ graus de liberdade.}$$

$$T^2 = \frac{(5)(5)}{5+5} [1,4 \quad -7,2] \frac{1}{134,8850} \begin{bmatrix} 12,1 & -26,9 \\ -26,9 & 70,95 \end{bmatrix} \begin{bmatrix} 1,4 \\ -7,2 \end{bmatrix}$$

$$T^2 = \frac{25}{10} [1,4 \quad -7,2] \begin{bmatrix} 1,5614 \\ -4,0664 \end{bmatrix} = 78,66$$

$$F = \frac{5+5-2-1}{(5+5-2)2} (78,66) = 34,41^* \quad * (P < 0,05)$$

$$F_{5\%}(2; 7) = 4,74$$

Como o valor calculado $F = 34,41 > 4,74$, o resultado é estatisticamente significativo ao nível de 5% de probabilidade. Assim, rejeita-se H_0 e conclui-se que os vetores de médias das duas subespécies diferem significativamente ($P < 0,05$) pelo teste T^2 de Hotelling.

3.4. EXERCÍCIOS

1. Sejam

$$\tilde{X} = \begin{bmatrix} X_1 \\ X_2 \\ \dots \\ X_p \end{bmatrix}, \quad \mu_X = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \dots \\ \mu_p \end{bmatrix}, \quad \text{Cov}(\tilde{X}) = \sum_X.$$

Para $\tilde{Y} = C' \tilde{X}$, mostrar que:

- i) $E(\tilde{Y}) = C' \mu_X$;
- ii) $\text{Cov}(\tilde{Y}) = \sum_Y = C' \sum_X C$;
- iii) $\text{Cov}(A \tilde{X}, B \tilde{X}) = A \sum_X B'$.

2. Seja $\tilde{X} \sim N_3(\tilde{x} : \mu_X ; \sum_X)$ em que,

$$\tilde{X} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}, \quad \mu_X = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}, \quad \sum_X = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{bmatrix}.$$

Qual é a distribuição de $\tilde{Y} = \begin{bmatrix} X_1 \\ X_3 \end{bmatrix}$?

3. Seja $\tilde{X} \sim N_5(\tilde{x} : \underline{\mu}_X, \underline{\Sigma}_X)$, em que:

$$\tilde{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_3 \\ X_4 \\ X_5 \end{bmatrix} = \begin{bmatrix} \tilde{X}_r \\ \vdots \\ \tilde{X}_s \end{bmatrix}, \quad \underline{\Sigma}_X = \begin{bmatrix} 4 & -1 & 1/2 & 1/2 & 0 \\ -1 & 3 & 1 & -1 & 0 \\ \hline 1/2 & 1 & 6 & 1 & -1 \\ 1/2 & -1 & 1 & 4 & 0 \\ 0 & 0 & -1 & 0 & 2 \end{bmatrix} = \begin{bmatrix} \underline{\Sigma}_{rr} & \underline{\Sigma}_{rs} \\ \hline \underline{\Sigma}_{sr} & \underline{\Sigma}_{ss} \end{bmatrix}.$$

Pede-se:

i) Sendo $\tilde{X}_r = C_1' \tilde{X}$ e $\tilde{X}_s = C_2' \tilde{X}$, mostrar que

$$\text{Cov}(\tilde{X}_r, \tilde{X}_s) = C_1' \underline{\Sigma}_X C_2;$$

ii) Mostrar as matrizes C_1' e C_2' , e calcular $\text{Cov}(\tilde{X}_r, \tilde{X}_s)$;

Observação: Note que $\text{Cov}(\tilde{X}_r, \tilde{X}_s) = \underline{\Sigma}_{rs}$.

iii) Mostrar que $\text{Cov}(A \tilde{X}_r, B \tilde{X}_s) = A \underline{\Sigma}_{rs} B'$;

iv) Para $A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$ e $B = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -2 \end{bmatrix}$, calcular $\text{Cov}(A \tilde{X}_r, B \tilde{X}_s)$.

4. **TEOREMA:** Se $\tilde{x}' = [x_1 \ x_2 \ \dots \ x_p]$ é um vetor aleatório com distribuição normal

multivariada cuja f.d.p. conjunta é $f_{x_1, x_2, \dots, x_p}(x_1, x_2, \dots, x_p) = K \exp(-\frac{1}{2}Q)$, onde Q é uma

forma quadrática, então o vetor de médias $E(\tilde{X}) = \underline{\mu}$ é o vetor que é a solução do sistema de

equações $\frac{\partial Q}{\partial \tilde{X}} = 0$.

Seja $\tilde{X} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}$, $\tilde{X} \sim N_3(\tilde{x} : \underline{\mu}; \underline{\Sigma})$, cuja f.d.p. é

$$f_{x_1, x_2, x_3}(x_1, x_2, x_3) = \frac{1}{(2\pi)^{3/2} |\underline{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}Q\right\},$$

em que:

$$Q = \frac{3}{2}x_1^2 + 2x_2^2 + x_3^2 - 3x_1x_2 + 2x_1x_3 - 2x_2x_3 + 10x_1 - 14x_2 + 8x_3 + 26$$

Pede-se:

- i) Obter a matriz Σ ;
- ii) Obter o vetor de médias $\underline{\mu}$;
- iii) Classificar a forma quadrática $Q = (\underline{x} - \underline{\mu})' \Sigma^{-1} (\underline{x} - \underline{\mu})$;
- iv) Calcular a probabilidade de que $X_1 > -1,08$.

5. Seja $\underline{X} \sim N_4(\underline{x} : \underline{\mu} ; \Sigma)$, em que:

$$\underline{X} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix}, \quad \underline{\mu} = \begin{bmatrix} 1 \\ -1 \\ 2 \\ 0 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} 2 & 0 & 1 & 0 \\ 0 & 3 & 0 & 2 \\ 1 & 0 & 5 & 0 \\ 0 & 2 & 0 & 3 \end{bmatrix}.$$

Pede-se:

- i) Encontrar a distribuição marginal de $\begin{bmatrix} X_1 \\ X_3 \end{bmatrix}$, isto é, obter $f_{X_1, X_3}(x_1, x_3)$;
- ii) Encontrar a distribuição condicional $f_{X_1, X_2 / X_3, X_4}(x_1, x_2 / x_3, x_4)$.

6. Mostrar que $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X_1, X_2}(x_1, x_2) dx_1 dx_2 = 1$, em que:

$$f_{X_1, X_2}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left(\frac{x_1^2}{\sigma_1^2} - \frac{2\rho x_1x_2}{\sigma_1\sigma_2} + \frac{x_2^2}{\sigma_2^2}\right)}$$

SUGESTÃO:

Teorema: $\Sigma = BB'$

$$\begin{array}{c} \text{Op. congruência em A} \\ \left[\begin{array}{c} \Sigma \\ \vdots \\ I \end{array} \right] \sim \dots \sim \dots \sim \left[\begin{array}{c} D \\ \vdots \\ C' \end{array} \right] \\ \text{Op. elementares em I} \end{array}$$

$$B = (C')^{-1} D^{1/2}$$

Usar a transformação: $\underline{\tilde{x}} = B \underline{\tilde{y}}$, em que:

$$\underline{\tilde{x}} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \text{ e } \underline{\tilde{y}} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}.$$

7. Na Tabela 3.3 são apresentados os dados sobre produtividade e altura de plantas de duas variedades de milho (A e B).

Tabela 3.3 - Produtividade e altura de plantas de duas variedades de milho

Variedade A		Variedade B	
Produtividade	Altura de plantas	Produtividade	Altura de plantas
8,7	1,10	6,4	1,90
5,9	1,90	5,5	1,75
6,2	1,98	5,4	1,78
5,8	2,92	4,6	1,79
6,8	2,00	5,9	1,90
6,2	2,01		

i) **Teste de um vetor de médias contra um padrão populacional**

Considerando os dados da variedade A como uma amostra aleatória de uma população normal bivariada, pede-se:

Testar a hipótese $H_0 : \underline{\mu} = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$ contra $H_a : \underline{\mu} \neq \begin{bmatrix} 5 \\ 1 \end{bmatrix}$, através do teste T^2 de Hotelling ao nível de significância $\alpha = 5\%$.

ii) **Teste dos vetores de médias de duas populações**

Supondo populações normais bivariadas e $\sum_A = \sum_B$, testar a hipótese de igualdade dos vetores de médias das duas variedades de milho, pelo teste T^2 de Hotelling, para $\alpha = 1\%$.

8. Os dados apresentados na Tabela 3.4 constituem parte de um experimento com progênies de Cupuaçuzeiro (*Theobroma grandiflorum* (Willd ex Spreng) Schum), obtidos em uma safra, numa fazenda da Universidade Federal do Amazonas, em Manaus-AM, em 1996-97.

Tabela 3.4 – Resultados experimentais*

Progênie A			Progênie B		
PMF	PMC	PFP	PMF	PMC	PFP
1,250	0,559	8,750	0,865	0,443	6,055
1,319	0,512	11,871	0,736	0,316	6,624
1,473	0,525	13,257	1,029	0,488	7,203
1,120	0,620	10,250	0,962	0,480	7,420
1,540	0,540	11,486	1,104	0,384	6,260
1,100	0,480	12,640	0,890	0,420	6,194
1,235	0,551	11,266	0,938	0,409	7,487

* PMF = peso médio de frutos (kg); PMC = peso médio da casca (kg); PFP = produção de frutos por planta (kg)

Pede-se:

i) **Teste de um vetor de médias contra um padrão populacional**

Considerando os dados da progênie A como uma amostra aleatória de uma população normal multivariada, testar a hipótese

$$H_o : \underline{\mu} = \begin{bmatrix} 1,000 \\ 0,500 \\ 10,000 \end{bmatrix} \text{ contra } H_a : \underline{\mu} \neq \begin{bmatrix} 1,000 \\ 0,500 \\ 10,000 \end{bmatrix}$$

através do teste T^2 de Hotelling ao nível de 1% de probabilidade.

ii) **Teste dos vetores de médias de duas populações** (supondo $\Sigma_A = \Sigma_B$)

Testar a hipótese de igualdade dos vetores de médias das duas progênies de Cupuaçuzeiro pelo teste T^2 de Hotelling para $\alpha = 1\%$.

3.5. REFERÊNCIAS

- ANDERSON, T. W. **An introduction to multivariate statistical analysis**. 2nd ed., New York: John Wiley and Sons, 1984. 379p.
- HARRIS, R. J. **A primer of multivariate statistics**. New York: Academic Press, 1975. 332p.
- JOHNSON, R. A.; WICHERN, D.W. **Applied multivariate statistical analysis**. 4th ed., USA, Upper Saddle River, New Jersey: Prentice Hall, Inc., 1998. 816p.
- MARDIA, K. V.; KENT, J. T.; BIBBY, J. M. **Multivariate analysis**. 6th ed., London: Academic Press, 1997. 518p.
- MORRISON, D. F. **Multivariate statistical methods**. 2nd ed., USA: McGraw-Hill Book Company, 1976. 415p.
- RENCHE, A. C. **Methods of multivariate analysis**. 2nd ed., USA: John Wiley and Sons, 2002. 738p.

CAPITULO 4

ANÁLISE DE VARIÂNCIA MULTIVARIADA E VARIÁVEIS CANÔNICAS

4.1. INTRODUÇÃO

Na análise de experimentos que envolvem variáveis aleatórias contínuas, medidas na mesma unidade experimental, pressupondo a multinormalidade, pode-se realizar uma análise multivariada com aplicações de certos testes de significância. Um ponto relevante da análise multivariada é o aproveitamento da informação conjunta das variáveis envolvidas.

A análise de variância multivariada (MANOVA) pode ser feita para qualquer delineamento experimental, sem dificuldade. É preciso trabalhar com todas as variáveis simultaneamente, obtendo para elas somas de quadrados (SQ) e somas de produtos (SP).

As condições para a realização da MANOVA são as seguintes:

- i) Modelo aditivo para efeitos de tratamentos, blocos (se houver) e erro;
- ii) Independência dos erros;
- iii) Igualdade da matriz de covariância Σ para todas as amostras;
- iv) Distribuição normal multivariada dos erros com variância Σ .

A seguir será apresentada uma derivação da análise de variância multivariada considerando o modelo com um fator em blocos completos casualizados.

4.2. ANÁLISE DE VARIÂNCIA MULTIVARIADA EM BLOCOS CASUALIZADOS

4.2.1. Modelo estatístico

O modelo estatístico para um delineamento em blocos casualizados com **b** blocos e **k** tratamentos em que são medidas **p** variáveis, é dado a seguir:

$$y_{ijr} = \mu_r + t_{ir} + b_{jr} + e_{ijr}$$

em que:

$$\begin{array}{l} i = 1, 2, \dots, k \\ j = 1, 2, \dots, b \\ r = 1, 2, \dots, p \end{array} \quad \sum_{i=1}^k t_{ir} = 0 \quad \text{e} \quad \sum_{j=1}^b b_{jr} = 0 \quad (r = 1, 2, \dots, p);$$

y_{ijr} é o valor observado da r -ésima variável, sob o i -ésimo tratamento no j -ésimo bloco;

μ_r é a média geral da r -ésima variável;

t_{ir} é o efeito do i -ésimo tratamento na r -ésima variável;

b_{jr} é o efeito do j -ésimo bloco na r -ésima variável;

e_{ijr} é o efeito aleatório associado à observação y_{ijr} .

Tem-se ainda que, os $\mathbf{e}_{ij} = [e_{ij1}, e_{ij2}, \dots, e_{ijp}]$ têm distribuição multinormal p -dimensional com vetor de médias nulo e uma matriz de covariância Σ comum a todas as combinações i e j . Os \mathbf{e}_{ij} correspondentes às diferentes unidades experimentais em cada bloco são independentemente distribuídos.

Na forma matricial, esse modelo pode ser assim escrito:

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \boldsymbol{\varepsilon}$$

em que:

\mathbf{Y} é a matriz das observações de dimensão $bk \times p$;

\mathbf{X} é a matriz do delineamento de dimensão $bk \times (1+k+b)$;

\mathbf{B} é a matriz dos parâmetros de dimensão $(1+k+b) \times p$;

$\boldsymbol{\varepsilon}$ é a matriz dos erros de dimensão $bk \times p$.

A seguir são apresentadas as matrizes \mathbf{Y} , \mathbf{X} , \mathbf{B} e $\boldsymbol{\varepsilon}$ para o modelo em questão.

$$\mathbf{Y} = \begin{bmatrix} y_{111} & y_{112} & \cdots & y_{11p} \\ y_{121} & y_{122} & \cdots & y_{12p} \\ \cdots & \cdots & \cdots & \cdots \\ y_{1b1} & y_{1b2} & \cdots & y_{1bp} \\ \hline y_{211} & y_{212} & \cdots & y_{21p} \\ y_{221} & y_{222} & \cdots & y_{22p} \\ \cdots & \cdots & \cdots & \cdots \\ y_{2b1} & y_{2b2} & \cdots & y_{2bp} \\ \hline \cdots & \cdots & \cdots & \cdots \\ \hline y_{k11} & y_{k12} & \cdots & y_{k1p} \\ y_{k21} & y_{k22} & \cdots & y_{k2p} \\ \cdots & \cdots & \cdots & \cdots \\ y_{kb1} & y_{kb2} & \cdots & y_{kbp} \end{bmatrix} = \begin{bmatrix} y_{\sim 1} & y_{\sim 2} & \cdots & y_{\sim p} \end{bmatrix}$$

$$X = \begin{bmatrix} 1 & 1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & 1 & 0 & \dots & 0 & 0 & 0 & \dots & 1 \\ \hline 1 & 0 & 1 & \dots & 0 & 1 & 0 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & 0 & 1 & \dots & 0 & 0 & 0 & \dots & 1 \\ \hline \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \hline 1 & 0 & 0 & \dots & 1 & 1 & 0 & \dots & 0 \\ 1 & 0 & 0 & \dots & 1 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & 0 & 0 & \dots & 1 & 0 & 0 & \dots & 1 \end{bmatrix}$$

$$B = \begin{bmatrix} \mu_1 & \mu_2 & \dots & \mu_p \\ \hline t_{11} & t_{12} & \dots & t_{1p} \\ t_{21} & t_{22} & \dots & t_{2p} \\ \dots & \dots & \dots & \dots \\ t_{k1} & t_{k2} & \dots & t_{kp} \\ \hline b_{11} & b_{12} & \dots & b_{1p} \\ b_{21} & b_{22} & \dots & b_{2p} \\ \dots & \dots & \dots & \dots \\ b_{b1} & b_{b2} & \dots & b_{bp} \end{bmatrix} = \begin{bmatrix} \beta_{\sim 1} & \beta_{\sim 2} & \dots & \beta_p \end{bmatrix}$$

$$\mathcal{E} = \begin{bmatrix} e_{111} & e_{112} & \dots & e_{11p} \\ e_{121} & e_{122} & \dots & e_{12p} \\ \dots & \dots & \dots & \dots \\ e_{1b1} & e_{1b2} & \dots & e_{1bp} \\ \hline e_{211} & e_{212} & \dots & e_{21p} \\ e_{221} & e_{222} & \dots & e_{22p} \\ \dots & \dots & \dots & \dots \\ e_{2b1} & e_{2b2} & \dots & e_{2bp} \\ \hline \dots & \dots & \dots & \dots \\ \hline e_{k11} & e_{k12} & \dots & e_{k1p} \\ e_{k21} & e_{k22} & \dots & e_{k2p} \\ \dots & \dots & \dots & \dots \\ e_{kb1} & e_{kb2} & \dots & e_{kbp} \end{bmatrix} = \begin{bmatrix} e_{\sim 11} \\ e_{\sim 12} \\ \dots \\ e_{\sim 1b} \\ \hline e_{\sim 21} \\ e_{\sim 22} \\ \dots \\ e_{\sim 2b} \\ \hline \dots \\ \hline e_{\sim k1} \\ e_{\sim k2} \\ \dots \\ e_{\sim kb} \end{bmatrix} = \begin{bmatrix} e_{\sim 1} & e_{\sim 2} & \dots & e_p \end{bmatrix}$$

logo:

$$Y = \begin{bmatrix} y_1 & y_2 & \cdots & y_p \end{bmatrix} = \begin{bmatrix} X\beta_1 & X\beta_2 & \cdots & X\beta_p \end{bmatrix} + \begin{bmatrix} e_1 & e_2 & \cdots & e_p \end{bmatrix}$$

e, portanto:

$$y_r = X\beta_r + e_r \text{ para a variável } r \text{ (} r = 1, 2, \dots, p \text{)}.$$

4.2.2. Matrizes de sistemas de equações normais

Considerando o modelo linear multivariado na forma

$$Y = XB + \varepsilon,$$

tem-se a matriz de sistemas de equações normais

$$X'XB^0 = X'Y$$

ou

$$X'X \begin{bmatrix} \beta_1^0 & \beta_2^0 & \cdots & \beta_p^0 \end{bmatrix} = X' \begin{bmatrix} y_1 & y_2 & \cdots & y_p \end{bmatrix}$$

em que

$$B^0 = (X'X)^{-} X'Y = \begin{bmatrix} \beta_1^0 & \beta_2^0 & \cdots & \beta_p^0 \end{bmatrix} \quad (\text{solução geral})$$

é a matriz das soluções de mínimos quadrados de $X'XB^0 = X'Y$, de dimensão $(1+k+b) \times p$,

$X'X$ é a mesma matriz do modelo univariado;

$(X'X)^{-}$ é uma inversa condicional de $X'X$;

$$\beta_r^{0'} = \begin{bmatrix} \mu_r^0 & t_{1r}^0 & \cdots & t_{kr}^0 & b_{1r}^0 & \cdots & b_{br}^0 \end{bmatrix};$$

$$X'Y = \begin{bmatrix} G_1 & G_2 & \cdots & G_p \\ \hline T_{11} & T_{12} & \cdots & T_{1p} \\ T_{21} & T_{22} & \cdots & T_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ \hline T_{k1} & T_{k2} & \cdots & T_{kp} \\ \hline B_{11} & B_{12} & \cdots & B_{1p} \\ B_{21} & B_{22} & \cdots & B_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ B_{b1} & B_{b2} & \cdots & B_{bp} \end{bmatrix} = \begin{bmatrix} X'y_1 & X'y_2 & \cdots & X'y_p \end{bmatrix}; e$$

$$G_r = \sum_{i=1}^k \sum_{j=1}^b y_{ijr}, \quad T_{ir} = \sum_{j=1}^b y_{ijr} \quad e \quad B_{jr} = \sum_{i=1}^k y_{ijr}.$$

Da mesma forma que no modelo univariado, sendo $\text{posto}(X'X) = k + b - 1$, o sistema admite infinitas soluções. Impondo as restrições

$$\sum_{i=1}^k t_{ir}^0 = 0 \text{ e } \sum_{j=1}^b b_{jr}^0 = 0 \quad (r=1, 2, \dots, p)$$

ou na forma matricial, para todo r ,

$$A' \beta_r^0 = \begin{bmatrix} 0 & 1 & 1 & \dots & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 & 1 & 1 & \dots & 1 \end{bmatrix} \beta_r^0 = 0,$$

o sistema terá solução única que será denotada por $\hat{\beta}_r$. Assim:

$$\hat{\beta}_r = (X'X + AA')^{-1} X' y_r,$$

onde $(X'X + AA')^{-1}$ é a mesma matriz do modelo univariado, sendo também uma inversa condicional de $X'X$. Então:

$$\hat{\beta}_r' = [\hat{\mu}_r \quad \hat{t}_{1r} \quad \dots \quad \hat{t}_{kr} \quad \hat{b}_{1r} \quad \dots \quad \hat{b}_{br}]$$

sendo:

$$\hat{\mu}_r = \frac{G_r}{kb}, \quad \hat{t}_{ir} = \frac{T_{ir}}{b} - \frac{G_r}{kb} \text{ e } \hat{b}_{jr} = \frac{B_{jr}}{k} - \frac{G_r}{kb}.$$

4.2.3. Estimabilidade

Para o modelo em questão, sejam os vetores c' e β_r' , ambos de dimensão $1 \times (1+k+b)$,

tal que:

$$c' = \begin{bmatrix} c_0 & c_1 & c_2 & \dots & c_k & c_{11} & c_{12} & \dots & c_b \end{bmatrix}$$

$$\beta_r' = \begin{bmatrix} \mu_r & t_{1r} & t_{2r} & \dots & t_{kr} & b_{1r} & b_{2r} & \dots & b_{br} \end{bmatrix}$$

Como no modelo univariado, um teste prático para verificar a estimabilidade da função paramétrica $c'\beta_r$ é o seguinte:

Se $c'H = c'$, para $H = (X'X)^- X'X$, então, $c'\beta_r$ é estimável; uma outra maneira, ainda mais

prática: se $c_0 = \sum_{i=1}^k c_i = \sum_{j=1}^b c_j$, então, $c'\beta_r$ é estimável. Por outro lado, o melhor estimador

linear imparcial (BLUE = "Best Linear Unbiased Estimator") de $c'\beta_r$ é dado por $c'\beta_r^0$. É

interessante lembrar que combinações lineares de funções estimáveis são estimáveis.

Da mesma forma que no modelo univariado, com as restrições impostas aos parâmetros e as mesmas restrições nas soluções (modelo restrito), todos os parâmetros tornam-se estimáveis e, portanto:

$$\hat{B} = (X'X + AA')^{-1} X'Y = \begin{bmatrix} \hat{\beta}_{\tilde{1}} & \hat{\beta}_{\tilde{2}} & \cdots & \hat{\beta}_{\tilde{p}} \end{bmatrix},$$

em que

$$\hat{\beta}_{\tilde{r}} = (X'X + AA')^{-1} X'y_{\tilde{r}} = [\hat{\mu}_r \quad \hat{t}_{1r} \quad \cdots \quad \hat{t}_{kr} \quad \hat{b}_{1r} \quad \cdots \quad \hat{b}_{br}]',$$

é o BLUE de B.

4.2.4. Decomposição das somas de quadrados e de produtos totais

Uma vez que $Y = XB + \varepsilon$, pelo método dos mínimos quadrados, ter-se-á:

$$\hat{\varepsilon}'\hat{\varepsilon} = Y'Y - \hat{B}'X'Y$$

isto é, uma matriz de somas de quadrados e de produtos residuais, que será denotada por **E**, com $n_e = n - \text{posto}(X)$ graus de liberdade. No modelo em questão, $n_e = kb - (k+b-1) = kb - k - b + 1 = (k-1)(b-1)$.

$$\begin{aligned} \hat{B}'X'Y &= \begin{bmatrix} \hat{\beta}_{\tilde{1}}' \\ \hat{\beta}_{\tilde{2}}' \\ \cdots \\ \hat{\beta}_{\tilde{p}}' \end{bmatrix} \begin{bmatrix} X'y_1 & X'y_2 & \cdots & X'y_p \end{bmatrix} \\ &= \begin{bmatrix} \hat{\beta}_{\tilde{1}}' X'y_1 & \hat{\beta}_{\tilde{1}}' X'y_2 & \cdots & \hat{\beta}_{\tilde{1}}' X'y_p \\ \hat{\beta}_{\tilde{2}}' X'y_1 & \hat{\beta}_{\tilde{2}}' X'y_2 & \cdots & \hat{\beta}_{\tilde{2}}' X'y_p \\ \cdots & \cdots & \cdots & \cdots \\ \hat{\beta}_{\tilde{p}}' X'y_1 & \hat{\beta}_{\tilde{p}}' X'y_2 & \cdots & \hat{\beta}_{\tilde{p}}' X'y_p \end{bmatrix} \end{aligned}$$

em que: $\hat{B}'X'Y$ é uma matriz de somas de quadrados e de produtos de parâmetros com $\text{posto}(X)$ graus de liberdade;

$$X'y_{\tilde{r}} = [G_r \quad T_{1r} \quad \cdots \quad T_{kr} \quad B_{1r} \quad \cdots \quad B_{br}]';$$

$$\hat{\beta}_{\tilde{r}}' X'y_{\tilde{r}} = \text{SQParâmetros}_r = \frac{G_r^2}{kb} + \frac{1}{b} \sum_{i=1}^k T_{ir}^2 - \frac{G_r^2}{kb} + \frac{1}{k} \sum_{j=1}^b B_{jr}^2 - \frac{G_r^2}{kb};$$

$$\hat{\beta}'_{\tilde{r}} X'_{\tilde{s}} y_s = SPParâmetros_{rs} = \frac{G_r G_s}{kb} + \frac{1}{b} \sum_{i=1}^k T_{ir} T_{is} - \frac{G_r G_s}{kb} + \frac{1}{k} \sum_{j=1}^b B_{jr} B_{js} - \frac{G_r G_s}{kb};$$

e $r, s = 1, 2, \dots, p$.

De forma semelhante ao modelo univariado, as somas de quadrados são dadas por:

$SQTotal_r = SQBlocos_r + SQTratamentos_r + SQResíduo_r$, em que:

$$SQTotal_r = \sum_{i=1}^k \sum_{j=1}^b y_{ijr}^2 - \frac{G_r^2}{kb},$$

$$SQBlocos_r = \frac{1}{k} \sum_{j=1}^b B_{jr}^2 - \frac{G_r^2}{kb},$$

$$SQTratamentos_r = \frac{1}{b} \sum_{i=1}^k T_{ir}^2 - \frac{G_r^2}{kb}.$$

As somas de produtos são dadas por:

$SPTotal_{r,s} = SPBlocos_{r,s} + SPTratamentos_{r,s} + SPResíduo_{r,s}$, em que:

$$SPTotal_{r,s} = \sum_{i=1}^k \sum_{j=1}^b y_{ijr} \cdot y_{ijs} - \frac{G_r G_s}{kb}$$

$$SPBlocos_{r,s} = \frac{1}{k} \sum_{j=1}^b B_{jr} B_{js} - \frac{G_r G_s}{kb}$$

$$SPTratamentos_{r,s} = \frac{1}{b} \sum_{i=1}^k T_{ir} T_{is} - \frac{G_r G_s}{kb}$$

logo, $\mathbf{A} = \mathbf{B} + \mathbf{H} + \mathbf{E}$, onde \mathbf{A} , \mathbf{B} , \mathbf{H} e \mathbf{E} são matrizes $p \times p$, de somas de quadrados e de produtos, sendo \mathbf{A} referente a totais, \mathbf{B} a blocos, \mathbf{H} a tratamentos e \mathbf{E} ao resíduo.

Na MANOVA, a hipótese de real interesse a ser testada, é a de igualdade dos vetores de médias de tratamentos, isto é, $H_0 : \mu_1 = \mu_2 = \dots = \mu_k$, ou ainda:

$$H_0 : \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \dots \\ \mu_{1p} \end{bmatrix} = \begin{bmatrix} \mu_{21} \\ \mu_{22} \\ \dots \\ \mu_{2p} \end{bmatrix} = \dots = \begin{bmatrix} \mu_{k1} \\ \mu_{k2} \\ \dots \\ \mu_{kp} \end{bmatrix}.$$

A hipótese H_0 , considerando o modelo restrito, também pode ser colocada da seguinte maneira: os vetores de efeitos de tratamentos são nulos, isto é, $H_0 : t_1 = t_2 = \dots = t_k = 0$, ou ainda:

$$H_0 : \begin{bmatrix} t_{11} \\ t_{12} \\ \dots \\ t_{1p} \end{bmatrix} = \begin{bmatrix} t_{21} \\ t_{22} \\ \dots \\ t_{2p} \end{bmatrix} = \dots = \begin{bmatrix} t_{k1} \\ t_{k2} \\ \dots \\ t_{kp} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \dots \\ 0 \end{bmatrix}.$$

No modelo linear geral multivariado, a hipótese H_0 pode ser expressa da seguinte maneira:

$$H_0 = C'BW = \phi.$$

Uma vez que na MANOVA, para testar $H_0 : t_1 = t_2 = \dots = t_k = \underline{0}$, W é uma matriz identidade de ordem p , pode-se escrever:

$$H_0 = C'B = \phi$$

em que C' , da mesma forma que no modelo univariado, é uma matriz de posto linha completo, constituída de funções estimáveis. No problema em questão, a matriz C' , para testar a igualdade dos vetores de médias dos k tratamentos, pode ser expressa de diversas maneiras. Uma delas de ordem $(k-1) \times (1+k+b)$, pode ser:

$$C' = \begin{bmatrix} 0 & 1 & -1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & -1 & \dots & 0 & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 1 & 0 & 0 & \dots & -1 & 0 & 0 & \dots & 0 \end{bmatrix}.$$

A matriz de somas de quadrados e de produtos devida à hipótese testada é a matriz H de tratamentos, que matricialmente é dada por:

i) Para $B^0 = (X'X)^{-1}X'Y$:

$$H = [C'B^0]'[C'(X'X)^{-1}C]^{-1}[C'B^0].$$

ii) Para $\hat{B} = (X'X + AA')^{-1}X'Y$:

$$H = [C'\hat{B}]'[C'(X'X + AA')^{-1}C]^{-1}[C'\hat{B}].$$

Fato: H é invariante para qualquer solução do sistema de equações normais.

A matriz A de somas de quadrados e somas de produtos totais com $n-1$ graus de liberdade, é dada por:

$$A = Y' \left[I_{(n)} - \frac{1}{n} \mathbf{1} \mathbf{1}' \right] Y$$

$$A = Y'Y - \frac{1}{n} Y' \mathbf{1} \mathbf{1}' Y, \quad \mathbf{1} \text{ é um vetor de } 1\text{'s } n \times 1$$

A matriz E de somas de quadrados e somas de produtos residual com $n - \text{posto}[X] = n_e$ graus de liberdade, é dada por:

$$E = Y' \left[I_{(n)} - X(X'X)^{-1}X' \right] Y$$

$$E = Y'Y - \hat{B}'X'Y$$

Na Tabela 4.1 é apresentado o esquema da análise de variância multivariada para o problema em questão.

Tabela 4.1 – Esquema da análise de variância multivariada para um experimento com um fator em blocos completos casualizados

Causas de variação	G.L.*	Matrizes de somas de quadrados e de produtos
Blocos	b - 1	B
Tratamentos	k - 1	H
Resíduo	$n_e = (b-1)(k-1)$	E
Total	bk-1	A

* $n = bk$ e $n_e = n - \text{posto}(X)$.

Com o desenvolvimento apresentado anteriormente, ficou evidente que uma MANOVA pode ser feita facilmente utilizando somatório ou matrizes. A derivação para o modelo com um fator inteiramente casualizado pode ser feita de modo análogo ao de blocos casualizados, basta suprimir o efeito de blocos b_{jr} do modelo anterior.

O esquema da análise de variância multivariada para um experimento com um fator inteiramente casualizado com **k** tratamentos, **r** repetições e **p** variáveis está apresentado na Tabela 4.2, cujo modelo estatístico é dado por:

$$y_{ijl} = \mu_l + t_{il} + e_{ijl}, \quad i = 1, 2, \dots, k; \quad j = 1, 2, \dots, r; \quad l = 1, 2, \dots, p$$

Tabela 4.2 – Esquema da análise de variância multivariada para um experimento com um fator inteiramente casualizado

Causas de variação	G.L.*	Matrizes de somas de quadrados e de produtos
Tratamentos	k-1	H
Resíduo	$n_e = n - k$	E
Total	n - 1	A

* $n = kr$ e $n_e = n - \text{posto}(X)$.

A seguir, é apresentada uma MANOVA com os cálculos realizados de uma maneira bastante simples e em detalhes, considerando dados de um experimento inteiramente casualizado com 8 tratamentos, 2 repetições e 2 variáveis (Tabela 4.3). Após o exemplo numérico, serão apresentados os quatro testes de hipóteses, isto é:

- i) Teste de Wilks;
- ii) Teste de Pillai;

iii) Teste de Hotelling-Lawley;

iv) Teste de Roy.

Tabela 4.3 – Resultados parciais de um experimento com aves de postura avaliadas na 10ª semana de idade

Tratamentos	Repetições	Variáveis*	
		X	Y
1	1	402,18	76,00
	2	397,55	76,55
2	1	421,76	91,08
	2	436,54	121,08
3	1	435,67	134,42
	2	417,31	145,78
4	1	381,81	87,70
	2	382,22	103,45
5	1	381,68	122,67
	2	350,94	108,50
6	1	442,53	66,49
	2	430,09	91,55
7	1	390,37	116,62
	2	413,48	126,92
8	1	416,18	104,47
	2	469,85	166,46

* X = Consumo de ração (g) e Y = Ganho de peso (g).

Todos os cálculos necessários para elaboração da Tabela da MANOVA estão apresentados a seguir.

k = 8 tratamentos; r = 2 repetições; p = 2 variáveis; n = 16 observações por variável

$$\sum X = 402,18 + 397,55 + \dots + 469,85 = 6570,16$$

$$\sum X^2 = (402,18)^2 + (397,55)^2 + \dots + (469,85)^2 = 2711038,3908$$

$$SQ_{Total}(X) = \sum X^2 - \frac{(\sum X)^2}{n}$$

$$= 2711038,3908 - \frac{(6570,16)^2}{16}$$

$$= 2711038,3908 - 2697937,6515 = 13100,7393$$

$$\sum Y = 76,00 + 76,55 + \dots + 166,46 = 1739,74$$

$$\sum Y^2 = (76,00)^2 + (76,55)^2 + \dots + (166,46)^2 = 200259,8974$$

$$\begin{aligned} \text{SQTotal}(Y) &= \sum Y^2 - \frac{(\sum Y)^2}{n} \\ &= 200259,8974 - \frac{(1739,74)^2}{16} \\ &= 200259,8974 - 189168,4542 = 11091,4432 \end{aligned}$$

$$\sum XY = (402,18)(76,00) + (397,55)(76,55) + \dots + (469,85)(166,46) = 718081,5271$$

$$\begin{aligned} \text{SPTotal}(X, Y) &= \sum XY - \frac{(\sum X)(\sum Y)}{n} \\ &= 718081,5271 - \frac{(6570,16)(1739,74)}{16} \\ &= 3683,3922 \end{aligned}$$

A matriz de somas de quadrados e de produtos total é:

$$A = \begin{bmatrix} \text{SQTotal}(X) & \text{SPTotal}(X, Y) \\ \text{SPTotal}(X, Y) & \text{SQTotal}(Y) \end{bmatrix} = \begin{bmatrix} 13100,7393 & 3683,3922 \\ 3683,3922 & 11091,4432 \end{bmatrix},$$

com $n - 1 = 16 - 1 = 15$ graus de liberdade.

A matriz de somas de quadrados e de produtos de tratamentos pode ser obtida facilmente a partir dos dados apresentados na Tabela 4.4.

Tabela 4.4 – Totais de tratamentos relativos ao consumo de ração (X) e ganho de peso (Y)

Tratamentos	Totais	
	X	Y
1	799,73	152,55
2	858,30	212,16
3	852,98	280,20
4	764,03	191,15
5	732,62	231,17
6	872,62	158,04
7	803,85	243,54
8	886,03	270,93
	6570,16	1739,74

$$\text{SQT}(X) = \frac{1}{2} [(799,73)^2 + (858,30)^2 + \dots + (886,03)^2] - \frac{(6570,16)^2}{16}$$

$$= \frac{1}{2}(5416985,3964) - 2697937,6515 = 10555,0467$$

$$\begin{aligned} \text{SQT}(Y) &= \frac{1}{2}[(152,55)^2 + (212,16)^2 + \dots + (270,93)^2] - \frac{(1739,74)^2}{16} \\ &= \frac{1}{2}(394464,7376) - 189168,4542 = 8063,9146 \end{aligned}$$

A soma de produtos de tratamentos é dada por:

$$\begin{aligned} \text{SPT}(X, Y) &= \frac{1}{2}[(799,73)(152,55) + (858,30)(212,16) + \dots + (886,03)(270,93)] \\ &\quad - \frac{(6570,16)(1739,74)}{16} \\ &= \frac{1}{2}(1432235,4371) - 714398,1348 = 1719,5837. \end{aligned}$$

A matriz de somas de quadrados e de produtos de tratamentos é, pois:

$$H = \begin{bmatrix} \text{SQT}(X) & \text{SPT}(X, Y) \\ \text{SPT}(X, Y) & \text{SQT}(Y) \end{bmatrix} = \begin{bmatrix} 10555,0467 & 1719,5837 \\ 1719,5837 & 8063,9146 \end{bmatrix},$$

com $q = k - 1 = 8 - 1 = 7$ graus de liberdade.

A matriz correspondente ao resíduo se obtém por subtração.

$$\begin{aligned} \text{SQR}(X) &= \text{SQTotal}(X) - \text{SQT}(X) \\ &= 13100,7393 - 10555,0467 = 2545,6926, \end{aligned}$$

$$\begin{aligned} \text{SQR}(Y) &= \text{SQTotal}(Y) - \text{SQT}(Y) \\ &= 11091,4432 - 8063,9146 = 3027,5286, \end{aligned}$$

$$\begin{aligned} \text{SPR}(X, Y) &= \text{SPTTotal}(X, Y) - \text{SPT}(X, Y) \\ &= 3683,3922 - 1719,5837 = 1963,8085. \end{aligned}$$

Logo,

$$E = \begin{bmatrix} 2545,6926 & 1963,8085 \\ 1963,8085 & 3027,5286 \end{bmatrix},$$

com $n_e = k(r-1) = kr - k = n - k = 16 - 8 = 8$ graus de liberdade.

Note que a matriz E poderia ser obtida diretamente como a seguir:

$$\begin{aligned} E = A - H &= \begin{bmatrix} 13100,7393 & 3683,3922 \\ 3683,3922 & 11091,4432 \end{bmatrix} - \begin{bmatrix} 10555,0467 & 1719,5837 \\ 1719,5837 & 8063,9146 \end{bmatrix} \\ &= \begin{bmatrix} 2545,6926 & 1963,8085 \\ 1963,8085 & 3027,5286 \end{bmatrix} \end{aligned}$$

O resultado da análise de variância multivariada (no caso bivariada) está apresentado na Tabela 4.5.

Tabela 4.5 – Análise de variância multivariada para os dados de consumo de ração (X) e ganho de peso (Y)

Causas de variação	G.L.	Matrizes de somas de quadrados e de produtos
Tratamentos	q = 7	$H = \begin{bmatrix} 10555,0467 & 1719,5837 \\ 1719,5837 & 8063,9146 \end{bmatrix}$
Resíduo	n _e = 8	$E = \begin{bmatrix} 2545,6926 & 1963,8085 \\ 1963,8085 & 3027,5286 \end{bmatrix}$
Total	15	$A = \begin{bmatrix} 13100,7393 & 3683,3922 \\ 3683,3922 & 11091,4432 \end{bmatrix}$

É interessante observar que, se o delineamento fosse o de blocos casualizados, além das matrizes A (Total) e H (Tratamentos), seria necessário calcular a matriz B (Blocos) e, neste caso, o número de graus de liberdade seria 1, 7 e 7, para blocos, tratamentos e resíduo, respectivamente.

4.3. TESTES DE HIPÓTESES

Para o modelo $Y = XB + \varepsilon$, a hipótese linear geral multivariada é:

$$H_0 = C'BW = \phi$$

com C' de posto linha completo, e ainda:

$$p = \text{posto}(W)$$

$$q = \text{posto}(C')$$

$$w = \max [\text{posto}(W), \text{posto}(C')]$$

$$H = [C'B^0W]'[C'(X'X)^-C]^{-1}[C'B^0W]$$

$$H = W'[C'B^0]'[C'(X'X)^-C]^{-1}[C'B^0]W$$

$$E = W'[Y'Y - B^0'X'Y]W$$

Uma vez que na MANOVA, W é uma matriz identidade de ordem p , pode-se escrever:

$$H_0 = C'B = \phi,$$

e assim,

$$H = [C'B^0]'[C'(X'X)^-C]^{-1}[C'B^0]$$

$$E = [Y'Y - B^0'X'Y]$$

Define-se ainda:

$$s = \min [\text{posto}(W), \text{posto}(C')] = \min(p, q)$$

$$m' = \frac{1}{2}(|p - q| - 1)$$

$$n' = \frac{1}{2}(n_e - p - 1)$$

4.3.1. Teste de Wilks

Na análise de variância multivariada, a hipótese de real interesse a ser testada é a de igualdade dos vetores de médias de tratamentos, isto é, $H_0 : \mu_1 = \mu_2 = \dots = \mu_k$, ou ainda,

$$H_0 : \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \dots \\ \mu_{1p} \end{bmatrix} = \begin{bmatrix} \mu_{21} \\ \mu_{22} \\ \dots \\ \mu_{2p} \end{bmatrix} = \dots = \begin{bmatrix} \mu_{k1} \\ \mu_{k2} \\ \dots \\ \mu_{kp} \end{bmatrix}$$

O teste de Wilks é um teste de significância muito usado na MANOVA; a estatística do teste é indicada pela letra grega Λ (lambda maiúsculo), assim definida:

$$\Lambda = \frac{\det(E)}{\det(H + E)} = \frac{|E|}{|H + E|}$$

Na presença de diferenças sistemáticas entre tratamentos, espera-se sempre obter $\Lambda < 1$; quanto menor o valor de Λ , mais significativo ele será.

Para avaliar a significância do valor de Λ calculado, pode-se usar a tabela própria para o teste de Wilks, ou, o que é mais comum, transformar o valor de Λ num valor correspondente de F e usar as tabelas de F já conhecidas.

O valor de Λ tabelado para o teste de Wilks é função de α , p, q e n_e .

$$\Lambda_{\text{Tabelado}} \rightarrow \Lambda_{(\alpha; p; q; n_e)}$$

Regra decisória: rejeita-se H_0 em nível de significância α se $\Lambda_{\text{calculado}} < \Lambda_{\text{Tabelado}}$. Caso contrário, não se rejeita H_0 .

A transformação de Λ em F se faz por meio das fórmulas a seguir.

Caso A: $p=2$, q e n_e quaisquer:

$$F_0 = \frac{n_e - 1}{q} \cdot \frac{1 - \sqrt{\Lambda}}{\sqrt{\Lambda}} \sim F[2q; 2(n_e - 1)]$$

Caso B: $q=1$, p e n_e quaisquer:

$$F_0 = \frac{n_e - p + 1}{p} \cdot \frac{1 - \Lambda}{\Lambda} \sim F[p; n_e - p + 1]$$

Caso C: $q=2$, p e n_e quaisquer:

$$F_0 = \frac{n_e - p + 1}{p} \cdot \frac{1 - \sqrt{\Lambda}}{\sqrt{\Lambda}} \sim F[2p; 2(n_e - p + 1)]$$

Caso geral: (HARRIS, 1975):

$$F_0 = \left[\left(\frac{1}{\Lambda} \right)^{1/b} - 1 \right] \frac{ab - c}{pq} \sim F(pq; ab - c)$$

ou

$$F_0 = \left[\frac{1 - \Lambda^{1/b}}{\Lambda^{1/b}} \right] \frac{ab - c}{pq} \sim F(pq; ab - c)$$

em que:

$$a = n_e - \frac{1}{2}(p - q + 1)$$

$$c = \frac{1}{2}(pq - 2)$$

$$b = \sqrt{\frac{p^2 q^2 - 4}{p^2 + q^2 - 5}}, \text{ se } (p^2 + q^2 - 5) > 0 \text{ ou } 1 \text{ caso contrário.}$$

A distribuição é exata se $\min(p, q) \leq 2$ (RAO, 1973, p. 556).

Como ilustração, o teste de Wilks aplicado no exemplo discutido anteriormente, virá:

$$\begin{aligned} |E| &= \begin{vmatrix} 2545,6926 & 1963,8085 \\ 1963,8085 & 3027,5286 \end{vmatrix} \\ &= (2545,6926)(3027,5286) - (1963,8085)(1963,8085) = 3850613,3287 \end{aligned}$$

No delineamento com um fator inteiramente casualizado, $H + E = A$. Assim,

$$\begin{aligned} |H + E| &= |A| = \begin{vmatrix} 13100,7393 & 3683,3922 \\ 3683,3922 & 11091,4432 \end{vmatrix} \\ &= (13100,7393)(11091,4432) - (3683,3922)(3683,3922) = 131738727,7300 \end{aligned}$$

$$\Lambda = \frac{3850613,3287}{131738727,7300} = 0,0292^{**}$$

Para $\alpha = 1\%$ e usando o **Caso A**, segue que:

$$F_0 = \frac{8-1}{7} \cdot \frac{1-\sqrt{0,0292}}{\sqrt{0,0292}} = 4,85^{**} \quad (** P < 0,01)$$

$$F_{1\%}(14; 14) = 3,70$$

Pela Tabela de Wilks, segue que: $\Lambda(1\%; 2; 7; 8) = 0,045317$. Como $\Lambda_{\text{calculado}} < \Lambda_{\text{Tabelado}}$, rejeita-se H_0 .

Verifica-se, pois, que o resultado é significativo ao nível de 1% de probabilidade. Assim, é rejeitada a hipótese de igualdade dos vetores de médias de tratamentos, cujas estimativas estão apresentadas na Tabela 4.6.

Tabela 4.6. Estimativas para os vetores de médias de tratamentos

Variáveis	Tratamentos							
	1	2	3	4	5	6	7	8
X	[399,865]	[429,150]	[426,490]	[382,015]	[366,310]	[436,310]	[401,925]	[443,015]
Y	[76,275]	[106,080]	[140,100]	[95,575]	[115,585]	[79,020]	[121,770]	[135,465]

X = Consumo de ração (g) e Y = Ganho de peso (g)

Observe que a estimativa para o vetor de médias do tratamento 1, inclui as médias obtidas para consumo de ração e ganho de peso, e é dado por:

$$\hat{\tilde{m}}_1 = \hat{\tilde{\mu}}_1 = \begin{bmatrix} 399,865 \\ 76,275 \end{bmatrix}.$$

4.3.2. Teste de Pillai

Define-se a estatística:

$$V = \text{traço} \left[H(H + E)^{-1} \right]$$

V se relaciona com a distribuição F, aproximadamente, por:

$$F_0 = \frac{2n'+s+1}{2m'+s+1} \cdot \frac{V}{s-V} \sim F(n_1; n_2),$$

em que:

$$n_1 = s(2m'+s+1) \quad \text{e} \quad n_2 = s(2n'+s+1).$$

Regra decisória: rejeita-se H_0 em nível de significância α se $F_0 > F_\alpha(n_1; n_2)$.

Como ilustração, o teste de Pillai aplicado no exemplo discutido anteriormente, virá:

$$p = 2, q = 7, s = 2, m' = 2, n' = 2,5$$

$$n_1 = 2[2(2) + 2 + 1] = 14$$

$$n_2 = 2[2(2,5) + 2 + 1] = 16$$

$$V = \text{traço} \left\{ \begin{bmatrix} 10555,0467 & 1719,5837 \\ 1719,5837 & 8063,9146 \end{bmatrix} \begin{bmatrix} 8,4192 \cdot 10^{-5} & -2,7959 \cdot 10^{-5} \\ -2,7959 \cdot 10^{-5} & 9,9444 \cdot 10^{-5} \end{bmatrix} \right\}$$

$$V = \text{traço} \begin{bmatrix} 0,8406 & -0,1241 \\ -0,0807 & 0,7538 \end{bmatrix} = 1,5944$$

$$F_0 = \frac{2(2,5) + 2 + 1}{2(2) + 2 + 1} \cdot \frac{1,5944}{2 - 1,5944} = 4,49^{**} \quad (** P < 0,01)$$

$$F_{1\%}(14; 16) = 3,45$$

Rejeita-se H_0 em nível de 1% de probabilidade.

4.3.3. Teste de Hotelling-Lawley

Define-se a estatística:

$$U = \text{traço}(E^{-1}H).$$

U se relaciona com a distribuição F, aproximadamente por (RENCHE, 2002):

1º CASO:

$$F_1 = \frac{2(sn'+1)}{s^2(2m'+s+1)} \cdot U \sim F[s(2m'+s+1), 2(sn'+1)]$$

2º CASO:

$$F_2 = \frac{U}{c} \sim F(a, b),$$

em que:

$$a = pq \quad b = 4 + \frac{a+2}{D-1} \quad c = \frac{a(b-2)}{b(n_e - p - 1)} \quad D = \frac{(n_e + q - p - 1)(n_e - 1)}{(n_e - p - 3)(n_e - p)}$$

3º CASO:

$$F_3 = \frac{[s(n_e - q - 1) + 1]}{spq} \cdot U \sim F[pq, s(n_e - q - 1)]$$

Fato: Versões mais antigas do **SAS** usavam o 1º CASO. A versão 9.1 usa o 2º CASO para $n' > 0$ e o 1º CASO para $n' \leq 0$.

Regra decisória: rejeita-se H_0 em nível de significância α se $F_{\text{calculado}} > F_{\text{Tabelado}}$. Caso contrário, não se rejeita H_0 .

Como ilustração, o teste de Hotelling-Lawley (**2º caso**) aplicado no exemplo discutido anteriormente, virá:

$$U = \text{traço} \left\{ \begin{bmatrix} 78,6245 \cdot 10^{-5} & -50,9998 \cdot 10^{-5} \\ -50,9998 \cdot 10^{-5} & 66,1113 \cdot 10^{-5} \end{bmatrix} \begin{bmatrix} 10555,0467 & 1719,5837 \\ 1719,5837 & 8063,9146 \end{bmatrix} \right\}$$

$$U = \text{traço} \begin{bmatrix} 7,4219 & -2,7606 \\ -4,2462 & 4,4542 \end{bmatrix} = 11,8761$$

$$p = 2 \quad q = 7 \quad n_e = 8$$

$$a = 2(7) = 14$$

$$D = \frac{(8 + 7 - 2 - 1)(8 - 1)}{(8 - 2 - 3)(8 - 2)} = \frac{42}{9}$$

$$b = 4 + \frac{14 + 2}{(42/9) - 1} = \frac{92}{11} \cong 8,3636$$

$$c = \frac{14[(92/11) - 2]}{(92/11)(8 - 2 - 1)} = \frac{980}{460} \cong 2,1304$$

$$F_0 = \frac{11,8761}{2,1304} = 5,57^{**} \quad (** P < 0,01)$$

$$F_{1\%}(14; 8,3636) = 5,33$$

Rejeita-se H_0 em nível de 1% de probabilidade.

4.3.4. Teste de Roy

Para o teste de Roy, inicialmente determina-se os autovalores da matriz $E^{-1}H$, isto é:

$$|E^{-1}H - \lambda I| = 0 \quad \text{ou} \quad |H - \lambda E| = 0.$$

Sendo λ_{\max} o maior autovalor de $E^{-1}H$, define-se a estatística:

$$\theta_0 = \frac{\lambda_{\max}}{1 + \lambda_{\max}}$$

λ_{\max} se relaciona com a distribuição F, aproximadamente, por:

$$F_0 = \lambda_{\max} \cdot \frac{(n_e - w + q)}{w} \sim F(w; n_e - w + q)$$

Esta aproximação tende a fornecer resultados demasiadamente significativos, indicando uma superestimação do teste F.

Regra decisória: rejeita-se H_0 em nível de significância α se $F_0 > F_{\alpha}(w; n_e - w + q)$.

Como ilustração, o teste de Roy aplicado no exemplo discutido anteriormente, virá:

$$|E^{-1}H - \lambda I| = \begin{vmatrix} 7,4219 - \lambda & -2,7606 \\ -4,2462 & 4,4542 - \lambda \end{vmatrix} = 0$$

$$(7,4219 - \lambda)(4,4542 - \lambda) - (-4,2462)(-2,7606) = 0$$

$$\lambda^2 - 11,8761\lambda + 21,3366 = 0$$

$$\lambda = \frac{11,8761 \pm \sqrt{(-11,8761)^2 - 4(1)(21,3366)}}{2} = \frac{11,8761 \pm 7,4629}{2}$$

$$\lambda_1 = 9,6695 \quad \text{e} \quad \lambda_2 = 2,2066$$

Este mesmo resultado seria obtido se fosse usada a equação $|H - \lambda E| = 0$.

$$F_0 = 9,6695 \cdot \frac{(8 - 7 + 7)}{7} = 11,05^{**} \quad (**P < 0,01)$$

$$F_{1\%}(7; 8) = 6,18$$

Rejeita-se H_0 em nível de 1% de probabilidade.

Outra alternativa para avaliar a significância do valor de θ_0 , é compará-lo com o limite de significância dados por ábacos e tabelas incluídas nos livros de Harris (1975) e Morrison (1976), entre outros. Os parâmetros da distribuição de θ_0 sob a hipótese nula são s , m' e n' , conforme definidos anteriormente.

O valor de θ tabelado para o teste de Roy é função de α , s , m' e n' .

$$\theta_{\text{Tabelado}} \rightarrow \theta_{\alpha}(s; m'; n')$$

Regra decisória: rejeita-se H_0 em nível de significância α se $\theta_0 > \theta_{\alpha}(s; m'; n')$. Caso contrário, não se rejeita H_0 .

$$\text{No exemplo, segue que: } \theta_0 = \frac{9,6695}{1 + 9,6695} = 0,906.$$

Para $\alpha = 1\%$, $s = 2$, $m' = 2$ e $n' = 2,5$. Pela Tabela apropriada, temos:

$$n' = 1 \rightarrow 0,977$$

$$n' = 3 \rightarrow 0,886$$

Por interpolação, virá:

$$2 \rightarrow 0,091$$

$$1,5 \rightarrow x \quad \therefore x = \frac{(1,5)(0,091)}{2} = 0,068$$

$\theta = 0,977 - 0,068 = 0,909$. Assim, não se rejeita H_0 em nível de significância de 1% de probabilidade ($P > 0,01$).

Para $\alpha = 5\%$, $s = 2$, $m' = 2$ e $n' = 2,5$, virá:

$$n' = 1 \rightarrow 0,947$$

$$n' = 3 \rightarrow 0,820$$

Por interpolação, tem-se que $\theta = 0,852$. Assim, rejeita-se H_0 em nível de significância de 5% de probabilidade ($P < 0,05$).

Os quatro testes multivariados apresentados podem produzir diferentes níveis descritivos, também chamados de níveis críticos (valor-p). O teste de Wilks é o mais comumente usado. Cabe ressaltar que, para todos os quatro testes apresentados, é preciso ter $n_e \geq p$. Quando $s = \min [\text{posto}(W), \text{posto}(C')] = 1$, todos os quatro testes estatísticos dão resultados equivalentes.

4.4. COMPARAÇÕES MÚLTIPLAS

4.4.1. Teste T^2 de Hotelling

Para comparar dois vetores de médias de tratamentos, pode-se usar o teste T^2 de Hotelling como apresentado a seguir: caso de p variáveis

<u>Tratamento i</u>	<u>Tratamento u</u>
r_i repetições	r_u repetições
$\tilde{m}_i = \begin{bmatrix} m_{i1} \\ m_{i2} \\ \dots \\ m_{ip} \end{bmatrix}$	$\tilde{m}_u = \begin{bmatrix} m_{u1} \\ m_{u2} \\ \dots \\ m_{up} \end{bmatrix}$

$$H_0 : \tilde{m}_i = \tilde{m}_u \quad \text{vs.} \quad H_a : \tilde{m}_i \neq \tilde{m}_u$$

A estatística T^2 de Hotelling é dada por:

$$T^2 = \frac{r_i \cdot r_u}{r_i + r_u} \left[\hat{\tilde{m}}_i - \hat{\tilde{m}}_u \right]' S^{-1} \left[\hat{\tilde{m}}_i - \hat{\tilde{m}}_u \right],$$

em que

$$S = \frac{1}{n_e} E \Rightarrow S^{-1} = n_e E^{-1}.$$

T^2 se relaciona com a distribuição F por:

$$\frac{n_e - p + 1}{n_e \cdot p} \cdot T^2 \sim F(p; n_e - p + 1)$$

Assim, pode-se escrever:

$$F_0 = \frac{(n_e - p + 1)}{p} \cdot \frac{r_i \cdot r_u}{r_i + r_u} \left[\hat{\tilde{m}}_i - \hat{\tilde{m}}_u \right]' E^{-1} \left[\hat{\tilde{m}}_i - \hat{\tilde{m}}_u \right]$$

Regra decisória: rejeita-se H_0 em nível de significância α se $F_0 > F_\alpha(p; n_e - p + 1)$. Caso contrário, não se rejeita H_0 .

Fato: Note que, para $p=1$, $\hat{m}_{\tilde{i}} - \hat{m}_{\tilde{u}} = \hat{m}_{i1} - \hat{m}_{u1} = \hat{Y}$, onde \hat{Y} é um contraste univariado.

Neste caso, para experimentos com um fator inteiramente casualizado, $t^2 = F_0$, como é mostrado a seguir.

$$F_0 = n_e \cdot \frac{r_i \cdot r_u}{r_i + r_u} \hat{Y} (\text{SQR})^{-1} \hat{Y}$$

$$F_0 = n_e \cdot \frac{r_i \cdot r_u}{r_i + r_u} \cdot \hat{Y}^2 \cdot \frac{1}{\text{SQR}}$$

$$F_0 = \frac{\hat{Y}^2}{\left(\frac{1}{r_i} + \frac{1}{r_u}\right) \cdot \frac{\text{SQR}}{n_e}} = \frac{\hat{Y}^2}{\left(\frac{1}{r_i} + \frac{1}{r_u}\right) \cdot \text{QMR}} = t^2,$$

pois, neste caso, para testar $H_0 : Y = 0$ vs. $H_a : Y \neq 0$ pelo teste t de Student, sabe-se que

$$t = \frac{\hat{Y}}{\sqrt{\hat{V}(\hat{Y})}} = \frac{\hat{Y}}{\sqrt{\left(\frac{1}{r_i} + \frac{1}{r_u}\right) \cdot \text{QMR}}} \sim t(n_e). \text{ Logo, } t^2 = F_0.$$

Se todos os tratamentos tiverem o mesmo número de repetições r , as estatísticas T^2 e F_0 terão expressões mais simples, como as apresentadas a seguir:

$$T^2 = \frac{r}{2} \left[\hat{m}_{\tilde{i}} - \hat{m}_{\tilde{u}} \right]' S^{-1} \left[\hat{m}_{\tilde{i}} - \hat{m}_{\tilde{u}} \right]$$

$$F_0 = \frac{(n_e - p + 1)r}{2p} \left[\hat{m}_{\tilde{i}} - \hat{m}_{\tilde{u}} \right]' E^{-1} \left[\hat{m}_{\tilde{i}} - \hat{m}_{\tilde{u}} \right].$$

Como ilustração, serão comparados os vetores de médias dos tratamentos 5 e 6, cujos valores estão apresentados na Tabela 4.6, pelo teste T^2 de Hotelling adotando $\alpha = 1\%$.

$$H_0 : m_{\tilde{5}} = m_{\tilde{6}} \quad \text{vs.} \quad H_a : m_{\tilde{5}} \neq m_{\tilde{6}}$$

$$\hat{m}_{\tilde{5}} = \begin{bmatrix} 366,310 \\ 115,585 \end{bmatrix} \quad \hat{m}_{\tilde{6}} = \begin{bmatrix} 436,3110 \\ 79,020 \end{bmatrix}$$

$$\hat{m}_{\tilde{5}} - \hat{m}_{\tilde{6}} = \begin{bmatrix} -70,000 \\ 36,565 \end{bmatrix}$$

$$F_0 = \frac{(8 - 2 + 1)r}{2(2)} \begin{bmatrix} -70,000 & 36,565 \end{bmatrix} \begin{bmatrix} 78,6245 \cdot 10^{-5} & -50,9998 \cdot 10^{-5} \\ -50,9998 \cdot 10^{-5} & 66,1113 \cdot 10^{-5} \end{bmatrix} \begin{bmatrix} -70,000 \\ 36,565 \end{bmatrix}$$

$$F_0 = \frac{14}{4} \begin{bmatrix} -0,0737 & 0,0599 \end{bmatrix} \begin{bmatrix} -70,000 \\ 36,565 \end{bmatrix} = 25,72^{**} \quad (** P < 0,01)$$

$$F_{1\%}(2; 7) = 9,55$$

Assim, rejeita-se H_0 e conclui-se que os vetores de médias dos tratamentos 5 e 6 diferem significativamente ($P < 0,01$) pelo teste T^2 de Hotelling.

4.4.2. Teste de Roy e Bose

Dentre os métodos propostos para comparações múltiplas, o princípio da União-Interseção de Roy é mais adequado segundo Morrison (1976), pois ele vem a ser a continuação lógica do teste do maior autovalor. Assim, para uma variável número 1, por exemplo, como testar um contraste entre médias $Y = c_1 m_{11} + c_2 m_{21} + \dots + c_I m_{I1}$, para I tratamentos, com

$$\sum_{i=1}^I c_i = 0. \text{ O processo consiste em calcular uma diferença mínima significativa (DMS), dada}$$

por:

$$DMS = \sqrt{\frac{\theta_\alpha}{1 - \theta_\alpha} \cdot \frac{SQR}{r} \sum_{i=1}^I c_i^2},$$

onde θ_α é o valor crítico de Roy em nível α de probabilidade, podendo ser obtido dos ábacos de Heck ou das tabelas de Pillai, com parâmetros s, m' e n' , conforme já definidos anteriormente. SQR é a soma de quadrados do resíduo para a variável em questão e r é o número de repetições (suposto igual para todos os tratamentos). Este método é devido a Roy e Bose (1953) e Roy (1957), e será chamado simplesmente de teste de Roy e Bose.

Para o caso univariado ($p = 1$), demonstra-se que

$$DMS = \sqrt{(I - 1)F \frac{QMR}{r} \sum_{i=1}^I c_i^2},$$

com $F_\alpha(I - 1; n_e)$, que é a expressão obtida por Scheffé (1953).

O modelo multivariado leva a um menor número de diferenças significativas que o univariado. Isto, porém, ocorre porque no caso do modelo multivariado o critério de rejeição é mais rigoroso, por levar em conta um nível de significância conjunto, e no univariado ele é tomado isoladamente por análise, não se sabendo o nível conjunto de probabilidade para todas as análises, que será tanto maior quanto maior for o número de variáveis.

Para os dados da Tabela 4.3, vamos comparar as médias duas a duas pelo teste de Roy e Bose.

Temos que: $\theta_\alpha(s; m'; n') = \theta_{5\%}(2; 2; 2,5) = 0,852$ (valor obtido por interpolação). Logo, a DMS para a variável X é:

$$DMS = \sqrt{\frac{0,852}{1-0,852} \cdot \frac{2545,6926}{2} [(1)^2 + (-1)^2]} = 121,0757$$

A DMS para a variável Y é:

$$DMS = \sqrt{\frac{0,852}{1-0,852} \cdot \frac{3027,5286}{2} [(1)^2 + (-1)^2]} = 132,018.$$

Os resultados obtidos para as comparações múltiplas, de acordo com o modelo multivariado estão apresentados na Tabela 4.7.

Tabela 4.7 – Médias dos tratamentos e resultados obtidos para as comparações múltiplas de acordo com o modelo multivariado*

Consumo de ração (X)		Ganho de peso (Y)	
Tratamentos	Médias	Tratamentos	Médias
8	443,015 a	3	140,100 a
6	436,310 a	8	135,465 a
2	429,150 a	7	121,770 a
3	426,490 a	5	115,585 a
7	401,925 a	2	106,080 a
1	399,865 a	4	95,575 a
4	382,015 a	6	79,020 a
5	366,310 a	1	76,275 a
DMS	121,057	DMS	132,018

* As médias seguidas de pelo menos uma mesma letra, em cada coluna, não diferem entre si em nível de 5% de probabilidade pelo teste de Roy e Bose.

Na realidade, para este exemplo, as diferenças entre médias de tratamentos foram não significativas ($P > 0,05$).

4.4.3. Intervalo de confiança simultâneo de Bonferroni

Se o teste multivariado indicar diferença entre vetores de médias de tratamentos, então uma recomendação é proceder ao cálculo dos intervalos de confiança de Bonferroni para todos os pares de tratamentos e todas as características. Estes intervalos utilizam somente os valores críticos do teste t de Student univariado.

Seja um experimento com um fator no delineamento inteiramente casualizado. O modelo estatístico é:

$y_{ijl} = \mu_l + t_{il} + e_{ijl}$, com $i = 1, 2, \dots, k$ (tratamentos), $j = 1, 2, \dots, r_i$ (repetições) e $l = 1, 2, \dots, p$ (variáveis). Os vetores de médias para os tratamentos i e i' são dados por:

$$\mathbf{m}_i = \begin{bmatrix} m_{i1} \\ m_{i2} \\ \dots \\ m_{ip} \end{bmatrix} \quad \text{e} \quad \mathbf{m}_{i'} = \begin{bmatrix} m_{i'1} \\ m_{i'2} \\ \dots \\ m_{i'p} \end{bmatrix}.$$

Seja m_{il} o l -ésimo componente de \mathbf{m}_i . A diferença entre $m_{il} - m_{i'l}$ é estimada por $\hat{m}_{il} - \hat{m}_{i'l}$.

Para médias independentes, segue que:

$$\hat{V}(\hat{m}_{il} - \hat{m}_{i'l}) = \left(\frac{1}{r_i} + \frac{1}{r_{i'}} \right) \cdot \text{QMR}_l, \text{ onde } \text{QMR}_l \text{ é o quadrado médio do resíduo da } l\text{-ésima}$$

variável.

Há p variáveis e $C_k^2 = \frac{k(k-1)}{2}$ comparações por variável. O valor crítico será

$t_{\alpha'/2}(n_e)$, em que,

$$\alpha' = \frac{\alpha}{m}$$

n_e é o número de graus de liberdade do resíduo;

$m = p \cdot \frac{k(k-1)}{2}$, é o número de intervalos de confiança simultâneo para diferenças entre duas

médias de tratamentos. O valor m também pode ser escolhido de acordo com o número de comparações de interesse.

Com uma confiança de pelo menos $1 - \alpha$, $m_{il} - m_{i'l}$ pertence a

$$\hat{m}_{il} - \hat{m}_{i'l} \pm t_{(\alpha'/2; n_e)} \sqrt{\left(\frac{1}{r_i} + \frac{1}{r_{i'}} \right) \cdot \text{QMR}_l}, \quad (1)$$

para todos componentes $l = 1, 2, \dots, p$ e todas diferenças $i' < i = 1, 2, \dots, k$.

Se $r_i = r$ para todo i , então:

$$\hat{m}_{il} - \hat{m}_{i'l} \pm t_{(\alpha'/2; n_e)} \sqrt{\frac{2\text{QMR}_l}{r}}. \quad (2)$$

A expressão (2) pode ser usada para experimentos com um fator em blocos casualizados completos.

Fato: Numa Tabela bilateral, entrando diretamente com α' , a Tabela já dará o quantil de ordem

$(1 - \frac{\alpha'}{2}) \cdot 100\%$ da distribuição t de Student.

Exemplo: De um experimento no delineamento inteiramente casualizado com três tratamentos ($r_1 = 271$, $r_2 = 138$ e $r_3 = 107$) e quatro variáveis, são dados:

$$\hat{m}_1 = \begin{bmatrix} 2,066 \\ 0,480 \\ 0,082 \\ 0,360 \end{bmatrix}, \quad \hat{m}_2 = \begin{bmatrix} 2,167 \\ 0,596 \\ 0,124 \\ 0,418 \end{bmatrix}, \quad \hat{m}_3 = \begin{bmatrix} 2,273 \\ 0,521 \\ 0,125 \\ 0,383 \end{bmatrix}, \quad E = \begin{bmatrix} 182,962 & 4,408 & 1,695 & 9,581 \\ 4,408 & 8,200 & 0,633 & 2,428 \\ 1,695 & 0,633 & 1,484 & 0,394 \\ 9,581 & 2,428 & 0,394 & 6,538 \end{bmatrix}$$

Neste caso, $k = 3$, $p = 4$ e $n_e = 513$. Poderíamos construir $m = p \cdot \frac{k(k-1)}{2} = 4 \cdot \frac{3(3-1)}{2} = 12$

intervalos de confiança simultâneos. Para ilustrar, serão apresentados os cálculos comparando o tratamento 1 com o 3, apenas para a variável 3 e $\alpha = 5\%$.

$$\begin{aligned} \hat{m}_{13} - \hat{m}_{33} \pm t_{\left(\frac{0,05}{2(12)}; 513\right)} \sqrt{\left(\frac{1}{r_1} + \frac{1}{r_3}\right) \cdot QMR_3} \\ (0,082 - 0,125) \pm 2,87 \sqrt{\left(\frac{1}{271} + \frac{1}{107}\right) \cdot \frac{1,484}{513}} = -0,043 \pm 2,87(0,00614) \\ = -0,43 \pm 0,18 \\ -0,061 \leq m_{13} - m_{33} \leq -0,025 \end{aligned}$$

Com 95% de confiança, dizemos que $m_{13} - m_{33}$ pertence ao intervalo $[-0,061, -0,025]$. Como o intervalo de confiança não inclui o zero, existe diferença significativa entre $m_{13} - m_{33}$.

4.4.4. Função discriminante linear de Fisher

Uma análise complementar interessante da análise de variância multivariada é o uso da função discriminante linear de Fisher ou primeira variável canônica. Para testar hipótese sobre efeito de tratamentos, essa função é tal que fornece o maior valor possível para o teste F, entre todas as combinações lineares que se façam das variáveis envolvidas (Harris, 1975).

Dada a MANOVA relativa às p variáveis, determina-se a função discriminante linear de Fisher. Vamos ilustrar o procedimento a partir do exemplo, cujos dados são apresentados na Tabela 4.3 e a MANOVA na Tabela 4.5.

O objetivo é obter a função discriminante $Z^* = a^*X + b^*Y$, sendo $\underline{v}' = [a^* \ b^*]$, tal que, $\underline{v}'\underline{v} = 1$; \underline{v} é o autovetor normalizado associado ao maior autovalor da matriz $E^{-1}H$.

i) Cálculo dos autovalores

$$\text{De } |E^{-1}H - \lambda I| = 0, \text{ obtém-se: } \lambda_1 = 9,6695 \quad \text{e} \quad \lambda_2 = 2,2066.$$

ii) Cálculo de um autovetor não normalizado associado ao maior autovalor, cujos coeficientes serão denotados por a e b

$$(E^{-1}H - \lambda I) \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 7,4219 - 9,6695 & -2,7606 \\ -4,2462 & 4,4542 - 9,6695 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} -2,2476 & -2,7606 \\ -4,2462 & -5,2153 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

ou seja,

$$\begin{cases} 2,2476 a + 2,7606 b = 0 \\ 4,2462 a + 5,2153 b = 0 \end{cases}$$

A matriz do sistema é singular neste caso, conforme resulta da equação $|E^{-1}H - \lambda I| = 0$. Logo, o sistema de equações é consistente e indeterminado. Para resolvê-lo, pode-se tomar arbitrariamente $b = 1$ e abandonar a segunda equação. Então,

$$2,2476 a + 2,7606 (1) = 0$$

$$a = -1,2282.$$

Uma solução seria, pois a função $Z = -1,2282 X + Y$.

iii) Cálculo do autovetor normalizado v

$$a^* = \frac{a}{\sqrt{a^2 + b^2}} = \frac{-1,2282}{\sqrt{(-1,2282)^2 + 1^2}} = -0,7755$$

$$b^* = \frac{b}{\sqrt{a^2 + b^2}} = \frac{1}{\sqrt{(-1,2282)^2 + 1^2}} = 0,6314.$$

Logo,

$$Z^* = -0,7755 X + 0,6314 Y$$

ou indiferentemente, trocando-se o sinal do segundo membro, teremos:

$$Z^* = 0,7755 X - 0,6314 Y.$$

Fatos:

i) Em geral, $E^{-1}H$ tem $s = \min(p; q)$ autovalores não zeros.

ii) Seja soma de quadrados entre tratamentos $\tilde{v}'H\tilde{v}$ e soma de quadrados dentro de tratamentos $\tilde{v}'E\tilde{v}$. A razão $\tilde{v}'H\tilde{v}/\tilde{v}'E\tilde{v}$ é maximizada. Sendo \tilde{v}' o vetor que maximiza esta razão, então os elementos deste vetor são os coeficientes da função discriminante linear de Fisher ou primeira variável canônica.

iii) Como $\tilde{v}'H\tilde{v}/\tilde{v}'E\tilde{v}$ trata-se de uma razão, então ela será invariante a qualquer padronização de \tilde{v} , ou seja, $\tilde{v}'H\tilde{v}/\tilde{v}'E\tilde{v}$ continuará constante, seja para \tilde{v} ou para $c\tilde{v}$, sendo c uma constante diferente de zero. Assim, pode-se obter a função discriminante tal que $\tilde{v}'\tilde{v} = 1$, $\tilde{v}'E\tilde{v} = 1$ ou $\tilde{v}'\frac{E}{n_e}\tilde{v} = 1$, etc. Em qualquer caso, a estatística $F = \tilde{v}'H\tilde{v}/\tilde{v}'E\tilde{v}$ será invariante.

Os escores obtidos em cada caso serão diferentes, mas todos os testes e interpretações serão equivalentes.

iv) A função discriminante linear de Fisher produz o maior F dentre todas as combinações lineares que se façam das \mathbf{p} variáveis sob análise. Ela capta a maior quantidade de informação contida nas \mathbf{p} variáveis.

Aplicando-se a função obtida aos dados da Tabela 4.3, obtém-se os dados apresentados na Tabela 4.8.

Tabela 4.8 – Valores da função discriminante $Z^* = 0,7755X - 0,6314Y$

Tratamentos	Repetições		Totais	Médias
	1	2		
1	263,90	259,97	523,87	261,93
2	269,57	262,09	531,66	265,83
3	252,99	231,58	484,57	242,28
4	240,72	231,09	471,81	235,90
5	218,54	203,65	422,19	211,09
6	301,20	275,73	576,93	288,46
7	229,09	240,52	469,61	234,80
8	256,78	259,26	516,04	258,02

$$G = \sum Z_{ij}^* = 3996,68, \quad \sum Z_{ij}^{*2} = 1007034,0484, \quad C = \frac{(3996,68)^2}{16} = 998340,6888$$

$$SQ_{Total} = 8693,3596, \quad SQ_{Tratamentos} = 7878,4853, \quad SQ_{Resíduo} = 814,8743$$

O resultado da análise de variância dos dados da Tabela 4.8 está apresentado na Tabela 4.9.

Tabela 4.9 – Análise de variância dos dados gerados pela função discriminante linear de Fisher

Causas de variação	G.L.	S.Q.	Q.M.	F
Tratamentos	$q = 7$	7878,4853	1125,4979	11,05**
Resíduo	$n_e = 8$	814,8743	101,8593	
Total	$n - 1 = 15$	8693,3596		

** $P < 0,01$ $F_{1\%}(7; 8) = 6,18$

Quando $w = q$, o valor de F pode ser calculado previamente pela aproximação dada para o teste de Roy (maior autovalor), cuja estatística é dada a seguir:

$$F_0 = \lambda_{\max} \cdot \frac{(n_e - w + q)}{w} \sim F(w; n_e - w + q) = F(q; n_e).$$

No exemplo, como $w = q = 7$, vira': $F_0 = 9,6695 \cdot \frac{(8 - 7 + 7)}{7} = 11,05$.

A importância relativa (IR) de uma função discriminante é avaliada pela porcentagem da variância total observada atribuída a ela, e é dada por:

$$\text{IR de } Z^* (\%) = \frac{\lambda_i}{\text{traço}(E^{-1}H)} \cdot 100 = \frac{\lambda_i}{\sum_{i=1}^p \lambda_i} \cdot 100.$$

No exemplo, a importância relativa da primeira função discriminante linear de Fisher é:

$$\text{IR de } Z^* (\%) = \frac{9,6695}{9,6695 + 2,2066} \cdot 100 = 81,42\%.$$

A análise baseada na função discriminante linear de Fisher será tanto mais eficiente quanto maior for a porcentagem da variância total atribuída a ela.

É interessante comparar as médias dos tratamentos relativas à função discriminante linear de Fisher. Assim, para testar qualquer contraste entre as médias, pode-se usar o teste de Roy e Bose apresentado anteriormente, cuja diferença mínima significativa (DMS) é dada por:

$$\text{DMS} = \sqrt{\frac{\theta_\alpha}{1 - \theta_\alpha} \cdot \frac{\text{SQR}}{r} \sum_{i=1}^I c_i^2},$$

onde θ_α é o valor crítico de Roy ao nível α de probabilidade, podendo ser obtido dos ábacos de Heck ou das tabelas de Pillai, com parâmetros s , m' e n' , conforme já definidos anteriormente. SQR é a soma de quadrados do resíduo e r é o número de repetições (suposto igual para todos os tratamentos).

No exemplo, a DMS para testar todo e qualquer contraste entre duas médias de tratamentos adotando-se $\alpha = 5\%$ é dada a seguir:

$\theta_{5\%}(2; 2; 2,5) = 0,852$, valor este obtido por interpolação. Assim,

$$\text{DMS} = \sqrt{\frac{0,852}{1 - 0,852} \cdot \frac{814,8743}{2} \cdot [(1)^2 + (-1)^2]} = 68,49.$$

Assim, todo contraste entre duas médias, cuja estimativa, em valor absoluto, exceder o valor 68,49 será declarado significativo em nível de 5% de probabilidade. Os resultados do teste são apresentados na Tabela 4.10.

Tabela 4.10 – Médias de tratamentos relativas à função discriminante linear de Fisher

$$Z^* = 0,7755X - 0,6314Y$$

Tratamentos	Médias*
5	211,09 a
7	234,80 a b
4	235,90 a b
3	242,28 a b
8	258,02 a b
1	261,93 a b
2	265,83 a b
6	288,46 b

* As médias seguidas de pelo menos uma mesma letra não diferem significativamente entre si, em nível de 5% de probabilidade pelo teste de Roy e Bose.

Pode-se concluir que o tratamento 5 diferiu significativamente ($P < 0,05$) apenas do tratamento 6 e, nenhum outro contraste entre duas médias foi significativo.

4.5. EXEMPLO DE UMA ANÁLISE DE VARIÂNCIA MULTIVARIADA DE UM EXPERIMENTO EM BLOCOS COMPLETOS CASUALIZADOS

Sejam os dados da Tabela 4.3, considerando agora como sendo dados de um experimento no delineamento em blocos completos casualizados (DBC) com $I = 8$ tratamentos e $r = 2$ repetições. Além das matrizes **A** (Total) e **H** (Tratamentos) já calculadas anteriormente para o experimento inteiramente casualizado, é preciso calcular a matriz **B**, relativa às somas de quadrados e de produtos de blocos.

Foi visto que:

$$A = \begin{bmatrix} 13100,7393 & 3683,3922 \\ 3683,3922 & 11091,4432 \end{bmatrix} \text{ e } H = \begin{bmatrix} 10555,0467 & 1719,5837 \\ 1719,5837 & 8063,9146 \end{bmatrix}.$$

A matriz B pode ser obtida matricialmente, ou ainda, de uma maneira bastante simples, a partir dos totais de blocos apresentados na Tabela 4.11.

Tabela 4.11 – Totais de blocos

Variáveis*	Repetições		Totais
	Bloco 1	Bloco 2	
X	$B_{11} = 3272,18$	$B_{12} = 3297,98$	6570,16
Y	$B_{21} = 799,45$	$B_{22} = 940,29$	1739,74

* X = Consumo de ração (g) e Y = Ganho de peso (g)

Assim virá:

$$SQB(X) = \frac{B_{11}^2 + B_{12}^2 + \dots + B_{1r}^2}{I} - C_{11}$$

$$SQB(X) = \frac{(3272,18)^2 + (3297,98)^2}{8} - \frac{(6570,16)^2}{16} = 41,6025$$

$$SQB(Y) = \frac{B_{21}^2 + B_{22}^2 + \dots + B_{2r}^2}{I} - C_{22}$$

$$SQB(Y) = \frac{(799,45)^2 + (940,29)^2}{8} - \frac{(1739,74)^2}{16} = 1239,7441$$

$$SPB(X, Y) = \frac{B_{11}B_{21} + B_{12}B_{22} + \dots + B_{1r}B_{2r}}{I} - C_{12}$$

$$SPB(X, Y) = \frac{(3272,18)(799,45) + (3297,98)(940,29)}{8} - \frac{(6570,16)(1739,74)}{16} = 227,1045$$

$$B = \begin{bmatrix} 41,6025 & 227,1045 \\ 227,1045 & 1239,7441 \end{bmatrix} \text{ e } E = A - B - H = \begin{bmatrix} 2504,0901 & 1736,7040 \\ 1736,7040 & 1787,7845 \end{bmatrix}.$$

O resultado da MANOVA está apresentado na Tabela 4.12.

