

FRANCISCO PAULO VILHENA ANTUNES BERNARDINO CARVALHO



# INFERÊNCIA PARA MODELOS COM COMMUTATIVE ORTHOGONAL BLOCK STRUCTURE

*Esta tese não inclui as críticas e sugestões feitas pelo Júri*

Orientador: Professora Auxiliar Maria Manuela Melo Oliveira  
Universidade de Évora

Co-Orientador: Professor Catedrático João Tiago Praça Nunes Mexia  
Universidade Nova de Lisboa – Faculdade de Ciências e Tecnologia

2009

# INFERÊNCIA PARA MODELOS COM COMMUTATIVE ORTHOGONAL BLOCK STRUCTURE

Orientador: Professora Auxiliar Maria Manuela Melo Oliveira  
Universidade de Évora



Co-Orientador: Professor Catedrático João Tiago Praça Nunes Mexia  
Universidade Nova de Lisboa – Faculdade de Ciências e Tecnologia

170 171

*Dissertação apresentada como requerimento parcial para a  
obtenção do grau de Doutor em Matemática, na especialidade  
de Estatística, na Universidade de Évora.*

U.E Serviços Académicos	N.º <u>63/22502</u>
<u>7/8/09</u>	Sector:
<u>Fimena</u>	<u>DEPG</u>

# Erratum

Página <i>Page</i>	linha <i>line</i>	onde se lê... <i>were you read...</i>	deve ler-se... <i>you should read...</i>
7	5	Álgebrsa	Álgebras
22	4	vetores	vectores
49	6	consideraá	considerará
49	12	$\sigma_{m+1}^2 I_{c_w+1}, \dots, \sigma_w I_{c_w}$	$\sigma_{m+1}^2 I_{c_w+1}, \dots, \sigma_w I_{c_w}$
50	15	$M^\circ = \{M_{m+1}, \dots, M_w\}$	$M^\circ = \{M_{m+1}, \dots, M_w, I_n\}$
51	4	$M_i = \sum_{j=1}^k c_{i,j} Q_j$	$M_i = \sum_{j=1}^k c_{i-m,j} Q_j$
51	7	$\gamma_j = \sum_{i=1}^{w-m} c_{i,j} \sigma_j^2$	$\gamma_j = \sum_{i=m+1}^w c_{i-m,j} \sigma_j^2$
54	20	$\tilde{\sigma}^2 = (C_2^\top)^+ \tilde{\gamma}(2)$	$\tilde{\sigma}^2 = \left(C_2^\top\right)^+ \tilde{\gamma}(2)$
54	22	$\tilde{\gamma}(1) = C_1^\top (C_2^\top)^+ \tilde{\gamma}(2)$	$\tilde{\gamma}(1) = C_1^\top \left(C_2^\top\right)^+ \tilde{\gamma}(2)$
57	6	$\tilde{\Psi}(1)$	$\tilde{\gamma}(1)$
59	13	$\tilde{\sigma}_i^2 = \frac{S_i}{p_i}$	$\hat{\sigma}_i^2 = \frac{S_i}{p_i}$
59	17	$\tilde{\sigma}_i^2 = \frac{S_i}{p_i} \hat{\sigma}_i^2$	$\hat{\sigma}_i^2 = \frac{g_i}{p_i} \tilde{\sigma}_i^2$
74	16	são independentes	são linearmente independentes
101	12/13	$K_2, K_3$ e $K_5$	$K_2, K_3$ e $K_5$