Tabela 4.12 – Análise de variância multivariada para os dados de consumo de ração (X) e ganho de peso (Y)

Causas de variação	G.L.	Matrizes de somas de quadrados e de produtos
Blocos	$r - 1 = 1$	$B = \begin{bmatrix} 41,6025 & 227,1045 \\ 227,1045 & 1239,7441 \end{bmatrix}$
Tratamentos	$I - 1 = q = 7$	$H = \begin{bmatrix} 10555,0467 & 1719,5837 \\ 1719,5837 & 8063,9146 \end{bmatrix}$
Resíduo	$(I-1)(r-1) = n_e = 7$	$E = \begin{bmatrix} 2504,0901 & 1736,7040 \\ 1736,7040 & 1787,7845 \end{bmatrix}$
Total	$Ir - 1 = 15$	$A = \begin{bmatrix} 13100,7393 & 3683,3922 \\ 3683,3922 & 11091,4432 \end{bmatrix}$

Os resultados para os quatro testes estão apresentados a seguir:

- i) Teste de Wilks: $\Lambda = 0,0125$, $F_0 = 6,80$, $F(14; 12)$, Valor-p=0,0010;
- ii) Teste de Pillai: $V = 1,6914$, $F_0 = 5,48$, $F(14; 14)$, Valor-p=0,0015;
- iii) Teste de Hotelling-Lawley (2º caso): $U=22,65$, $F_0 = 9,14$, $F(14; 6,8571)$, Valor-p=0,0037;
- iv) Teste de Roy (aproximação): $\lambda_{\max} = 19,82$, $F_0 = 19,82$, $F(7; 7)$, Valor-p=0,0004;

4.6. RESULTADOS OBTIDOS ATRAVÉS DO PROGRAMA SAS – EXEMPLO 1

Com os dados da Tabela 4.3, a seguir são apresentados os resultados da MANOVA, entre outros, com o uso do programa SAS.

(a) Programa:

```

OPTIONS NODATE NONUMBER;
/* MANOVA E PRIMEIRA VARIÁVEL CANÔNICA */
DATA AVE1;
  INPUT TRAT REP X Y;
CARDS;
1 1 402.18 76.00
1 2 397.55 76.55
2 1 421.76 91.08
2 2 436.54 121.08
3 1 435.67 134.42
3 2 417.31 145.78
4 1 381.81 87.70
4 2 382.22 103.45
5 1 381.68 122.67
5 2 350.94 108.50
6 1 442.53 66.49
6 2 430.09 91.55
7 1 390.37 116.62
7 2 413.48 126.92
8 1 416.18 104.47
8 2 469.85 166.46
;
PROC PRINT;
PROC GLM;
CLASS TRAT REP;
MODEL X Y = TRAT; /* Experimento no delineamento inteiramente casualizado */
CONTRAST 'Y1=M1 - M2' TRAT 1 -1;
CONTRAST 'Y2=M5 - M6' TRAT 0 0 0 1 -1;
CONTRAST 'Y3=M1 + M2 -2M3' TRAT 1 1 -2;
MANOVA H = TRAT / PRINTH PRINTE SHORT;
PROC GLM;
CLASS TRAT REP;
MODEL X Y = TRAT REP; /* Experimento em blocos casualizados */
MANOVA H = TRAT / PRINTH PRINTE SHORT;
RUN;

*****
* ANOVA PARA OS DADOS GERADOS PELA PRIMEIRA FUNÇÃO *
* DISCRIMINANTE CANÔNICA DE FISHER *
*****;

DATA AVE1;
  SET AVE1;
  FDF = 0.7755*X - 0.6314*Y;
RUN;

PROC ANOVA DATA=AVE1;
  CLASS TRAT REP;
  MODEL FDF = TRAT;
  MEANS TRAT;
RUN;

```

(b) Resultado da Análise

The SAS System

Obs	TRAT	REP	X	Y
1	1	1	402.18	76.00
2	1	2	397.55	76.55
3	2	1	421.76	91.08
4	2	2	436.54	121.08
5	3	1	435.67	134.42
6	3	2	417.31	145.78
7	4	1	381.81	87.70
8	4	2	382.22	103.45
9	5	1	381.68	122.67
10	5	2	350.94	108.50
11	6	1	442.53	66.49
12	6	2	430.09	91.55
13	7	1	390.37	116.62
14	7	2	413.48	126.92
15	8	1	416.18	104.47
16	8	2	469.85	166.46

The SAS System

The GLM Procedure

Class Level Information

Class	Levels	Values
TRAT	8	1 2 3 4 5 6 7 8
REP	2	1 2

Number of observations 16

The SAS System

The GLM Procedure

Dependent Variable: X

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	10555.04660	1507.86380	4.74	0.0220
Error	8	2545.69260	318.21157		
Corrected Total	15	13100.73920			

R-Square	Coeff Var	Root MSE	X Mean
0.805683	4.344122	17.83849	410.6350

Source	DF	Type I SS	Mean Square	F Value	Pr > F
TRAT	7	10555.04660	1507.86380	4.74	0.0220

Source	DF	Type III SS	Mean Square	F Value	Pr > F
TRAT	7	10555.04660	1507.86380	4.74	0.0220

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
Y1=M1 - M2	1	857.611225	857.611225	2.70	0.1393
Y2=M5 - M6	1	4900.000000	4900.000000	15.40	0.0044
Y3=M1 + M2 -2M3	1	191.440408	191.440408	0.60	0.4603

The SAS System
The GLM Procedure

Dependent Variable: Y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	8063.91458	1151.98780	3.04	0.0708
Error	8	3027.52860	378.44107		
Corrected Total	15	11091.44318			

R-Square	Coeff Var	Root MSE	Y Mean
0.727039	17.89101	19.45356	108.7338

Source	DF	Type I SS	Mean Square	F Value	Pr > F
TRAT	7	8063.914575	1151.987796	3.04	0.0708

Source	DF	Type III SS	Mean Square	F Value	Pr > F
TRAT	7	8063.914575	1151.987796	3.04	0.0708

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
Y1=M1 - M2	1	888.338025	888.338025	2.35	0.1640
Y2=M5 - M6	1	1336.999225	1336.999225	3.53	0.0970
Y3=M1 + M2 -2M3	1	3191.214675	3191.214675	8.43	0.0198

The SAS System
The GLM Procedure
Multivariate Analysis of Variance

E = Error SSCP Matrix

	X	Y
X	2545.6926	1963.80855
Y	1963.80855	3027.5286

Partial Correlation Coefficients from the Error SSCP Matrix / Prob > |r|

DF = 8	X	Y
X	1.000000	0.707379 0.0330
Y	0.707379 0.0330	1.000000

The SAS System

The GLM Procedure
Multivariate Analysis of Variance

H = Type III SSCP Matrix for TRAT

	X	Y
X	10555.0466	1719.58365
Y	1719.58365	8063.914575

Characteristic Roots and Vectors of: E Inverse * H, where
H = Type III SSCP Matrix for TRAT
E = Error SSCP Matrix

Characteristic Root	Percent	Characteristic Vector X	V'EV=1 Y
9.66948907	81.42	0.02716805	-0.02211979
2.20656428	18.58	0.00693852	0.01310834

MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall TRAT Effect

H = Type III SSCP Matrix for TRAT
E = Error SSCP Matrix

S=2 M=2 N=2.5

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.02922917	4.85	14	14	0.0028
Pillai's Trace	1.59441453	4.49	14	16	0.0026
Hotelling-Lawley Trace	11.87605334	5.57	14	8.3636	0.0087
Roy's Greatest Root	9.66948907	11.05	7	8	0.0015

NOTE: F Statistic for Roy's Greatest Root is an upper bound.

NOTE: F Statistic for Wilks' Lambda is exact.

H = Contrast SSCP Matrix for Y1=M1 - M2

	X	Y
X	857.611225	872.839425
Y	872.839425	888.338025

The SAS System

The GLM Procedure
Multivariate Analysis of Variance

Characteristic Roots and Vectors of: E Inverse * H, where
H = Contrast SSCP Matrix for Y1=M1 - M2
E = Error SSCP Matrix

Characteristic Root	Percent	Characteristic Vector X	V'EV=1 Y
0.37129126	100.00	0.01284133	0.00782682
0.00000000	0.00	-0.02492682	0.02449193

MANOVA Test Criteria and Exact F Statistics for the Hypothesis of No Overall Y1=M1 - M2 Effect

H = Contrast SSCP Matrix for Y1=M1 - M2

E = Error SSCP Matrix

S=1 M=0 N=2.5

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.72923968	1.30	2	7	0.3312
Pillai's Trace	0.27076032	1.30	2	7	0.3312
Hotelling-Lawley Trace	0.37129126	1.30	2	7	0.3312
Roy's Greatest Root	0.37129126	1.30	2	7	0.3312

H = Contrast SSCP Matrix for Y2=M5 - M6

	X	Y
X	4900	-2559.55
Y	-2559.55	1336.999225

Characteristic Roots and Vectors of: E Inverse * H, where

H = Contrast SSCP Matrix for Y2=M5 - M6

E = Error SSCP Matrix

Characteristic Root	Percent	Characteristic Vector X	V'EV=1 Y
7.34724849	100.00	0.02718433	-0.02208882
0.00000000	0.00	0.00687445	0.01316045

The SAS System

The GLM Procedure

Multivariate Analysis of Variance

MANOVA Test Criteria and Exact F Statistics for the Hypothesis of No Overall Y2=M5 - M6 Effect

H = Contrast SSCP Matrix for Y2=M5 - M6

E = Error SSCP Matrix

S=1 M=0 N=2.5

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.11979996	25.72	2	7	0.0006
Pillai's Trace	0.88020004	25.72	2	7	0.0006
Hotelling-Lawley Trace	7.34724849	25.72	2	7	0.0006
Roy's Greatest Root	7.34724849	25.72	2	7	0.0006

H = Contrast SSCP Matrix for Y3=M1 + M2 -2M3

	X	Y
X	191.44040833	781.618475
Y	781.618475	3191.214675

Characteristic Roots and Vectors of: E Inverse * H, where

H = Contrast SSCP Matrix for Y3=M1 + M2 -2M3

E = Error SSCP Matrix

Characteristic Root	Percent	Characteristic Vector X	V'EV=1 Y
1.46302540	100.00	-0.01482496	0.02504259
0.00000000	0.00	0.02380056	-0.00582943

MANOVA Test Criteria and Exact F Statistics for the
Hypothesis of No Overall Y3=M1 + M2 -2M3 Effect

H = Contrast SSCP Matrix for Y3=M1 + M2 -2M3

E = Error SSCP Matrix

S=1 M=0 N=2.5

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.40600475	5.12	2	7	0.0426
Pillai's Trace	0.59399525	5.12	2	7	0.0426
Hotelling-Lawley Trace	1.46302540	5.12	2	7	0.0426
Roy's Greatest Root	1.46302540	5.12	2	7	0.0426

The SAS System
The GLM Procedure
Class Level Information

Class	Levels	Values
TRAT	8	1 2 3 4 5 6 7 8
REP	2	1 2

Number of observations 16

The SAS System
The GLM Procedure

Dependent Variable: X

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	10596.64910	1324.58114	3.70	0.0508
Error	7	2504.09010	357.72716		
Corrected Total	15	13100.73920			

R-Square	Coeff Var	Root MSE	X Mean
0.808859	4.605958	18.91368	410.6350

Source	DF	Type I SS	Mean Square	F Value	Pr > F
TRAT	7	10555.04660	1507.86380	4.22	0.0385
REP	1	41.60250	41.60250	0.12	0.7431

Source	DF	Type III SS	Mean Square	F Value	Pr > F
TRAT	7	10555.04660	1507.86380	4.22	0.0385
REP	1	41.60250	41.60250	0.12	0.7431

The SAS System
The GLM Procedure

Dependent Variable: Y

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	9303.65868	1162.95733	4.55	0.0303
Error	7	1787.78450	255.39779		
Corrected Total	15	11091.44318			

R-Square	Coeff Var	Root MSE	Y Mean
0.838814	14.69752	15.98117	108.7338

Source	DF	Type I SS	Mean Square	F Value	Pr > F
TRAT	7	8063.914575	1151.987796	4.51	0.0325
REP	1	1239.744100	1239.744100	4.85	0.0634

Source	DF	Type III SS	Mean Square	F Value	Pr > F
TRAT	7	8063.914575	1151.987796	4.51	0.0325
REP	1	1239.744100	1239.744100	4.85	0.0634

The SAS System
The GLM Procedure
Multivariate Analysis of Variance

E = Error SSCP Matrix

	X	Y
X	2504.0901	1736.70405
Y	1736.70405	1787.7845

Partial Correlation Coefficients from the Error SSCP Matrix / Prob > |r|

DF = 7	X	Y
X	1.000000	0.820811 0.0125
Y	0.820811 0.0125	1.000000

The SAS System
The GLM Procedure
Multivariate Analysis of Variance

H = Type III SSCP Matrix for TRAT

	X	Y
X	10555.0466	1719.58365
Y	1719.58365	8063.914575

Characteristic Roots and Vectors of: E Inverse * H, where

H = Type III SSCP Matrix for TRAT
E = Error SSCP Matrix

Characteristic Root	Percent	Characteristic Vector X	V'EV=1 Y
19.8161301	87.47	-0.03321799	0.03969103
2.8385084	12.53	0.01097928	0.01179023

MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall TRAT Effect

H = Type III SSCP Matrix for TRAT
E = Error SSCP Matrix

S=2 M=2 N=2

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.01251519	6.80	14	12	0.0010
Pillai's Trace	1.69144247	5.48	14	14	0.0015
Hotelling-Lawley Trace	22.65463856	9.14	14	6.8571	0.0037
Roy's Greatest Root	19.81613013	19.82	7	7	0.0004

NOTE: F Statistic for Roy's Greatest Root is an upper bound.

NOTE: F Statistic for Wilks' Lambda is exact.

The SAS System
The ANOVA Procedure
Class Level Information

Class	Levels	Values
TRAT	8	1 2 3 4 5 6 7 8
REP	2	1 2

Number of observations 16

The SAS System
The ANOVA Procedure

Dependent Variable: FDF

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	7878.625234	1125.517891	11.05	0.0015
Error	8	814.792299	101.849037		
Corrected Total	15	8693.417533			

R-Square	Coeff Var	Root MSE	FDF Mean
0.906275	4.040157	10.09203	249.7930

Source	DF	Anova SS	Mean Square	F Value	Pr > F
TRAT	7	7878.625234	1125.517891	11.05	0.0015

The SAS System

The ANOVA Procedure

Level of TRAT	N	-----FDF----- Mean	Std Dev
1	2	261.935273	2.7844698
2	2	265.826913	5.2892365
3	2	242.283855	15.1397813
4	2	235.906578	6.8070306
5	2	211.093036	10.5301861
6	2	288.465177	18.0100833
7	2	234.807260	8.0740175
8	2	258.025532	1.7540484

4.7. ANÁLISE DE VARIÂNCIA MULTIVARIADA DE UM EXPERIMENTO INTEIRAMENTE CASUALIZADO COM DESENVOLVIMENTO MATRICIAL – EXEMPLO 2

Sejam os dados de um experimento no delineamento inteiramente casualizado (DIC) apresentados na Tabela 4.13.

Tabela 4.13 – Resultados experimentais (DIC)*

1. Testemunha		2. Turfa Fermentada		3. Turfa Não-Fermentada	
X_1	X_2	X_1	X_2	X_1	X_2
4,63	0,95	6,03	1,08	4,71	0,96
4,38	0,89	5,96	1,05	4,81	0,93
4,94	1,01	6,16	1,08	4,49	0,87
4,96	1,23	6,33	1,19	4,43	0,82
4,48	0,94	6,08	1,08	4,56	0,91

X_1 = Teor de Nitrogênio, X_2 = Teor de Fósforo

* Fonte: Gomes (2009)

► **O modelo linear** : $Y = XB + \varepsilon$

$$y_{ijk} = \mu_k + t_{ik} + e_{ijk},$$

em que:

$i = 1, 2, \dots, I$ ($I = 3$ tratamentos)

$j = 1, 2, \dots, r$ ($r = 5$ repetições)

$k = 1, 2, \dots, p$ ($p = 2$ variáveis)

$$\begin{bmatrix} 4,63 & 0,95 \\ 4,38 & 0,89 \\ 4,94 & 1,01 \\ 4,96 & 1,23 \\ 4,48 & 0,94 \\ \hline 6,03 & 1,08 \\ 5,96 & 1,05 \\ 6,16 & 1,08 \\ 6,33 & 1,19 \\ 6,08 & 1,08 \\ \hline 4,71 & 0,96 \\ 4,81 & 0,93 \\ 4,49 & 0,87 \\ 4,43 & 0,82 \\ 4,56 & 0,91 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ \hline 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ \hline 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \mu_1 & \mu_2 \\ \hline t_{11} & t_{12} \\ t_{21} & t_{22} \\ t_{31} & t_{32} \end{bmatrix} + \begin{bmatrix} e_{111} & e_{112} \\ e_{121} & e_{122} \\ e_{131} & e_{132} \\ e_{141} & e_{142} \\ e_{151} & e_{152} \\ \hline e_{211} & e_{212} \\ e_{221} & e_{222} \\ e_{231} & e_{232} \\ e_{241} & e_{242} \\ e_{251} & e_{252} \\ \hline e_{311} & e_{312} \\ e_{321} & e_{322} \\ e_{331} & e_{332} \\ e_{341} & e_{342} \\ e_{351} & e_{352} \end{bmatrix}$$

► O sistema de equações normais

$$X'XB^0 = X'Y$$

$$\begin{bmatrix} 15 & 5 & 5 & 5 \\ 5 & 5 & 0 & 0 \\ 5 & 0 & 5 & 0 \\ 5 & 0 & 0 & 5 \end{bmatrix} \begin{bmatrix} \mu_1^0 & \mu_2^0 \\ t_{11}^0 & t_{12}^0 \\ t_{21}^0 & t_{22}^0 \\ t_{31}^0 & t_{32}^0 \end{bmatrix} = \begin{bmatrix} 76,95 & 14,99 \\ 23,39 & 5,02 \\ 30,56 & 5,48 \\ 23,00 & 4,49 \end{bmatrix}$$

► Solução do sistema

✓ Solução Geral:

$$B^0 = (X'X)^- X'Y$$

Uma $(X'X)^-$ poderia ser: $(X'X)^- = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1/5 & 0 & 0 \\ 0 & 0 & 1/5 & 0 \\ 0 & 0 & 0 & 1/5 \end{bmatrix}$

✓ Usando a restrição $\sum_{i=1}^I r_i \hat{t}_{ik} = 0$, para $k = 1, 2, \dots, p$, virá:

$$A\hat{B} = \phi \quad \begin{bmatrix} 0 & 5 & 5 & 5 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{\mu}_1 & \hat{\mu}_2 \\ \hat{t}_{11} & \hat{t}_{12} \\ \hat{t}_{21} & \hat{t}_{22} \\ \hat{t}_{31} & \hat{t}_{32} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$\underbrace{(X'X - A)}_M \hat{B} = X'Y \Rightarrow M\hat{B} = X'Y \therefore \hat{B} = M^{-1}X'Y$$

$$\begin{bmatrix} 15 & 0 & 0 & 0 \\ 5 & 5 & 0 & 0 \\ 5 & 0 & 5 & 0 \\ 5 & 0 & 0 & 5 \end{bmatrix} \begin{bmatrix} \hat{\mu}_1 & \hat{\mu}_2 \\ \hat{t}_{11} & \hat{t}_{12} \\ \hat{t}_{21} & \hat{t}_{22} \\ \hat{t}_{31} & \hat{t}_{32} \end{bmatrix} = \begin{bmatrix} 76,95 & 14,99 \\ 23,39 & 5,02 \\ 30,56 & 5,48 \\ 23,00 & 4,49 \end{bmatrix}$$

$$\hat{B} = M^{-1}X'Y = \begin{bmatrix} 1/15 & 0 & 0 & 0 \\ -1/15 & 1/5 & 0 & 0 \\ -1/15 & 0 & 1/5 & 0 \\ -1/15 & 0 & 0 & 1/5 \end{bmatrix} \begin{bmatrix} 76,95 & 14,99 \\ 23,39 & 5,02 \\ 30,56 & 5,48 \\ 23,00 & 4,49 \end{bmatrix} = \begin{bmatrix} 5,1300 & 0,9993 \\ -0,4520 & 0,0047 \\ 0,9820 & 0,0967 \\ -0,5300 & -0,1013 \end{bmatrix}$$

► Matrizes de somas de quadrados e de produtos

i) Parâmetros

$$\hat{B}'X'Y = \begin{bmatrix} 5,1300 & -0,4520 & 0,9820 & -0,5300 \\ 0,9993 & 0,0047 & 0,0967 & -0,1013 \end{bmatrix} \begin{bmatrix} 76,95 & 14,99 \\ 23,39 & 5,02 \\ 30,56 & 5,48 \\ 23,00 & 4,49 \end{bmatrix}$$

$$\hat{B}'X'Y = \begin{bmatrix} 402,00114 & 77,63132 \\ 77,63132 & 15,07818 \end{bmatrix},$$

com graus de liberdade = Posto $[X] = 3$ (1 para correção e 2 para tratamentos)

ii) Resíduo

$$\text{Na MANOVA } H_0 : C'BW = \phi \Rightarrow H_0 : C'B = \phi \quad (W = I_{(p)})$$

$$E = Y' \left[I_{(n)} - X(X'X)^{-1}X' \right] Y$$

$$E = Y'Y - \hat{B}'X'Y, \text{ com } n - \text{Posto}[X] = n_e \text{ graus de liberdade.}$$

$$E = \begin{bmatrix} 402,4591 & 77,8038 \\ 77,8038 & 15,1729 \end{bmatrix} - \begin{bmatrix} 402,00114 & 77,63132 \\ 77,63132 & 15,07818 \end{bmatrix}$$

$$E = \begin{bmatrix} 0,45796 & 0,17248 \\ 0,17248 & 0,09472 \end{bmatrix}, \text{ com } n_e = 12 \text{ graus de liberdade.}$$

iii) Tratamentos

Para $H_0 : C'B = \phi$, virá:

$$H = [C'B^0]' [C'(X'X)^- C]^{-1} [C'B^0]$$

com a solução adotada, virá:

$$H = [C'\hat{B}]' [C'M^{-1}C]^{-1} [C'\hat{B}]$$

C' é de posto linha completo

$$\text{Uma } C' \text{ pode ser: } C' = \begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 1 & 1 & -2 \end{bmatrix} \Rightarrow \text{ não é única. Os dois contrastes podem ser ortogonais ou não}$$

$$C'\hat{B} = \begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 1 & 1 & -2 \end{bmatrix} \begin{bmatrix} 5,1300 & 0,9993 \\ -0,4520 & 0,0047 \\ 0,9820 & 0,0967 \\ -0,5300 & -0,1013 \end{bmatrix} = \begin{bmatrix} -1,4340 & -0,0920 \\ 1,5900 & 0,3040 \end{bmatrix}$$

$$C'M^{-1}C = \begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 1 & 1 & -2 \end{bmatrix} \begin{bmatrix} \frac{1}{15} & 0 & 0 & 0 \\ -\frac{1}{15} & \frac{1}{5} & 0 & 0 \\ -\frac{1}{15} & 0 & \frac{1}{5} & 0 \\ -\frac{1}{15} & 0 & 0 & \frac{1}{5} \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 1 & 1 \\ -1 & 1 \\ 0 & -2 \end{bmatrix} = \begin{bmatrix} \frac{2}{5} & 0 \\ 0 & \frac{6}{5} \end{bmatrix}$$

$$H = \begin{bmatrix} -1,4340 & 1,5900 \\ -0,0920 & 0,3040 \end{bmatrix} \begin{bmatrix} \frac{5}{2} & 0 \\ 0 & \frac{5}{6} \end{bmatrix} \begin{bmatrix} -1,4340 & -0,0920 \\ 1,5900 & 0,3040 \end{bmatrix}$$

$$H = \begin{bmatrix} -3,5850 & 1,3250 \\ -0,2300 & 0,253333 \end{bmatrix} \begin{bmatrix} -1,4340 & -0,0920 \\ 1,5900 & 0,3040 \end{bmatrix}$$

$$H = \begin{bmatrix} 7,24764 & 0,73262 \\ 0,73262 & 0,09817 \end{bmatrix}, \text{ com } q = \text{Posto } [C'] = 2 \text{ graus de liberdade.}$$

iv) Total

A matriz **A** de somas de quadrados e de produtos totais com n-1 graus de liberdade é dada por:

$$A = Y' \left[I_{(n)} - \frac{1}{n} \mathbf{1} \mathbf{1}' \right] Y$$

$$A = Y' Y - \frac{1}{n} Y' \mathbf{1} \mathbf{1}' Y, \quad \mathbf{1} \text{ é um vetor de } 1\text{'s } n \times 1 \quad Y' \mathbf{1} = \begin{bmatrix} G_1 \\ G_2 \end{bmatrix}$$

$$A = \begin{bmatrix} 402,4591 & 77,8038 \\ 77,8038 & 15,1729 \end{bmatrix} - \frac{1}{15} \begin{bmatrix} 76,95 \\ 14,99 \end{bmatrix} \begin{bmatrix} 76,95 & 14,99 \end{bmatrix}$$

$$A = \begin{bmatrix} 7,70560 & 0,90510 \\ 0,90510 & 0,19289 \end{bmatrix}, \quad \text{com 14 graus de liberdade.}$$

► Um desdobramento da matriz H

$$Y_1 = m_1 - m_2 \quad \rightarrow \quad H_1 \quad \text{com } 1 \text{ grau de liberdade}$$

$$Y_2 = m_1 + m_2 - 2m_3 \quad \rightarrow \quad H_2 \quad \text{com } 1 \text{ grau de liberdade}$$

$H = H_1 + H_2$, pois os contrastes são ortogonais.

$$a) H_1 = \left[C_1' \hat{B} \right]' \left[C_1' M^{-1} C_1 \right]^{-1} \left[C_1' \hat{B} \right]$$

$$C_1' \hat{B} = \begin{bmatrix} 0 & 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} 5,1300 & 0,9993 \\ -0,4520 & 0,0047 \\ 0,9820 & 0,0967 \\ -0,5300 & -0,1013 \end{bmatrix} = \begin{bmatrix} -1,4340 & -0,0920 \end{bmatrix}$$

$$C_1' M^{-1} C_1 = \begin{bmatrix} 0 & 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} 1/15 & 0 & 0 & 0 \\ -1/15 & 1/5 & 0 & 0 \\ -1/15 & 0 & 1/5 & 0 \\ -1/15 & 0 & 0 & 1/5 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \end{bmatrix} = \frac{2}{5}$$

$$H_1 = \begin{bmatrix} -1,4340 \\ -0,0920 \end{bmatrix} \left(\frac{2}{5} \right)^{-1} \begin{bmatrix} -1,4340 & -0,0920 \end{bmatrix}$$

$$H_1 = \begin{bmatrix} 5,14089 & 0,32982 \\ 0,32982 & 0,02116 \end{bmatrix}, \quad \text{com } 1 \text{ grau de liberdade.}$$

$$b) H_2 = [C_2' \hat{B}]' [C_2' M^{-1} C_2]^{-1} [C_2' \hat{B}]$$

$$C_2' \hat{B} = [0 \quad 1 \quad 1 \quad -2] \begin{bmatrix} 5,1300 & 0,9993 \\ -0,4520 & 0,0047 \\ 0,9820 & 0,0967 \\ -0,5300 & -0,1013 \end{bmatrix} = [1,5900 \quad 0,3040]$$

$$C_2' M^{-1} C_2 = [0 \quad 1 \quad 1 \quad -2] \begin{bmatrix} 1/15 & 0 & 0 & 0 \\ -1/15 & 1/5 & 0 & 0 \\ -1/15 & 0 & 1/5 & 0 \\ -1/15 & 0 & 0 & 1/5 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 1 \\ -2 \end{bmatrix} = \frac{6}{5}$$

$$H_2 = \begin{bmatrix} 1,5900 \\ 0,3040 \end{bmatrix} \left(\frac{6}{5} \right)^{-1} [1,5900 \quad 0,3040]$$

$$H_2 = \begin{bmatrix} 2,10675 & 0,40280 \\ 0,40280 & 0,07701 \end{bmatrix}, \text{ com 1 grau de liberdade.}$$

Os resultados da MANOVA são apresentados na Tabela 4.14.

FATOS:

- A matriz H_2 poderia ter sido obtida por diferença, isto é, $H_2 = H - H_1$, pois os contrastes são ortogonais;
- A soma de quadrado para um contraste Y_i pode ser obtida da maneira usual;
- A soma de produto para um contraste Y_i pode ser obtida da seguinte maneira:

Para $Z = X_1 + X_2$, tem-se que:

$$SPY_i(X_1, X_2) = \frac{1}{2} [SQY_i(Z) - SQY_i(X_1) - SQY_i(X_2)]$$

Tabela 4.14 – Análise de variância multivariada

Causas de Variação	G.L.	Matrizes	Λ
$Y_1 = m_1 - m_2$ $\sim \quad \sim \quad \sim$	1	$H_1 = \begin{bmatrix} 5,14089 & 0,32982 \\ 0,32982 & 0,02116 \end{bmatrix}$	0,034373**
$Y_2 = m_1 + m_2 - 2 m_3$ $\sim \quad \sim \quad \sim \quad \sim$	1	$H_2 = \begin{bmatrix} 2,10675 & 0,40280 \\ 0,40280 & 0,07701 \end{bmatrix}$	0,124463**
Tratamentos	(2)	$H = \begin{bmatrix} 7,24764 & 0,73262 \\ 0,73262 & 0,09817 \end{bmatrix}$	0,020428**
Resíduo	12	$E = \begin{bmatrix} 0,45796 & 0,17248 \\ 0,17248 & 0,09472 \end{bmatrix}$	

** $P < 0,01$ ► **Teste de Wilks para a igualdade dos vetores de médias de tratamentos**

$$H_0 : m_1 = m_2 = m_3$$

$$\Lambda = \frac{|E|}{|H + E|} = 0,020428^{**} \quad ** (P < 0,01)$$

✓ Pela tabela: $\Lambda_\alpha(p; q; n_e)$

em que:

p = Posto (W)

q = Posto (C') n_e = número de graus de liberdade do resíduo

$$\Lambda_{1\%}(2; 2; 12) = 0,314111$$

✓ Caso A: p = 2, q e n_e quaisquer

$$F_0 = \frac{n_e - 1}{q} \cdot \frac{1 - \sqrt{\Lambda}}{\sqrt{\Lambda}} \sim F[2q; 2(n_e - 1)]$$

$$F_0 = \frac{12 - 1}{2} \cdot \frac{1 - \sqrt{0,020428}}{\sqrt{0,020428}} = 32,98^{**} \quad ** (P < 0,01)$$

$$F_{1\%}(4; 22) = 4,31$$

As estimativas para os vetores de médias de tratamentos são apresentados na Tabela 4.15.

Tabela 4.15 – Estimativas para os vetores de médias de tratamentos

Variáveis	Tratamentos		
	1	2	3
Teor de N (X_1)	$\begin{bmatrix} 4,678 \end{bmatrix}$	$\begin{bmatrix} 6,112 \end{bmatrix}$	$\begin{bmatrix} 4,600 \end{bmatrix}$
Teor de P (X_2)	$\begin{bmatrix} 1,004 \end{bmatrix}$	$\begin{bmatrix} 1,096 \end{bmatrix}$	$\begin{bmatrix} 0,898 \end{bmatrix}$

► **Teste de Wilks para os contrastes**

$$i) \underset{\sim}{Y}_1 = \underset{\sim}{m}_1 - \underset{\sim}{m}_2$$

$$H_0 : \underset{\sim}{Y}_1 = \underset{\sim}{0} \left(H_0 : C_1' B = \phi \right)$$

$$\checkmark \text{ Pela tabela: } \Lambda = \frac{|E|}{|H_1 + E|} = 0,034373^{**} \quad ** (P < 0,01)$$

$$\Lambda_1(\alpha; p; q; n_e)$$

em que:

p = Posto (W)

q = Posto (C_1')

n_e = número de graus de liberdade do resíduo

$$\Lambda_1(1\%; 2; 1; 12) = 0,432876$$

$$\checkmark \underline{\Lambda_1 \rightarrow F \Rightarrow \text{Caso B: } q = 1, p \text{ e } n_e \text{ quaisquer}}$$

OBS.: Para posto de (C') = q = 1, o SAS usa o caso B.

$$F_0 = \frac{n_e - p + 1}{p} \cdot \frac{1 - \Lambda_1}{\Lambda_1} \sim F[p; n_e - p + 1]$$

$$F_0 = \frac{12 - 2 + 1}{2} \cdot \frac{1 - 0,034373}{0,034373} = 154,51^{**} \quad ** (P < 0,01)$$

$$F_{1\%}(2; 11) = 7,21$$

$$ii) \underset{\sim}{Y}_2 = \underset{\sim}{m}_1 + \underset{\sim}{m}_2 - 2\underset{\sim}{m}_3$$

$$H_0 : \underset{\sim}{Y}_2 = \underset{\sim}{0} \left(H_0 : C_2' B = \phi \right)$$

$$\checkmark \text{ Pela tabela: } \Lambda_2 = \frac{|E|}{|H_2 + E|} = 0,124463^{**} \quad ** (P < 0,01)$$

$$\Lambda_2 (1\%; 2; 1; 12) = 0,432876$$

$$\checkmark \underline{\Lambda_2 \rightarrow F \Rightarrow \text{Caso B}}$$

$$F_0 = \frac{12 - 2 + 1}{2} \cdot \frac{1 - 0,124463}{0,124463} = 38,69^{**} \quad ** (P < 0,01)$$

$$F_{1\%} (2; 11) = 7,21$$

A seguir é apresentado o programa SAS para o exemplo em questão.

4.8. RESULTADOS OBTIDOS ATRAVÉS DO PROGRAMA SAS – EXEMPLO 2

(a) Programa

```
options nodate nonumber;
data Exemplo;
input trat rep x1 x2;
cards;
1 1 4.63 0.95
1 2 4.38 0.89
1 3 4.94 1.01
1 4 4.96 1.23
1 5 4.48 0.94
2 1 6.03 1.08
2 2 5.96 1.05
2 3 6.16 1.08
2 4 6.33 1.19
2 5 6.08 1.08
3 1 4.71 0.96
3 2 4.81 0.93
3 3 4.49 0.87
3 4 4.43 0.82
3 5 4.56 0.91
;
proc glm;
class trat rep;
model x1 x2 = trat; /* Experimento no delineamento inteiramente casualizado */
contrast 'y1=m1-m2' trat 1 -1 0;
contrast 'y2=m1+m2-2m3' trat 1 1 -2;
manova H = trat / printH printE short;
proc glm;
class trat rep;
model x1 x2 = rep trat; /* Experimento em Blocos Casualizados */
contrast 'y1=m1-m2' trat 1 -1 0;
contrast 'y2=m1+m2-2m3' trat 1 1 -2;
manova H = trat / printH printE short;
run;
```


(b) Resultado da Análise

The SAS System

The GLM Procedure
Class Level Information

Class	Levels	Values
trat	3	1 2 3
rep	5	1 2 3 4 5

Number of observations 15

The SAS System
The GLM Procedure

Dependent Variable: x1

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	7.24764000	3.62382000	94.96	<.0001
Error	12	0.45796000	0.03816333		
Corrected Total	14	7.70560000			

R-Square	Coeff Var	Root MSE	x1 Mean
0.940568	3.808078	0.195354	5.130000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
trat	2	7.24764000	3.62382000	94.96	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
trat	2	7.24764000	3.62382000	94.96	<.0001

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
y1=m1-m2	1	5.14089000	5.14089000	134.71	<.0001
y2=m1+m2-2m3	1	2.10675000	2.10675000	55.20	<.0001

The SAS System
The GLM Procedure

Dependent Variable: x2

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	0.09817333	0.04908667	6.22	0.0140
Error	12	0.09472000	0.00789333		
Corrected Total	14	0.19289333			

	R-Square	Coeff Var	Root MSE	x2 Mean	
	0.508951	8.890370	0.088844	0.999333	
Source	DF	Type I SS	Mean Square	F Value	Pr > F
trat	2	0.09817333	0.04908667	6.22	0.0140
Source	DF	Type III SS	Mean Square	F Value	Pr > F
trat	2	0.09817333	0.04908667	6.22	0.0140
Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
y1=m1-m2	1	0.02116000	0.02116000	2.68	0.1275
y2=m1+m2-2m3	1	0.07701333	0.07701333	9.76	0.0088

The SAS System

The GLM Procedure
Multivariate Analysis of Variance

E = Error SSCP Matrix

	x1	x2
x1	0.45796	0.17248
x2	0.17248	0.09472

Partial Correlation Coefficients from the Error SSCP Matrix / Prob > |r|

DF = 12	x1	x2
x1	1.000000	0.828141 0.0005
x2	0.828141 0.0005	1.000000

The SAS System
The GLM Procedure
Multivariate Analysis of Variance

H = Type III SSCP Matrix for trat

	x1	x2
x1	7.24764	0.73262
x2	0.73262	0.098173333

Characteristic Roots and Vectors of: E Inverse * H, where
H = Type III SSCP Matrix for trat
E = Error SSCP Matrix

Characteristic Root	Percent	Characteristic Vector x1	V'EV=1 x2
34.7579162	98.95	2.61059062	-4.30107181
0.3689930	1.05	-0.36728225	3.88633434

MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall trat Effect

H = Type III SSCP Matrix for trat
E = Error SSCP Matrix

S=2 M=-0.5 N=4.5

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.02042803	32.98	4	22	<.0001
Pillai's Trace	1.24157026	9.82	4	24	<.0001
Hotelling-Lawley Trace	35.12690920	94.48	4	12.235	<.0001
Roy's Greatest Root	34.75791615	208.55	2	12	<.0001

NOTE: F Statistic for Roy's Greatest Root is an upper bound.

NOTE: F Statistic for Wilks' Lambda is exact.

H = Contrast SSCP Matrix for y1=m1-m2
x1 x2

x1	5.14089	0.32982
x2	0.32982	0.02116

The SAS System

The GLM Procedure
Multivariate Analysis of Variance

Characteristic Roots and Vectors of: E Inverse * H, where

H = Contrast SSCP Matrix for y1=m1-m2

E = Error SSCP Matrix

Characteristic Root	Percent	Characteristic Vector x1	V'EV=1 x2
28.0924118	100.00	2.62579731	-4.49168618
0.0000000	0.00	-0.23509174	3.66436468

MANOVA Test Criteria and Exact F Statistics for the Hypothesis of No Overall y1=m1-m2 Effect

H = Contrast SSCP Matrix for y1=m1-m2

E = Error SSCP Matrix

S=1 M=0 N=4.5

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.03437322	154.51	2	11	<.0001
Pillai's Trace	0.96562678	154.51	2	11	<.0001
Hotelling-Lawley Trace	28.09241176	154.51	2	11	<.0001
Roy's Greatest Root	28.09241176	154.51	2	11	<.0001

H = Contrast SSCP Matrix for y2=m1+m2-2m3

x1 x2

x1	2.10675	0.4028
x2	0.4028	0.0770133333

Characteristic Roots and Vectors of: E Inverse * H, where
 H = Contrast SSCP Matrix for $y_2=m_1+m_2-2m_3$
 E = Error SSCP Matrix

Characteristic Root	Percent	Characteristic Vector x1	V'EV=1 x2
7.03449744	100.00	2.47926880	-3.40996438
0.00000000	0.00	-0.89627332	4.68774532

The SAS System

The GLM Procedure
 Multivariate Analysis of Variance

MANOVA Test Criteria and Exact F Statistics for the Hypothesis of No Overall $y_2=m_1+m_2-2m_3$ Effect

H = Contrast SSCP Matrix for $y_2=m_1+m_2-2m_3$
 E = Error SSCP Matrix

S=1 M=0 N=4.5

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.12446329	38.69	2	11	<.0001
Pillai's Trace	0.87553671	38.69	2	11	<.0001
Hotelling-Lawley Trace	7.03449744	38.69	2	11	<.0001
Roy's Greatest Root	7.03449744	38.69	2	11	<.0001

The SAS System
 The GLM Procedure

Class Level Information

Class	Levels	Values
trat	3	1 2 3
rep	5	1 2 3 4 5
Number of observations		15

The SAS System
 The GLM Procedure

Dependent Variable: x1

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	7.34090667	1.22348444	26.84	<.0001
Error	8	0.36469333	0.04558667		
Corrected Total	14	7.70560000			

R-Square	Coeff Var	Root MSE	x1 Mean
0.952672	4.161995	0.213510	5.130000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
rep	4	0.09326667	0.02331667	0.51	0.7298
trat	2	7.24764000	3.62382000	79.49	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
rep	4	0.09326667	0.02331667	0.51	0.7298
trat	2	7.24764000	3.62382000	79.49	<.0001

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
y1=m1-m2	1	5.14089000	5.14089000	112.77	<.0001
y2=m1+m2-2m3	1	2.10675000	2.10675000	46.21	0.0001

The SAS System
The GLM Procedure

Dependent Variable: x2

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	0.12520000	0.02086667	2.47	0.1184
Error	8	0.06769333	0.00846167		
Corrected Total	14	0.19289333			

R-Square	Coeff Var	Root MSE	x2 Mean
0.649063	9.204868	0.091987	0.999333

Source	DF	Type I SS	Mean Square	F Value	Pr > F
rep	4	0.02702667	0.00675667	0.80	0.5586
trat	2	0.09817333	0.04908667	5.80	0.0277

Source	DF	Type III SS	Mean Square	F Value	Pr > F
rep	4	0.02702667	0.00675667	0.80	0.5586
trat	2	0.09817333	0.04908667	5.80	0.0277

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
y1=m1-m2	1	0.02116000	0.02116000	2.50	0.1525
y2=m1+m2-2m3	1	0.07701333	0.07701333	9.10	0.0166

The SAS System

The GLM Procedure
Multivariate Analysis of Variance

E = Error SSCP Matrix

	x1	x2
x1	0.3646933333	0.13198
x2	0.13198	0.0676933333

Partial Correlation Coefficients from the Error SSCP Matrix / Prob > |r|

DF = 8	x1	x2
x1	1.000000	0.839984 0.0046
x2	0.839984 0.0046	1.000000

The SAS System
The GLM Procedure
Multivariate Analysis of Variance

H = Type III SSCP Matrix for trat

	x1	x2
x1	7.24764	0.73262
x2	0.73262	0.098173333

Characteristic Roots and Vectors of: E Inverse * H, where
H = Type III SSCP Matrix for trat
E = Error SSCP Matrix

Characteristic Root	Percent	Characteristic Vector x1	V'EV=1 x2
45.2877693	98.84	3.02024021	-5.33765541
0.5309980	1.16	-0.43736057	4.65653643

MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall trat Effect

H = Type III SSCP Matrix for trat
E = Error SSCP Matrix

S=2 M=-0.5 N=2.5

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.01411104	25.96	4	14	<.0001
Pillai's Trace	1.32522731	7.86	4	16	0.0011
Hotelling-Lawley Trace	45.81876731	78.16	4	7.4839	<.0001
Roy's Greatest Root	45.28776929	181.15	2	8	<.0001

NOTE: F Statistic for Roy's Greatest Root is an upper bound.
NOTE: F Statistic for Wilks' Lambda is exact.

H = Contrast SSCP Matrix for y1=m1-m2

	x1	x2
x1	5.14089	0.32982
x2	0.32982	0.02116

The SAS System

The GLM Procedure
Multivariate Analysis of Variance

Characteristic Roots and Vectors of: E Inverse * H, where
 H = Contrast SSCP Matrix for $y_1=m_1-m_2$
 E = Error SSCP Matrix

Characteristic Root	Percent	Characteristic Vector x1	V'EV=1 x2
36.9620127	100.00	3.03881135	-5.57123954
0.0000000	0.00	-0.28064353	4.37437846

MANOVA Test Criteria and Exact F Statistics for the Hypothesis of No Overall $y_1=m_1-m_2$ Effect

H = Contrast SSCP Matrix for $y_1=m_1-m_2$
 E = Error SSCP Matrix

S=1 M=0 N=2.5

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.02634212	129.37	2	7	<.0001
Pillai's Trace	0.97365788	129.37	2	7	<.0001
Hotelling-Lawley Trace	36.96201269	129.37	2	7	<.0001
Roy's Greatest Root	36.96201269	129.37	2	7	<.0001

H = Contrast SSCP Matrix for $y_2=m_1+m_2-2m_3$

	x1	x2
x1	2.10675	0.4028
x2	0.4028	0.077013333

Characteristic Roots and Vectors of: E Inverse * H, where
 H = Contrast SSCP Matrix for $y_2=m_1+m_2-2m_3$
 E = Error SSCP Matrix

Characteristic Root	Percent	Characteristic Vector x1	V'EV=1 x2
8.85675461	100.00	2.84900542	-4.17710881
0.00000000	0.00	-1.09375653	5.72063448

The SAS System

The GLM Procedure
Multivariate Analysis of VarianceMANOVA Test Criteria and Exact F Statistics for the Hypothesis of No Overall $y_2=m_1+m_2-2m_3$ EffectH = Contrast SSCP Matrix for $y_2=m_1+m_2-2m_3$

E = Error SSCP Matrix

S=1 M=0 N=2,5

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.10145327	31.00	2	7	0.0003
Pillai's Trace	0.89854673	31.00	2	7	0.0003
Hotelling-Lawley Trace	8.85675461	31.00	2	7	0.0003
Roy's Greatest Root	8.85675461	31.00	2	7	0.0003

4.9. EXPERIMENTO FATORIAL MULTIVARIADO

Os dados da Tabela 4.16, referem-se a um experimento em 4 blocos completos casualizados, sobre consorciação de 4 cultivares de milho com 4 de sorgo, constituindo um fatorial 4 x 4.

Tabela 4.16 – Resultados experimentais para os cultivares de milho (M) e sorgo (S)

S	M	Milho (X_1)					Sorgo (X_2)		
S ₁	M ₁	2,32	2,62	1,45	1,71	1,51	1,20	1,66	1,35
	M ₂	2,14	1,74	2,38	1,64	1,12	0,69	1,57	1,48
	M ₃	4,72	3,27	3,59	2,27	2,56	2,51	2,24	1,64
	M ₄	2,67	2,64	2,56	2,73	1,44	1,28	0,93	1,08
S ₂	M ₁	2,03	1,51	2,12	2,59	1,12	1,35	1,50	2,32
	M ₂	2,63	1,56	1,68	1,93	0,84	2,24	1,14	1,41
	M ₃	3,94	3,44	3,26	2,39	2,20	0,84	1,41	1,20
	M ₄	2,70	3,26	3,05	3,02	1,60	1,36	1,34	1,62
S ₃	M ₁	2,41	2,20	1,95	2,45	2,20	2,30	2,77	2,40
	M ₂	2,04	3,88	1,94	1,47	2,68	2,08	2,00	2,36
	M ₃	2,94	2,33	2,74	3,76	5,78	3,21	4,05	3,01
	M ₄	2,51	2,20	3,39	3,51	2,81	3,20	1,76	1,74
S ₄	M ₁	1,74	1,21	1,76	1,88	4,81	4,50	4,47	4,86
	M ₂	1,07	2,02	1,95	1,23	4,90	2,98	3,98	3,24
	M ₃	2,42	2,63	2,04	2,46	3,22	3,00	3,08	2,72
	M ₄	2,65	2,92	3,09	2,67	2,70	2,46	2,64	2,42

Os resultados da MANOVA com o teste de Wilks são apresentados na Tabela 4.17.

Tabela 4.17 – Análise de variância multivariada para as variáveis X_1 e X_2

Causas de variação	G.L.	Matrizes	Λ
Blocos	3	$B = \begin{bmatrix} 0,3323 & 0,6667 \\ 0,6667 & 1,7645 \end{bmatrix}$	0,1836**
Sorgo (S)	3	$S = \begin{bmatrix} 2,5850 & -7,9276 \\ -7,9276 & 47,4431 \end{bmatrix}$	
Milheto (M)	3	$M = \begin{bmatrix} 14,7836 & 0,1798 \\ 0,1798 & 5,7649 \end{bmatrix}$	0,3195**
Interação S x M	9	$SM = \begin{bmatrix} 2,2609 & -0,7824 \\ -0,7824 & 15,7113 \end{bmatrix}$	0,3678**
Resíduo	45	$E = \begin{bmatrix} 13,8304 & -2,0874 \\ -2,0874 & 11,9212 \end{bmatrix}$	
Total	63	$A = \begin{bmatrix} 33,7922 & -9,9509 \\ -9,9509 & 82,6050 \end{bmatrix}$	

** $P < 0,01$

$$\text{Teste de Wilks: } \Lambda_S = \frac{|E|}{|S + E|}, \Lambda_M = \frac{|E|}{|M + E|} \text{ e } \Lambda_{SM} = \frac{|E|}{|SM + E|}$$

Todos os cálculos necessários para elaboração da Tabela da MANOVA são apresentados a seguir de uma maneira bastante simples.

$p = 2$ variáveis

Sorgo: $a = 4$ níveis

Milheto: $b = 4$ níveis

Blocos: $r = 4$ repetições

$n = 64$ observações por variável

$$\sum X_1 = 2,32 + 2,62 + \dots + 2,67 = 157,02$$

$$\sum X_1^2 = (2,32)^2 + (2,62)^2 + \dots + (2,67)^2 = 419,0310$$

$$\begin{aligned} \text{SQTotal}(X_1) &= \sum X_1^2 - \frac{(\sum X_1)^2}{n} \\ &= 419,0310 - \frac{(157,02)^2}{64} = 419,0310 - 385,2387 = 33,7923 \end{aligned}$$

$$\sum X_2 = 1,51 + 1,20 + \dots + 2,42 = 148,08$$

$$\sum X_2^2 = (1,51)^2 + (1,20)^2 + \dots + (2,42)^2 = 425,2252$$

$$\text{SQTotal}(X_2) = \sum X_2^2 - \frac{(\sum X_2)^2}{n}$$

$$= 425,2252 - \frac{(148,08)^2}{64} = 425,2252 - 342,6201 = 82,6051$$

$$\sum X_1 X_2 = (2,32)(1,51) + (2,62)(1,20) + \dots + (2,67)(2,42) = 353,3541$$

$$\begin{aligned} \text{SPTotal}(X_1 X_2) &= \sum X_1 X_2 - \frac{(\sum X_1)(\sum X_2)}{n} \\ &= 353,3541 - \frac{(157,02)(148,08)}{64} = 353,3541 - 363,3050 = -9,9509 \end{aligned}$$

$$\text{Logo, } A = \begin{bmatrix} 33,7922 & -9,9509 \\ -9,9509 & 82,6050 \end{bmatrix}.$$

Blocos: Tabela auxiliar

Variáveis	Totais de blocos				Totais
	B ₁	B ₂	B ₃	B ₄	
X ₁	40,93	39,43	38,95	37,71	157,02
X ₂	41,49	35,20	36,54	34,85	148,08

$$\begin{aligned} \text{SQB}(X_1) &= \frac{1}{16} [(40,93)^2 + \dots + (37,71)^2] - \frac{(157,02)^2}{64} \\ &= 385,5710 - 385,2387 = 0,3323 \end{aligned}$$

$$\begin{aligned} \text{SQB}(X_2) &= \frac{1}{16} [(41,49)^2 + \dots + (34,85)^2] - \frac{(148,08)^2}{64} \\ &= 344,3846 - 342,6201 = 1,7645 \end{aligned}$$

$$\begin{aligned} \text{SPB}(X_1, X_2) &= \frac{1}{16} [(40,93)(41,49) + \dots + (37,71)(34,85)] - \frac{(157,02)(148,08)}{64} \\ &= 363,9717 - 363,3050 = 0,6667 \end{aligned}$$

$$\text{Logo, } B = \begin{bmatrix} 0,3323 & 0,6667 \\ 0,6667 & 1,7645 \end{bmatrix}.$$

Fator S: Tabela auxiliar

Fator S	Totais	
	X ₁	X ₂
1	40,45	24,26
2	41,11	23,49
3	41,72	44,35
4	33,74	55,98
	157,02	148,08

$$\begin{aligned} \text{SQS}(X_1) &= \frac{1}{16} [(40,45)^2 + (41,11)^2 + (41,72)^2 + (33,74)^2] - \frac{(157,02)^2}{64} \\ &= 387,8238 - 385,2387 = 2,5851 \end{aligned}$$

$$\begin{aligned} \text{SQS}(X_2) &= \frac{1}{16} [(24,26)^2 + (23,49)^2 + (44,35)^2 + (55,98)^2] - \frac{(148,08)^2}{64} \\ &= 390,0632 - 342,6201 = 47,4431 \end{aligned}$$

$$\begin{aligned} \text{SPS}(X_1, X_2) &= \frac{1}{16} [(40,45)(24,26) + \dots + (33,74)(55,98)] - \frac{(157,02)(148,08)}{64} \\ &= 355,3774 - 363,3050 = -7,9276 \end{aligned}$$

$$\text{Logo, } S = \begin{bmatrix} 2,5850 & -7,9276 \\ -7,9276 & 47,4431 \end{bmatrix}.$$

Fator M: Tabela auxiliar

Fator M	Totais	
	X ₁	X ₂
1	31,95	40,32
2	31,30	34,71
3	48,20	42,67
4	45,57	30,38
	157,02	148,08

$$\begin{aligned} \text{SQM}(X_1) &= \frac{1}{16} [(31,95)^2 + (31,30)^2 + (48,20)^2 + (45,57)^2] - \frac{(157,02)^2}{64} \\ &= 400,0223 - 385,2387 = 14,7836 \end{aligned}$$

$$\begin{aligned} \text{SQM}(X_2) &= \frac{1}{16} [(40,32)^2 + (34,71)^2 + (42,67)^2 + (30,38)^2] - \frac{(148,08)^2}{64} \\ &= 348,3850 - 342,6201 = 5,7649 \end{aligned}$$

$$\begin{aligned} \text{SPM}(X_1, X_2) &= \frac{1}{16} [(31,95)(40,32) + \dots + (45,57)(30,38)] - \frac{(157,02)(148,08)}{64} \\ &= 363,4848 - 363,3050 = 0,1798 \end{aligned}$$

$$\text{Logo, } M = \begin{bmatrix} 14,7836 & 0,1798 \\ 0,1798 & 5,7649 \end{bmatrix}.$$

Interação SxM: Tabela auxiliar

S \ M	1	2	3	4
	X ₁	X ₂	X ₁	X ₂
1	8,10	5,72	7,90	4,86
2	8,25	6,29	7,80	5,63
3	9,01	9,67	9,33	9,12
4	6,59	18,64	6,27	15,10

$$SQS_{xM}(X_1) = SQS, M(X_1) - SQS(X_1) - SQM(X_1)$$

$$= \frac{1}{4}[(8,10)^2 + (7,90)^2 + \dots + (9,55)^2 + (11,33)^2] - \frac{(157,02)^2}{64} - 2,5851 - 14,7836$$

$$= 404,8683 - 385,2387 - 2,5851 - 14,7836 = 2,2609$$

$$SQS_{xM}(X_2) = SQS, M(X_2) - SQS(X_2) - SQM(X_2)$$

$$= \frac{1}{4}[(5,72)^2 + (4,86)^2 + \dots + (12,02)^2 + (10,22)^2] - \frac{(148,08)^2}{64} - 47,4431 - 5,7649$$

$$= 411,5394 - 342,6201 - 47,4431 - 5,7649 = 15,7113$$

$$SPS_{xM}(X_1, X_2) = SPS, M(X_1, X_2) - SPS(X_1, X_2) - SPM(X_1, X_2)$$

$$= \frac{1}{4}[(8,10)(5,72) + (7,90)(4,86) + \dots + (9,55)(12,02) + (11,33)(10,22)] - \frac{(157,02)(148,08)}{64}$$

$$- \frac{(157,02)(148,08)}{64} - (-7,9276) - 0,1798$$

$$= 354,7748 - 363,3050 - (-7,9276) - 0,1798 = -0,7824$$

$$\text{Logo, } SM = \begin{bmatrix} 2,2609 & -0,7824 \\ -0,7824 & 15,7113 \end{bmatrix}.$$

$$E = A - B - S - M - S_{xM} = \begin{bmatrix} 13,8304 & -2,0874 \\ -2,0874 & 11,9212 \end{bmatrix}.$$

4.9.1. Resultados obtidos através do programa SAS

(a) Programa

```

OPTIONS NODATE NONUMBER;
/* EXEMPLO DE UM FATORIAL MULTIVARIADO */
DATA SORGMIL;
  TITLE1 'FATORIAL MULTIVARIADO';
  INPUT SORGO MILHO REP X1 X2;
CARDS;
1 1 1 2.32 1.51
1 1 2 2.62 1.20
1 1 3 1.45 1.66
1 1 4 1.71 1.35
1 2 1 2.14 1.12
1 2 2 1.74 0.69
1 2 3 2.38 1.57
1 2 4 1.64 1.48
1 3 1 4.72 2.56
1 3 2 3.27 2.51
1 3 3 3.59 2.24
1 3 4 2.27 1.64
1 4 1 2.67 1.44
1 4 2 2.64 1.28
1 4 3 2.56 0.93
1 4 4 2.73 1.08
2 1 1 2.03 1.12
2 1 2 1.51 1.35
2 1 3 2.12 1.50
2 1 4 2.59 2.32
2 2 1 2.63 0.84
2 2 2 1.56 2.24
2 2 3 1.68 1.14
2 2 4 1.93 1.41
2 3 1 3.94 2.20
2 3 2 3.44 0.84
2 3 3 3.26 1.41
2 3 4 2.39 1.20
2 4 1 2.70 1.60
2 4 2 3.26 1.36
2 4 3 3.05 1.34
2 4 4 3.02 1.62
3 1 1 2.41 2.20
3 1 2 2.20 2.30
3 1 3 1.95 2.77
3 1 4 2.45 2.40
3 2 1 2.04 2.68
3 2 2 3.88 2.08
3 2 3 1.94 2.00
3 2 4 1.47 2.36
3 3 1 2.94 5.78
3 3 2 2.33 3.21
3 3 3 2.74 4.05
3 3 4 3.76 3.01

```

3	4	1	2.51	2.81
3	4	2	2.20	3.20
3	4	3	3.39	1.76
3	4	4	3.51	1.74
4	1	1	1.74	4.81
4	1	2	1.21	4.50
4	1	3	1.76	4.47
4	1	4	1.88	4.86
4	2	1	1.07	4.90
4	2	2	2.02	2.98
4	2	3	1.95	3.98
4	2	4	1.23	3.24
4	3	1	2.42	3.22
4	3	2	2.63	3.00
4	3	3	2.04	3.08
4	3	4	2.46	2.72
4	4	1	2.65	2.70
4	4	2	2.92	2.46
4	4	3	3.09	2.64
4	4	4	2.67	2.42

;

PROC GLM;

CLASS SORGO MILHO REP X1 X2;

MODEL X1 X2 = REP SORGO MILHO SORGO*MILHO;

MANOVA H = SORGO MILHO SORGO*MILHO / PRINTH PRINTE SHORT;

RUN;

(b) Resultado da Análise

FATORIAL MULTIVARIADO
The GLM Procedure
Class Level Information

Class	Levels	Values
SORGO	4	1 2 3 4
MILHO	4	1 2 3 4
REP	4	1 2 3 4

FATORIAL MULTIVARIADO
The GLM Procedure

Dependent Variable: X1

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	18	19.96181250	1.10898958	3.61	0.0002
Error	45	13.83043125	0.30734292		
Corrected Total	63	33.79224375			

R-Square 0.590722 Coeff Var 22.59626 Root MSE 0.554385 X1 Mean 2.453438

Source	DF	Type I SS	Mean Square	F Value	Pr > F
REP	3	0.33226875	0.11075625	0.36	0.7819
SORGO	3	2.58503125	0.86167708	2.80	0.0505
MILHO	3	14.78358125	4.92786042	16.03	<.0001
SORGO*MILHO	9	2.26093125	0.25121458	0.82	0.6031

Source	DF	Type III SS	Mean Square	F Value	Pr > F
REP	3	0.33226875	0.11075625	0.36	0.7819
SORGO	3	2.58503125	0.86167708	2.80	0.0505
MILHO	3	14.78358125	4.92786042	16.03	<.0001
SORGO*MILHO	9	2.26093125	0.25121458	0.82	0.6031

FATORIAL MULTIVARIADO
The GLM Procedure

Dependent Variable: X2

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	18	70.68383750	3.92687986	14.82	<.0001
Error	45	11.92126250	0.26491694		
Corrected Total	63	82.60510000			

R-Square 0.855684 Coeff Var 22.24531 Root MSE 0.514701 X2 Mean 2.313750

Source	DF	Type I SS	Mean Square	F Value	Pr > F
REP	3	1.76453750	0.58817917	2.22	0.0988
SORGO	3	47.44306250	15.81435417	59.70	<.0001
MILHO	3	5.76488750	1.92162917	7.25	0.0005
SORGO*MILHO	9	15.71135000	1.74570556	6.59	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
REP	3	1.76453750	0.58817917	2.22	0.098
SORGO	3	47.44306250	15.81435417	59.70	<.000
MILHO	3	5.76488750	1.92162917	7.25	0.0005
SORGO*MILHO	9	15.71135000	1.74570556	6.59	<.0001

FATORIAL MULTIVARIADO
The GLM Procedure
Multivariate Analysis of Variance

E = Error SSCP Matrix

	X1	X2
X1	13.83043125	-2.0874375
X2	-2.0874375	11.9212625

Partial Correlation Coefficients from the Error SSCP Matrix / Prob > |r|

DF = 45	X1	X2
X1	1.000000	-0.162568 0.2804
X2	-0.162568 0.2804	1.000000

FATORIAL MULTIVARIADO
The GLM Procedure
Multivariate Analysis of Variance

H = Type III SSCP Matrix for SORGO

	X1	X2
X1	2.58503125	-7.92764375
X2	-7.92764375	47.4430625

Characteristic Roots and Vectors of: E Inverse * H, where

H = Type III SSCP Matrix for SORGO

E = Error SSCP Matrix

Characteristic Root	Percent	Characteristic Vector X1	V'EV=1 X2
3.97993235	97.70	0.00210178	0.28998630
0.09359613	2.30	0.27251176	0.04548365

MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall SORGO Effect

H = Type III SSCP Matrix for SORGO

E = Error SSCP Matrix

Statistic	S=2 Value	M=0 F Value	N=21 Num DF	Den DF	Pr > F
Wilks' Lambda	0.18361983	19.56	6	88	<.0001
Pillai's Trace	0.88477971	11.90	6	90	<.0001
Hotelling-Lawley Trace	4.07352848	29.55	6	56.923	<.0001
Roy's Greatest Root	3.97993235	59.70	3	45	<.0001

NOTE: F Statistic for Roy's Greatest Root is an upper bound.

NOTE: F Statistic for Wilks' Lambda is exact.

H = Type III SSCP Matrix for MILHO
X1 X2

X1	14.78358125	0.179825
X2	0.179825	5.7648875

FATORIAL MULTIVARIADO
The GLM Procedure
Multivariate Analysis of Variance

Characteristic Roots and Vectors of: E Inverse * H, where
H = Type III SSCP Matrix for MILHO
E = Error SSCP Matrix

Characteristic Root	Percent	Characteristic Vector X1	V'EV=1 X2
1.12938129	70.62	0.26968594	0.08888237
0.46993649	29.38	-0.03919914	0.27975121

MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall MILHO Effect

H = Type III SSCP Matrix for MILHO
E = Error SSCP Matrix

S=2 M=0 N=21

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.31948318	11.28	6	88	<.0001
Pillai's Trace	0.85007852	11.09	6	90	<.0001
Hotelling-Lawley Trace	1.59931778	11.60	6	56.923	<.0001
Roy's Greatest Root	1.12938129	16.94	3	45	<.0001

NOTE: F Statistic for Roy's Greatest Root is an upper bound.

NOTE: F Statistic for Wilks' Lambda is exact.

H = Type III SSCP Matrix for SORGO*MILHO

	X1	X2
X1	2.26093125	-0.78240625
X2	-0.78240625	15.71135

Characteristic Roots and Vectors of: E Inverse * H, where
H = Type III SSCP Matrix for SORGO*MILHO
E = Error SSCP Matrix

Characteristic Root	Percent	Characteristic Vector X1	V'EV=1 X2
1.33882264	89.18	0.03632082	0.29340290
0.16244353	10.82	0.27008864	0.00869227

FATORIAL MULTIVARIADO
The GLM Procedure
Multivariate Analysis of Variance

MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall SORGO*MILHO Effect

H = Type III SSCP Matrix for SORGO*MILHO
E = Error SSCP Matrix

	S=2	M=3	N=21		
Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.36781620	3.17	18	88	0.0002
Pillai's Trace	0.71217760	2.77	18	90	0.0008
Hotelling-Lawley Trace	1.50126617	3.61	18	69.649	<.0001
Roy's Greatest Root	1.33882264	6.69	9	45	<.0001

NOTE: F Statistic for Roy's Greatest Root is an upper bound.

NOTE: F Statistic for Wilks' Lambda is exact.

4.10. VARIÁVEIS CANÔNICAS

Em uma análise de variância multivariada (MANOVA) com k tratamentos, usualmente testa-se a hipótese $H_0 : \underline{\mu}_1 = \underline{\mu}_2 = \dots = \underline{\mu}_k$, a qual é equivalente ao teste de que não há diferença entre os vetores de efeitos de tratamentos $\underline{t}_1, \underline{t}_2, \dots, \underline{t}_k$.

Considerando um estudo com k tratamentos, então os vetores $\underline{\mu}_i = [\mu_{i1} \ \mu_{i2} \ \dots \ \mu_{ip}]'$ com $i = 1, 2, \dots, k$, podem ser considerados como sendo as coordenadas de pontos no espaço p -dimensional. Quando H_0 é verdadeira, os vetores $\underline{\mu}_1, \underline{\mu}_2, \dots, \underline{\mu}_k$ são idênticos. Assim, se d é a dimensão do hiperplano gerado por $\underline{\mu}_1, \underline{\mu}_2, \dots, \underline{\mu}_k$, então H_0 verdadeira implica em $d = 0$. Em qualquer caso, tem-se que:

$$d \leq \min(p, q) = t, \text{ com } q = k - 1.$$

Se H_0 é rejeitada, é de importância determinar a real dimensionalidade d , onde $d = 0, 1, \dots, t$.

Se $d = t$, não há nenhuma restrição sobre os $\underline{\mu}_i$'s, e $d < t$ ocorre, se e somente se, houver exatamente $s = t - d$ relações independentes linearmente entre os k vetores de médias.

Considere agora a equação $|H^* - \lambda n_e \sum| = 0$, onde H^* é definida em função das médias verdadeiras e $E^* = n_e \sum$ substitui o estimador E com n_e graus de liberdade. O posto de H^* e, portanto, o número de raízes não nulas da equação determinantal, dependerá da dimensionalidade dos $\underline{\mu}_i$'s e, inversamente, o número de raízes não nulas determinará a dimensionalidade das médias. Dimensionalidade, portanto, é a ordem do hiperplano gerado pelas diferentes médias dos tratamentos. Se $d = 0$, então as médias são coincidentes, se $d = 1$, então as médias são colineares, etc.

Claro está, que a dimensionalidade em termos de médias populacionais, é o número de autovalores não nulos de $E^{*-1}H^*$. Entretanto, quando se trabalha com dados observados, um

autovalor pode ser muito pequeno sem propriamente ser nulo. Um teste para verificação da dimensionalidade torna-se, então, necessário. A aproximação mais adequada, nesse caso, é aquela proposta por Bartlett (1947). O teste é feito sequencialmente para $\mathbf{d} = 0$, $\mathbf{d} = 1$, etc., até que um resultado não significativo apareça. Se até $\mathbf{d} - 1$ obtiver resultados significativos, mas em \mathbf{d} não, infere-se, então, que a dimensionalidade é \mathbf{d} . A estatística proposta por Bartlett (1947) é dada por:

$$D_d^2 = \left(n_e - \frac{p-q+1}{2} \right) \sum_{j=d+1}^p \ln(1 + \lambda_j),$$

em que:

λ_j com $j=1, 2, \dots, p$ são autovalores da matriz $E^{-1}H$.

A estatística D_d^2 , assintoticamente, tem distribuição qui-quadrado (χ^2) com $f = (p - d)(q - d)$ graus de liberdade, isto é, $D_d^2 \sim \chi_f^2$.

Assumindo que \mathbf{d} é a dimensionalidade do hiperplano gerado pelas médias verdadeiras, pode-se olhar de modo mais aprofundado a separação das médias dentro do plano. Seja dessa maneira, \underline{v}_j o autovetor associado ao autovalor λ_j de $E^{-1}H$, onde \underline{v}_j é normalizado de tal modo

que $\underline{v}_j' \frac{E}{n_e} \underline{v}_j = 1$ e é denominado j -ésimo **vetor canônico**.

A projeção de um ponto \underline{X} (vetor de observações) sobre o hiperplano estimado pode ser representada em termos de coordenadas canônicas \mathbf{d} -dimensional ($\underline{v}_1' \underline{X}, \dots, \underline{v}_d' \underline{X}$). Em particular, as **médias canônicas** dos \mathbf{k} -tratamentos, $\hat{m}c_i = [\underline{v}_1' \hat{m}_i, \dots, \underline{v}_d' \hat{m}_i]'$, com $i=1, 2, \dots, k$, representam a projeção do grupo de médias sobre esse hiperplano e podem ser usadas para estudar as diferenças entre grupos (tratamentos).

O vetor \underline{v}_j é o **vetor canônico** para a j -ésima **variável canônica** $VC_j = \underline{v}_j' \underline{X}$. As variáveis canônicas são funções discriminantes ótimas, ou seja, maximizam a variação entre tratamentos (expressa aqui por H) em relação à variação residual (expressa aqui por E), sujeitas à restrição de que não são correlacionadas entre si. Os gráficos para $\mathbf{d} = 1$ ou $\mathbf{d} = 2$ envolvendo as médias canônicas podem representar uma ajuda importante na discriminação dos tratamentos.

Note que,

$$\frac{\lambda_1 + \lambda_2 + \dots + \lambda_d}{\text{traço}(\mathbf{E}^{-1}\mathbf{H})} = \frac{\lambda_1 + \lambda_2 + \dots + \lambda_d}{\lambda_1 + \lambda_2 + \dots + \lambda_p},$$

representa a proporção da variação entre tratamentos que é explicada pelas primeiras **d** variáveis canônicas.

Fatos: i) $\tilde{\mathbf{v}}_j' \frac{\mathbf{E}}{n_e} \tilde{\mathbf{v}}_j = 1$;

ii) $\tilde{\mathbf{v}}_j' \frac{\mathbf{E}}{n_e} \tilde{\mathbf{v}}_{j'} = 0$, $j \neq j'$.

Uma vez que as variáveis canônicas são normalizadas por $\tilde{\mathbf{v}}_j' \frac{\mathbf{E}}{n_e} \tilde{\mathbf{v}}_j = 1$, então se

$\hat{\mathbf{m}}\mathbf{c}_i = [\tilde{\mathbf{v}}_1' \hat{\mathbf{m}}_i, \tilde{\mathbf{v}}_2' \hat{\mathbf{m}}_i, \dots, \tilde{\mathbf{v}}_d' \hat{\mathbf{m}}_i]'$, então sob normalidade, segue que:

$\hat{\mathbf{m}}\mathbf{c}_i \sim N_d \left\{ [\tilde{\mathbf{v}}_1' \mu_i, \tilde{\mathbf{v}}_2' \mu_i, \dots, \tilde{\mathbf{v}}_d' \mu_i]', \left(\frac{1}{r_i}\right) \mathbf{I}_{(d)} \right\}$, onde r_i é o número de repetições para o

tratamento *i*.

A região de confiança (aproximada) para a *i*-ésima verdadeira média canônica $\mu\mathbf{c}_i = [\tilde{\mathbf{v}}_1' \mu_i, \tilde{\mathbf{v}}_2' \mu_i, \dots, \tilde{\mathbf{v}}_d' \mu_i]'$ ao nível de confiança $100(1-\alpha)\%$ será dada através do disco de

raio $\left\{ \frac{1}{r_i} \chi_{(d;\alpha)}^2 \right\}^{1/2}$ construído em torno da média $\hat{\mathbf{m}}\mathbf{c}_i$.

No caso de **p** variáveis observadas (X_1, X_2, \dots, X_p) se obtêm, em geral, **p** autovalores distintos, todos positivos. Postos em ordem decrescente, cada um deles nos dá uma variável canônica e, a cada variável canônica corresponde uma porcentagem do total. Nos casos favoráveis, considera-se então, apenas um número reduzido **d** de variáveis canônicas, com **d** bem menor do que **p**. Verifica-se, pois, que este método pode levar a uma simplificação considerável dos problemas em que seja grande o número de variáveis observadas. A seguir são apresentados dois exemplos de aplicação.

Exemplo 1 (Mardia, et al., 1997)

Considerando dados de um experimento inteiramente casualizado com 3 tratamentos e 50 repetições, com 4 variáveis (X_1, X_2, X_3, X_4), obteve-se os resultados da MANOVA apresentados na Tabela 4.18. Outros resultados são apresentados na sequência.

Tabela 4.18 – Resultados da análise de variância multivariada

Causas de variação	G.L.	Matrizes de somas de quadrados e de produtos	Λ
Tratamentos	q = 2	$H = \begin{bmatrix} 63,21 & -19,95 & 165,25 & 71,28 \\ -19,95 & 11,35 & -57,24 & -22,93 \\ 165,25 & -57,24 & 437,11 & 186,78 \\ 71,28 & -22,93 & 186,78 & 80,41 \end{bmatrix}$	0,0234**
Resíduo	n _e = 147	$E = \begin{bmatrix} 38,6 & 13,63 & 24,62 & 5,64 \\ 13,63 & 16,96 & 8,12 & 4,81 \\ 24,62 & 8,12 & 27,22 & 6,27 \\ 5,64 & 4,81 & 6,27 & 6,16 \end{bmatrix}$	

** P < 0,01

• Vetores de médias observadas

$$\hat{\tilde{m}}_1 = \begin{bmatrix} 5,006 \\ 3,428 \\ 1,462 \\ 0,246 \end{bmatrix}, \quad \hat{\tilde{m}}_2 = \begin{bmatrix} 5,936 \\ 2,770 \\ 4,260 \\ 1,326 \end{bmatrix}, \quad \hat{\tilde{m}}_3 = \begin{bmatrix} 6,588 \\ 2,974 \\ 5,552 \\ 2,026 \end{bmatrix}$$

• Autovalores de $E^{-1}H$

$$\lambda_1 = 32,1877, \quad \lambda_2 = 0,2853, \quad \lambda_3 = 0, \quad \lambda_4 = 0$$

• Autovetores normalizados por $\tilde{v}_j' \frac{E}{n_e} \tilde{v}_j = 1$

$$d \leq \min(p, q) = t \Rightarrow d \leq 2$$

$$\tilde{v}_1' = [0,8281 \quad 1,5362 \quad -2,2006 \quad -2,8092]$$

$$\tilde{v}_2' = [-0,0194 \quad -2,1715 \quad -0,9239 \quad -2,8250]$$

Primeira variável canônica:

$$VC_1 = 0,8281X_1 + 1,5362X_2 - 2,2006X_3 - 2,8092X_4$$

Segunda variável canônica:

$$VC_2 = -0,0194X_1 - 2,1715X_2 + 0,9239X_3 - 2,8250X_4$$

• Teste de dimensionalidade

Vimos que:

$$D_d^2 = \left(n_e - \frac{p-q+1}{2} \right) \sum_{j=d+1}^p \ln(1 + \lambda_j) \sim \chi^2 \text{ com } f = (p-d)(q-d) \text{ graus de liberdade.}$$

Assim, para d = 0, virá:

$$D_0^2 = \left(147 - \frac{4-2+1}{2} \right) \sum_{j=1}^4 \ln(1 + \lambda_j)$$

$$= 145,5 (\ln 33,1877 + \ln 1,2853)$$

$$= 145,5 (3,7532) = 546,09$$

$$f = (4 - 0)(2 - 0) = 8 \quad \chi^2_{1\%}(8) = 20,90$$

Como $546,09 > 20,90$, temos que o resultado é significativo ($P < 0,01$).

Para $d = 1$, virá:

$$D_1^2 = \left(147 - \frac{4 - 2 + 1}{2} \right) \sum_{j=2}^4 \ln(1 + \lambda_j)$$

$$= 145,5 (\ln 1,2853)$$

$$= 145,5 (0,2510) = 36,52$$

$$f = (4 - 1)(2 - 1) = 3 \quad \chi^2_{1\%}(3) = 11,34$$

Como $36,52 > 11,34$, temos que o resultado é significativo ($P < 0,01$).

Sendo $\lambda_3 = \lambda_4 = 0$ e pelos resultados do teste, podemos concluir que a dimensionalidade é 2. Uma vez que $t = 2$, ambas as variáveis canônicas são necessárias e, a dimensão não pode ser reduzida.

• Médias canônicas

As médias canônicas para os três tratamentos (com as variáveis não centradas na média) são:

$$\text{Tratamento 1: } \hat{\mathbf{m}} \mathbf{c}_1 = \begin{bmatrix} \hat{\mathbf{v}}_1' \hat{\mathbf{m}}_1 \\ \hat{\mathbf{v}}_2' \hat{\mathbf{m}}_1 \end{bmatrix} = \begin{bmatrix} 5,503 \\ -6,885 \end{bmatrix}$$

$$\text{Tratamento 2: } \hat{\mathbf{m}} \mathbf{c}_2 = \begin{bmatrix} \hat{\mathbf{v}}_1' \hat{\mathbf{m}}_2 \\ \hat{\mathbf{v}}_2' \hat{\mathbf{m}}_2 \end{bmatrix} = \begin{bmatrix} -3,929 \\ -5,940 \end{bmatrix}$$

$$\text{Tratamento 3: } \hat{\mathbf{m}} \mathbf{c}_3 = \begin{bmatrix} \hat{\mathbf{v}}_1' \hat{\mathbf{m}}_3 \\ \hat{\mathbf{v}}_2' \hat{\mathbf{m}}_3 \end{bmatrix} = \begin{bmatrix} -7,885 \\ -7,180 \end{bmatrix}$$

Na Figura 4.1, temos a dispersão dos três tratamentos em relação às duas variáveis canônicas. A região de confiança para as três médias canônicas a 99% de confiança é dada pelos

$$\text{círculos de raio } \left\{ \frac{1}{50} \chi^2_{(2;0,01)} \right\}^{1/2} = (9,21/50)^{1/2} = 0,429.$$

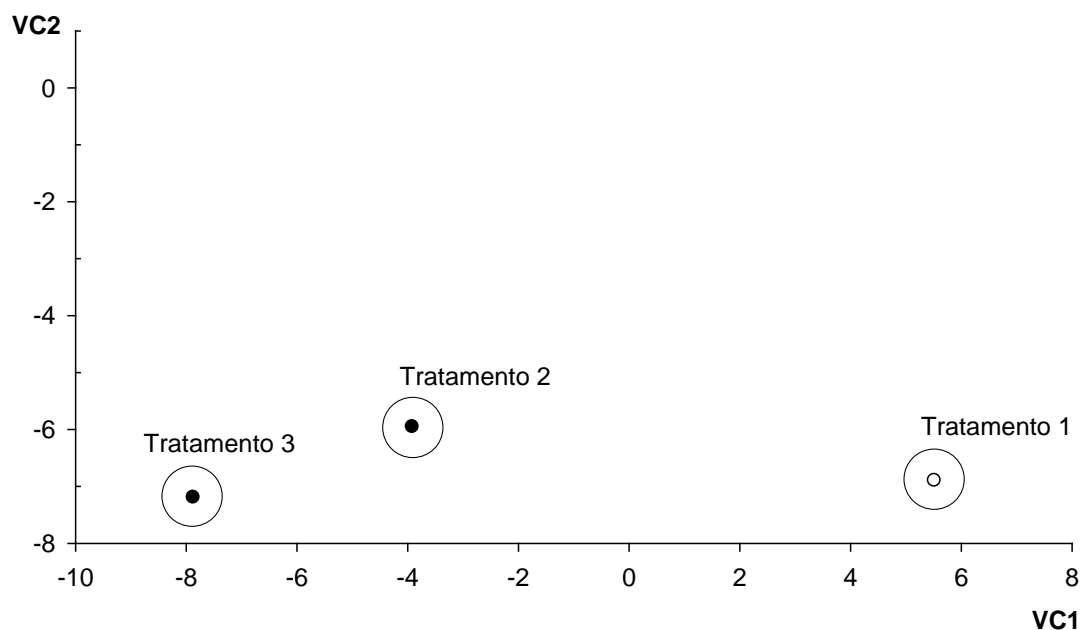


Figura 4.1 – Dispersão dos três tratamentos em relação às duas variáveis canônicas

Pela Figura 4.1, verifica-se que os três tratamentos são totalmente diferentes, mas os tratamentos 2 e 3 estão mais próximos um do outro. É importante observar que, na construção do diagrama de dispersão, deve-se usar a mesma escala para VC_1 e VC_2 , pois, em certos casos, o uso de escalas diferentes pode proporcionar impressionantes equívocos na interpretação.

Os escores para os três tratamentos (com as variáveis não centradas na média), obtidos a partir das duas variáveis canônicas são apresentados na Tabela 4.19.

Tabela 4.19 – Escores para os três tratamentos

Tratamentos	VC_1	VC_2
1	5,503	-6,885
2	-3,929	-5,940
3	-7,885	-7,180
Variância	47,31	0,42
Covariância	0,00	

$$s^2(\text{VC}_1) = \frac{107,8933 - \frac{(-6,311)^2}{3}}{3-1} = 47,31 = \lambda_1^*$$

$$s^2(\text{VC}_2) = \frac{134,2392 - \frac{(-20,005)^2}{3}}{3-1} = 0,42 = \lambda_2^*$$

$$\text{Cov}(\text{VC}_1, \text{VC}_2) = \frac{42,0644 - \frac{(-6,311)(-20,005)}{3}}{3-1} = 0$$

λ_1^* e λ_2^* são autovalores da matriz $S^{-1}T$,

em que,

$S = \frac{E}{n_e}$ é a matriz de covariância residual;

$T = \frac{(H/q)}{r} = \frac{H}{r \cdot q}$ é a matriz de covariância entre médias de tratamentos;

r é o número de repetições;

q é o número de graus de liberdade para tratamentos.

Os autovalores da matriz $E^{-1}H$ foram $\lambda_1 = 32,1877$ e $\lambda_2 = 0,2853$. Note que

$\lambda_i^* = \frac{n_e}{r \cdot q} \lambda_i$. Este fato não traz nenhum problema, pois pode-se obter os autovalores de

$E^{-1}H$ ou $S^{-1}T$, que os resultados e conclusões serão exatamente os mesmos. Mardia et al. (1997), assim como o programa SAS, dentre outros, apresentam os autovalores da matriz $E^{-1}H$. Cruz, et al. (2012), apresentaram um estudo sobre divergência genética por variáveis canônicas, a partir das matrizes T e S , onde as variáveis canônicas foram estimadas a partir de dados transformados, cujas funções de transformação foram estabelecidas pelo processo de condensação pivotal. Se tivessem obtido os autovalores da matriz $E^{-1}H$, chegariam aos mesmos resultados e conclusões.

Exemplo 2

A seguir é apresentado um estudo por variáveis canônicas, considerando os dados apresentados na Tabela 4.13 e os resultados da MANOVA na Tabela 4.14.

Os vetores de médias observadas foram:

$$\hat{m}_1 = \begin{bmatrix} 4,678 \\ 1,004 \end{bmatrix}, \quad \hat{m}_2 = \begin{bmatrix} 6,112 \\ 1,096 \end{bmatrix}, \quad \hat{m}_3 = \begin{bmatrix} 4,600 \\ 0,898 \end{bmatrix}.$$

• Autovalores de $E^{-1}H$

$$\lambda_1 = 34,7579 \quad \text{e} \quad \lambda_2 = 0,3690$$

• **Autovetores normalizados por** $\tilde{\mathbf{v}}_j' \frac{\mathbf{E}}{\mathbf{n}_e} \tilde{\mathbf{v}}_j = 1$

Seja $\tilde{\mathbf{a}}_j$ um autovetor associado a λ_j e $\tilde{\mathbf{a}}_j' \frac{\mathbf{E}}{\mathbf{n}_e} \tilde{\mathbf{a}}_j = \mathbf{K}_j$. Então, o vetor canônico $\tilde{\mathbf{v}}_j$ é

dado por:

$$\tilde{\mathbf{v}}_j = \frac{1}{\sqrt{\mathbf{K}_j}} \cdot \tilde{\mathbf{a}}_j.$$

Assim, virá:

$$\tilde{\mathbf{v}}_1' = [9,043351 \quad -14,899349]$$

$$\tilde{\mathbf{v}}_2' = [-1,272303 \quad 13,462657].$$

Pode-se verificar que: (i) $\tilde{\mathbf{v}}_j' \frac{\mathbf{E}}{\mathbf{n}_e} \tilde{\mathbf{v}}_j = 1$

$$(ii) \tilde{\mathbf{v}}_j' \frac{\mathbf{E}}{\mathbf{n}_e} \tilde{\mathbf{v}}_{j'} = 0, \quad \text{para } j \neq j'$$

Um resumo da análise está apresentado na Tabela 4.20.

Tabelas 4.20 – Variáveis canônicas estabelecidas pela combinação linear das duas variáveis originais X1 e X2

Variáveis Canônicas	Autovalores de $\mathbf{E}^{-1}\mathbf{H}$	Coeficientes de ponderação		Proporção da variância	Proporção da variância acumulada
		X ₁	X ₂		
VC ₁	34,5779	9,043351	-14,899349	0,9895	0,9805
VC ₂	0,3690	-1,272303	13,462657	0,0105	1,0000

• Médias canônicas

As médias canônicas para os três tratamentos (com as variáveis não centradas na média) são:

$$\text{Tratamento 1: } \hat{\mathbf{m}}_1 \mathbf{c}_1 = \begin{bmatrix} \tilde{\mathbf{v}}_1' \hat{\mathbf{m}}_1 \\ \tilde{\mathbf{v}}_2' \hat{\mathbf{m}}_1 \end{bmatrix} = \begin{bmatrix} 27,3458 \\ 7,5647 \end{bmatrix}$$

$$\text{Tratamento 2: } \hat{\mathbf{m}}_2 \mathbf{c}_2 = \begin{bmatrix} \tilde{\mathbf{v}}_1' \hat{\mathbf{m}}_2 \\ \tilde{\mathbf{v}}_2' \hat{\mathbf{m}}_2 \end{bmatrix} = \begin{bmatrix} 38,9433 \\ 6,9788 \end{bmatrix}$$

$$\text{Tratamento 3: } \hat{\mathbf{m}}_3 \mathbf{c}_3 = \begin{bmatrix} \mathbf{v}_1' \hat{\mathbf{m}}_3 \\ \mathbf{v}_2' \hat{\mathbf{m}}_3 \end{bmatrix} = \begin{bmatrix} 28,2198 \\ 6,2369 \end{bmatrix}$$

Os escores para os três tratamentos (com as variáveis não centradas na média), obtidos a partir das duas variáveis canônicas são apresentados na Tabela 4.21.

Tabela 4.21 – Escores para os três tratamentos

Tratamentos	VC ₁	VC ₂
1	27,3458	7,5647
2	38,9433	6,9788
3	28,2198	6,2369
Variância	41,7099	0,4428
Covariância	0,00	

4.10.1. Sobre o descarte de variáveis

Na análise por variáveis canônicas, pode-se adotar o critério de avaliação da importância das variáveis X_1, X_2, \dots, X_p a partir dos coeficientes de ponderação de cada variável nas últimas variáveis canônicas que, por estimação, retêm proporção mínima da variação total. Como os coeficientes (elementos dos autovetores) são influenciados pela escala de avaliação das variáveis, recomenda-se que a avaliação seja feita a partir de coeficientes associados às variáveis padronizadas, isto é:

$$\partial_j = c_j \sqrt{\hat{\sigma}_j^2},$$

em que,

∂_j é o coeficiente de ponderação associado à variável padronizada

$$x_j \left(x_j = \frac{X_j}{\sqrt{\hat{\sigma}_j^2}} \text{ ou } x_j = \frac{X_j - \bar{X}_j}{\sqrt{\hat{\sigma}_j^2}} \right);$$

c_j é o coeficiente de ponderação associado à variável original X_j ;

$\hat{\sigma}_j^2 = \hat{\sigma}_{jj}$ é o quadrado médio do resíduo associado à j -ésima variável.

Considerando todos os autovalores não nulos, identificam-se as variáveis de menor importância para a divergência entre os tratamentos avaliados, como sendo aquelas cujos coeficientes de ponderação ∂_j , são de maior magnitude, em valor absoluto, nas últimas variáveis canônicas. Assim, se VC_d é a variável canônica de menor importância relativa, dada por:

$$VC_d = \partial_{d1}x_1 + \partial_{d2}x_2 + \dots + \partial_{dp}x_p,$$

em que, x_1, x_2, \dots, x_p são as variáveis originais padronizadas, então identifica-se a variável de menor importância como sendo aquela associada ao maior dos elementos (em valor absoluto) $\partial_{d1}, \partial_{d2}, \dots, \partial_{dp}$. A segunda variável de menor importância é identificada com o mesmo critério, pelos coeficientes da variável canônica VC_{d-1} , e assim sucessivamente.

Baseado no princípio que a importância ou variância das variáveis canônicas decresce da primeira para a última, tem-se que as últimas variáveis canônicas são responsáveis pela explicação de uma fração muito pequena da variância total. Assim, a variável X_j que domina (aquela que possui maior coeficiente de ponderação ∂_j em valor absoluto) a variável canônica de menor autovalor (menor variância) deve ser a menos importante para explicar a variância total e, portanto, passível de descarte. A razão é que variáveis com maior ∂_j em valor absoluto nas variáveis canônicas de menor variância, representam variação praticamente insignificante.

Quando em uma variável canônica de menor variância, o maior coeficiente de ponderação ∂_j (em valor absoluto) estiver associado a uma variável já previamente descartada, tem-se optado por não fazer nenhum outro descarte com base nos coeficientes daquela variável canônica, mas prosseguir a identificação da importância das variáveis na outra variável canônica e de variância imediatamente superior (Cruz, et al., 2012).

Apesar de o exemplo em questão incluir somente duas variáveis, como ilustração, será utilizado o critério descrito anteriormente, cujos coeficientes ∂_j são apresentados na Tabela 4.22.

Tabela 4.22 – Coeficientes de ponderação associados às variáveis padronizadas

Variáveis canônicas	Coeficientes de ponderação (∂_j) associados a:	
	x_1	x_2
VC_1	1,766650	-1,323696
VC_2	-0,248549	1,196057

Pela análise, pode-se concluir que a variável de menor importância é a X_2 , com maior coeficiente de ponderação em VC_2 . Neste tipo de análise, consideram-se de menor importância aquelas variáveis que são relativamente invariantes ou que apresentam redundância. Como X_2 apresenta estatística F ($F_{\text{Treat}} = 6,22$) inferior à de X_1 ($F_{\text{Treat}} = 94,96$), ela se torna comparativamente menos importante, sendo, portanto, indicado o seu descarte.

• Teste de dimensionalidade

Assim, para $d = 0$, virá:

$$D_0^2 = \left(12 - \frac{2 - 2 + 1}{2} \right) \sum_{j=1}^2 \ln(1 + \lambda_j)$$

$$= 11,5 (\ln 35,7579 + \ln 1,3690) = 44,74^{**}$$

** (P < 0,01)

$$f = (2 - 0)(2 - 0) = 4 \quad \chi^2_{1\%}(4) = 13,28$$

Para d = 1, virá:

$$D_1^2 = \left(12 - \frac{2-2+1}{2}\right) \ln 1,3690 = 3,61^{n.s.} \quad n.s. (P > 0,05)$$

$$f = (2 - 1)(2 - 1) = 1 \quad \chi^2_{5\%}(1) = 3,84$$

Pelos resultados do teste, pode-se concluir que a dimensionalidade é 1. Assim, somente a primeira variável canônica seria necessária, e a dimensão poderia ser reduzida.

A título de ilustração, vamos continuar considerando as duas variáveis canônicas. Na Figura 4.2 apresenta-se a dispersão dos três tratamentos em relação às duas variáveis canônicas. A região de confiança para as três médias canônicas a 95% de confiança é dada pelos círculos

de raio $\left\{ \frac{1}{5} \chi^2_{(2;0,05)} \right\}^{1/2} = (5,99/5)^{1/2} = 1,094$.

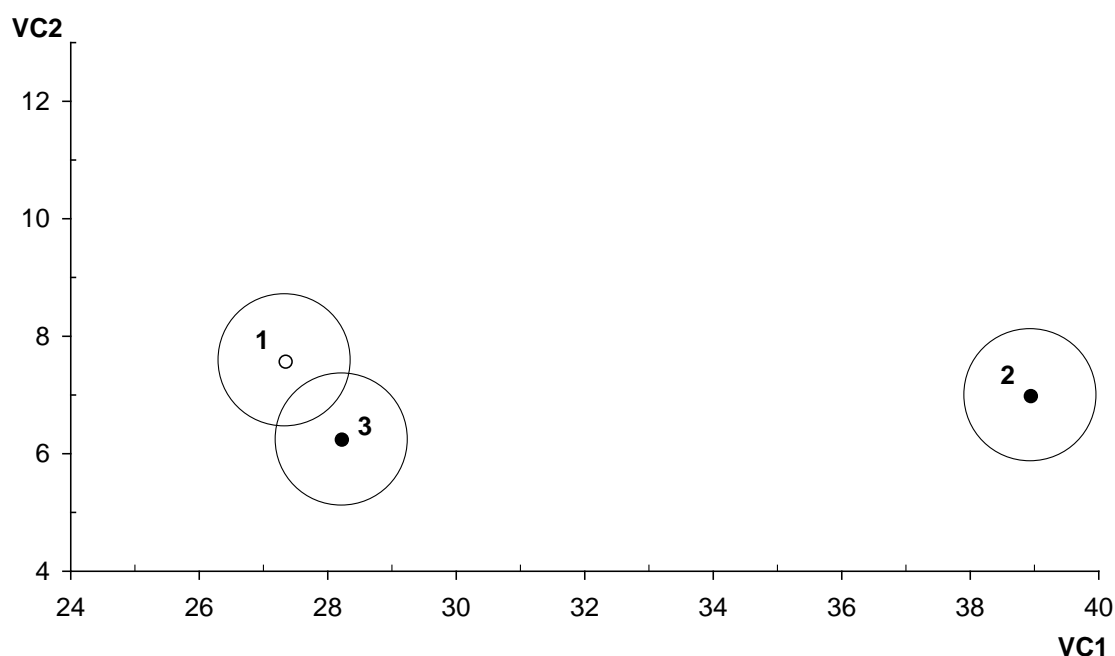


Figura 4.2 – Dispersão dos três tratamentos em relação às duas variáveis canônicas

• Medidas de dissimilaridade

i) Distância euclidiana

A distância euclidiana entre dois pontos \tilde{X}_i e $\tilde{X}_{i'}$, no espaço p-dimensional é definida

como:

$$d_{ii'} = \left[\left(\tilde{X}_i - \tilde{X}_{i'} \right)' \left(\tilde{X}_i - \tilde{X}_{i'} \right) \right]^{\frac{1}{2}} = \left[\sum_{j=1}^p (\tilde{X}_{ij} - \tilde{X}_{i'j})^2 \right]^{\frac{1}{2}}.$$

Assim, a distância gráfica com as variáveis não centradas na média, entre cada par de tratamentos na Figura 4.2, dada pela distância euclidiana é:

$$dVC_{ii'} = \sqrt{(VC_{i1} - VC_{i'1})^2 + (VC_{i2} - VC_{i'2})^2}.$$

$$dVC_{12} = 11,61$$

$$dVC_{13} = 1,59$$

$$dVC_{23} = 10,75$$

Na Tabela 4.23 são apresentadas as distâncias entre os pares de tratamentos.

Tabela 4.23 – Dissimilaridade entre tratamentos, baseada nas distâncias euclidianas obtidas a partir dos escores das duas variáveis canônicas

Tratamentos	2	3
1	11,61	1,59
2		10,75

ii) Distância de Mahalanobis (D^2)

• Médias observadas

\bar{X}_{ij} , $i = 1, 2, \dots, k$ (k tratamentos) e $j = 1, 2, \dots, p$ (p variáveis)

$\bar{\tilde{X}}_i = \hat{m}_i$ é o vetor de médias observadas para o tratamento i.

• Desvios

$d_j = \bar{X}_{ij} - \bar{X}_{i'j}$ representa a diferença entre as médias observadas dos tratamentos i e i' para a variável j.

• Distância

Assim, a distância de Mahalanobis (na verdade é a distância ao quadrado) entre $\bar{\tilde{X}}_i$ e

$\bar{\tilde{X}}_{i'}$, é dada por:

$$D_{ii'}^2 = \tilde{\delta}' S^{-1} \tilde{\delta},$$

em que,

$$\tilde{\delta}' = \begin{bmatrix} d_1 & d_2 & \dots & d_p \end{bmatrix} \text{ e } S = \frac{E}{n_e},$$

ou ainda,

$$D_{ii'}^2 = \left(\bar{\tilde{X}}_i - \bar{\tilde{X}}_{i'} \right)' S^{-1} \left(\bar{\tilde{X}}_i - \bar{\tilde{X}}_{i'} \right).$$

Sob multinormalidade e homogeneidade das matrizes de covariâncias, D^2 é chamada distância generalizada de Mahalanobis (ao quadrado).

No exemplo, $D_{12}^2 = 134,8436$, $D_{13}^2 = 2,5268$ e $D_{23}^2 = 115,5433$.

Um dos objetivos da análise por variáveis canônicas é proporcionar uma simplificação estrutural nos dados, de modo que a diferença entre tratamentos, influenciada a princípio por um conjunto de \mathbf{p} -dimensional (\mathbf{p} é o número de variáveis consideradas no estudo), possa ser avaliada por um complexo no espaço bi ou tridimensional de fácil interpretação geométrica. A eficácia de sua utilização depende do grau de distorção provocado nas distâncias entre tratamentos quando se passa do espaço \mathbf{p} -dimensional para o \mathbf{d} -dimensional ($\mathbf{d} < \mathbf{p}$).

Como as distâncias gráficas, em relação a eixos que representam as primeiras variáveis canônicas, são influenciadas pelas variações entre (variâncias e covariâncias entre tratamentos) e dentro (variâncias e covariâncias residuais), pode-se quantificar o grau de distorção destas distâncias comparando o seu total com o total das distâncias de Mahalanobis, isto é:

$$\text{Grau de distorção} = 1 - \theta,$$

sendo:

$$\theta = \frac{\sum_{i < i'} \sum d_{ii'}^2}{\sum_{i < i'} \sum D_{ii'}^2},$$

em que:

$d_{ii'}^2$ = quadrado da distância euclidiana estimada a partir dos escores de \mathbf{d} variáveis canônicas;

$D_{ii'}^2$ = distância de Mahalanobis estimada a partir de \mathbf{p} variáveis originais.

Pode-se demonstrar que θ é também a proporção acumulada da variância explicada pelas \mathbf{d} variáveis canônicas.

No exemplo em questão, considerando as duas variáveis canônicas, tem-se que o grau de distorção é zero, pois, neste caso, $\mathbf{d} = \mathbf{p}$, e $\theta = 1$.

$$\theta = \frac{(11,61)^2 + (1,59)^2 + (10,75)^2}{134,8436 + 2,5268 + 115,5433} = \frac{252,8827}{252,9168} \cong 1.$$

Assim, $\theta = 100\%$, que é exatamente a porcentagem de variância acumulada pelas duas variáveis canônicas. A seguir estão apresentados o programa e alguns resultados para os dados da Tabela 4.13, utilizando o PROC CANDISC do Sistema SAS.

4.10.2. Resultados obtidos através do programa SAS

(a) Programa

```

OPTIONS NODATE NONUMBER;
/* EXEMPLO DE VARIÁVEL CANONICA */
DATA VARCAN;
  INPUT TRAT REP X1 X2;
CARDS;
1 1 4.63 0.95
1 2 4.38 0.89
1 3 4.94 1.01
1 4 4.96 1.23
1 5 4.48 0.94
2 1 6.03 1.08
2 2 5.96 1.05
2 3 6.16 1.08
2 4 6.33 1.19
2 5 6.08 1.08
3 1 4.71 0.96
3 2 4.81 0.93
3 3 4.49 0.87
3 4 4.43 0.82
3 5 4.56 0.91
;
PROC CANDISC DATA=VARCAN OUT=CAN ALL;
CLASS TRAT;
VAR X1 X2;
PROC PRINT; VAR TRAT X1 X2 CAN1 CAN2;
/* A média de CAN por TRAT dá as medias canônicas com as variáveis
   centradas na média */
RUN;
PROC PLOT;
PLOT CAN2*CAN1 = TRAT / VPOS=20 HPOS=60;
RUN;
/* Supondo DBC, usamos MANOVA e o comando CANONICAL */
PROC GLM DATA=VARCAN;
CLASS REP TRAT;
MODEL X1 X2 = REP TRAT;
MANOVA H = TRAT / PRINTE PRINTH CANONICAL SHORT;
RUN;

```

(b) Resultado da Análise

The SAS System
The CANDISC Procedure

Observations	15	DF Total	14
Variables	2	DF Within Classes	12
Classes	3	DF Between Classes	2

Class Level Information

TRAT	Variable Name	Frequency	Weight	Proportion
1	_1	5	5.0000	0.333333
2	_2	5	5.0000	0.333333
3	_3	5	5.0000	0.333333

The SAS System

The CANDISC Procedure
Within-Class SSCP Matrices

TRAT = 1

Variable	X1	X2
X1	0.2784800000	0.1145400000
X2	0.1145400000	0.0711200000

TRAT = 2

Variable	X1	X2
X1	0.0806800000	0.0285400000
X2	0.0285400000	0.0117200000

TRAT = 3

Variable	X1	X2
X1	0.0988000000	0.0294000000
X2	0.0294000000	0.0118800000

The SAS System
The CANDISC Procedure
Pooled Within-Class SSCP Matrix

Variable	X1	X2
X1	0.4579600000	0.1724800000
X2	0.1724800000	0.0947200000

Between-Class SSCP Matrix

Variable	X1	X2
X1	7.247640000	0.732620000
X2	0.732620000	0.098173333

Total-Sample SSCP Matrix

Variable	X1	X2
X1	7.705600000	0.905100000
X2	0.905100000	0.192893333

The SAS System
The CANDISC Procedure
Within-Class Covariance Matrices

TRAT = 1, DF = 4

Variable	X1	X2
X1	0.0696200000	0.0286350000
X2	0.0286350000	0.0177800000

TRAT = 2, DF = 4

Variable	X1	X2
X1	0.0201700000	0.0071350000
X2	0.0071350000	0.0029300000

TRAT = 3, DF = 4

Variable	X1	X2
X1	0.0247000000	0.0073500000
X2	0.0073500000	0.0029700000

The SAS System
The CANDISC Procedure
Pooled Within-Class Covariance Matrix, DF = 12

Variable	X1	X2
X1	0.0381633333	0.0143733333
X2	0.0143733333	0.0078933333

Between-Class Covariance Matrix, DF = 2

Variable	X1	X2
X1	0.7247640000	0.0732620000
X2	0.0732620000	0.0098173333

Total-Sample Covariance Matrix, DF = 14

Variable	X1	X2
X1	0.5504000000	0.0646500000
X2	0.0646500000	0.0137780952

The SAS System
The CANDISC Procedure
Simple Statistics
Total-Sample

Variable	N	Sum	Mean	Variance	Standard Deviation
X1	15	76.95000	5.13000	0.55040	0.7419
X2	15	14.99000	0.99933	0.01378	0.1174

TRAT = 1

Variable	N	Sum	Mean	Variance	Standard Deviation
X1	5	23.39000	4.67800	0.06962	0.2639
X2	5	5.02000	1.00400	0.01778	0.1333

TRAT = 2

Variable	N	Sum	Mean	Variance	Standard Deviation
X1	5	30.56000	6.11200	0.02017	0.1420
X2	5	5.48000	1.09600	0.00293	0.0541

TRAT = 3

Variable	N	Sum	Mean	Variance	Standard Deviation
X1	5	23.00000	4.60000	0.02470	0.1572
X2	5	4.49000	0.89800	0.00297	0.0545

The SAS System
The CANDISC Procedure

Pairwise Squared Distances Between Groups

$$D^2(i|j) = (\bar{X}_i - \bar{X}_j)' \text{COV}^{-1} (\bar{X}_i - \bar{X}_j)$$

Squared Distance to TRAT

From TRAT	1	2	3
1	0	134.84358	2.52685
2	134.84358	0	115.54332
3	2.52685	115.54332	0

F Statistics, NDF=2, DDF=11 for Squared Distance to TRAT

From TRAT	1	2	3
1	0	154.50826	2.89534
2	154.50826	0	132.39339
3	2.89534	132.39339	0

Prob > Mahalanobis Distance for Squared Distance to TRAT

From TRAT	1	2	3
1	1.0000	<.0001	0.0977
2	<.0001	1.0000	<.0001
3	0.0977	<.0001	1.0000

The SAS System
The CANDISC Procedure
Univariate Test Statistics

F Statistics, Num DF=2, Den DF=12

Variable	Total Standard Deviation	Pooled Standard Deviation	Between Standard Deviation	R-Square	R-Square / RSq)	F Value	Pr > F
X1	0.7419	0.1954	0.8513	0.9406	15.8259	94.96	<.0001
X2	0.1174	0.0888	0.0991	0.5090	1.0365	6.22	0.0140

Average R-Square

Unweighted 0.7247597
Weighted by Variance 0.9300272

Multivariate Statistics and F Approximations

S=2 M=-0.5 N=4.5

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.02042803	32.98	4	22	<.0001
Pillai's Trace	1.24157026	9.82	4	24	<.0001
Hotelling-Lawley Trace	35.12690920	94.48	4	12.235	<.0001
Roy's Greatest Root	34.75791615	208.55	2	12	<.0001

NOTE: F Statistic for Roy's Greatest Root is an upper bound.

NOTE: F Statistic for Wilks' Lambda is exact.

The SAS System
The CANDISC Procedure

	Canonical Correlation	Adjusted Canonical Correlation	Approximate Standard Error	Squared Canonical Correlation
1	0.985918	0.984855	0.007474	0.972034
2	0.519169	.	0.195225	0.269536

Test of H0: The canonical correlations in
the current row and all

Eigenvalues of Inv(E)*H that follow are zero
= CanRsq/(1-CanRsq)

Likelihood Approximate

	Eigenvalue	Difference	Proportion	Cumulative	Ratio	F Value	Num DF	Den DF	Pr > F
1	34.7579	34.3889	0.9895	0.9895	0.02042803	32.98	4	22	<.0001
2	0.3690		0.0105	1.0000	0.73046390	4.43	1	12	0.0571

The SAS System
The CANDISC Procedure
Total Canonical Structure

Variable	Can1	Can2
X1	0.977347	0.211641
X2	0.583786	0.811908

Between Canonical Structure

Variable	Can1	Can2
X1	0.993561	0.113296
X2	0.806782	0.590850

Pooled Within Canonical Structure

Variable	Can1	Can2
X1	0.670428	0.741974
X2	0.139317	0.990248

The SAS System
The CANDISC Procedure

Total-Sample Standardized Canonical Coefficients

Variable	Can1	Can2
X1	6.709167106	-0.943908247
X2	-1.748887635	1.580248455

Pooled Within-Class Standardized Canonical Coefficients

Variable	Can1	Can2
X1	1.766658254	-0.248549972
X2	-1.323724289	1.196082138

Raw Canonical Coefficients

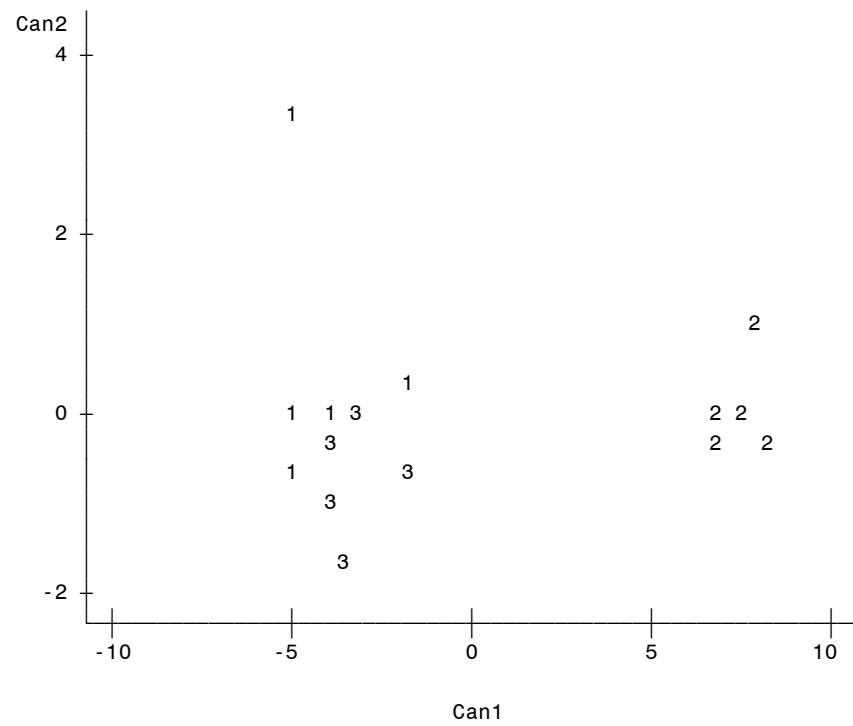
Variable	Can1	Can2
X1	9.04335119	-1.27230305
X2	-14.89934979	13.46265706

Class Means on Canonical Variables

TRAT	Can1	Can2
1	-4.157125035	0.637906711
2	7.440300385	0.051988589
3	-3.283175350	-0.689895300

The SAS System

Plot of Can2*Can1. Symbol is value of TRAT.



The SAS System
The GLM Procedure
Class Level Information

Class	Levels	Values
REP	5	1 2 3 4 5
TRAT	3	1 2 3

Number of observations 15

The SAS System
The GLM Procedure

Dependent Variable: X1

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	7.34090667	1.22348444	26.84	<.0001
Error	8	0.36469333	0.04558667		
Corrected Total	14	7.70560000			

R-Square	Coeff Var	Root MSE	X1 Mean
0.952672	4.161995	0.213510	5.130000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
REP	4	0.09326667	0.02331667	0.51	0.7298
TRAT	2	7.24764000	3.62382000	79.49	<.0001

Source	DF	Type III SS	Mean Square	F Value	Pr > F
REP	4	0.09326667	0.02331667	0.51	0.7298
TRAT	2	7.24764000	3.62382000	79.49	<.0001

The SAS System
The GLM Procedure

Dependent Variable: X2

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	0.12520000	0.02086667	2.47	0.1184
Error	8	0.06769333	0.00846167		
Corrected Total	14	0.19289333			

R-Square	Coeff Var	Root MSE	X2 Mean
0.649063	9.204868	0.091987	0.999333

Source	DF	Type I SS	Mean Square	F Value	Pr > F
REP	4	0.02702667	0.00675667	0.80	0.5586
TRAT	2	0.09817333	0.04908667	5.80	0.0277

Source	DF	Type III SS	Mean Square	F Value	Pr > F
REP	4	0.02702667	0.00675667	0.80	0.5586
TRAT	2	0.09817333	0.04908667	5.80	0.0277

The SAS System
The GLM Procedure
Multivariate Analysis of Variance

E = Error SSCP Matrix

	X1	X2
X1	0.3646933333	0.13198
X2	0.13198	0.0676933333

Partial Correlation Coefficients from the Error SSCP Matrix / Prob > |r|

DF = 8	X1	X2
X1	1.000000	0.839984
		0.0046
X2	0.839984	1.000000
	0.0046	

The SAS System
The GLM Procedure
Multivariate Analysis of Variance

H = Type III SSCP Matrix for TRAT

	X1	X2
X1	7.24764	0.73262
X2	0.73262	0.0981733333

Canonical Analysis

H = Type III SSCP Matrix for TRAT

E = Error SSCP Matrix

	Canonical Correlation	Adjusted Canonical Correlation	Approximate Standard Error	Squared Canonical Correlation
1	0.989139	0.987997	0.006832	0.978396
2	0.588924	.	0.206550	0.346831

Test of H0: The canonical correlations in
the current row and all
that follow are zero

Eigenvalues of Inv(E)*H
= CanRsqr/(1-CanRsqr)

	Eigenvalue	Difference	Proportion	Cumulative	Likelihood Ratio	Approximate F Value	Num DF	Den DF	Pr > F
1	45.2878	44.7568	0.9884	0.9884	0.01411104	25.96	4	14	<.0001
2	0.5310	0.0116		1.0000	0.65316871	4.25	1	8	0.0732

Canonical Structure

	Total		Between		Within	
	Can1	Can2	Can1	Can2	Can1	Can2
X1	0.9732	0.2168	0.9924	0.1232	0.6574	0.7535
X2	0.5683	0.8150	0.8009	0.5988	0.1433	0.9897

Canonical Coefficients

	Standardized		Raw	
	Can1	Can2	Can1	Can2
X1	6.33761265	-0.91774881	8.5425293	-1.2370425
X2	-1.77210772	1.54597544	-15.0971694	13.1706740

The SAS System

The GLM Procedure
Multivariate Analysis of Variance

MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall TRAT Effect

H = Type III SSCP Matrix for TRAT

E = Error SSCP Matrix

S=2 M=-0.5 N=2.5

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.01411104	25.96	4	14	<.0001
Pillai's Trace	1.32522731	7.86	4	16	0.0011
Hotelling-Lawley Trace	45.81876731	78.16	4	7.4839	<.0001
Roy's Greatest Root	45.28776929	181.15	2	8	<.0001

NOTE: F Statistic for Roy's Greatest Root is an upper bound.

NOTE: F Statistic for Wilks' Lambda is exact.

Observação: Escores com as variáveis centradas na média

Can1= 8,5425293 (X1-5,1300) - 15,0971694 (X2-0,9993)

Can2= -1,2370425 (X1-5,1300) + 13,1706740 (X2-0,9993)

Estas equações geram os escores para as n=15 observações. A média de Can por tratamento dá as médias canônicas com as variáveis centradas na média.

4.11. EXERCÍCIOS

1. Na Tabela 4.24 são apresentados os dados experimentais de um ensaio de fertilização, no delineamento inteiramente casualizado (DIC) com 3 tratamentos e 6 repetições, onde X_1 é o teor de nitrogênio e X_2 é o teor de fósforo.

Tabela 4.24 – Resultados experimentais sobre teor de nitrogênio (X_1) e teor de fósforo (X_2)

1. Testemunha		2. Turfa fermentada		3. Turfa não fermentada	
X_1	X_2	X_1	X_2	X_1	X_2
4,63	0,95	6,03	1,08	4,71	0,96
4,38	0,89	5,96	1,05	4,81	0,93
4,94	1,01	6,16	1,08	4,49	0,87
4,96	1,23	6,33	1,19	4,43	0,82
4,48	0,94	6,08	1,08	4,56	0,91
4,40	0,98	6,20	1,05	4,50	0,88

Pede-se: (Usar $\alpha = 5\%$)

- Fazer a análise de variância multivariada (MANOVA) e testar a hipótese de igualdade dos vetores de médias de tratamentos pelo teste de Wilks;
- A título de ilustração, aplicar os testes de Pillai, Hotelling-Lawley e Roy no item (a);
- Comparar os vetores de médias dos tratamentos 1 e 2 pelo teste T^2 de Hotelling;
- Comparar as médias de tratamentos, para cada variável pelo teste de Roy e Bose;
- Exibir Matricialmente:

- O modelo linear: $Y = X B + \varepsilon$
 - O sistema de equações normais: $X' X B^0 = X' Y$
 - Uma solução para o sistema de equações: $B^0 = (X' X)^{-1} X' Y$, para qualquer $(X' X)^{-1}$;
 - $B^0' X' Y$ e $Y' Y$
 - $E = Y' Y - B^0' X' Y$
 - H para $C' B = \phi$
- $$H = [C' B^0]' [C' (X' X)^{-1} C]^{-1} [C' B^0]$$
- $A = Y' Y - \frac{1}{n} Y' \underline{1} \underline{1}' Y$, onde $\underline{1}$ é um vetor de 1's $n \times 1$.

Note que, neste caso, $E = A - H$

- Obter a Função Discriminante Linear de Fisher $(\underline{v}' X)$ tal que $\underline{v}' \underline{v} = 1$;

g) Faça uma ANOVA dos dados gerados por $\underline{y}'\underline{X}$ e compare as médias pelo teste de Roy e Bose.

2. Considerando os dados do exercício 1 (Tabela 4.24), mas admitindo que cada linha horizontal se refere a um bloco (Delineamento em Blocos Completos Casualizados - DBC), resolva as mesmas questões propostas em 1. Além disso, obtenha a matriz B para blocos, utilizando a seguinte expressão:

$$B = [C' B^0]' [C' (X' X)^{-1} C]^{-1} [C' B^0]$$

Note que, neste caso, $E = A - B - H$

3. Considerando os dados do exercício 1 e DIC (Tabela 4.24), pede-se:

- a) Os autovalores de $E^{-1}H$;
- b) As duas variáveis canônicas $VC_j = \underline{y}'_j \underline{X}$ com $j=1,2$, onde \underline{y}_j é o vetor canônico para

a j -ésima variável canônica, tal que $\underline{y}'_j [E/n_e] \underline{y}_j = 1$;

- c) As médias canônicas;
- d) A Tabela com os escores dos três tratamentos, obtidos a partir das duas variáveis canônicas;
- e) O teste de dimensionalidade;
- f) O gráfico mostrando a dispersão dos três tratamentos em relação às duas variáveis canônicas;
- g) A distância gráfica entre cada par de tratamentos baseada na distância euclidiana $(dVC_{ii'})$;
- h) A distância gráfica entre cada par de tratamentos baseada na distância de Mahalanobis $(D_{ii'}^2)$;
- i) O grau de distorção $= 1 - \theta$.

4. Considerando os dados do exercício 1 (Tabela 4.24) e supondo o delineamento em blocos completos casualizados, resolva todos os itens pedidos no exercício 3.

OBSERVAÇÃO: Em todos os exercícios, quando pertinente, faça os devidos comentários e interpretações dos resultados.

4.12. REFERÊNCIAS

- ANDERSON, T. W. **An introduction to multivariate statistical analysis**. 2nd ed., New York: John Wiley and Sons, 1984. 379p.
- BARTLETT, M. S. **Multivariate analysis**. **J. R. Stat. Soc.**, Série B, n.9, p.176-197, 1947.
- COOLEY, W.W. & LOHNES, P.R. **Multivariate data analysis**. New York: John Wiley and Sons, 1971. 364p.
- CRUZ, C. D.; REGAZZI, A. J.; CARNEIRO, P. C. S. **Modelos biométricos aplicados ao melhoramento genético**. v.1, 3. ed., Viçosa: UFV, 2012. 514p.
- GOMES, F. P. **Curso de estatística experimental**. 15. ed., Piracicaba: FEALQ, 2009. 451p.
- HARRIS, R. J. **A primer of multivariate statistics**. New York: Academic Press, 1975. 332p.
- JOHNSON, R. A.; WICHERN, D.W. **Applied multivariate statistical analysis**. 4th ed., USA, Upper Saddle River, New Jersey: Prentice Hall, Inc., 1998. 816p.
- KARSON, M. J. **Multivariate statistical methods**. Ames, Iowa: The Iowa State University Press, 1982. 307p.
- KHATTREE, R.; NAIK, D. N. **Applied multivariate statistical with SAS Software**. 2. ed. A copublication of Cary, NC, SAS Institute Inc. and New York: John Wiley and Sons, 1999. 338p.
- KHATTREE, R.; NAIK, D.N. **Multivariate data reduction and discrimination with SAS Software**. USA: SAS Institute Inc., Cary, NC, 2000. 558p.
- LACHENBRUCH, P. A. **Discriminant analysis**. New York: Hafner Press, 1975. 128p.
- MAHALANOBIS, P. C. On the generalized distance in statistic. **Proc. Nat. Inst. Sci.**, v.2, p.49-55, 1936.
- MANLY, B. F. J. **Multivariate statistical methods. A primer**. London: Chapman and Hall, 1986. 159p.
- MARDIA, K. V.; KENT, J. T.; BIBBY, J. M. **Multivariate analysis**. 6th ed., London: Academic Press, 1997. 518p.
- MORRISON, D. F. **Multivariate statistical methods**. 2nd ed., USA: McGraw-Hill Book Company, 1976. 415p.
- RAO, C. R. **Advanced statistical methods in biometric research**. New York: John Wiley and Sons, 1952. 390p.
- RAO, C. R. **Linear Statistical Inference and its Applications**. 2nd ed., New York: John Wiley and Sons, 1973. 624p.
- RENCHE, A. C. **Methods of multivariate analysis**. 2nd ed., USA: John Wiley and Sons, 2002. 738p.
- ROY, S. N.; Bose, R. C. Simultaneous confidence interval estimation. **The Annals of Mathematical Statistics**, n.24, p.513-536, 1953.

ROY, S. N. On a heuristic method of test construction and its use in multivariate analysis. **The Annals of Mathematical Statistics**, n.24, p.220-238, 1953.

ROY, S. N. **Some aspects of multivariate analysis**. New York: John Wiley and Sons, 1957. 486 p.

SAS INSTITUTE INC. **Statistical Analysis System**. Version 9.1, USA: Cary, NC, 2004.

SCHEFFÉ, H. A method for judging all contrast in the analysis of variance. **Biometrika**, n.40, p.87-104, 1953.

TATSUOKA, M. M. **Multivariate analysis: techniques for educational and psychological research**. New York: John Wiley and Sons, 1971. 310p.

CAPÍTULO 5

ANÁLISE DE COMPONENTES PRINCIPAIS

5.1. INTRODUÇÃO

A análise de componentes principais (ACP) é uma técnica de análise multivariada que consiste em transformar um conjunto original de variáveis em outro conjunto, os componentes principais, de dimensões equivalentes, porém com propriedades importantes, as quais serão descritas adiante.

Cada componente principal é uma combinação linear de todas as variáveis originais, são independentes entre si e estimados com o propósito de reter, em ordem de estimação, o máximo de informação, em termos de variação total, contida nos dados iniciais.

A análise de componentes principais é associada à ideia de redução da massa de dados. Procura-se redistribuir a variação observada nas variáveis (eixos ortogonais) de forma a obter um conjunto ortogonal de eixos não correlacionados. Esta técnica de análise de dados tem como principal objetivo a redução da dimensionalidade do conjunto original de variáveis, com a menor perda de informação possível, além de permitir o agrupamento de indivíduos (tratamentos, genótipos, etc.) similares, mediante exames visuais em dispersões gráficas no espaço bi ou tridimensional de fácil interpretação geométrica.

Embora, historicamente, as técnicas de análise multivariada, que constam da Literatura Estatística, tenham sido desenvolvidas para resolver problemas específicos, principalmente, de Biologia e Psicologia, elas podem ser utilizadas para resolver uma gama muito grande de problemas práticos nas diversas áreas do conhecimento. Na maior parte das vezes, os objetivos destes problemas práticos são atingidos com a utilização de mais de uma destas técnicas em uma sequência de análise e processamento de dados.

Desta forma, é muito importante ter uma visão conjunta de todas ou quase todas as técnicas de análise multivariada e, em destaque, a técnica de componentes principais.

5.2. COMPONENTES PRINCIPAIS

Considere a situação na qual observamos as variáveis X_1, X_2, \dots, X_p em cada uma de n unidades experimentais (indivíduos, tratamentos, genótipos, etc.). Este conjunto de $n \times p$ medidas dá origem a uma matriz X , de ordem $n \times p$ do tipo apresentada na Tabela 5.1.

Tabela 5.1 – Matriz de dados de n indivíduos e p variáveis

Indivíduos	Variáveis					
	X_1	X_2	...	X_j	...	X_p
1	X_{11}	X_{12}	...	X_{1j}	...	X_{1p}
2	X_{21}	X_{22}	...	X_{2j}	...	X_{2p}
...
i	X_{i1}	X_{i2}	...	X_{ij}	...	X_{ip}
...
n	X_{n1}	X_{n2}	...	X_{nj}	...	X_{np}

A estrutura de interdependência destas variáveis é representada pela matriz de covariância \mathbf{S} (matriz de variâncias e covariâncias) ou pela matriz de correlação \mathbf{R} . Intuitivamente percebemos que, quanto maior for o número de variáveis e quanto mais geral for a estrutura de interdependências entre elas (algumas têm variâncias grande, outras têm variâncias média, outras variâncias pequena e as correlações entre elas assumem valores muito diferentes entre si), mais difícil será comparar indivíduos baseando-se nos valores destas variáveis. Mas, se hipoteticamente tivermos uma situação em que, as variáveis são não correlacionadas, e apenas uma ou duas delas tem variância muito grande e as outras variâncias são bem pequenas, então para comparar dois indivíduos, podemos observar apenas estas duas variáveis desprezando todas as outras, pois não irão ajudar a distinguir os indivíduos.

Pois bem, esta situação é realmente hipotética e dificilmente iremos encontrá-la na prática. Então o nosso objetivo é transformar um conjunto de variáveis X_1, X_2, \dots, X_p , que apresenta uma estrutura de interdependência complicada, em um conjunto de variáveis Y_1, Y_2, \dots, Y_p , que sejam não correlacionadas e suas variâncias ordenadas para que seja possível comparar os indivíduos usando apenas aquelas que apresentem a maior variância.

Este é um problema matemático que pode ser formalmente apresentado da seguinte forma: Dado um conjunto de variáveis originais X_1, X_2, \dots, X_p , queremos encontrar outro conjunto de variáveis Y_1, Y_2, \dots, Y_p , tal que:

$$(a) \ Y_i = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{ip}X_p = \sum_{j=1}^p a_{ij}X_j$$

$$\text{com } i = 1, 2, \dots, p \quad \text{e} \quad \sum_{j=1}^p a_{ij}^2 = 1$$

$$(b) \text{Var}(Y_1) \geq \text{Var}(Y_2) \geq \dots \geq \text{Var}(Y_p)$$

$$(c) \text{Cov}(Y_i, Y_{i'}) = 0 \quad \text{para } i \neq i'$$

$$(d) \sum_{i=1}^p \text{Var}(Y_i) = \sum_{i=1}^p \text{Var}(X_i)$$

Este problema pode ser resolvido a partir da matriz de covariância \mathbf{S} (é a matriz de variâncias e covariâncias ou também chamada matriz de dispersão) ou a partir da matriz de correlação \mathbf{R} .

Quando as variáveis são medidas em escalas diferentes, é conveniente padronizar as variáveis X_j ($j = 1, 2, \dots, p$) e neste caso a estrutura de dependência de X_j é dada por \mathbf{R} .

▪ Padronização com média zero e variância 1:

$$x_{ij} = \frac{X_{ij} - \bar{X}_j}{s(X_j)}, \quad \text{com } i = 1, 2, \dots, n \text{ e } j = 1, 2, \dots, p$$

▪ Padronização com variância 1:

$$x_{ij} = \frac{X_{ij}}{s(X_j)}, \quad \text{com } i = 1, 2, \dots, n \text{ e } j = 1, 2, \dots, p$$

em que,

$$\bar{X}_j = \frac{\sum_{i=1}^n X_{ij}}{n}, \quad \text{com } j = 1, 2, \dots, p$$

$s(X_j)$ é o desvio padrão da variável X_j , dado por:

$$s(X_j) = \sqrt{\hat{\text{Var}}(X_j)}$$

$$\hat{\text{Var}}(X_j) = \frac{\text{SQD}(X_j)}{n-1} = \frac{\sum_{i=1}^n (X_{ij} - \bar{X}_j)^2}{n-1} = \frac{\sum_{i=1}^n X_{ij}^2 - \frac{(\sum_{i=1}^n X_{ij})^2}{n}}{n-1}, \quad \text{para } j = 1, 2, \dots, p,$$

em que SQD é a soma de quadrados dos desvios.

Temos ainda que a covariância entre as variáveis X_j e $X_{j'}$ é dada por:

$$\text{Côv}(X_j, X_{j'}) = \frac{\text{SPD}(X_j, X_{j'})}{n-1} = \frac{\sum_{i=1}^n (X_{ij} - \bar{X}_j)(X_{ij'} - \bar{X}_{j'})}{n-1} = \frac{\sum_{i=1}^n X_{ij}X_{ij'} - \frac{(\sum_{i=1}^n X_{ij})(\sum_{i=1}^n X_{ij'})}{n}}{n-1},$$

em que SPD é a soma de produtos dos desvios.

Considerando as variáveis originais X_1, X_2, \dots, X_p , a matriz \mathbf{R} (que é igual à matriz \mathbf{S} entre as variáveis padronizadas x_1, x_2, \dots, x_p) é dada por:

$$\mathbf{R} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{bmatrix},$$

em que

$$r_{jj'} = r(X_j, X_{j'}) = \text{Cov}(x_j, x_{j'}) = \frac{\text{Cov}(X_j, X_{j'})}{\sqrt{\hat{\text{Var}}(X_j) \cdot \hat{\text{Var}}(X_{j'})}.$$

Simplificando o termo n-1, a correlação entre as variáveis X_j e $X_{j'}$ pode ser estimada pela seguinte expressão:

$$r_{jj'} = \frac{\text{SPD}(X_j, X_{j'})}{\sqrt{\text{SQD}(X_j) \cdot \text{SQD}(X_{j'})}}.$$

Note que \mathbf{R} é uma matriz simétrica de dimensão $\mathbf{p} \times \mathbf{p}$.

Fazendo-se a padronização dos dados apresentados na Tabela 5.1, teremos os novos dados apresentados na Tabela 5.2. Certamente, a matriz de correlação das variáveis X_j é igual à matriz de covariância das variáveis padronizadas.

Tabela 5.2 – Matriz de dados padronizados de \mathbf{n} indivíduos e \mathbf{p} variáveis

Indivíduos	Variáveis					
	X_1	X_2	...	X_j	...	X_p
1	X_{11}	X_{12}	...	X_{1j}	...	X_{1p}
2	X_{21}	X_{22}	...	X_{2j}	...	X_{2p}
...
i	X_{i1}	X_{i2}	...	X_{ij}		X_{ip}
...
n	X_{n1}	X_{n2}	...	X_{nj}	...	X_{np}

5.2.1. Solução utilizando a matriz de correlação R

Para a realização da análise, geralmente feita com dados padronizados, considera-se que x_{ij} é o valor padronizado da j -ésima variável ($j=1, 2, \dots, p$) avaliado no i -ésimo indivíduo ($i=1, 2, \dots, n$) e R a matriz de covariância ou de correlação entre esses caracteres (a matriz de correlação entre os caracteres baseada nos dados originais é igual à matriz de covariância baseada nos dados padronizados). A técnica dos componentes principais consiste em transformar o conjunto de p variáveis ($x_{i1}, x_{i2}, \dots, x_{ip}$) em um novo conjunto ($Y_{i1}, Y_{i2}, \dots, Y_{ip}$), que são funções lineares dos x_i 's e independentes entre si.

As seguintes propriedades são verificadas:

a) Se Y_{i1} é um componente principal, então Y_{i1} é uma combinação linear das variáveis x_i 's, como descrito a seguir:

$$Y_{i1} = a_1 x_{i1} + a_2 x_{i2} + \dots + a_p x_{ip}$$

b) Se Y_{i2} é outro componente principal, então Y_{i2} é outra combinação linear das variáveis x_i 's, ou seja:

$$Y_{i2} = b_1 x_{i1} + b_2 x_{i2} + \dots + b_p x_{ip}$$

c) Entre todos os componentes, Y_{i1} apresenta a maior variância, Y_{i2} a segunda maior, e assim sucessivamente.

Também são consideradas as restrições:

$$\sum_{j=1}^p a_j^2 = \sum_{j=1}^p b_j^2 = 1$$

$$\sum_{j=1}^p a_j b_j = 0, \text{ ou seja, os componentes principais são não correlacionados.}$$

Com base na propriedade c, objetiva-se, por exemplo, em estudos sobre divergência genética por meio dos componentes principais, avaliar a possibilidade de estudar a dispersão dos indivíduos em sistemas de eixos cartesianos nos quais o aproveitamento da variabilidade disponível seja maximizada. O problema estatístico consiste fundamentalmente em estimar os coeficientes de ponderação dos caracteres em cada componente e a variância a eles associada.

Sendo Y_{i1} (ou simplesmente Y_1) o primeiro componente principal, sua variância é dada por:

$$V(Y_{i1}) = V(Y_1) = \sum_{j=1}^p a_{ij}^2 r_{jj} + \sum_{j \neq j'} \sum_{j'} a_j a_{j'} r_{jj'} = \sum_{j=1}^p \sum_{j'=1}^p a_j a_{j'} r_{jj'} ,$$

em que $r_{jj'}$ é o elemento da j -ésima linha e da j' -ésima coluna de R , lembrando que $r_{jj} = 1$.

Sob forma matricial, tem-se:

$$V(Y_1) = \tilde{a}' R \tilde{a}$$

em que \tilde{a} é um vetor $1 \times p$ de elementos a_j ($j = 1, 2, \dots, p$) .

Objetiva-se obter o vetor \tilde{a} de forma que a variância de Y_1 seja maximizada, impondo-se a restrição no conjunto de soluções de \tilde{a} por meio de $\tilde{a}'\tilde{a} = 1$. Expressando a variância de Y_1 pela função ω_1 e incorporando a restrição pelo multiplicador λ_1 de Lagrange, tem-se:

$$\omega_1 = \tilde{a}' R \tilde{a} + \lambda_1 (1 - \tilde{a}'\tilde{a}) .$$

Por diferenciação, encontra-se:

$$\partial \omega_1 = 2 \partial \tilde{a}' R \tilde{a} - 2 \lambda_1 \partial \tilde{a}' \tilde{a}, \text{ ou}$$

$$\frac{\partial \omega_1}{\partial \tilde{a}} = 2(R - \lambda_1 I) \tilde{a} .$$

Fazendo $\partial \omega_1 / \partial \tilde{a} = 0$, tem-se:

$$(R - \lambda_1 I) \tilde{a} = 0 . \quad (1)$$

A solução do sistema deve ser tal que $\tilde{a} \neq 0$. Assim, é necessário que o determinante da matriz $[R - \lambda_1 I]$ seja nulo ($|R - \lambda_1 I| = 0$) para que o sistema se torne indeterminado e a solução possa ser escolhida entre aquelas que satisfaçam a condição $\tilde{a}'\tilde{a} = 1$.

Sendo λ_1 os valores que satisfazem $|R - \lambda_1 I| = 0$, então, por definição, λ_1 são as raízes características (ou autovalores) de R e \tilde{a} o vetor característico (ou autovetor) associado.

Para o primeiro componente principal, o valor de λ_1 deve ser o maior dos p autovalores estimados, pois pré-multiplicando (1) por \tilde{a}' , verifica-se que:

$$\tilde{a}' R \tilde{a} - \tilde{a}' \tilde{a} \lambda_1 = 0 ,$$

logo:

$$\lambda_1 = \tilde{a}' R \tilde{a} = V(Y_1).$$

Como o vetor \tilde{a} foi escolhido para maximizar $V(Y_1)$, tem-se que λ_1 assume, nesta condição, o valor máximo entre os elementos do conjunto de autovalores de R .

A variância do segundo componente principal é dada por:

$$V(Y_{i2}) = V(Y_2) = \tilde{b}' R \tilde{b}.$$

Na obtenção do vetor \tilde{b} , cujos elementos são os coeficientes b_j ($j = 1, 2, \dots, p$), devem-se considerar as restrições $\tilde{b}'\tilde{b} = 1$ e $\tilde{b}'\tilde{a} = \tilde{a}'\tilde{b} = 0$, as quais são incorporadas na função de maximização por meio dos multiplicadores λ_2 e θ de Lagrange. Assim, é estabelecido que:

$$\omega_2 = \tilde{b}' R \tilde{b} + \lambda_2 (1 - \tilde{b}'\tilde{b}) + \theta \tilde{a}'\tilde{b}.$$

A restrição $\tilde{b}'\tilde{b} = 1$ é necessária para garantir a unicidade de \tilde{b} , ao passo que $\tilde{a}'\tilde{b} = 0$ garante que Y_1 e Y_2 sejam ortogonais.

A solução que maximiza ω_2 é obtida pela derivação de ω_2 em relação a \tilde{b} , dada por:

$$\frac{\partial \omega_2}{\partial \tilde{b}} = 2(R - \lambda_2 I)\tilde{b} + \theta \tilde{a}.$$

Fazendo $\partial \omega_2 / \partial \tilde{b} = 0$, tem-se:

$$2(R - \lambda_2 I)\tilde{b} + \theta \tilde{a} = 0. \quad (2)$$

Pré-multiplicando (2) por \tilde{a}' , obtém-se:

$$2\tilde{a}' R \tilde{b} + \theta = 0. \quad (3)$$

Pré-multiplicando (1) por \tilde{b}' , obtém-se:

$$\tilde{b}' R \tilde{a} = \tilde{a}' R \tilde{b} = 0. \quad (4)$$

Substituindo (4) em (3), conclui-se que $\theta = 0$. Assim, (2) pode ser simplificado para:

$$(R - \lambda_2 I)\underline{b} = \underline{0} \quad \text{e} \quad |R - \lambda_2 I| = 0,$$

em que se conclui que λ_2 é o segundo maior autovalor de R e \underline{b} o seu autovetor associado. Os demais componentes principais são estimados de maneira análoga à descrita para os dois primeiros.

5.2.2. Solução utilizando a matriz de covariância S

Os componentes principais também poderiam ser obtidos a partir da matriz de covariância, tal como descrito a seguir.

Considerando as variáveis originais X_1, X_2, \dots, X_p , a matriz S é dada por:

$$S = \begin{bmatrix} \hat{\text{Var}}(X_1) & \text{C\hat{ov}}(X_1, X_2) & \cdots & \text{C\hat{ov}}(X_1, X_p) \\ \text{C\hat{ov}}(X_1, X_2) & \hat{\text{Var}}(X_2) & \cdots & \text{C\hat{ov}}(X_2, X_p) \\ \cdots & \cdots & \cdots & \cdots \\ \text{C\hat{ov}}(X_1, X_p) & \text{C\hat{ov}}(X_2, X_p) & \cdots & \hat{\text{Var}}(X_p) \end{bmatrix}$$

Note que S é uma matriz simétrica de ordem $p \times p$, pois $\text{C\hat{ov}}(X_j, X_{j'}) = \text{C\hat{ov}}(X_{j'}, X_j)$.

Para determinar os **componentes principais** a partir da matriz S é dado o seguinte procedimento:

(a) A solução é obtida resolvendo-se a equação característica da matriz S .

$$\det(S - \lambda I) = 0 \quad \text{ou} \quad |S - \lambda I| = 0$$

Se a matriz S for positiva definida, a equação $|S - \lambda I| = 0$ terá p raízes positivas chamadas autovalores ou raízes características da matriz S .

Sejam $\lambda_1 > \lambda_2 > \cdots > \lambda_p$ as p soluções. A cada autovalor λ_i corresponde um autovetor ou vetor característico

$$\underline{a}_{\sim i} = \begin{bmatrix} a_{i1} \\ a_{i2} \\ \vdots \\ a_{ip} \end{bmatrix}$$

$$\text{com } \sum_{j=1}^p a_{ij}^2 = 1 \quad (\tilde{\mathbf{a}}_i' \tilde{\mathbf{a}}_i = 1)$$

$$\text{e } \sum_{j=1}^p a_{ij} a_{kj} = 0 \quad \text{para } i \neq k \quad (\tilde{\mathbf{a}}_i' \tilde{\mathbf{a}}_k = 0 \quad \text{para } i \neq k).$$

As condições anteriores significam que cada autovetor é normalizado, isto é, a soma dos quadrados dos coeficientes é igual a 1, e ainda, são ortogonais entre si.

(b) Para cada autovalor λ_i determina-se o autovetor normalizado $\tilde{\mathbf{a}}_i$, a partir da solução do sistema de equações dado a seguir:

$$(\mathbf{S} - \lambda_i \mathbf{I}) \tilde{\mathbf{a}}_i^* = \mathbf{0}$$

$$\tilde{\mathbf{a}}_i^* = \begin{bmatrix} a_{i1}^* \\ a_{i2}^* \\ \vdots \\ a_{ip}^* \end{bmatrix} \text{ é um autovetor não normalizado}$$

$\mathbf{0}$ é um vetor nulo de dimensão $\mathbf{p} \times \mathbf{1}$.

O autovetor normalizado é dado por:

$$\tilde{\mathbf{a}}_i = \begin{bmatrix} a_{i1} \\ a_{i2} \\ \vdots \\ a_{ip} \end{bmatrix} = \frac{\tilde{\mathbf{a}}_i^*}{\|\tilde{\mathbf{a}}_i^*\|} = \frac{1}{\sqrt{a_{i1}^{*2} + a_{i2}^{*2} + \cdots + a_{ip}^{*2}}} \begin{bmatrix} a_{i1}^* \\ a_{i2}^* \\ \vdots \\ a_{ip}^* \end{bmatrix}$$

$$\tilde{\mathbf{a}}_i' \tilde{\mathbf{a}}_i = 1$$

Tomando-se os elementos do vetor $\tilde{\mathbf{a}}_i$ assim determinados, como sendo os coeficientes de Y_i , segue que o **i-ésimo componente principal** é dado por:

$$Y_i = a_{i1} X_1 + a_{i2} X_2 + \cdots + a_{ip} X_p.$$

Temos ainda que:

$$(i) \hat{\text{Var}}(Y_i) = \lambda_i$$

Portanto $\hat{\text{Var}}(Y_1) > \hat{\text{Var}}(Y_2) > \dots > \hat{\text{Var}}(Y_p)$

$$(ii) \sum_{i=1}^p \hat{\text{Var}}(X_i) = \sum_{i=1}^p \lambda_i = \sum_{i=1}^p \hat{\text{Var}}(Y_i)$$

$$(iii) \text{C}\hat{\text{ov}}(Y_i, Y_j) = 0, \text{ desde que } \sum_{j=1}^p a_{ij} a_{kj} = 0.$$

Nesta metodologia a contribuição de cada componente principal Y_i é medida em termos de variância. Portanto o quociente (expresso em porcentagem)

$$\text{IR}(\%) = \frac{\hat{\text{Var}}(Y_i)}{\sum_{i=1}^p \hat{\text{Var}}(Y_i)} \cdot 100 = \frac{\lambda_i}{\sum_{i=1}^p \lambda_i} \cdot 100 = \frac{\lambda_i}{\text{traço}(S)} \cdot 100,$$

que denominaremos de **importância relativa**, representa a proporção da variância total explicada pelo componente principal Y_i . Assim, a importância relativa de um componente principal Y_i é então avaliada pela porcentagem da variância total que ele explica. A soma dos primeiros k autovalores dividida pela soma de todos os p autovalores, isto é, $(\lambda_1 + \dots + \lambda_k) / (\lambda_1 + \dots + \lambda_p)$ representa a proporção da variância total explicada pelos primeiros k componentes principais, ou seja, a proporção da informação retida na redução de p para k dimensões. Se a obtenção dos componentes principais foi feita a partir da matriz de correlação R , virá:

$$\text{IR}(\%) = \frac{\hat{\text{Var}}(Y_i)}{\sum_{i=1}^p \hat{\text{Var}}(Y_i)} \cdot 100 = \frac{\lambda_i}{\sum_{i=1}^p \lambda_i} \cdot 100 = \frac{\lambda_i}{\text{traço}(R)} \cdot 100 = \frac{\lambda_i}{p} \cdot 100.$$

Quantos componentes principais usar?

i) Com a medida que acabamos de introduzir, podemos tomar a decisão sobre quantos componentes principais iremos utilizar para avaliar os dados, isto é, quantos componentes serão utilizados para diferenciar os indivíduos. Esta não é uma decisão Estatística, pois não temos um Modelo Estatístico com algum termo representando o erro aleatório. Fazendo uma revisão sobre a utilização de componentes principais nas diversas áreas de aplicação, verificamos que para interpretar os dados com sucesso, basta escolher os primeiros componentes que acumulem uma porcentagem de variância explicada igual ou superior a 80% admitindo que a distorção das coordenadas de cada indivíduo, no gráfico de dispersão cujos eixos são os componentes principais, será considerada aceitável e as inferências no estudo da dissimilaridade, satisfatórias. Assim, ficamos com Y_1, \dots, Y_k tal que:

$$\frac{\hat{\text{Var}}(Y_1) + \dots + \hat{\text{Var}}(Y_k)}{\sum_{i=1}^p \hat{\text{Var}}(Y_i)} \cdot 100 \geq 80\%, \quad \text{onde } k < p.$$

O sucesso da metodologia é medido pelo valor de \mathbf{k} . Se $\mathbf{k} = 1$, diremos que o método está reduzindo ao máximo a dimensão inicial. Neste caso, podemos então comparar os indivíduos em uma escala linear. Se $\mathbf{k} = 2$ podemos localizar cada indivíduo em um plano cartesiano onde os dois eixos representam os dois componentes. Neste caso, ainda fica muito simples a comparação dos indivíduos. Se \mathbf{k} for igual a 3 recomenda-se a dispersão gráfica em relação aos três eixos representativos dos três primeiros componentes principais. Se k for maior do que 3 a comparação dos indivíduos passa a ser mais complicada, recomendando-se a representação em gráficos de dispersão envolvendo a combinação do primeiro componente com os demais.

Como se sabe, o número de autovalores não nulos proporciona a dimensão do espaço em que se encontram as observações; um autovalor nulo revela a existência de uma dependência linear entre as variáveis originais. Um autovetor correspondente ao autovalor nulo revela a forma da dependência linear. Uma solução para o dilema quando um autovalor é igual a zero é remover uma ou mais variáveis.

Se, por exemplo, os autovalores estimados de uma matriz de variâncias e covariâncias forem: 10, 8, 6, 5, 1, 0; saber-se-á imediatamente que a dimensionalidade do espaço multivariado não é 6. Mas seria tetravariado ou pentavariado?

Observe os percentuais da variância observada atribuídos a cada componente principal:

$$\% \text{ componente principal 1} = \frac{\lambda_1}{\sum_{j=1}^p \lambda_j} \cdot 100 = \frac{10}{30} \cdot 100 = 33,33 \%$$

$$\% \text{ componente principal 2} = \frac{\lambda_2}{30} \cdot 100 = \frac{8}{30} \cdot 100 = 26,67 \%$$

$$\% \text{ componente principal 3} = \frac{\lambda_3}{30} \cdot 100 = \frac{6}{30} \cdot 100 = 20,00 \%$$

$$\% \text{ componente principal 4} = \frac{\lambda_4}{30} \cdot 100 = \frac{5}{30} \cdot 100 = 16,67 \%$$

$$\% \text{ componente principal 5} = \frac{\lambda_5}{30} \cdot 100 = \frac{1}{30} \cdot 100 = 3,33 \%$$

Este último componente provavelmente não representa uma característica (dimensionalidade) do espaço multivariado, pois provavelmente é resultado de erros nas estimativas. Neste caso se

considerarmos os três primeiros componentes principais, explicaremos 80 % da variabilidade total dos dados.

ii) Critério de Kaiser (1958)

Pelo critério de Kaiser (1958) consideramos os componentes principais com variância superior à média das variâncias das variáveis originais. No exemplo anterior, temos:

$$\text{Média das variâncias das variáveis } X_j = \frac{30}{6} = 5.$$

Neste caso optaríamos pelos três primeiros componentes principais. Esse procedimento, quando os componentes principais são extraídos da matriz de correlação, equivale a considerar os componentes principais associados a autovalores superiores a 1, pois nesse caso,

$$\sum_{j=1}^p \lambda_j = \text{traço}(\mathbf{R}) = p = \text{número de variáveis}.$$

A regra apresentada anteriormente está associada à idéia de redução da massa de dados, e pode ser adequada em muitas situações. Entretanto, não convém perder de vista a estrutura do espaço de variáveis sob análise.

Dois testes estatísticos podem ser feitos com relação ao número de componentes principais: o de esfericidade da distribuição conjunta e o da irrelevância dos componentes remanescentes. Entretanto, estes testes não serão apresentados.

iii) Um outro critério é manter na análise os componentes principais que acumulem pelo menos uma certa porcentagem de variabilidade de cada uma das variáveis originais, digamos 50%. A porcentagem de variância de cada uma das variáveis originais (X_j) explicada por cada um dos componentes principais (Y_i) é igual ao quadrado da correlação entre X_j e Y_i . Assim, para o componente $Y_i = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{ip}X_p$, por exemplo, temos que,

$$\text{Corr}(X_j, Y_i) = r_{X_j Y_i} = a_{ij} \frac{\sqrt{\hat{\text{Var}}(Y_i)}}{\sqrt{\hat{\text{Var}}(X_j)}} = \sqrt{\lambda_i} \cdot \frac{a_{ij}}{\sqrt{\hat{\text{Var}}(X_j)}}.$$

Vamos ilustrar a aplicação deste critério a partir de alguns resultados de uma análise por componentes principais baseada numa matriz de correlação envolvendo quatro variáveis (Tabela 5.3). Pode-se verificar que, com dois componentes principais, a menor porcentagem de variância individual explicada é a da variável X_2 , com 73,22%. Note que se a opção for por somente um

componente, a variância das variáveis X_2 e X_3 teriam somente 6,72% e 11,84% de explicação, respectivamente.

Tabela 5.3 - % de explicação das variâncias individuais (acumuladas entre parêntesis) e autovalores obtidos de uma matriz de correlação R ($p = 4$ variáveis)

Variável	Componentes Principais			
	Y_1	Y_2	Y_3	Y_4
X_1	78,40	9,65	1,93	10,02
	(78,40)	(88,05)	(89,98)	(100,00)
X_2	6,72	66,50	26,47	0,31
	(6,72)	(73,22)	(99,69)	(100,00)
X_3	11,84	61,92	25,45	0,79
	(11,84)	(73,76)	(99,21)	(100)
X_4	83,52	5,13	0,82	10,53
	(83,52)	(88,65)	(89,47)	(100,00)
Autovalores	1,804734	1,432056	0,546756	0,216454

iv) “Scree Plot” dos autovalores

O “Scree Plot” dos autovalores é um gráfico em que representam-se os autovalores, o qual pode auxiliar na escolha do número de componentes principais a serem considerados. Comumente, a diferença entre os primeiros autovalores é grande e diminui para os últimos. A sugestão é fazer o corte quando a variação passa a ser pequena. Os autovalores da Tabela 5.3 estão representados na Figura 5.1.

Pode-se observar na Figura 5.1, que uma maior mudança ocorre em torno do terceiro autovalor. Assim, os autovalores após o segundo maior, são relativamente pequenos. Neste caso, se considerarmos os dois primeiros componentes principais, explicaremos 80,92% da variabilidade total dos dados.

Para que a análise dos dados por meio de componentes principais seja completa é necessário saber o significado de cada componente. Portanto, a seguir, mostraremos como se interpreta os componentes.

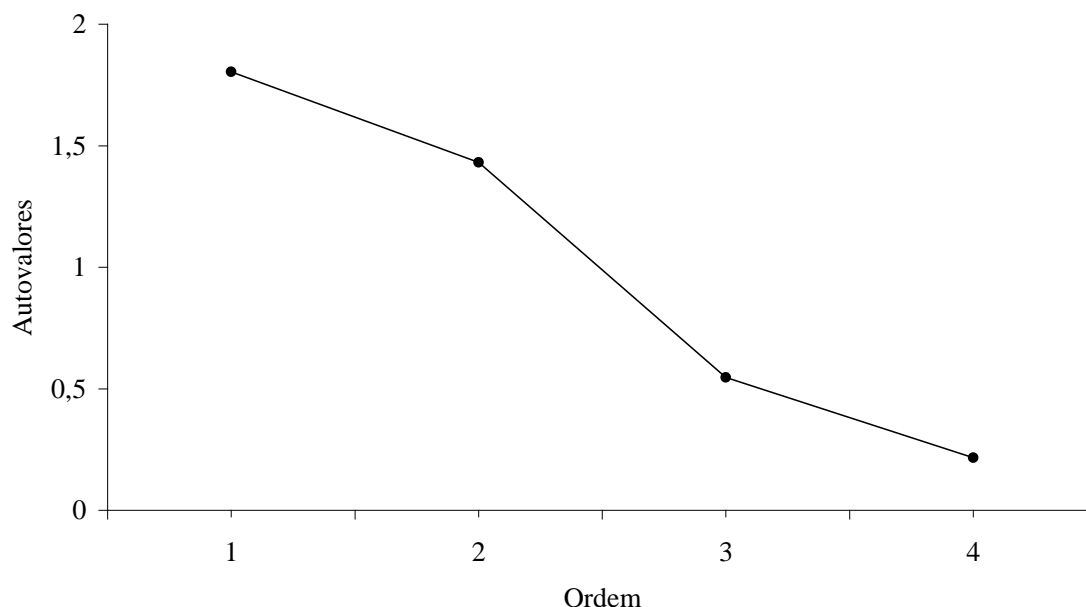


Figura 5.1 – Scree plot dos autovalores

A interpretação de cada componente principal é feita verificando o grau de importância ou influência que cada variável X_j tem sobre o componente. Esta importância é dada pelos coeficientes e/ou pela correlação entre cada X_j e o componente $Y_i = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{ip}X_p$ que está sendo interpretado. Assim, para o componente Y_1 , por exemplo, temos que,

$$\text{Corr}(X_j, Y_1) = r_{X_j Y_1} = a_{1j} \frac{\sqrt{\hat{\text{Var}}(Y_1)}}{\sqrt{\hat{\text{Var}}(X_j)}} = \sqrt{\lambda_1} \cdot \frac{a_{1j}}{\sqrt{\hat{\text{Var}}(X_j)}}.$$

Portanto, para comparar a importância de X_1, X_2, \dots, X_p sobre Y_1 , basta examinar os coeficientes a_{1j} ($j = 1, 2, \dots, p$) e/ou as correlações $r_{X_j Y_1}$ ($j = 1, 2, \dots, p$). O mesmo procedimento é feito para os demais componentes.

As correlações são medidas das contribuições individuais de cada variável e não consideram a contribuição multivariada das demais variáveis; os coeficientes são medidas das contribuições multivariadas. Johnson e Wichern (1998) sugerem que as interpretações sejam feitas com base nas duas medidas, proporcionando as visões univariada e multivariada. Esses autores indicam que, na prática, não é comum encontrar-se diferenças importantes. Rencher (2002) recomenda que somente os coeficientes e não as correlações devem ser usadas para interpretar os

componentes. As duas medidas de importância, a primeira multivariada (coeficientes) e a segunda univariada (correlações), frequentemente dão resultados similares.

Se a análise de componentes principais for realizada com as variáveis padronizadas, a interpretação feita com base nos coeficientes ou nas correlações é exatamente a mesma, pois $\hat{\text{Var}}(x_j) = 1$.

As informações de uma análise de componentes principais podem ser condensadas na Tabela 5.4. Esta Tabela nos dá informações que nos permite decidir pelo número de componentes Y_i que iremos utilizar na análise e interpretá-los.

Quando o objetivo da análise é construir índices, prática muito comum em Economia e Biologia, a análise termina aí. Neste caso devemos lembrar que a unidade de medida dos componentes principais é uma combinação linear das unidades de medida de cada variável observada X_j e, com isto, os componentes principais podem se acompanhados de unidade de medida sem sentido. Para estas situações é aconselhado fazer uma padronização nas variáveis X_j .

Quando o objetivo da análise é comparar indivíduos (tratamentos, genótipos, etc.) ou agrupá-los, então devemos calcular para cada indivíduo os seus valores (escores) para cada componente principal que será utilizado na análise. Isto equivale a substituir a matriz de dados originais, cuja dimensão é $\mathbf{n} \times \mathbf{p}$, por uma matriz $\mathbf{n} \times \mathbf{k}$, onde \mathbf{k} é o número de componentes principais escolhidos. A Tabela 5.5 ilustra este fato.

Tabela 5.4 – Componentes principais obtidos da análise de **p** variáveis (X_1, X_2, \dots, X_p)

Componente principal	Variância (Autovalor λ_i)	Coeficientes de ponderação associados às variáveis				Correlação entre X_j e Y_i				Porcentagem da variância de Y_i	Porcentagem acumulada da variância dos Y_i
		X_1	X_2	\dots	X_p	X_1	X_2	\dots	X_p		
Y_1	λ_1	a_{11}	a_{12}	\dots	a_{1p}	$\sqrt{\lambda_1} \cdot \frac{a_{11}}{s_1}$	$\sqrt{\lambda_1} \cdot \frac{a_{12}}{s_2}$	\dots	$\sqrt{\lambda_1} \cdot \frac{a_{1p}}{s_p}$	$\left(\lambda_1 / \sum_{i=1}^p \lambda_i \right) \cdot 100$	$\left(\lambda_1 / \sum_{i=1}^p \lambda_i \right) \cdot 100$
Y_2	λ_2	a_{21}	a_{22}	\dots	a_{2p}	$\sqrt{\lambda_2} \cdot \frac{a_{21}}{s_1}$	$\sqrt{\lambda_2} \cdot \frac{a_{22}}{s_2}$	\dots	$\sqrt{\lambda_2} \cdot \frac{a_{2p}}{s_p}$	$\left(\lambda_2 / \sum_{i=1}^p \lambda_i \right) \cdot 100$	$\left[(\lambda_1 + \lambda_2) / \sum_{i=1}^p \lambda_i \right] \cdot 100$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
Y_p	λ_p	a_{p1}	a_{p2}	\dots	a_{pp}	$\sqrt{\lambda_p} \cdot \frac{a_{p1}}{s_1}$	$\sqrt{\lambda_p} \cdot \frac{a_{p2}}{s_2}$	\dots	$\sqrt{\lambda_p} \cdot \frac{a_{pp}}{s_p}$	$\left(\lambda_p / \sum_{i=1}^p \lambda_i \right) \cdot 100$	$\left[(\lambda_1 + \lambda_2 + \dots + \lambda_p) / \sum_{i=1}^p \lambda_i \right] \cdot 100$

FATOS: $s_j = \sqrt{\hat{\text{Var}}(X_j)}$, $j=1, 2, \dots, p$

$$\sum_{i=1}^p \lambda_i = \sum_{i=1}^p \hat{\text{Var}}(Y_i) = \sum_{i=1}^p \hat{\text{Var}}(X_i) = \text{traço da matriz } \mathbf{S}$$

Tabela 5.5 – Escores relativos a **n** tratamentos, obtidos em relação aos **k** primeiros componentes principais

Tratamentos (Indivíduos)	Variáveis				Escores para os componentes			
	X_1	X_2	...	X_p	Y_1	Y_2	...	Y_k
1	X_{11}	X_{12}	...	X_{1p}	Y_{11}	Y_{12}	...	Y_{1k}
2	X_{21}	X_{22}	...	X_{2p}	Y_{21}	Y_{22}	...	Y_{2k}
\vdots	\vdots	\vdots		\vdots	\vdots	\vdots		\vdots
n	X_{n1}	X_{n2}	...	X_{np}	Y_{n1}	Y_{n2}	...	Y_{nk}

Assim, teremos que:

$$Y_{11} = a_{11}X_{11} + a_{12}X_{12} + \cdots + a_{1p}X_{1p}$$

$$Y_{21} = a_{11}X_{21} + a_{12}X_{22} + \cdots + a_{1p}X_{2p}$$

$$\vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots$$

$$Y_{n1} = a_{11}X_{n1} + a_{12}X_{n2} + \cdots + a_{1p}X_{np}$$

Os escores com as variáveis centradas na média mudam, mas as conclusões não se alteram, e são dados por:

$$Y_{i1} = a_{11}(X_{i1} - \bar{X}_1) + a_{12}(X_{i2} - \bar{X}_2) + \cdots + a_{1p}(X_{ip} - \bar{X}_p).$$

Para os escores dos demais componentes os cálculos são feitos de modo análogo. Pelo programa SAS, para a análise baseada na matriz **S**, os escores são dados com as variáveis originais centradas na média, enquanto que, para a análise baseada na matriz **R**, os escores são dados com as variáveis padronizadas com média 0 e variância 1.

Através da dispersão gráfica dos escores dos indivíduos em relação aos dois primeiros componentes principais, tem-se que, quanto maior a proximidade entre dois indivíduos, maior é a similaridade entre eles. A composição dos grupos de indivíduos pode ser feita de forma subjetiva (visualmente). Na realidade, a análise pode ser complementada utilizando-se uma outra técnica de análise multivariada como, por exemplo, a Análise de Agrupamento (“Cluster Analysis”).

Cabem aqui duas observações importantes. A primeira é sobre o fato de que utilizando a matriz de correlação **R**, os componentes serão ausentes de unidade de medida. A segunda é que se o pesquisador fizer duas análises de componentes principais, com os mesmos dados, uma utilizando a matriz de covariância **S** (dados originais) e outra utilizando a matriz **R**, ele certamente poderá obter resultados diferentes. Esta técnica não é invariante à mudança de escala.

5.2.3. Algumas considerações

(1ª) A análise deve ser baseada na matriz **S** ou na matriz **R**?

Esta questão pode ser desdobrada nas duas seguintes:

- Até que ponto padronizar variáveis é abrir mão de características relevantes dessas variáveis?
- Até que ponto não padronizar é introduzir distorção devido a influências das escalas e unidades de medida?

Obviamente, não existem respostas universais a estas questões. A decisão entre padronizar e não padronizar tem de se basear em conhecimento do fenômeno sob análise, em definições claras das variáveis e, também, em um pouco de intuição.

(2ª) Como um dos objetivos da Análise de Componentes Principais é a redução da dimensionalidade do espaço original, quando os últimos componentes principais explicam proporções reduzidas da variância, e tomamos apenas os dois primeiros componentes ($> 80\%$), tem-se a vantagem que permite sintetizar em um único plano “quase toda” a informação original.

Apesar de termos sugerido escolher os primeiros componentes que acumulem uma porcentagem de variância explicada igual ou superior a 80% , é difícil estabelecer uma porcentagem de variância explicada a partir da qual, e de forma geral, pode-se decidir qual o número de componentes que vamos considerar. Isto depende do número de variáveis originais, da magnitude das correlações de **R**, do estudo da significância estatística dos componentes e, sobretudo, da experiência e conhecimento que o pesquisador possui sobre as variáveis consideradas.

(3ª) Diversas proposições sobre o **descarte de variáveis** têm sido apresentadas na literatura. Na análise de componentes principais, tem-se adotado o critério de avaliação da importância das variáveis a partir das correlações entre cada variável e os últimos componentes que, por estimação, retêm proporção mínima da variação total. Quando se trabalha com as variáveis padronizadas, basta examinar os coeficientes de ponderação (autovetores normalizados) das variáveis associados aos últimos componentes. Baseado no princípio de que a importância relativa dos componentes principais decresce do primeiro para o último, os últimos componentes são responsáveis pela explicação de uma fração mínima da variância total disponível. Assim, a variável que domina (aquela que possui maior correlação em valor absoluto) o componente de menor autovalor é considerada de menor importância, sendo possível o seu descarte em estudos futuros. A razão é que variáveis altamente correlacionadas com os componentes principais de menor variância, representam variação praticamente insignificante. Para **S** ou **R** positiva semidefinida, utilizar este critério a partir do componente principal associado ao menor autovalor não nulo. Normalmente consideram-se os últimos autovetores associados a autovalores da matriz de correlação inferior a $0,7$, conforme recomendações de Jolliffe (1972, 1973) e Mardia et al. (1997). Quando em um componente de menor variância o maior coeficiente de ponderação (em valor absoluto) está associado a uma

variável já previamente descartada, tem-se optado por não fazer nenhum outro descarte com base nos coeficientes daquele componente, mas prosseguir a identificação da importância das variáveis no outro componente de variância imediatamente superior (CRUZ et al., 2012). Por outro lado, as variáveis de maiores pesos (em valor absoluto) nos primeiros autovetores são consideradas de maior importância para o componente quando os primeiros componentes principais explicam uma fração considerável da variação disponível, normalmente limitada em valor mínimo de 80% (CRUZ et al., 2014).

5.2.4. Exemplos de aplicação

5.2.4.1. Exemplo 1 - Componentes principais a partir de uma matriz de covariância 2x2

Considere os dados da Tabela 5.6, referentes a duas variáveis X_1 e X_2 mensuradas em uma amostra constituída de cinco indivíduos (Tratamentos).

Tabela 5.6 – Observações relativas a duas variáveis (X_1 e X_2) avaliadas em cinco tratamentos

Tratamentos	Variáveis	
	X_1	X_2
1	102	96
2	104	87
3	101	62
4	93	68
5	100	77

Vamos obter os componentes principais relativos às variáveis X_1 e X_2 , cuja matriz de covariância estimada é:

$$S = \begin{bmatrix} 17,50 & 31,50 \\ 31,50 & 190,50 \end{bmatrix}$$

Solução:

A equação característica é:

$$|S - \lambda I| = 0$$

$$\begin{vmatrix} 17,50 - \lambda & 31,50 \\ 31,50 & 190,50 - \lambda \end{vmatrix} = 0 ,$$

isto é:

$$(17,50 - \lambda)(190,50 - \lambda) - (31,50)^2 = 0$$

$$\lambda^2 - 208,00 \lambda + 2341,50 = 0$$

Os autovalores (raízes próprias ou raízes características) obtidos dessa equação são:

$$\lambda = \frac{208,00 \pm \sqrt{(-208,00)^2 - (4)(1)(2341,50)}}{(2)(1)} = \frac{208,00 \pm 184,1141}{2}$$

$$\lambda_1 = 196,0571$$

$$\lambda_2 = 11,9430$$

A soma dessas duas raízes dá 208,00, que é exatamente o coeficiente do segundo termo da equação, com sinal trocado.

Pode-se observar também que, $196,0571 + 11,9430 = 208 = \text{traço}(\mathbf{S})$. Sobre esse total os autovalores correspondem, em porcentagem, a:

$$\frac{196,0571}{208} \cdot 100 = 94,26 \% \quad \text{para } \lambda_1 \text{ e}$$

$$\frac{11,9430}{208} \cdot 100 = 5,74 \% \quad \text{para } \lambda_2.$$

Então o primeiro componente principal, relativo a λ_1 , explica 94,26% da variação, em comparação com apenas 5,74% para o segundo componente.