# Inferência para Modelos com Commutative Orthogonal Block Structure

## Resumo

Os modelos com Orthogonal Block Structure (OBS) foram introduzidos por Nelder (1965a)[35], (1965b)[36]. Estes modelos desempenham um papel importante no delineamento e análise de experiências, ver por exemplo Caliński & Kageyama (2000), (2003). Um modelo pertence a esta classe se tiver matriz de covariância  $V$ , dada pela combinação linear de matrizes de projecção ortogonal mutuamente ortogonais conhecidas  $\mathbf{P}_1, \dots, \mathbf{P}_w$ . O objectivo deste trabalho é o estudo da inferência para um caso particular importante dos modelos com OBS. Se a matriz  $\mathbf{T}$  de projecção ortogonal sobre o espaço varrido pelo vector médio comutar com as  $\mathbf{P}_1, \dots, \mathbf{P}_w$ , o modelo terá Commutative Orthogonal Block Structure (COBS). Para perceber a importância destes modelos bastará pensar que, comutando  $\mathbf{T}$  e  $\mathbf{V}$ , os estimadores de mínimos quadrados dos respectivos vectores estimáveis são Best Linear Unbiased Estimator, ver Zmysłony (1978). Após apresentarmos resultados preliminares que nos serão úteis, estudaremos a teoria para os modelos com COBS, particularizando em seguida para: modelos estritamente associados e modelos com cross-nesting equilibrado. Para cada um destes casos particulares, apresentaremos um exemplo de forma a demonstrar a utilização e aplicabilidade.



# Inference for Models with Commutative Orthogonal Block Structure

## Abstract

Models with Orthogonal Block Structure (OBS) were introduced by Nelder (1965a)[35], (1965b)[36]. These models play an important role in the design and analysis of experiments, e. g. Caliński & Kageyama (2000), (2003). A model belongs to this class if it has covariance matrix  $\mathbf{V}$ , given by the linear combination of known mutually orthogonal orthogonal projection matrices  $\mathbf{P}_1, \dots, \mathbf{P}_w$ . The objective of this work is the study of inference for an important particular case of models with OBS. If  $\mathbf{T}$ , the matrix of orthogonal projection on the range space spanned by the mean vector, commutes with  $\mathbf{P}_1, \dots, \mathbf{P}_w$ , the model will have Commutative Orthogonal Block Structures (COBS). To understand the importance of these models, we just have to think that, commuting  $\mathbf{T}$  e  $\mathbf{V}$ , the least square estimators of the estimable vectors are Best Linear Unbiased Estimators, e. g. Zmysłony (1978). After presenting preliminary results that will be useful to our work, we will study the theory for models with COBS, in particular to: strictly associated models and models with balanced cross-nesting. For each of these cases we will present an example to show their use and applicability.



# Agradecimentos

Gostaria de aqui publicamente agradecer a todos quantos de uma forma directa ou indirecta contribuíram para que este trabalho fosse uma realidade.

Em primeiro lugar gostaria de agradecer à Professora Doutora Maria Manuela Oliveira e ao Professor Doutor João Tiago Mexia, que mais que orientadores, foram amigos, tornando esta tarefa um fardo bem mais leve. Sem eles o trabalho não existiria. Foram incansáveis na disponibilidade do seu tempo e paciência para estarem ao meu lado no esclarecimento de dúvidas, incentivos e correcções.

Ao Doutor Ricardo Covas, que para além de Colega me incentivou, ajudou e teve paciência o suficiente para rever o trabalho e sugerir alterações.

Ao Instituto Politécnico de Tomar, na pessoa do seu Presidente, Professor Dr António Pires da Silva e à Escola Superior de Gestão de Tomar, na pessoa da sua Directora, Professora Coordenadora Dr<sup>a</sup> Maria da Conceição Fortunato, que permitiram que me ausentasse para apresentação de comunicações e demais deslocações durante o período que compreendeu a elaboração deste trabalho.

À minha Família, Xana, Diogo e Francisca. Durante o tempo que dediquei a este trabalho, foram estes os principais prejudicados com as minhas ausências, faltas de atenção e indisponibilidade. A eles, o meu mais profundo agradecimento.

Não querendo cometer nenhuma injustiça, quero prestar o meu agradecimento a todos quantos de uma forma ou de outra ajudaram neste trabalho, quer por simples observações, incentivos, enfim aqueles pequenos momentos que motivam nas alturas em que tudo parece estar a dar mal. A todos, muito obrigado.