Vamos calcular o primeiro componente principal, correspondente ao maior autovalor $\lambda_1 = 196,0571$. Ele será dado pelo autovetor normalizado associado a λ_1 .

$$[\mathbf{S} - \lambda_1 \mathbf{I}] \mathbf{a}_1^* = \mathbf{0}$$

$$\begin{bmatrix} 17,50 - 196,0571 & 31,50 \\ 31,50 & 190,50 - 196,0571 \end{bmatrix} \begin{bmatrix} a_{11}^* \\ a_{12}^* \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} -178,5571 & 31,50 \\ 31,50 & -5,5571 \end{bmatrix} \begin{bmatrix} a_{11}^* \\ a_{12}^* \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{cases} -178,5571 a_{11}^* + 31,50 a_{12}^* = 0 \\ 31,50 a_{11}^* - 5,5571 a_{12}^* = 0 \end{cases}$$

Este sistema de equações é indeterminado, uma vez que temos, em virtude de $|\mathbf{S} - \lambda \mathbf{I}| = 0$:

$$\begin{vmatrix} -178,5571 & 31,50 \\ 31,50 & -5,5571 \end{vmatrix} = 0$$

Podemos, pois, abandonar uma das equações (por exemplo, a segunda) e dar um valor arbitrário, não nulo, a uma das incógnitas (por exemplo, $a_{12}^* = 1$). Assim virá:

$$-178,5571a_{11}^* + 31,50(1) = 0$$

$$178,5571a_{11}^* = 31,50 \Rightarrow a_{11}^* = 0,1764$$

$$\tilde{a}_1^* = \begin{bmatrix} 0,1764 \\ 1 \end{bmatrix} \quad \text{e} \quad \left\| \tilde{a}_1^* \right\| = \sqrt{(0,1764)^2 + (1)^2} = 1,0154$$

O autovetor normalizado é:

$$\tilde{a}_1 = \begin{bmatrix} a_{11} \\ a_{21} \end{bmatrix} = \frac{1}{\left\| \tilde{a}_1^* \right\|} \tilde{a}_1^* = \frac{1}{1,0154} \begin{bmatrix} 0,1764 \\ 1 \end{bmatrix}$$

$$\tilde{a}_1 = \begin{bmatrix} 0,1737 \\ 0,9848 \end{bmatrix}. \quad \text{Note que } \tilde{a}_1' \tilde{a}_1 = 1.$$

Logo, o primeiro componente principal é $Y_1 = 0,1737 X_1 + 0,9848 X_2$, responsável por 94,26% da variação.

O segundo componente principal é dado pelo sistema de equações relativo ao outro autovalor $\lambda_2 = 11,9430$:

$$[S - \lambda_2 I] \tilde{a}_2^* = 0$$

$$\begin{bmatrix} 17,50 - 11,9430 & 31,50 \\ 31,50 & 190,50 - 11,9430 \end{bmatrix} \begin{bmatrix} a_{21}^* \\ a_{22}^* \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 5,5570 & 31,50 \\ 31,50 & 178,5570 \end{bmatrix} \begin{bmatrix} a_{21}^* \\ a_{22}^* \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{cases} 5,5570a_{21}^* + 31,50a_{22}^* = 0 \\ 31,50a_{21}^* + 178,5570a_{22}^* = 0 \end{cases}$$

Dando o mesmo procedimento adotado anteriormente, teremos:

$$\tilde{a}_2^* = \begin{bmatrix} a_{21}^* \\ a_{22}^* \end{bmatrix} = \begin{bmatrix} -5,6685 \\ 1 \end{bmatrix} \quad \text{e} \quad \left\| \tilde{a}_2^* \right\| = \sqrt{(-5,6685)^2 + (1)^2} = 5,7560$$

O autovetor normalizado é:

$$\tilde{a}_2 = \begin{bmatrix} a_{21} \\ a_{22} \end{bmatrix} = \frac{1}{\left\| \tilde{a}_2^* \right\|} \tilde{a}_2^* = \frac{1}{5,7560} \begin{bmatrix} -5,6685 \\ 1 \end{bmatrix}$$

$$\tilde{a}_2 = \begin{bmatrix} -0,9848 \\ 0,1737 \end{bmatrix}. \quad \text{Note que } \tilde{a}_2' \tilde{a}_2 = 1.$$

O segundo componente principal, responsável por 5,74% da variação e ortogonal ao primeiro, é, portanto:

$$Y_2 = -0,9848 X_1 + 0,1737 X_2$$

FATOS: (i) Os componentes são ortogonais:

$$\tilde{a}_1' \tilde{a}_2 = [0,1737 \quad 0,9848] \begin{bmatrix} -0,9848 \\ 0,1737 \end{bmatrix} = 0$$

(ii) Cada componente admite na realidade **duas soluções**, cada uma delas obtida da outra pela multiplicação de seu segundo membro por (-1). Por exemplo, Y_2 também poderia ter o seguinte valor: $Y_2 = 0,9848 X_1 - 0,1737 X_2$.

Os escores relativos aos cinco tratamentos, obtidos em relação aos dois componentes principais estão apresentados na Tabela 5.7.

Tabela 5.7 – Escores relativos aos cinco tratamentos obtidos em relação aos dois componentes principais

Tratamentos	Variáveis		Escores para os componentes	
	X_1	X_2	Y_1	Y_2
1	102	96	112,26	83,77
2	104	87	103,74	87,31
3	101	62	78,60	88,69
4	93	68	83,12	79,77
5	100	77	93,20	85,10

Obtenção dos escores: $Y_1 = 0,1737 X_1 + 0,9848 X_2$

$$Y_{11} = 0,1737(102) + 0,9848(96) = 112,26$$

$$Y_{12} = 0,1737(104) + 0,9848(87) = 103,74, \text{ etc.}$$

$$Y_2 = 0,9848 X_1 - 0,1737 X_2$$

$$Y_{21} = 0,9848(102) - 0,1737(96) = 83,77$$

$$Y_{22} = 0,9848(104) - 0,1737(87) = 87,31, \text{ etc.}$$

Pode-se também fazer o estudo obtendo-se os escores com as variáveis centradas na média, por exemplo, $Y_1 = 0,1737(X_1 - \bar{X}_1) + 0,9848(X_2 - \bar{X}_2)$. A interpretação será exatamente a mesma.

Com os escores dos dois componentes principais podemos localizar cada indivíduo (tratamento) em um plano cartesiano onde os dois eixos representam os dois componentes. Isto pode ser visto na Figura 5.2.

As escalas dos eixos Y_1 e Y_2 devem ser a mesma, porque, dependendo do caso, numa interpretação visual, escalas diferentes poderão proporcionar impressionantes equívocos.

No Capítulo 6, mostraremos como completar esta análise através da Técnica de Análise de Agrupamento.

As informações importantes de uma análise de componentes principais condensadas na Tabela 5.4, com os dados do exemplo 5.1, estão apresentadas numericamente na Tabela 5.8.

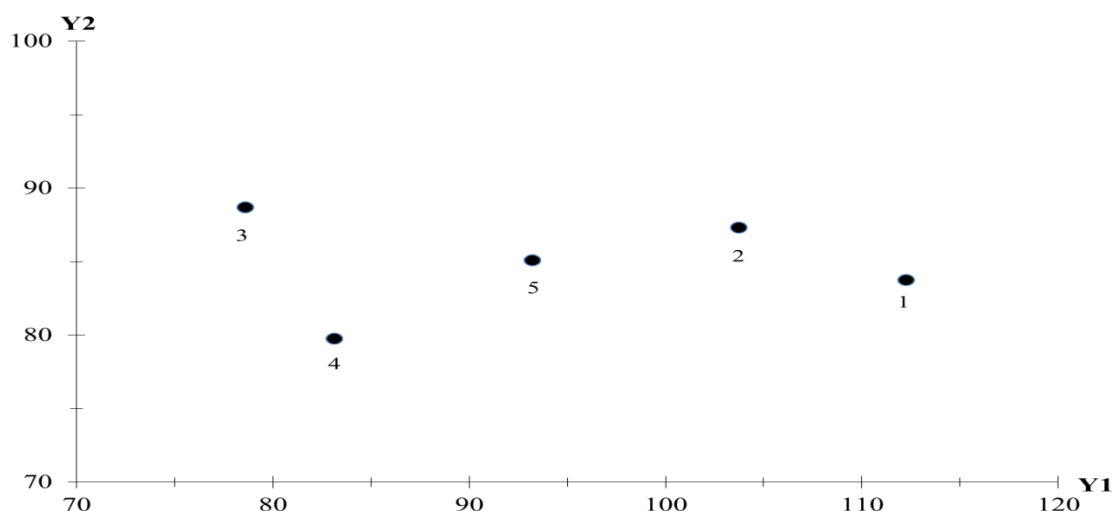


Figura 5.2 – Dispersão de cinco tratamentos em relação aos componentes principais Y_1 e Y_2

Tabela 5.8 – Componentes principais obtidos da análise de duas variáveis X_1 e X_2

Componente principal	Variância (Autovalor)	Coeficientes associados às variáveis		Correlações entre X_j e Y_i		Variância de Y_i (%)	% Acumulada da variância de Y_i
		X_1	X_2	X_1	X_2		
Y_1	196,0571	0,1737	0,9848	0,581	0,999	94,26	94,26
Y_2	11,9430	0,9848	-0,1737	0,813	-0,043	5,74	100,00

Ilustração do cálculo das correlações entre X_j e Y_i apresentadas na Tabela 5.8.

$$r_{X_1 Y_1} = \sqrt{\lambda_1} \cdot \frac{a_{11}}{\sqrt{\hat{\text{Var}}(X_1)}} = \sqrt{196,0571} \cdot \frac{0,1737}{\sqrt{17,50}} = 0,581$$

ou

$$r_{X_1 Y_1} = \frac{\text{Cov}(X_1, Y_1)}{\sqrt{\hat{\text{Var}}(X_1) \cdot \hat{\text{Var}}(Y_1)}} = \frac{34,06}{\sqrt{(17,50)(196,0571)}} = 0,581.$$

$$r_{X_2 Y_1} = \sqrt{\lambda_1} \cdot \frac{a_{12}}{\sqrt{\hat{V}\text{ar}(X_2)}} = \sqrt{196,0571} \cdot \frac{0,9848}{\sqrt{190,50}} = 0,999$$

$$\text{ou } r_{X_2 Y_1} = \frac{\text{C}\hat{\text{ov}}(X_2, Y_1)}{\sqrt{\hat{V}\text{ar}(X_2) \cdot \hat{V}\text{ar}(Y_1)}} = \frac{193,08}{\sqrt{(190,50)(196,0571)}} = 0,999.$$

$$r_{X_1 Y_2} = \sqrt{\lambda_2} \cdot \frac{a_{21}}{\sqrt{\hat{V}\text{ar}(X_1)}} = \sqrt{11,9430} \cdot \frac{0,9848}{\sqrt{17,50}} = 0,813$$

$$\text{ou } r_{X_1 Y_2} = \frac{\text{C}\hat{\text{ov}}(X_1, Y_2)}{\sqrt{\hat{V}\text{ar}(X_1) \cdot \hat{V}\text{ar}(Y_2)}} = \frac{11,76}{\sqrt{(17,50)(11,9430)}} = 0,813.$$

$$r_{X_2 Y_2} = \sqrt{\lambda_2} \cdot \frac{a_{22}}{\sqrt{\hat{V}\text{ar}(X_2)}} = \sqrt{11,9430} \cdot \frac{(-0,1737)}{\sqrt{190,50}} = -0,043$$

$$\text{ou } r_{X_2 Y_2} = \frac{\text{C}\hat{\text{ov}}(X_2, Y_2)}{\sqrt{\hat{V}\text{ar}(X_2) \cdot \hat{V}\text{ar}(Y_2)}} = \frac{-2,07}{\sqrt{(190,50)(11,9430)}} = -0,043.$$

$$\text{Fato: } \text{C}\hat{\text{ov}}(X_j, Y_k) = a_{kj} \lambda_k = \frac{\sum_{i=1}^n X_{ij} Y_{ik} - \frac{(\sum_{i=1}^n X_{ij})(\sum_{i=1}^n Y_{ik})}{n}}{n-1}$$

$$\text{C}\hat{\text{ov}}(X_1, Y_2) = a_{21} \lambda_2 = (0,9848)(11,9430) = 11,76$$

$$\text{ou } \text{C}\hat{\text{ov}}(X_1, Y_2) = \frac{42511,08 - \frac{(500)(424,64)}{5}}{5-1} = 11,76.$$

Como a importância relativa das variáveis pode ser avaliada pela magnitude da correlação destas variáveis a partir dos últimos componentes, no exemplo em consideração, identifica-se a variável X_1 com maior correlação em Y_2 (0,813), como a de menor importância no estudo realizado. Neste exemplo, interpretando o coeficiente de ponderação da variável X_1 com maior valor em Y_2 (0,9848) chegaríamos à mesma conclusão. Cabe ressaltar que trabalhamos com apenas duas variáveis para facilitar os cálculos, mas na prática, em geral, trabalha-se com um número bem maior de variáveis, fazendo com que a discussão dos resultados fique muito mais interessante.

5.2.4.2. Exemplo 2 - Componentes principais a partir de uma matriz de correlação 2x2

Vamos considerar os mesmos dados do exemplo 1, que estão novamente apresentados na Tabela 5.9.

Tabela 5.9 – Observações relativas a duas variáveis avaliadas em cinco tratamentos

Tratamentos (n=5)	Variáveis originais		Variáveis padronizadas*	
	X_1	X_2	x_1	x_2
1	102	96	24,3827	6,9554
2	104	87	24,8608	6,3033
3	101	62	24,1436	4,4920
4	93	68	22,2313	4,9268
5	100	77	23,9046	5,5788
Variância	17,50	190,50	1	1

$$* x_{ij} = \frac{X_{ij}}{s(X_j)}. \text{ Poderíamos ter utilizado } x_{ij} = \frac{X_{ij} - \bar{X}_j}{s(X_j)}.$$

Neste caso, $r_{jj'} = r(X_j, X_{j'}) = \text{Cov}(x_j, x_{j'})$.

$$\text{A matriz de correlação é } \mathbf{R} = \begin{bmatrix} 1 & 0,5456 \\ 0,5456 & 1 \end{bmatrix}.$$

Solução:

A equação característica é:

$$|\mathbf{R} - \lambda \mathbf{I}| = 0$$

$$\begin{vmatrix} 1 - \lambda & 0,5456 \\ 0,5456 & 1 - \lambda \end{vmatrix} = 0,$$

isto é:

$$\lambda^2 - 2\lambda + 0,7023 = 0$$

Os autovalores (raízes próprias ou raízes características) obtidos dessa equação são:

$$\lambda_1 = 1,5456 \quad \text{e} \quad \lambda_2 = 0,4544$$

Pode-se observar que $1,5456 + 0,4544 = 2 = \text{traço}(\mathbf{R})$.

Obtenção do primeiro componente principal:

$$[\mathbf{R} - \lambda_1 \mathbf{I}] \tilde{\mathbf{a}}_1^* = \mathbf{0}$$

$$\begin{bmatrix} 1 - 1,5456 & 0,5456 \\ 0,5456 & 1 - 1,5456 \end{bmatrix} \begin{bmatrix} a_{11}^* \\ a_{12}^* \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} -0,5456 & 0,5456 \\ 0,5456 & -0,5456 \end{bmatrix} \begin{bmatrix} a_{11}^* \\ a_{12}^* \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{cases} -0,5456 a_{11}^* + 0,5456 a_{12}^* = 0 \\ 0,5456 a_{11}^* - 0,5456 a_{12}^* = 0 \end{cases}$$

Fazendo-se $a_{11}^* = 1$ na primeira equação, tem-se $a_{12}^* = 1$.

$$\tilde{a}_1^* = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad \text{e} \quad \|\tilde{a}_1^*\| = \sqrt{1^2 + 1^2} = \sqrt{2}$$

O autovetor normalizado é:

$$\tilde{a}_1 = \begin{bmatrix} a_{11} \\ a_{21} \end{bmatrix} = \frac{1}{\|\tilde{a}_1^*\|} \tilde{a}_1^* = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0,7071 \\ 0,7071 \end{bmatrix}.$$

Logo, $Y_1 = 0,7071x_1 + 0,7071x_2$.

Obtenção do segundo componente principal:

$$[R - \lambda_2 I] \tilde{a}_2^* = 0$$

$$\begin{bmatrix} 1 - 0,4544 & 0,5456 \\ 0,5456 & 1 - 0,4544 \end{bmatrix} \begin{bmatrix} a_{21}^* \\ a_{22}^* \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 0,5456 & 0,5456 \\ 0,5456 & 0,5456 \end{bmatrix} \begin{bmatrix} a_{21}^* \\ a_{22}^* \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{cases} 0,5456 a_{21}^* + 0,5456 a_{22}^* = 0 \\ 0,5456 a_{21}^* + 0,5456 a_{22}^* = 0 \end{cases}$$

Fazendo-se $a_{21}^* = -1$ na primeira equação, tem-se $a_{22}^* = 1$.

$$\tilde{a}_2^* = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \quad \text{e} \quad \|\tilde{a}_2^*\| = \sqrt{(-1)^2 + 1^2} = \sqrt{2}$$

O autovetor normalizado é:

$$\tilde{a}_2 = \begin{bmatrix} a_{21} \\ a_{22} \end{bmatrix} = \frac{1}{\|\tilde{a}_2^*\|} \tilde{a}_2^* = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} -0,7071 \\ 0,7071 \end{bmatrix}.$$

Logo, $Y_2 = -0,7071x_1 + 0,7071x_2$.

As informações importantes desta análise estão apresentadas na Tabela 5.10 e os escores relativos a cada tratamento para os dois primeiros componentes estão na Tabela 5.11.

Tabela 5.10 – Componentes principais obtidos da análise de duas variáveis x_1 e x_2

Componente principal	Variância (Autovalor)	Coeficientes associados às variáveis		Correlações entre x_j e Y_i		Variância de Y_i (%)	% Acumulada da variância de Y_i
		x_1	x_2	x_1	x_2		
Y_1	1,5456	0,7071	0,7071	0,879	0,879	77,28	77,28
Y_2	0,4544	-0,7071	0,7071	-0,476	0,476	22,72	100,00

Tabela 5.11 – Escores relativos aos cinco tratamentos obtidos em relação aos dois componentes principais

Tratamentos	Variáveis		Escores para os componentes	
	x_1	x_2	Y_1	Y_2
1	24,3827	6,9554	22,61	-12,32
2	24,8608	6,3033	22,04	-13,12
3	24,1436	4,4920	20,25	-13,90
4	22,2313	4,9268	19,20	-12,24
5	23,9046	5,5788	20,85	-12,96

A localização de cada tratamento pode ser vista na Figura 5.3.

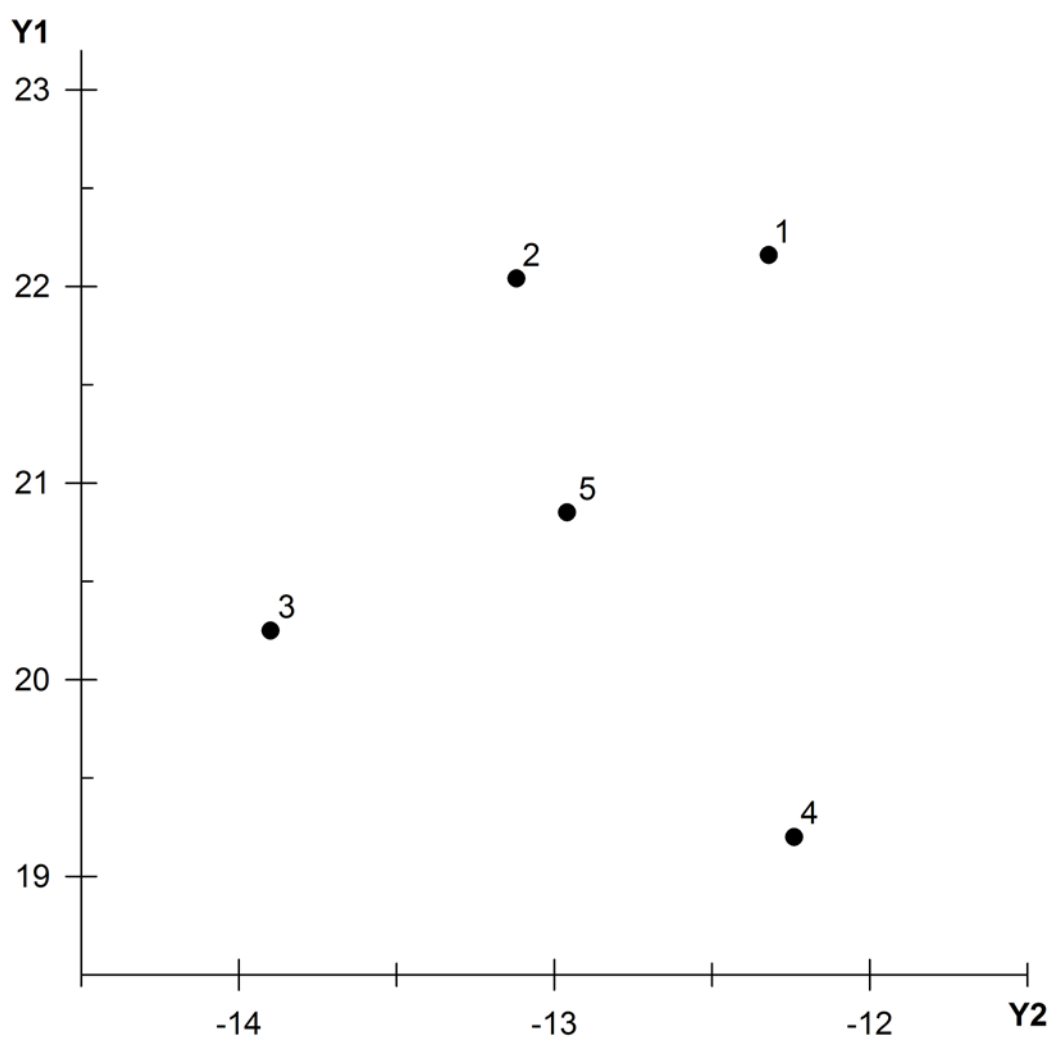


Figura 5.3 – Dispersão de cinco tratamentos em relação aos componentes principais Y_1 e Y_2

Pode-se também fazer o estudo obtendo-se os escores com as variáveis padronizadas com média 0 e variância 1, por exemplo, $Y_1 = 0,7071 \left(\frac{X_1 - \bar{X}_1}{s(X_1)} \right) + 0,7071 \left(\frac{X_2 - \bar{X}_2}{s(X_2)} \right)$. Esta padronização é mais utilizada. A interpretação será exatamente a mesma.

5.2.4.3. Exemplo 3 - Componentes principais a partir de uma matriz de correlação 3x3

Considere os dados da Tabela 5.12, relativas a três variáveis X_1 , X_2 e X_3 mensuradas em nove tratamentos.

Tabela 5.12 – Observações relativas a três variáveis avaliadas em nove tratamentos

Tratamentos	Variáveis originais			Variáveis padronizadas*		
	X_1	X_2	X_3	x_1	x_2	x_3
1	456,6667	2,0467	50,8667	14,4537	13,5453	41,7351
2	524,0000	2,3200	53,0667	16,5849	15,3541	43,5401
3	481,3333	1,8300	50,8000	15,2344	12,1112	41,6803
4	514,6667	2,2133	51,3000	16,2894	14,6479	42,0906
5	490,0000	1,9900	50,8667	15,5087	13,1701	41,7351
6	516,3333	2,1000	53,5000	16,3422	13,8981	43,8956
7	501,3333	1,9800	51,2667	15,8674	13,1039	42,0633
8	439,6667	1,9700	49,7000	13,9157	13,0377	40,7778
9	534,0000	2,1967	52,4000	16,9014	14,5381	42,9931
Média	495,3333	2,0719	51,5296	15,6775	13,7121	42,2789
Desvio Padrão	31,5951	0,1511	1,2188	1	1	1

$$* x_{ij} = \frac{X_{ij}}{s(X_j)}. \text{ Poderíamos ter utilizado } x_{ij} = \frac{X_{ij} - \bar{X}_j}{s(X_j)}.$$

Vamos proceder à análise de componentes principais, mas na realidade, cada observação da Tabela 5.12 é média de dados provenientes de um delineamento experimental com nove tratamentos em blocos casualizados com três repetições. Neste caso, o mais indicado seria fazer uma análise baseada em **Variáveis Canônicas**. Este assunto foi apresentado no Capítulo 4.

A matriz de correlação é:

$$\mathbf{R} = \begin{bmatrix} 1 & 0,6593 & 0,8255 \\ 0,6593 & 1 & 0,6591 \\ 0,8255 & 0,6591 & 1 \end{bmatrix}$$

Solução:

Os autovalores e autovetores podem ser facilmente obtidos utilizando-se um programa computacional apropriado. Entretanto, apesar de bastante trabalhoso, como exercício, vamos obter as raízes da equação do terceiro grau de acordo com Harris (1975).

$$|\mathbf{R} - \lambda \mathbf{I}| = 0$$

$$\begin{vmatrix} 1-\lambda & 0,6593 & 0,8255 \\ 0,6593 & 1-\lambda & 0,6591 \\ 0,8255 & 0,6591 & 1-\lambda \end{vmatrix} = 0$$

$$(1-\lambda)^3 + (0,6593)(0,6591)(0,8255) + (0,8255)(0,6593)(0,6591) - [(0,8255)^2(1-\lambda) + (0,6591)^2(1-\lambda) + (0,6593)^2(1-\lambda)] = 0$$

$$(1-3\lambda+3\lambda^2-\lambda^3) + 0,7174 - (1,5505)(1-\lambda) = 0$$

$$-\lambda^3 + 3\lambda^2 - 1,4495\lambda + 0,1669 = 0$$

ou ainda,

$$\lambda^3 - 3\lambda^2 + 1,4495\lambda - 0,1669 = 0$$

Dada a equação cúbica da forma:

$$\lambda^3 + p\lambda^2 + q\lambda + r = 0,$$

(a) Elimina-se o termo λ^2 através da substituição $\lambda = x - (p/3)$, e assim teremos:

$$x^3 + ax + b = 0,$$

em que,

$$a = q - (p^2/3) \quad \text{e} \quad b = \frac{2p^3 - 9pq + 27r}{27}$$

$$a = 1,4495 - \frac{(-3)^2}{3} = -1,5505$$

$$b = \frac{2(-3)^3 - 9(-3)(1,4495) + 27(-0,1669)}{27} = -0,7174$$

(b) Calcula-se o discriminante que é dado por:

$$d = \frac{b^2}{4} + \frac{a^3}{27} = \frac{(-0,7174)^2}{4} + \frac{(-1,5505)^3}{27} = -0,009389$$

Como $d < 0$, existem três raízes reais e distintas.

$$\text{Uma raiz é } x_1 = 2\left(\frac{-a}{3}\right)^{1/2} \cdot \cos\left(\frac{\theta}{3}\right),$$

em que

$$\cos \theta = \frac{-2,59808b}{(-a)^{3/2}} = \frac{(-2,59808)(-0,7174)}{(1,5505)^{3/2}} = 0,9654.$$

Logo, $\theta = 15,1159^\circ$

e

$$x_1 = 2\left(\frac{1,5505}{3}\right)^{1/2} \cdot \cos\left(\frac{15,1159^\circ}{3}\right) = 1,3478 \cos(5,0386^\circ) = 1,4322.$$

Como $\lambda = x - \frac{p}{3}$, tem-se que:

$$\lambda = 1,4322 - \frac{(-3)}{3} = 2,4322.$$

A outra raiz é $x_2 = 2\left(\frac{-a}{3}\right)^{1/2} \cdot \cos\left(\frac{\theta}{3} + 120^\circ\right)$

$$x_2 = 2\left(\frac{1,5505}{3}\right)^{1/2} \cdot \cos(125,0386^\circ)$$

$$x_2 = 1,4378(-0,5741) = -0,8254.$$

Logo, $\lambda = -0,8254 - \frac{(-3)}{3} = 0,1746.$

A outra raiz é $x_3 = 2\left(\frac{-a}{3}\right)^{1/2} \cdot \cos\left(\frac{\theta}{3} + 240^\circ\right)$

$$x_3 = 2\left(\frac{1,5505}{3}\right)^{1/2} \cdot \cos(245,0386^\circ)$$

$$x_3 = 1,4378(-0,4220) = -0,6067.$$

Logo, $\lambda = -0,6067 - \frac{(-3)}{3} = 0,3933.$

Assim, os autovalores em ordem decrescente de magnitude são:

$$\lambda_1 = 2,4322$$

$$\lambda_2 = 0,3933$$

$$\lambda_3 = 0,1746$$

Observe que, $2,4322 + 0,3933 + 0,1746 = 3 = \text{traço}(\mathbf{R}).$

Obtenção do primeiro componente principal:

$$[\mathbf{R} - \lambda_1 \mathbf{I}] \tilde{\mathbf{a}}_1^* = \mathbf{0}$$

$$\begin{bmatrix} 1-2,4322 & 0,6593 & 0,8255 \\ 0,6593 & 1-2,4322 & 0,6591 \\ 0,8255 & 0,6591 & 1-2,4322 \end{bmatrix} \begin{bmatrix} a_{11}^* \\ a_{12}^* \\ a_{13}^* \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\begin{cases} -1,4322 a_{11}^* + 0,6593 a_{12}^* + 0,8255 a_{13}^* = 0 \\ 0,6593 a_{11}^* - 1,4322 a_{12}^* + 0,6591 a_{13}^* = 0 \\ 0,8255 a_{11}^* + 0,6591 a_{12}^* - 1,4322 a_{13}^* = 0 \end{cases}$$

O sistema de equações homogêneas anterior é indeterminado. Então podemos tomar, por exemplo, $a_{13}^* = 1$ e considerar somente duas das três equações, como a seguir:

$$\begin{cases} -1,4322 a_{11}^* + 0,6593 a_{12}^* = -0,8255 \\ 0,6593 a_{11}^* - 1,4322 a_{12}^* = -0,6591 \end{cases}$$

Resolvendo matricialmente, teremos:

$$\begin{bmatrix} a_{11}^* \\ a_{12}^* \end{bmatrix} = \frac{1}{1,6165} \begin{bmatrix} -1,4322 & -0,6593 \\ -0,6593 & -1,4322 \end{bmatrix} \begin{bmatrix} -0,8255 \\ -0,6591 \end{bmatrix} = \begin{bmatrix} 1,0002 \\ 0,9206 \end{bmatrix}.$$

$$\text{Logo, } \tilde{a}_1^* = \begin{bmatrix} 1,0002 \\ 0,9206 \\ 1 \end{bmatrix} \text{ e } \|\tilde{a}_1^*\| = \sqrt{(1,0002)^2 + (0,9206)^2 + 1^2} = \sqrt{2,8479}.$$

O autovetor normalizado é:

$$\tilde{a}_1 = \begin{bmatrix} a_{11} \\ a_{12} \\ a_{13} \end{bmatrix} = \frac{\tilde{a}_1^*}{\|\tilde{a}_1^*\|} = \begin{bmatrix} 0,5927 \\ 0,5455 \\ 0,5926 \end{bmatrix}. \text{ Note que } \tilde{a}_1' \tilde{a}_1 = 1.$$

Assim, $Y_1 = 0,5927 x_1 + 0,5455 x_2 + 0,5926 x_3$.

Obtenção do segundo componente principal:

$$[R - \lambda_2 I] \tilde{a}_2^* = 0$$

$$\begin{bmatrix} 1-0,3933 & 0,6593 & 0,8255 \\ 0,6593 & 1-0,3933 & 0,6591 \\ 0,8255 & 0,6591 & 1-0,3933 \end{bmatrix} \begin{bmatrix} a_{21}^* \\ a_{22}^* \\ a_{23}^* \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\begin{cases} 0,6067 a_{21}^* + 0,6593 a_{22}^* + 0,8255 a_{23}^* = 0 \\ 0,6593 a_{21}^* + 0,6067 a_{22}^* + 0,6591 a_{23}^* = 0 \\ 0,8255 a_{21}^* + 0,6591 a_{22}^* + 0,6067 a_{23}^* = 0 \end{cases}$$

Dando o mesmo procedimento anterior e, tomando-se $a_{23}^* = 1$, teremos:

$$\begin{cases} 0,6067 a_{21}^* + 0,6593 a_{22}^* = -0,8255 \\ 0,6593 a_{21}^* + 0,6067 a_{22}^* = -0,6591 \end{cases}$$

Resolvendo matricialmente, teremos:

$$\begin{bmatrix} a_{21}^* \\ a_{22}^* \end{bmatrix} = \frac{1}{-0,0666} \begin{bmatrix} 0,6067 & -0,6593 \\ -0,6593 & 0,6067 \end{bmatrix} \begin{bmatrix} 0,8255 \\ 0,6591 \end{bmatrix} = \begin{bmatrix} -0,9953 \\ 2,1678 \end{bmatrix}.$$

$$\text{Logo, } \tilde{a}_2^* = \begin{bmatrix} -0,9953 \\ 2,1678 \\ 1 \end{bmatrix} \text{ e } \|\tilde{a}_2^*\| = \sqrt{(-0,9953)^2 + (2,1678)^2 + 1^2} = \sqrt{6,6899}.$$

O autovetor normalizado é:

$$\tilde{a}_2 = \begin{bmatrix} a_{21} \\ a_{22} \\ a_{23} \end{bmatrix} = \frac{\tilde{a}_2^*}{\|\tilde{a}_2^*\|} = \begin{bmatrix} -0,3848 \\ 0,8381 \\ -0,3866 \end{bmatrix}. \text{ Note que } \tilde{a}_2' \tilde{a}_2 = 1.$$

Assim, $Y_2 = -0,3848 x_1 + 0,8381 x_2 - 0,3866 x_3$.

Obtenção do terceiro componente principal:

$$[R - \lambda_3 I] \tilde{a}_3^* = 0$$

$$\begin{bmatrix} 1-0,1746 & 0,6593 & 0,8255 \\ 0,6593 & 1-0,1746 & 0,6591 \\ 0,8255 & 0,6591 & 1-0,1746 \end{bmatrix} \begin{bmatrix} a_{31}^* \\ a_{32}^* \\ a_{33}^* \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\begin{cases} 0,8254 a_{31}^* + 0,6593 a_{32}^* + 0,8255 a_{33}^* = 0 \\ 0,6593 a_{31}^* + 0,8254 a_{32}^* + 0,6591 a_{33}^* = 0 \\ 0,8255 a_{31}^* + 0,6591 a_{32}^* + 0,8254 a_{33}^* = 0 \end{cases}$$

Com o mesmo procedimento dado anteriormente e, tomando-se $a_{33}^* = 1$, teremos:

$$\begin{cases} 0,8254 a_{31}^* + 0,6593 a_{32}^* = -0,8255 \\ 0,6593 a_{31}^* + 0,8254 a_{32}^* = -0,6591 \end{cases}$$

Resolvendo matricialmente, teremos:

$$\begin{bmatrix} a_{31}^* \\ a_{32}^* \end{bmatrix} = \frac{1}{0,2466} \begin{bmatrix} 0,8254 & -0,6593 \\ -0,6593 & 0,8254 \end{bmatrix} \begin{bmatrix} -0,8255 \\ -0,6591 \end{bmatrix} = \begin{bmatrix} -1,0009 \\ 0,0009 \end{bmatrix}.$$

$$\text{Logo, } \tilde{a}_3^* = \begin{bmatrix} -1,0009 \\ 0,0009 \\ 1 \end{bmatrix} \text{ e } \|\tilde{a}_3^*\| = \sqrt{(-1,0009)^2 + (0,0009)^2 + 1^2} = \sqrt{2,0018}.$$

O autovetor normalizado é:

$$\tilde{a}_3 = \begin{bmatrix} a_{31} \\ a_{32} \\ a_{33} \end{bmatrix} = \frac{a_3^*}{\|a_3^*\|} = \begin{bmatrix} -0,7074 \\ 0,0006 \\ 0,7068 \end{bmatrix}. \text{ Note que } \tilde{a}_3' \tilde{a}_3 = 1.$$

Assim, $Y_3 = -0,7074 x_1 + 0,0006 x_2 + 0,7068 x_3$.

As informações importantes desta análise estão apresentadas na Tabela 5.13.

Os escores relativos a cada tratamento, para os dois primeiros componentes estão apresentados na Tabela 5.14. A localização de cada tratamento em relação aos dois primeiros componentes principais pode ser vista na Figura 5.4. Quanto maior a proximidade entre dois tratamentos, maior é a similaridade entre eles. A composição dos grupos de tratamentos pode ser feita de forma subjetiva (visualmente). Neste caso, com base na Figura 5.4, pode-se considerar, por exemplo, a existência de três grupos. No grupo I estariam os tratamentos 2, 4, 6 e 9; no grupo II estariam os tratamentos 1, 3, 5, e 7; no grupo III o tratamento 8.

Tabela 5.13 – Componentes principais obtidos da análise de três variáveis x_1 , x_2 e x_3

Componente principal	Variância (Autovalor)	Coeficientes associados às variáveis			Correlação entre x_i e Y_i			Variância de Y_i (%)	% Acumulada da variância de Y_i
		x_1	x_2	x_3	x_1	x_2	x_3		
Y_1	2,4322	0,5927	0,5455	0,5926	0,924	0,851	0,924	81,07	81,07
Y_2	0,3933	-0,3848	0,8381	-0,3866	-0,242	0,526	-0,242	13,11	94,18
Y_3	0,1746	-0,7074	0,0006	0,7068	-0,295	0,0002	0,295	5,82	100,00

Tabela 5.14 – Escores relativos aos nove tratamentos em relação aos dois primeiros componentes principais

Tratamentos (n=9)	Escore para os Componentes	
	Y_1^*	Y_2
1	40,69	-10,34
2	44,01	-10,35
3	40,34	-11,82
4	42,59	-10,26
5	41,11	-11,06
6	43,28	-11,61
7	41,48	-11,38
8	39,52	-10,19
9	43,42	-10,94
Variância: $\hat{V}(Y_i)$	2,4322	0,3933
Covariância: $Cov(Y_1, Y_2)$	0	

*Ilustração: $Y_{11} = 0,5927(14,4537) + 0,5455(13,5433) + 0,5926(41,7351) = 40,69$.

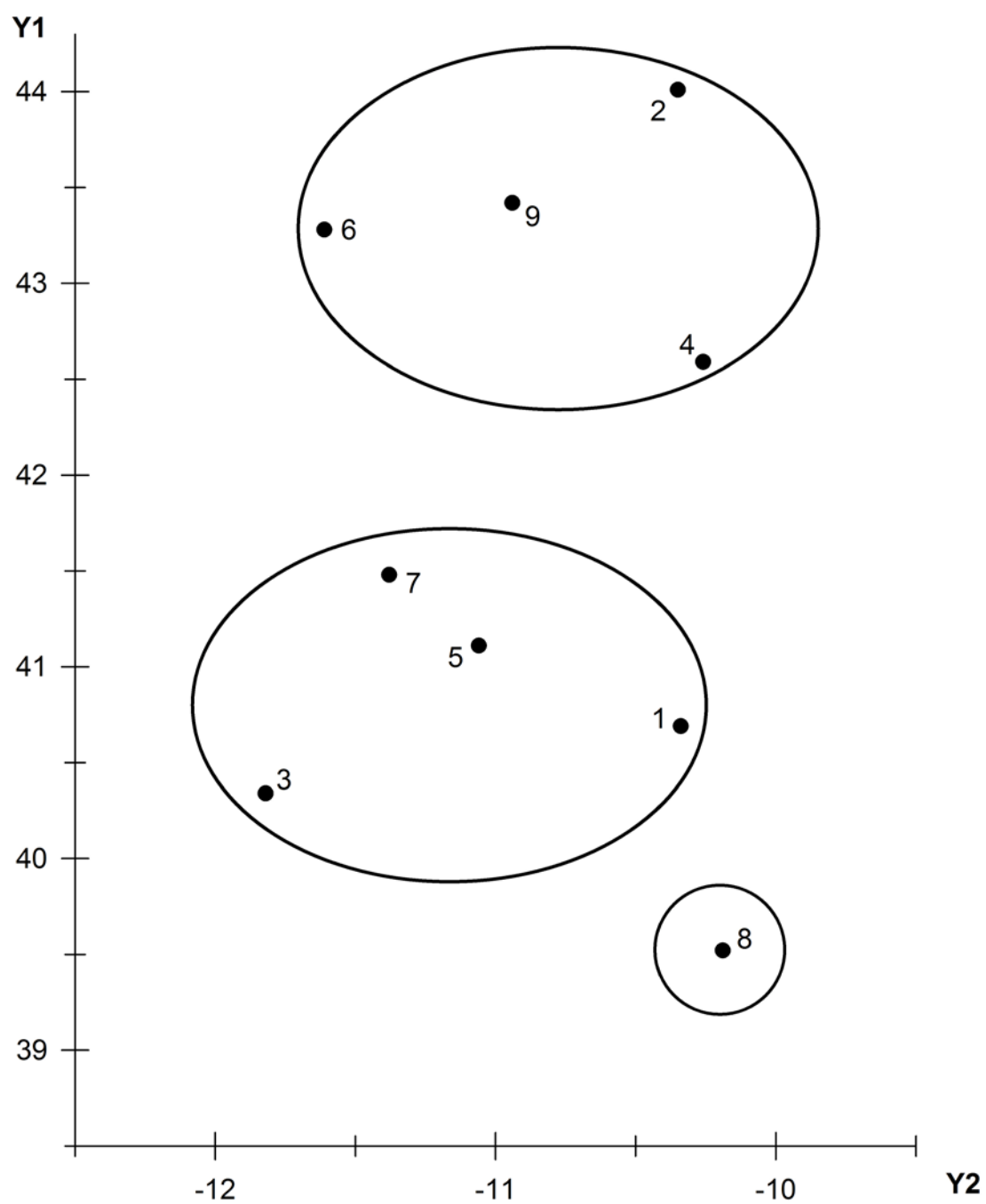


Figura 5.4 – Dispersão de nove tratamentos em relação aos dois primeiros componentes principais

Considere agora a Figura 5.5 onde plotamos os escores, porém, com escalas muito diferentes para os eixos Y_1 e Y_2 .

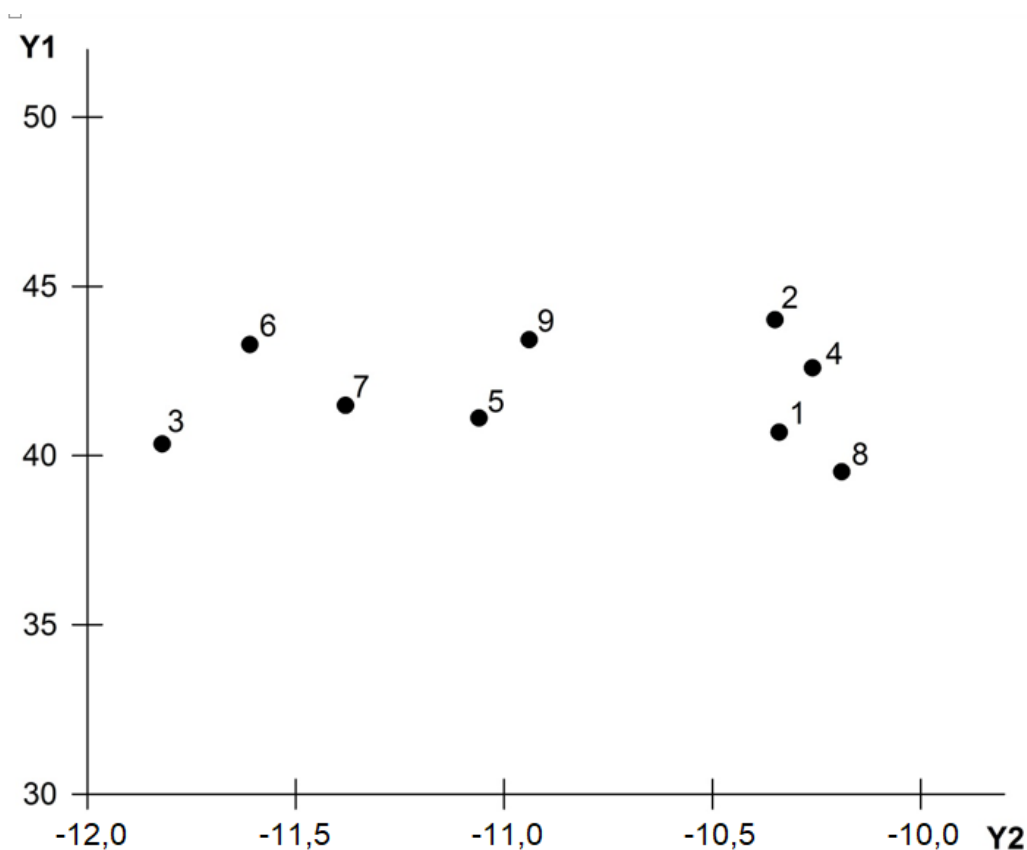


Figura 5.5 - Dispersão de nove tratamentos em relação aos dois primeiros componentes principais

Devido à diferença de escala para os eixos Y_1 e Y_2 , a Figura 5.5 não reflete visualmente a verdadeira dissimilaridade entre os tratamentos. Assim, se erroneamente fosse feita uma interpretação com base na Figura 5.5, certamente cometeríamos um erro grosseiro.

Pela Figura 5.5, visualmente poderíamos considerar os tratamentos 1, 2, 4 e 8 como pertencentes a um mesmo grupo, o que estaria errado, pois, pela Figura correta, isto é, Figura 5.4, os tratamentos 2 e 8 são os mais divergentes, fato este que pode ser comprovado com o cálculo de distâncias gráficas (Tabela 5.15).

As distâncias gráficas apresentadas na Figura 5.3, podem ser estimadas pela distância euclidiana.

Sendo Y_{ij} o escore no i -ésimo tratamento ($i = 1, 2, \dots, n$) e j -ésimo componente principal ($j = 1, 2, \dots, k$), define-se a distância euclidiana entre dois tratamentos i e i' por meio da expressão:

$$d_{cp_{ii'}} = \sqrt{\sum_{j=1}^k (Y_{ij} - Y_{i'j})^2}.$$

Assim, para os dois primeiros componentes, teremos:

$$dcp_{ii'} = \sqrt{(Y_{i1} - Y_{i'1})^2 + (Y_{i2} - Y_{i'2})^2}.$$

A distância euclidiana é uma medida de dissimilaridade. Pela expressão anterior, as distâncias foram obtidas com os dados da Tabela 5.14 e estão apresentadas na Tabela 5.15. Como ilustração, temos que a estimativa de dcp_{12} é:

$$dcp_{12} = \sqrt{(40,69 - 44,01)^2 + [(-10,34) - (-10,35)]^2} = 3,32.$$

Tabela 5.15 – Dissimilaridade entre tratamentos baseada na distância euclidiana, obtida a partir dos escores dos dois primeiros componentes principais

Tratamentos	2	3	4	5	6	7	8	9
1	3,32	1,52	1,90	0,833	2,88	1,31	1,18	2,79
2		3,95	1,42	2,98	1,46	2,73	4,49	0,834
3			2,74	1,08	2,95	1,22	1,82	3,20
4				1,68	1,52	1,58	3,07	1,07
5					2,24	0,49	1,812	2,31
6						1,814	4,02	0,68
7							2,29	1,99
8								3,97

Um dos objetivos do uso dos componentes principais em estudos sobre a divergência entre tratamentos é avaliar a dissimilaridade deles em gráficos de dispersão que têm os primeiros componentes como eixos de referência. Como nesta técnica é feita uma simplificação do espaço p -dimensional para o bi ou tridimensional, há certas distorções nas distâncias gráficas. Entretanto, há entre as estimativas das distâncias gráficas ou distâncias euclidianas baseadas nos escores dos primeiros componentes principais e a distância euclidiana baseada nos dados das p variáveis (variáveis originais ou padronizadas de acordo com a análise) uma relação matemática dada por:

$$\theta = \frac{\sum_{i < i'} \sum dcp_{ii'}^2}{p \sum_{i < i'} \sum d_{ii'}^2}$$

em que:

$dcp_{ii'}^2$: Quadrado da distância euclidiana estimada a partir dos escores de k componentes principais

$$d_{ii'}^2 = \frac{1}{p} \sum_{j=1}^p (x_{ij} - x_{i'j})^2, \text{ isto é o quadrado da distância euclidiana média estimada a}$$

partir das **p** variáveis (variáveis originais ou padronizadas de acordo com a análise)

A estatística $1 - \theta$ mede o grau de distorção proporcionado pela técnica dos componentes principais ao passar do espaço **p**-dimensional para o **k**-dimensional (**k** < **p**). Deduz-se ainda que θ é a proporção acumulada, explicada pelos **k** componentes principais.

Na Tabela 5.16 estão apresentadas as distâncias euclidianas médias entre pares de nove tratamentos, considerando as variáveis padronizadas.

Tabela 5.16 - Distâncias euclidianas médias entre pares de nove tratamentos, considerando as

$$\text{variáveis padronizadas } (x_{ij} = \frac{X_{ij}}{s(X_j)})$$

Tratamentos	2	3	4	5	6	7	8	9
1	1,9211	0,9432	1,2532	0,6465	1,6692	0,8758	0,6985	1,6891
2		2,2948	0,9464	1,7498	0,8766	1,6082	2,5897	0,5959
3			1,6037	0,6323	1,7633	0,7148	1,0664	1,8612
4				1,9865	1,1288	0,9242	1,8212	0,6327
5					1,4015	0,2833	1,0757	1,3408
6						1,1851	2,3344	0,7157
7							1,3498	1,1533
8								2,3147

$$\sum_{i < i'} \sum d_{ii'}^2 = 72,00$$

Ilustração do cálculo:

$$d_{12} = \sqrt{\frac{1}{3} [(14,4537 - 16,5849)^2 + (13,5453 - 15,3541)^2 + (41,7351 - 43,5401)^2]}$$

$$d_{12} = \sqrt{\frac{1}{3} (11,07179588)} = 1,9211, \text{ etc.}$$

No exemplo em questão, temos que:

$$\theta = \frac{\sum_{i < i'} \sum d_{cp_{ii'}}^2}{p \sum_{i < i'} \sum d_{ii'}^2} = \frac{203,10}{3(72,00)} = 0,9403 \quad \text{ou} \quad 94,03\% .$$

$$1 - \theta = 1 - 0,9403 = 0,0597 \quad \text{ou} \quad 5,97\% .$$

Assim, o grau de distorção das distâncias ao passar do espaço tri para o bidimensional foi 5,97%. O valor de θ deveria ser 0,9418 de acordo com a Tabela 5.13. A pequena diferença observada deve-se a problemas de aproximação de decimais.

Baseado na Figura 5.4, vimos que formamos os grupos de forma subjetiva (visualmente). Esta análise será complementada com a elaboração de um dendrograma (gráfico em forma de árvore) e a determinação do número de grupos, cujo assunto é tratado em Análise de Agrupamento apresentada no Capítulo 6.

5.2.5. Resultados obtidos através do programa SAS

5.2.5.1. Exemplo 1

(a) Programa:

```

OPTIONS NODATE NONUMBER;
/* COMPONENTES PRINCIPAIS USANDO S */
DATA EXEMPLO1;
INPUT TRAT X1 X2;
CARDS;
1 102 96
2 104 87
3 101 62
4 93 68
5 100 77
;

PROC PRINCOMP COV DATA=EXEMPLO1 OUT=SAIDA;
VAR X1 X2;
TITLE 'ACP COM A MATRIZ DE COVARIÂNCIAS';

PROC CORR DATA=SAIDA;
VAR X1 X2;
WITH PRIN1 PRIN2;
TITLE 'GERA CORRELAÇÕES DOS CP COM AS VARIÁVEIS X ';

PROC PRINT;
VAR TRAT PRIN1 PRIN2 X1 X2;
TITLE 'ESCORES DOS COMPONENTES PRINCIPAIS';
RUN;
```

(b) Resultado da Análise:

ACP COM A MATRIZ DE COVARIÂNCIAS

The PRINCOMP Procedure

Observations 5
Variables 2

Simple Statistics

	X1	X2
Mean	100.0000000	78.00000000
Std	4.1833001	13.80217374

Covariance Matrix

	X1	X2
X1	17.5000000	31.5000000
X2	31.5000000	190.5000000

Total Variance 208

Eigenvalues of the Covariance Matrix

	Eigenvalue	Difference	Proportion	Cumulative
1	196.057048	184.114095	0.9426	0.9426
2	11.942952		0.0574	1.0000

Eigenvectors

	Prin1	Prin2
X1	0.173731	0.984793
X2	0.984793	-.173731

GERA CORRELAÇÕES DOS CP COM AS VARIÁVEIS X

The CORR Procedure

2 With Variables: Prin1 Prin2
2 Variables: X1 X2

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
Prin1	5	0	14.00204	0	-15.58296	18.07374
Prin2	5	0	3.45586	0	-5.15624	3.76450
X1	5	100.00000	4.18330	500.00000	93.00000	104.00000
X2	5	78.00000	13.80217	390.00000	62.00000	96.00000

Pearson Correlation Coefficients, N = 5
Prob > |r| under H0: Rho=0

	X1	X2
Prin1	0.58150 0.3038	0.99905 <.0001
Prin2	0.81355 0.0939	-0.04350 0.9446

ESCORES DOS COMPONENTES PRINCIPAIS

Obs	TRAT	Prin1	Prin2	X1	X2
1	1	18.0737	-1.15758	102	96
2	2	9.5581	2.37559	104	87
3	3	-15.5830	3.76450	101	62
4	4	-11.0641	-5.15624	93	68
5	5	-0.9848	0.17373	100	77

Note que os escores aqui foram obtidos com as variáveis centradas na média, por exemplo, $Y_1 = \text{Prin1} = 0,173731(X_1 - \bar{X}_1) + 0,9844797(X_2 - \bar{X}_2)$.

5.2.5.2. Exemplo 2

(a) Programa:

```

OPTIONS NODATE NONUMBER;
/* COMPONENTES PRINCIPAIS USANDO R */
DATA EXEMPLO2;
INPUT TRAT X1 X2;
CARDS;
1 102 96
2 104 87
3 101 62
4 93 68
5 100 77
;

PROC PRINCOMP DATA=EXEMPLO2 OUT=SAIDA;
VAR X1 X2;
TITLE 'ACP COM A MATRIZ DE CORRELAÇÃO';

PROC CORR DATA=SAIDA;
VAR X1 X2;
WITH PRIN1 PRIN2;
TITLE 'GERA CORRELAÇÕES DOS CP COM AS VARIÁVEIS X: rYX=rYx';

PROC PRINT;
VAR TRAT PRIN1 PRIN2 X1 X2;
TITLE 'ESCORES DOS COMPONENTES PRINCIPAIS';
RUN;
```

(b) Resultado da Análise:

ACP COM A MATRIZ DE CORRELAÇÃO

The PRINCOMP Procedure

Observations 5
Variables 2

Simple Statistics

	X1	X2
Mean	100.0000000	78.00000000
Std	4.1833001	13.80217374

Correlation Matrix

	X1	X2
X1	1.0000	0.5456
X2	0.5456	1.0000

Eigenvalues of the Correlation Matrix

	Eigenvalue	Difference	Proportion	Cumulative
1	1.54556191	1.09112382	0.7728	0.7728
2	0.45443809		0.2272	1.0000

Eigenvectors

	Prin1	Prin2
X1	0.707107	0.707107
X2	0.707107	-.707107

GERA CORRELAÇÕES DOS CP COM AS VARIÁVEIS X: $r_{YX}=r_{Yx}$

The CORR Procedure

2 With Variables: Prin1 Prin2
2 Variables: X1 X2

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
Prin1	5	0	1.24321	0	-1.69553	1.26023
Prin2	5	0	0.67412	0	-0.67090	0.98874
X1	5	100.00000	4.18330	500.00000	93.00000	104.00000
X2	5	78.00000	13.80217	390.00000	62.00000	96.00000

Pearson Correlation Coefficients, N = 5

Prob > |r| under H0: Rho=0

	X1	X2
Prin1	0.87908 0.0496	0.87908 0.0496
Prin2	0.47667 0.4169	-0.47667 0.4169

ESCORES DOS COMPONENTES PRINCIPAIS

Obs	TRAT	Prin1	Prin2	X1	X2
1	1	1.26023	-0.58411	102	96
2	2	1.13721	0.21504	104	87
3	3	-0.65067	0.98874	101	62
4	4	-1.69553	-0.67090	93	68
5	5	-0.05123	0.05123	100	77

Note que os escores aqui foram obtidos com as variáveis padronizadas com média 0 e

variância 1, por exemplo, $Y_1 = \text{Prin1} = 0,707107 \left(\frac{X_1 - \bar{X}_1}{s(X_1)} \right) + 0,707107 \left(\frac{X_2 - \bar{X}_2}{s(X_2)} \right)$.

5.2.5.3. Exemplo 3

(a) Programa:

```

OPTIONS NODATE NONUMBER;
/* COMPONENTES PRINCIPAIS - OUTRO EXEMPLO */
DATA EXEMPLO3;
INPUT TRAT X1 X2 X3;
CARDS;
1 456.6667 2.0467 50.8667
2 524.0000 2.3200 53.0667
3 481.3333 1.8300 50.8000
4 514.6667 2.2133 51.3000
5 490.0000 1.9900 50.8667
6 516.3333 2.1000 53.5000
7 501.3333 1.9800 51.2667
8 439.6667 1.9700 49.7000
9 534.0000 2.1967 52.4000
;
PROC PRINCOMP DATA=EXEMPLO3 OUT=SAIDA;
VAR X1-X3;
TITLE 'ACP COM A MATRIZ DE CORRELAÇÃO';

PROC CORR DATA=SAIDA;
VAR X1-X3;
WITH PRIN1-PRIN3;
TITLE 'GERA CORRELAÇÕES DOS CP COM AS VARIÁVEIS X: rYX=rYx';

PROC PRINT; VAR TRAT PRIN1-PRIN3 X1-X3;
TITLE 'ESCORES DOS COMPONENTES PRINCIPAIS';
RUN;

```

(b) Resultado da Análise:

ACP COM A MATRIZ DE CORRELAÇÃO

The PRINCOMP Procedure

Observations	9
Variables	3

Simple Statistics

	X1	X2	X3
Mean	495.3333333	2.071855556	51.52964444
Std	31.5950830	0.151079318	1.21877020

Correlation Matrix

	X1	X2	X3
X1	1.0000	0.6594	0.8255
X2	0.6594	1.0000	0.6592
X3	0.8255	0.6592	1.0000

Eigenvalues of the Correlation Matrix

	Eigenvalue	Difference	Proportion	Cumulative
1	2.43240748	2.03928407	0.8108	0.8108
2	0.39312342	0.21865431	0.1310	0.9418
3	0.17446910		0.0582	1.0000

Eigenvectors

	Prin1	Prin2	Prin3
X1	0.592654	-.385270	-.707339
X2	0.545521	0.838097	0.000581
X3	0.592595	-.386212	0.706875

GERA CORRELAÇÕES DOS CP COM AS VARIÁVEIS X: $r_{YX}=r_{YX}$

The CORR Procedure

3 With Variables: Prin1 Prin2 Prin3
3 Variables: X1 X2 X3

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
Prin1	9	0	1.55962	0	-2.30158	2.18108
Prin2	9	0	0.62700	0	-0.93974	0.69356
Prin3	9	0	0.41769	0	-0.56548	0.67276
X1	9	495.33333	31.59508	4458	439.66670	534.00000
X2	9	2.07186	0.15108	18.64670	1.83000	2.32000
X3	9	51.52964	1.21877	463.76680	49.70000	53.50000

Pearson Correlation Coefficients, N = 9
Prob > |r| under H0: Rho=0

	X1	X2	X3
Prin1	0.92431 0.0004	0.85080 0.0036	0.92422 0.0004
Prin2	-0.24156 0.5312	0.52548 0.1463	-0.24215 0.5302
Prin3	-0.29545 0.4402	0.00024 0.9995	0.29526 0.4405

A porcentagem de variância de cada uma das variáveis originais (X_j) explicada por cada um dos componentes principais ($Y_i=Prin_i$) é igual ao quadrado da correlação entre X_j e Y_i .

ESCORES DOS COMPONENTES PRINCIPAIS

Obs	TRAT	Prin1	Prin2	Prin3	X1	X2	X3
1	1	-1.13847	0.54203	0.48106	456.667	2.0467	50.8667
2	2	2.18108	0.53992	0.25065	524.000	2.3200	53.0667
3	3	-1.49068	-0.93974	-0.11069	481.333	1.8300	50.8000
4	4	0.76172	0.62167	-0.56548	514.667	2.2133	51.3000
5	5	-0.71795	-0.17897	-0.26542	490.000	1.9900	50.8667
6	6	1.45357	-0.72433	0.67276	516.333	2.1000	53.5000
7	7	-0.34698	-0.49940	-0.28718	501.333	1.9800	51.2667
8	8	-2.30158	0.69356	0.18468	439.667	1.9700	49.7000
9	9	1.59928	-0.05474	-0.36038	534.000	2.1967	52.4000

Note que os escores aqui foram obtidos com as variáveis padronizadas com média 0 e variância 1, por exemplo:

$$Y_1 = \text{Prin1} = 0,592654 \left(\frac{X_1 - \bar{X}_1}{s(X_1)} \right) + 0,545521 \left(\frac{X_2 - \bar{X}_2}{s(X_2)} \right) + 0,592595 \left(\frac{X_3 - \bar{X}_3}{s(X_3)} \right).$$

5.3. EXERCÍCIOS

Considere os dados da Tabela 5.17, referentes a três variáveis X_1 , X_2 e X_3 mensuradas em uma amostra constituída de cinco indivíduos.

Tabela 5.17 – Resultados experimentais

Indivíduos	Variáveis		
	X_1	X_2	X_3
A	102	17	4
B	104	19	5
C	101	21	7
D	93	18	1
E	100	15	3

Pede-se:

1. Faça uma Análise de Componentes Principais a partir da matriz de covariância **S** incluindo todas as informações e interpretações de interesse nesta análise.
2. Faça a mesma análise a partir da matriz de correlação **R**.
3. Faça ambas as análises utilizando o Programa SAS.

5.4. REFERÊNCIAS

- CRUZ, C. D.; CARNEIRO, P. C. S.; REGAZZI, A. J. **Modelos biométricos aplicados ao melhoramento genético**. v.2, 3. ed. rev. e ampl., Viçosa: UFV, 2014. 668p.
- CRUZ, C. D.; REGAZZI, A. J.; CARNEIRO, P. C. S. **Modelos biométricos aplicados ao melhoramento genético**. v.1, 3. ed., Viçosa: UFV, 2012. 514p.
- JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. 4th ed., USA, Upper Saddle River, New Jersey: Prentice Hall, Inc., 1998. 816p.
- JOLLIFFE, I. T. Discarding variables in a principal component analysis. I. Artificial data. **Appl. Stat.**, Série C, v.21, p. 160-173, 1972.
- JOLLIFFE, I.T. Discarding variables in a principal component analysis; II. Real data. **Appl. Stat.**, Série C, v.22, p. 21-31, 1973.
- KAISER, H. F. The varimax criterion for analytic rotation in factor analysis. **Psychometrika**, **23**, 187-200. 1958.
- KHATTREE, R.; NAIK, D. N. **Multivariate data reduction and discrimination with SAS Software**. USA: SAS Institute Inc., Cary, NC, 2000. 558p.
- MARDIA, K. V.; KENT, J. T.; BIBBY, J. M. **Multivariate analysis**. 6th ed., London: Academic Press, 1997. 518p.
- MORRISON, D. F. **Multivariate statistical methods**. 2nd ed., USA: McGraw-Hill Book Company, 1976. 415p.
- RAO, C. R. **Advanced statistical methods in biometric research**. New York: John Wiley and Sons, 1952. 390p.
- RENCER, A. C. **Methods of multivariate analysis**. 2nd ed., USA: John Wiley and Sons, 2002. 738p.

CAPÍTULO 6

ANÁLISE DE AGRUPAMENTO

6.1. INTRODUÇÃO

Em muitos estudos é comum efetuarem-se medidas de natureza distinta (**p** variáveis) de um conjunto de unidades amostrais. De posse dessas medidas, objetiva-se encontrar a melhor maneira de descrever os padrões de similaridade mútuas do conjunto de unidades.

Dentre os diferentes métodos que podem ser usados com este propósito, destaca-se a Análise de Agrupamento (“Cluster Analysis”).

A análise de agrupamento tem por objetivo resolver o problema de, dado um conjunto de **n** unidades amostrais (tratamentos, genótipos, objetos, indivíduos, entidades biológicas, etc.), as quais são medidas segundo **p** variáveis, obter um esquema que possibilite reunir as unidades em um número de grupos tal que exista homogeneidade dentro e heterogeneidade entre os grupos. Mais especificamente, as técnicas de análise de agrupamento objetivam dividir um conjunto de observação em um número de grupos homogêneos, segundo algum critério conveniente de homogeneidade.

A análise de agrupamento constitui uma metodologia numérica multivariada com o objetivo de propor uma estrutura classificatória ou do reconhecimento da existência de grupos. Do ponto de vista biológico a análise de agrupamento é extremamente fascinante, pois seu resultado final é um gráfico em forma de árvore denominado dendrograma ou fenograma. Este representa uma síntese gráfica dos resultados o que, como em toda síntese, ocasiona uma perda de informação. Mas, mesmo com esta perda de informação, este gráfico nos será de grande utilidade para a classificação, comparação e discussão de agrupamentos de entidades biológicas. A aplicação desta técnica de análise tem sido intensa em áreas como Ciências Biológicas, Medicina, Geologia, Ciências Sociais, dentre outras.

As técnicas de agrupamento foram desenvolvidas por analistas de problemas aplicados, dando ênfase à solução de problemas específicos, ao contrário de outros campos da Análise Multivariada, que se acham teoricamente mais fundamentados. Na análise de agrupamento as afirmativas empíricas nem sempre têm respaldo teórico. Muitas técnicas são propostas, mas não há, ainda, uma teoria generalizada e amplamente aceita.

A maioria das técnicas de agrupamento começa com o cálculo da matriz de similaridade ou de dissimilaridade que, pode-se chamar de matriz de proximidade, entre os elementos observados. Desta forma as técnicas de agrupamento podem ser vistas como forma de resumir informações sobre os elementos que compõem a matriz de proximidade.

Um problema destas técnicas é a determinação da medida de proximidade mais conveniente, uma vez que as técnicas baseadas em diferentes medidas de proximidade nem sempre levam aos mesmos resultados. Além disso, qualquer processo de agrupamento envolve uma técnica numérica e a intensidade do agrupamento resultante é uma propriedade não só dos dados utilizados, mas também, do processo ao qual estes dados são submetidos. Por isso, seria interessante, processar o conjunto de dados por mais de uma técnica e comparar os seus resultados.

A análise de agrupamento relaciona-se com outras técnicas multivariadas. É comum, quando se tem um grande número de variáveis, reduzir a dimensão através da análise fatorial (“Factor Analysis”), análise por variáveis canônicas ou análise de componentes principais, utilizando-se os escores dos primeiros fatores, variáveis ou componentes para se proceder à análise de agrupamento. Além disso, para verificar a adequação da partição obtida, uma vez tendo-se o número de grupos e conhecidos os seus componentes, é comum a utilização da análise discriminante.

O processo de agrupamento envolve basicamente duas etapas. A primeira relaciona-se com a estimação de uma medida de similaridade ou de dissimilaridade entre os indivíduos e a segunda, com a adoção de uma técnica de agrupamento para a formação dos grupos.

6.2. MEDIDAS DE SIMILARIDADE E DE DISSIMILARIDADE

Um conceito fundamental na utilização das técnicas de Análise de Agrupamento é a escolha de um critério que meça a distância entre dois objetos, ou que quantifique o quanto eles são parecidos. Cabe observar que tecnicamente pode-se dividir em duas categorias: medidas de similaridade e de dissimilaridade. Na primeira quanto maior o valor observado mais parecidos são os objetos. Já para a segunda quanto maior for o valor observado menos parecidos (mais dissimilares ou divergentes) serão os objetos. Coeficiente de correlação é um exemplo de medida de similaridade, enquanto que a distância euclidiana é um exemplo de dissimilaridade.

A maioria dos algoritmos de Análise de Agrupamento está programado para operarem com o conceito de distância (dissimilaridade).

Neste capítulo, serão tratadas apenas as variáveis quantitativas, entretanto uma vasta opção de medidas de dissimilaridade para variáveis qualitativas binárias e multicategóricas está disponível na literatura.

Seja ${}_nX_p = [X_{ij}]$ a matriz de dados, onde X_{ij} representa o valor da j -ésima variável referente ao i -ésimo indivíduo, onde $i = 1, 2, \dots, n$ e $j = 1, 2, \dots, p$. Assim cada vetor linha representa uma unidade amostral (indivíduo, tratamento, etc.) e cada vetor coluna uma variável, como apresentado na Tabela 6.1.

Segundo Maxwell (1977), o primeiro estágio em muitos métodos de análise de agrupamento é a conversão da matriz $n \times p$ de dados em uma matriz $n \times n$ de similaridades ou de dissimilaridades individuais que são medidas da relação entre pares de indivíduos, dado o valor de um conjunto de p variáveis. Alta similaridade indica que dois indivíduos são comuns em relação ao conjunto de variáveis, enquanto que alta dissimilaridade indica o contrário.

Tabela 6.1 – Matriz de dados de n indivíduos e p variáveis

Indivíduos	Variáveis					
	X_1	X_2	...	X_j	...	X_p
1	X_{11}	X_{12}	...	X_{1j}	...	X_{1p}
2	X_{21}	X_{22}	...	X_{2j}	...	X_{2p}
...
i	X_{i1}	X_{i2}	...	X_{ij}	...	X_{ip}
...
i'	$X_{i'1}$	$X_{i'2}$...	$X_{i'j}$...	$X_{i'p}$
...
n	X_{n1}	X_{n2}	...	X_{nj}	...	X_{np}

Se os dados estão todos no mesmo padrão de medida, as variabilidades de cada variável são homogêneas ou quase homogêneas, então, podem ser usados os dados originais. Caso isto não ocorra, há necessidade de transformar esses dados. A transformação mais comum é a padronização das variáveis (pode-se usar Z ou x minúsculo), ou seja,

$$Z_{ij} = x_{ij} = \frac{X_{ij} - \bar{X}_j}{s(X_j)} \text{ ou } Z_{ij} = x_{ij} = \frac{X_{ij}}{s(X_j)}.$$

A seguir, serão apresentadas algumas medidas de dissimilaridade e as expressões matemáticas usadas na determinação dos coeficientes de distância. Estas medidas serão dadas em função das variáveis originais. Caso sejam usadas as variáveis transformadas, utilizam-se as mesmas fórmulas substituindo-se X_{ij} por Z_{ij} .

6.2.1. Medidas de dissimilaridade

Embora, na realidade, a distância seja uma medida de dissimilaridade, às vezes ela é referida como uma medida de semelhança, pois quanto maior seu valor, menos parecidos são os indivíduos (ou unidades amostrais).

Segundo Johnson e Wichern (1998), qualquer medida de distância entre dois pontos P e Q, denotada por $d(P, Q)$ é válida se as seguintes propriedades forem satisfeitas, em que R é qualquer ponto intermediário:

- i) $d(P, Q) = d(Q, P)$ (simetria);
- ii) $d(P, Q) > 0$, se $P \neq Q$;
- iii) $d(P, Q) = 0$, se e somente se $P = Q$;
- iv) $d(P, Q) \leq d(P, R) + d(R, Q)$ (desigualdade triangular).

O termo dissimilaridade apareceu em função de que à medida que $d(P, Q)$ cresce, diz-se que a divergência entre P e Q aumenta, ou seja, tornam-se cada vez mais dissimilares. As principais medidas de dissimilaridade usadas em análise de agrupamento são apresentadas a seguir.

6.2.1.1. Distância euclidiana

A distância euclidiana entre os indivíduos i e i' (Tabela 6.1) é dada por:

$$d_{ii'} = \left[\sum_{j=1}^p (X_{ij} - X_{i'j})^2 \right]^{1/2}$$

Outra forma de apresentar a distância é utilizando a linguagem matricial, dada por:

$$d_{ii'} = \left[\left(\begin{matrix} X_i \\ \sim \end{matrix} - \begin{matrix} X_{i'} \\ \sim \end{matrix} \right)' \left(\begin{matrix} X_i \\ \sim \end{matrix} - \begin{matrix} X_{i'} \\ \sim \end{matrix} \right) \right]^{1/2}$$

em que

$$\begin{matrix} X_i \\ \sim \end{matrix} = [X_{i1} \ X_{i2} \ \dots \ X_{ip}]' \text{ e } \begin{matrix} X_{i'} \\ \sim \end{matrix} = [X_{i'1} \ X_{i'2} \ \dots \ X_{i'p}]'$$

A distância euclidiana não preserva a ordem das distâncias com a mudança de escala, por isso é comum proceder-se à padronização das variáveis antes de se obter o valor da distância.

6.2.1.2. Distância euclidiana média

A distância euclidiana cresce à medida que cresce o número de caracteres (variáveis). Uma maneira razoável de contornar esse problema é dividir esse valor pela raiz quadrada do número de variáveis, obtendo-se a distância euclidiana média que é dada por:

$$\Delta_{ii'} = \frac{1}{\sqrt{p}} d_{ii'}$$

Consideremos os dados apresentados na Tabela 6.2.

Tabela 6.2 - Resultados experimentais

Indivíduo	Altura (X_1)	Peso (X_2)	Idade (X_3)
A	180	79	30
B	175	75	25
C	170	70	28
D	167	63	21
E	180	71	18
F	165	60	28
Média	172,8	69,7	25,0
Desvio padrão	6,5	7,1	4,6

Usando as variáveis altura e peso, a distância euclidiana entre A e B é:

$$d_{AB} = d(A, B) = [(180 - 175)^2 + (79 - 75)^2]^{1/2} = (41)^{1/2} = 6,40.$$

Usando como critério as duas variáveis anteriores mais a idade, a distância euclidiana passa a ser:

$$d_{AB} = d(A, B) = [(180 - 175)^2 + (79 - 75)^2 + (30 - 25)^2]^{1/2} = (66)^{1/2} = 8,12.$$

Considerando as variáveis altura e peso, a distância euclidiana média entre A e B é:

$$\Delta_{AB} = \frac{1}{\sqrt{2}} (41)^{1/2} = \left(\frac{41}{2} \right)^{1/2} = 4,53.$$

Considerando as variáveis altura, peso e idade, a distância euclidiana média passa a ser:

$$\Delta_{AB} = \frac{1}{\sqrt{3}} (66)^{1/2} = \left(\frac{66}{3} \right)^{1/2} = 4,69.$$

Observe que neste último coeficiente (distância euclidiana média) embora aumente o número de coordenadas a magnitude dos coeficientes são comparáveis.

6.2.1.3. Distância ponderada

Neste caso, considera-se, no cálculo da distância euclidiana, a diferença de precisão de cada variável mensurada, de forma que aquelas que foram medidas com menor precisão venham contribuir menos para o valor da diversidade entre pares de acessos. Para o cálculo dessa distância, é necessário dispor da informação da variação residual, obtida em análise prévia de variância em modelos estatísticos apropriados, de forma que a distância possa ser calculada da seguinte maneira:

$$d_{ii'}^2 = \left[\sum_{j=1}^p \frac{(X_{ij} - X_{i'j})^2}{\hat{\sigma}_j^2} \right]$$

ou

$$d_{ii'}^2 = \left(\bar{\tilde{X}}_i - \bar{\tilde{X}}_{i'} \right)' \Delta^{-1} \left(\bar{\tilde{X}}_i - \bar{\tilde{X}}_{i'} \right)$$

em que:

$d_{ii'}^2$: quadrado da distância ponderada entre os indivíduos i e i' ;

Δ : matriz diagonal cujos elementos são as variâncias residuais;

$\hat{\sigma}_j^2$: quadrado médio do resíduo associado à j -ésima variável;

e

X_{ij} : média do i -ésimo indivíduo em relação à j -ésima variável.

6.2.1.4. Distância de Gower

Esta distância é calculada por meio de:

$$d_{ii'} = \sum_{j=1}^p \frac{|X_{ij} - X_{i'j}|}{R_j}$$

em que:

R_j : amplitude de variação verificada na j -ésima variável.

Esta medida de distância tem, como vantagem, proporcionar valores de distância que variam no intervalo de 0 a 1.

6.2.1.5. Distância de Mahalanobis

A distância de Mahalanobis entre as unidades amostrais (tratamento, indivíduos, etc.) i e i' , é dada por:

$$D_{ii'}^2 = \left(\bar{\tilde{X}}_i - \bar{\tilde{X}}_{i'} \right)' S^{-1} \left(\bar{\tilde{X}}_i - \bar{\tilde{X}}_{i'} \right)$$

em que, \mathbf{S} é a matriz de dispersão amostral comum a todas as unidades. No caso de delineamentos experimentais $\mathbf{S} = \frac{\mathbf{E}}{n_e}$, isto é, \mathbf{S} é a matriz de covariância residual, $\bar{\mathbf{X}}_i$ e $\bar{\mathbf{X}}_{i'}$ são os vetores p -dimensionais de médias i e i' , respectivamente, com $i \neq i'$ e $i, i' = 1, 2, \dots, n$.

Embora na realidade, $D_{ii'}^2$ seja o quadrado da distância de Mahalanobis, ela será chamada simplesmente de distância de Mahalanobis. Esta medida é muito importante quando existem repetições dentro das unidades amostrais, isto é, quando a unidade amostral é um conjunto de indivíduos e, principalmente quando os caracteres são correlacionados.

Tendo-se repetições, pode-se trabalhar com a média e utilizar a distância euclidiana média como medida de dissimilaridade. Entretanto, este procedimento deve ser evitado, por ser menos preciso que aquele cuja medida de dissimilaridade baseia-se na distância de Mahalanobis. Quando as variáveis são correlacionadas, a distância D^2 apresenta vantagem em relação à distância euclidiana por levar em consideração a correlação residual entre as variáveis, e é equivalente a ela para o caso de correlações nulas entre as variáveis, considerando-se as variáveis padronizadas.

A distância de Mahalanobis considera a variabilidade dentro de cada unidade amostral, e não somente a medida de tendência central, sendo, portanto, uma medida mais aceitável quando as unidades amostrais constituem um conjunto de indivíduos. Supondo distribuição multinormal p -dimensional e homogeneidade das matrizes de covariâncias das unidades amostrais, $D_{ii'}^2$ é denominada distância generalizada de Mahalanobis.

6.2.2. Medidas de similaridade

Uma medida de similaridade entre dois objetos P e Q , denotada por $s(P, Q)$, deve satisfazer as seguintes propriedades:

- i) $s(P, Q) = s(Q, P)$;
- ii) $|s(P, Q)| \geq 0$; e
- iii) $s(P, Q)$ cresce à medida que a semelhança entre P e Q cresce.

A medida de similaridade e a medida de dissimilaridade caminham em sentidos opostos, pois enquanto a primeira cresce à medida que aumenta a semelhança entre os dois objetos em comparação, a segunda decresce.

Existe uma gama muito grande de coeficientes de similaridade, e dentre os mais empregados está o coeficiente de correlação Momento-Produto de Pearson, que para os indivíduos i e i' é definido por:

$$r_{ii'} = \frac{\sum_j X_{ij} X_{i'j} - \frac{1}{p} \left(\sum_j X_{ij} \right) \left(\sum_j X_{i'j} \right)}{\sqrt{\left[\sum_j X_{ij}^2 - \frac{1}{p} \left(\sum_j X_{ij} \right)^2 \right] \left[\sum_j X_{i'j}^2 - \frac{1}{p} \left(\sum_j X_{i'j} \right)^2 \right]}}$$

com $-1 \leq r_{ii'} \leq 1$.

6.3. MÉTODOS DE AGRUPAMENTO

Como no processo de agrupamento é desejável ter informações relativas a cada par de indivíduos (tratamentos), o número de estimativas de medidas de dissimilaridade é relativamente grande, o que torna impraticável o reconhecimento de grupos homogêneos pelo simples exame visual daquelas estimativas. Para realizar esta tarefa, faz-se uso dos métodos de agrupamento.

Existe grande número de métodos de agrupamento disponível, dos quais o pesquisador tem que decidir qual o mais adequado ao seu trabalho, uma vez que as variadas técnicas podem levar a diferentes padrões de agrupamentos. De acordo com Everitt (1980), os métodos de agrupamento podem ser classificados em cinco grupos, dos quais os mais importantes são:

- i) Métodos Hierárquicos: os indivíduos são reunidos em grupos e o processo repete-se em diferentes níveis até formar uma árvore.
- ii) Métodos de Otimização: os grupos são formados pela otimização de um critério de agrupamento. Os grupos são mutuamente exclusivos, formando uma partição do conjunto de entidades.

6.3.1. Métodos hierárquicos

Nos métodos hierárquicos os indivíduos são classificados em grupos em diferentes etapas, de modo hierárquico, produzindo uma árvore de classificação. Os métodos hierárquicos podem ainda ser subdivididos em dois tipos: **aglomerativos**, onde através de fusões sucessivas dos n indivíduos, vão sendo obtidos $n-1$, $n-2$, etc. grupos, até reunir todos os indivíduos num único grupo; **divisivos**, partem de um único grupo e por divisões sucessivas vão sendo obtidos

2, 3, etc. grupos. O que caracteriza estes processos é que a reunião de dois agrupamentos numa certa etapa produz um dos agrupamentos da etapa superior, caracterizando o processo hierárquico.

O procedimento básico de todos os métodos hierárquicos aglomerativos de agrupamento é similar. Eles iniciam com o cálculo de uma matriz de proximidade entre as entidades e finalizam com um dendrograma (gráfico em forma de árvore), mostrando as fusões sucessivas dos indivíduos, o que culmina no estágio onde todos os indivíduos estão em um único grupo.

Abordaremos somente quatro métodos ou algoritmos de agrupamento, dentre os métodos hierárquicos, ou seja:

- i) Método do vizinho mais próximo ou método da ligação simples (“Single Linkage Method”);
- ii) Método do vizinho mais distante ou método da ligação completa (“Complete Linkage Method”);
- iii) Método da ligação média entre grupos ou UPGMA (“Unweighted Pair Group Method with Arithmetic Average”);
- iv) Método da variância mínima ou método de Ward.