# Glossário

OBS	Orthogonal Block Structure
COBS	Commutative Orthogonal Block Structure
SCOBS	Segregated Commutative Orthogonal Block Structure
SOAM	Segregated Orthogonal Associated Model
MPO	matriz de projecção ortogonal
AJC	álgebra de Jordan comutativa
$\mathcal{A}$	álgebra
$bp(\mathcal{A})$	base principal da álgebra de Jordan comutativa $\mathcal{A}$
$dim(\mathcal{A})$	dimensão da álgebra $\mathcal{A}$
$\alpha$	vector $\alpha$
$\mathbf{A}$	matriz $\mathbf{A}$
$R(\cdot)$	espaço-imagem
$\boxplus$	soma directa ortogonal de sub-espacos
$car(\mathbf{M})$	característica da matriz $\mathbf{M}$
$\mathbf{M}^+$	matriz inversa de Moore-Penrose de $\mathbf{M}$
$det(\mathbf{M})$	determinante da matriz $\mathbf{M}$
$\tau$	relação de equivalência
$\mathbf{A}^\top$	matriz transposta de $\mathbf{A}$
$\alpha^\top$	vector transposto de $\alpha$

$\mathbf{M}^{-1}$	matriz inversa de $\mathbf{M}$
$\mathbf{I}_n$	matriz identidade de ordem $n$
$\mathbf{0}_{n \times m}$	matriz nula com $n$ linhas e $m$ colunas
$\mathbf{0}_n$	vector nulo com $n$ linhas
$\tilde{\boldsymbol{\beta}}$	estimador do vector $\boldsymbol{\beta}$
$\tilde{\gamma}$	estimador de $\gamma$
$\perp$	ortogonal
$\#(\mathcal{M})$	cardinal de $\mathcal{M}$
$\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$	distribuição normal com vector médio $\boldsymbol{\mu}$ e matriz de variância $\mathbf{V}$
$\chi_{g_i}^2$	distribuição $\chi^2$ com $g_i$ graus de liberdade
$\mathcal{A}(\mathbf{P})$	família de matrizes diagonalizadas por $\mathbf{P}$
$\mathcal{A}(\mathcal{M})$	álgebra de Jordan comutativa gerada por $\mathcal{M}$

*À minha Família*



# Conteúdo

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>I</b>	<b>Estruturas de Suporte</b>	<b>5</b>
<b>2</b>	<b>Resultados Preliminares</b>	<b>7</b>
2.1	Resultados sobre Matrizes . . . . .	7
2.1.1	Álgebras de Jordan Comutativas . . . . .	7
2.1.2	Álgebras geradas . . . . .	11
2.1.3	Sub-famílias . . . . .	15
2.1.4	Produto de Kronecker . . . . .	17
2.1.5	Operações Binárias . . . . .	19
2.2	Mínimos Quadrados . . . . .	25
2.3	Estatísticas Suficientes e Completas . . . . .	28
2.4	Variáveis Pivot Indutoras . . . . .	36
2.4.1	Medidas Induzidas . . . . .	36
2.4.2	Densidades . . . . .	40
2.4.3	Indução para vectores estimáveis . . . . .	44
<b>II</b>	<b>Modelos</b>	<b>47</b>
<b>3</b>	<b>Modelos Mistos</b>	<b>49</b>

3.1	Estrutura Algébrica . . . . .	49
3.2	Projecções Ortogonais e Estimadores . . . . .	53
3.3	Normalidade . . . . .	55
3.3.1	Estimadores de Máxima Verosimilhança . . . . .	57
3.4	Componentes de variância e vectores estimáveis . . . . .	60
3.4.1	Componentes variância . . . . .	60
3.4.2	Vectores Estimáveis . . . . .	61
<b>4</b>	<b>Modelos Estritamente Associados</b>	<b>63</b>
4.1	Estrutura Algébrica . . . . .	63
4.2	Inferência . . . . .	66
4.2.1	Modelos de efeitos fixos . . . . .	67
4.2.2	Modelos de efeitos mistos . . . . .	69
4.2.3	Modelos de efeitos aleatórios . . . . .	72
4.3	Factoriais de Base Prima . . . . .	72
4.3.1	Caso de efeitos fixos . . . . .	76
4.3.2	Caso de efeitos aleatórios . . . . .	78
4.3.3	Caso de efeitos mistos . . . . .	79
4.3.4	Repetições . . . . .	79
4.3.5	Aplicação: Estudo do efeito da lidocaína no músculo cardíaco de uma raça de cães . . . . .	81
<b>5</b>	<b>Modelos com Cross-Nesting Equilibrado</b>	<b>87</b>
5.1	Estrutura Algébrica . . . . .	89
5.1.1	Testes . . . . .	95
5.2	Operações Binárias . . . . .	96
5.3	Aplicação: cimentos odontológicos . . . . .	98
5.3.1	Estrutura Algébrica . . . . .	99
5.4	Simulações . . . . .	108