6.3.1.1. Método do vizinho mais próximo

Segundo Sokal e Sneath (1963) o método do vizinho mais próximo foi introduzido em Taxonomia Numérica por Florek et al. em 1951. Nesse método, grupos inicialmente constituídos cada um de um indivíduo simplesmente (n grupos), são reunidos de acordo com a proximidade dos elementos, e então os indivíduos mais próximos são fundidos. Um indivíduo candidato a um agrupamento apresenta uma distância a este agrupamento igual à sua menor distância com relação aos membros do agrupamento. Ou seja, o coeficiente de distância entre dois agrupamentos J e K será dado por:

$$d_{J,K} = \min_{\substack{j \in J \\ k \in K}} d_{jk}.$$

Assim, a distância entre um indivíduo k e um grupo, formado pelos indivíduos i e j , dada por:

$$d_{(ij)k} = \min \{d_{ik}; d_{jk}\}$$

A distância entre dois grupos é dada por:

$$d_{(ij)(kl)} = \min\{d_{ik}; d_{il}; d_{jk}; d_{jl}\}$$

6.3.1.2. Método do vizinho mais distante

Segundo Sokal e Sneath (1973) o método do vizinho mais distante foi introduzido por Sorensen em 1948. Esse método é exatamente o oposto do método do vizinho mais próximo, uma vez que a distância entre grupos é agora definida como a distância entre os pares de indivíduos mais distantes, ou seja, o coeficiente de distância entre dois agrupamentos J e K será dado por:

$$d_{J,K} = \max_{\substack{j \in J \\ k \in K}} d_{jk}$$

Assim, a distância entre um indivíduo k e um grupo, formado pelos indivíduos i e j, dada por:

$$d_{(ij)k} = \max\{d_{ik}; d_{jk}\}$$

A distância entre dois grupos é dada por:

$$d_{(ij)(kl)} = \max\{d_{ik}; d_{il}; d_{jk}; d_{jl}\}$$

6.3.1.3. Método da ligação média entre grupos

Este método não ponderado de agrupamento aos pares usando as médias aritméticas, mais conhecido como UPGMA, é a técnica de agrupamento mais usada. Seu desenvolvimento foi devido a Sokal e Michener (1958). O método tem sido utilizado com maior frequência em ecologia e sistemática (JAMES; McCULLOCH, 1990) e em taxionomia numérica (SNEATH; SOKAL, 1973). Nesse caso, a distância entre dois agrupamentos J e K será definida como:

$$d_{J,K} = \frac{1}{n_J \cdot n_K} \sum_{\substack{j \in J \\ k \in K}} d_{jk}$$

Assim, a distância entre um indivíduo k e um grupo formado pelos indivíduos i e j é dada por:

$$d_{(ij)k} = \text{média}(d_{ik}; d_{jk}) = \frac{d_{ik} + d_{jk}}{2}$$

A distância entre dois grupos é fornecida por:

$$d_{(ij)(klm)} = \text{média}(d_{ik}; d_{il}; d_{im}; d_{jk}; d_{jl}; d_{jm}) = \frac{d_{ik} + d_{il} + d_{im} + d_{jk} + d_{jl} + d_{jm}}{6}$$

Exemplo 1

Apenas a título de ilustração, vamos considerar uma única variável cujos dados estão apresentados a seguir:

Indivíduos	Variável (X)
1	2
2	7
3	12
4	9
5	1

- i) Vamos obter o dendrograma considerando o método do vizinho mais próximo como algoritmo de agrupamento e a distância euclidiana como medida de dissimilaridade.

Usando a distância euclidiana para cálculo da distância entre os elementos, obtém-se a matriz de distância D_1 , dada por:

$$D_1 = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} & \left[\begin{array}{ccccc} 0 & 5 & 10 & 7 & 1 \\ & 0 & 5 & 2 & 6 \\ & & 0 & 3 & 11 \\ & (sim.) & & 0 & 8 \\ & & & & 0 \end{array} \right] \end{matrix}$$

Como $d_{15} = 1$ é a menor distância em D_1 , os indivíduos 1 e 5 são agrupados. As distâncias entre este grupo e os grupos individuais 2, 3 e 4 são obtidas de D_1 como segue:

$$d_{(15)2} = \min \{d_{12}, d_{25}\} = d_{12} = 5$$

$$d_{(15)3} = \min \{d_{13}, d_{35}\} = d_{13} = 10$$

$$d_{(15)4} = \min \{d_{14}, d_{45}\} = d_{14} = 7$$

Com isso obtém-se D_2 dada por:

$$D_2 = \begin{array}{c} (15) \\ (15) \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} (15) \quad 2 \quad 3 \quad 4 \\ \left[\begin{array}{cccc} 0 & 5 & 10 & 7 \\ & 0 & 5 & 2 \\ & (\text{sim.}) & 0 & 3 \\ & & & 0 \end{array} \right] \end{array}$$

O menor valor em D_2 é $d_{24} = 2$, então os indivíduos 2 e 4 são agrupados formando um 2º grupo, e as distâncias, analogamente ao caso anterior, são dadas por:

$$d_{(15)3} = 10$$

$$\begin{aligned} d_{(15)(24)} &= \min \{d_{(15)2}, d_{(15)4}\} \\ &= \min \{d_{12}, d_{14}, d_{25}, d_{45}\} \\ &= d_{(15)2} = d_{12} = 5 \end{aligned}$$

$$d_{(24)3} = \min \{d_{23}, d_{34}\} = d_{34} = 3$$

Assim obtém-se D_3 dada por:

$$D_3 = \begin{array}{c} (15) \quad (24) \quad 3 \\ (15) \\ (24) \\ 3 \end{array} \begin{array}{c} (15) \quad (24) \quad 3 \\ \left[\begin{array}{ccc} 0 & 5 & 10 \\ & 0 & 3 \\ & (\text{sim.}) & 0 \end{array} \right] \end{array}$$

O menor valor em D_3 é $d_{(24)3} = 3$ e, então, o indivíduo 3 é incluído no grupo de 2 e 4. No último passo agrupa-se (15) e (234), formando um grupo só.

$$\begin{aligned} d_{(15)(234)} &= \min \{d_{(15)(24)}, d_{(15)3}\} \\ &= \min \{d_{(15)2}, d_{(15)3}, d_{(15)4}\} \\ &= d_{(15)2} = 5 \end{aligned}$$

$$D_4 = \begin{array}{c} (15) \quad (234) \\ (15) \\ (234) \end{array} \begin{array}{c} (15) \quad (234) \\ \left[\begin{array}{cc} 0 & 5 \\ & (\text{sim.}) \quad 0 \end{array} \right] \end{array}$$

O dendrograma está apresentado na Figura 6.1.

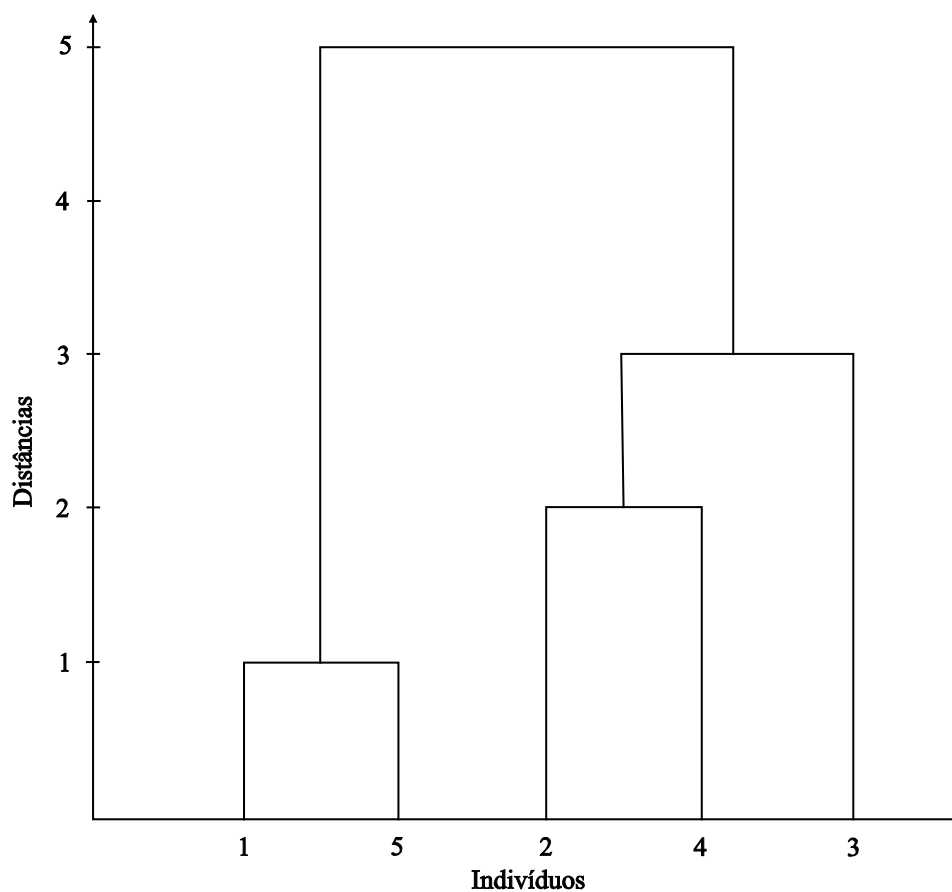


Figura 6.1 - Dendrograma obtido com o método do vizinho mais próximo

Um resumo do método do vizinho mais próximo aplicado aos dados do exemplo 1 está apresentado a seguir:

Passo	Junção	Nível
1	1,5	1
2	2,4	2
3	24,3	3
4	15,234	5

Observando a Figura 6.1, nota-se que o maior salto é observado na última etapa, sugerindo a existência de dois grupos homogêneos: (1, 5) e (2, 3, 4).

Algumas observações interessantes devem ser feitas:

a) No método do vizinho mais próximo, a distância mínima entre indivíduos está incluída no dendrograma. No exemplo, a menor distância se verifica entre os acessos 1 e 5 e vale 1,0.

- b) A distância máxima entre os indivíduos não está representada no dendrograma. O valor da distância no último nível de fusão (5) é inferior à maior medida de dissimilaridade estimada, que foi de 11, estabelecida entre os indivíduos 3 e 5.
- c) A disposição dos indivíduos na vertical não deve ser considerada como um padrão fixo e, portanto, sem significado prático.
- d) O agrupamento explora pouco a diversidade entre os indivíduos, tendo como referência de dissimilaridade total um pequeno valor da distância no último nível de fusão. Em certos casos, não é capaz de discernir grupos pobremente separados.
- e) O padrão de agrupamento geralmente é do formato de encadeamento ou do “tipo escada”. O padrão de agrupamento é contínuo, sem clara distinção de formação de grupos discretos.

- ii) Vamos obter o dendrograma considerando o método do vizinho mais distante como algoritmo de agrupamento e a distância euclidiana como medida de dissimilaridade.

A matriz D_1 é a mesma obtida em (i), isto é,

$$D_1 = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} & \begin{bmatrix} 0 & 5 & 10 & 7 & 1 \\ & 0 & 5 & 2 & 6 \\ & & 0 & 3 & 11 \\ & & & 0 & 8 \\ & & & & 0 \end{bmatrix} \end{matrix}$$

O menor valor em D_1 é $d_{15} = 1$, então os indivíduos 1 e 5 são agrupados e as distâncias agora são dadas por:

$$d_{(15)2} = \max \{d_{12}, d_{25}\} = d_{25} = 6$$

$$d_{(15)3} = \max \{d_{13}, d_{35}\} = d_{35} = 11$$

$$d_{(15)4} = \max \{d_{14}, d_{45}\} = d_{45} = 8$$

$$D_2 = \begin{matrix} & \begin{matrix} (15) & 2 & 3 & 4 \end{matrix} \\ \begin{matrix} (15) \\ 2 \\ 3 \\ 4 \end{matrix} & \begin{bmatrix} 0 & 6 & 11 & 8 \\ & 0 & 5 & 2 \\ & & 0 & 3 \\ & & & 0 \end{bmatrix} \end{matrix}$$

O menor valor em D_2 é $d_{24} = 2$, então os indivíduos 2 e 4 são agrupados e as distâncias são dadas por:

$$\begin{aligned} d_{(15)3} &= 11 \text{ (como antes)} \\ d_{(15)(24)} &= \max \{d_{(15)2}, d_{(15)4}\} \\ &= \max \{d_{12}, d_{25}, d_{14}, d_{45}\} \\ &= d_{(15)4} = d_{45} = 8 \\ d_{(24)3} &= \max \{d_{23}, d_{34}\} = d_{23} = 5 \end{aligned}$$

Assim obtém-se D_3 dada por:

$$D_3 = \begin{array}{cc} & \begin{array}{ccc} (15) & (24) & 3 \end{array} \\ \begin{array}{c} (15) \\ (24) \\ 3 \end{array} & \begin{bmatrix} 0 & 8 & 11 \\ & 0 & 5 \\ & & 0 \end{bmatrix} \end{array}$$

O menor valor em D_3 é $d_{(24)3} = 5$ e, então, o indivíduo 3 é incluído no grupo 2 e 4.

Para o último estágio tem-se que:

$$\begin{aligned} d_{(15)(234)} &= \max \{d_{(15)(24)}, d_{(15)3}\} \\ &= \max \{d_{12}, d_{14}, d_{25}, d_{45}, d_{13}, d_{35}\} \\ &= d_{(15)3} = d_{35} = 11 \end{aligned}$$

e então,

$$D_4 = \begin{array}{cc} & \begin{array}{cc} (15) & (234) \end{array} \\ \begin{array}{c} (15) \\ (234) \end{array} & \begin{bmatrix} 0 & 11 \\ \text{(sim.)} & 0 \end{bmatrix} \end{array}$$

Assim, agrupa-se (15) e (234), formando um grupo só.

O dendrograma está apresentado na Figura 6.2.

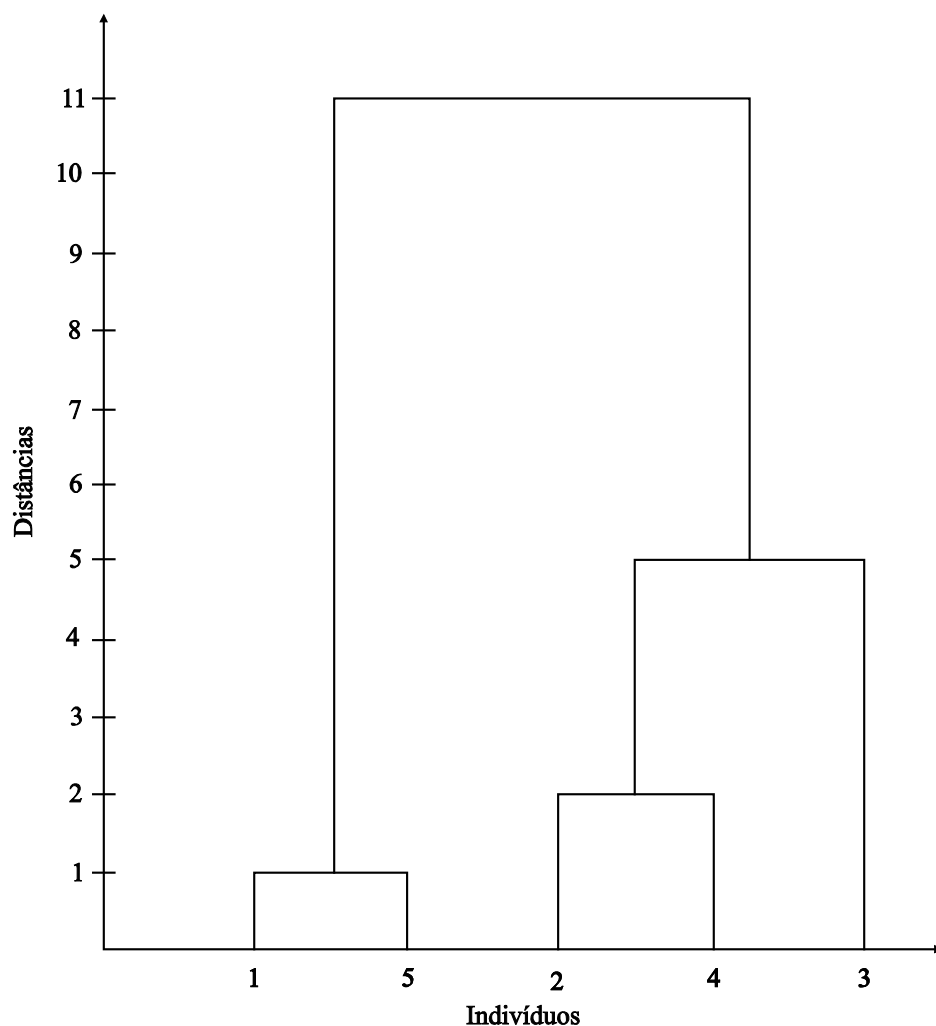


Figura 6.2 - Dendrograma obtido com o método do vizinho mais distante

Um resumo do método do vizinho mais distante aplicado aos dados do exemplo 1 está apresentado a seguir:

Passo	Junção	Nível
1	1,5	1
2	2,4	2
3	24,3	5
4	15,234	11

Algumas observações interessantes devem ser feitas:

- a) No método do vizinho mais distante, a distância mínima entre acessos está incluída no dendrograma. No exemplo, a menor distância se verifica entre os acessos 1 e 5 e vale 1,0.
- b) A distância máxima entre os indivíduos também está explicitamente representada no dendrograma. O valor da distância no último nível de fusão (11,0) é exatamente a maior medida de dissimilaridade estimada, que foi entre os acessos 3 e 5.
- c) A disposição dos acessos na vertical não deve ser considerada um padrão fixo e, portanto, sem significado prático. Nenhum método de agrupamento hierárquico tem por objetivo estabelecer projeção linear dos indivíduos quanto às suas dissimilaridades. A disposição dos acessos é decorrente de um arranjo gráfico apropriado, com vista à melhor estética.
- d) O agrupamento explora adequadamente a diversidade entre os indivíduos, tendo como referência de dissimilaridade total, expressa no último nível de fusão, o valor máximo da distância entre acessos. Há formação de grupos compactos e discretos.

iii) Vamos obter o dendrograma considerando o método UPGMA como algoritmo de agrupamento e a distância euclidiana como medida de dissimilaridade.

A matriz D_1 é a mesma obtida em (i), isto é,

$$D_1 = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} & \begin{bmatrix} 0 & 5 & 10 & 7 & 1 \\ & 0 & 5 & 2 & 6 \\ & & 0 & 3 & 11 \\ & & & 0 & 8 \\ & & & & 0 \end{bmatrix} \end{matrix}$$

O menor valor em D_1 é $d_{15} = 1$, então os indivíduos 1 e 5 são agrupados e as distâncias agora são dadas por:

$$d_{(15)2} = \text{média} \{d_{12}, d_{25}\} = (5+6)/2 = 5,5$$

$$d_{(15)3} = \text{média} \{d_{13}, d_{35}\} = (10+11)/2 = 10,5$$

$$d_{(15)4} = \text{média} \{d_{14}, d_{45}\} = (7+8)/2 = 7,5$$

$$D_2 = \begin{matrix} & \begin{matrix} (15) & 2 & 3 & 4 \end{matrix} \\ \begin{matrix} (15) \\ 2 \\ 3 \\ 4 \end{matrix} & \begin{bmatrix} 0 & 5,5 & 10,5 & 7,5 \\ & 0 & 5 & 2 \\ & & 0 & 3 \\ & & & 0 \end{bmatrix} \end{matrix}$$

O menor valor em D_2 é $d_{24} = 2$, então os indivíduos 2 e 4 são agrupados e as distâncias são dadas por:

$$d_{(15)3} = 10,5 \text{ (como antes)}$$

$$d_{(15)(24)} = \text{média } \{d_{12}, d_{14}, d_{25}, d_{45}\} = (5 + 7 + 6 + 8)/4 = 6,5$$

$$d_{(24)3} = \text{média } \{d_{23}, d_{34}\} = (5 + 3)/2 = 4$$

Assim obtém-se D_3 dada por:

$$D_3 = \begin{array}{c} \begin{array}{cc} & \begin{array}{ccc} (15) & (24) & 3 \end{array} \\ \begin{array}{c} (15) \\ (24) \\ 3 \end{array} & \left[\begin{array}{ccc} 0 & 6,5 & 10,5 \\ & 0 & 4 \\ & & 0 \end{array} \right] \end{array}$$

O menor valor em D_3 é $d_{(24)3} = 4$ e, então, o indivíduo 3 é incluído no grupo 2 e 4.

Para o último estágio tem-se que:

$$\begin{aligned} d_{(15)(234)} &= \text{média } \{d_{12}, d_{13}, d_{14}, d_{25}, d_{35}, d_{45}\} \\ &= (5 + 10 + 7 + 6 + 11 + 8)/6 = 7,8 \end{aligned}$$

e então,

$$D_4 = \begin{array}{c} \begin{array}{cc} & \begin{array}{cc} (15) & (234) \end{array} \\ \begin{array}{c} (15) \\ (234) \end{array} & \left[\begin{array}{cc} 0 & 7,8 \\ \text{(sim.)} & 0 \end{array} \right] \end{array}$$

Assim, agrupa-se (15) e (234), formando um grupo só.

O dendrograma está apresentado na Figura 6.3.

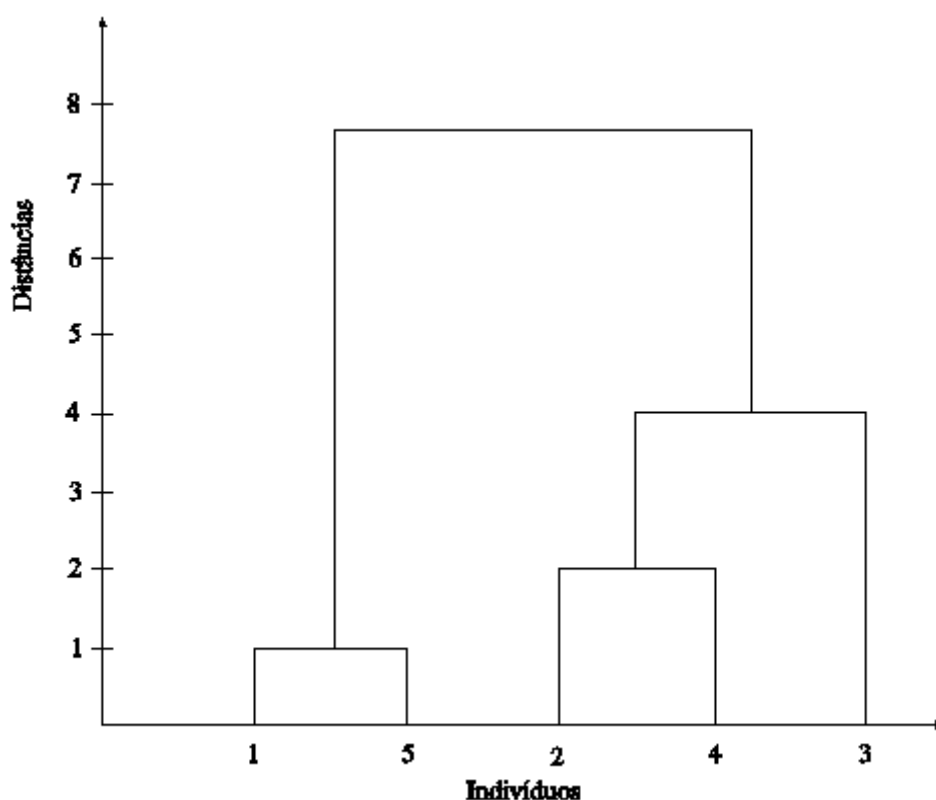


Figura 6.3 - Dendrograma obtido com o método UPGMA

Um resumo do método UPGMA aplicado aos dados do exemplo 1 está apresentado a seguir:

Passo	Junção	Nível
1	1,5	1
2	2,4	2
3	24,3	4
4	15,234	7,8

Algumas observações interessantes devem ser feitas:

- No método UPGMA, a distância mínima entre os indivíduos está incluída no dendrograma. No exemplo, a menor distância se verifica entre os acessos 1 e 5 e vale 1,0.
- A distância máxima entre os indivíduos também não está representada no dendrograma. O valor da distância no último nível de fusão (7,8) está abaixo da maior medida de dissimilaridade

estimada, que foi entre os indivíduos 3 e 5, e acima do nível de fusão estabelecido pelo método do vizinho mais próximo (5,0). Assim, nesta metodologia evita-se caracterizar a dissimilaridade pelos valores extremos (máximo ou mínimo) de dissimilaridade entre indivíduos.

c) A disposição dos indivíduos na vertical não deve ser considerada um padrão fixo e, portanto, sem significado prático. A disposição dos indivíduos é decorrente de um arranjo gráfico apropriado, com vista à melhor estética.

d) O agrupamento explora adequadamente a diversidade entre os indivíduos, porém adota, como referência de dissimilaridade total, um valor inferior à distância máxima.

6.3.1. 4. Método da variância mínima ou método de Ward

Ward (1963) usa a soma de quadrados dos desvios como a função objetivo (FO). Os valores da função objetivo mostram o erro associado a cada passo do agrupamento. Neste método, consideram-se, para a formação inicial do grupo, aqueles indivíduos que proporcionam a menor soma de quadrados dos desvios. Admite-se que, em qualquer estágio, há perda de informações em razão do agrupamento realizado, o qual pode ser quantificado pela razão entre a soma de quadrados dos desvios dentro do grupo em formação e a soma de quadrados total dos desvios. A soma de quadrados dos desvios dentro é calculada considerando apenas os indivíduos dentro do grupo em formação, e a soma de quadrados dos desvios total é calculada considerando todos os indivíduos disponíveis para análise de agrupamento.

O agrupamento é feito a partir das somas de quadrados dos desvios entre indivíduos ou, alternativamente, a partir do quadrado da distância euclidiana, uma vez que se verifica a relação:

$$SQD_{ii'} = \frac{1}{2}d_{ii'}^2$$

em que:

$$SQD_{ii'} = \sum_{j=1}^p SQD_{j(ii')}$$

sendo $SQD_{j(ii')}$ a soma de quadrados dos desvios, para a j -ésima variável, considerando os indivíduos i e i' ; e

$$d_{ii'}^2 = \sum_{j=1}^p (X_{ij} - X_{i'j})^2$$

em que:

$d_{ii'}^2$: quadrado da distância euclidiana entre os indivíduos i e i' ;

p : número de variáveis avaliadas;

X_{ij} : valor da variável j para o indivíduo i .

A soma de quadrados dos desvios total é dada por:

$$SQD_{Total} = \frac{1}{n} \sum_{i < j}^n \sum_{i'}^n d_{ii'}^2$$

sendo n o número de indivíduos a serem agrupados.

Nesta análise de agrupamento, identifica-se na matriz D (cujos elementos são os quadrados das distâncias euclidianas - $d_{ii'}^2$) ou na matriz S (cujos elementos são as somas dos quadrados dos desvios - $SQD_{ii'}$) o par de indivíduos que proporciona menor soma de quadrados dos desvios. Com estes indivíduos agrupados, uma nova matriz de dissimilaridade, de dimensão inferior, é recalculada, considerando que:

$$SQD_{(ijk)} = \frac{1}{k} d_{(ijk)}^2 \quad (k \text{ é o número de indivíduos no grupo, que, neste caso, é igual a 3}).$$

$$d_{(ijk)}^2 = d_{(ij)}^2 + d_{(ij)k}^2 = d_{ij}^2 + d_{ik}^2 + d_{jk}^2$$

e ainda que:

$$SQD_{(ijkm)} = \frac{1}{k} d_{(ijkm)}^2 \quad (k \text{ é o número de indivíduos no grupo, que, neste caso, é igual a 4}).$$

$$d_{(ijkm)}^2 = d_{ij}^2 + d_{ik}^2 + d_{jk}^2 + d_{im}^2 + d_{jm}^2 + d_{km}^2$$

e assim sucessivamente.

No procedimento, realiza-se a análise de agrupamento fornecendo os $n-1$ passos de agrupamento para que seja formado o dendrograma. Pelo programa SAS (2004) o dendrograma é construído utilizando o R^2 denominado R^2 semi parcial que será definido mais adiante.

O valor da função objetivo no passo i [$FO(i)$] é dado pela soma de quadrados dos desvios acumulada. No último passo ($i = u$), $FO(u)$ é a soma de quadrados dos desvios total.

$$\text{O } R^2 \text{ semi parcial no passo } i \text{ é dado por: } R_{sp(i)}^2 = \frac{\Delta(i)}{FO(u)},$$

em que: $\Delta(i) = FO(i) - FO(i-1)$. Os valores de $\Delta(i)$ indicam a mudança no termo erro a cada junção. $R_{sp(i)}^2$ é a diminuição na proporção da variância contabilizada resultante da união de dois grupos.

$$\text{A perda de informação no passo } i \text{ é dada por: } PI_i = \frac{FO(i)}{FO(u)}.$$

A proporção da variância considerada pelos grupos no passo i é dada por:

$$R_i^2 = 1 - \frac{FO(i)}{FO(u)} = 1 - PI_i, \text{ ou ainda } R_i^2 = 1 - \sum_{j=1}^i R_{sp(j)}^2$$

Ilustração

Será considerado o agrupamento de cinco indivíduos avaliados em relação a duas variáveis (X_1 e X_2), conforme Tabela 6.3.

Tabela 6.3 - Valores para as variáveis hipotéticas X_1 e X_2 , avaliadas em cinco indivíduos

Indivíduos	X_1	X_2
1	1	2
2	2	4
3	7	3
4	9	5
5	12	7
Total	31	21

A soma de quadrados dos desvios, para cada variável, é dada por:

$$SQD_1 = \sum_{i=1}^g X_{i1}^2 - \frac{\left(\sum_{i=1}^g X_{i1}\right)^2}{g} = (1^2 + \dots + 12^2) - \frac{31^2}{5} = 86,80$$

$$SQD_2 = \sum_{i=1}^g X_{i2}^2 - \frac{\left(\sum_{i=1}^g X_{i2}\right)^2}{g} = (2^2 + \dots + 7^2) - \frac{21^2}{5} = 14,80$$

As dissimilaridades, expressas pelos quadrados das distâncias euclidianas, são dadas por:

Dissimilaridade ($d_{ii'}^2$)	Apenas X_1	Apenas X_2	Total
12	$(1-2)^2 = 1$	$(2-4)^2 = 4$	5
13	$(1-7)^2 = 36$	$(2-3)^2 = 1$	37
14	$(1-9)^2 = 64$	$(2-5)^2 = 9$	73
15	$(1-12)^2 = 121$	$(2-7)^2 = 25$	146
23	$(2-7)^2 = 25$	$(4-3)^2 = 1$	26
24	$(2-9)^2 = 49$	$(4-5)^2 = 1$	50
25	$(2-12)^2 = 100$	$(4-7)^2 = 9$	109
34	$(7-9)^2 = 4$	$(3-5)^2 = 4$	8
35	$(7-12)^2 = 25$	$(3-7)^2 = 16$	41
45	$(9-12)^2 = 9$	$(5-7)^2 = 4$	13
Total	434	74	508

Verifica-se que:

$$a) \text{ Para a variável } X_1, \quad SQD_1 = \frac{1}{n} \sum_{i < j} \sum_{i' < j'} d_{ii'(1)}^2 = \frac{434}{5} = 86,80$$

b) Para a variável X_2 , $SQD_2 = \frac{1}{n} \sum_{i < j}^n \sum_j^n d_{ii'(2)}^2 = \frac{74}{5} = 14,80$

c) Para o conjunto de variáveis X_1 e X_2

$$SQD \text{ Total} = FO(u) = \sum_{j=1}^p SQD_j, \text{ para o exemplo } SQD = SQD_1 + SQD_2 = 101,60$$

$$SDist = \sum_i^n \sum_{i' < i}^n \left[\sum_{j=1}^p (X_{ij} - X_{i'j})^2 \right] = 508$$

$$\text{sendo } SQD \text{ Total} = FO(u) = \frac{SDist}{n} = \frac{508}{5} = 101,60$$

Assim, a dissimilaridade entre os indivíduos, dada pela soma de quadrados dos desvios, pode ser apresentada de duas formas:

$$S = \begin{bmatrix} 0 & & & & \\ SQD_{12} & 0 & & & \\ SQD_{13} & SQD_{23} & 0 & & \\ \dots & \dots & \dots & \dots & \\ SQD_{1g} & SQD_{2g} & SQD_{3g} & \dots & 0 \end{bmatrix} \quad \text{ou } S = \frac{1}{2} D = \frac{1}{2} \begin{bmatrix} 0 & & & & \\ d_{12}^2 & 0 & & & \\ d_{13}^2 & d_{23}^2 & 0 & & \\ \dots & \dots & \dots & \dots & \\ d_{1g}^2 & d_{2g}^2 & d_{3g}^2 & \dots & 0 \end{bmatrix}$$

Para o exemplo em consideração, tem-se:

$$S = \begin{bmatrix} 0,00 & & & & \\ 2,50 & 0,00 & & & \\ 18,50 & 13,00 & 0,00 & & \\ 36,50 & 25,00 & 4,00 & 0,00 & \\ 73,00 & 54,50 & 20,50 & 6,50 & 0,00 \end{bmatrix} \quad \text{e } D = \begin{bmatrix} 0,00 & & & & \\ 5,00 & 0,00 & & & \\ 37,00 & 26,00 & 0,00 & & \\ 73,00 & 50,00 & 8,00 & 0,00 & \\ 146,00 & 109,00 & 41,00 & 13,00 & 0,00 \end{bmatrix}$$

Para realizar o agrupamento, consideram-se os seguintes passos:

Passo 1: Identifica-se na matriz D ou S o elemento de menor valor. O par de indivíduos que proporciona este menor valor é agrupado. Neste exemplo, os indivíduos mais similares são o 1 e o 2. $\text{Min } SQD_{12} = 2,50$. Tem-se, então:

$$\text{Soma de quadrados dos desvios: } SQD_{12} = \frac{1}{2} d_{12}^2 = 2,50$$

Soma de quadrados dos desvios acumulada: SQD_a

$$FO(1) = SQD_a = SQD_{12} + SQD_3 + SQD_4 + SQD_5 = 2,50 + 0 + 0 + 0 = 2,50$$

$$\Delta(1) = FO(1) - FO(0) = 2,50 - 0 = 2,50$$

Passo 2: Deve ser estimada a soma de quadrados dos desvios envolvendo os indivíduos do grupo em formação, com cada um dos demais não agrupados. Dessa forma, tem-se:

$$SQD_{(123)} = \frac{1}{3}d_{(12/3)}^2 = \frac{1}{3}[d_{12}^2 + d_{(12)3}^2] = \frac{1}{3}(d_{12}^2 + d_{13}^2 + d_{23}^2) = \frac{1}{3}(5 + 37 + 26) = 22,67$$

$$SQD_{(124)} = \frac{1}{3}d_{(12/4)}^2 = \frac{1}{3}[d_{12}^2 + d_{(12)4}^2] = \frac{1}{3}(d_{12}^2 + d_{14}^2 + d_{24}^2) = \frac{1}{3}(5 + 73 + 50) = 42,67$$

$$SQD_{(125)} = \frac{1}{3}d_{(12/5)}^2 = \frac{1}{3}[d_{12}^2 + d_{(12)5}^2] = \frac{1}{3}(d_{12}^2 + d_{15}^2 + d_{25}^2) = \frac{1}{3}(5 + 146 + 109) = 86,67$$

Deve ser ressaltado que estas somas de quadrados poderiam ser calculadas da forma convencional. Assim, por exemplo, tem-se:

$$SQD_{(123)} = \left[(1^2 + 2^2 + 7^2) - \frac{10^2}{3} \right] + \left[(2^2 + 4^2 + 3^2) - \frac{9^2}{3} \right] = 20,67 + 2,00 = 22,67$$

Tem-se uma nova matriz de dissimilaridade, dada por:

$$S = \begin{matrix} & \begin{matrix} (1,2) & 3 & 4 & 5 \end{matrix} \\ \begin{matrix} (1,2) \\ 3 \\ 4 \\ 5 \end{matrix} & \begin{bmatrix} 0,00 & & & \\ 22,67 & 0,00 & & \\ 42,67 & 4,00 & 0,00 & \\ 86,67 & 20,50 & 6,50 & 0,00 \end{bmatrix} \end{matrix}$$

Verifica-se agora que os indivíduos mais similares são o 3 e 4, tendo-se:

$$\text{Soma de quadrados dos desvios: Min } SQD_{34} = \frac{1}{2}d_{34}^2 = 4,00$$

$$FO(2) = SQDa = SQD_{12} + SQD_{34} + SQD_5 = 2,50 + 4,00 + 0 = 6,50$$

$$\Delta(2) = FO(2) - FO(1) = 6,50 - 2,50 - 0 = 4,00$$

Passo 3: Deve ser estimada a soma de quadrados do desvio envolvendo os indivíduos e grupos ou entre dois grupos formados. Assim, tem-se:

$$SQD_{(1234)} = \frac{1}{4}d_{(12/34)}^2 = \frac{1}{4}(d_{12}^2 + d_{13}^2 + d_{14}^2 + d_{23}^2 + d_{24}^2 + d_{34}^2) = \frac{1}{4}(5 + 37 + 73 + 26 + 50 + 8) = 49,75$$

$$SQD_{(345)} = \frac{1}{3}d_{(34/5)}^2 = \frac{1}{3}[d_{34}^2 + d_{(34)5}^2] = \frac{1}{3}(d_{34}^2 + d_{35}^2 + d_{45}^2) = \frac{1}{3}(8 + 41 + 13) = 20,67$$

Tem-se uma nova matriz de dissimilaridade, dada por:

$$S = \begin{matrix} & \begin{matrix} (12) & (34) & 5 \end{matrix} \\ \begin{matrix} (12) \\ (34) \\ 5 \end{matrix} & \begin{bmatrix} 0,00 & & \\ 49,75 & 0,00 & \\ 86,67 & 20,67 & 0,00 \end{bmatrix} \end{matrix}$$

Verifica-se agora que os mais similares são aqueles do grupo formado pelos indivíduos 3, 4 e 5, tendo-se:

Soma de quadrados dos desvios: $\text{Min SQD}_{345} = 20,67$

$\text{FO}(3) = \text{SQDa} = \text{SQD}_{12} + \text{SQD}_{345} = 2,50 + 20,67 = 23,17$

$\Delta(3) = \text{FO}(3) - \text{FO}(2) = 23,17 - 6,50 = 16,67$

Passo 4: Por fim, calcula-se a soma de quadrados dos desvios entre os dois grupos restantes, ou seja, o grupo estabelecido pelos indivíduos 1 e 2 e o grupo formado pelos indivíduos 3, 4 e 5. Assim, tem-se:

$\text{SQD}_{(12345)} = \text{SQD Total} = \text{FO}(4) = 101,60$

$\Delta(4) = \text{FO}(4) - \text{FO}(3) = 101,60 - 23,17 = 78,43$

As fusões produzidas em cada passo e as estatísticas de interesse estão descritas a seguir:

Passo i	Entidades		FO(i)	$\Delta(i)$	$R_{sp(i)}^2$	PI_i	$R_i^2 = 1 - PI_i$
	X	Y					
1	1	2	2,50	2,50	0,0246	0,0246	0,9754
2	3	4	6,50	4,00	0,0394	0,0640	0,9360
3	(3,4)	5	23,17	16,67	0,1641	0,2280	0,7720
4	(3,4,5)	(1,2)	101,60	78,43	0,7719	1,0000	0,0000

A correspondente representação gráfica é mostrada na Figura 6.4.

Método de agrupamento: Método de Ward

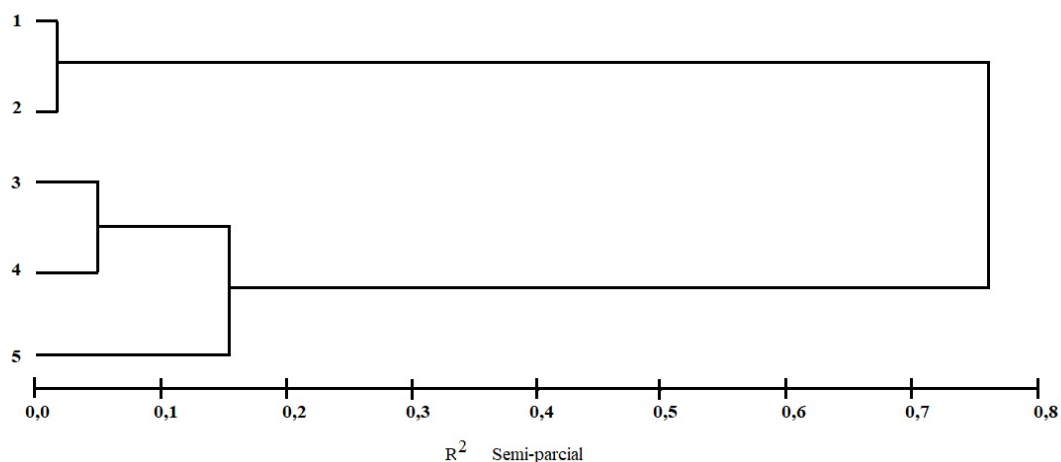


Figura 6.4 - Dendrograma obtido pelo método de Ward, a partir das medidas de dissimilaridade entre cinco indivíduos

• Comentário sobre o número de grupos

Os algoritmos vistos produzem grupos que constituem uma proposição sobre a organização básica e desconhecida dos dados. Entretanto eles esbarram em uma dificuldade que é a determinação do número ideal de grupos. Uma sugestão é o exame do dendrograma (gráfico em forma de árvore) em busca de grandes alterações dos níveis de similaridade para as sucessivas fusões. A escala vertical à esquerda, indica o nível de similaridade. No eixo horizontal são marcados os indivíduos numa ordem conveniente. As linhas verticais, partindo dos indivíduos têm altura correspondente ao nível em que os indivíduos são considerados semelhantes. A grande vantagem do dendrograma é mostrar graficamente o quanto é necessário “relaxar” o nível de similaridade para considerar grupos próximos. Observando a Figura 6.2, nota-se que o maior salto é observado na última etapa, sugerindo a existência de dois grupos homogêneos: (1, 5) e (2, 3, 4).

Existem outros procedimentos para solucionar o problema de determinar o número de grupos, um dos que será destacado é o proposto por Mojena (1977). O autor sugeriu um procedimento baseado no tamanho relativo dos níveis de fusões (distâncias) no dendrograma. A proposta é selecionar o número de grupos no estágio j que, primeiramente, satisfizer a seguinte inequação:

$$\alpha_j > \theta_k$$

em que:

α_j é o valor de distâncias dos níveis de fusão correspondentes ao estágio j ($j = 1, 2, \dots, n$); e

θ_k é o valor referencial de corte, dado por: $\theta_k = \bar{\alpha} + k\hat{\sigma}_\alpha$,

sendo $\bar{\alpha}$ e $\hat{\sigma}_\alpha$ a média e o desvio-padrão não-viesado dos valores de α , respectivamente; e k é uma constante. A adoção de valores de k em torno de 2,75 e 3,50 é sugerida por Mojena (1977). No entanto, Milligan e Cooper (1985) sugeriram o valor de $k = 1,25$ como regra de parada na definição do número de grupos.

Como ilustração, serão considerados os resultados obtidos pelo método UPGMA, dados por:

Passo	Junção	Nível (α_j)
1	1,5	1
2	2,4	2
3	24,3	4
4	15,234	7,8

Assim, tem-se:

$$\bar{\alpha} = \frac{1}{n-1} \sum_{j=1}^{n-1} \alpha_j = 3,7$$

$$\hat{\sigma}_\alpha = \sqrt{\left(\sum_{j=1}^{n-1} \alpha_j^2 - \frac{1}{n-1} \left(\sum_{j=1}^{n-1} \alpha_j \right)^2 \right) / (n-2)} = 3,0$$

Para $k = 1,25$ tem-se: $\theta_k = 3,7 + 1,25(3,0) = 7,45$

Assim, o valor α_j que supera θ_j ocorre no estágio 4, de modo que poderia ser estabelecido um grupo formado pelos indivíduos 1 e 5 e outro pelos indivíduos 2,3 e 4. A representação gráfica é mostrada na Figura 6.5.

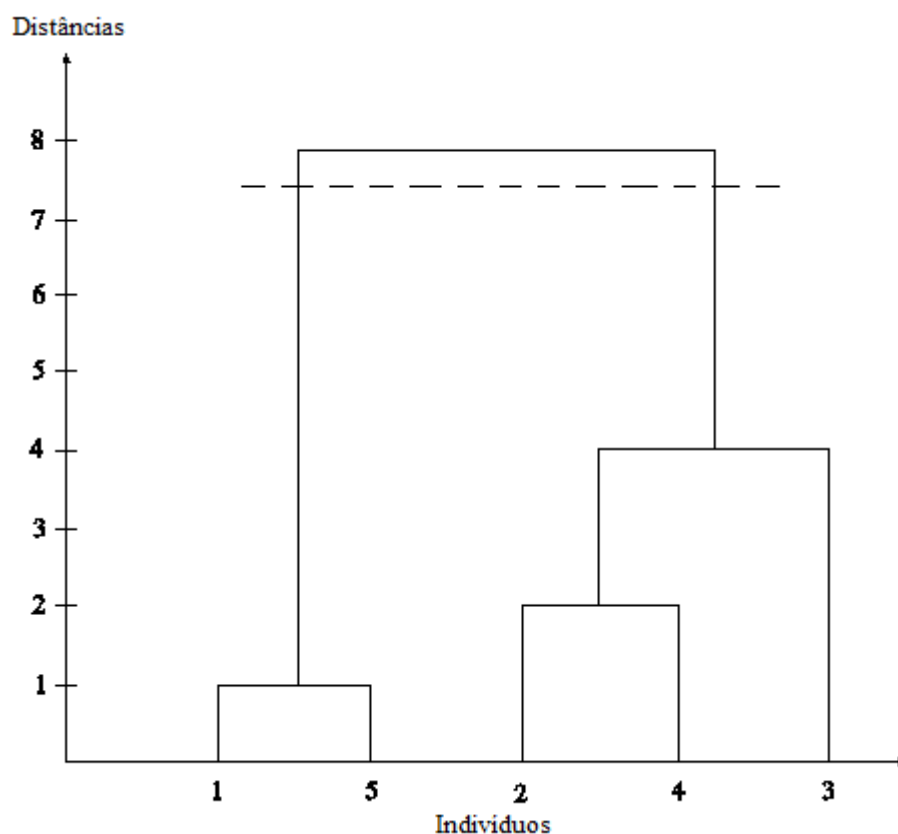


Figura 6.5 - Dendrograma obtido com o método UPGMA

No exemplo 1 apresentado anteriormente, foi considerado apenas uma variável original para facilitar o entendimento dos três algoritmos apresentados. Na prática, certamente, tem-se um maior número de variáveis e, em geral, trabalha-se com as variáveis padronizadas, principalmente quando os dados não estão todos no mesmo padrão de medida.

• Ajuste dos agrupamentos

De acordo com Rohlf (1963, 1967), pouco se sabe sobre qual técnica é mais apropriada para certo tipo de dados. Independente do método usado para resumir os dados, é importante que sejam efetuadas medidas do grau de ajuste entre a matriz original dos coeficientes de distância e a matriz resultante do processo de agrupamento (Rohlf, 1970). Quanto maior o grau de ajuste, menor será a distorção ocasionada pelo método.

Sokal e Rohlf (1962) propuseram o coeficiente de correlação cofenética (CCC) para medir o grau de ajuste entre a matriz de dissimilaridade (matriz fenética F) e a matriz resultante da simplificação proporcionada pelo método de agrupamento (matriz cofenética C). Usam-se somente os valores encontrados acima da diagonal das referidas matrizes.

O CCC é um coeficiente de correlação momento-produto calculado considerando a seguinte expressão:

$$r_{\text{cof}} = \frac{\sum_{j=1}^{n-1} \sum_{j'=j+1}^n (c_{jj'} - \bar{c})(f_{jj'} - \bar{f})}{\sqrt{\sum_{j=1}^{n-1} \sum_{j'=j+1}^n (c_{jj'} - \bar{c})^2} \sqrt{\sum_{j=1}^{n-1} \sum_{j'=j+1}^n (f_{jj'} - \bar{f})^2}}$$

em que

$$\bar{c} = \frac{2}{n(n-1)} \sum_{j=1}^{n-1} \sum_{j'=j+1}^n c_{jj'} \text{ e } \bar{f} = \frac{2}{n(n-1)} \sum_{j=1}^{n-1} \sum_{j'=j+1}^n f_{jj'}.$$

Quanto maior o valor obtido para o CCC, menor será a distorção provocada pelo agrupamento dos indivíduos. Segundo Rohlf (1970), na prática fenogramas também chamado dendrogramas com CCC menor que 0,7 indicariam a inadequação do método de agrupamento para resumir a informação do conjunto de dados. O exemplo 2 apresentado a seguir, ilustra principalmente como estimar o coeficiente de correlação cofenética.

Exemplo 2

Dados sobre medidas efetuadas no sangue de seis indivíduos estão apresentados na Tabela 6.4.

Tabela 6.4 - Valores das variáveis (características) medidas no sangue (mg)

Características	Indivíduos					
	1	2	3	4	5	6
Hematócrito	40	45	30	35	20	22
Triptofano	50	52	35	45	30	32
Creatinina	10	15	25	30	20	22
Uréia	20	18	30	28	25	24

i) Vamos obter o dendrograma considerando o método do vizinho mais próximo como algoritmo de agrupamento e a distância euclidiana média como medida de dissimilaridade.

A distância euclidiana média entre os indivíduos 1 e 2 é dada por (Notação: $d_{ii'} = \Delta_{ii'}$).

$$d_{12} = \sqrt{\frac{1}{4} [(40 - 45)^2 + (50 - 52)^2 + (10 - 15)^2 + (20 - 18)^2]} = 3,8078$$

As demais distâncias são obtidas de modo análogo e a matriz de distância D_1 (matriz fenética F) é dada por:

$$D_1 = F = \begin{bmatrix} - & 3,8078 & 12,7475 & 11,3357 & 15,2069 & 14,2126 \\ - & - & 13,7658 & 10,8857 & 17,1973 & 15,9216 \\ - & - & - & 6,2048 & 6,6143 & 5,4313 \\ - & - & - & - & 11,8215 & 10,2225 \\ - & - & - & - & - & 1,8027 \\ - & - & - & - & - & - \end{bmatrix}$$

O menor valor em D_1 é $d_{56} = 1,8027$, então os indivíduos 5 e 6 são agrupados e as distâncias agora são dadas por:

$$d_{(56)1} = \min \{d_{15}, d_{16}\} = d_{16} = 14,2126$$

$$d_{(56)2} = \min \{d_{25}, d_{26}\} = d_{26} = 15,9216$$

$$d_{(56)3} = \min \{d_{35}, d_{36}\} = d_{36} = 5,4313$$

$$d_{(56)4} = \min \{d_{45}, d_{46}\} = d_{46} = 10,2225$$

Com isso obtém-se D_2 dada por:

$$D_2 = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & (56) \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ (56) \end{matrix} & \begin{bmatrix} - & 3,8078 & 12,7475 & 11,3357 & 14,2126 \\ - & - & 13,7658 & 10,8857 & 15,9216 \\ - & - & - & 6,2048 & 5,4313 \\ - & - & - & - & 10,2225 \\ - & - & - & - & - \end{bmatrix} \end{matrix} \quad (56)$$

O menor valor em D_2 é $d_{12} = 3,8078$, então os indivíduos 1 e 2 são agrupados formando um 2º grupo e as distâncias, analogamente ao caso anterior, são dadas por:

$$\begin{aligned} d_{(12)(56)} &= \min \{d_{(12)5}, d_{(12)6}\} \\ &= \min \{d_{15}, d_{25}, d_{16}, d_{26}\} \\ &= d_{16} = 14,2126 \end{aligned}$$

$$d_{(12)3} = \min \{d_{13}, d_{23}\} = d_{13} = 12,7475$$

$$d_{(12)4} = \min \{d_{14}, d_{24}\} = d_{24} = 10,8857$$

$$d_{(56)3} = \min \{d_{35}, d_{36}\} = d_{36} = 5,4313$$

$$d_{(56)4} = \min \{d_{45}, d_{46}\} = d_{46} = 10,2225$$

Assim obtém-se D_3 dada por:

$$D_3 = \begin{matrix} & \begin{matrix} (12) & 3 & 4 & (56) \end{matrix} \\ \begin{matrix} (12) \\ 3 \\ 4 \\ (56) \end{matrix} & \begin{bmatrix} - & 12,7475 & 10,8857 & 14,2126 \\ - & - & 6,2048 & 5,4313 \\ - & - & - & 10,2225 \\ - & - & - & - \end{bmatrix} \end{matrix}$$

O menor valor em D_3 é $d_{(56)3} = 5,4313$ e, então, o indivíduo 3 é incluído no grupo de 5 e 6. As distâncias são dadas por:

$$\begin{aligned} d_{(12)(356)} &= \min \{d_{(12)3}, d_{(12)(56)}\} \\ &= d_{(12)3} = 12,7475 \\ d_{(356)4} &= \min \{d_{34}, d_{45}, d_{46}\} = d_{34} = 6,2048 \end{aligned}$$

Assim obtém-se D_4 dada por:

$$D_4 = \begin{matrix} & (12) & (356) & 4 \\ \begin{matrix} (12) \\ (356) \\ 4 \end{matrix} & \begin{bmatrix} - & 12,7475 & 10,8857 \\ - & - & 6,2048 \\ - & - & - \end{bmatrix} \end{matrix}$$

O menor valor em D_4 é $d_{(356)4} = 6,2048$ e, então, o indivíduo 4 é incluído no grupo de 3, 5 e 6. No último passo agrupa-se (12) e (3456), formando um grupo só.

$$\begin{aligned} d_{(12)(3456)} &= \min \{d_{(12)(356)}, d_{(12)4}\} \\ &= d_{(12)4} = 10,8857 \end{aligned}$$

Então,

$$D_5 = \begin{matrix} & (12) & (3456) \\ \begin{matrix} (12) \\ (3456) \end{matrix} & \begin{bmatrix} - & 10,8857 \\ - & - \end{bmatrix} \end{matrix}$$

Um resumo do método do vizinho mais próximo aplicado aos dados do exemplo 2 está apresentado a seguir:

Passos	Junção	Nível	Elemento da matriz cofenética
1º	5,6	1,8027	$c_{56}=1,8027$
2º	1,2	3,8078	$c_{12}=3,8078$
3º	56,3	5,4313	$c_{35}=c_{36}=5,4313$
4º	356,4	6,2048	$c_{34}=c_{45}=c_{46}=6,2048$
5º	12,3456	10,8857	$c_{13}=c_{14}=c_{15}=c_{16}=c_{23}=c_{24}=c_{25}=c_{26}=10,8857$

O dendrograma está apresentado na Figura 6.6.

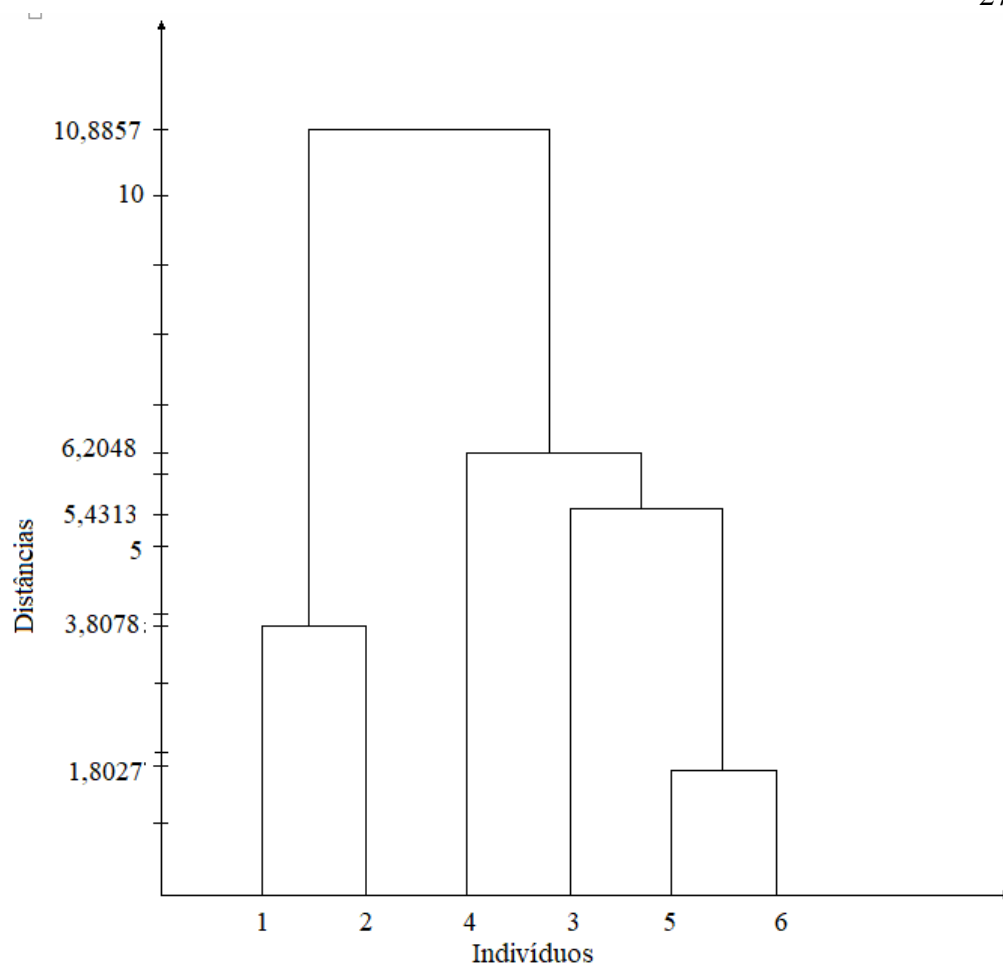


Figura 6.6 - Dendrograma obtido com o método do vizinho mais próximo

ii) Vamos obter o coeficiente de correlação cofenética e concluir

A matriz cofenética C é dada por:

$$C = \begin{bmatrix} - & 3,8078 & 10,8857 & 10,8857 & 10,8857 & 10,8857 \\ - & - & 10,8857 & 10,8857 & 10,8857 & 10,8857 \\ - & - & - & 6,2048 & 5,4313 & 5,4313 \\ - & - & - & - & 6,2048 & 6,2048 \\ - & - & - & - & - & 1,8027 \\ - & - & - & - & - & - \end{bmatrix}$$

Para o cálculo do coeficiente de correlação cofenética pode-se usar a fórmula apresentada anteriormente, mas uma maneira mais simples e equivalente de cálculo é dada a seguir:

F	C	
3,8078	3,8078	
12,7475	10,8857	
11,3357	10,8857	
15,2069	10,8857	$s_F^2 = 22,016707$
14,2126	10,8857	$s_C^2 = 10,367788$
13,7658	10,8857	$s_{FC} = 13,538655$
10,8857	10,8857	$r_{\text{cof}} = r_{FC} = \frac{\text{Cov}(F, C)}{\sqrt{\hat{V}(F) \cdot \hat{V}(C)}}$
17,1973	10,8857	
15,9216	10,8857	$r_{\text{cof}} = \frac{13,538655}{\sqrt{(22,016707)(10,367788)}}$
6,2048	6,2048	
6,6143	5,4313	$r_{\text{cof}} = 0,896$
5,4313	5,4313	
11,8215	6,2048	
10,2225	6,2048	
1,8027	1,8027	

Como $r_{\text{cof}} = 0,896 > 0,7$, concluímos que o método de agrupamento utilizado foi adequado para resumir a informação do conjunto de dados.

6.3.2 - Métodos de otimização

Nos métodos de otimização realiza-se a partição do conjunto de indivíduos (tratamentos) em subgrupos não vazios e mutuamente exclusivos por meio da maximização ou minimização de alguma medida preestabelecida. Vamos apresentar um método de otimização proposto por Tocher, citado por Rao (1952). Este método é bastante empregado no melhoramento genético.

No método de Tocher adota-se o critério de que a média das medidas de dissimilaridade dentro de cada grupo deve ser menor que as distâncias médias entre quaisquer grupos. O método requer a obtenção da matriz de dissimilaridade, sobre a qual é identificado o par de indivíduos

mais similares. Estes indivíduos formarão o grupo inicial. A partir daí é avaliada a possibilidade de inclusão de novos indivíduos, adotando-se o critério anteriormente citado.

A entrada de um indivíduo em um grupo sempre aumenta o valor médio da distância dentro do grupo. Assim, pode-se tomar a decisão de incluir o indivíduo em um grupo por meio da comparação entre o acréscimo no valor médio da distância dentro do grupo e um nível máximo permitido, que pode ser estabelecido arbitrariamente, ou adotar, como tem sido geralmente feito, o valor máximo da medida de dissimilaridade encontrado no conjunto das menores distâncias envolvendo cada indivíduo. A seguir vamos apresentar dois exemplos de aplicação.

Exemplo 3

Considere os dados da Tabela 5.12 do Capítulo 5 e as variáveis padronizadas x_1 , x_2 , x_3 mensuradas em nove tratamentos.

- i) Vamos obter o dendrograma considerando o método do vizinho mais próximo como algoritmo de agrupamento e a distância euclidiana média como medida de dissimilaridade.

Para os tratamentos 1 e 2, obtém-se a seguinte estimativa da medida de dissimilaridade (Notação: $d_{ii'} = \Delta_{ii'}$):

$$d_{12} = \sqrt{\frac{1}{3} \left[(14,4537 - 16,5949)^2 + (13,5453 - 15,3541)^2 + (41,7351 - 43,5401)^2 \right]}$$

$$d_{12} = 1,9211$$

As demais distâncias euclidianas médias, entre pares de tratamento, estão apresentadas na Tabela 6.5.

Tabela 6.5 – Distâncias euclidianas médias entre pares de nove tratamentos considerando as

variáveis padronizadas ($x_{ij} = Z_{ij} = \frac{X_{ij}}{s(X_j)}$)

Tratamentos	2	3	4	5	6	7	8	9
1	1,9211	0,9432	1,2532	0,6465	1,6692	0,8758	0,6985	1,6891
2		2,2948	0,9464	1,7498	0,8766	1,6082	2,5897	0,5959
3			1,6037	0,6323	1,7633	0,7148	1,0664	1,8612
4				0,9865	1,1288	0,9242	1,8212	0,6327
5					1,4015	0,2833	1,0757	1,3408
6						1,1851	2,3344	0,7157
7							1,3498	1,1533
8								2,3147

$$\sum_{i < i'} \sum d_{ii'}^2 = 72,00$$

Os seguintes passos são considerados:

Passo 1: Entidades mais similares: tratamentos 5 e 7

Distância entre entidades: 0,2833

Nova matriz de dissimilaridade:

	(2)	(3)	(4)	(5, 7)	(6)	(8)	(9)
(1)	1,9211	0,9432	1,2532	0,6465	1,6692	0,6985	1,6891
(2)		2,2948	0,9464	1,6082	0,8766	2,5897	0,5959
(3)			1,6037	0,6323	1,7633	1,0644	1,8612
(4)				0,9242	1,1288	1,8212	0,6327
(5, 7)					1,1851	1,0757	1,1533
(6)						2,3344	0,7157
(8)							2,3147

Ilustração: $d_{(57)1} = \min \{d_{15}; d_{17}\}$

$= \min \{0,6465; 0,8758\} = 0,6465$

etc.

Passo 2: Entidade mais similares: tratamento 2 e 9

Distância entre entidades: 0,5959

Nova matriz de dissimilaridade

	(2, 9)	(3)	(4)	(5, 7)	(6)	(8)
(1)	1,6891	0,9432	1,2532	0,6465	1,6692	0,6985
(2, 9)		1,8612	0,6327	1,1533	0,7157	2,3147
(3)			1,6037	0,6323	1,7633	1,0664
(4)				0,9242	1,1288	1,8212
(5, 7)					1,1851	1,0757
(6)						2,3344

Passo 3: Entidades mais similares: tratamento 3 e o grupo (5, 7)

Distância entre entidades: 0,6323

Nova matriz de dissimilaridade:

	(2, 9)	(3, 5, 7)	(4)	(6)	(8)
(1)	1,6891	0,6465	1,2532	1,6692	0,6985
(2, 9)		1,1533	0,6327	0,7157	2,3147
(3, 5, 7)			0,9242	1,1851	1,0644
(4)				1,1288	1,8212
(6)					2,3344

$$\text{Ilustração: } d_{(3,5,7)1} = \min \{d_{13}; d_{(5,7)1}\} \\ = \min \{0,9432; 0,6465\} = 0,6465$$

ou

$$d_{(3,5,7)1} = \min \{d_{13}; d_{15}; d_{17}\} \\ = \min \{0,9432; 0,6465; 0,8758\} = 0,6465$$

Passo 4: Entidades mais similares: tratamentos 4 e o grupo (2, 9)

Distância entre entidades: 0,6327

Nova matriz de dissimilaridade:

	(2, 4, 9)	(3, 5, 7)	(6)	(8)
(1)	1,2532	0,6465	1,6692	0,6985
(2, 4, 9)		0,9242	0,7157	1,8212
(3, 5, 7)			1,1851	1,0664
(6)				2,3344

Passo 5: Entidades mais similares: tratamento 1 e o grupo (3, 5, 7)

Distância entre entidades: 0,6465

Nova matriz de dissimilaridade:

	(2, 4, 9)	(6)	(8)
(1, 3, 5, 7)	0,9242	1,1851	0,6985
(2, 4, 9)		0,7157	1,8212
(6)			2,3344

Passo 6: Entidades mais similares: tratamento 8 e o grupo (1, 3, 5, 7)

Distância entre entidades: 0,6985

Nova matriz de dissimilaridade:

	(2, 4, 9)	(6)
(1, 3, 5, 7, 8)	0,9242	1,1851
(2, 4, 9)		0,7157

Passo 7: Entidades mais similares: tratamento 6 e o grupo (2, 4, 9)

Distância entre entidades: 0,7157

Nova matriz de dissimilaridade:

	(2, 4, 6, 9)
(1, 3, 5, 7, 8)	0,9242

Passo 8: Formação do grupo final: (1, 2, 3, 4, 5, 6, 7, 8, 9)

Distância entre entidades: 0,9242

Com base nos cálculos, é estabelecido o dendrograma apresentado na Figura 6.7. Nesta Figura, as distâncias entre os indivíduos foram convertidas em porcentagens, tomando, para esse fim, o valor obtido na formação do grupo final (igual a 0,9242, no exemplo) igual a 100%. Assim, teremos: 0,9242 (100%); 0,7157 (77,44%); 0,6985 (75,58%); 0,6465 (69,95%); 0,6327 (68,46%); 0,6323 (68,41%); 0,5959 (64,48%); 0,2833 (30,65%).

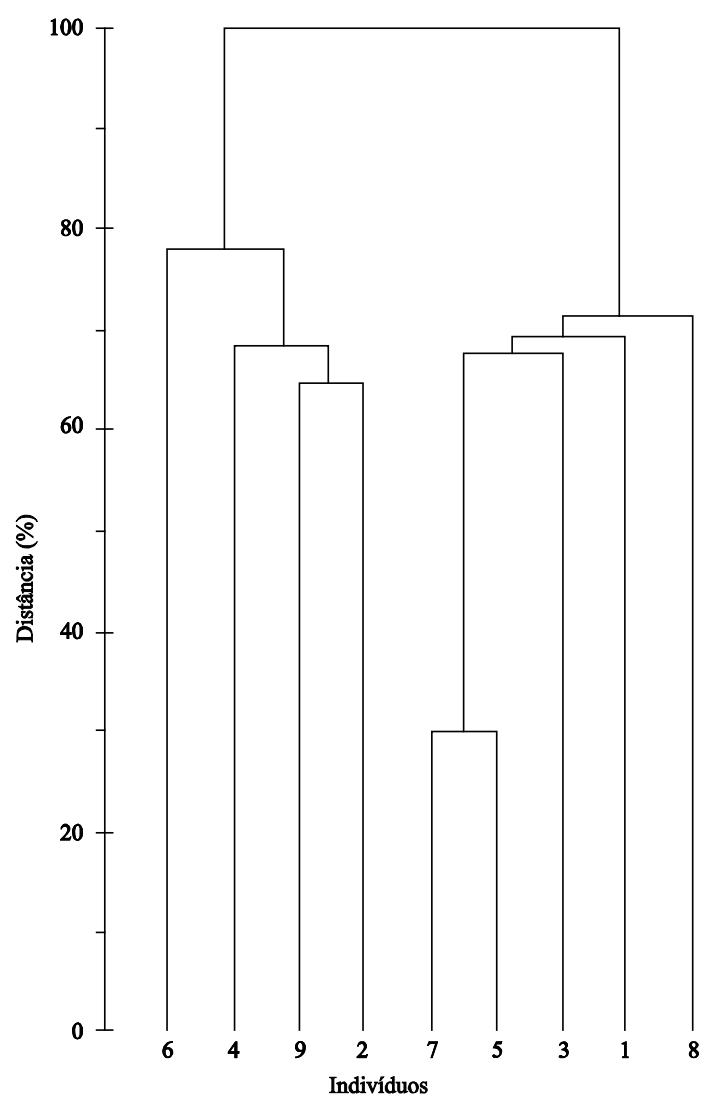


Figura 6.7 - Dendrograma ilustrativo da similaridade entre nove indivíduos, obtido pelo método do vizinho mais próximo, baseado na distância euclidiana média

Um resumo do método do vizinho mais próximo aplicado aos dados do exemplo 3 está apresentado a seguir:

Passo	Junção	Nível
1	5,7	0,2833
2	2,9	0,5959
3	57,3	0,6323
4	29,4	0,6327
5	357,1	0,6465
6	1357,8	0,6985
7	249,6	0,7157
8	13578,2469	0,9242

A análise do dendrograma (Figura 6.7) permite ao pesquisador verificar o grau de similaridade entre indivíduos, indivíduos e grupos similares, ou entre dois grupos distintos. O estabelecimento dos grupos, neste caso, é feito de forma subjetiva, tendo por base **as mudanças acentuadas de níveis**, associadas ao conhecimento prévio que o pesquisador tem dos indivíduos (tratamentos) avaliados.

Com base na Figura 6.7, pode-se pressupor, por exemplo, a existência dos grupos: (I) 5 e 7; (II) 2, 4 e 9; (III) 1, 3 e 8; (IV) 6.

- ii) Aplicar o método de Tocher para o agrupamento dos nove tratamentos, cuja medida de dissimilaridade é expressa pela distância euclidiana média calculada no item (i).

Os seguintes passos são envolvidos:

- Estimação da maior distância, dentre o conjunto de menores distâncias, entre cada tratamento.

As menores distâncias em relação a cada um dos tratamentos são dadas a seguir:

Tratamentos	1	2	3	4	5	6	7	8	9
$d_{ii'}$ mínimo	0,6465	0,5959	0,6323	0,6327	0,2833	0,7157	0,2833	0,6985	0,5959

Maior distância entre as menores: $\theta = 0,7157$

Assim é estabelecido $\theta = 0,7157$ como o limite de acréscimo, na média da distância intragrupo, para a formação ou inclusão de um novo elemento no grupo.

- **Formação do Grupo I**

O grupo I é formado inicialmente pelos tratamentos 5 e 7, cuja medida de dissimilaridade é a menor de todas. A distância entre eles é de 0,2833, menor que o mínimo estabelecido.

Avalia-se a possibilidade de inclusão de um tratamento em um grupo, verificando se a distância deste tratamento em relação ao grupo, dividida pelo número de tratamentos que já o constitui, é inferior ao máximo permitido, ou seja:

$$\text{Se } \frac{d_{(\text{Grupo}) i}}{g} \leq \theta \Rightarrow \text{inclui-se o tratamento } i \text{ no grupo.}$$

$$\text{Se } \frac{d_{(\text{Grupo}) i}}{g} > \theta \Rightarrow \text{o tratamento } i \text{ não deve ser incluído no grupo.}$$

g = número de tratamentos que constituem o grupo que está sendo considerado.

Uma vez formado o grupo, calculam-se as medidas de dissimilaridade entre este grupo e os demais tratamentos, por meio de:

$$d_{(ij)k} = d_{ik} + d_{jk}$$

em que d é qualquer medida de dissimilaridade. Assim, para o exemplo em questão, tem-se:

$$d_{(57)1} = d_{15} + d_{17} = 0,6465 + 0,8758 = 1,5223$$

$$d_{(57)2} = d_{25} + d_{27} = 1,7498 + 1,6082 = 3,3580$$

$$d_{(57)3} = d_{35} + d_{37} = 0,6323 + 0,7148 = 1,3471$$

$$d_{(57)4} = d_{45} + d_{47} = 0,9865 + 0,9242 = 1,9107$$

$$d_{(57)6} = d_{56} + d_{67} = 1,4015 + 1,1851 = 2,5866$$

$$d_{(57)8} = d_{58} + d_{78} = 1,0757 + 1,3498 = 2,4255$$

$$d_{(57)9} = d_{59} + d_{79} = 1,3408 + 1,1533 = 2,4941$$

Com base nas distâncias entre tratamentos–grupo, constata-se que o tratamento 3 é o mais similar. O valor $\frac{d_{(\text{Grupo}) i}}{g} = \frac{d_{(57)3}}{2} = \frac{1,3471}{2} = 0,6735$. Como o valor 0,6735 é inferior a θ , a inclusão do tratamento 3 no grupo I será permitida.

As novas distâncias dos tratamentos não incluídos, em relação ao grupo em formação (3, 5 e 7), são:

$$d_{(357)1} = d_{(57)1} + d_{13} = 1,5223 + 0,9432 = 2,4655$$

$$d_{(357)2} = d_{(57)2} + d_{23} = 3,3580 + 2,2948 = 5,6528$$

$$d_{(357)4} = d_{(57)4} + d_{34} = 1,9107 + 1,6037 = 3,5144$$

$$d_{(357)6} = d_{(57)6} + d_{36} = 2,5866 + 1,7633 = 4,3499$$

$$d_{(357)8} = d_{(57)8} + d_{38} = 2,4255 + 1,0664 = 3,4919$$

$$d_{(357)9} = d_{(57)9} + d_{39} = 2,4941 + 1,8612 = 4,3553$$

Agora o tratamento 1 é o mais similar ao grupo que está sendo formado. A inclusão deste tratamento é avaliada por meio de:

$$\frac{d_{(357)1}}{3} = \frac{2,4655}{3} = 0,8218.$$

Com $0,8218 > \theta$, a inclusão do tratamento 1 não é permitida.

O grupo I fica então estabelecido apenas com os tratamentos 3, 5 e 7.

- **Formação do Grupo II**

A matriz de medidas de dissimilaridade entre os tratamentos ainda não agrupados é apresentada a seguir:

	(2)	(4)	(6)	(8)	(9)
(1)	1,9211	1,2532	1,6692	0,6985	1,6891
(2)		0,9464	0,8766	2,5897	0,5959
(4)			1,1288	1,8212	0,6327
(6)				2,3344	0,7157
(8)					2,3147

Dentre os tratamentos que ainda não foram agrupados, os mais similares são o 2 e 9. A distância entre eles (0,5959) é menor que o valor máximo preestabelecido (θ) para a formação de grupos. Assim, o grupo II fica inicialmente formado com estes tratamentos.

As distâncias dos demais tratamentos em relação ao grupo II, em formação, são dadas por:

$$d_{(29)1} = d_{12} + d_{19} = 1,9211 + 1,6891 = 3,6102$$

$$d_{(29)4} = d_{24} + d_{49} = 0,9464 + 0,6327 = 1,5791$$

$$d_{(29)6} = d_{26} + d_{69} = 0,8766 + 0,7157 = 1,5923$$

$$d_{(29)8} = d_{28} + d_{89} = 2,5897 + 2,3147 = 4,9044$$

O tratamento 4 destaca-se como o mais similar em relação ao grupo II. Sua inclusão neste grupo é avaliada por:

$$\frac{d_{(29)4}}{2} = \frac{1,5791}{2} = 0,7895.$$

Como $0,7895 > \theta$, a inclusão do tratamento 4 não é permitida. Logo, o grupo II é estabelecido somente pelos tratamentos 2 e 9.

- **Formação do Grupo III**

A matriz de dissimilaridade entre os tratamentos ainda não agrupados é apresentada a seguir:

	(4)	(6)	(8)
(1)	1,2532	1,6692	0,6985
(4)		1,1288	1,8212
(6)			2,3344

Na matriz anterior, os tratamentos mais similares são o 1 e 8. A distância entre eles (0,6985) é menor que o valor máximo preestabelecido (θ) para a formação de grupos. Assim o grupo III fica inicialmente formado com estes tratamentos.

As distâncias dos demais tratamentos em relação ao grupo III, em formação, são dadas por:

$$d_{(18)4} = d_{14} + d_{48} = 1,2532 + 1,8212 = 3,0744$$

$$d_{(18)6} = d_{16} + d_{68} = 1,6692 + 2,3344 = 4,0036$$

O tratamento 4 é o mais similar em relação ao grupo III. Sua inclusão neste grupo é avaliada por:

$$\frac{d_{(18)4}}{2} = \frac{3,0744}{2} = 1,5372$$

Como $1,5372 > \theta$, a inclusão do tratamento 4 não é permitida. Logo, o grupo III incluirá apenas os tratamentos 1 e 8. Os demais tratamentos não serão mais agrupados, pois sua dissimilaridade é superior ao máximo preestabelecido para formação de grupos.

• Resultado final

O agrupamento dos tratamentos com base no método proposto por Tocher permitiu o estabelecimento dos grupos apresentados na Tabela 6.6.

Tabela 6.6 - Grupos de tratamentos estabelecidos pelo método de Tocher, com base na dissimilaridade expressa pela distância euclidiana média

Grupos	Tratamentos	Distância Média
I	5, 7 e 3	0,5435
II	2 e 9	0,5959
III	1 e 8	0,6985
IV	4	—
V	6	—

A distância média dentro do grupo é a média das distâncias entre cada par de tratamentos que o constitui. Pelo critério adotado, esta distância é sempre menor que as distâncias médias intergrupos, que são obtidas pela média das distâncias entre pares de tratamentos pertencentes aos diferentes grupos. Para o exemplo em questão, tem-se:

$$d_{I,II} = \frac{1}{6} (d_{25} + d_{27} + d_{23} + d_{59} + d_{79} + d_{39}) = 1,6680.$$

As demais distâncias médias intergrupos estão apresentadas na Tabela 6.7.

Tabela 6.7 - Distâncias médias intergrupos

Grupos	II	III	IV	V
I	1,6680	0,9929	1,1715	1,4499
II		2,1286	0,7895	0,7961
III			1,5372	2,0018
IV				1,1288

Exemplo 4

Neste exemplo vamos ver uma aplicação de Análise de Agrupamento na área de Melhoramento Genético, a partir de dados provenientes de um delineamento experimental, apresentado por CRUZ, et al. (2012).

Será estudada a divergência genética entre oito cultivares, avaliados em relação a quatro caracteres (X_1 , X_2 , X_3 e X_4), num experimento em blocos ao acaso, cujas observações experimentais são apresentadas na Tabela 6.8.

A distância euclidiana média pode ser obtida por meio das observações individuais dos progenitores, sem a necessidade de experimentos que envolvam delineamentos experimentais. Quando se dispõe de ensaios planejados, pode-se, alternativamente, ignorar o delineamento e calcular os valores $d_{ii'}$, a partir das médias das repetições do ensaio. Este último procedimento deve, entretanto, ser evitado, por ser menos preciso que aquele cuja medida de dissimilaridade baseia-se na distância generalizada de Mahalanobis.

Vamos obter o dendrograma considerando o método do vizinho mais próximo como algoritmo de agrupamento e a distância generalizada de Mahalanobis como medida de dissimilaridade. Numa segunda etapa vamos aplicar o Método de Tocher para o agrupamento dos oito cultivares (progenitores).

A distância generalizada de Mahalanobis entre os progenitores i e i' , é dada por:

$$D_{ii'}^2 = \left(\bar{\tilde{X}}_i - \bar{\tilde{X}}_{i'} \right)' S^{-1} \left(\bar{\tilde{X}}_i - \bar{\tilde{X}}_{i'} \right)$$

em que,

S é a matriz de covariância residual das unidades amostrais;

$\bar{\tilde{X}}_i$ e $\bar{\tilde{X}}_{i'}$, são os vetores p -dimensionais de médias dos progenitores i e i' ,

respectivamente, com $i \neq i'$ e $i, i' = 1, 2, \dots, n$.

Tabela 6.8 - Observações relativas a quatro caracteres (X_1 , X_2 , X_3 e X_4) avaliados em oito progenitores, em quatro blocos

Variável X_1								
Bloco/Progenitor	1	2	3	4	5	6	7	8
I	50,2	41,8	39,2	33,8	35,6	53,4	43,8	50,6
II	41,4	47,2	37,6	49,6	31,4	50,2	46,8	57,8
III	36,2	39,6	38,8	35,4	33,2	49,6	41,4	41,8
IV	39,8	46,6	33,6	41,8	29,8	57,8	43,6	46,8
Média	41,90	43,80	37,30	40,15	32,50	52,75	43,90	49,25

Variável X_2								
Bloco/Progenitor	1	2	3	4	5	6	7	8
I	20,5	19,5	19,0	20,0	20,0	19,2	19,5	19,7
II	20,6	20,1	18,5	20,3	20,8	19,5	20,4	19,8
II	20,5	19,3	18,1	20,6	20,3	20,3	20,7	20,1
IV	19,6	20,1	19,3	20,3	19,9	19,9	20,3	20,5
Média	20,30	19,75	18,725	20,30	20,25	19,725	20,225	20,025

Variável X_3								
Bloco/Progenitor	1	2	3	4	5	6	7	8
I	3,9	3,7	4,5	4,3	4,1	4,2	4,3	4,2
II	4,0	3,6	4,6	4,4	4,0	4,5	4,3	4,0
III	3,8	3,6	4,6	4,2	4,2	4,3	4,2	4,3
IV	3,9	3,7	4,7	4,3	4,1	4,5	4,3	4,1
Média	3,900	3,650	4,600	4,300	4,100	4,375	4,275	4,150

Bloco/ Progenitor	Variável X ₄							
	1	2	3	4	5	6	7	8
I	104,9	88,0	80,0	80,8	60,0	96,4	91,5	91,8
II	84,3	106,5	71,3	106,5	52,5	98,8	99,7	84,8
III	77,0	89,8	77,5	83,3	53,0	99,1	83,3	70,6
IV	76,5	108,7	69,5	95,9	51,0	107,2	89,5	81,5
Média	85,675	98,250	74,575	91,625	54,125	100,375	91,000	82,175

Para o cálculo de D^2 , supõe-se a existência de distribuição multinormal p-dimensional e a homogeneidade da matriz de covariância residual das unidades amostrais, restringindo-se, portanto, o seu uso. No entanto, considerável robustez para a violação dessas hipóteses já foi demonstrada, que faz da distância generalizada de Mahalanobis uma opção de grande utilidade, principalmente pelo fato de D^2 ter grande analogia com outras técnicas multivariadas.

Por construção, a distância generalizada de Mahalanobis corresponde à distância euclidiana apenas quando as correlações entre os caracteres forem estatisticamente nulas.

A matriz **S** pode ser obtida a partir da análise de variância multivariada apresentada na Tabela 6.9.

$$S = \frac{E}{n_e} = \begin{bmatrix} 18,7459 & 0,0151 & -0,0146 & 30,8056 \\ 0,0151 & 0,1879 & -0,0036 & 0,8517 \\ -0,0146 & -0,0036 & 0,0096 & 0,0871 \\ 30,8056 & 0,8517 & 0,0871 & 72,1533 \end{bmatrix}$$

$$S^{-1} = \begin{bmatrix} 0,2217 & 0,4724 & 1,4508 & -0,1020 \\ 0,4724 & 6,7022 & 5,8672 & -0,2879 \\ 1,4508 & 5,8672 & 116,3802 & -0,8304 \\ -0,1020 & -0,2879 & -0,8304 & 0,0618 \end{bmatrix}$$

Tabela 6.9 - Análise de variância multivariada para as variáveis X_1 , X_2 , X_3 e X_4

Causas de variação	G.L.	Matrizes de somas de quadrados e de produtos			
Blocos	3	B =	$\begin{bmatrix} 140,0976 & -1,8627 & 0,4287 & 224,0214 \\ -1,8627 & 0,6054 & 0,0453 & -4,6224 \\ 0,4287 & 0,0453 & 0,0135 & 0,7587 \\ 224,0214 & -4,6224 & 0,7587 & 370,2030 \end{bmatrix}$		
Tratamentos	7	H =	$\begin{bmatrix} 1147,5744 & 2,9001 & -0,9590 & 2042,2990 \\ 2,9001 & 7,9856 & -2,0503 & 5,1401 \\ -0,9590 & -2,0503 & 2,4136 & -18,7845 \\ 2042,2990 & 5,1401 & -18,7845 & 6249,6091 \end{bmatrix}$		
Resíduo	n _e = 21	E =	$\begin{bmatrix} 393,6639 & 0,3171 & -0,3066 & 646,9176 \\ 0,3171 & 3,9459 & -0,0756 & 17,8857 \\ -0,3066 & -0,0756 & 0,2016 & 1,8291 \\ 646,9176 & 17,8857 & 1,8291 & 1515,2193 \end{bmatrix}$		

A distância entre os cultivares 1 e 2, dada por D_{12}^2 , é estimada a partir de:

$$\bar{\tilde{X}}_1 = \begin{bmatrix} 41,900 \\ 20,300 \\ 3,900 \\ 85,675 \end{bmatrix}, \quad \bar{\tilde{X}}_2 = \begin{bmatrix} 43,800 \\ 19,750 \\ 3,650 \\ 98,250 \end{bmatrix}, \quad \bar{\tilde{X}}_1 - \bar{\tilde{X}}_2 = \begin{bmatrix} -1,900 \\ 0,550 \\ 0,250 \\ -12,575 \end{bmatrix}$$

$$\text{Assim, } D_{12}^2 = \left(\bar{\tilde{X}}_1 - \bar{\tilde{X}}_2 \right)' S^{-1} \left(\bar{\tilde{X}}_1 - \bar{\tilde{X}}_2 \right) = 23,4550.$$

De modo análogo, obtêm-se as demais estimativas de distância, cujos valores encontram-se na Tabela 6.10.

Tabela 6.10 - Distância generalizada de Mahalanobis entre pares de progenitores

Progenitores	2	3	4	5	6	7	8
1	23,4550	62,9440	17,6295	29,7385	34,5228	15,4890	29,6096
2		125,0260	56,5964	100,4269	91,2130	63,5913	97,4834
3			26,2044	47,6977	22,0013	23,6783	44,0254
4				37,4413	17,5122	3,1624	35,2571
5					40,0172	25,9657	15,6553
6						7,6556	13,7396
7							17,6554

D^2 máximo: 125,0260 (progenitores 2 e 3)

D^2 mínimo: 3,1624 (progenitores 4 e 7)

Total de $D^2 \left(\sum_{i < i'} \sum_{ii'} D_{ii'}^2 \right)$: 1125,3940

i) Método do Vizinho mais Próximo

Neste método identifica-se, na matriz de dissimilaridade, progenitores mais similares, os quais são reunidos, formando o grupo inicial. Calculam-se então as distâncias daquele grupo em relação aos demais progenitores e, nos estádios mais avançados, em relação a outros grupos já formados.

O processo de identificação das entidades (grupos ou progenitores individuais) mais similares se repete sobre a nova matriz de dissimilaridade, cuja dimensão é reduzida a cada passo e finaliza quando todos os progenitores são reunidos em um único grupo.

A distância entre um progenitor k e um grupo formado pelos progenitores i e j é dada por:

$$d_{(ij)k} = \min\{d_{ik}; d_{jk}\},$$

ou seja, $d_{(ij)k}$ é dada pelo menor elemento do conjunto das distâncias dos pares de progenitores (i e k) e (j e k).

A distância entre dois grupos é dada por:

$$d_{(ij)(kl)} = \min\{d_{ik}; d_{il}; d_{jk}; d_{jl}\},$$

ou seja, a distância entre dois grupos formados, respectivamente, pelos progenitores (i e j) e (k e l) é dada pelo menor elemento do conjunto, cujos elementos são as distâncias entre os pares de progenitores (i e k), (i e l), (j e k) e (j e l).

Aplicação:

Será aplicado o método do vizinho mais próximo para o agrupamento de oito progenitores, cujas medidas de dissimilaridade são expressas pelos valores de D^2 , dadas na Tabela 6.9.

Os seguintes passos são considerados:

Passo 1: Entidades mais similares: progenitores 4 e 7

Distância entre entidades: 3,1624

Nova matriz de dissimilaridade:

	(2)	(3)	(4, 7)	(5)	(6)	(8)
(1)	23,4550	62,9440	15,4890	29,7385	34,5228	29,6096
(2)		125,0260	56,5964	100,4269	91,2130	97,4834
(3)			23,6783	47,6977	22,0013	44,0254
(4, 7)				25,9657	7,6556	17,6554
(5)					40,0172	15,6553
(6)						13,7396

Passo 2: Entidades mais similares: progenitor 6 e o grupo (4, 7)

Distância entre entidades: 7,6556

Nova matriz de dissimilaridade:

	(2)	(3)	(4, 6, 7)	(5)	(8)
(1)	23,4550	62,9440	15,4890	29,7385	29,6096
(2)		125,0260	56,5964	100,4269	97,4834
(3)			22,0013	47,6977	44,0254
(4, 6, 7)				25,9657	13,7396
(5)					15,6553

Passo 3: Entidades mais similares: progenitor 8 e o grupo (4, 6, 7)

Distância entre entidades: 13,7396

Nova matriz de dissimilaridade:

	(2)	(3)	(4, 6, 7, 8)	(5)
(1)	23,4550	62,9440	15,4890	29,7385
(2)		125,0260	56,5964	100,4269
(3)			22,0013	47,6977
(4, 6, 7, 8)				15,6553

Passo 4: Entidades mais similares: progenitor 1 e o grupo (4, 6, 7, 8)

Distância entre entidades: 15,4890

Nova matriz de dissimilaridade:

	(2)	(3)	(5)
(1, 4, 6, 7, 8)	23,4550	22,0013	15,6553
(2)		125,0260	100,4269
(3)			47,6977

Passo 5: Entidades mais similares: progenitor 5 e o grupo (1, 4, 6, 7, 8)

Distância entre entidades: 15,6553

Nova matriz de dissimilaridade:

	(2)	(3)
(1, 4, 5, 6, 7, 8)	23,4550	22,0013
(2)		125,0260

Passo 6: Entidades mais similares: progenitor 3 e o grupo (1, 4, 5, 6, 7, 8)

Distância entre entidades: 22,0013

Nova matriz de dissimilaridade:

	(2)
(1, 3, 4, 5, 6, 7, 8)	23,4550

Passo 7: Formação do grupo final: 1, 2, 3, 4, 5, 6, 7, 8.

Distância entre entidades: 23,4550.

Com base nos cálculos, é estabelecido o dendrograma apresentado na Figura 6.8. Nesta Figura as distâncias entre os indivíduos são convertidas em percentagens, tomando, para esse fim, o valor obtido na formação do grupo final (igual a 23,4550, no exemplo) igual a 100%.

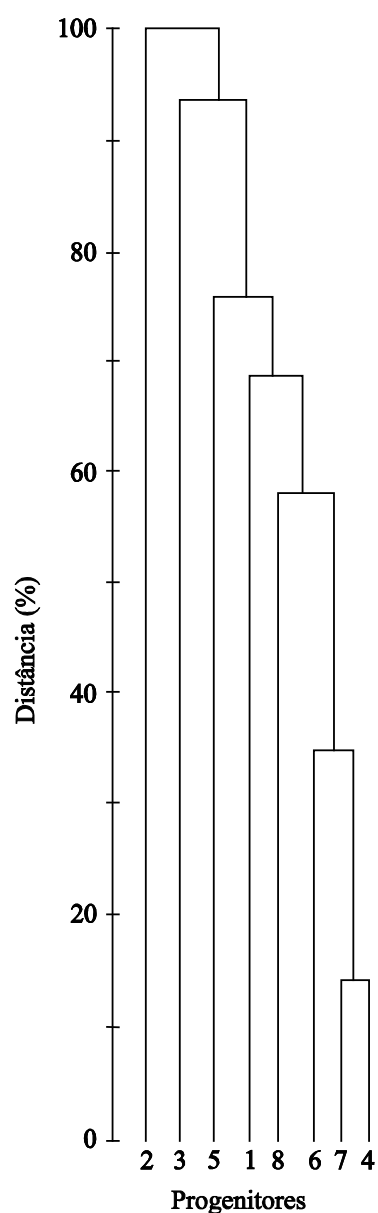


Figura 6.8 - Dendrograma ilustrativo da similaridade entre oito progenitores, obtido pelo método do vizinho mais próximo, baseado nas distâncias generalizadas de Mahalanobis

A análise do dendrograma (Figura 6.8) permite ao pesquisador verificar o grau de similaridade entre progenitores, progenitores e grupos similares, ou entre dois grupos distintos.

O estabelecimento dos grupos é feito de forma subjetiva, tendo por base **as mudanças acentuadas de níveis**, associadas ao conhecimento prévio que o pesquisador tem do material avaliado. Dentro deste princípio, pode-se pressupor, por exemplo, a existência dos grupos: (I) p_4 e p_7 ; (II) p_6 ; (III) p_8 , p_1 e p_5 ; (IV) p_3 e p_2 .

ii) Método de Tocher

Será aplicado o método de Tocher para o agrupamento de oito progenitores, cujas medidas de dissimilaridade são expressas pela distância generalizada de Mahalanobis (D^2), já apresentadas na Tabela 6.10.

Os seguintes passos são envolvidos:

Estimação da maior distância, dentre o conjunto de menores distâncias, entre cada progenitor.

As menores distâncias em relação a cada um dos progenitores são enfatizadas a seguir:

Progenitor	1	2	3	4	5	6	7	8
D^2 mínimo	15,49	23,46	22,00	3,16	15,66	7,66	3,16	13,74

Assim, é estabelecido $\theta = 23,4550 \cong 23,46$ como o limite de acréscimo, na média da distância intragrupo, para a formação ou inclusão de um novo elemento no grupo.

• Formação do Grupo I

O grupo I é formado inicialmente pelos progenitores 4 e 7, cuja medida de dissimilaridade é a menor de todas. A distância entre eles é de 3,1624, menor que o mínimo estabelecido.

Uma vez formado o grupo, calculam-se as medidas de dissimilaridade entre esse grupo e os demais progenitores. Como o critério de inclusão de novos progenitores em um grupo baseia-se no acréscimo da distância média do grupo, é apropriado o cálculo da distância entre progenitor-grupo por meio de:

$$d_{(ij)k} = \{d_{ik} + d_{jk}\}$$

em que d é qualquer medida de dissimilaridade. Assim, para o exemplo em questão, tem-se:

$$D^2_{(47)1} = D^2_{14} + D^2_{17} = 17,6295 + 15,4890 = 33,1185$$

$$D^2_{(47)2} = D^2_{24} + D^2_{27} = 56,5964 + 63,5913 = 120,1877$$

$$D^2_{(47)3} = D^2_{34} + D^2_{37} = 26,2044 + 23,6783 = 49,8831$$

$$D^2_{(47)5} = D^2_{45} + D^2_{57} = 37,4413 + 25,9657 = 63,4070$$

$$D^2_{(47)6} = D^2_{46} + D^2_{67} = 17,5122 + 7,6556 = 25,1678$$

$$D^2_{(47)8} = D^2_{48} + D^2_{78} = 35,2571 + 17,6554 = 52,9125$$

Com base nas distâncias entre progenitores-grupo, constata-se que o progenitor 6 é o mais similar. O acréscimo médio no valor de D^2 no grupo original, pela inclusão deste progenitor, é dado por:

$$\frac{D^2_{(4,7,6)} - D^2_{(4,7)}}{C_{3,2} - C_{2,2}} = \frac{D^2_{(47)6}}{2} = 12,5839$$

$C_{3,2}$ e $C_{2,2}$: combinação de três, dois a dois, e combinação de dois, dois a dois, respectivamente.

Como o valor 12,5839 é inferior a θ , a inclusão do progenitor no grupo I será permitida.

Uma maneira mais simples para se avaliar a possibilidade de inclusão de um progenitor em um grupo é verificar se a distância deste progenitor em relação ao grupo, dividida pelo número de progenitores que já o constitui, é inferior ao máximo permitido, ou seja:

$$\text{Se } \frac{D^2_{(\text{Grupo})i}}{g} \leq \theta \Rightarrow \text{inclui-se o progenitor } i \text{ no grupo.}$$

$$\text{Se } \frac{D^2_{(\text{Grupo})i}}{g} > \theta \Rightarrow \text{o progenitor } i \text{ não deve ser incluído no grupo.}$$

g = número de progenitores que constituem o grupo original.

As novas distâncias dos progenitores não incluídos, em relação ao grupo em formação (4, 6 e 7), são:

$$D^2_{(467)1} = D^2_{(47)1} + D^2_{16} = 33,1185 + 34,5228 = 67,6413$$

$$D^2_{(467)2} = D^2_{(47)2} + D^2_{26} = 120,1877 + 91,2130 = 211,4007$$

$$D^2_{(467)3} = D^2_{(47)3} + D^2_{36} = 49,8831 + 22,0013 = 71,8844$$

$$D^2_{(467)5} = D^2_{(47)5} + D^2_{56} = 63,4070 + 40,0172 = 103,4242$$

$$D^2_{(467)8} = D^2_{(47)8} + D^2_{68} = 52,9125 + 13,7396 = 66,6521$$

Agora o progenitor 8 é o mais similar ao grupo que está sendo formado. A inclusão deste progenitor é avaliada por meio de:

$$\frac{D^2_{(467)8}}{3} = \frac{66,6521}{3} = 22,2174$$

Como $22,2174 < \theta$, a inclusão do progenitor 8 é permitida.

As novas distâncias entre o grupo considerado e os demais progenitores ainda não incluídos são:

$$D^2_{(4678)1} = D^2_{(467)1} + D^2_{18} = 67,6413 + 29,6096 = 97,2509$$

$$D^2_{(4678)2} = D^2_{(467)2} + D^2_{28} = 211,4007 + 97,4834 = 308,8841$$

$$D^2_{(4678)3} = D^2_{(467)3} + D^2_{38} = 71,8844 + 44,0254 = 115,9098$$

$$D^2_{(4678)5} = D^2_{(467)5} + D^2_{58} = 103,4242 + 15,6553 = 119,0795$$

Com base nestas distâncias, fica evidenciado que o progenitor mais similar é o 1. A inclusão do mesmo no grupo em formação é avaliada por:

$$\frac{D^2_{(4678)1}}{4} = \frac{97,2509}{4} = 24,3127$$

Como $24,3127 > \theta$, a inclusão do progenitor 1 não é permitida.

O grupo I fica então estabelecido apenas com os progenitores 4, 6, 7 e 8.

• Formação do Grupo II

A matriz de medidas de dissimilaridade entre os progenitores ainda não agrupados é apresentada a seguir:

	(2)	(3)	(5)
(1)	23,4550	62,9440	29,7385
(2)		125,0260	100,4269
(3)			47,6977

Dentre os progenitores que ainda não foram agrupados, os mais similares são o 1 e 2. A distância entre eles (23,4550) é igual ao valor máximo preestabelecido (θ) para a formação de grupo. Assim, o grupo II fica inicialmente formado com estes progenitores.

As distâncias dos demais progenitores em relação ao grupo II, em formação, são dadas por:

$$D^2_{(12)3} = D^2_{13} + D^2_{23} = 187,9700$$

$$D^2_{(12)5} = D^2_{15} + D^2_{25} = 130,1654$$

O progenitor 5 destaca-se como o mais similar em relação ao grupo II. Sua inclusão neste grupo é avaliada por:

$$\frac{D^2_{(12)5}}{2} = \frac{130,1654}{2} = 65,0827$$

Como $65,0827 > \theta$, a inclusão do progenitor 5 não é permitida.

Logo, o grupo II é estabelecido somente pelos progenitores 1 e 2. Os demais progenitores não serão mais agrupados, pois suas dissimilaridades são superiores ao máximo preestabelecido para formação de grupos.

Resultado final

O agrupamento dos progenitores com base no método proposto por Tocher permite o estabelecimento dos grupos apresentados na Tabela 6.11.

Tabela 6.11 - Grupos de progenitores estabelecidos pelo método de Tocher, com base na dissimilaridade expressa pela distância generalizada de Mahalanobis

Grupo	Progenitores	Distância média
I	4, 7, 6 e 8	15,8304
II	1, 2	23,4550
III	5	—
IV	3	—

A distância média dentro do grupo é a média das distâncias entre cada par de progenitores que o constitui. Pelo critério adotado, esta distância é sempre menor que as distâncias médias intergrupos, que são obtidas pela média das distâncias entre pares de progenitores pertencentes aos diferentes grupos. Para o exemplo em questão, tem-se:

$$D_{I,II}^2 = \frac{1}{8} (D_{14}^2 + D_{24}^2 + \dots + D_{18}^2 + D_{28}^2) = 50,7669$$

As demais distâncias intergrupos são dadas na Tabela 6.12.

Tabela 6.12 - Distâncias médias intergrupos

Grupo	II	III	IV
I	50,7669	29,7699	28,9776
II		68,0825	93,9850
III			47,6977

6.4. EXEMPLO UTILIZANDO O SAS

(a) Programa:

```
options ls=64 ps=45 nodate nonumber;
title 'Output Agrupamento ';
data Mahalanobis (type=distance);
input trat1 trat2 trat3 trat4 trat5 trat6 trat7 trat $;
lines;
    0  3.63  4.57  3.57  4.57 11.87  6.39 trat1
  3.63    0  1.49  0.14  2.19 10.13  3.83 trat2
  4.57  1.49    0  0.71  0.18  3.93  0.64 trat3
  3.57  0.14  0.71    0  1.25  7.89  2.53 trat4
  4.57  2.19  0.18  1.25    0  2.96  0.26 trat5
11.87 10.13  3.93  7.89  2.96    0  1.53 trat6
  6.39  3.83  0.64  2.53  0.26  1.53    0 trat7
;
proc cluster data=Mahalanobis noeigen method=single
nonorm out=tree1;
id trat;
var trat1 trat2 trat3 trat4 trat5 trat6 trat7;
title2 'Mahalanobis Data: Sigle Linkage Clustering';
run;
filename gsasfile "dendro.graph";
goptions reset=all gaccess=gsasfile autofeed dev=pslmono;
goptions horigin=1in vorigin=2in;
goptions hsize=6in vsize=8in;
proc tree data=tree1 out=adair nclusters=3;
id trat;
title1 h=1.2 'Dendrogram: Single Linkage Method';
title2 j=1 'Output 1';
run;
proc sort data=adair;
by cluster;
run;
proc print data=adair;
by cluster;
title1 "Output 1";
title2 "3-Clustes Solution: Sigle Linkage Clustering";
title3 "Unstandardized Data: Using Mahalanobis Distance Matrix";
run;
```

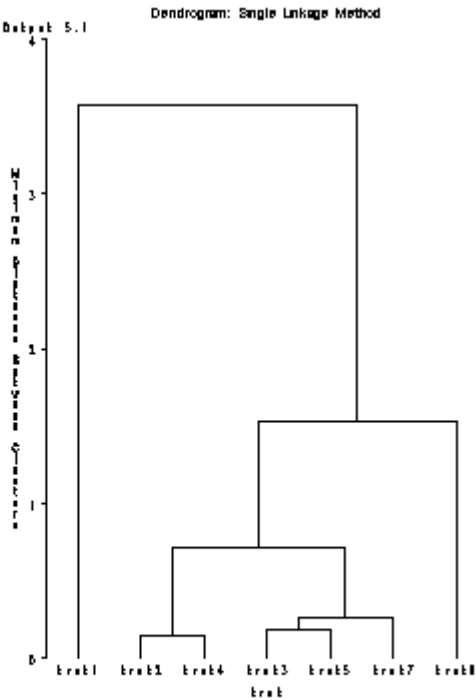
(b) Resultado da análise:

```
Output Agrupamento
Mahalanobis Data: Sigle Linkage Clustering
```

```
The CLUSTER Procedure
Single Linkage Cluster Analysis
```

Cluster History

NCL	--Clusters Joined---		FREQ	Min Dist	T i e
6	trat2	trat4	2	0.14	
5	trat3	trat5	2	0.18	
4	CL5	trat7	3	0.26	
3	CL6	CL4	5	0.71	
2	CL3	trat6	6	1.53	
1	trat1	CL2	7	3.57	



Output 1

3-Clustes Solution: Sigle Linkage Clustering
Unstandardized Data: Using Mahalanobis Distance Matrix

----- CLUSTER=1 -----

Obs	trat	CLUSNAME
1	trat2	CL3
2	trat4	CL3
3	trat3	CL3
4	trat5	CL3
5	trat7	CL3

----- CLUSTER=2 -----		
Obs	trat	CLUSNAME
6	trat6	trat6
----- CLUSTER=3 -----		
Obs	trat	CLUSNAME
7	trat1	trat1

6.5. EXERCÍCIO

De um experimento com 7 tratamentos e 3 repetições, no delineamento Inteiramente Casualizado, em que foram avaliadas 4 características (X_1, X_2, X_3, X_4) , obteve-se a seguinte matriz de distâncias (Distância de Mahalanobis):

$$D^2 = \begin{matrix} & \begin{matrix} 2 & 3 & 4 & 5 & 6 & 7 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{matrix} & \begin{bmatrix} 3,63 & 4,57 & 3,57 & 4,57 & 11,87 & 6,39 \\ & 1,49 & 0,14 & 2,19 & 10,13 & 3,83 \\ & & 0,71 & 0,18 & 3,93 & 0,64 \\ & & & 1,25 & 7,89 & 2,53 \\ & & & & 2,96 & 0,26 \\ & & & & & 1,53 \end{bmatrix} \end{matrix}$$

Pede-se:

- Obter o dendrograma considerando o método do vizinho mais próxima como algoritmo de agrupamento;
- Aplicar o método de Tocher para o agrupamento dos sete tratamentos, cujas medidas de dissimilaridade são expressas pela distância de Mahalanabis (D^2), apresentadas na matriz acima.

6.6. REFERÊNCIAS

CRUZ, C. D.; REGAZZI, A. J.; CARNEIRO, P. C. S. **Modelos biométricos aplicados ao melhoramento genético**. v. 1, 3. ed., Viçosa: UFV, 2012. 514p.

- EVERITT, B. S. **Cluster analysis**. Cambridge, Edward Arnold: University Press, 1993. 170p.
- JAMES, F. C.; McCULLOCH, C. E. Multivariate analysis in ecology and systematics: Panacea of pandora's box? **Annual Review Ecology Systematic**, v. 21, p.129-166, 1990.
- JOHNSON, R. A.; WICHERN, D.W. **Applied multivariate statistical analysis**. 4th ed., USA, Upper Saddle River, New Jersey: Prentice Hall, Inc., 1998. 816p.
- MAXWELL, A. E. **Multivariate analysis in behavioural research**. London, Chapman and Hall, 1977. 237p.
- MILLIGAN, G. W.; COOPER, M. C. An examination of procedures for determining the number of clusters in a data set. **Psychometrika**, Williamsburg, v. 50, p. 159-179, 1985.
- MOJENA, R. Hierarchical grouping methods and stopping rules: an evaluation. **Computer Journal**, London, v. 20, p. 359-363, 1977.
- RAO, C. R. **Advanced statistical methods in biometric research**. New York: John Wiley and Sons, 1952. 390p.
- ROHLF, F. J. Classification of Aedes by numerical taxonomic methods (Diptera: Culicidae). **Ann. Ent. Soc. Am.**, v.56, p. 798 – 804, 1963 a.
- ROHLF, F. J. Congruence of larval and adult classifications in Aedes (Diptera: Culicidae). **Syst. Zool.**, v.12, p. 97 – 117, 1963 b.
- ROHLF, F. J. Correlated characteres in numerical taxonomy. **Syst. Zool.**, v.16, p. 109 – 126, 1967.
- ROHLF, F. J. Adaptative hierarquical clustering schemes. **Syst. Zool.**, v.19, p. 58 – 82, 1970.
- SAS INSTITUTE INC. **Statistical Analysis System**. Version 9.1, USA: Cary, NC, 2004.
- SNEATH, P. H. A.; SOKAL, R. R. **Numerical taxonomy: the principles and practice of numerical classification**. San Francisco: W. H. Freeman, 1973. 573p.
- SOKAL, R. R.; MICHENER, C. D. A statistical methods for evaluating systematic relationships. **Kansas Univ. Sci. Bull.**, v. 38, p. 1409-1438, 1958.
- SOKAL, R. R.; ROHLF, F. J. The comparison of dendograms by objective methods. **Taxonomy**, v.11, p. 33 – 40, 1962.
- SOKAL, R. R.; SNEATH, P. H. A. **Principles of numerical taxonomy**. San Francisco: W.H. Freeman, 1963. 450 p.
- WARD, J. H. Hierarchical grouping to optimize an objective function. **Journal of the American Statistical Association**, v. 58, n. 301, p. 236-244. 1963.

CAPÍTULO 7

ANÁLISE DISCRIMINANTE

7.1. INTRODUÇÃO

O campo biológico apresenta uma infinidade de situações, nas quais, o pesquisador está interessado no estudo de um número p de características ou variáveis, sendo que na maioria dos casos interessa-nos o comportamento simultâneo de todas as variáveis e não o estudo isolado de cada uma delas. Neste sentido, a análise multivariada exerce um papel fundamental dentro da estatística matemática, assim como em aplicações.

Muito pouco era conhecido no campo da análise multivariada, em termos de teoria e aplicações, antes das investigações pioneiras realizadas por R. A. Fisher. Muitas contribuições de outros autores neste campo foram inspiradas no trabalho de Fisher, sendo este considerado “O Arquiteto da Análise Multivariada”, segundo Rao (1964).

Uma das finalidades da análise discriminante que vamos abordar neste capítulo, consiste em obter funções que permitam classificar um indivíduo \underline{X} (uma observação \underline{X}), com base em medidas de um número p de características do mesmo, em uma de várias populações π_i , ($i = 1, 2, \dots, g$), distintas, buscando minimizar a probabilidade de má classificação, isto é, minimizar a probabilidade de classificar erroneamente um indivíduo, em uma população π_i , quando ele realmente pertence à população π_j , ($i \neq j$) $i, j = 1, 2, \dots, g$.

Assim por exemplo, no campo médico, a fim de diagnosticar uma dada doença, o médico pode dispor de uma série de exames realizados no indivíduo e, com base nos resultados procurar classificá-lo como Portador ou Não Portador da dada doença. Entretanto, muitas vezes, com os resultados individuais de cada exame torna-se impraticável a tomada de uma decisão que tornem mínimos os riscos de uma classificação errônea. Desta forma necessita-se de um critério baseado em todas as características conjuntamente, características estas que se julgue de importância, que discrimine o melhor possível os grupos Portador e Não Portador, de modo que seja possível classificar um novo indivíduo em um dos dois grupos com mínimas chances de erro.

O problema de discriminação entre dois ou mais grupos, visando uma posterior classificação, foi inicialmente abordado por Fisher (1936) com a obtenção de uma combinação linear das características observadas que apresenta maior poder de discriminação entre os grupos. A esta combinação linear chamamos Função Discriminante Linear de Fisher, que é a base de todo estudo na Análise Discriminante. Tal função tem a propriedade de minimizar as

probabilidades de má classificação, quando as populações são normalmente distribuídas com parâmetros μ_i e Σ_i conhecidos. Entretanto tal situação pode não ocorrer na prática, necessitando-se, portanto de estimativas e métodos de estimação dessas probabilidades ótimas.

Vamos focalizar o problema de classificação de uma variável contínua normal multivariada, em uma de duas ou mais populações, supondo os parâmetros populacionais desconhecidos, uma vez que muito dificilmente, se conhece os vetores de médias e matrizes de covariâncias, tornando necessário estimá-los.

7.2. CLASSIFICAÇÃO EM UMA DE DUAS POPULAÇÕES

7.2.1. Função discriminante linear de Fisher

Uma das soluções apresentadas ao problema da classificação de uma observação em uma de duas populações foi sugerida por Fisher (1936), através da utilização de uma função linear do vetor aleatório \tilde{X} , a qual se caracteriza por produzir separação máxima entre as duas populações.

Consideremos μ_i e Σ_i parâmetros conhecidos, isto é, o vetor de médias \mathbf{p} -variado e a matriz, de ordem p , de covariâncias comum das populações π_i , ($i = 1, 2$), respectivamente. Admite-se, também, que em cada população o vetor de observações \tilde{X} tem distribuição $N_p(\mu_i, \Sigma)$. Assim, \tilde{X} , em uma determinada população π_i , apresenta a seguinte função densidade de probabilidade:

$$f_i(\tilde{x}) = |2\pi\Sigma|^{-1/2} e^{-\frac{1}{2}\left[\left(\tilde{x}-\mu_i\right)'\Sigma^{-1}\left(\tilde{x}-\mu_i\right)\right]}$$

Tomando um par de populações π_1 e π_2 , aloca-se um indivíduo, com vetor de observações \tilde{x} , em π_1 se:

$$\frac{f_1(\tilde{x})}{f_2(\tilde{x})} > 1$$

o que, significa que:

$$\left(\tilde{x}-\mu_1\right)'\Sigma^{-1}\left(\tilde{x}-\mu_1\right) < \left(\tilde{x}-\mu_2\right)'\Sigma^{-1}\left(\tilde{x}-\mu_2\right)$$

ou

$$D_1^2 < D_2^2$$

sendo D_i^2 a distância generalizada de Mahalanobis, calculada tomando as informações do indivíduo a ser classificado e a média da i -ésima população. Assim, a análise discriminante consiste, portanto, em alocar um indivíduo com um conjunto de observações \tilde{x} numa população π_i adotando-se o critério de que a distância de Mahalanobis seja mínima.

Será considerada novamente a inequação:

$$\left(\tilde{x} - \tilde{\mu}_1\right)' \Sigma^{-1} \left(\tilde{x} - \tilde{\mu}_1\right) - \left(\tilde{x} - \tilde{\mu}_2\right)' \Sigma^{-1} \left(\tilde{x} - \tilde{\mu}_2\right) < 0$$

Expandindo-a, tem-se:

$$\left(\tilde{x}' \Sigma^{-1} \tilde{x} - \tilde{x}' \Sigma^{-1} \tilde{\mu}_1 - \tilde{\mu}_1' \Sigma^{-1} \tilde{x} + \tilde{\mu}_1' \Sigma^{-1} \tilde{\mu}_1\right) - \left(\tilde{x}' \Sigma^{-1} \tilde{x} - \tilde{x}' \Sigma^{-1} \tilde{\mu}_2 - \tilde{\mu}_2' \Sigma^{-1} \tilde{x} + \tilde{\mu}_2' \Sigma^{-1} \tilde{\mu}_2\right) < 0$$

tendo-se, para os escalares, a seguinte igualdade:

$$\tilde{x}' \Sigma^{-1} \tilde{\mu}_1 = \tilde{\mu}_1' \Sigma^{-1} \tilde{x}$$

$$\tilde{x}' \Sigma^{-1} \tilde{\mu}_2 = \tilde{\mu}_2' \Sigma^{-1} \tilde{x}$$

logo:

$$-2 \tilde{x}' \Sigma^{-1} \tilde{\mu}_1 + \tilde{\mu}_1' \Sigma^{-1} \tilde{\mu}_1 + 2 \tilde{x}' \Sigma^{-1} \tilde{\mu}_2 + \tilde{\mu}_2' \Sigma^{-1} \tilde{\mu}_2 < 0$$

Arranjando os elementos, somando e subtraindo os escalares $\tilde{\mu}_1' \Sigma^{-1} \tilde{\mu}_2 = \tilde{\mu}_2' \Sigma^{-1} \tilde{\mu}_1$ na

expressão anterior, tem-se:

$$-2 \tilde{x}' \Sigma^{-1} \left(\tilde{\mu}_1 - \tilde{\mu}_2\right) + \left(\tilde{\mu}_1' \Sigma^{-1} \tilde{\mu}_1 + \tilde{\mu}_1' \Sigma^{-1} \tilde{\mu}_2\right) - \left(\tilde{\mu}_2' \Sigma^{-1} \tilde{\mu}_2 + \tilde{\mu}_2' \Sigma^{-1} \tilde{\mu}_1\right) < 0$$

$$-2 \tilde{x}' \Sigma^{-1} \left(\tilde{\mu}_1 - \tilde{\mu}_2\right) + \tilde{\mu}_1' \Sigma^{-1} \left(\tilde{\mu}_1 + \tilde{\mu}_2\right) - \tilde{\mu}_2' \Sigma^{-1} \left(\tilde{\mu}_1 + \tilde{\mu}_2\right) < 0$$

$$-2 \tilde{x}' \Sigma^{-1} \left(\tilde{\mu}_1 - \tilde{\mu}_2\right) + \left(\tilde{\mu}_1 - \tilde{\mu}_2\right)' \Sigma^{-1} \left(\tilde{\mu}_1 + \tilde{\mu}_2\right) < 0$$

e, finalmente:

$$\tilde{x}' \Sigma^{-1} \left(\tilde{\mu}_1 - \tilde{\mu}_2\right) - \left(\tilde{\mu}_1 - \tilde{\mu}_2\right)' \Sigma^{-1} \frac{1}{2} \left(\tilde{\mu}_1 + \tilde{\mu}_2\right) > 0$$

Mardia et al. (1979) relatam que a decisão de alocar um indivíduo com vetor de observações \tilde{x} em π_1 ao invés de π_2 ocorre quando:

$$\alpha'(\tilde{x} - \tilde{u}) > 0$$

em que:

$$\tilde{\alpha} = \Sigma^{-1} \begin{pmatrix} \tilde{\mu}_1 - \tilde{\mu}_2 \end{pmatrix}$$

e

$$\tilde{u} = \frac{1}{2} \begin{pmatrix} \tilde{\mu}_1 + \tilde{\mu}_2 \end{pmatrix}$$

Demonstra-se que a função linear do vetor aleatório $\tilde{X}' = [X_1, X_2, \dots, X_p]$, que produz separação máxima entre as duas populações, é estabelecida pela seguinte expressão:

$$\tilde{l}' \tilde{X} = [\tilde{\mu}_1 - \tilde{\mu}_2]' \Sigma^{-1} \tilde{X},$$

a qual é denominada **Função Discriminante Linear de Fisher** (FDLF).

Note que $\tilde{l}' = [\tilde{\mu}_1 - \tilde{\mu}_2]' \Sigma^{-1}$.

Vamos usar a seguinte notação para a FDLF: $D(\tilde{X}) = \tilde{l}' \tilde{X} = [\tilde{\mu}_1 - \tilde{\mu}_2]' \Sigma^{-1} \tilde{X}$.

Seja $D(\tilde{x}_0) = \tilde{l}' \tilde{x}_0 = [\tilde{\mu}_1 - \tilde{\mu}_2]' \Sigma^{-1} \tilde{x}_0$ o valor da função discriminante para uma nova observação \tilde{x}_0 e seja

$$m = \frac{1}{2} [\tilde{\mu}_1 - \tilde{\mu}_2]' \Sigma^{-1} [\tilde{\mu}_1 + \tilde{\mu}_2] = \frac{1}{2} [\tilde{l}' \tilde{\mu}_1 + \tilde{l}' \tilde{\mu}_2]$$

$$m = \frac{1}{2} \left(D(\tilde{\mu}_1) + D(\tilde{\mu}_2) \right)$$

o ponto médio entre as duas médias populacionais univariadas $D(\tilde{\mu}_1)$ e $D(\tilde{\mu}_2)$.

A regra de classificação é:

$$\text{Alocar } \tilde{x}_0 \text{ em } \pi_1 \text{ se } D(\tilde{x}_0) = [\tilde{\mu}_1 - \tilde{\mu}_2]' \Sigma^{-1} \tilde{x}_0 \geq m$$

$$\text{Alocar } \tilde{x}_0 \text{ em } \pi_2 \text{ se } D(\tilde{x}_0) = [\tilde{\mu}_1 - \tilde{\mu}_2]' \Sigma^{-1} \tilde{x}_0 < m$$

Alternativamente, podemos subtrair m de $D(\tilde{x}_0)$ e comparar o resultado com zero.

Neste caso, a regra é a seguinte:

$$\text{Alocar } \tilde{x}_0 \text{ em } \pi_1 \text{ se } D(\tilde{x}_0) - m \geq 0$$

$$\text{Alocar } \tilde{x}_0 \text{ em } \pi_2 \text{ se } D(\tilde{x}_0) - m < 0$$

As situações colocadas anteriormente, valiosas do ponto de vista teórico, são de pouca utilidade na prática, uma vez que muito dificilmente se conhece os vetores de médias e matrizes

de covariâncias. O caminho usual, neste caso, é supor normalidade da população e estimar os parâmetros μ_i e Σ . Os processos de obtenção de regras de discriminação continuam os mesmo descritos anteriormente, utilizando-se as estimativas ao invés dos parâmetros. Assim $\bar{1}$ e m podem ser estimados das observações que já foram corretamente classificadas.

Sejam duas populações normais multivariadas π_1 e π_2 . Suponha, então, que nós temos n_1 observações da variável aleatória multivariada $\tilde{X}' = [X_1, X_2, \dots, X_p]$ de π_1 e n_2 medidas destas quantidades de π_2 . As respectivas matrizes de dados são:

$$\begin{aligned} \underset{(p \times n_1)}{X_1} &= \begin{bmatrix} \tilde{x}_{11}, \tilde{x}_{12}, \dots, \tilde{x}_{1n_1} \end{bmatrix} \\ \underset{(p \times n_2)}{X_2} &= \begin{bmatrix} \tilde{x}_{21}, \tilde{x}_{22}, \dots, \tilde{x}_{2n_2} \end{bmatrix} \end{aligned}$$

Sejam $\tilde{x}_{11}, \tilde{x}_{12}, \dots, \tilde{x}_{1n_1}$ observações de $N_p(\mu_1; \Sigma)$ e

$\tilde{x}_{21}, \tilde{x}_{22}, \dots, \tilde{x}_{2n_2}$ observações de $N_p(\mu_2; \Sigma)$.

Das matrizes de dados, os vetores de médias e as matrizes de covariâncias amostrais são dados por:

$$\underset{(p \times 1)}{\bar{x}_1} = \frac{1}{n_1} \sum_{j=1}^{n_1} \tilde{x}_{1j}$$

$$\underset{(p \times 1)}{\bar{x}_2} = \frac{1}{n_2} \sum_{j=1}^{n_2} \tilde{x}_{2j}$$

$$\underset{(p \times p)}{S_1} = \frac{A_1}{n_1 - 1} = \frac{1}{n_1 - 1} \sum_{j=1}^{n_1} (\tilde{x}_{1j} - \bar{x}_1)(\tilde{x}_{1j} - \bar{x}_1)'$$

$$\underset{(p \times p)}{S_2} = \frac{A_2}{n_2 - 1} = \frac{1}{n_2 - 1} \sum_{j=1}^{n_2} (\tilde{x}_{2j} - \bar{x}_2)(\tilde{x}_{2j} - \bar{x}_2)'$$

Uma vez que foi assumido que as populações de origem têm a mesma matriz de covariância Σ , as matrizes de covariância amostral S_1 e S_2 são combinadas ("pooled"), obtendo assim uma única matriz comum de covariância amostral, que é dada por:

$$S_c = \frac{A_1}{(n_1 - 1) + (n_2 - 1)} + \frac{A_2}{(n_1 - 1) + (n_2 - 1)}$$

$$S_c = \left[\frac{n_1 - 1}{(n_1 - 1) + (n_2 - 1)} \right] S_1 + \left[\frac{n_2 - 1}{(n_1 - 1) + (n_2 - 1)} \right] S_2$$

$$S_c = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{(n_1 + n_2 - 2)}.$$

Em particular, S_c é um estimador não tendencioso de Σ se as matrizes de dados X_1 e X_2 contém amostras aleatórias das populações π_1 e π_2 , respectivamente.

Destaca-se o fato de que se deve ter $n_1 + n_2 - 2 > p$; caso contrário, S_c é singular e não apresenta inversa comum. Quando se dispõe de várias populações (π_i , com $i = 1, 2, \dots, g$), cada uma com n_i indivíduos a partir dos quais são estimadas as matrizes de covariâncias S_i , obtém-se S_c por meio de:

$$S_c = \frac{\sum_{i=1}^g (n_i - 1)S_i}{\sum_{i=1}^g n_i - p}$$

Os parâmetros μ_1, μ_2 e Σ da Função Discriminante Linear de Fisher são substituídos pelas quantidades amostrais \bar{x}_1, \bar{x}_2 e S_c , respectivamente, resultando no seguinte:

Função Discriminante Linear Amostral de Fisher:

$$D(\tilde{x}) = \hat{1}' \tilde{x} = [\bar{x}_1 - \bar{x}_2]' S_c^{-1} \tilde{x}$$

Fato: É preciso ter $(n_1 + n_2 - 2) > p$, senão S_c é singular e a inversa S_c^{-1} não existe.

O ponto médio, \hat{m} , entre as duas médias amostrais univariadas, $D(\bar{x}_1) = \hat{1}' \bar{x}_1$ e

$D(\bar{x}_2) = \hat{1}' \bar{x}_2$ é dado por:

$$\begin{aligned} \hat{m} &= \frac{1}{2} [\bar{x}_1 - \bar{x}_2]' S_c^{-1} [\bar{x}_1 + \bar{x}_2] = \frac{1}{2} (\hat{1}' \bar{x}_1 + \hat{1}' \bar{x}_2) \\ &= \frac{1}{2} \left(D(\bar{x}_1) + D(\bar{x}_2) \right). \end{aligned}$$

A regra de classificação baseada na Função Discriminante Linear de Fisher, obtida a partir dos dados amostrais é apresentada a seguir:

Alocar \tilde{x}_0 em π_1 se

$$D(\tilde{x}_0) = [\bar{\tilde{x}}_1 - \bar{\tilde{x}}_2]' S_c^{-1} \tilde{x}_0 \geq \hat{m}$$

$$\text{ou } D(\tilde{x}_0) - \hat{m} \geq 0$$

Alocar \tilde{x}_0 em π_2 se

$$D(\tilde{x}_0) < \hat{m}$$

$$\text{ou } D(\tilde{x}_0) - \hat{m} < 0$$

7.2.2. Avaliação da função discriminante

- **Teste de Significância**

A distância amostral de Mahalanobis entre as duas populações é dada por:

$$D^2 = [\bar{\tilde{x}}_1 - \bar{\tilde{x}}_2]' S_c^{-1} [\bar{\tilde{x}}_1 - \bar{\tilde{x}}_2] = D(\bar{\tilde{x}}_1) - D(\bar{\tilde{x}}_2).$$

A distribuição de D^2 pode ser usada para verificar se existem diferenças significativas entre as duas populações, isto é, se os dois vetores de médias diferem significativamente.

Suponha que as populações π_1 e π_2 são normais multivariadas com uma matriz de covariância comum Σ . Então, um teste para $H_0: \mu_1 = \mu_2$ vs. $H_a: \mu_1 \neq \mu_2$ é efetuado considerando a estatística

$$F_0 = \left(\frac{n_1 + n_2 - p - 1}{(n_1 + n_2 - 2)p} \right) \left(\frac{n_1 \cdot n_2}{n_1 + n_2} \right) D^2,$$

onde n_1 e n_2 são os tamanhos das amostras em π_1 e π_2 , respectivamente, e p é o número de variáveis. Sob a hipótese nula, F_0 tem distribuição F com $v_1 = p$ e $v_2 = n_1 + n_2 - p - 1$ graus de liberdade. Se H_0 for rejeita, pode-se concluir que a separação entre as duas populações π_1 e π_2 é significativa.

Comentário: Separação significativa não implica necessariamente em boa classificação. A eficácia de um método de classificação pode ser avaliada independentemente de qualquer teste de separação. Por outro lado, se a separação não for significativa, a procura de uma regra de classificação útil será provavelmente infrutífera.

Na realidade o teste apresentado anteriormente é o teste T^2 de Hotelling para igualdade dos vetores de médias de duas populações normais.

7.2.3. Métodos de estimação das probabilidades de má classificação

Existem vários métodos de estimação das probabilidades de má classificação. Vamos apresentar apenas dois deles, supondo ainda que os custos de má classificação ($C(j/i)$): é o custo de classificarmos erroneamente uma observação de π_i em π_j ; $i \neq j$) bem como as probabilidades “a priori” de uma observação provir da população π_i , $i = 1, 2$, sejam iguais.

7.2.3.1. Método devido a Okamoto (1963)

Este método é dependente da suposição de normalidade das populações. Ele fornece as probabilidades aproximadas de má classificação.

Sendo $D^2 = [\bar{x}_1 - \bar{x}_2]' S_c^{-1} [\bar{x}_1 - \bar{x}_2]$ a distância amostral de Mahalanobis e, considerando as suposições anteriores, temos que a probabilidade de má classificação para as duas populações são iguais e estimadas por:

$$\hat{P}(2/1) = \hat{P}(1/2) = \phi\left(-\frac{D}{2}\right),$$

em que $\phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx$ e $\hat{P}(j/i)$ é a estimativa da probabilidade de classificarmos uma observação de π_i em π_j , $i \neq j$. Tal procedimento de estimação das probabilidades de má classificação é citado na literatura como “Método O”.

7.2.3.2. Método devido a Smith (1947)

Este método não depende da suposição de normalidade das populações. Smith (1947) sugere que as observações usadas para a construção da função discriminante, sejam reusadas para a estimação de $P(j/i)$. Assim se $D(\underline{x})$ é construída utilizando-se as $n_1 + n_2$ observações e se $m_i \leq n_i$ delas, são má classificadas em π_i através de $D(\underline{x})$, temos que:

$$\hat{P}(j/i) = \hat{P}_i = \frac{m_i}{n_i}, i = 1, 2$$

Assim, $\hat{P}(2/1) = \hat{P}_1 = \frac{m_1}{n_1}$ e $\hat{P}(1/2) = \hat{P}_2 = \frac{m_2}{n_2}$.

De todos os métodos citados na literatura, este parece ser o mais fraco no sentido de que as estimativas diferem consideravelmente do valor real, estes fixados em experimentos amostrais.

7.2.4. Exemplo

Como um exemplo ilustrativo para a obtenção da Função Discriminante Linear Amostral de Fisher considere os dados relativos às variáveis:

X_1 = número médio de cerdas próximas à cabeça

X_2 = número médio de cerdas junto às patas

São dados de duas raças de insetos, apresentados por Hoel (1971), cujos resultados estão apresentados na Tabela 7.1.

Tabela 7.1 – Dados relativos ao número médio de cerdas próximas à cabeça (X_1) e número médio de cerdas junto às patas (X_2) em duas raças de insetos*

Raça A		Raça B	
X_1	X_2	X_1	X_2
6,36	5,24	6,00	4,88
5,92	5,12	5,60	4,64
5,92	5,36	5,64	4,96
6,44	5,64	5,76	4,80
6,40	5,16	5,96	5,08
6,56	5,56	5,72	5,04
6,64	5,36	5,64	4,96
6,68	4,96	5,44	4,88
6,72	5,48	5,04	4,44
6,76	5,60	4,56	4,04
6,72	5,08	5,48	4,20
		5,76	4,80

* Fonte: Hoel (1971)

Neste exemplo temos $p = 2$ variáveis, $n_1 = 11$ e $n_2 = 12$. A matriz X_1 ($p \times n_1$) é constituída pelos elementos da Raça A (π_1) e a matriz X_2 ($p \times n_2$) é constituída pelos elementos da Raça B (π_2).

Com base nos dados da Tabela 7.1, temos:

Raça A

$$\bar{\tilde{x}}_1 = \begin{bmatrix} \bar{x}_{11} \\ \bar{x}_{12} \end{bmatrix} = \begin{bmatrix} 6,46545 \\ 5,32364 \end{bmatrix}$$

$$S_1 = \begin{bmatrix} 0,091287 & 0,011258 \\ 0,011258 & 0,052625 \end{bmatrix}$$

Raça B

$$\bar{\tilde{x}}_2 = \begin{bmatrix} \bar{x}_{21} \\ \bar{x}_{22} \end{bmatrix} = \begin{bmatrix} 5,55000 \\ 4,72667 \end{bmatrix}$$

$$S_2 = \begin{bmatrix} 0,160327 & 0,107418 \\ 0,107418 & 0,111661 \end{bmatrix}$$

Supondo $\Sigma_1 = \Sigma_2 = \Sigma$, a matriz comum de covariância amostral é dada por:

$$S_c = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{(n_1 + n_2 - 2)}$$

$$S_c = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{(n_1 + n_2 - 2)} = \begin{bmatrix} 0,127451 & 0,061627 \\ 0,061627 & 0,083549 \end{bmatrix}$$

$|S_c| = 0,00685051647$. Assim a inversa de S_c é:

$$S_c^{-1} = \begin{bmatrix} 12,196015 & -8,995964 \\ -8,995964 & 18,604583 \end{bmatrix}.$$

- **Função Discriminante Linear Amostral de Fisher**

$$D(\tilde{x}) = \hat{\tilde{l}}' \tilde{x} = [\bar{\tilde{x}}_1 - \bar{\tilde{x}}_2]' S_c^{-1} \tilde{x}$$

$$\hat{\tilde{l}}' = \begin{bmatrix} 6,46545 - 5,55000 \\ 5,32364 - 4,72667 \end{bmatrix}' S_c^{-1}$$

$$\hat{\tilde{l}}' = \begin{bmatrix} 0,91545 & 0,59697 \end{bmatrix}' \begin{bmatrix} 12,196015 & -8,995964 \\ -8,995964 & 18,604583 \end{bmatrix}$$

$$\hat{\tilde{l}}' = \begin{bmatrix} 5,794521 & 2,871023 \end{bmatrix}$$

Assim,

$$D(\underline{x}) = \begin{bmatrix} 5,794521 & 2,871023 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$D(\underline{x}) = 5,794521 x_1 + 2,871023 x_2$$

- **Classificação de novos indivíduos**

Um novo indivíduo ou uma nova observação \underline{x}_0 pertence à Raça A (π_1) ou à Raça B

(π_2)?

$$\text{Calcula-se inicialmente } \hat{m} = \frac{1}{2} \left(D(\bar{\underline{x}}_1) + D(\bar{\underline{x}}_2) \right)$$

$$D(\bar{\underline{x}}_1) = \hat{1}' \bar{\underline{x}}_1$$

$$D(\bar{\underline{x}}_1) = \begin{bmatrix} 5,794521 & 2,871023 \end{bmatrix} \begin{bmatrix} 6,46545 \\ 5,32364 \end{bmatrix}$$

$$D(\bar{\underline{x}}_1) = 52,748479$$

$$D(\bar{\underline{x}}_2) = \hat{1}' \bar{\underline{x}}_2$$

$$D(\bar{\underline{x}}_2) = \begin{bmatrix} 5,794521 & 2,871023 \end{bmatrix} \begin{bmatrix} 5,55000 \\ 4,72667 \end{bmatrix}$$

$$D(\bar{\underline{x}}_2) = 45,729970$$

$$\hat{m} = \frac{1}{2} (52,748479 + 45,729970) = 49,239$$

A regra de classificação é a seguinte:

Um novo indivíduo \underline{x}_0 será classificado em π_1 , isto é, na Raça A se

$$D(\underline{x}_0) = \hat{1}' \underline{x}_0 \geq 49,239 \text{ ou } D(\underline{x}_0) - 49,239 \geq 0. \text{ Esse novo indivíduo } \underline{x}_0 \text{ será classificado}$$

em π_2 , isto é, na Raça B se $D(\underline{x}_0) = \hat{1}' \underline{x}_0 < 49,239$ ou $D(\underline{x}_0) - 49,239 < 0$.

Por exemplo, para um novo indivíduo, o número médio de cerdas próximas à cabeça e junto às patas foi 6,21 e 5,31, respectivamente. Em qual das duas raças este indivíduo seria classificado?

$$D(\bar{x}_0) = \hat{1}' \bar{x}_0 = [5,794521 \quad 2,871023] \begin{bmatrix} 6,21 \\ 5,31 \end{bmatrix}$$

$$D(\bar{x}_0) = 51,229107$$

Como $D(\bar{x}_0) = 51,229107 > 49,239$, o indivíduo seria classificado na Raça A.

- Teste de significância para verificar se a separação entre as duas populações π_1 e π_2 é significativa.

$$H_0 : \mu_1 = \mu_2 \text{ vs. } H_a : \mu_1 \neq \mu_2$$

$$F_0 = \left(\frac{n_1 + n_2 - p - 1}{(n_1 + n_2 - 2)p} \right) \left(\frac{n_1 \cdot n_2}{n_1 + n_2} \right) D^2 \sim F(p; n_1 + n_2 - p - 1)$$

$$D^2 = [\bar{x}_1 - \bar{x}_2]' S_c^{-1} [\bar{x}_1 - \bar{x}_2] = D(\bar{x}_1) - D(\bar{x}_2)$$

$$D^2 = 52,748479 - 45,729970 = 7,018509$$

$$F_0 = \left(\frac{11 + 12 - 2 - 1}{(11 + 12 - 2)2} \right) \left(\frac{(11)(12)}{11 + 12} \right) 7,018509 = 19,18^{**}$$

$$F_{1\%}(2; 11 + 12 - 2 - 1) = F_{1\%}(2; 20) = 5,85$$

** Rejeita-se H_0 e concluímos que a separação entre as duas populações π_1 e π_2 é significativa ($P < 0,01$).

- Estimação das probabilidades de má classificação

(a) Método devido Okamoto (1963)

Vimos que $D^2 = 7,018509$. Considerando as suposições apresentadas anteriormente, temos que a estimativa da probabilidade de má classificação para as duas populações são iguais e dadas por:

$$\begin{aligned}
 \hat{P}(2/1) &= \hat{P}(1/2) = \phi\left(-\frac{D}{2}\right) \\
 &= \phi\left(-\frac{\sqrt{7,018509}}{2}\right) \\
 &= \phi(-1,32) = 0,0934
 \end{aligned}$$

Assim, por este método, existe uma probabilidade de 9,34% de errarmos ao classificar um indivíduo em π_1 (Raça A) ou em π_2 (Raça B).

(b) Método devido a Smith (1947)

Considere inicialmente os dados apresentados na Tabela 7.2.

Tabela 7.2 – Valores de $D(\underline{x})$ para as duas raças de insetos, dados pela função discriminante

$$D(\underline{x}) = 5,794521 x_1 + 2,871023 x_2$$

Raça A	Raça B
51,8973	48,7777
49,0032*	45,7709
49,6922	46,9214
53,5093	47,1574
51,8994	49,1201
53,9749	47,6146
53,8643	46,9214
52,9477	45,5328
54,6724	41,9517
55,2487	38,0219
53,5240	43,8123
	47,1574

*Erro de classificação

Estimativas das probabilidades de má classificação:

Raça A (π_1)

$$\hat{P}(2/1) = \hat{P}_1 = \frac{m_1}{n_1} = \frac{1}{11} = 0,0909 \\ = 9,09 \%$$

Raça B (π_2)

$$\hat{P}(1/2) = \hat{P}_2 = \frac{m_2}{n_2} = \frac{0}{12} = 0 \\ = 0 \%$$

Um resumo destes resultados está representado na Tabela 7.3.

Tabela 7.3 – Distribuição dos elementos segundo a designação a priori e através da função discriminante

Designação a priori	Função Discriminante		Total	Classificação errada (%)
	Raça A	Raça B		
Raça A	10	1	11	9,09
Raça B	0	12	12	0,00
Total	10	13	23	-

Note que a média dos escores da Raça A e da Raça B apresentados na Tabela 7.2, são iguais a $D(\bar{x}_1)$ e $D(\bar{x}_2)$, respectivamente, isto é:

$$\frac{580,2334}{11} = 52,7485 = D(\bar{x}_1) \text{ e } \frac{548,7596}{12} = 45,7299 = D(\bar{x}_2).$$

Fato: Vimos que a Função Discriminante Linear Amostral de Fisher para o exemplo em questão foi:

$$D(\underline{x}) = \hat{\underline{l}}' \underline{x} = 5,794521 x_1 + 2,871023 x_2$$

O vetor $\hat{\underline{l}} = \begin{bmatrix} \hat{l}_1 & \hat{l}_2 \end{bmatrix}$ pode ser multiplicado por uma constante tal que torne \hat{l}_1 , ou então

\hat{l}_2 , igual a 1. Neste exemplo, se dividirmos os valores de $\hat{\underline{l}}$ pelo valor 5,794521, teremos:

$$D(\underline{x}) = x_1 + 0,495472 x_2$$

Este procedimento tem a vantagem de facilitar o cálculo dos escores da função discriminante, o que é desnecessário pelas facilidades computacionais de hoje. É importante salientar que esta última função leva obrigatoriamente à mesma classificação obtida com a função anterior.

Também poderíamos utilizar o vetor normalizado tal que $\hat{\tilde{1}} \hat{\tilde{1}} = 1$.

7.3. CLASSIFICAÇÃO EM UMA DE VÁRIAS POPULAÇÕES: FUNÇÕES DISCRIMINANTES DE ANDERSON (1958)

7.3.1. Introdução

Sejam $\pi_1, \pi_2, \dots, \pi_g$ um grupo de g populações. Um ponto importante além de assumir alguma distribuição para obter as funções discriminantes, é o estabelecimento das probabilidades a priori para as várias populações. Há casos em que a probabilidade de um determinado indivíduo (uma observação \underline{x}) pertencer a uma determinada população é muito diferente dele pertencer a outra. De forma que a experiência do pesquisador nesse momento é extremamente importante. Além desses pontos é preciso considerar os custos de má classificação, pois há casos em que, pela natureza do fenômeno, precisam-se estudar profundamente as variáveis realmente necessárias para que o uso dessa técnica venha a ter sucesso. A teoria está bem fundamentada, mas o importante é o pesquisador fornecer as informações adequadamente.

7.3.2. Estimação das funções discriminantes e classificação

Na análise discriminante proposta por Anderson (1958), consideram-se as informações de indivíduos sabidamente pertencentes a diferentes populações. A partir dessas informações são geradas funções, que são combinações lineares das características avaliadas e têm por finalidade promover a melhor discriminação entre os indivíduos, alocando-os em suas devidas populações. Estas funções, uma vez estimadas, passam a ser de grande utilidade por permitirem classificar novos indivíduos, de comportamento desconhecido, nas populações já conhecidas. A eficácia das variáveis utilizadas em promover a discriminação também é avaliada, permitindo conhecer a adequação da função estimada.

Nesta situação, tem-se uma observação \underline{x} e deseja-se classificá-la em uma dentre g populações, $g > 2$. Vamos admitir que a discriminação a ser feita envolva custos idênticos de má classificação e probabilidades (p_i = probabilidade de que uma observação seja classificada na população i) conhecidas para as várias populações ($\sum_{i=1}^g p_i = 1$).

Para o estabelecimento da função discriminante de Anderson, novamente se considera que, para uma população π_i ($i=1,2,\dots,g$), o vetor de observações \tilde{X} tem distribuição $N_p(\mu_i, \Sigma)$, com a seguinte função densidade de probabilidade:

$$f_i(\tilde{x}) = |2\pi\Sigma|^{-1/2} e^{-\frac{1}{2}\left[\left(\tilde{x}-\mu_i\right)'\Sigma^{-1}\left(\tilde{x}-\mu_i\right)\right]}$$

Anderson (1958) relata que é fundamental considerar na classificação de observações a probabilidade a priori inerente às várias populações avaliadas em um determinado estudo. Há casos em que a probabilidade de um determinado indivíduo, com vetor de observações \tilde{x} , pertencer a uma determinada população já é, de certa forma, conhecida pelo pesquisador, a partir de sua experiência e do conhecimento que se tem sobre as populações estudadas e do indivíduo a ser classificado.

Como ilustração, pode-se imaginar o estabelecimento de funções discriminantes entre duas populações π_1 e π_2 a partir de um conjunto de variáveis mensuráveis. Entretanto, pode-se supor que, além dos caracteres avaliados e da função discriminante estimada, o pesquisador tem conhecimento de que a população π_1 é típica de uma região de baixada, enquanto a π_2 é típica de uma região de encosta. Assim, se é desejado classificar um indivíduo desconhecido a partir de um vetor de observações, mas se sabe que este indivíduo foi coletado numa região de baixada, é de se pressupor que a sua maior probabilidade é a de pertencer à população π_1 . Dessa forma, a probabilidade a priori deve ser computada, fazendo com que a discriminação seja feita de forma mais eficaz.

Será admitido que a probabilidade de uma observação pertencer a uma determinada população é p_i ($\sum_{i=1}^g p_i = 1$), conhecida a priori. Assim, pode-se estabelecer a função discriminante, dada pela probabilidade de \tilde{x} pertencer a π_i , por meio do logaritmo da função densidade de probabilidade de \tilde{x} , de forma que se tenha:

$$D_i(\tilde{x}) = -\frac{1}{2}[p\ln(2\pi) + \ln|\Sigma_i|] - \frac{1}{2}\left[\left(\tilde{x}-\mu_i\right)'\Sigma_i^{-1}\left(\tilde{x}-\mu_i\right)\right] + \ln(p_i)$$

Supondo a homogeneidade das matrizes de covariâncias ($\Sigma_1 = \Sigma_2 = \dots \Sigma_g = \Sigma$) e retirando-se os componentes constantes da função $D_i(\tilde{x})$ que são dispensáveis na discriminação entre populações, tem-se:

$$D_i(\tilde{x}) = \ln(p_i) - \frac{1}{2} \left(\tilde{x} - \mu_i \right)' \Sigma^{-1} \left(\tilde{x} - \mu_i \right)$$

ou

$$D_i(\tilde{x}) = \ln(p_i) - \frac{1}{2} \left(\tilde{x}' \Sigma^{-1} \tilde{x} - \tilde{x}' \Sigma^{-1} \mu_i - \mu_i' \Sigma^{-1} \tilde{x} + \mu_i' \Sigma^{-1} \mu_i \right)$$

Sendo $\tilde{x}' \Sigma^{-1} \tilde{x}$ uma constante, ela também deve ser excluída da função, por ser dispensável na discriminação. Além disso, verifica-se na expressão anterior que:

$$\tilde{x}' \Sigma^{-1} \mu_i = \mu_i' \Sigma^{-1} \tilde{x}$$

logo:

$$D_i(\tilde{x}) = \ln(p_i) - \frac{1}{2} \left(-2 \tilde{x}' \Sigma^{-1} \mu_i + \mu_i' \Sigma^{-1} \mu_i \right)$$

de forma que se tenha:

$$\begin{aligned} D_i(\tilde{x}) &= \ln(p_i) + \left(\tilde{x} - \frac{1}{2} \mu_i \right)' \Sigma^{-1} \mu_i \\ &= \ln(p_i) + \tilde{x}' \Sigma^{-1} \mu_i - \frac{1}{2} \mu_i' \Sigma^{-1} \mu_i \\ &= \ln(p_i) + \mu_i' \Sigma^{-1} \tilde{x} - \frac{1}{2} \mu_i' \Sigma^{-1} \mu_i \quad \text{para } i = 1, 2, \dots, g \end{aligned}$$

que é a função discriminante de Anderson (1958). Esta técnica tem por finalidade classificar novos indivíduos, de comportamento desconhecido, em populações já conhecidas. Como exemplo, serão consideradas três populações, em que:

π_1, π_2 e π_3 : populações 1, 2 e 3, respectivamente;

μ_1, μ_2 e μ_3 : vetor de médias dos p caracteres avaliados em π_1, π_2 e π_3 ,

respectivamente;

Σ_1, Σ_2 e Σ_3 : matriz de covariâncias entre os caracteres avaliados em π_1, π_2 e π_3 , respectivamente. A matriz de covariâncias comum é denotada por Σ ;

p_1, p_2 e p_3 : probabilidades, *a priori*, de os indivíduos pertencerem a π_1, π_2 e π_3 , respectivamente; e

\tilde{x} : vetor de variáveis representativas dos caracteres envolvidos na análise.

As funções discriminantes são obtidas por meio de:

$$D_1(\tilde{x}) = \ln(p_1) + \left(\tilde{x} - \frac{1}{2} \mu_1 \right)' \Sigma^{-1} \mu_1$$

$$D_2(\underline{x}) = \ln(p_2) + \left(\underline{x} - \frac{1}{2} \underline{\mu}_2 \right)' \Sigma^{-1} \underline{\mu}_2$$

e

$$D_3(\underline{x}) = \ln(p_3) + \left(\underline{x} - \frac{1}{2} \underline{\mu}_3 \right)' \Sigma^{-1} \underline{\mu}_3$$

De forma genérica, tem-se:

$$D_i(\underline{x}) = \ln(p_i) - \frac{1}{2} \left(-2 \underline{x}' \Sigma^{-1} \underline{\mu}_i + \underline{\mu}_i' \Sigma^{-1} \underline{\mu}_i \right) = k_i + \underline{x}' \Sigma^{-1} \underline{\mu}_i$$

ou ainda:

$$D_i(\underline{x}) = k_i + \alpha_{i1}x_1 + \alpha_{i2}x_2 + \cdots + \alpha_{ip}x_p$$

em que:

$$k_i = \ln(p_i) - \frac{1}{2} \underline{\mu}_i' \Sigma^{-1} \underline{\mu}_i : \text{constante associada à função discriminante;}$$

α_{ij} : coeficiente de ponderação da j -ésima variável ($j = 1, 2, \dots, p$) na i -ésima função discriminante; e

x_j : valor representativo do escore da j -ésima variável da observação que se deseja classificar em uma das populações em estudo.

Classifica-se o t -ésimo indivíduo, com vetor de observações \underline{x} , na população π_i se e somente se $D_i(\underline{x})$ for o maior entre os elementos do conjunto $\{D_1(\underline{x}), D_2(\underline{x}), D_3(\underline{x})\}$.

Utilizando as funções discriminantes e os dados das próprias populações π_1 , π_2 e π_3 , estima-se a taxa de erro aparente apresentada mais adiante. Ela mede a eficiência da função discriminante em classificar os indivíduos, corretamente, nas populações previamente estabelecidas.

Como vimos anteriormente, se for considerada que a densidade associada à π_i seja normal multivariada, então a função discriminante é dada por:

$$D_i(\underline{x}) = -\frac{1}{2} [p \ln(2\pi) + \ln |\Sigma_i|] - \frac{1}{2} \left[\left(\underline{x} - \underline{\mu}_i \right)' \Sigma_i^{-1} \left(\underline{x} - \underline{\mu}_i \right) \right] + \ln(p_i)$$

Supondo a igualdade das matrizes de covariâncias, então os componentes constantes para todo i podem ser retirados e a função discriminante será:

$$D_i(\underline{x}) = \underline{l}_i' \underline{x} - \frac{1}{2} \underline{l}_i' \underline{\mu}_i + \ln(p_i), \text{ para } i = 1, 2, \dots, g,$$

em que,

$$\underline{l}_i = \Sigma^{-1} \underline{\mu}_i.$$

Uma questão importante a ser abordada refere-se à questão da igualdade das matrizes de covariâncias ($\Sigma_1 = \Sigma_2 = \dots = \Sigma_g = \Sigma$). Como se pode ver através da expressão para $D_i(\underline{x})$ sem a pressuposição de igualdade das matrizes de covariância nas g populações, a regra de decisão discriminante torna-se muito mais complicada uma vez que a função discriminante é quadrática em \underline{x} . Com validade da pressuposição, a função discriminante é linear em \underline{x} , o que torna a regra muito mais simples.

A regra de decisão quanto à população de \underline{x} é a seguinte:

Classificar \underline{x} em π_i , se e somente se,

$$D_i(\underline{x}) = \max (D_1(\underline{x}), D_2(\underline{x}), \dots, D_g(\underline{x})).$$

Na prática, raramente se conhece os parâmetros μ_i e Σ . Os processos de obtenção de regras de discriminação continuam os mesmos descritos anteriormente, utilizando-se as estimativas ao invés dos parâmetros.

Supondo a homogeneidade das matrizes de covariâncias, a função discriminante é dada por:

$$D_i(\underline{x}) = \hat{\underline{l}}_i' \underline{x} - \frac{1}{2} \hat{\underline{l}}_i' \bar{\underline{x}}_i + \ln(p_i), \text{ para } i = 1, 2, \dots, g,$$

em que,

$$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_p \end{bmatrix}, \quad \bar{\underline{x}}_i = \begin{bmatrix} \bar{x}_{i1} \\ \bar{x}_{i2} \\ \dots \\ \bar{x}_{ip} \end{bmatrix}, \quad \hat{\underline{l}}_i = S_c^{-1} \bar{\underline{x}}_i$$

S_c é a matriz comum de covariância amostral, dada por:

$$S_c = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2 + \dots + (n_g - 1)S_g}{(n_1 - 1) + (n_2 - 1) + \dots + (n_g - 1)}$$

$$S_c = \frac{\sum_{i=1}^g (n_i - 1)S_i}{\sum_{i=1}^g n_i - g} = \frac{\sum_{i=1}^g (n_i - 1)S_i}{n - g}$$

A observação \underline{x} será classificada em π_i , se e somente se,

$$D_i(\underline{x}) = \max(D_1(\underline{x}), D_2(\underline{x}), \dots, D_g(\underline{x})).$$

A regra anterior pode ser utilizada mesmo para o caso particular de $g = 2$.

Comentário: Pode-se calcular a probabilidade de má classificação para cada grupo. Uma maneira prática é fazer a análise de consistência, isto é, resubstituir os dados para fazer a reclassificação. A soma dos casos desfavoráveis em cada grupo fornece uma estimativa da probabilidade de má classificação para o mesmo. Assim teremos:

$$\hat{P}_i = \frac{m_i}{n_i}, \quad i = 1, 2, \dots, g$$

m_i = número de observação má classificadas em π_i

n_i = número de observações em π_i

A soma de todos os casos desfavoráveis fornece a taxa de erro aparente (TEA), isto é:

$$TEA = \frac{\sum_{i=1}^g m_i}{\sum_{i=1}^g n_i}.$$

Esta taxa é subestimada, uma vez que os mesmos dados são utilizados para obtenção das funções discriminantes e da probabilidade de má classificação. Quando a taxa de erro aparente é alta deve-se concluir que:

- As populações analisadas não são suficientemente diferenciadas para que possam ser distinguidas por meio das funções discriminantes.
- As populações analisadas, apesar de serem diferenciadas, não puderam ser distinguidas em razão da quantidade e qualidade das variáveis consideradas na discriminação.

Um dos caminhos para resolver o problema de subestimação da probabilidade de má classificação consiste em dividir o conjunto de dados em duas partes: uma amostra para obtenção das funções discriminantes e outra para avaliá-las. Entretanto, essa sugestão necessita de uma amostra grande da população (ou grupo) para obter estimativas satisfatórias de má classificação. Há ainda situações em que a função discriminante construída com parte dos dados disponíveis pode estimar coeficientes inadequados à outra amostra.

Um método alternativo para estimar a taxa de erro é a *validação cruzada*. Neste método, omite-se o primeiro indivíduo da análise e gera-se a(s) função(ões) discriminante(s) usando os $(\sum_{i=1}^g n_i - 1)$ indivíduos restantes. Classifica-se então o indivíduo omitido. Este procedimento é repetido para todos os indivíduos $(\sum_{i=1}^g n_i)$. Finalmente, o número de classificações erradas para cada população é contabilizado e a taxa de erro de cada população é computada. A taxa de erro aparente (global) pode ser obtida com as médias ponderadas dessas proporções, considerando as probabilidades *a priori* como pesos. Com este procedimento, tem-se um menor viés. O interessante é que, nesta técnica, a cada iteração, uma função discriminante diferente está sendo avaliada; contudo, com a eliminação de apenas um indivíduo, espera-se que essa alteração não afete significativamente a função discriminante, a menos que este indivíduo seja uma informação muito discrepante da amostra sob análise.

7.3.3. Distância de Mahalanobis entre dois grupos

Para i e j fixados $i \neq j = 1, 2, \dots, g$, estima-se a distância de Mahalanobis entre os grupos i e j por meio da seguinte expressão:

$$D_{ij}^2 = (\bar{\mathbf{x}}_i - \bar{\mathbf{x}}_j)' \mathbf{S}_c^{-1} (\bar{\mathbf{x}}_i - \bar{\mathbf{x}}_j).$$

Quanto maior for os D_{ij}^2 's melhor, pois isto indica que os grupos são bastante distintos e, portanto a classificação de um novo indivíduo num dos grupos se fará com a máxima probabilidade de acerto.

7.3.4. Exemplo

Um exemplo ilustrativo para a obtenção da Função Discriminante de Anderson é apresentado a seguir.

Com o objetivo de ilustrarmos a aplicação da Análise Discriminante foram utilizados dados de duas espécies de abelhas (adaptados de CAMARGO, 1977) e uma terceira população foi obtida por simulação, cujos resultados estão apresentados nas Tabelas 7.4, 7.5 e 7.6.

As três populações foram:

$$\pi_1 = \textit{Partamona testacea}$$

$\pi_2 = \textit{Partamona pseudomusarum}$

$\pi_3 = \text{Simulação}$

Foram considerados seis caracteres quantitativos contínuos, sendo que as medidas de cada variável correspondem a uma escala observada ao microscópio com aumento de 20 vezes. Para se transformar estas medidas em milímetros basta multiplicá-los pelo fator de correção $c = 0,04$.

As variáveis observadas foram:

X_1 = comprimento do clipeo;

X_2 = largura máxima da cabeça;

X_3 = distância máxima interorbital (maior distância entre as órbitas interna dos olhos compostos);

X_4 = comprimento do olho composto;

X_5 = comprimento da área malar (entre a órbita inferior do olho composto e a base da mandíbula, junto ao ângulo lateral do clipeo);

X_6 = comprimento do flagelo.

Tabela 7.4 – Dados da população π_1 - *Partamona testacea*

Amostra	Variáveis					
	X_1	X_2	X_3	X_4	X_5	X_6
1	17,0	62,0	41,5	40,0	4,0	46,0
2	17,5	59,5	40,0	39,0	4,5	45,0
3	18,0	62,0	41,0	41,0	4,5	46,5
4	19,0	64,5	42,5	42,0	4,5	48,0
5	17,0	56,0	38,0	35,5	4,5	44,0
6	17,5	62,0	41,5	41,0	4,5	46,0

7	19,0	61,5	40,0	41,0	4,0	47,5
8	17,0	61,5	41,5	40,0	4,5	45,5
9	18,5	63,0	41,5	41,0	4,0	46,0
10	18,5	64,0	42,0	42,0	4,5	48,5
11	16,0	60,5	41,0	40,0	4,0	45,0
12	17,0	58,0	38,0	39,5	3,5	44,0
13	17,0	59,0	40,0	39,5	4,0	44,0
14	17,5	61,0	41,0	39,0	4,5	44,0
15	17,0	58,5	38,5	40,0	4,0	44,0
16	18,5	61,0	41,0	40,0	4,5	45,5
17	17,5	62,5	42,0	41,0	4,0	46,0
18	17,0	59,5	40,5	39,5	4,0	43,0
19	19,5	66,0	43,5	43,0	4,5	49,0
20	17,0	60,5	41,0	39,0	4,0	45,0
21	17,0	59,0	40,0	39,0	4,5	44,0
22	17,5	59,0	40,0	39,0	4,0	45,0
23	18,5	60,0	41,0	39,0	4,5	46,5
24	19,0	63,0	41,5	41,0	4,5	48,0
25	19,0	65,0	42,0	43,5	4,0	47,0
26	19,5	64,5	43,0	42,5	4,5	48,0
27	16,5	61,0	41,0	40,0	4,5	46,0
28	17,5	60,5	41,5	39,5	4,5	45,0

29	16,5	56,5	39,5	37,0	4,5	43,0
30	17,5	58,5	40,5	38,5	4,5	45,0
31	18,0	62,5	42,5	40,5	4,5	47,0
32	19,0	64,0	41,5	42,0	4,0	47,0

Tabela 7.5 – Dados da população $\pi_2 = \textit{Partamona pseudomusarum}$

Amostra	Variáveis					
	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆
1	16,5	61,0	38,5	40,0	3,0	41,0
2	16,5	57,0	37,5	38,0	3,5	41,5
3	16,0	57,0	37,0	39,0	3,0	39,5
4	16,5	54,0	35,0	36,5	3,0	37,0
5	16,5	60,5	40,0	41,0	3,5	42,5
6	17,0	61,0	40,0	40,0	3,5	43,0
7	17,5	63,0	41,0	42,0	4,0	44,0
8	17,5	62,5	41,0	42,0	3,5	44,0
9	16,5	60,5	40,0	40,0	3,5	43,5
10	16,0	61,0	39,5	41,0	3,5	43,0
11	17,5	59,0	38,5	39,5	3,5	41,0
12	17,0	60,5	40,5	40,5	4,5	43,5
13	17,0	62,0	40,0	41,5	3,5	43,5
14	17,5	62,0	40,0	42,0	3,5	43,5
15	16,5	61,5	40,0	41,5	3,5	42,0

16	15,5	55,0	37,5	36,0	3,5	38,5
17	17,0	58,0	37,5	39,0	3,5	41,0
18	16,5	58,0	38,0	39,0	3,5	41,0
19	16,5	61,0	41,0	40,5	3,5	43,5
20	17,5	62,0	40,5	40,5	3,5	44,0
21	17,0	61,0	40,5	40,5	4,0	43,0
22	17,0	62,0	40,5	41,5	3,5	44,0
23	17,0	59,0	38,5	39,0	4,0	41,5
24	16,5	60,0	40,0	40,0	4,0	41,0
25	16,5	61,5	40,5	41,0	3,5	43,5
26	16,5	61,0	39,5	40,0	3,5	41,0
27	17,0	60,0	40,0	40,0	3,5	43,5
28	16,5	61,0	39,5	41,0	3,5	43,5
29	17,0	61,5	40,5	41,5	3,0	43,0
30	17,0	62,0	40,0	42,0	3,5	42,5
31	15,0	56,0	36,5	37,5	3,0	39,5
32	16,5	60,5	38,5	40,5	3,5	41,5

Tabela 7.6 – Dados da população π_3 = Simulação

Amostra	Variáveis					
	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆
1	16,0	60,0	38,0	40,5	3,0	39,0
2	15,0	57,5	36,0	39,5	3,0	38,5

3	15,0	59,0	37,5	41,0	2,5	39,0
4	14,0	54,0	34,0	38,0	3,0	36,0
5	17,0	59,5	37,0	41,5	2,5	39,5
6	15,0	57,0	37,0	39,5	3,0	39,5
7	16,5	58,0	36,5	40,0	3,5	39,5
8	15,5	56,5	37,0	38,5	3,5	39,0
9	14,5	55,0	35,5	38,0	3,5	37,0
10	16,5	60,0	37,0	40,5	3,0	39,5
11	15,0	57,0	37,0	39,5	3,0	38,0
12	15,0	59,0	37,0	40,5	3,0	39,0
13	15,0	58,0	36,5	40,0	3,0	38,0
14	15,0	58,0	36,0	41,0	2,5	39,5
15	15,5	57,0	37,0	39,5	3,5	38,5
16	16,0	56,5	36,5	39,5	3,0	39,0
17	15,0	58,0	36,5	40,0	2,5	38,0
18	16,0	59,0	38,5	41,5	3,0	42,0
19	16,0	63,0	40,0	43,0	3,0	41,0
20	15,0	56,5	35,0	38,0	3,0	37,5
21	15,5	58,5	38,0	40,5	3,5	38,5
22	15,0	58,0	36,0	40,5	3,0	38,5
23	16,5	61,0	39,0	41,5	3,0	41,5
24	16,0	59,0	37,0	40,5	3,0	39,5

25	15,0	57,5	37,0	39,5	3,5	38,0
26	15,5	59,0	37,5	41,0	3,5	40,0
27	15,5	61,5	39,5	42,0	3,0	41,0
28	15,5	61,5	39,0	42,5	3,0	4,0
29	15,5	58,0	37,0	39,5	2,5	39,0
30	14,5	56,5	36,0	39,0	3,0	38,5
31	16,0	59,0	38,0	40,5	3,5	40,5
32	15,0	56,0	35,0	39,0	2,5	37,5

Temos: $\mathbf{p} = 6$ variáveis

$\mathbf{g} = 3$ grupos

$\mathbf{n}_1 = \mathbf{n}_2 = \mathbf{n}_3 = 32$

- **Vetores de médias**

$$\bar{\mathbf{x}}_1 = \begin{bmatrix} 17,765625 \\ 61,125000 \\ 40,937500 \\ 40,140625 \\ 4,281250 \\ 45,750000 \end{bmatrix}, \bar{\mathbf{x}}_2 = \begin{bmatrix} 16,703125 \\ 60,062500 \\ 39,296875 \\ 40,125000 \\ 3,515625 \\ 42,125000 \end{bmatrix}, \bar{\mathbf{x}}_3 = \begin{bmatrix} 15,4375 \\ 58,2500 \\ 37,0156 \\ 40,1719 \\ 3,0312 \\ 39,0781 \end{bmatrix}$$

- **Matrizes de covariâncias**

$$S_1 = \begin{bmatrix} 0,90297379 & 1,72379030 & 0,71068548 & 1,09043780 & 0,051915323 & 1,2379032 \\ & 5,93548390 & 2,7822581 & 3,7157258 & 0,076612968 & 3,3870968 \\ & & 1,6572581 & 1,5574597 & 0,12298387 & 1,5403226 \\ & & & 2,6973286 & -0,040826613 & 2,0685484 \\ & \text{Simétrica} & & & 0,079637097 & 0,11290323 \\ & & & & & 2,5645161 \end{bmatrix}$$

$$S_2 = \begin{bmatrix} 0,3203125 & 0,76108871 & 0,45388105 & 0,50604839 & 0,069304436 & 0,57862903 \\ & 4,9959677 & 3,03131048 & 3,266129 & 0,2328629 & 3,4516129 \\ & & 2,1751512 & 1,9133065 & 0,22908266 & 2,3326613 \\ & & & 2,3387097 & 0,12701613 & 2,250000 \\ & \text{Simétrica} & & & 0,10458669 & 0,23185484 \\ & & & & & 3,0483871 \end{bmatrix}$$

$$S_3 = \begin{bmatrix} 0,44758065 & 0,7983871 & 0,49294355 & 0,46270161 & 0,010080645 & 0,593750 \\ & 3,6935484 & 2,2217742 & 2,2217742 & -0,072580645 & 2,0685484 \\ & & 1,7174899 & 1,3520665 & 0,055947581 & 1,4745464 \\ & & & 1,5420867 & -0,0781250 & 1,3571069 \\ & \text{Simétrica} & & & 0,11189516 & 0,00554435 \\ & & & & & 1,7598286 \end{bmatrix}$$

- **Obtenção das Funções Discriminantes de Anderson**

Vamos supor a distribuição normal multivariada, homogeneidade das matrizes de covariâncias, custos idênticos de má classificação e probabilidades a priori iguais ($p_1 = p_2 = p_3 = 1/3$).