<i>CONTEÚDO</i>	xiii
<b>III Considerações Finais</b>	<b>113</b>
<b>6 Considerações finais e trabalho futuro</b>	<b>115</b>



# Lista de Tabelas

4.1	Níveis da enzima . . . . .	83
4.2	Teste $F$ e $p$ – <i>values</i> da experiência . . . . .	84
4.3	Teste $F$ e $p$ – <i>values</i> . . . . .	86
5.1	Exemplo 1 . . . . .	88
5.2	Exemplo 2 . . . . .	88
5.3	Médias dos discos $d_1, d_2, d_3, d_4$ e $d_5$ . . . . .	100
5.4	Índices utilizados no modelo . . . . .	101
5.5	$F$ tests . . . . .	109



# Lista de Figuras

2.1	Curvas de nível . . . . .	44
4.1	Combinações de tratamento num delineamento $3^3$ . . . . .	82
5.1	Função densidade . . . . .	110
5.2	Curvas de nível . . . . .	112



# Capítulo 1

## Introdução

Os modelos com estrutura ortogonal por blocos, Orthogonal Block Structure (OBS) desempenham um papel importante no delineamento e análise de experiências, ver por exemplo Caliński & Kageyama (2000) [2] e (2002) [3].

Nestes modelos a matriz de covariância do vector  $\mathbf{y}$  das observações pode ser escrita como uma combinação linear

$$\mathbf{V} = \sum_{j=1}^w \gamma_j \mathbf{P}_j \quad (1.1)$$

de matrizes de projecção ortogonal, mutuamente ortogonais, conhecidas. A inferência realizada nestes modelos assenta nas projecções ortogonais de  $\mathbf{y}$  sobre os espaços imagem  $R(\mathbf{P}_j)$ ,  $j = 1, \dots, w$ , das matrizes de projecção ortogonal  $\mathbf{P}_j$ ,  $j = 1, \dots, w$ . Estes espaços imagem serão os estratos.

Nesta dissertação vamos estudar uma classe particular de modelos com OBS, os modelos com COBS (Commutative Orthogonal Block Structure).

Sendo  $\mathbf{T}$  a matriz de projecção ortogonal sobre o espaço  $\Omega$  varrido pelo vector médio  $\boldsymbol{\mu}$  de  $\mathbf{y}$ , um modelo com OBS tem COBS se a matriz  $\mathbf{T}$  comuta com as matrizes  $\mathbf{P}_1, \dots, \mathbf{P}_w$ . No estudo destes modelos utilizaremos as Álgebras de Jordan Comutativas (AJC), que desempenharão um papel central.

As AJC consideradas serão espaços lineares constituídos por matrizes simétricas que comutam. Estes espaços contêm os quadrados das respectivas matrizes. Estas

estruturas foram introduzidas por Jordan *et al* (1934) [22] numa reestruturação da teoria da Mecânica Quântica. Posteriormente estas álgebras foram redescobertas por Seely (1970a) [40], (1970b) [41], (1971a) [42], (1971b) [44] e (1979) [43], que as utilizou no âmbito da Inferência Estatística Linear. Nesta construção há vários outros trabalhos importantes, ver por exemplo Seely & Zyskind (1971) [44], Drygas & Zmysłony (1992) [10], Vanleuwen *et al* (1998)[47] e (1999)[46] e Fonseca *et al* (2006) [15], (2007) [16], (2008) [17].