A matriz comum de covariância amostral é dada por:

$$S_c = \frac{\sum_{j=1}^g (n_j - 1) S_i}{\sum_{j=1}^g (n_j - 1)} = \frac{\sum_{j=1}^g (n_j - 1) S_j}{\sum_{j=1}^g n_j - g} = \frac{\sum_{j=1}^g (n_j - 1) S_j}{n - g}$$

No exemplo, temos $g = 3$ e $n_1 = n_2 = n_3 = 32$. Neste caso, temos que:

$$S_c = \frac{S_1 + S_2 + S_3}{3}$$

Fato: S_c é a matriz de covariância “dentro”. Assim, S_c nada mais é que a matriz de covariância residual, que pode ser obtida através de uma MANOVA, supondo no exemplo $g = 3$ tratamentos

e 32 repetições, no modelo inteiramente casualizado. Neste caso, $S_c = \frac{E}{n_e}$,

em que,

E = matriz de somas de quadrados e de produtos residuais,

n_e = número de graus de liberdade do resíduo.

$$S_c = \begin{bmatrix} 0,55695565 & 1,09442204 & 0,55250336 & 0,68640793 & 0,04376680 & 0,80342741 \\ & 4,8750000 & 2,67237903 & 3,06787633 & 0,07896507 & 2,96908603 \\ & & 1,84996640 & 1,60761090 & 0,13600470 & 1,78251010 \\ & \text{Simétrica} & & 2,19270833 & 0,00268817 & 1,89188510 \\ & & & & 0,09870632 & 0,11676747 \\ & & & & & 2,45757727 \end{bmatrix}$$

e sua inversa S_c^{-1} é:

$$S_c^{-1} = \begin{bmatrix} 3,9731249 & -1,1087852 & 1,2898645 & 0,2239671 & -1,4785608 & -0,99704192 \\ & 3,5952817 & -2,2984413 & -2,8574907 & 1,0546546 & -0,16438247 \\ & & 4,0246875 & 0,65556366 & -3,2129114 & -0,91601434 \\ & & & 4,3790779 & 1,8195269 & -0,55401234 \\ & \text{Simétrica} & & & 14,999156 & -0,57379537 \\ & & & & & 2,0495997 \end{bmatrix}$$

$$\hat{\tilde{z}}_1 = \begin{bmatrix} 12,659804 \\ -11,735831 \\ 17,835467 \\ 14,374641 \\ 17,670595 \\ 3,813935 \end{bmatrix}, \hat{\tilde{z}}_2 = \begin{bmatrix} 12,243059 \\ -10,773788 \\ 18,074271 \\ 16,644022 \\ 13,960027 \\ -0,431046 \end{bmatrix}, \hat{\tilde{z}}_3 = \begin{bmatrix} 10,046354 \\ -10,787614 \\ 15,804457 \\ 21,056219 \\ 15,816970 \\ -2,774488 \end{bmatrix}$$

$$\hat{\tilde{z}}_1' \bar{x}_1 = 1064,84290$$

$$\hat{\tilde{z}}_2' \bar{x}_2 = 966,42084$$

$$\hat{\tilde{z}}_3' \bar{x}_3 = 897,11451$$

$$p_1 = p_2 = p_3 = \frac{1}{3}. \text{ Logo, } \ln(p_i) = \ln \frac{1}{3} = -1,0986123$$

As funções discriminantes são:

$$D_j(\tilde{x}) = \hat{\tilde{z}}_j' \tilde{x} - \frac{1}{2} \hat{\tilde{z}}_j' \bar{x}_j + \ln(p_j)$$

Para a população π_1 :

$$D_1(\tilde{x}) = 12,659804 x_1 - 11,735831 x_2 + 17,835467 x_3 + 14,374641 x_4 + 17,670595 x_5 + 3,813935 x_6 - 533,52007$$

Para a população π_2 :

$$D_2(\tilde{x}) = 12,243059 x_1 - 10,773788 x_2 + 18,074271 x_3 + 16,644022 x_4 + 13,960027 x_5 - 0,431046 x_6 - 484,30903$$

Para a população π_3 :

$$D_3(\underline{x}) = 10,046354 x_1 - 10,787614 x_2 + 15,804457 x_3 + 21,056219 x_4 + 15,816970 x_5 - 2,774488 x_6 - 449,65587$$

- **Classificação**

Por exemplo, se quisermos classificar a primeira observação de π_1 através das funções anteriores, teremos:

$$\underline{x}_1' = \begin{bmatrix} 17,0 & 62,0 & 41,5 & 40,0 & 4,0 & 46,0 \end{bmatrix}$$

$$D_1(\underline{x}_1) = 515,35597$$

$$D_2(\underline{x}_1) = 507,70321$$

$$D_3(\underline{x}_1) = 486,07521$$

Como $\max \{D_1(\underline{x}_1), D_2(\underline{x}_1), D_3(\underline{x}_1)\} = D_1(\underline{x}_1)$, concluímos que a observação $\underline{x}_1 \in \pi_1$, o que realmente é verdadeiro.

Para a primeira observação de π_2 , teremos:

$$\underline{x}_2' = \begin{bmatrix} 16,5 & 61,0 & 38,5 & 40,0 & 3,0 & 41,0 \end{bmatrix}$$

$$D_1(\underline{x}_2) = 430,51523$$

$$D_2(\underline{x}_2) = 446,32786$$

$$D_3(\underline{x}_2) = 442,48175$$

Como $\max \{D_1(\underline{x}_2), D_2(\underline{x}_2), D_3(\underline{x}_2)\} = D_2(\underline{x}_2)$, concluímos que a observação $\underline{x}_2 \in \pi_2$.

- **Análise de consistência**

Utilizando o procedimento anterior, reclassificamos as 96 observações das populações π_1, π_2 e π_3 , e verificamos que apenas a observação número 18 da população π_1 foi classificada erroneamente, sendo alocada em π_2 . Todas as demais foram classificadas corretamente. Os resultados estão apresentados na Tabela 7.7.

Tabela 7.7 – Distribuição dos elementos segundo a designação a priori e através da função discriminante

Designação a priori	Função discriminante			Total	Classificação errada (%)
	População 1	População 2	População 3		
População 1	31	1	0	32	3,12
População 2	0	32	0	32	0,00
População 3	0	0	32	32	0,00
Total	31	33	32	96	

- **Taxa de erro aparente: TEA**

$$n_1 = n_2 = n_3 = 32, \sum_{i=1}^3 n_i = 96$$

$$m_1 = 1, m_2 = 0, m_3 = 0$$

$$TEA = \frac{1+0+0}{96} = \frac{1}{96} = 0,0104$$

$$\text{ou } TEA = 1,04\%$$

Esta taxa é subestimada.

7.4. EXEMPLO UTILIZANDO O SAS (Dados da Tabela 7.1)

(a) Programa:

```
options nodate nonumber;
data racas;
input racas $ x1 x2;
cards;
A 6.36 5.24
A 5.92 5.12
A 5.92 5.36
A 6.44 5.64
A 6.40 5.16
A 6.56 5.56
A 6.64 5.36
A 6.68 4.96
A 6.72 5.48
A 6.76 5.60
A 6.72 5.08
B 6.00 4.88
B 5.60 4.64
B 5.64 4.96
B 5.76 4.80
B 5.96 5.08
B 5.72 5.04
B 5.64 4.96
B 5.44 4.88
B 5.04 4.44
B 4.56 4.04
B 5.48 4.20
B 5.76 4.80
;
proc discrim data=racas method=normal pool=test manova wcov pcov
bcov list listerr crosslisterr;
class racas;
var x1 x2;
run;
data racasm;
set racas;
d=5.794521*x1+2.871023*x2;
proc print;
run;
```

(b) Resultado da Análise:

```

The SAS System
The DISCRIM Procedure

Total Sample Size      23      DF Total      22
Variables              2      DF Within Classes  21
Classes                2      DF Between Classes  1

Number of Observations Read      23
Number of Observations Used      23
Class Level Information
Variable
racas  Name      Frequency      Weight      Proportion      Prior
      A      A      11      11.0000      0.478261      0.500000
      B      B      12      12.0000      0.521739      0.500000

The SAS System
The DISCRIM Procedure
Within-Class Covariance Matrices
racas = A,      DF = 10

Variable      x1      x2
x1      0.0912872727      0.0112581818
x2      0.0112581818      0.0526254545
-----
racas = B,      DF = 11

Variable      x1      x2
x1      0.1603272727      0.1074181818
x2      0.1074181818      0.1116606061
-----

The SAS System
The DISCRIM Procedure
Pooled Within-Class Covariance Matrix,      DF = 21
Variable      x1      x2
x1      0.1274510823      0.0616277056
x2      0.0616277056      0.0835486291

Between-Class Covariance Matrix,      DF = 1
Variable      x1      x2
x1      0.4182363980      0.2727327720
x2      0.2727327720      0.1778495732

Within Covariance Matrix Information
Covariance      Natural Log of the
racas      Matrix Rank      Determinant of the
A      2      Covariance Matrix
B      2      -5.05717
Pooled      2      -4.98345

The SAS System
The DISCRIM Procedure
Test of Homogeneity of Within Covariance Matrices
Notation: K      = Number of Groups

P      = Number of Variables

N      = Total Number of Observations - Number of Groups

N(i) = Number of Observations in the i'th Group - 1

```

$$V = \frac{\sum_{i=1}^N | \text{Within SS Matrix}(i) |^{N(i)/2}}{N/2}$$

$$RHO = 1.0 - \frac{\sum_{i=1}^N \frac{1}{N(i)} - \frac{1}{N}}{\frac{2P + 3P - 1}{6(P+1)(K-1)}}$$

$$DF = .5(K-1)P(P+1)$$

Under the null hypothesis: $-2 RHO \ln \left[\frac{\sum_{i=1}^N \frac{PN/2}{N(i)}}{V} \right]$

is distributed approximately as Chi-Square(DF).

Chi-Square	DF	Pr > ChiSq
4.147930	3	0.2459

Since the Chi-Square value is not significant at the 0.1 level, a pooled covariance

matrix will be used in the discriminant function.

Reference: Morrison, D.F. (1976) Multivariate Statistical Methods p252.

The SAS System

The DISCRIM Procedure

Pairwise Generalized Squared Distances Between Groups

$$D(i|j) = \frac{1}{2} (\bar{X}_i - \bar{X}_j)' \text{COV}^{-1} (\bar{X}_i - \bar{X}_j)$$

Generalized Squared Distance to racas

From		
racas	A	B
A	0	7.01854
B	7.01854	0

The SAS System

The DISCRIM Procedure

Multivariate Statistics and Exact F Statistics

	S=1	M=0	N=9		
Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.34268764	19.18	2	20	<.0001
Pillai's Trace	0.65731236	19.18	2	20	<.0001
Hotelling-Lawley Trace	1.91810930	19.18	2	20	<.0001
Roy's Greatest Root	1.91810930	19.18	2	20	<.0001

Linear Discriminant Function

$$\text{Constant} = -\frac{1}{2} \sum_j \bar{X}_j' \text{COV}^{-1} \bar{X}_j \quad \text{Coefficient Vector} = \text{COV}^{-1} \sum_j \bar{X}_j$$

Linear Discriminant Function for racas

Variable	A	B
Constant	-208.90747	-159.66832
x1	30.96121	25.16664
x2	40.88119	38.01025

The SAS System

The DISCRIM Procedure

Classification Results for Calibration Data: WORK.RACAS
 Resubstitution Results using Linear Discriminant Function

Generalized Squared Distance Function

$$D_j(X) = \frac{1}{2} (X - \bar{X}_j)' \text{COV}_j^{-1} (X - \bar{X}_j)$$

Posterior Probability of Membership in Each racas

$$\Pr(j|X) = \frac{\exp(-.5 D_j(X))}{\sum_k \exp(-.5 D_k(X))}$$

Posterior Probability of Membership in racas

Obs	From racas	Classified into racas	A	B
1	A	A	0.9345	0.0655
2	A	B	0.4413	0.5587
3	A	A	0.6113	0.3887
4	A	A	0.9862	0.0138
5	A	A	0.9346	0.0654
6	A	A	0.9913	0.0087
7	A	A	0.9903	0.0097
8	A	A	0.9761	0.0239
9	A	A	0.9956	0.0044
10	A	A	0.9976	0.0024
11	A	A	0.9864	0.0136
12	B	B	0.3866	0.6134
13	B	B	0.0302	0.9698
14	B	B	0.0897	0.9103
15	B	B	0.1109	0.8891
16	B	B	0.4703	0.5297
17	B	B	0.1646	0.8354
18	B	B	0.0897	0.9103
19	B	B	0.0240	0.9760
20	B	B	0.0007	0.9993
21	B	B	0.0000	1.0000
22	B	B	0.0044	0.9956
23	B	B	0.1109	0.8891

* Misclassified observation

The SAS System

The DISCRIM Procedure

Classification Summary for Calibration Data: WORK.RACAS
 Resubstitution Summary using Linear Discriminant Function

Generalized Squared Distance Function

$$D_j(X) = \frac{1}{2} (X - \bar{X}_j)' \text{COV}_j^{-1} (X - \bar{X}_j)$$

Posterior Probability of Membership in Each racas

$$\Pr(j|X) = \frac{\exp(-.5 D_j^2(X))}{\sum_k \exp(-.5 D_k^2(X))}$$

Number of Observations and Percent Classified into racas

From racas	A	B	Total
A	10 90.91	1 9.09	11 100.00
B	0 0.00	12 100.00	12 100.00
Total	10 43.48	13 56.52	23 100.00
Priors	0.5	0.5	

Error Count Estimates for racas

	A	B	Total
Rate	0.0909	0.0000	0.0455
Priors	0.5000	0.5000	

The SAS System

The DISCRIM Procedure

Classification Results for Calibration Data: WORK.RACAS
Cross-validation Results using Linear Discriminant Function

Generalized Squared Distance Function

$$D_j^2(X) = (X - \bar{X}_j)' \text{COV}^{-1}(X) (X - \bar{X}_j)$$

Posterior Probability of Membership in Each racas

$$\Pr(j|X) = \frac{\exp(-.5 D_j^2(X))}{\sum_k \exp(-.5 D_k^2(X))}$$

Posterior Probability of Membership in racas

Obs	From racas	Classified into racas		A	B
2	A	B	*	0.3515	0.6485
3	A	B	*	0.3646	0.6354
16	B	A	*	0.5189	0.4811

* Misclassified observation

The SAS System

The DISCRIM Procedure

Classification Summary for Calibration Data: WORK.RACAS
Cross-validation Summary using Linear Discriminant Function

Generalized Squared Distance Function

$$D_j^2(X) = (X - \bar{X}_j)' \text{COV}^{-1}(X) (X - \bar{X}_j)$$

Posterior Probability of Membership in Each racas

$$\Pr(j|X) = \frac{\exp(-.5 \sum_j D_j^2(X))}{\sum_k \exp(-.5 \sum_k D_k^2(X))}$$

Number of Observations and Percent Classified into racas

From racas	A	B	Total
A	9 81.82	2 18.18	11 100.00
B	1 8.33	11 91.67	12 100.00
Total	10 43.48	13 56.52	23 100.00
Priors	0.5	0.5	

Error Count Estimates for racas

	A	B	Total
Rate	0.1818	0.0833	0.1326
Priors	0.5000	0.5000	

The SAS System

Obs	racas	x1	x2	d
1	A	6.36	5.24	51.8973
2	A	5.92	5.12	49.0032
3	A	5.92	5.36	49.6922
4	A	6.44	5.64	53.5093
5	A	6.40	5.16	51.8994
6	A	6.56	5.56	53.9749
7	A	6.64	5.36	53.8643
8	A	6.68	4.96	52.9477
9	A	6.72	5.48	54.6724
10	A	6.76	5.60	55.2487
11	A	6.72	5.08	53.5240
12	B	6.00	4.88	48.7777
13	B	5.60	4.64	45.7709
14	B	5.64	4.96	46.9214
15	B	5.76	4.80	47.1574
16	B	5.96	5.08	49.1201
17	B	5.72	5.04	47.6146
18	B	5.64	4.96	46.9214
19	B	5.44	4.88	45.5328
20	B	5.04	4.44	41.9517
21	B	4.56	4.04	38.0219
22	B	5.48	4.20	43.8123
23	B	5.76	4.80	47.1574

A seguir é apresentado um teste para igualdade de matrizes de covariâncias.

TESTE PARA IGUALDADE DE MATRIZES DE COVARIÂNCIAS

Bartlett (1937) propôs a seguinte extensão do teste univariado para igualdade de variâncias para testar

$$H_0 : \sum_1 = \dots = \sum_g$$

Sejam:

g = número de grupos (ou g populações)

p = número de variáveis

$N_i = n_i - 1, i = 1, \dots, g$

S_i é o estimador não viesado de \sum_i com N_i graus de liberdade.

Sob H_0 , a matriz comum de covariância amostral é

$$S = \frac{1}{\sum_{i=1}^g N_i} \sum_{i=1}^g N_i S_i.$$

A estatística do teste é

$$M = \left(\sum_{i=1}^g N_i \right) \ln |S| - \sum_{i=1}^g N_i \ln |S_i|.$$

Korin (1969) e Pearson e Hartley (1972) dão tabelas ao nível de 5% para o valor crítico de M no caso de n_i iguais. Para n_i desiguais, Box (1949) mostrou que, com um fator de escala

$$C^{-1} = 1 - \frac{2p^2 + 3p - 1}{6(p+1)(g-1)} \left(\sum_{i=1}^g \frac{1}{N_i} - \frac{1}{\sum_{i=1}^g N_i} \right),$$

MC^{-1} tem aproximadamente distribuição de qui-quadrado com $v = \frac{1}{2}(g-1)p(p+1)$ graus de liberdade. A aproximação parece ser boa se g e p não excedem quatro ou cinco, e cada N_i é maior que 20.

Regra de decisão: Se $MC^{-1} \geq \chi_{\alpha}^2(v)$, rejeita-se H_0 . Caso contrário, não se rejeita H_0 .

O teste é muito sensível à suposição de normalidade e não deve ser usado quando há razões para duvidar da normalidade multivariada. Veja Morrison (1976) para mais detalhes.

O mesmo teste anterior utilizando a notação do SAS (O **PROC DISCRIM** tem esta opção) é dado a seguir:

K = número de grupos

P = número de variáveis

$N_i = n_i - 1$, sendo n_i o número de observações do i -ésimo grupo

$$N = \sum_{i=1}^K N_i = \sum_{i=1}^K n_i - K$$

Seja A_i a matriz de soma de quadrados e de produtos para o grupo i . Então,

$$S_i = \frac{A_i}{N_i} \Rightarrow A_i = N_i S_i$$

Sob H_0 , a matriz comum de covariância amostral é $S = \frac{A_1 + A_2 + \dots + A_K}{\sum_{i=1}^K N_i}$

Sejam:

$$V = \frac{\prod_{i=1}^K |N_i S_i|^{N_i/2}}{|NS|^{N/2}}$$

$$RH_0 = 1 - \left[\sum_{i=1}^K \frac{1}{N_i} - \frac{1}{N} \right] \frac{2P^2 + 3P - 1}{6(P+1)(K-1)}$$

Número de graus de liberdade: $v = \frac{1}{2}(K-1)P(P+1)$

Sob H_0 ,

$$-2RH_0 \ln \left[\frac{N^{PN/2} \cdot V}{\prod_{i=1}^K N_i^{PN_i/2}} \right]$$

tem aproximadamente distribuição de qui-quadrado com v graus de liberdade.

Observação: Com aplicação deste teste aos dados da Tabela 7.1, o resultado obtido pelo SAS é o seguinte:

$$\chi^2_{\text{Calc.}} = 4,147930 \text{ com 3 graus de liberdade. Valor-P} = 0,2459$$

Chi-Square	DF	Pr > ChiSq
4.147930	3	0.2459

Since the Chi-Square value is not significant at the 0.1 level, a pooled covariance matrix will be used in the discriminant function. Reference: Morrison, D.F. (1976) Multivariate Statistical Methods p.252.

7.5. Análise Discriminante baseada em componentes principais

A análise de componentes principais, para fins de discriminação de populações, visa estabelecer funções que permitam discriminar graficamente um conjunto de indivíduos distribuídos em grupos (ou populações) previamente conhecidos, denominadas funções

discriminantes ou componentes principais populacionais. O número de componentes principais necessário para caracterizar as diferenças entre os grupos é igual ao mínimo entre $(g - 1, p)$, em que g representa o número de grupos (ou populações) e p o número de caracteres. Os componentes principais discriminantes são obtidos em ordem decrescente de importância. O primeiro explica o máximo da variância entre os grupos, e o segundo, ortogonal ao primeiro, o máximo da variância remanescente, e assim sucessivamente, até se esgotar a variância na matriz $S^{-1}T$, em que S é a matriz de covariâncias conjunta dentro de grupos e T , a matriz de covariâncias entre grupos. Isto nada mais é que variáveis canônicas, que são funções discriminantes ótimas, assunto já apresentado anteriormente.

Para essa análise, realizada a partir da matriz de covariâncias entre populações (T), são obtidos componentes principais, independentes entre si e capazes de reterem, em ordem de estimação, o máximo da variância originalmente existente entre as populações. Obtidas as funções que melhor discriminam as populações, são estimados os escores a partir dos valores dos indivíduos que pertencem a cada população, de forma que se possa, por meio de uma análise de dispersão gráfica, avaliar as dissimilaridades entre estes indivíduos e as populações a que pertencem.

Para análise discriminante com base na técnica dos componentes principais, em que se ignora o delineamento experimental ou não há condições de obter estimativas da matriz de dispersão residual, considera-se o modelo:

$$Y_{ijk} = \mu_k + P_{ik} + I_{ijk}$$

em que:

Y_{ijk} : valor, para uma variável k , do j -ésimo indivíduo dentro da i -ésima população, ou grupo; ($k = 1, 2, \dots, p$);

μ_k : média geral da k -ésima variável;

P_{ik} : efeito da i -ésima população ($i = 1, 2, \dots, g$), considerando a variável k ; e

I_{ijk} : efeito do j -ésimo indivíduo dentro da i -ésima população ($j = 1, 2, \dots, n_i$), considerando a variável k .

A partir dessas informações são obtidas as médias dentro de cada grupo, dadas por:

$$X_{ik} = \frac{\sum_{j=1}^{n_i} Y_{ijk}}{n_i}$$

em que n_i é o número de indivíduos dentro da população i .

Os valores a serem utilizados em análises gráficas são resultantes da padronização, feita por meio de:

$$x_{ik} = \frac{X_{ik}}{\hat{\sigma}_{xk}}$$

Obtém-se a matriz de covariâncias (B), com base nas médias destes grupos, ou, quando os dados são padronizados, a matriz de correlação (R). A partir da matriz R (ou B) obtêm-se os autovalores (λ_k) e os autovetores correspondentes (α_k), como descrito a seguir:

$$\det(R - \lambda I) = 0 \Rightarrow \lambda_1, \lambda_2, \dots, \lambda_p$$

e

$$(R - \lambda_k I) \alpha_k = 0, \text{ com } k = 1, 2, \dots, p.$$

O número de autovalores não nulos da matriz R é dado por:

$$\eta \leq \min \{p, f\}$$

sendo **p** o número de variáveis consideradas na análise e **f** (**f** = **g**-1) os graus de liberdade associados às variâncias e covariâncias que deram origem a esta matriz.

Nesta análise, a importância relativa (IR) de cada componente é calculada por meio da expressão:

$$IR_k = \frac{\lambda_k}{\text{Traço}(R)}$$

São obtidos os escores considerando as médias dos grupos como referência e também os valores dos indivíduos. Admitindo que existam η componentes principais, considera-se que cada um seja estabelecido por:

$$CP_1 = a_{11}x_1 + a_{21}x_2 + \dots + a_{p1}x_p$$

...

$$CP_\eta = a_{1\eta}x_1 + a_{2\eta}x_2 + \dots + a_{p\eta}x_p$$

Assim, utilizando o primeiro componente principal como exemplo, serão obtidos os escores a partir da expressão:

$$CP_{1i} = a_{11}x_{i1} + a_{21}x_{i2} + \dots + a_{p1}x_{ip}$$

em que x_{ik} representa a média padronizada da população i, para a variável k.

Neste caso, verifica-se que:

$$V(CP_\ell) = \lambda_\ell, \text{ para } \ell = 1, 2, \dots, \eta.$$

Também são calculados os escores para cada indivíduo. Utilizando o primeiro componente principal como exemplo, estes escores são obtidos por:

$$CP_{1ij} = a_{11}y_{ij1} + a_{21}y_{ij2} + \dots + a_{p1}y_{ijp}$$

Os valores de y são obtidos por meio de:

$$y_{ijk} = \frac{Y_{ijk}}{\hat{\sigma}_{xk}}$$

Verifica-se, portanto, que não se trata de uma padronização desses valores, mas uma mudança de escala, tornando-os apropriados para a dispersão gráfica. Nessa análise também deve ser identificada a ordem de variáveis associadas aos maiores elementos identificados no último até o primeiro autovetor. As variáveis de maiores pesos nos últimos autovetores são consideradas de menor importância para estudo da diversidade entre as populações. Normalmente, consideram-se os últimos autovetores associados a autovalores da matriz de correlação inferior a 0,7. As variáveis de maiores pesos nos primeiros autovetores são consideradas de maior importância para o estudo da diversidade quando o autovalor explica uma fração considerável da variação disponível, normalmente limitado em valor mínimo de 80%.

O conjunto de cargas totais associado às variáveis é obtido considerando que:

$$\tilde{C}'_{\ell} = \alpha'_{\ell} R$$

em que \tilde{C}'_{ℓ} é o vetor linha de ordem p, cujos elementos, denotados $C_{k\ell}$, são utilizados para obtenção da estimativa da carga da variável k no ℓ -ésimo componente principal, por meio de:

$$\text{Carga}_{k\ell} = \frac{C_{k\ell}}{\sqrt{\lambda_{\ell}}}$$

Ilustração

Será considerada, como ilustração, a obtenção de funções discriminantes, baseadas em componentes principais, a partir das médias padronizadas de 15 genótipos de milho-pipoca, avaliados em relação aos caracteres altura da planta (AP, em metros), da espiga (AE, em metros), número de espigas (NE), prolificidade (PROL), capacidade de expansão (CE), proporção de plantas quebradas (QUE), proporção de plantas doentes (DOEN), peso de cem grãos (PCG) e produção de grãos (PROD, em kg por parcela). Considera-se, neste caso, o estudo de três populações, nas quais a população 1 é representada pelos cultivares de 1 a 5, a população 2, pelos cultivares de 6 a 10, e a população 3, pelos cultivares de 11 a 15, conforme especificado na Tabela 7.8.

As médias das populações estão representadas na Tabela 7.9. São também apresentados os valores padronizados a serem utilizados na obtenção das funções discriminantes.

Para a análise considerada, têm-se nove características ($p = 9$), porém os graus de liberdade associados às variâncias é $g - 1 = 2$ (g é o número de populações avaliadas, que, no exemplo, é igual a 3).

A partir dos valores padronizados, é obtida a matriz de correlação mostrada a seguir:

$$R = \begin{bmatrix} 1 & 0,9966 & 0,5354 & 0,5917 & -0,8263 & 0,9981 & 0,9204 & 0,3628 & 0,7800 \\ & 1 & 0,6028 & 0,6558 & -0,8696 & 0,9998 & 0,9493 & 0,2852 & 0,7262 \\ & & 1 & 0,9977 & -0,9181 & 0,5861 & 0,8230 & -0,5928 & -0,1108 \\ & & & 1 & -0,9430 & 0,6400 & 0,8598 & -0,5365 & -0,0428 \\ & & & & 1 & -0,8592 & -0,9807 & 0,2251 & -0,2921 \\ & & & & & 1 & 0,9426 & 0,3051 & 0,7403 \\ & & & & & & 1 & -0,0304 & 0,4733 \\ & & & & & & & 1 & 0,8661 \\ & & & & & & & & 1 \end{bmatrix}$$

Tabela 7.8 - Média de 15 genótipos de milho-pipoca, pertencentes a três populações, para nove características agronômicas*

Popula- ções	AP	AE	NE	PROL	CE	QUE	DOEN	PCG	PROD
1	1,82	1,20	33,84	1,06	14,17	0,32	0,19	22,01	3339,00
1	2,90	1,43	24,46	1,06	14,88	0,27	0,15	21,86	3950,33
1	2,52	1,16	29,81	1,60	13,63	0,23	0,15	17,87	3512,67
1	2,02	0,97	24,94	1,08	11,96	0,33	0,16	16,40	2845,33
1	3,48	1,53	39,37	1,18	13,71	0,38	0,10	21,93	3022,67
2	1,39	0,83	23,94	0,73	25,88	0,11	0,08	14,08	2142,00
2	1,62	0,67	20,64	0,94	35,66	0,12	0,09	14,63	1806,33
2	1,70	0,86	27,32	0,70	31,26	0,09	0,08	14,73	1981,33
2	1,35	0,87	20,51	0,90	34,62	0,09	0,09	14,03	1734,67
2	1,60	1,11	24,74	0,84	28,68	0,10	0,10	13,60	1818,33
3	1,71	1,05	19,87	0,62	28,84	0,20	0,13	36,17	4105,67
3	2,04	1,25	18,84	0,63	28,92	0,26	0,08	33,63	2901,33
3	2,25	1,08	15,76	0,60	37,75	0,17	0,09	31,77	4343,67
3	2,39	0,81	16,80	0,72	30,65	0,21	0,10	31,38	3565,33
3	2,08	1,09	21,27	0,53	31,29	0,18	0,10	31,00	3811,67
DP	0,5824	0,2359	6,4005	0,2841	9,1306	0,0937	0,0345	8,2426	906,9794

* Fonte: Cruz, et al. (2014)

Tabela 7.9 - Médias originais e padronizadas de populações de milho-pipoca, para nove características agronômicas

Popula- ções	AP	AE	NE	PROL	CE	QUE	DOEN	PCG	PROD
1	2,548	1,258	30,484	1,196	13,670	0,306	0,150	20,014	3334,000
2	1,532	0,868	23,430	0,822	31,220	0,102	0,088	14,214	1896,532
3	2,094	1,056	18,508	0,620	31,490	0,204	0,100	32,790	3745,534
DP*	0,5090	0,1950	6,0195	0,2922	10,2113	0,1020	0,0329	9,5038	970,7800
Médias padronizadas									
1	5,006	6,451	5,064	4,093	1,339	3,000	4,559	2,106	3,434
2	3,010	4,451	3,892	2,813	3,057	1,000	2,675	1,496	1,954
3	4,114	5,415	3,075	2,122	3,084	2,000	3,039	3,450	3,858

* Desvio padrão calculado com os três valores

Assim, são estimadas duas funções discriminantes, cujas variâncias são representadas pelos autovalores de R, e os coeficientes de ponderação, representados pelos elementos dos autovetores associados a estes autovalores, mostrados na Tabela 7.10. Neste exemplo, em que foram consideradas apenas 3 populações, verifica-se, como esperado, que os dois primeiros componentes explicam 100% da variação disponível. Também constata-se que, à exceção de PCG no primeiro componente, todas as variáveis têm peso com magnitude relativamente próxima, e alguns apresentam sinais contrários. A importância de PCG é maior no segundo componente, indicando que, ao utilizar os dois componentes, todas as variáveis terão pesos consideráveis na discriminação das populações em estudo. A análise das cargas fatoriais também permite as mesmas conclusões obtidas pelos elementos dos autovetores.

Na Figura 7.1 está representada a dispersão dos centróides das populações estudadas. Estes centróides são estabelecidos pelos escores dos dois primeiros componentes principais. Percebe-se que eles se encontram bem dispersos, indicando ser possível discriminá-las a partir das nove características agronômicas mensuradas.

Na Figura 7.2 é apresentada a dispersão dos genótipos em torno dos centróides, cujas coordenadas foram estabelecidas pelas funções discriminantes, obtidas com base em componentes principais, e estimadas com o objetivo de maximizar as diferenças entre as populações analisadas. As populações 2 e 3 mostram ser mais homogêneas, pois seus cultivares estão bem próximos aos seus respectivos centróides. A população 1 é mais heterogênea, porém mantém características que a tornam possível de ser distinguida das outras duas consideradas na análise.

Tabela 7.10 - Estimativas de autovalores e coeficientes de ponderação das características altura da planta (AP, em metros), da espiga (AE, em metros), número de espigas (NE), prolificidade (PROL), capacidade de expansão (CE), proporção de plantas quebradas (QUE), proporção de plantas doentes (DOEN), peso de cem grãos (PCG) e produção de grãos (PROD, em kg por parcela), avaliados em cultivares de milho-pipoca

λ_j	$\lambda_j(\%)$ acumu- lada	AP	AE	NE	PROL	CE	QUE	DOEN	PCG	PROD
Autovetores										
6,2598	69,55	0,3785	0,3877	0,3111	0,3275	-0,3851	0,3856	0,3986	0,0176	0,2149
2,7402	100,00	0,1941	0,1466	-0,3792	-0,3463	0,1618	0,1588	-0,0450	0,6035	0,5094
Cargas fatoriais										
-	-	0,9469	0,9701	0,7784	0,8194	-0,9634	0,9648	0,9972	0,0441	0,5376
-	-	0,3213	0,2427	-0,6277	-0,5732	0,2678	0,2628	-0,0744	0,9990	0,8432

Escores	
C1	C2
10,545	2,088
5,718	1,218
6,993	4,418
V(C1)=6,26 V(C2)=2,74	
Cov(C1,C2) =0	

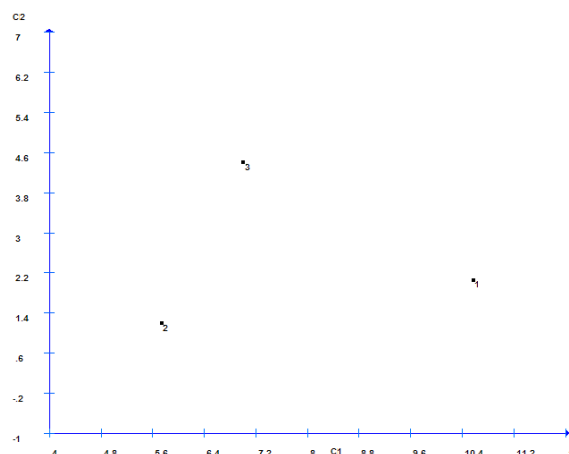


Figura 7.1 - Dispersão gráfica dos centróides de três populações de milho-pipoca, em relação a funções discriminantes, estimadas com base em componentes principais, estabelecidos pela combinação linear entre nove características agrônômicas.

Escore	
C1	C2
10,43	1,82
10,64	3,29
10,50	1,40
9,32	1,67
11,83	2,25
5,65	1,27
5,28	1,21
5,77	1,24
5,34	1,18
6,55	1,18
7,30	4,50
7,25	4,19
6,64	5,00
6,76	4,15
7,02	4,25

$V(C1)=4,79$ $V(C2)=2,15$
 $Cov(C1,C2) =0$

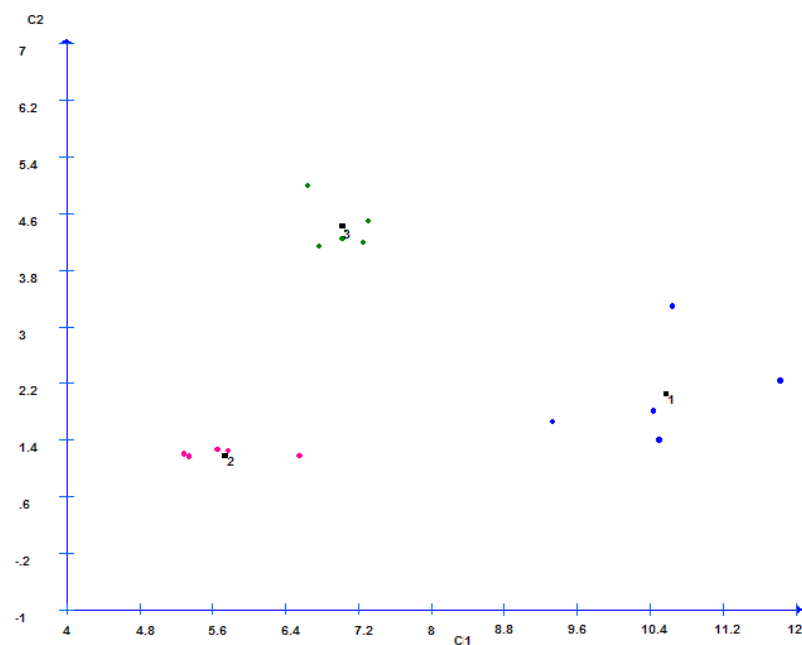


Figura 7.2 - Dispersão gráfica de genótipos em torno dos centróides de três populações de milho-pipoca, em relação a funções discriminantes, estimadas com base em componentes principais, estabelecidos pela combinação linear entre nove características agronômicas.

7.6. EXERCÍCIOS

1. Considerando apenas os dados das espécies *Iris setosa*, *Iris versicolour*, apresentados na Tabela 7.11, pede-se:
 - a) Obter a Função Discriminante Linear Amostral de Fisher;
 - b) Estabelecer a regra de decisão para uma observação \underline{x}_0 ;
 - c) Seja $\underline{x}'_0 = [5,0 \ 3,6 \ 1,4 \ 0,2]$. Em qual das duas espécies este indivíduo seria classificado?
 - d) A separação entre as duas populações é significativa? Usar $\alpha = 5\%$ e concluir.
2. Considerando os dados das três espécies, apresentadas na Tabela 7.11, pede-se:
 - a) Obter as Funções Discriminantes de Anderson. (Usar $p_1 = p_2 = p_3 = 1/3$).
 - b) A observação $\underline{x}'_0 = [5,0 \ 3,6 \ 1,4 \ 0,2]$ seria classificada em qual espécie?

VARIÁVEIS:

X_1 – Sépala alongada
 X_2 – Sépala larga
 X_3 – Pétala alongada
 X_4 – Pétala larga

ESPÉCIES

1 – *Iris setosa*
 2 – *Iris versicolour*
 3 – *Iris virginica*

Tabela 7.11 – Medidas de três tipos de *Iris*

Tipo	X ₁	X ₂	X ₃	X ₄
1	5,1	3,5	1,4	0,2
1	4,9	3,0	1,4	0,2
1	4,7	3,2	1,3	0,2
1	4,6	3,1	1,5	0,2
1	5,0	3,6	1,4	0,2
1	5,4	3,9	1,7	0,4
1	4,6	3,4	1,4	0,3
1	5,0	3,4	1,5	0,2
1	4,4	2,9	1,4	0,2
1	4,9	3,1	1,5	0,1
1	5,4	3,7	1,5	0,2
1	4,8	3,4	1,6	0,2
1	4,8	3,0	1,4	0,1
1	4,3	3,0	1,1	0,1
1	5,8	4,0	1,2	0,2
1	5,7	4,4	1,5	0,4
1	5,4	3,9	1,3	0,4
1	5,1	3,5	1,4	0,3
1	5,7	3,8	1,7	0,3
1	5,1	3,8	1,5	0,3
1	5,4	3,4	1,7	0,2
1	5,1	3,7	1,5	0,4
1	4,6	3,6	1,0	0,2
1	5,1	3,3	1,7	0,5
1	4,8	3,4	1,9	0,2
1	5,0	3,0	1,6	0,2
1	5,0	3,4	1,6	0,4
1	5,2	3,5	1,5	0,2
1	5,2	3,4	1,4	0,2
1	4,7	3,2	1,6	0,2
1	4,8	3,1	1,6	0,2
1	5,4	3,4	1,5	0,4
1	5,2	4,1	1,5	0,1
1	5,5	4,2	1,4	0,2
1	4,9	3,1	1,5	0,2
1	5,0	3,2	1,2	0,2
1	5,5	3,5	1,3	0,2
1	4,9	3,6	1,4	0,1
1	4,4	3,0	1,3	0,2
1	5,1	3,4	1,5	0,2
1	5,0	3,5	1,3	0,3
1	4,5	2,3	1,3	0,3
1	4,4	3,2	1,3	0,2
1	5,0	3,5	1,6	0,6
1	5,1	3,8	1,9	0,4
1	4,8	3,0	1,4	0,3
1	5,1	3,8	1,6	0,2
1	4,6	3,2	1,4	0,2
1	5,3	3,7	1,5	0,2
1	5,0	3,3	1,4	0,2

2	7,0	3,2	4,7	1,4
2	6,4	3,2	4,5	1,5
2	6,9	3,1	4,9	1,5
2	5,5	2,3	4,0	1,3
2	6,5	2,8	4,6	1,5
2	5,7	2,8	4,5	1,3
2	6,3	3,3	4,7	1,6
2	4,9	2,4	3,3	1,0
2	6,6	2,9	4,6	1,3
2	5,2	2,7	3,9	1,4
2	5,0	2,0	3,5	1,0
2	5,9	3,0	4,2	1,5
2	6,0	2,2	4,0	1,0
2	6,1	2,9	4,7	1,4
2	5,6	2,9	3,6	1,3
2	6,7	3,1	4,4	1,4
2	5,6	3,0	4,5	1,5
2	5,8	2,7	4,1	1,0
2	6,2	2,2	4,5	1,5
2	5,6	2,5	3,9	1,1
2	5,9	3,2	4,8	1,8
2	6,1	2,8	4,0	1,3
2	6,3	2,5	4,9	1,5
2	6,1	2,8	4,7	1,2
2	6,4	2,9	4,3	1,3
2	6,6	3,0	4,4	1,4
2	6,8	2,8	4,8	1,4
2	6,7	3,0	5,0	1,7
2	6,0	2,9	4,5	1,5
2	5,7	2,6	3,5	1,0
2	5,5	2,4	3,8	1,1
2	5,5	2,4	3,7	1,0
2	5,8	2,7	3,9	1,2
2	6,0	2,7	5,1	1,6
2	5,4	3,0	4,5	1,5
2	6,0	3,4	4,5	1,6
2	6,7	3,1	4,7	1,5
2	6,3	2,3	4,4	1,3
2	5,6	3,0	4,1	1,3
2	5,5	2,5	4,0	1,3
2	5,5	2,6	4,4	1,2
2	6,1	3,0	4,6	1,4
2	5,8	2,6	4,0	1,2
2	5,0	2,3	3,3	1,0
2	5,6	2,7	4,2	1,3
2	5,7	3,0	4,2	1,2
2	5,7	2,9	4,2	1,3
2	6,2	2,9	4,3	1,3
2	5,1	2,5	3,0	1,1
2	5,7	2,8	4,1	1,3
3	6,3	3,3	6,0	2,5
3	5,8	2,7	5,1	1,9
3	7,1	3,0	5,9	2,1
3	6,3	2,9	5,6	1,8
3	6,5	3,0	5,8	2,2

3	7,6	3,0	6,6	2,1
3	4,9	2,5	4,5	1,7
3	7,3	2,9	6,3	1,8
3	6,7	2,5	5,8	1,8
3	7,2	3,6	6,1	2,5
3	6,5	3,2	5,1	2,0
3	6,4	2,7	5,3	1,9
3	6,8	3,0	5,5	2,1
3	5,7	2,5	5,0	2,0
3	5,8	2,8	5,1	2,4
3	6,4	3,2	5,3	2,3
3	6,5	3,0	5,5	1,8
3	7,7	3,8	6,7	2,2
3	7,7	2,6	6,9	2,3
3	6,0	2,2	5,0	1,5
3	6,9	3,2	5,7	2,3
3	5,6	2,8	4,9	2,0
3	7,7	2,8	6,7	2,0
3	6,3	2,7	4,9	1,8
3	6,7	3,3	5,7	2,1
3	7,2	3,2	6,0	1,8
3	6,2	2,8	4,8	1,8
3	6,1	3,0	4,9	1,8
3	6,4	2,8	5,6	2,1
3	7,2	3,0	5,8	1,6
3	7,4	2,8	6,1	1,9
3	7,9	3,8	6,4	2,0
3	6,4	2,8	5,6	2,2
3	6,3	2,8	5,1	1,5
3	6,1	2,6	5,6	1,4
3	7,7	3,0	6,1	2,3
3	6,3	3,4	5,6	2,4
3	6,4	3,1	5,5	1,8
3	6,0	3,0	4,8	1,8
3	6,9	3,1	5,4	2,1
3	6,7	3,1	5,6	2,4
3	6,9	3,1	5,1	2,3
3	5,8	2,7	5,1	1,9
3	6,8	3,2	5,9	2,3
3	6,7	3,3	5,7	2,5
3	6,7	3,0	5,2	2,3
3	6,3	2,5	5,0	1,9
3	6,5	3,0	5,2	2,0
3	6,2	3,4	5,4	2,3
3	5,9	3,0	5,1	1,8

7.7. REFERÊNCIAS

ANDERSON, T. W. **An introduction to multivariate statistical analysis**. New York: John Wiley and Sons, 1958. 374 p.

- BARTLETT, M. S. The statistical conception of mental factors. **British Journal of Psychology**, v.28, p. 97-104. 1937.
- BOX, G. E. P. A general distribution theory for a class of likelihood criteria. **Biometrika**, v.36, p. 317-346. 1949.
- CAMARGO, J. M. F. O grupo *Partamona* (*Partamona testacea* (Klung): suas espécies, distribuição e diferenciação geográfica (Meliponinae, Apidae, Hymenoptera). Curitiba: PR, 1977, 206 p. (Dissertação de Mestrado).
- CRUZ, C. D.; CARNEIRO, P. C. S.; REGAZZI, A. J.; **Modelos biométricos aplicados ao melhoramento genético**. v.2, 3. ed., Viçosa: UFV, 2014. 668p.
- CRUZ, C. D.; REGAZZI, A. J.; CARNEIRO, P. C. S. **Modelos biométricos aplicados ao melhoramento genético**. v.1, 3. ed., Viçosa: UFV, 2012. 514p.
- FISHER, R. A. The use of multiple measurements in taxonomic problems. **Ann. Eugenics**, v.7, p. 179 – 188. 1936.
- HOEL, P. G. **Introduction to mathematical statistics**. 4. ed., New York: John Wiley and Sons, 1971. 383p.
- JOHNSON, R. A.; WICHERN, D. W. **Applied multivariate statistical analysis**. 4th ed., USA, Upper Saddle River, New Jersey: Prentice Hall, Inc., 1998. 816p.
- KHATTREE, R.; NAIK, D. N. **Multivariate data reduction and discrimination with SAS Software**. USA: SAS Institute Inc., Cary, NC, 2000. 558p.
- KORIN, B. P. On testing the equality of k covariance matrices. **Biometrika**, v.56, p. 216-218. 1969.
- MARDIA, K. V.; KENT, J. T.; BIBBY, J. M. **Multivariate analysis**. 6th ed., Londres: Academic Press, 1997. 518p.
- MORRISON, D. F. **Multivariate statistical methods**. 2nd ed., USA: McGraw-Hill Book Company, 1976. 415p.
- OKAMOTO, M. An asymptotic expansion for the distribution of the linear discriminant function. **Ann. Math. Statist.**, v.34, p. 1286–1301. 1963. Correction: **Ann. Math. Statist.**, v.39, p. 1358–1359. 1968.
- PEARSON, E. S.; HARTLEY, H. O. **Biometrika Tables for Statisticians**, v.2, Cambridge University Press, Cambridge. 1972.
- RAO, C.R. Sir Ronald Aylmer Fisher. The architect of multivariate analysis. **Biometrics**, v.20, n.2, p. 286–300. 1964.
- RENCER, A. C. **Methods of multivariate analysis**. 2nd ed., USA: John Wiley and Sons, 2002. 738p.
- SMITH, C. A. B. Some examples of discrimination. **Ann. Eugenics**, v.13, p. 228–237. 1947.

CAPÍTULO 8

CORRELAÇÕES CANÔNICAS

8.1. INTRODUÇÃO

A análise de correlação de Pearson pode ser generalizada para a correlação de uma variável sobre um conjunto de variáveis, o que constitui a análise de correlação múltipla.

Quando se estuda regressão linear múltipla, coloca-se o problema de encontrar-se a combinação linear de X_1, X_2, \dots, X_p que mais se correlaciona com a variável Y . O coeficiente de correlação entre Y e X_1, X_2, \dots, X_p é o coeficiente de correlação simples entre Y e \hat{Y} , sendo \hat{Y} o estimador obtido da regressão linear de Y sobre o conjunto de variáveis. Se $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \dots + \hat{\beta}_p X_p$, então

$$r_{Y, \hat{Y}} = \frac{\text{Cov}(Y, \hat{Y})}{\sqrt{\hat{V}(Y) \cdot \hat{V}(\hat{Y})}} \quad (1)$$

é o coeficiente de correlação múltipla entre Y e X_1, X_2, \dots, X_p . A expressão (1) ao quadrado nos dá o coeficiente de determinação múltipla (R^2 para $p \geq 2$). Para $p = 1$, $(r_{Y, \hat{Y}})^2$ é o coeficiente de determinação simples (r^2). Uma generalização adicional conduz ao conceito de **correlação canônica**.

Correlação canônica é uma técnica para analisar as relações entre dois conjuntos de variáveis, na qual se estima a máxima correlação entre combinações lineares das variáveis.

Alguns exemplos de aplicações da Análise de Correlação Canônica:

- \underline{X} - Conjunto de variáveis de solo
 \underline{Y} - Conjunto de variáveis da planta
- \underline{X} - Conjunto de variáveis da parte aérea da planta
 \underline{Y} - Conjunto de variáveis do sistema radicular
- \underline{X} - Caracteres agronômicos
 \underline{Y} - Caracteres fisiológicos
- \underline{X} - Conjunto de variáveis referentes às medidas físicas de uma criança
 \underline{Y} - Conjunto de variáveis referentes às medidas mentais de uma criança

8.2. ESTIMAÇÃO DAS CORRELAÇÕES CANÔNICAS E DOS PARES CANÔNICOS

A análise de correlação canônica caracteriza-se por avaliar relações entre dois conjuntos influenciados, no mínimo, por duas variáveis. De maneira genérica considera-se que o primeiro conjunto é estabelecido por p variáveis e o segundo por q . O número de correlações canônicas é igual ao menor número de variáveis que constituir um dos conjuntos (p ou q), e sua magnitude sempre decresce com a ordem em que são estimadas. Entretanto, o primeiro coeficiente é sempre maior ou igual, em valor absoluto, a qualquer coeficiente de correlação simples ou múltipla, entre as variáveis do primeiro e do segundo grupo.

Na análise de correlação canônica têm-se dois conjuntos de variáveis:

$$\tilde{X} = \begin{bmatrix} X_1 & X_2 & \cdots & X_p \end{bmatrix}' \text{ e}$$

$$\tilde{Y} = \begin{bmatrix} Y_1 & Y_2 & \cdots & Y_q \end{bmatrix}'.$$

Assim, procura-se estabelecer a forma e a dimensão do relacionamento entre esses dois conjuntos.

São calculadas inicialmente duas combinações lineares, uma sobre cada conjunto de variáveis, de forma que o coeficiente de correlação de Pearson entre elas seja máximo.

As combinações lineares são denominadas **variáveis canônicas** e os pares são denominados **pares canônicos**. Em seguida são calculadas duas outras variáveis canônicas, uma sobre cada conjunto de variáveis, com a condição de serem ortogonais às primeiras e apresentarem o máximo coeficiente de correlação de Pearson entre si. E assim por diante.

Sejam U_1 e V_1 as variáveis que constituem o primeiro par canônico:

$$U_1 = a_{11}X_1 + a_{21}X_2 + \cdots + a_{p1}X_p = \tilde{a}'_1 \tilde{X}$$

$$V_1 = b_{11}Y_1 + b_{21}Y_2 + \cdots + b_{q1}Y_q = \tilde{b}'_1 \tilde{Y}$$

Nesse caso, o problema consiste em determinar os vetores \tilde{a}_1 e \tilde{b}_1 , de forma a tornar máximo o coeficiente de correlação entre U_1 e V_1 .

Sejam agora, U_2 e V_2 as variáveis canônicas que constituem o segundo par canônico:

$$U_2 = a_{12}X_1 + a_{22}X_2 + \cdots + a_{p2}X_p = \tilde{a}'_2 \tilde{X}$$

$$V_2 = b_{12}Y_1 + b_{22}Y_2 + \cdots + b_{q2}Y_q = \tilde{b}'_2 \tilde{Y}$$

Aqui, o problema consiste em determinar os vetores \underline{a}_2 e \underline{b}_2 de forma a se obter o máximo coeficiente de correlação entre U_2 e V_2 , obedecida a condição de ortogonalidade entre U_2 e U_1 , e entre V_2 e V_1 .

O terceiro par canônico é expresso por:

$$U_3 = a_{13}X_1 + a_{23}X_2 + \cdots + a_{p3}X_p = \underline{a}'_3 \underline{X}$$

$$V_3 = b_{13}Y_1 + b_{23}Y_2 + \cdots + b_{q3}Y_q = \underline{b}'_3 \underline{Y}$$

Os vetores \underline{a}_3 e \underline{b}_3 são determinados de forma a se obter o máximo coeficiente de correlação entre U_3 e V_3 , sendo U_3 ortogonal a U_1 e U_2 , e V_3 ortogonal a V_1 e V_2 .

Os coeficientes de correlação canônica entre as variáveis dos pares sucessivos são obviamente decrescentes. Cada coeficiente de correlação canônica representa a magnitude de variação comum entre os dois conjuntos de variáveis; a composição do par canônico correspondente representa o padrão de variação (comum) a que se refere o coeficiente.

A análise de correlação canônica se baseia na determinação de variáveis canônicas ortogonais em cada conjunto de variáveis, por isso as variáveis em cada conjunto devem ser linearmente independentes, ou o que é a mesma coisa:

- i) O posto da matriz de correlação das variáveis X_1, X_2, \dots, X_p deve ser p ;
- ii) O posto da matriz de correlação das variáveis Y_1, Y_2, \dots, Y_q deve ser q .

Se esta condição não for satisfeita de início, será necessário descartar as variáveis que são combinações lineares das demais (variáveis redundantes).

Define-se como a primeira correlação canônica aquela que maximiza a relação entre $U_1 = \underline{a}'_1 \underline{X}$ e $V_1 = \underline{b}'_1 \underline{Y}$. As variáveis canônicas U_1 e V_1 constituem o primeiro par canônico

associado àquela correlação canônica, que é expressa por: $r_1 = \frac{\text{Cov}(U_1, V_1)}{\sqrt{\hat{V}(U_1) \cdot \hat{V}(V_1)}}$,

em que,

$$\text{Cov}(U_1, V_1) = \underline{a}'_1 S_{12} \underline{b}_1$$

$$\hat{V}(U_1) = \underline{a}'_1 S_{11} \underline{a}_1$$

$$\hat{V}(V_1) = \underline{b}'_1 S_{22} \underline{b}_1$$

S_{11} : matriz $p \times p$ de covariância amostral entre as variáveis X_1, X_2, \dots, X_p ;

S_{22} : matriz $q \times q$ de covariância amostral entre as variáveis Y_1, Y_2, \dots, Y_q ;

S_{12} : matriz $p \times q$ de covariância amostral entre as variáveis dos dois conjuntos.

Normalmente utilizam-se as variáveis padronizadas (média zero e variância 1). Neste caso, tem-se $S_{11} = R_{11}$, $S_{22} = R_{22}$ e $S_{12} = R_{12}$, em que R representa uma matriz de correlação amostral.

A análise de correlação canônica visa encontrar combinações lineares $U = \underline{a}' \underline{X}$ e $V = \underline{b}' \underline{Y}$

de tal modo que $r^2(U, V)$ seja máxima. Por definição, $r^2(\underline{a}' \underline{X}, \underline{b}' \underline{Y}) = \frac{(\underline{a}' S_{12} \underline{b})^2}{(\underline{a}' S_{11} \underline{a})(\underline{b}' S_{22} \underline{b})}$, e o

problema da correlação canônica pode ser resolvido encontrando-se a solução para o problema:

$$\begin{aligned} &\max_{\underline{a}, \underline{b}} \left(\underline{a}' S_{12} \underline{b} \right)^2 \\ &\underline{a}, \underline{b} \end{aligned}$$

sujeito às restrições:

$$\begin{cases} \underline{a}' S_{11} \underline{a} = 1 \\ \underline{b}' S_{22} \underline{b} = 1 \end{cases}$$

As duas restrições referem-se ao fato de que a correlação é invariante à mudança de escala.

O problema anteriormente formulado é resolvido pelo Método dos Multiplicadores de Lagrange, para obtenção dos vetores \underline{a} e \underline{b} .

Seja, então, o Lagrangeano,

$L(\underline{a}, \underline{b}, \lambda_1, \lambda_2) = (\underline{a}' S_{12} \underline{b})^2 - \lambda_1 (\underline{a}' S_{11} \underline{a} - 1) - \lambda_2 (\underline{b}' S_{22} \underline{b} - 1)$, em que λ_1 e λ_2 são os multiplicadores de Lagrange. Ou, no caso das variáveis padronizadas, virá:

$$L(\underline{a}, \underline{b}, \lambda_1, \lambda_2) = (\underline{a}' R_{12} \underline{b})^2 - \lambda_1 (\underline{a}' R_{11} \underline{a} - 1) - \lambda_2 (\underline{b}' R_{22} \underline{b} - 1).$$

Então,

$$\frac{\partial L}{\partial \underline{a}} = 2(\underline{a}' R_{12} \underline{b}) R_{12} \underline{b} - 2\lambda_1 R_{11} \underline{a} = 0$$

$$\frac{\partial L}{\partial \underline{b}} = 2(\underline{a}' R_{12} \underline{b}) R_{21} \underline{a} - 2\lambda_2 R_{22} \underline{b} = 0$$

ou ainda,

$$\begin{cases} (\underline{a}' \mathbf{R}_{12} \underline{b}) \mathbf{R}_{12} \underline{b} - \lambda_1 \mathbf{R}_{11} \underline{a} = 0 \\ (\underline{a}' \mathbf{R}_{12} \underline{b}) \mathbf{R}_{21} \underline{a} - \lambda_2 \mathbf{R}_{22} \underline{b} = 0 \end{cases} \quad \begin{matrix} (2) \\ (3) \end{matrix}$$

Pré-multiplicando a equação (2) por \underline{a}' e a equação (3) por \underline{b}' , virá:

$$\begin{cases} \underline{a}' (\underline{a}' \mathbf{R}_{12} \underline{b}) \mathbf{R}_{12} \underline{b} - \lambda_1 \underline{a}' \mathbf{R}_{11} \underline{a} = 0 \\ \underline{b}' (\underline{a}' \mathbf{R}_{12} \underline{b}) \mathbf{R}_{21} \underline{a} - \lambda_2 \underline{b}' \mathbf{R}_{22} \underline{b} = 0 \end{cases}$$

$$\begin{cases} (\underline{a}' \mathbf{R}_{12} \underline{b}) (\underline{a}' \mathbf{R}_{12} \underline{b}) - \lambda_1 \underline{a}' \mathbf{R}_{11} \underline{a} = 0 \\ (\underline{a}' \mathbf{R}_{12} \underline{b}) (\underline{a}' \mathbf{R}_{12} \underline{b}) - \lambda_2 \underline{b}' \mathbf{R}_{22} \underline{b} = 0 \end{cases}$$

$$\begin{cases} (\underline{a}' \mathbf{R}_{12} \underline{b})^2 - \lambda_1 \underline{a}' \mathbf{R}_{11} \underline{a} = 0 \\ (\underline{a}' \mathbf{R}_{12} \underline{b})^2 - \lambda_2 \underline{b}' \mathbf{R}_{22} \underline{b} = 0 \end{cases}$$

Logo,

$$\lambda_1 \underline{a}' \mathbf{R}_{11} \underline{a} = \lambda_2 \underline{b}' \mathbf{R}_{22} \underline{b} = (\underline{a}' \mathbf{R}_{12} \underline{b})^2$$

Como, $\underline{a}' \mathbf{R}_{11} \underline{a} = \underline{b}' \mathbf{R}_{22} \underline{b} = 1$, segue que:

$$\lambda_1 = \lambda_2 = \lambda = (\underline{a}' \mathbf{R}_{12} \underline{b})^2 \Rightarrow \underline{a}' \mathbf{R}_{12} \underline{b} = \lambda^{1/2} = \delta$$

Assim, sob as restrições descritas, a raiz quadrada do multiplicador de Lagrange λ , mede a máxima correlação entre as combinações lineares das variáveis dos dois conjuntos.

Voltando às equações (2) e (3) originadas pelas derivadas, tem-se o sistema homogêneo de incógnitas \underline{a} , \underline{b} e δ :

$$\begin{cases} -\delta^2 \mathbf{R}_{11} \underline{a} + \delta \mathbf{R}_{12} \underline{b} = 0 \\ \delta \mathbf{R}_{21} \underline{a} - \delta^2 \mathbf{R}_{22} \underline{b} = 0 \end{cases}$$

Considerando-se \underline{a} e \underline{b} como incógnitas, o sistema acima terá solução diferente da trivial, se e somente se,

$$\begin{vmatrix} -\delta^2 \mathbf{R}_{11} & \delta \mathbf{R}_{12} \\ \delta \mathbf{R}_{21} & -\delta^2 \mathbf{R}_{22} \end{vmatrix} = 0.$$

Sabe-se que, por exemplo, se A é positiva definida, então, por operações elementares que não alterem o determinante, pode-se escrever:

$$\begin{aligned}
 \det \begin{bmatrix} A & B \\ C & D \end{bmatrix} &= \det \left\{ \begin{bmatrix} I & \phi \\ \phi & I \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} \right\} \\
 &= \det \left\{ \begin{bmatrix} A & \phi \\ \phi & I \end{bmatrix} \begin{bmatrix} A^{-1} & \phi \\ \phi & I \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} \right\} \\
 &= \det \left\{ \begin{bmatrix} A & \phi \\ \phi & I \end{bmatrix} \begin{bmatrix} I & A^{-1}B \\ C & D \end{bmatrix} \right\} \\
 &= \det \left\{ \begin{bmatrix} A & \phi \\ \phi & I \end{bmatrix} \begin{bmatrix} I & A^{-1}B \\ \phi & D - CA^{-1}B \end{bmatrix} \right\} \\
 &= \det \begin{bmatrix} A & B \\ \phi & D - CA^{-1}B \end{bmatrix}.
 \end{aligned}$$

Logo, $\det \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \det(A) \cdot \det(D - CA^{-1}B)$, de onde se conclui que, para o caso em questão, o

determinante pode ser escrito como:

$$\begin{aligned}
 \det(-\delta^2 R_{11}) \cdot \det(-\delta^2 R_{22} + \delta R_{21} \delta^{-2} R_{11}^{-1} \delta R_{12}) &= 0 \\
 \Rightarrow \det(-\delta^2 R_{22} + R_{21} R_{11}^{-1} R_{12}) &= 0 \\
 \det(R_{22}^{-1} R_{21} R_{11}^{-1} R_{12} - \lambda I) &= 0.
 \end{aligned}$$

Como $(\underline{a}' R_{12} \underline{b})^2 = \lambda$ e sendo λ um autovalor de $R_{22}^{-1} R_{21} R_{11}^{-1} R_{12}$, conclui-se que a correlação máxima é obtida quando λ é o maior autovalor de $R_{22}^{-1} R_{21} R_{11}^{-1} R_{12}$. Uma vez que a equação foi resolvida em relação a R_{22} , ou seja, em relação ao conjunto de variáveis Y então \underline{b} é o autovetor associado ao autovalor λ . Uma vez conhecido \underline{b} e voltando ao sistema de equações homogêneas pode-se agora, determinar o vetor \underline{a} . Assim, da equação

$$-\delta^2 R_{11} \underline{a} + \delta R_{12} \underline{b} = \underline{0}$$

tem-se que:

$$-\delta(\delta R_{11} \underline{a} - R_{12} \underline{b}) = \underline{0}$$

$$\delta R_{11} \underline{a} - R_{12} \underline{b} = \underline{0}$$

$$\lambda^{1/2} \mathbf{R}_{11} \mathbf{a} - \mathbf{R}_{12} \mathbf{b} = \mathbf{0} \quad (4)$$

Pré-multiplicando a equação (4) por \mathbf{R}_{11}^{-1} , virá:

$$\lambda^{1/2} \mathbf{a} = \mathbf{R}_{11}^{-1} \mathbf{R}_{12} \mathbf{b}$$

$$\mathbf{a} = \frac{1}{\sqrt{\lambda}} \mathbf{R}_{11}^{-1} \mathbf{R}_{12} \mathbf{b}$$

Pode-se concluir que os autovalores da matriz $\mathbf{R}_{22}^{-1} \mathbf{R}_{21} \mathbf{R}_{11}^{-1} \mathbf{R}_{12}$ são os quadrados das correlações canônicas.

Foi visto que:

i) Equação característica para determinação dos autovalores da matriz $\mathbf{R}_{22}^{-1} \mathbf{R}_{21} \mathbf{R}_{11}^{-1} \mathbf{R}_{12}$

$$\left| \mathbf{R}_{22}^{-1} \mathbf{R}_{21} \mathbf{R}_{11}^{-1} \mathbf{R}_{12} - \lambda \mathbf{I} \right| = 0$$

ii) \mathbf{b} é o autovetor da matriz $\mathbf{R}_{22}^{-1} \mathbf{R}_{21} \mathbf{R}_{11}^{-1} \mathbf{R}_{12}$, que é obtido resolvendo-se a equação:

$$\left[\mathbf{R}_{22}^{-1} \mathbf{R}_{21} \mathbf{R}_{11}^{-1} \mathbf{R}_{12} - \lambda \mathbf{I} \right] \mathbf{b} = \mathbf{0}$$

iii) O vetor \mathbf{a} pode ser obtido de:

$$\mathbf{a} = \frac{1}{\sqrt{\lambda}} \mathbf{R}_{11}^{-1} \mathbf{R}_{12} \mathbf{b}.$$

$$\text{Veja agora como obter a equação } \left[\mathbf{R}_{11}^{-1} \mathbf{R}_{12} \mathbf{R}_{22}^{-1} \mathbf{R}_{21} - \lambda \mathbf{I} \right] \mathbf{a} = \mathbf{0}.$$

Pode-se verificar que:

$$\begin{cases} \delta \mathbf{R}_{12} \mathbf{b} - \delta^2 \mathbf{R}_{11} \mathbf{a} = \mathbf{0} \\ \delta \mathbf{R}_{21} \mathbf{a} - \delta^2 \mathbf{R}_{22} \mathbf{b} = \mathbf{0} \end{cases}$$

ou ainda,

$$\begin{cases} \mathbf{R}_{12} \mathbf{b} = \delta \mathbf{R}_{11} \mathbf{a} \\ \mathbf{R}_{21} \mathbf{a} = \delta \mathbf{R}_{22} \mathbf{b} \end{cases} \quad (5)$$

Multiplicando (5) por δ , virá:

$$\mathbf{R}_{12} \delta \mathbf{b} = \delta^2 \mathbf{R}_{11} \mathbf{a} \quad (7)$$

Pré-multiplicando (6) por \mathbf{R}_{22}^{-1} , virá:

$$\mathbf{R}_{22}^{-1}\mathbf{R}_{21}\mathbf{a} = \delta\mathbf{b} \quad (8)$$

Substituindo (8) em (7), virá:

$$\mathbf{R}_{12}\mathbf{R}_{22}^{-1}\mathbf{R}_{21}\mathbf{a} = \delta^2\mathbf{R}_{11}\mathbf{a} \quad (9)$$

Agora, basta pré-multiplicar (9) por \mathbf{R}_{11}^{-1} para obter:

$$\mathbf{R}_{11}^{-1}\mathbf{R}_{12}\mathbf{R}_{22}^{-1}\mathbf{R}_{21}\mathbf{a} = \lambda\mathbf{a}, \text{ pois } \delta^2 = \lambda$$

Assim, virá:

$$\left[\mathbf{R}_{11}^{-1}\mathbf{R}_{12}\mathbf{R}_{22}^{-1}\mathbf{R}_{21} - \lambda\mathbf{I} \right] \mathbf{a} = \mathbf{0}$$

Além disso, da equação

$$-\delta^2\mathbf{R}_{22}\mathbf{b} + \delta\mathbf{R}_{21}\mathbf{a} = \mathbf{0}$$

tem-se que:

$$-\delta(\delta\mathbf{R}_{22}\mathbf{b} - \mathbf{R}_{21}\mathbf{a}) = \mathbf{0}$$

$$\delta\mathbf{R}_{22}\mathbf{b} - \mathbf{R}_{21}\mathbf{a} = \mathbf{0}$$

$$\lambda^{1/2}\mathbf{R}_{22}\mathbf{b} - \mathbf{R}_{21}\mathbf{a} = \mathbf{0} \quad (10)$$

Pré-multiplicando a equação (10) por \mathbf{R}_{22}^{-1} , virá:

$$\lambda^{1/2}\mathbf{b} = \mathbf{R}_{22}^{-1}\mathbf{R}_{21}\mathbf{a}$$

$$\mathbf{b} = \frac{1}{\sqrt{\lambda}}\mathbf{R}_{22}^{-1}\mathbf{R}_{21}\mathbf{a}$$

Foi visto que:

i) Equação característica para determinação dos autovalores da matriz $\mathbf{R}_{11}^{-1}\mathbf{R}_{12}\mathbf{R}_{22}^{-1}\mathbf{R}_{21}$:

$$\left| \mathbf{R}_{11}^{-1}\mathbf{R}_{12}\mathbf{R}_{22}^{-1}\mathbf{R}_{21} - \lambda\mathbf{I} \right| = 0$$

ii) \mathbf{a} é o autovetor da matriz $\mathbf{R}_{11}^{-1}\mathbf{R}_{12}\mathbf{R}_{22}^{-1}\mathbf{R}_{21}$, que é obtido resolvendo-se a equação:

$$\left[\mathbf{R}_{11}^{-1}\mathbf{R}_{12}\mathbf{R}_{22}^{-1}\mathbf{R}_{21} - \lambda\mathbf{I} \right] \mathbf{a} = \mathbf{0}$$

iii) O vetor \mathbf{b} pode ser obtido de:

$$\mathbf{b} = \frac{1}{\sqrt{\lambda}}\mathbf{R}_{22}^{-1}\mathbf{R}_{21}\mathbf{a}$$

Note que as matrizes em cujos autovalores há interesse são quadradas de ordem p ou q , conforme o caso.

$$R = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix}, R_{21} = R_{12}'$$

$$\left. \begin{array}{l} R_{11}^{-1} \text{ é } p \times p \\ R_{12} \text{ é } p \times q \end{array} \right\} \text{Então } R_{11}^{-1} R_{12} \text{ é } p \times q \quad \left. \begin{array}{l} R_{22}^{-1} \text{ é } q \times q \\ R_{21} \text{ é } q \times p \end{array} \right\} \text{Então } R_{22}^{-1} R_{21} \text{ é } q \times p \quad \left. \vphantom{\begin{array}{l} R_{11}^{-1} \text{ é } p \times p \\ R_{12} \text{ é } p \times q \\ R_{22}^{-1} \text{ é } q \times q \\ R_{21} \text{ é } q \times p \end{array}} \right\} \text{Portanto } R_{11}^{-1} R_{12} R_{22}^{-1} R_{21} \text{ é quadrada de ordem } p.$$

$$\left. \begin{array}{l} R_{22}^{-1} \text{ é } q \times q \\ R_{21} \text{ é } q \times p \end{array} \right\} \text{Então } R_{22}^{-1} R_{21} \text{ é } q \times p \quad \left. \begin{array}{l} R_{11}^{-1} \text{ é } p \times p \\ R_{12} \text{ é } p \times q \end{array} \right\} \text{Então } R_{11}^{-1} R_{12} \text{ é } p \times q \quad \left. \vphantom{\begin{array}{l} R_{22}^{-1} \text{ é } q \times q \\ R_{21} \text{ é } q \times p \\ R_{11}^{-1} \text{ é } p \times p \\ R_{12} \text{ é } p \times q \end{array}} \right\} \text{Portanto } R_{22}^{-1} R_{21} R_{11}^{-1} R_{12} \text{ é quadrada de ordem } q.$$

Observe que os autovalores (λ) podem ser igualmente calculados a partir de duas equações características distintas, vale dizer, a partir de duas matrizes diferentes, uma de ordem p e outra de ordem q . É claro que se $p=q$ e as variáveis X_1, X_2, \dots, X_p , bem como as variáveis Y_1, Y_2, \dots, Y_q são linearmente independentes, existirão $p=q$ autovalores não-nulos e $p=q$ pares canônicos. Entretanto se, por exemplo, $p < q$, existirão $q-p$ autovalores nulos da matriz $R_{22}^{-1} R_{21} R_{11}^{-1} R_{12}$ e apenas p pares canônicos.

Os autovetores \underline{a} e \underline{b} devem ser normalizados por $\underline{a}' R_{11} \underline{a} = 1$ e $\underline{b}' R_{22} \underline{b} = 1$, respectivamente.

Mas geralmente, se existem p variáveis X e q variáveis Y , com X e Y variáveis padronizadas (média zero e variância 1), então para $k = \min(p, q)$, pode-se escrever:

$$\begin{aligned} \text{a) } U_1 &= \underline{a}_1' \underline{X} = a_{11} X_1 + a_{21} X_2 + \dots + a_{p1} X_p \\ U_2 &= \underline{a}_2' \underline{X} = a_{12} X_1 + a_{22} X_2 + \dots + a_{p2} X_p \\ \dots & \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \\ U_k &= \underline{a}_k' \underline{X} = a_{1k} X_1 + a_{2k} X_2 + \dots + a_{pk} X_p \end{aligned}$$

$$V_1 = \tilde{b}'_1 \tilde{Y} = b_{11}Y_1 + b_{21}Y_2 + \dots + b_{q1}Y_q$$

$$V_2 = \tilde{b}'_2 \tilde{Y} = b_{12}Y_1 + b_{22}Y_2 + \dots + b_{q2}Y_q$$

$$\dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots$$

$$V_k = \tilde{b}'_k \tilde{Y} = b_{1k}Y_1 + b_{2k}Y_2 + \dots + b_{qk}Y_q$$

onde os coeficiente são escolhidos de tal modo que $r_1^2(U_1, V_1) > r_2^2(U_2, V_2) > \dots > r_k^2(U_k, V_k)$, isto é, $\lambda_1 > \lambda_2 > \dots > \lambda_k$.

A primeira correlação canônica (r_1) entre as combinações lineares $U_1 = \tilde{a}'_1 \tilde{X}$ e $V_1 = \tilde{b}'_1 \tilde{Y}$

é dada por: $r_1 = \sqrt{\lambda_1}$.

b) Fica evidenciado, então, que o coeficiente de correlação canônica de ordem j é a raiz quadrada do autovalor λ_j , isto é:

$$r_j = \frac{\tilde{a}'_j R_{12} \tilde{b}_j}{\left[\left(\tilde{a}'_j R_{11} \tilde{a}_j \right) \left(\tilde{b}'_j R_{22} \tilde{b}_j \right) \right]^{1/2}} = \tilde{a}'_j R_{12} \tilde{b}_j = \sqrt{\lambda_j}, \text{ pois } \tilde{a}'_j R_{11} \tilde{a}_j \text{ e } \tilde{b}'_j R_{22} \tilde{b}_j \text{ são iguais a } 1.$$

Na realidade, r_j independe da escala de \tilde{a}_j e \tilde{b}_j .

A maior das k correlações canônicas é superior a qualquer uma das correlações individuais das variáveis do primeiro conjunto com as variáveis do segundo conjunto. A análise de correlação canônica se baseia no cálculo de pares canônicos de correlação máxima.

c) Variáveis canônicas originadas de um mesmo conjunto de variáveis são não correlacionadas entre si e têm variância unitária, isto é:

- $\hat{V}(U_j) = 1$, para $j = 1, 2, \dots, k$
 $\text{Cov}(U_i, U_j) = 0$, para $i \neq j$
- $\hat{V}(V_j) = 1$, para $j = 1, 2, \dots, k$
 $\text{Cov}(V_i, V_j) = 0$, para $i \neq j$

E ainda,

$$\text{Cov}(U_j, V_j) = r_j \text{ para } j = 1, 2, \dots, k \text{ e } \text{Cov}(U_i, V_j) = 0 \text{ para } i \neq j.$$

d) Para cada par canônico, o coeficiente de correlação canônica mede a intensidade da correlação, enquanto a composição das variáveis canônicas exprime a natureza da associação.

- e) A análise de correlação canônica é útil à pesquisa das relações entre dois conjuntos de variáveis.

8.2.1. Autovalores e Autovetores da matriz $R_{11}^{-1}R_{12}R_{22}^{-1}R_{21}$

Um problema associado à essa matriz é que ela não é simétrica e as metodologias numéricas usuais disponíveis para determinação dos autovalores e autovetores referem-se a uma matriz simétrica. Dadas as características dessa matriz, produto de uma matriz simétrica positiva definida por outra simétrica, no entanto, é possível resolver o problema através dos processos numéricos usuais disponíveis para matrizes simétricas. Dentre várias alternativas, uma delas é apresentada a seguir.

Necessita-se determinar λ e \underline{a} tal que $R_{11}^{-1}R_{12}R_{22}^{-1}R_{21}\underline{a} = \lambda \underline{a}$.

Considere $G = R_{11}^{-1}$ e $H = R_{12}R_{22}^{-1}R_{21}$. Neste caso, pode-se usar os seguintes fatos:

- a) Se G é uma matriz real e simétrica positiva definida, então existe T não singular, tal que $G = TT'$, em que T é obtida por meio do produto: $(C')^{-1}D^{1/2}$. As matrizes C' e D são, por sua vez, obtidas por operações de congruência em G e elementares em I justaposta a G . Esquematicamente, tem-se:

$$[G : I] \sim \dots \sim [D : C']$$

em que:

\sim : significa operações de congruência em G e elementares em I ;

I : matriz identidade;

D : matriz diagonal;

C : matriz tal que $C'GC = D$.

Comentário: As operações elementares feitas nas linhas de G são também efetuadas na matriz I justaposta à direita de G . Porém, as mesmas operações efetuadas passo a passo nas colunas de G não o são em I .

- b) Se λ é autovalor da matriz não-simétrica GH , resultante do produto de matrizes simétricas G e H , então λ é também autovalor de $T'HT$, real e simétrica. Se $\underline{\theta}$ é autovetor associado à $T'HT$, então $T\underline{\theta} = \underline{a}$ é autovetor associado à GH .

8.3. TESTES DE SIGNIFICÂNCIA

Inicialmente, considere a hipótese $H_0 : \sum_{12} = \phi$, a qual significa que os dois conjuntos de variáveis são não correlacionados um com o outro.

Sendo

$$\Sigma = \begin{bmatrix} \sum_{11} & \sum_{12} \\ \sum_{21} & \sum_{22} \end{bmatrix}_{p+q},$$

então, sob normalidade, a estatística para testar a hipótese

$$H_0 : \sum_{12} = \phi \quad \text{vs.} \quad H_a : \sum_{12} \neq \phi$$

é dada por:

$$\Lambda = \prod_{j=1}^k (1 - r_j^2),$$

em que, r_1, \dots, r_k são os coeficientes de correlação canônica amostral e $k = \min(p, q)$. Segundo Mardia et. al. (1997), usando uma aproximação de Bartlett, tem-se que:

$$-\left[n - \frac{1}{2}(p + q + 3) \right] \ln \prod_{j=1}^k (1 - r_j^2) \sim \chi_{pq}^2, \text{ assintoticamente para } n \text{ grande.}$$

Segundo Mardia et. al. (1997), Bartlett propôs uma estatística similar para testar a hipótese de que apenas s ($s < k = \min(p, q)$) dos coeficientes de correlação canônica populacional são não-zeros. Este teste é baseado na estatística

$$-\left[n - \frac{1}{2}(p + q + 3) \right] \ln \prod_{j=s+1}^k (1 - r_j^2) \sim \chi_{(p-s)(q-s)}^2$$

assintoticamente, a qual é usada para testar a significância de $s + 1$ a k coeficientes de correlação canônica.

Fazendo-se $s = 0$ na expressão anterior, testa-se a hipótese de que os k coeficientes de correlação canônica são nulos, isto é: $H_0 : \rho_1 = \rho_2 = \dots = \rho_k = 0$. A estatística utilizada é,

$$\chi^2 = -\left[n - \frac{1}{2}(p + q + 3) \right] \ln \prod_{j=1}^k (1 - r_j^2)$$

que tem aproximadamente uma distribuição de Qui-quadrado com pq graus de liberdade.

Se a hipótese H_0 pode ser rejeitada, a contribuição do primeiro par canônico pode ser removida e a significância dos $k - 1$ ($s = 1$) pares canônicos remanescentes pode ser testada por:

$$\chi^2 = - \left[n - \frac{1}{2}(p+q+3) \right] \ln \prod_{j=2}^k (1-r_j^2)$$

com $(p-1)(q-1)$ graus de liberdade. Em geral, com s pares canônicos removidos, testa-se a hipótese $H_0 : \rho_s > 0$ e $\rho_{s+1} = \rho_{s+2} = \dots = \rho_k = 0$ através da estatística dada por:

$$\chi^2 = - \left[n - \frac{1}{2}(p+q+3) \right] \ln \prod_{j=s+1}^k (1-r_j^2), \text{ com } (p-s)(q-s) \text{ graus de liberdade.}$$

Para uma interpretação dos testes apresentados, suponha um problema envolvendo as variáveis X_1, X_2 e X_3 ($p=3$) e Y_1, Y_2, Y_3 e Y_4 ($q=4$), em que foram estimadas as correlações canônicas r_1, r_2 e r_3 ($k=3$). Os resultados dos testes foram os seguintes:

Para $s=0$ e $s=1$, as hipóteses H_0 foram rejeitadas, enquanto que para $s=2$, não se rejeitou H_0 . Isto significa que o primeiro e o segundo pares canônicos são estatisticamente relevantes.

Outro teste sobre correlações canônicas é o Teste da Razão de Verossimilhança, que é uma aproximação F da estatística Λ de Wilks, chamada F de Rao (1973, p.556), apresentado a seguir.

TESTE DA RAZÃO DE VEROSSIMILHANÇA

• Sejam $\tilde{X}' = \begin{bmatrix} X_1, & X_2, & \dots, & X_p \end{bmatrix} \Rightarrow p$ variáveis

$\tilde{Y}' = \begin{bmatrix} Y_1, & Y_2, & \dots, & Y_q \end{bmatrix} \Rightarrow q$ variáveis

• Sejam s correlações canônicas, em que $s = \min(p, q)$: r_1, r_2, \dots, r_s .

• H_0 : A k -ésima correlação canônica e todas as que se seguem são zero

H_a : Não H_0

• $\Lambda_k = \prod_{i=k}^s (1-r_i^2)$

FATO:

$$k=1 \Rightarrow \Lambda_1 = \prod_{i=1}^s (1-r_i^2)$$

$$k=2 \Rightarrow \Lambda_2 = \prod_{i=2}^s (1-r_i^2)$$

...

$$k=s \Rightarrow \Lambda_s = (1-r_s^2)$$

A aproximação F para Λ_k é dada por:

$$F_{\text{calculado}} = \frac{1 - \Lambda_k^{1/t}}{\Lambda_k^{1/t}} \cdot \frac{df_2}{df_1} = \left[\left(\frac{1}{\Lambda_k} \right)^{1/t} - 1 \right] \cdot \frac{df_2}{df_1},$$

em que,

$$df_1 = (p - k + 1)(q - k + 1)$$

$$df_2 = wt - \frac{1}{2}[(p - k + 1)(q - k + 1)] + 1$$

$$w = n - \frac{1}{2}(p + q + 3)$$

$$t = \sqrt{\frac{(p - k + 1)^2 (q - k + 1)^2 - 4}{(p - k + 1)^2 + (q - k + 1)^2 - 5}},$$

e quando $(p - k + 1)(q - k + 1) = 2$, então $t = 1$.

$$\bullet F_{\text{tabelado}} \Rightarrow F_{\alpha}(df_1, df_2).$$

8.4. EXEMPLOS

Exemplo 1

Num estudo sobre índices de produção e preço para o período 1919 a 1939 ($n = 21$), desenvolvido por Tintner (citado por CLEMENTE, 1990), foram consideradas as seguintes variáveis:

X_1 - produção de bens duráveis

X_2 - produção de bens não duráveis

X_3 - produção mineral

X_4 - produção agrícola

Y_1 - preços agrícolas

Y_2 - preços dos alimentos

Y_3 - outros preços

Trabalhando-se com as variáveis padronizadas $(x_1, x_2, x_3, x_4$ e $y_1, y_2, y_3)$, obteve-se a seguinte matriz de correlação:

$$R = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix}, R \text{ é simétrica e } R_{21} = R_{12}'$$

$$R = \begin{bmatrix} 1,000000 & 0,495951 & 0,872836 & 0,481240 & -0,436385 & -0,427250 & -0,203390 \\ & 1,000000 & 0,768279 & 0,709807 & 0,425782 & 0,429576 & 0,584220 \\ & & 1,000000 & 0,712358 & -0,038273 & -0,043762 & 0,138680 \\ & & & 1,000000 & 0,261010 & 0,267098 & 0,378452 \\ & & & & 1,000000 & 0,987285 & 0,904598 \\ & & & & & 1,000000 & 0,914394 \\ & & & & & & 1,000000 \end{bmatrix}$$

A inversa de R_{22} é:

$$R_{22}^{-1} = \begin{bmatrix} 39,60728 & -38,69915 & -0,44240 \\ -38,69915 & 43,91373 & -5,14728 \\ -0,44240 & -5,14728 & 6,10683 \end{bmatrix}$$

A inversa de R_{11} é:

$$R_{11}^{-1} = \begin{bmatrix} 6,58473 & 2,37155 & -8,35031 & 1,09623 \\ 2,37155 & 3,66215 & -4,50477 & -0,53169 \\ -8,35031 & -4,50477 & 13,41800 & -2,34240 \\ 1,09623 & -0,53169 & -2,34240 & 2,51847 \end{bmatrix}$$

Pós-multiplicando a matriz R_{11}^{-1} por R_{12} , obtém-se:

$$R_{11}^{-1}R_{12} = \begin{bmatrix} -1,25800 & -1,13634 & -0,69691 \\ 0,55800 & 0,61505 & 0,83121 \\ 0,60096 & 0,41968 & 0,04092 \\ 0,04223 & 0,07842 & 0,09469 \end{bmatrix}$$

Pré-multiplicando a matriz acima por R_{21} , tem-se:

$$R_{21}R_{11}^{-1}R_{12} = \begin{bmatrix} 0,77450 & 0,76211 & 0,68112 \\ 0,76211 & 0,75227 & 0,67831 \\ 0,68112 & 0,67831 & 0,66885 \end{bmatrix},$$

$$\text{e } R_{22}^{-1}R_{21}R_{11}^{-1}R_{12} = \begin{bmatrix} 0,88147 & 0,77281 & 0,43159 \\ -0,01126 & 0,05057 & -0,01459 \\ -0,10595 & -0,06701 & 0,29178 \end{bmatrix}.$$

Note que a matriz acima não é simétrica.

A equação característica para determinação dos autovalores da matriz $R_{22}^{-1}R_{21}R_{11}^{-1}R_{12}$ é:

$$\left| R_{22}^{-1}R_{21}R_{11}^{-1}R_{12} - \lambda I \right| = 0$$

Os autovalores são: $\lambda_1 = 0,7799$

$$\lambda_2 = 0,3818$$

$$\lambda_3 = 0,0622$$

As estimativas dos coeficientes de correlação canônica são, portanto:

$$r_1 = \sqrt{0,7799} = 0,883$$

$$r_2 = \sqrt{0,3818} = 0,618$$

$$r_3 = \sqrt{0,0622} = 0,249$$

Agora, podem ser calculados os autovetores da matriz $R_{22}^{-1}R_{21}R_{11}^{-1}R_{12}$ que são os autovetores \tilde{b}_1 , \tilde{b}_2 e \tilde{b}_3 das três variáveis canônicas de índices de preços.

Resolvendo-se a equação

$$\left[R_{22}^{-1}R_{21}R_{11}^{-1}R_{12} - \lambda_j I \right] \tilde{b}_j = 0, \quad j = 1, 2, 3$$

$$\left[R_{22}^{-1}R_{21}R_{11}^{-1}R_{12} \right] \tilde{b}_j = \lambda_j \tilde{b}_j, \text{ com } \tilde{b}_j \text{ normalizado por } \tilde{b}_j' R_{22} \tilde{b}_j = 1, \text{ obteve-se:}$$

$$\tilde{b}_1 = \begin{bmatrix} 1,2510 \\ -0,0139 \\ -0,2697 \end{bmatrix}; \quad \tilde{b}_2 = \begin{bmatrix} -1,9602 \\ -0,0363 \\ 2,3344 \end{bmatrix}; \quad \tilde{b}_3 = \begin{bmatrix} -5,8481 \\ 6,6267 \\ -0,7645 \end{bmatrix}$$

As variáveis canônicas de índice de preços são:

$$V_1 = 1,2510y_1 - 0,0139y_2 - 0,2697y_3$$

$$V_2 = -1,9602y_1 - 0,0363y_2 + 2,3344y_3$$

$$V_3 = -5,8481y_1 + 6,6267y_2 - 0,7645y_3.$$

Os coeficientes das variáveis canônicas do primeiro conjunto podem ser obtidos calculando-se os autovetores de $R_{11}^{-1}R_{12}R_{22}^{-1}R_{21}$. Entretanto, como visto anteriormente, também podem ser calculados como:

$$\tilde{a}_j = \frac{1}{\sqrt{\lambda_j}} R_{11}^{-1} R_{12} \tilde{b}_j,$$

sendo \underline{a}_j o vetor dos coeficientes da variável canônica de ordem j . Assim, obteve-se:

$$\underline{a}_1 = \begin{bmatrix} -1,5513 \\ 0,5269 \\ 0,8322 \\ 0,0297 \end{bmatrix}; \quad \underline{a}_2 = \begin{bmatrix} 1,4247 \\ 1,3340 \\ -1,7765 \\ 0,2192 \end{bmatrix}; \quad \underline{a}_3 = \begin{bmatrix} 1,4415 \\ 0,7099 \\ -3,0660 \\ 0,8032 \end{bmatrix}$$

As equações das variáveis canônicas de índices de produção são:

$$U_1 = -1,5513x_1 + 0,5269x_2 + 0,8322x_3 + 0,0297x_4$$

$$U_2 = 1,4247x_1 + 1,3340x_2 - 1,7765x_3 + 0,2192x_4$$

$$U_3 = 1,4415x_1 + 0,7099x_2 - 3,0660x_3 + 0,8032x_4.$$

É interessante ressaltar que sempre haverá duas soluções, uma delas obtida da outra multiplicada por -1, mas as conclusões serão exatamente as mesmas.

O teste de significância para $H_0 : \rho_1 = \rho_2 = \rho_3 = 0$ pode ser realizado pela estatística Qui-quadrado (χ^2), como apresentado a seguir:

$$\chi^2 = -\left[n - \frac{1}{2}(p + q + 3) \right] \ln \prod_{j=1}^k (1 - r_j^2)$$

$$\chi^2 = -\left[21 - \frac{1}{2}(4 + 3 + 3) \right] \ln[(1 - 0,7799)(1 - 0,3818)(1 - 0,0622)]$$

$$\chi^2 = -16 \ln 0,1276 = 32,941, \text{ associado a } pq = 12 \text{ graus de liberdade.}$$

Para $\alpha = 1\%$, $\chi^2_{1\%}(12) = 26,22$. Logo, rejeite-se H_0 .

Comentário: Na prática, em estudos de correlações genéticas pode ocorrer $|r_j| \geq 1$. Para aplicação do teste neste caso, utiliza-se o artifício de assumir como 0,999 o coeficiente de correlação cuja estimativa em valor absoluto for maior ou igual a 1. CRUZ et al. (2012), trabalhando com uma matriz de correlações genotípicas entre caracteres agrônômicos obtiveram $r_1 = 1,0122$.

Uma vez que a hipótese anterior foi rejeitada, avalia-se a significância de uma nova hipótese, dada por ($s = 1$), como a seguir:

$H_0 : \rho_1 > 0 \text{ e } \rho_2 = \rho_3 = 0$, por meio da estatística

$$\chi^2 = -\left[n - \frac{1}{2}(p + q + 3) \right] \ln \prod_{j=s+1}^k (1 - r_j^2)$$

$$\chi^2 = -\left[21 - \frac{1}{2}(4 + 3 + 3)\ln \prod_{j=2}^3 (1 - r_j^2)\right]$$

$$\chi^2 = -16\ln[(1 - 0,3818)(1 - 0,0622)]$$

$\chi^2 = -16\ln 0,5797 = 8,724$, associado a $(p - s)(q - s) = (4 - 1)(3 - 1) = 6$ graus de liberdade.

Para $\alpha = 5\%$, $\chi_{5\%}^2(6) = 12,59$. Logo, a este nível, não se rejeita H_0 .

Uma vez que para $s = 1$, não se rejeitou H_0 , não há necessidade de aplicar o teste para $s = 2$, pois certamente o resultado será não significativo.

Esses resultados indicam que somente o primeiro par canônico é estatisticamente relevante. Assim, de maneira resumida, têm-se os resultados apresentados na Tabela 8.1.

Tabela 8.1 - Correlações canônicas e pares canônicos estimados entre índices de produção (Grupo I) e índices de preços (Grupo II)

Variáveis	Pares Canônicos		
	1 ^o	2 ^o	3 ^o
Produção de bens duráveis	-1,5513	1,4247	1,4415
Produção de bens não-duráveis	0,5269	1,3340	0,7099
Produção mineral	0,8322	-1,7765	-3,0660
Produção agrícola	0,0297	0,2192	0,8032
Preços agrícolas	1,2510	-1,9602	-5,8481
Preços dos alimentos	-0,0139	-0,0363	6,6267
Outros preços	-0,2697	2,3344	-0,7645
r	0,883	0,618	0,249
χ^2	32,941**	8,724 n.s.	--

** Significativo ao nível de 1% de probabilidade.

n.s. Não significativo ao nível de 5% de probabilidade.

Conclui-se que os grupos considerados não são independentes. Somente a primeira correlação canônica é significativa ($P < 0,01$), ou ainda, somente o primeiro par canônico é estatisticamente relevante. Esses resultados mostram que os índices de produção e de preços são fortemente correlacionados, pois a correlação do primeiro par canônico quase atinge 0,9. O exame das variáveis que compõem o primeiro par canônico apresenta, do lado dos preços, predomínio

absoluto dos preços agrícolas, enquanto do lado da produção o índice de produção agrícola é o que apresenta contribuição desprezível.

A interpretação através das correlações entre as variáveis originais e as canônicas fornece, em geral, um caminho mais adequado. Entretanto, existem problemas reais de interpretação quando as variáveis de pelo menos um dos conjuntos são altamente correlacionadas.

Quando a relação entre dois conjuntos de variáveis for hipotetizada e precisar ser testada, essa é a técnica adequada.

• Resultados obtidos através do programa SAS para o exemplo 1

(a) Programa

```

OPTIONS NODATE NONUMBER;
/* CORRELACAO CANONICA - EXEMPLO 1 */
TITLE 'CANONICAL CORRELATION ANALYSIS';
DATA EST746 (TYPE=CORR);
  _TYPE_='CORR';
  INPUT _NAME_ $ X1 X2 X3 X4 Y1 Y2 Y3;
  CARDS;
X1 1.0 . . . . .
X2 0.495951 1.0 . . . . .
X3 0.872836 0.768279 1.0 . . . . .
X4 0.481240 0.709807 0.712358 1.0 . . . . .
Y1 -0.436385 0.425782 -0.038273 0.261010 1.0 . . . . .
Y2 -0.427250 0.429576 -0.043762 0.267098 0.987285 1.0 . . . . .
Y3 -0.203390 0.584220 0.138680 0.378452 0.904598 0.914394 1.0
;
PROC CANCORR DATA=EST746 EDF=20 VPREFIX=PROD WPREFIX=PRECO;
VAR X1 X2 X3 X4; WITH Y1 Y2 Y3;
RUN;

```

(b) Resultados da análise

CANONICAL CORRELATION ANALYSIS

The CANCORR Procedure

Canonical Correlation Analysis

	Canonical Correlation	Adjusted Canonical Correlation	Approximate Standard Error	Squared Canonical Correlation
1	0.883123	0.853965	0.049214	0.779906
2	0.617744	0.556755	0.138277	0.381607
3	0.249178	0.187119	0.209723	0.062090

Test of H0: The canonical correlations in
the current row and all
that follow are zero

Eigenvalues of $\text{Inv}(E) * H$
= $\text{CanRsqr} / (1 - \text{CanRsqr})$

Likelihood Approximate									
	Eigenvalue	Difference	Proportion	Cumulative	Ratio	F Value	Num DF	Den DF	Pr > F
1	3.5435	2.9264	0.8383	0.8383	0.12765374	3.66	12	37.332	0.0011
2	0.6171	0.5509	0.1460	0.9843	0.57999688	1.57	6	30	0.1917
3	0.0662		0.0157	1.0000	0.93791027	0.53	2	16	0.5988

Multivariate Statistics and F Approximations

S=3 M=0 N=6

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.12765374	3.66	12	37.332	0.0011
Pillai's Trace	1.22360317	2.76	12	48	0.0063
Hotelling-Lawley Trace	4.22681202	4.68	12	20.545	0.0011
Roy's Greatest Root	3.54351658	14.17	4	16	<.0001

NOTE: F Statistic for Roy's Greatest Root is an upper bound.

The CANCELL Procedure Canonical Correlation Analysis Raw Canonical Coefficients for the VAR Variables

	PROD1	PROD2	PROD3
X1	-1.551310093	1.4246720254	-1.445010018
X2	0.5269117661	1.3340140273	-0.713574808
X3	0.8322942557	-1.776008648	3.0721280951
X4	0.0296367088	0.2189183442	-0.804276068

Raw Canonical Coefficients for the WITH Variables

	PREC01	PREC02	PREC03
Y1	1.2520135371	-1.955160212	5.849537802
Y2	-0.015184782	-0.042126607	-6.626592134
Y3	-0.269434275	2.3351189857	0.7625334623

Standardized Canonical Coefficients for the VAR Variables

	PROD1	PROD2	PROD3
X1	-1.5513	1.4247	-1.4450
X2	0.5269	1.3340	-0.7136
X3	0.8323	-1.7760	3.0721
X4	0.0296	0.2189	-0.8043

Standardized Canonical Coefficients for the WITH Variables

	PREC01	PREC02	PREC03
Y1	1.2520	-1.9552	5.8495
Y2	-0.0152	-0.0421	-6.6266
Y3	-0.2694	2.3351	0.7625

Canonical Structure

Correlations Between the VAR Variables and Their Canonical Variables

	PROD1	PROD2	PROD3
X1	-0.5493	0.6415	0.4955
X2	0.4180	0.8315	0.3591
X3	-0.0958	0.6483	0.6897
X4	0.2500	0.5863	0.1823

Correlations Between the WITH Variables and Their Canonical Variables

	PREC01	PREC02	PREC03
Y1	0.9933	0.1156	-0.0030
Y2	0.9745	0.1628	-0.1542
Y3	0.8492	0.5280	-0.0053

Correlations Between the VAR Variables and the Canonical Variables of the WITH Variables

	PREC01	PREC02	PREC03
X1	-0.4851	0.3963	0.1235
X2	0.3692	0.5137	0.0895
X3	-0.0846	0.4005	0.1719
X4	0.2208	0.3622	0.0454

Correlations Between the WITH Variables and the Canonical Variables of the VAR Variables

	PROD1	PROD2	PROD3
Y1	0.8772	0.0714	-0.0008
Y2	0.8606	0.1006	-0.0384
Y3	0.7500	0.3261	-0.0013

Exemplo 2

(a) Programa

```
DM 'output; clear; log; clear;';
options FORMDLIM='-' nodate nonumber;
data timmnh1;
input N NA SS SAT PEA RAV; /* Fonte: TIMM, N.H.(1975) */
cards;
```

```
1 12 16 49 48 8
2 30 27 47 76 13
3 16 16 11 40 13
4 17 8 9 52 9
5 26 17 69 63 15
6 34 25 35 82 14
7 23 18 6 71 21
8 19 14 8 68 8
9 16 13 49 74 11
10 26 25 8 70 15
11 35 24 47 70 15
```

```

12 15 14 6 61 11
13 27 21 14 54 12
14 20 17 30 55 13
15 26 22 4 54 10
16 14 8 24 40 14
17 35 27 19 66 13
18 14 16 45 54 10
19 27 26 22 64 14
20 18 10 16 47 16
21 14 18 32 48 16
22 26 26 37 52 14
23 23 23 47 74 19
24 11 8 5 57 12
25 15 17 6 57 10
26 28 21 60 80 11
27 34 23 58 78 13
28 23 11 6 70 16
29 12 8 16 47 14
30 32 32 45 94 19
31 25 14 9 63 11
32 29 21 69 76 16
33 23 24 35 59 11
34 19 12 19 55 8
35 18 18 58 74 14
36 31 26 58 71 17
37 15 14 79 54 14

```

```

;
PROC CANCORR DATA=TIMMNH1 OUT=CSCORES OUTSTAT=CANSTAT;
VAR SAT PEA RAV;
WITH NA SS;
PROC PRINT DATA=CSCORES;
TITLE 'ESCORES CANONICOS';
PROC PRINT DATA=CANSTAT;
TITLE 'ESTATISTICAS DA CORRELAÇÃO CANONICA';
RUN;

```

(b) Resultado da análise

ESTATISTICAS DA CORRELAÇÃO CANONICA										
The CANCORR Procedure										
Canonical Correlation Analysis										
	Canonical	Adjusted	Approximate	Squared						
	Correlation	Canonical	Standard	Canonical						
		Correlation	Error	Correlation						
1	0.688892	0.661186	0.087571	0.474573						
2	0.193591	0.133346	0.160420	0.037477						
					Test of H0: The canonical correlations in					
					the current row and all					
					that follow are zero					
Eigenvalues of Inv(E)*H										
= CanRsq/(1-CanRsq)										
Likelihood Approximate										
Eigenvalue	Difference	Proportion	Cumulative	Ratio	F Value	Num	DF	Den	DF	Pr > F
1	0.9032	0.8643	0.9587	0.9587	0.50573565	4.33	6	64		0.0010
2	0.0389		0.0413	1.0000	0.96252272	0.64	2	33		0.5325

Multivariate Statistics and F Approximations

S=2 M=0 N=15

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.50573565	4.33	6	64	0.0010
Pillai's Trace	0.51205005	3.79	6	66	0.0027
Hotelling-Lawley Trace	0.94214963	4.95	6	40.936	0.0007
Roy's Greatest Root	0.90321312	9.94	3	33	<.0001

NOTE: F Statistic for Roy's Greatest Root is an upper bound.

NOTE: F Statistic for Wilks' Lambda is exact.

The CANCERR Procedure
Canonical Correlation Analysis

Raw Canonical Coefficients for the VAR Variables

	V1	V2
SAT	0.0023512268	0.0449702569
PEA	0.0717779903	-0.047200647
RAV	0.059151479	0.1009824063

Raw Canonical Coefficients for the WITH Variables

	W1	W2
NA	0.1070291672	-0.200934083
SS	0.0417878556	0.2554802154

Standardized Canonical Coefficients for the VAR Variables

	V1	V2
SAT	0.0520	0.9939
PEA	0.8991	-0.5912
RAV	0.1831	0.3125

Standardized Canonical Coefficients for the WITH Variables

	W1	W2
NA	0.7752	-1.4554
SS	0.2662	1.6274

Observação: A interpretação baseada na correlação entre cada variável e as variáveis canônicas é o critério mais amplamente recomendado, mas ele é o menos usado. Por várias razões, Rencher (2002), recomenda que a interpretação deva ser feita com base nos coeficientes canônicos padronizados e não nos coeficientes canônicos brutos. Existe controvérsia. No SAS, entrando com a matriz R os coeficientes canônicos brutos coincidem com os padronizados.

The CANCELL Procedure
Canonical Structure

Correlations Between the VAR Variables and Their Canonical Variables

	V1	V2
SAT	0.4236	0.8410
PEA	0.9833	-0.1123
RAV	0.5131	0.3129

Correlations Between the WITH Variables and Their Canonical Variables

	W1	W2
NA	0.9869	-0.1614
SS	0.8826	0.4701

Correlations Between the VAR Variables and the Canonical Variables of the WITH Variables

	W1	W2
SAT	0.2918	0.1628
PEA	0.6774	-0.0217
RAV	0.3534	0.0606

Correlations Between the WITH Variables and the Canonical Variables of the VAR Variables

	V1	V2
NA	0.6799	-0.0312
SS	0.6080	0.0910

ESCORES CANONICOS										
Obs	N	NA	SS	SAT	PEA	RAV	V1	V2	W1	W2
1	1	12	16	49	48	8	-1.31991	0.95926	-1.21018	1.47774
2	2	30	27	47	76	13	0.98093	0.05261	1.17601	0.67121
3	3	16	16	11	40	13	-1.68772	0.13291	-0.78206	0.67400
4	4	17	8	9	52	9	-1.06770	-0.92737	-1.00933	-1.57077
5	5	26	17	69	63	15	0.21785	1.85753	0.33002	-1.07986
6	6	34	25	35	82	14	1.44253	-0.66925	1.52056	-0.64349
7	7	23	18	6	71	21	0.99885	-0.74730	0.05072	-0.22157
8	8	19	14	8	68	8	0.01925	-1.82853	-0.54455	-0.43976
9	9	16	13	49	74	11	0.72377	0.03499	-0.90742	-0.09244
10	10	26	25	8	70	15	0.57687	-1.21606	0.66432	0.96399
11	11	35	24	47	70	15	0.66856	0.53778	1.58580	-1.09990
12	12	15	14	6	61	11	-0.31044	-1.28512	-0.97266	0.36398
13	13	27	21	14	54	12	-0.73493	-0.49397	0.60420	-0.25887
14	14	20	17	30	55	13	-0.56638	0.27933	-0.31216	0.12575
15	15	26	22	4	54	10	-0.87674	-1.14564	0.53896	0.19755
16	16	14	8	24	40	14	-1.59801	0.81850	-1.33042	-0.96797
17	17	35	27	19	66	13	0.19731	-0.73455	1.71116	-0.33346
18	18	14	16	45	54	10	-0.78034	0.69814	-0.99612	1.07587

19	19	27	26	22	64	14	0.11996	-0.40425	0.81314	1.01853
20	20	18	10	16	47	16	-0.99607	0.33030	-0.81873	-1.26074
21	21	14	18	32	48	16	-0.88667	1.00263	-0.91254	1.58683
22	22	26	26	37	52	14	-0.70610	0.83671	0.70611	1.21947
23	23	23	23	47	74	19	1.19228	0.75291	0.25966	1.05583
24	24	11	8	5	57	12	-0.54076	-1.04031	-1.65151	-0.36517
25	25	15	17	6	57	10	-0.65671	-1.19730	-0.84730	1.13042
26	26	28	21	60	80	11	1.18030	0.24646	0.71123	-0.45980
27	27	34	23	58	78	13	1.15035	0.45289	1.43698	-1.15445
28	28	23	11	6	70	16	0.63132	-1.20502	-0.24179	-2.00993
29	29	12	8	16	47	14	-1.11437	0.12834	-1.54448	-0.56610
30	30	32	32	45	94	19	2.62314	-0.28104	1.59901	1.54674
31	31	25	14	9	63	11	-0.15983	-1.24461	0.09763	-1.64536
32	32	29	21	69	76	16	1.21011	1.34491	0.81826	-0.66074
33	33	23	24	35	59	11	-0.38581	0.11342	0.30145	1.31131
34	34	19	12	19	55	8	-0.88800	-0.72025	-0.62812	-0.95072
35	35	18	18	58	74	14	0.92239	0.74267	-0.48443	0.78310
36	36	31	26	58	71	17	0.88451	1.18722	1.24126	0.21480
37	37	15	14	79	54	14	-0.46380	2.63106	-0.97266	0.36398

ESTATÍSTICAS DA CORRELAÇÃO CANÔNICA

Obs	_TYPE_	_NAME_	SAT	PEA	RAV	NA	SS
1	MEAN		31.2703	62.6486	13.2432	22.3784	18.3784
2	STD		22.1003	12.5260	3.0948	7.2432	6.3699
3	N		37.0000	37.0000	37.0000	37.0000	37.0000
4	CORR	SAT	1.0000	0.3703	0.2114	0.2617	0.3341
5	CORR	PEA	0.3703	1.0000	0.3548	0.6720	0.5876
6	CORR	RAV	0.2114	0.3548	1.0000	0.3390	0.3404
7	CORR	NA	0.2617	0.6720	0.3390	1.0000	0.7951
8	CORR	SS	0.3341	0.5876	0.3404	0.7951	1.0000
9	CANCORR		0.6889	0.1936	.	.	.
10	SCORE	V1	0.0520	0.8991	0.1831	0.0000	0.0000
11	SCORE	V2	0.9939	-0.5912	0.3125	0.0000	0.0000
12	SCORE	W1	0.0000	0.0000	0.0000	0.7752	0.2662
13	SCORE	W2	0.0000	0.0000	0.0000	-1.4554	1.6274
14	RAWSCORE	V1	0.0024	0.0718	0.0592	0.0000	0.0000
15	RAWSCORE	V2	0.0450	-0.0472	0.1010	0.0000	0.0000
16	RAWSCORE	W1	0.0000	0.0000	0.0000	0.1070	0.0418
17	RAWSCORE	W2	0.0000	0.0000	0.0000	-0.2009	0.2555
18	STRUCTUR	V1	0.4236	0.9833	0.5131	0.6799	0.6080
19	STRUCTUR	V2	0.8410	-0.1123	0.3129	-0.0312	0.0910
20	STRUCTUR	W1	0.2918	0.6774	0.3534	0.9869	0.8826
21	STRUCTUR	W2	0.1628	-0.0217	0.0606	-0.1614	0.4701

8.5. EXERCÍCIOS

Num estudo envolvendo $n=20$ linhagens de feijão, foram avaliadas as seguintes variáveis:

(a) Componentes primários da produção:

X_1 = número de vagens por planta

X_2 = número de grãos por vagem

X_3 = peso médio de grãos

(b) Componentes secundários da produção:

Y_1 = área foliar

Y_2 = número de folhas por planta

Y_3 = altura da planta

Y_4 = peso total da palhada

A matriz das estimativas dos coeficientes de correlação genotípica entre os sete caracteres agronômicos avaliados na cultura de feijão, está apresentada na Tabela 8.2.

Tabela 8.2 – Estimativas dos coeficientes de correlação genotípica entre sete caracteres agronômicos, avaliados na cultura de feijão *

Caracteres	X_1	X_2	X_3	Y_1	Y_2	Y_3	Y_4
X_1	1,0000	0,6924	-0,4851	0,5205	0,8573	0,6965	0,3510
X_2		1,0000	-0,8878	0,6206	0,7511	0,8314	-0,0815
X_3			1,0000	-0,3455	-0,5971	-0,7860	0,4015
Y_1				1,0000	0,6394	0,4790	0,5271
Y_2					1,0000	0,7001	0,3841
Y_3						1,0000	0,1480
Y_4							1,0000

*Fonte: Cruz, et al. (2012)

Pede-se (usar α até 5%):

1. Faça um estudo utilizando correlações canônicas. Como teste de significância, usar a aproximação Qui-quadrado do teste da razão de verossimilhança (Lambda de Wilks). **Discutir os resultados.**

2. Rodar o exemplo no programa SAS. No teste para as correlações, o SAS apresenta uma aproximação F (chamada F de Rao) da estatística Lambda de Wilks dada por Rao (1973, p. 556). Lembre-se que no programa, será preciso informar EDF=19. Faça o teste à mão, reproduzindo os resultados obtidos pelo programa.

8.6. REFERÊNCIAS

- CLEMENTE, A. **Pesquisas de variáveis múltiplas**. Curitiba: Scientia et Labor, UFPR, 1990. 204p.
- CRUZ, C. D.; REGAZZI, A. J.; CARNEIRO, P. C. S. **Modelos biométricos aplicados ao melhoramento genético**. v.1, 3. ed., Viçosa: UFV, 2012. 514p.
- JOHNSON, R. A.; WICHERN, D.W. **Applied multivariate statistical analysis**. 4th ed., USA, Upper Saddle River, New Jersey: Prentice Hall, Inc., 1998. 816p.
- KHATTREE, R.; NAIK, D. N. **Applied multivariate statistical with SAS Software**. 2nd ed. A copublication of Cary, NC, SAS Institute Inc. and New York: John Wiley and Sons, 1999. 338p.
- KHATTREE, R.; NAIK, D. N. **Multivariate data reduction and discrimination with SAS Software**. USA: SAS Institute Inc., Cary, NC, 2000. 558p.
- MARDIA, K. V.; KENT, J. T.; BIBBY, J. M. **Multivariate analysis**. 6th ed., Londres: Academic Press, 1997. 518p.
- RAO, R.C. **Advanced statistical methods in biometric research**. New York: John Wiley and Sons, 1952. 390p.
- RAO, R.C. **Linear statistical inference and its applications**. New York: John Wiley and Sons, 1973. 687p.
- RENCER, A. C. **Methods of multivariate analysis**. 2nd ed., USA: John Wiley and Sons, 2002. 738p.
- TIMM, N.H. **Multivariate analysis with applications in education and psychology**. Belmont, California: Wadsworth Publishing Company, 1975. 689p.

CAPÍTULO 9

ANÁLISE DE FATORES

9.1. INTRODUÇÃO

A análise de fatores ou análise fatorial (AF) é uma técnica de análise multivariada que começou a ser desenvolvida no início do século 20, a partir das tentativas de Karl Pearson, Charles Spearman, entre outros, para definir e medir a inteligência (Johnson e Wichern, 1998). Seu desenvolvimento e, principalmente, sua utilização, foram limitados durante muitos anos, devido à complexidade dos cálculos envolvidos associados à técnica. A partir do advento de processamento de dados computadorizados, foi renovado o interesse por aspectos teóricos e computacionais da análise de fatores.

O objetivo principal da análise de fatores é descrever, se possível, as relações de covariância entre muitas variáveis com base num pequeno número de quantidades aleatórias, não observáveis, chamadas **fatores** (Johnson e Wichern, 1998). Os fatores podem ser não-correlacionados (fatores ortogonais) ou correlacionados (fatores oblíquos). As variáveis são agrupadas por meio de suas correlações, ou seja, aquelas pertencentes a um mesmo grupo serão fortemente correlacionadas entre si, mas pouco correlacionadas com as variáveis de outro grupo. Cada grupo de variáveis representará um fator, responsável pelas correlações observadas.

Admite-se que a relação entre variáveis observadas e fatores é linear. A análise de fatores será encarada aqui como uma técnica estatística exploratória destinada a “resumir” as informações contidas em um conjunto de variáveis em um conjunto de fatores, com o número de fatores sendo geralmente bem menor que o número de variáveis observadas.

A análise de fatores ocupa-se em explicitar a estrutura de relações. Pode-se dizer que a análise de componentes principais (ACP) é a análise da variação, enquanto a análise de fatores (AF) é a análise da covariação.

As etapas da análise de fatores podem ser resumidas em:

- i) Determinação da matriz de correlações entre todas as variáveis;
- ii) Extração dos fatores necessários para representar os dados;
- (iii) Transformação (rotação) dos fatores, de modo a torná-los mais facilmente interpretáveis;
- (iv) Determinação dos escores fatoriais.

É importante ressaltar que a análise de fatores também pode ser empregada a partir da matriz de covariâncias.

9.2. O MODELO DA ANÁLISE DE FATORES ORTOGONAL COM m FATORES COMUNS

Considerando um conjunto de p variáveis com n observações para cada variável, obtém-se o arranjo de valores $[X_{ij}]$, $i = 1, 2, \dots, n$, $j = 1, 2, \dots, p$, que podem ser dispostos de acordo com a Tabela 9.1.

Tabela 9.1 – Conjunto de n observações para cada uma das p variáveis

Indivíduos	Variáveis			
	X_1	X_2	\dots	X_p
1	X_{11}	X_{12}	\dots	X_{1p}
2	X_{21}	X_{22}	\dots	X_{2p}
\dots	\dots	\dots	\dots	\dots
n	X_{n1}	X_{n2}	\dots	X_{np}

Seja $\tilde{X} = [X_1 \ X_2 \ \dots \ X_p]'$ um vetor aleatório observável $p \times 1$, com média $\underline{\mu}$ e matriz de covariância Σ . O modelo da AF postula que \tilde{X} é linearmente dependente de m variáveis aleatórias não observáveis F_1, F_2, \dots, F_m chamadas **fatores comuns** e p variáveis aleatórias não observáveis (fontes de variação adicionais) $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p$, chamadas de **erros** ou, algumas vezes, **fatores específicos**.

O modelo da análise de fatores é:

$$X_1 - \mu_1 = \ell_{11}F_1 + \ell_{12}F_2 + \dots + \ell_{1m}F_m + \varepsilon_1$$

$$X_2 - \mu_2 = \ell_{21}F_1 + \ell_{22}F_2 + \dots + \ell_{2m}F_m + \varepsilon_2$$

$$\vdots \qquad \qquad \qquad \vdots$$

$$X_p - \mu_p = \ell_{p1}F_1 + \ell_{p2}F_2 + \dots + \ell_{pm}F_m + \varepsilon_p$$

Utilizando a notação matricial, pode-se escrever: $\tilde{X} - \underline{\mu} = L\tilde{F} + \tilde{\varepsilon}$,

em que,

$$\tilde{\mathbf{X}} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix} \text{ é um vetor } p \times 1 \text{ de variáveis aleatórias observáveis;}$$

$$\tilde{\boldsymbol{\mu}} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix} \text{ é um vetor } p \times 1 \text{ de médias;}$$

$$\mathbf{L} = \begin{bmatrix} \ell_{11} & \ell_{12} & \cdots & \ell_{1m} \\ \ell_{21} & \ell_{22} & \cdots & \ell_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ \ell_{p1} & \ell_{p2} & \cdots & \ell_{pm} \end{bmatrix} \text{ é uma matriz } p \times m \text{ de cargas fatoriais ("Factor Loadings");}$$

$$\tilde{\mathbf{F}} = \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_m \end{bmatrix} \text{ é um vetor } m \times 1 \text{ de variáveis aleatórias não observáveis;}$$

$$\tilde{\boldsymbol{\varepsilon}} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_p \end{bmatrix} \text{ é um vetor } p \times 1 \text{ de erros aleatórios não observáveis.}$$

E ainda,

μ_i é a média da i -ésima variável aleatória X_i ;

ε_i é o i -ésimo fator específico;

F_j é o j -ésimo fator comum;

ℓ_{ij} é a carga da i -ésima variável no j -ésimo fator.

Note que o i -ésimo fator específico ε_i está associado com a i -ésima resposta X_i . Os p desvios $X_1 - \mu_1, X_2 - \mu_2, \dots, X_p - \mu_p$ são expressos em termos de $p+m$ variáveis aleatórias não observáveis $F_1, F_2, \dots, F_m, \varepsilon_1, \varepsilon_2, \dots, \varepsilon_p$, e com algumas pressuposições acerca dos vetores aleatórios $\tilde{\mathbf{F}}$ e $\tilde{\boldsymbol{\varepsilon}}$, o modelo da análise de fatores implica numa estrutura de covariância para $\tilde{\mathbf{X}}$ que pode ser avaliada.

Os vetores aleatórios não observáveis \tilde{F} e $\tilde{\varepsilon}$ devem satisfazer as seguintes condições:

$$\tilde{F} \text{ e } \tilde{\varepsilon} \text{ são independentes} \Rightarrow \text{Cov}(\tilde{\varepsilon}, \tilde{F}) = E(\tilde{\varepsilon} \tilde{F}') = \underset{(p \times m)}{\mathbf{0}}, \text{ pois,}$$

$$E(\tilde{F}) = \underset{(m \times 1)}{\mathbf{0}}, \text{ Cov}(\tilde{F}) = E[\tilde{F} \tilde{F}'] = \underset{(m \times m)}{\mathbf{I}}$$

$$E(\tilde{\varepsilon}) = \underset{(p \times 1)}{\mathbf{0}}, \text{ Cov}(\tilde{\varepsilon}) = E[\tilde{\varepsilon} \tilde{\varepsilon}'] = \underset{(p \times p)}{\Psi} = \begin{bmatrix} \psi_1 & 0 & \dots & 0 \\ 0 & \psi_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \psi_p \end{bmatrix}.$$

O modelo da análise de fatores ortogonal implica numa estrutura de covariância para \tilde{X} .

Do modelo anterior, segue que:

$$\begin{aligned} (\tilde{X} - \underset{\sim}{\mu}) (\tilde{X} - \underset{\sim}{\mu})' &= (\tilde{L} \tilde{F} + \tilde{\varepsilon}) (\tilde{L} \tilde{F} + \tilde{\varepsilon})' = (\tilde{L} \tilde{F} + \tilde{\varepsilon}) ((\tilde{L} \tilde{F})' + \tilde{\varepsilon}') \\ &= \tilde{L} \tilde{F} (\tilde{L} \tilde{F})' + \tilde{\varepsilon} (\tilde{L} \tilde{F})' + \tilde{L} \tilde{F} \tilde{\varepsilon}' + \tilde{\varepsilon} \tilde{\varepsilon}' \\ &= \tilde{L} (\tilde{F} \tilde{F}') \tilde{L}' + (\tilde{\varepsilon} \tilde{F}') \tilde{L}' + \tilde{L} (\tilde{F} \tilde{\varepsilon}') + \tilde{\varepsilon} \tilde{\varepsilon}'. \end{aligned}$$

Assim, virá:

$$\begin{aligned} \text{i) } \Sigma &= \text{Cov}(\tilde{X}) = E \left[(\tilde{X} - \underset{\sim}{\mu}) (\tilde{X} - \underset{\sim}{\mu})' \right] \\ &= E(\tilde{L} \tilde{F} \tilde{F}') \tilde{L}' + E(\tilde{\varepsilon} \tilde{F}') \tilde{L}' + E(\tilde{L} \tilde{F} \tilde{\varepsilon}') + E(\tilde{\varepsilon} \tilde{\varepsilon}') \\ &= \tilde{L} \tilde{L}' + \Psi. \end{aligned}$$

$$\text{Logo, } \text{Var}(X_i) = \ell_{i1}^2 + \dots + \ell_{im}^2 + \psi_i = \sum_{j=1}^m \ell_{ij}^2 + \psi_i;$$

$$\text{Cov}(X_i, X_k) = \ell_{i1} \ell_{k1} + \dots + \ell_{im} \ell_{km} = \sum_{j=1}^m \ell_{ij} \ell_{kj}.$$

$$\begin{aligned} \text{ii) } \text{Cov}(\tilde{X}, \tilde{F}') &= E \left[(\tilde{X} - \underset{\sim}{\mu}) (\tilde{F} - \underset{\sim}{0})' \right] = E \left[(\tilde{L} \tilde{F} + \tilde{\varepsilon}) \tilde{F}' \right] \\ &= E(\tilde{L} \tilde{F} \tilde{F}' + \tilde{\varepsilon} \tilde{F}') = \tilde{L} E(\tilde{F} \tilde{F}') + E(\tilde{\varepsilon} \tilde{F}') \\ &= \tilde{L}. \end{aligned}$$

$$\text{Logo, } \text{Cov}(X_i, F_j) = \ell_{ij}.$$

Por definição, a i -ésima **comunalidade** (h_i^2) é a porção da variância da i -ésima variável devida aos m fatores comuns. A porção de $\text{Var}(X_i) = \sigma_{ii}$ devida ao fator específico ε_i é chamada de **variância específica**. Logo,

$$\underbrace{\sigma_{ii}}_{\text{Var}(X_i)} = \underbrace{\ell_{i1}^2 + \ell_{i2}^2 + \dots + \ell_{im}^2}_{\text{Comunalidade}} + \underbrace{\psi_i}_{\text{Variância Específica}} = h_i^2 + \psi_i, \text{ para } i = 1, 2, \dots, p.$$

Assim,

$$h_i^2 = \ell_{i1}^2 + \ell_{i2}^2 + \dots + \ell_{im}^2,$$

isto é, a *i*-ésima **comunalidade** é a soma dos quadrados das cargas da *i*-ésima variável nos **m** fatores comuns.

Para verificar a relação $\Sigma = LL' + \Psi$ para dois fatores, considere a matriz de covariância a seguir:

$$\Sigma = \begin{bmatrix} 19 & 30 & 2 & 12 \\ 30 & 57 & 5 & 23 \\ 2 & 5 & 38 & 47 \\ 12 & 23 & 47 & 68 \end{bmatrix}.$$

A igualdade,

$$\begin{bmatrix} 19 & 30 & 2 & 12 \\ 30 & 57 & 5 & 23 \\ 2 & 5 & 38 & 47 \\ 12 & 23 & 47 & 68 \end{bmatrix} = \begin{bmatrix} 4 & 1 \\ 7 & 2 \\ -1 & 6 \\ 1 & 8 \end{bmatrix} \begin{bmatrix} 4 & 7 & -1 & 1 \\ 1 & 2 & 6 & 8 \end{bmatrix} + \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}.$$

Assim,

$$\Sigma = LL' + \Psi,$$

em que,

$$L = \begin{bmatrix} \ell_{11} & \ell_{12} \\ \ell_{21} & \ell_{22} \\ \ell_{31} & \ell_{32} \\ \ell_{41} & \ell_{42} \end{bmatrix} = \begin{bmatrix} 4 & 1 \\ 7 & 2 \\ -1 & 6 \\ 1 & 8 \end{bmatrix}; \Psi = \begin{bmatrix} \psi_1 & 0 & 0 & 0 \\ 0 & \psi_2 & 0 & 0 \\ 0 & 0 & \psi_3 & 0 \\ 0 & 0 & 0 & \psi_4 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}.$$

A comunalidade de X_1 é:

$$h_1^2 = \ell_{11}^2 + \ell_{12}^2 = 4^2 + 1^2 = 17.$$

A variância de X_1 pode ser decomposta como:

$$\sigma_{11} = (\ell_{11}^2 + \ell_{12}^2) + \psi_1 = h_1^2 + \psi_1$$

ou

$$\underbrace{19}_{\text{Variância}} = \underbrace{4^2 + 1^2}_{\text{Comunalidade}} + \underbrace{2}_{\text{Variância Específica}} = 17 + 2.$$

- O modelo fatorial assume que $p + p(p-1)/2$ variâncias e covariâncias para \tilde{X} podem ser reproduzidas a partir de $p \cdot m$ cargas fatoriais ℓ_{ij} e p variâncias específicas ψ_i .
- Para $m = p \Rightarrow \Sigma = LL'$, $\Psi = \Phi$
- $m < p \Rightarrow$ a análise de fatores é mais útil.
- Seja ${}_m T_m$ uma matriz ortogonal $\Rightarrow TT' = T'T = I$, e o modelo a seguir:

$$\tilde{X} - \tilde{\mu} = LTT'F + \tilde{\varepsilon} = L^* \tilde{F}^* + \tilde{\varepsilon},$$

em que,

$$L^* = LT \text{ e } \tilde{F}^* = T'F.$$

Assim, virá:

$$E(\tilde{F}^*) = T'E(F) = 0$$

$$\text{Cov}(\tilde{F}^*) = T'\text{Cov}(F)T = \underset{(m \times m)}{I}$$

- \tilde{F} e $\tilde{F}^* = T'\tilde{F}$ têm as mesmas propriedades estatísticas.
- L e L^* , em geral são distintas.
- $\Sigma = LL' + \Psi = LTT'L' + \Psi = (L^*)(L^*)' + \Psi$.

$$\text{Logo, } LL' = (L^*)(L^*)'.$$

As Comunalidades não são alteradas pela escolha da matriz ortogonal T .

- A **rotação ortogonal** dos fatores comuns permite, em geral, uma melhor definição dos mesmos, facilitando as interpretações.
- Cargas rotacionadas (L^*) . Os escores fatoriais são estimados utilizando a matriz \hat{L}^* .
- Variáveis Padronizadas: $\Sigma \rightarrow \rho$, isto é, a matriz de covariância entre as variáveis padronizadas é igual à matriz de correlação entre as variáveis originais.

$$\rho = LL' + \Psi \Rightarrow \rho_{ii} = \text{Corr}(X_i, X_i) = \sum_{j=1}^m \ell_{ij}^2 + \psi_i = 1,$$

$$\rho_{ik} = \text{Corr}(X_i, X_k) = \sum_{j=1}^m \ell_{ij} \ell_{kj},$$

$$\text{e } \text{Corr}(\tilde{X}, \tilde{F}') = L, \quad \text{Corr}(X_i, F_j) = \ell_{ij}$$

9.3. MÉTODOS DE ESTIMAÇÃO DAS CARGAS FATORIAIS

9.3.1. Método do componente principal

- Decomposição Espectral de Σ (ou ρ)

$$\Sigma = \sum_{i=1}^p \lambda_i \mathbf{e}_i \mathbf{e}_i' = \lambda_1 \mathbf{e}_1 \mathbf{e}_1' + \lambda_2 \mathbf{e}_2 \mathbf{e}_2' + \dots + \lambda_p \mathbf{e}_p \mathbf{e}_p' ,$$

$$\text{com } \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0, \mathbf{e}_i' \mathbf{e}_i = 1$$

$$\mathbf{e}_i' \mathbf{e}_{i'} = 0, i \neq i'.$$

Para $m = p \Rightarrow$

$$\Sigma = \begin{bmatrix} \sqrt{\lambda_1} \mathbf{e}_1 & \sqrt{\lambda_2} \mathbf{e}_2 & \dots & \sqrt{\lambda_p} \mathbf{e}_p \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} \mathbf{e}_1' \\ \dots \\ \sqrt{\lambda_2} \mathbf{e}_2' \\ \dots \\ \vdots \\ \dots \\ \sqrt{\lambda_p} \mathbf{e}_p' \end{bmatrix} = \mathbf{L} \mathbf{L}' + \Phi = \mathbf{L} \mathbf{L}'$$

$$\Psi = \Phi$$

A j -ésima coluna da matriz de cargas é $\underbrace{\sqrt{\lambda_j} \mathbf{e}_j}_{\text{ACP}}$

- Na análise de fatores ($m < p$) obtém-se a aproximação:

$$\Sigma \doteq \begin{bmatrix} \sqrt{\lambda_1} \mathbf{e}_1 & \sqrt{\lambda_2} \mathbf{e}_2 & \dots & \sqrt{\lambda_p} \mathbf{e}_m \end{bmatrix} \begin{bmatrix} \sqrt{\lambda_1} \mathbf{e}_1' \\ \dots \\ \sqrt{\lambda_2} \mathbf{e}_2' \\ \dots \\ \vdots \\ \dots \\ \sqrt{\lambda_p} \mathbf{e}_m' \end{bmatrix} + \begin{bmatrix} \psi_1 & 0 & \dots & 0 \\ 0 & \psi_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \psi_p \end{bmatrix} = \mathbf{L} \mathbf{L}' + \Psi ,$$

em que $\psi_i = \sigma_{ii} - \sum_{j=1}^m e_{ij}^2$, para $i = 1, 2, \dots, p$.

- Na AF é comum usar as variáveis centradas na média ou padronizadas (média **0** e variância **1**).
- A AF em termos amostrais: S (ou R)

A matriz de cargas fatoriais estimadas $\hat{L} = (\hat{\ell}_{ij})$, $i = 1, 2, \dots, p$; $j = 1, 2, \dots, m$, é dada por:

$$\hat{L} = [\sqrt{\hat{\lambda}_1} \hat{e}_1 : \sqrt{\hat{\lambda}_2} \hat{e}_2 : \dots : \sqrt{\hat{\lambda}_m} \hat{e}_m]$$

As variâncias específicas são obtidas da diagonal da matriz $S - \hat{L}\hat{L}'$, assim:

$$\hat{\Psi} = \begin{bmatrix} \hat{\psi}_1 & 0 & \dots & 0 \\ 0 & \hat{\psi}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \hat{\psi}_p \end{bmatrix}, \text{ com } \hat{\psi}_i = s_{ii} - \underbrace{\sum_{j=1}^m \hat{\ell}_{ij}^2}_{\text{Comunalidade Estimada } (h_i^2)}$$

- Matriz Residual: $[S - (\hat{L}\hat{L}' + \hat{\Psi})]$

Os elementos da diagonal são iguais a zero. Se os demais elementos são de pequena magnitude, pode-se inferir empiricamente, que a aproximação feita é adequada e, portanto, que o modelo com m fatores comuns é apropriado.

- Como definir o n° de fatores comuns?

- a partir de considerações teóricas;
- basear em resultados experimentais anteriores;
- autovalores estimados (igual a ACP)

$$\text{- AF baseada em } S \Rightarrow \sum_{j=1}^m \hat{\lambda}_j / \text{tr}(S)$$

$$\text{- AF baseada em } R \Rightarrow \sum_{j=1}^m \hat{\lambda}_j / p$$

- Uma modificação do Método dos Componentes Principais é o chamado Método dos Fatores Principais. Trabalha-se com a matriz de correlação amostral reduzida $\rightarrow R_r$ é obtida a partir de R substituindo os elementos da diagonal por estimativas das comunalidades das p variáveis.

9.3.2. Método da máxima verossimilhança

É o método que tem melhor fundamentação estatística se for feita a pressuposição de que os fatores comuns e os fatores específicos (\underline{F} e $\underline{\xi}$) têm distribuição normal (Johnson e Wichern, 1998). Este método não será abordado neste texto.

9.4. ROTAÇÃO DOS FATORES

Uma transformação ortogonal corresponde a uma rotação rígida dos eixos das coordenadas e, portanto, uma transformação ortogonal das cargas dos fatores e, consequentemente dos fatores, é chamada **Rotação dos Fatores**.

- **Objetivo:** Obter a matriz de **cargas rotacionadas** $\hat{L}_{p \ m}^* = \hat{L} T$, onde $T T' = T' T = I$.

- Tem-se ainda que:
$$\begin{cases} \hat{L} \hat{L}' + \hat{\Psi} = \hat{L} T T' \hat{L}' + \hat{\Psi} = \hat{L} * \hat{L}' * + \hat{\Psi} \\ S - \hat{L} \hat{L}' - \hat{\Psi} = S - \hat{L} * \hat{L}' * - \hat{\Psi} \end{cases}$$

- A rotação ortogonal não altera a communalidade das variáveis e nem a variância específica.
- Obtém-se \hat{L}^* de maneira que os valores absolutos dos elementos de cada coluna dessa matriz se aproximem, na medida do possível, de zero ou 1. Isto facilita a interpretação, pois cada fator deverá apresentar correlação relativamente forte com uma ou mais variáveis e correlação relativamente fraca com as demais variáveis.
- Modalidades de Rotação:

$$(i) \text{ Ortogonal } \left\{ \begin{array}{l} \text{Varimax} \\ \text{Quartimax} \\ \text{Equamax} \end{array} \right\} \begin{array}{l} \text{Fatores Ortogonais} \\ \text{(Não correlacionados)} \end{array}$$

$$(ii) \text{ Oblíqua } \left\{ \begin{array}{l} \text{Obimin} \\ \text{Promax} \\ \text{Tandem} \end{array} \right\} \begin{array}{l} \text{Fatores não ortogonais} \\ \text{(Correlacionados)} \end{array}$$

Um dos critérios mais usados para obter a matriz $T_{m \ m}$ de transformação ortogonal é o Varimax (Harman, 1976). Como o nome sugere, trata-se de rotação que torna máxima uma variância.

9.5. ESCORES FATORIAIS

Os valores estimados dos fatores comuns são chamados **escores fatoriais**, que são frequentemente usados para propósitos de diagnóstico ou em algum outro tipo de análise.

- **Métodos de Estimação:**

- Método dos mínimos quadrados ponderados
- Método da regressão

Estes dois métodos não serão apresentados aqui e podem ser vistos em Johnson e Wichern (1998).

Se as cargas fatoriais são estimadas pelo método dos componentes principais, é comum estimar os escores fatoriais usando o método dos mínimos quadrados ordinário. Assim,

$$\hat{\tilde{f}}_j = (\hat{\mathbf{L}}' \hat{\mathbf{L}})^{-1} \hat{\mathbf{L}}' (\mathbf{x}_j - \bar{\mathbf{x}})$$

ou $\hat{\tilde{f}}_j = (\hat{\mathbf{L}}'_z \hat{\mathbf{L}}_z)^{-1} \hat{\mathbf{L}}'_z \mathbf{z}_j$, para dados padronizados.

• Para $\hat{\mathbf{L}} = [\sqrt{\hat{\lambda}_1} \hat{\mathbf{e}}_1 : \sqrt{\hat{\lambda}_2} \hat{\mathbf{e}}_2 : \dots : \sqrt{\hat{\lambda}_m} \hat{\mathbf{e}}_m]$, virá:

$$\hat{\tilde{f}}_j = \begin{bmatrix} \frac{1}{\sqrt{\hat{\lambda}_1}} \hat{\mathbf{e}}_1 (\mathbf{x}_j - \bar{\mathbf{x}}) \\ \frac{1}{\sqrt{\hat{\lambda}_2}} \hat{\mathbf{e}}_2 (\mathbf{x}_j - \bar{\mathbf{x}}) \\ \vdots \\ \frac{1}{\sqrt{\hat{\lambda}_m}} \hat{\mathbf{e}}_m (\mathbf{x}_j - \bar{\mathbf{x}}) \end{bmatrix}, \quad j = 1, 2, \dots, n.$$

• Escores padronizados: $\hat{\tilde{f}}_j = (\hat{\mathbf{L}}'_z \hat{\mathbf{L}}_z)^{-1} \hat{\mathbf{L}}'_z \mathbf{z}_j \Rightarrow \begin{cases} \frac{1}{n} \sum_{j=1}^n \hat{\tilde{f}}_j = \phi \\ \frac{1}{n-1} \sum_{j=1}^n \hat{\tilde{f}}_j \hat{\tilde{f}}_j' = \mathbf{I} \end{cases}$

• Usando $\hat{\mathbf{L}}_z^* = \hat{\mathbf{L}}_z \mathbf{T} \Rightarrow \hat{\tilde{f}}_j^* = (\hat{\mathbf{L}}_z^{*'} \hat{\mathbf{L}}_z^*)^{-1} \hat{\mathbf{L}}_z^{*'} \mathbf{z}_j$, $\hat{\tilde{f}}_j^* = \mathbf{T}' \hat{\tilde{f}}_j$.

9.6. COMPARAÇÃO ENTRE AS TÉCNICAS ANÁLISE DE FATORES E ANÁLISE DE COMPONENTES PRINCIPAIS

A Análise de Fatores (AF) e a Análise de Componentes Principais (ACP) têm como principal objetivo a redução da dimensionalidade. Por causa desta similaridade de objetivos entre as duas técnicas, muitos autores consideram a ACP como uma forma alternativa da AF. Este tipo de consideração pode se tornar confuso, pois existem diferenças entre as duas técnicas.

As diferenças entre as duas técnicas podem ser colocadas da seguinte maneira: (1) Na AF as variáveis são expressas como combinações lineares dos fatores, enquanto que os componentes principais são funções lineares das variáveis; (2) na ACP a ênfase é explicar a

variância total $\sum_i s_{ii}$, em contraste com a AF onde o objetivo é tentar explicar as covariâncias;

(3) a ACP essencialmente não requer pressuposições, enquanto que a AF requer várias pressuposições importantes ou fundamentais; (4) os componentes principais são únicos (se assumirmos que S possui autovalores distintos), enquanto que os fatores estão sujeitos a uma rotação arbitrária; e (5) se o número de fatores for alterado, os fatores (estimados) também se alteram. Isto não ocorre em ACP.

A possibilidade de se fazer uma rotação visando uma melhor interpretação dos fatores é uma das vantagens da AF sobre a ACP. Se o objetivo for encontrar e descrever alguns fatores de interesse, a AF pode ser mais útil e, portanto, preferível se o modelo de fatores se ajusta bem aos dados e se a interpretação dos fatores rotacionados for de agrado. Por outro lado, se o objetivo for definir um menor número de variáveis a serem utilizadas em uma outra análise, iríamos ordinariamente preferir a ACP, embora em alguns casos este objetivo também possa ser alcançado com a AF.

9.7. EXEMPLO

Na Tabela 9.2 são apresentados resultados médios de sete características de três cultivares de milho em cinco épocas de semeadura. Os experimentos foram conduzidos em São Gabriel D'Oeste e Sidrolândia no Estado do Mato Grosso do Sul.

Tabela 9.2 – Resultados médios de sete características de três cultivares de milho em cinco épocas de semeadura em experimentos conduzidos em São Gabriel D'Oeste e Sidrolândia no Estado do Mato Grosso do Sul

Cultivar	Local	Época	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇
C ₆₀₆	S.G.D.	10/11	60	173	81	51	59	1,165	7,305
		30/11	56	192	94	50	61	1,217	7,813
		20/12	57	186	90	57	58	1,011	7,065
		09/01	53	185	94	49	50	1,017	5,975
		29/01	52	172	72	54	55	1,002	6,077
	Sidrol.	10/11	53	200	105	48	52	1,065	7,582
		30/11	55	200	105	49	56	1,145	7,315
		20/12	54	194	96	48	56	1,148	6,452
		09/01	52	170	88	48	45	0,942	4,825
		29/01	51	180	83	46	42	0,907	4,133
Hatã ₁₀₀₀	S.G.D.	10/11	63	209	109	50	66	1,313	9,135
		30/11	58	229	128	50	66	1,305	8,337
		20/12	60	213	118	56	61	1,078	7,272
		09/01	56	220	116	55	58	1,061	6,515
		29/01	53	215	123	52	57	1,121	6,035
	Sidrol.	10/11	58	218	120	47	60	1,269	9,043
		30/11	56	218	118	46	59	1,282	8,295
		20/12	57	223	121	48	57	1,193	7,000
		09/01	55	208	113	46	49	1,073	5,210
		29/01	52	217	118	45	48	1,062	5,332
Ag ₁₀₆	S.G.D.	10/11	67	213	112	50	59	1,174	8,437
		30/11	63	224	124	58	67	1,163	8,690
		20/12	66	221	128	58	62	1,074	8,025
		09/01	61	236	130	55	60	1,083	7,003
		29/01	57	215	123	52	53	1,020	5,385
	Sidrol.	10/11	62	225	138	50	57	1,146	9,267
		30/11	59	238	135	49	61	1,230	8,502
		20/12	62	230	134	48	53	1,118	6,773
		09/01	59	216	126	47	48	1,024	6,105
		29/01	58	225	129	47	44	0,940	4,590

X₁=número de dias para florescimento; X₂=altura das plantas; X₃=altura das espigas; X₄=estande final; X₅=número de espigas; X₆=prolificidade; X₇=rendimento de grãos

A partir dos dados apresentados na Tabela 9.2 foi feito um estudo por meio de análise de fatores utilizando o programa SAS apresentado a seguir.

(a) Programa

```
DM 'output; clear; log; clear;';
options FORMDLIM='- ' nodate nonumber;
data milho; /* Analise de Fatores - Exemplo 1 */
input cultivar local epoca df ap ae sf ne pr rg;
cards;
1 1 1 60 173 81 51 59 1.165 7.305
1 1 2 56 192 94 50 61 1.217 7.813
1 1 3 57 186 90 57 58 1.011 7.065
1 1 4 53 185 94 49 50 1.017 5.975
1 1 5 52 172 72 54 55 1.002 6.077
1 2 1 53 200 105 48 52 1.065 7.582
1 2 2 55 200 105 49 56 1.145 7.315
1 2 3 54 194 96 48 56 1.148 6.452
1 2 4 52 170 88 48 45 0.942 4.825
1 2 5 51 180 83 46 42 0.907 4.133
2 1 1 63 209 109 50 66 1.313 9.135
2 1 2 58 229 128 50 66 1.305 8.337
2 1 3 60 213 118 56 61 1.078 7.272
2 1 4 56 220 116 55 58 1.061 6.515
2 1 5 53 215 123 52 57 1.121 6.035
2 2 1 58 218 120 47 60 1.269 9.043
2 2 2 56 218 118 46 59 1.282 8.295
2 2 3 57 223 121 48 57 1.193 7.000
2 2 4 55 208 113 46 49 1.073 5.210
2 2 5 52 217 118 45 48 1.062 5.332
3 1 1 67 213 112 50 59 1.174 8.437
3 1 2 63 224 124 58 67 1.163 8.690
3 1 3 66 221 128 58 62 1.074 8.025
3 1 4 61 236 130 55 60 1.083 7.003
3 1 5 57 215 123 52 53 1.020 5.385
3 2 1 62 225 138 50 57 1.146 9.267
3 2 2 59 238 135 49 61 1.230 8.502
3 2 3 62 230 134 48 53 1.118 6.773
3 2 4 59 216 126 47 48 1.024 6.105
3 2 5 58 225 129 47 44 0.940 4.590
;
```

```

/*PROGRAMA DE ANÁLISE DE FATORES PELO MÉTODO DOS COMPONENTES PRINCIPAIS*/
PROC FACTOR NFACTOR=3 METHOD=P SCORE SCREE R=V OUTSTAT=SAIDA1 OUT=RESULT;
VAR DF AP AE SF NE PR RG;
PROC PRINT DATA=RESULT;
VAR FACTOR1-FACTOR3;
PROC PRINT DATA=SAIDA1;
RUN;
PROC CORR DATA=RESULT; /* reproduz as cargas rotacionadas do PROC FACTOR */
VAR DF AP AE SF NE PR RG;
WITH FACTOR1 FACTOR2 FACTOR3;
RUN;

```

(b) Resultado da análise

```

The SAS System
The FACTOR Procedure
Initial Factor Method: Principal Components
Prior Communality Estimates: ONE
Eigenvalues of the Correlation Matrix: Total = 7 Average = 1

```

	Eigenvalue	Difference	Proportion	Cumulative
1	3.81597646	2.29207627	0.5451	0.5451
2	1.52390019	0.41385628	0.2177	0.7628
3	1.11004391	0.70782585	0.1586	0.9214
4	0.40221806	0.28296128	0.0575	0.9789
5	0.11925679	0.09294996	0.0170	0.9959
6	0.02630682	0.02400906	0.0038	0.9997
7	0.00229777		0.0003	1.0000

3 factors will be retained by the NFACTOR criterion.

Comentário: Os três primeiros autovalores da matriz R foram superiores a 1 e conseguiram explicar 92,14% da variação total dos dados. Os dois primeiros autovalores explicam 76,28% da variação total.

Pode-se observar na Figura 9.1, que uma maior mudança ocorre em torno de $i=4$. Assim, os autovalores após o terceiro maior, são todos relativamente pequenos. Neste caso, parece que três (ou talvez quatro) fatores são suficientes para descreverem as relações de covariância entre as variáveis avaliadas.

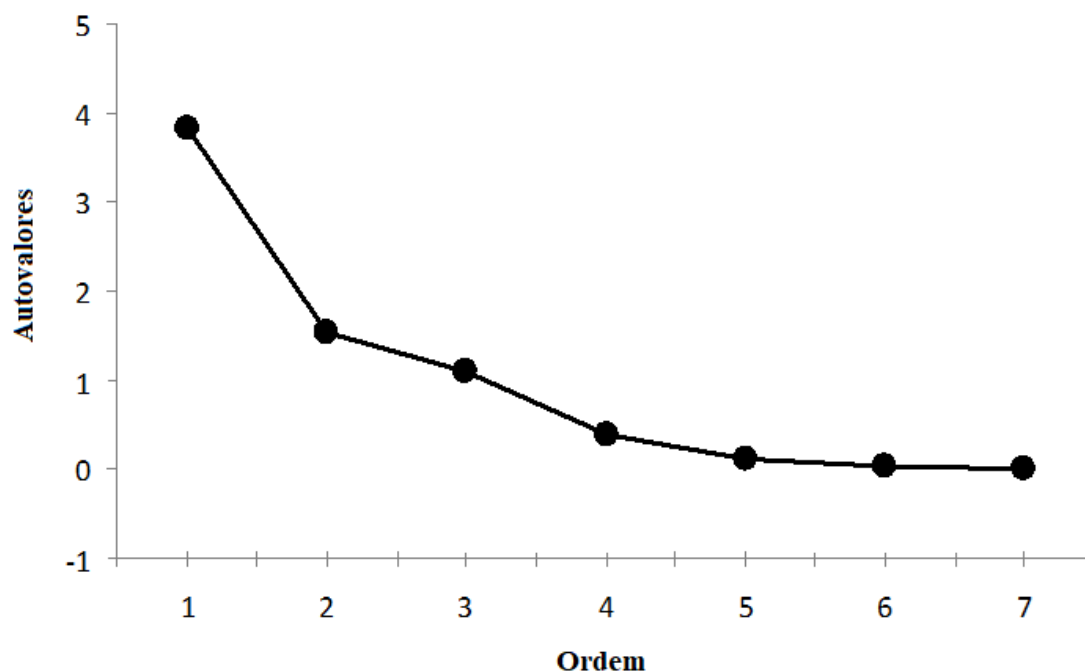


Figura 9.1 - “Scree plot” dos autovalores ordenados do maior ao menor

The SAS System			
The FACTOR Procedure			
Initial Factor Method: Principal Components			
Factor Pattern			
	Factor1	Factor2	Factor3
df	0.79338	-0.00288	0.30456
ap	0.71370	-0.65620	0.13228
ae	0.65644	-0.71727	0.17278
sf	0.37789	0.52395	0.73425
ne	0.86426	0.45076	-0.04310
pr	0.77062	0.13072	-0.60164
rg	0.87329	0.28988	-0.25884
= \hat{L}			

Comentário: \hat{L} é a matriz de cargas fatoriais (“Factor Loadings”) não rotacionadas.

Variance Explained by Each Factor		
Factor1	Factor2	Factor3
3.8159765	1.5239002	1.1100439

Final Communality Estimates: Total = 6.449921

df	ap	ae	sf	ne	pr	rg
0.72222110	0.95745355	0.97523957	0.95644381	0.95198779	0.97291227	0.91366247

The SAS System

The FACTOR Procedure

Rotation Method: **Varimax**

Orthogonal Transformation Matrix

		1	2	3
Matriz T \Rightarrow	1	0.74560	0.57188	0.34210
	2	0.37722	-0.78539	0.49078
	3	-0.54935	0.23687	0.80131

$$\hat{L}_p^* = \hat{L}_m T$$

Rotated Factor Pattern

		Factor1	Factor2	Factor3
<u>Comunalidade (%)</u>				
72,22	df	0.42315	0.52812	0.51405
95,45	ap	0.21193	0.95486	0.02811
97,52	ae	0.12395	0.97967	0.01100
95,64	sf	0.07604	-0.02147	0.97478
95,20	ne	0.83810	0.13003	0.48236
97,29	pr	0.95440	0.19552	-0.15431
91,37	rg	0.90266	0.21044	0.23361
<u>Total: 92,14</u>				

Variance Explained by Each Factor

Factor1	Factor2	Factor3
2.6732027	2.2503003	1.5264177

Final Communality Estimates: Total = 6.449921

df	ap	ae	sf	ne	pr	rg
0.72222110	0.95745355	0.97523957	0.95644381	0.95198779	0.97291227	0.91366247

Comentário: As comunalidades expressam a porcentagem da variância de cada variável que é explicada pelos três fatores comuns. As cargas fatoriais em \hat{L}^* representam os coeficientes de correlação entre cada fator (escores fatoriais) e cada uma das sete variáveis (Tabela 9.3).

Tabela 9.3 - Cargas fatoriais, comunalidade de cada variável e porcentagem da variância total correspondente a cada fator após a rotação pelo método varimax, na análise de fatores aplicada à matriz de correlação entre sete variáveis associadas à cultura do milho

Variáveis	Cargas fatoriais			Comunalidade (%)
	F ₁	F ₂	F ₃	
X ₁ (Nº dias para florescimento)	0,4232	0,5281	0,5140	72,22
X ₂ (Altura das plantas)	0,2119	0,9549	0,0281	95,75
X ₃ (Altura de espigas)	0,1240	0,9797	0,0110	97,52
X ₄ (Estande final)	0,0760	-0,0215	0,9748	95,64
X ₅ (Número de espigas)	0,8381	0,1300	0,4824	95,20
X ₆ (Prolificidade)	0,9544	0,1955	-0,1543	97,29
X ₇ (Rendimento de grãos)	0,9027	0,2104	0,2336	91,37
% da variância explicada pelos fatores rotacionados	38,19	32,15	21,80	92,14

Pelos resultados apresentados na Tabela 9.3 verifica-se que o primeiro fator (F₁), correlacionou-se mais fortemente e positivamente com X₅ (número de espigas), X₆ (prolificidade) e X₇ (rendimento de grãos). Portanto, F₁ é um fator que reflete a “**produtividade das cultivares de milho**”.

O segundo fator (F₂) correlacionou-se mais fortemente com X₂ (altura das plantas) e X₃ (altura das espigas). Trata-se, portanto de um fator associado ao “**crescimento das plantas de milho**”.

O terceiro fator (F₃) correlacionou-se mais fortemente apenas com X₄ (estande final), mostrando ser um fator que descreve a “**população de plantas das cultivares de milho**”.

A variável X₁ (número de dias para florescimento) correlacionou-se fracamente com os três fatores, apresentando a mais baixa comunalidade.

Observação: É usual tentar dar nomes aos fatores. Em muitos casos, isto requer certo grau de imaginação. Muitas vezes, esta técnica de análise multivariada é desprezada pela dificuldade em se nomear cada um dos fatores obtidos e de se fazer as interpretações corretas.

```

The SAS System

The FACTOR Procedure

Rotation Method: Varimax

Scoring Coefficients Estimated by Regression

Squared Multiple Correlations of the Variables with Each Factor

      Factor1      Factor2      Factor3
1.0000000      1.0000000      1.0000000

```

Standardized Scoring Coefficients			
	Factor1	Factor2	Factor3
df	0.00358	0.18537	0.29005
ap	-0.08845	0.47338	-0.05186
ae	-0.13480	0.50492	-0.04742
sf	-0.15984	-0.05672	0.73265
ne	0.30177	-0.11199	0.19154
pr	0.48068	-0.08027	-0.32312
rg	0.37048	-0.07375	-0.01520
	$\underbrace{a_1}$	$\underbrace{a_2}$	$\underbrace{a_3}$

Comentário: Estes coeficientes aplicados aos dados padronizados (média 0 e variância 1) geram os escores fatoriais apresentados a seguir.

ESCORES FATORIAIS			
Obs	Factor1	Factor2	Factor3
1	0.83897	-1.78170	0.41428
2	1.15061	-1.20563	-0.24318
3	-0.34424	-1.28415	1.75651
4	-0.66400	-1.05834	-0.33113
5	-0.45802	-2.20330	0.86249
6	-0.04639	-0.51625	-0.69814
7	0.38708	-0.55887	-0.48457
8	0.31278	-0.94818	-0.70795
9	-1.37660	-1.42337	-0.44929
10	-1.77719	-1.21978	-0.89829
11	1.97206	-0.28218	-0.00341
12	1.48778	0.57748	-0.41802
13	-0.14216	0.20864	1.50501
14	-0.53350	0.26741	0.99238
15	-0.34254	0.25537	-0.00887
16	1.47010	0.21950	-1.03018
17	1.34227	0.13663	-1.41781
18	0.38325	0.52655	-0.75843
19	-0.78057	0.19475	-1.08201
20	-0.88372	0.46072	-1.51833
21	0.80442	0.33774	0.46924
22	0.70920	0.50930	1.96299

23	-0.10948	0.86847	2.29207
24	-0.38989	1.20080	1.24263
25	-1.14672	0.60899	0.45687
26	0.54595	1.16524	0.04075
27	0.91143	1.19035	-0.51245
28	-0.32436	1.42417	-0.36030
29	-0.98707	0.92829	-0.56099
30	-2.00943	1.40135	-0.51187

Com os dados da Tabela 9.4, a seguir foram obtidos os escores relativos à observação número 1.

Tabela 9.4 – Dados para obtenção dos escores relativos à observação número 1

X_i	\bar{X}_i	s_i	z_i
$X_1 = 60$	57,5	4,232469	0,590672
$X_2 = 173$	208,8333	19,294601	-1,857167
$X_3 = 81$	112,3667	17,748207	-1,7673316
$X_4 = 51$	50,3	3,752241	0,186555
$X_5 = 59$	55,9667	6,462002	0,469406
$X_6 = 1,165$	1,1116	0,107628	0,496153
$X_7 = 7,305$	6,9831	1,419083	0,226837

Considerando a última coluna da Tabela 9.4 como sendo dados de um vetor \tilde{z} , virá:

Escore para o Fator 1 = $a'_1 \tilde{z} = 0,83897$.

Escore para o Fator 2 = $a'_2 \tilde{z} = -1,78170$.

Escore para o Fator 3 = $a'_3 \tilde{z} = 0,41428$.

Na Tabela 9.5 são apresentados os escores fatoriais para os três fatores que descreveram a cultura do milho nos diversos tratamentos estudados.

Tabela 9.5 – Escores fatoriais para os três fatores que descreveram a cultura do milho nos diversos tratamentos estudados

Cultivar	Local	Época	Obs. n°	Fator 1	Fator 2	Fator 3
C ₆₀₆	S.G.D.	10/11	1	0,83897	-1,78170	0,41428
		30/11	2	1,15061	-1,20563	-0,24318
		20/12	3	-0,34424	-1,28415	1,75651
		09/01	4	-0,66400	-1,05834	-0,33113
		29/01	5	-0,45802	-2,20330	0,86249
	Sidrol.	10/11	6	-0,04639	-0,51625	-0,69814
		30/11	7	0,38708	-0,55887	-0,48457
		20/12	8	0,31278	-0,94818	-0,70795
		09/01	9	-1,37660	-1,42337	-0,44929
		29/01	10	-1,77719	-1,21978	-0,89829
Hatã ₁₀₀₀	S.G.D.	10/11	11	1,97206	-0,28218	-0,00341
		30/11	12	1,48778	0,57748	-0,41802
		20/12	13	-0,14217	0,20864	1,50501
		09/01	14	-0,53350	0,26741	0,99238
		29/01	15	-0,34254	0,25537	-0,00887
	Sidrol.	10/11	16	1,47010	0,21950	-1,03018
		30/11	17	1,34227	0,13663	-1,41781
		20/12	18	0,38325	0,52655	-0,75843
		09/01	19	-0,78057	0,19475	-1,08201
		29/01	20	-0,88372	0,46072	-1,51833
Ag ₁₀₆	S.G.D.	10/11	21	0,80442	0,33774	0,46924
		30/11	22	0,70919	0,50930	1,96299
		20/12	23	-0,10949	0,86847	2,29207
		09/01	24	-0,38989	1,20080	1,24263
		29/01	25	-1,14672	0,60899	0,45687
	Sidrol.	10/11	26	0,54594	1,16524	0,04075
		30/11	27	0,91142	1,19035	-0,51245
		20/12	28	-0,32436	1,42417	-0,36030
		09/01	29	-0,98707	0,92829	-0,56099
		29/01	30	-2,00943	1,40135	-0,51187

S.G.D. (São Gabriel D'Oeste); Sidrol. (Sidrolândia)

Na Tabela 9.5 pode-se observar que o fator F₁ “produtividade das cultivares de milho” apresentou valores positivos nas primeiras épocas de semeadura e valores negativos nas últimas,

indicando que o retardamento do plantio prejudicou a produtividade. Os maiores valores de F_1 se deram com a Cultivar Hatã₁₀₀₀ nas duas primeiras épocas de semeadura nos dois locais. Os menores valores de F_1 se deram em Sidrolândia nas duas épocas.

Quanto ao fator F_2 “crescimento das plantas de milho”, estabeleceu-se um contraste entre a Cultivar C₆₀₆, com valores negativos, e as cultivares Hatã₁₀₀₀ e Ag₁₀₆ com valores positivos. Portanto, a Cultivar C₆₀₆ apresentou plantas com menor crescimento que as Cultivares Hatã₁₀₀₀ e Ag₁₀₆. A Cultivar Ag₁₀₆ apresentou plantas com maior crescimento.

O fator F_3 “população de plantas das cultivares de milho” estabeleceu um contraste entre os dois locais. São Gabriel D’Oeste apresentou valores positivos enquanto que Sidrolândia teve valores negativos.

The SAS System									
Obs	_TYPE_	_NAME_	df	ap	ae	sf	ne	pr	rg
1	MEAN		57.5000	208.833	112.367	50.3000	55.9667	1.1116	6.9831
2	STD		4.2325	19.295	17.748	3.7522	6.4620	0.1076	1.4191
3	N		30.0000	30.000	30.000	30.0000	30.0000	30.0000	30.0000
4	CORR	df	1.0000	0.523	0.520	0.4136	0.5806	0.3987	0.6347
5	CORR	ap	0.5230	1.000	0.964	0.0564	0.3512	0.3988	0.3774
6	CORR	ae	0.5199	0.964	1.000	0.0211	0.2491	0.3111	0.3261
7	CORR	sf	0.4136	0.056	0.021	1.0000	0.5693	-0.0657	0.2754
8	CORR	ne	0.5806	0.351	0.249	0.5693	1.0000	0.7790	0.8514
9	CORR	pr	0.3987	0.399	0.311	-0.0657	0.7790	1.0000	0.8204
10	CORR	rg	0.6347	0.377	0.326	0.2754	0.8514	0.8204	1.0000
11	COMMUNAL		0.7222	0.957	0.975	0.9564	0.9520	0.9729	0.9137
12	PRIORS		1.0000	1.000	1.000	1.0000	1.0000	1.0000	1.0000
13	EIGENVAL		3.8160	1.524	1.110	0.4022	0.1193	0.0263	0.0023
14	UNROTATE	Factor1	0.7934	0.714	0.656	0.3779	0.8643	0.7706	0.8733
15	UNROTATE	Factor2	-0.0029	-0.656	-0.717	0.5240	0.4508	0.1307	0.2899
16	UNROTATE	Factor3	0.3046	0.132	0.173	0.7342	-0.0431	-0.6016	-0.2588
17	TRANSFOR	Factor1	0.7456	0.377	-0.549
18	TRANSFOR	Factor2	0.5719	-0.785	0.237
19	TRANSFOR	Factor3	0.3421	0.491	0.801
20	PATTERN	Factor1	0.4232	0.212	0.124	0.0760	0.8381	0.9544	0.9027
21	PATTERN	Factor2	0.5281	0.955	0.980	-0.0215	0.1300	0.1955	0.2104
22	PATTERN	Factor3	0.5141	0.028	0.011	0.9748	0.4824	-0.1543	0.2336
23	SCORE	Factor1	0.0036	-0.088	-0.135	-0.1598	0.3018	0.4807	0.3705
24	SCORE	Factor2	0.1854	0.473	0.505	-0.0567	-0.1120	-0.0803	-0.0738
25	SCORE	Factor3	0.2901	-0.052	-0.047	0.7327	0.1915	-0.3231	-0.0152

The SAS System								
The CORR Procedure								
3 With Variables:		Factor1	Factor2	Factor3				
7	Variables:	df	ap	ae	sf	ne	pr	rg
Simple Statistics								
Variable	N	Mean	Std Dev		Sum	Minimum		Maximum
Factor1	30	0	1.00000		0	-2.00943		1.97206
Factor2	30	0	1.00000		0	-2.20330		1.42417
Factor3	30	0	1.00000		0	-1.51833		2.29207
df	30	57.50000	4.23247		1725	51.00000		67.00000
ap	30	208.83333	19.29460		6265	170.00000		238.00000
ae	30	112.36667	17.74821		3371	72.00000		138.00000
sf	30	50.30000	3.75224		1509	45.00000		58.00000
ne	30	55.96667	6.46200		1679	42.00000		67.00000
pr	30	1.11160	0.10763		33.34800	0.90700		1.31300
rg	30	6.98310	1.41908		209.49300	4.13300		9.26700
Pearson Correlation Coefficients, N = 30								
Prob > r under H0: Rho=0								
	df	ap	ae	sf	ne	pr		rg
Factor1	0.42315	0.21193	0.12395	0.07604	0.83810	0.95440		0.90266
	0.0198	0.2609	0.5140	0.6896	<.0001	<.0001		<.0001
Factor2	0.52812	0.95486	0.97967	-0.02147	0.13003	0.19552		0.21044
	0.0027	<.0001	<.0001	0.9103	0.4934	0.3005		0.2643
Factor3	0.51405	0.02811	0.01100	0.97478	0.48236	-0.15431		0.23361
	0.0037	0.8828	0.9540	<.0001	0.0069	0.4155		0.2141

Comentário: Note que estas correlações são exatamente as cargas fatoriais rotacionadas dadas por $\hat{L}^* = \hat{L}T$.

9.8. EXERCÍCIO

Num estudo sobre preferência do consumidor, foi tomada uma amostra aleatória de 30 consumidores para avaliar vários atributos de um novo produto. Os dados foram tabulados, obtidos a partir de uma escala apropriada, considerando os seguintes atributos:

X_1 = Gosto

X_2 = Bom preço

X_3 = Sabor

X_4 = Adequado para lanche

X_5 = Fornece muita energia

A matriz das estimativas dos coeficientes de correlação entre as cinco variáveis está apresentada na Tabela 9.6.

Tabela 9.6 – Matriz de correlação entre as variáveis

Variáveis	X_1	X_2	X_3	X_4	X_5
X_1	1,00				
X_2	0,02	1,00			
X_3	0,96	0,13	1,00		
X_4	0,42	0,71	0,50	1,00	
X_5	0,01	0,85	0,11	0,79	1,00

Fonte: Adaptado de Johnson e Wichern (1998)

Pede-se:

Faça uma análise de fatores utilizando o Método dos Componentes Principais. Como rotação ortogonal use o método Varimax. A partir das informações dadas e dos resultados da análise, faça as devidas interpretações.

9.9. REFERÊNCIAS

HARMAN, H. H. **Modern factor analysis**. 3. ed. Chicago: The University of Chicago Press, 1976. 474p.

JOHNSON, R. A.; WICHERN, D.W. **Applied multivariate statistical analysis**. 4th ed., USA, Upper Saddle River, New Jersey: Prentice Hall, Inc., 1998. 816p.

MARDIA, K. V.; KENT, J. T.; BIBBY, J. M. **Multivariate analysis**. 6th ed., Londres: Academic Press, 1997. 518p.

RENCER, A. C. **Methods of multivariate analysis**. 2nd ed., USA: John Wiley and Sons, 2002. 738p.