O presente trabalho encontra-se organizado em duas partes. Na primeira parte, serão apresentados os resultados algébricos e estatísticos necessários para o desenvolvimento do trabalho, capítulos 2 e 3 e na segunda parte, a aplicação dos modelos desenvolvidos a duas classes particulares de modelos, capítulos 4 e 5.

No segundo capítulo apresentamos alguns dos resultados de que nos serviremos no decorrer deste trabalho. Estes resultados agrupam-se em duas áreas: álgebras de matrizes, com destaque para as AJC, e resultados sobre estatísticas e estimação. Nesta segunda área há que salientar a indução de medidas de probabilidade nos espaços paramétricos. Como veremos, isso permitirá a construção de regiões de confiança para os parâmetros dos modelos. Os resultados são bastante recentes, veja-se Covas (2007) [7] e Fonseca (2007) [12].

A este capítulo preliminar segue-se o estudo dos Modelos Lineares Mistos, capítulo 3. Começaremos por apresentar uma teoria geral destes modelos com especial incidência para aqueles que têm COBS.

Na segunda parte segue-se o estudo de duas classes particulares de modelos, modelos estritamente associados e modelos com *cross-nesting*.

No estudo destes modelos procuraremos obter Best Linear Unbiased Estimator (BLUE) para vectores estimáveis utilizando, para isso, uma forma muito geral do teorema de Gauss-Markov, ver Zmysłony (1978) [51].

Admitindo a normalidade das observações, obteremos ainda estatísticas suficientes e completas. Estas estatísticas permitir-nos-ão:

- Obter Uniformly Minimum Variance Unbiased Estimator (UMVUE) para os parâmetros relevantes do modelo;
- Induzir medidas de probabilidade nos espaços paramétricos. Conseguem-se assim obter regiões de confiança e, por dualidade, testam-se hipóteses para os parâmetros de interesse.

O estudo dos modelos estritamente associados considerados nesta dissertação complementa o tratamento dos factoriais de base prima, ver Jesus *et al* (2009) [20] e dos modelos com *cross-nesting* equilibrado, ver Fonseca *et al* (2003) [14].

Para estas duas classes de modelos desenvolver-se-á a sua estrutura algébrica e calcular-se-ão estimadores, regiões de confiança e testes de hipóteses por dualidade.

Serão ainda apresentadas duas aplicações que demonstrarão a aplicabilidade e vantagens dos modelos desenvolvidos, secção 4.3.5 e secção 5.3.

Finalmente, apresentam-se algumas considerações finais e um conjunto de linhas de trabalho a desenvolver no futuro.



# Parte I

## Estruturas de Suporte



# Capítulo 2

## Resultados Preliminares

### 2.1 Resultados sobre Matrizes

#### 2.1.1 Álgebras de Jordan Comutativas

As Álgebras de Jordan Comutativas (AJC) são estruturas que foram introduzidas por Jordan<sup>1</sup> *et al* (1934)[22] para obter uma nova formulação para a Mecânica Quântica. Mais tarde, Seely (1970)[40], (1971) [42], e Seely & Zyskind (1971) [44], redescobriram estas álgebras aplicando-as para realizar inferência estatística. Apesar de se associar a Jordan a autoria destas estruturas, pensa-se que estas já seriam intuitivas de Arquimedes<sup>2</sup> e, que Gauss<sup>3</sup> nos seus trabalhos não publicados, teria já uma teoria completa desenvolvida. Ver a este respeito, por exemplo, McCrimmon (1978)[27].

A utilização destas estruturas levou, na inferência estatística, a resultados impor-

---

<sup>1</sup>Pascual Jordan (1902–1980). Matemático alemão com contributos significativos na mecânica quântica e teoria de grupos quânticos. É muitas vezes confundido com Camille Jordan, matemático francês a que está associado o *Teorema da Curva de Jordan*, e com o geodesista alemão Wilhelm Jordan, a que está associado o processo de eliminação *Gauss-Jordan*.

<sup>2</sup>Arquimedes (c. 287 a.C.– c. 212 a.C.), matemático, físico e inventor.

<sup>3</sup>Carl Friedrich Gauss (1777–1855). *Príncipe dos matemáticos*, por muitos considerado o maior génio da história da matemática. Matemático, astrónomo e físico alemão.

tantes, ver por exemplo VanLeeuwen (1998) [47], VanLeeuwen (1999) [46] e Fonseca *et al* (2006) [15]. Seely (1970) [41] chamou a estas estruturas espaços quadráticos comutativos uma vez que elas são espaços lineares que contêm o quadrado das suas matrizes. No entanto, por uma questão de prioridade, vamos chamá-las de álgebras de Jordan. Aquelas álgebras que iremos usar, serão comutativas, uma vez que as suas matrizes comutam.

Vejamos agora alguns resultados sobre estas estruturas.

**Definição 2.1 (Álgebras de Jordan Comutativas).** *Álgebras de Jordan Comutativas são espaços vectoriais constituídos por matrizes simétricas que comutam e contêm os quadrados das suas matrizes.*

**Lemma 2.1 (Base Principal).** *Toda a álgebra de Jordan comutativa tem uma e uma só base constituída por matrizes de projecção ortogonal mutuamente ortogonais: a base principal.*

Dada a álgebra de Jordan comutativa  $\mathcal{A}$ , a respectiva base principal será representada por  $bp(\mathcal{A})$ , tendo-se  $Q = \{Q_1, \dots, Q_v\} = bp(\mathcal{A})$  com  $v = \dim(\mathcal{A})$ , ver Seely (1971) [42].

Para além da base principal, é de particular interesse para o nosso estudo, as consequências da sua existência. De uma forma resumida, podemos enunciar então algumas dessas consequências. Para uma análise mais completa, ver Fonseca *et al* (2006) [15] e Jesus *et al* (2009)[20].

1. O produto de matrizes numa álgebra de Jordan comutativa pertence à álgebra de Jordan comutativa.
2. Polinómios definidos em matrizes numa álgebra de Jordan comutativa pertencem à álgebra de Jordan comutativa.
3. Matrizes de projecção ortogonal (MPO) pertencentes a uma álgebra de Jordan comutativa são somas de uma ou mais matrizes da base principal.

4. Dada a matriz  $\mathbf{M} = \sum_{j=1}^v c_j \mathbf{Q}_j$ , seja  $\mathcal{C} = \{j : c_j \neq 0\}$ , tendo-se  $R(\mathbf{M}) = \bigoplus_{j \in \mathcal{C}} R(\mathbf{Q}_j)$ , onde  $R(\cdot)$  indica o espaço-imagem e  $\bigoplus$  a soma directa ortogonal de sub-espacos, logo

$$\text{car}(\mathbf{M}) = \sum_{j \in \mathcal{C}} g_j$$

com  $g_j = \text{car}(\mathbf{Q}_j)$ ,  $j = 1, \dots, v$ .

5. Se uma MPO com característica 1 pertence a uma álgebra de Jordan comutativa  $\mathcal{A}$ , então também pertence à  $bp(\mathcal{A})$ . Esta observação aplica-se em particular à matriz  $\frac{1}{n} \mathbf{J}$  sendo regulares as álgebras de Jordan comutativas que contêm essa matriz, onde  $\mathbf{J} = \mathbf{1}_n \mathbf{1}_n^\top$ .

6. A unidade da álgebra de Jordan comutativa  $\mathcal{A}$  (elemento neutro) com base principal  $bp(\mathcal{A}) = \{\mathbf{Q}_1, \dots, \mathbf{Q}_v\}$  é dada por

$$\mathbf{U} = \sum_{j=1}^v \mathbf{Q}_j.$$

7. Dada a matriz  $\mathbf{M} = \sum_{j=1}^v a_j \mathbf{Q}_j \in \mathcal{A}$  e uma vez que as matrizes  $\mathbf{Q}_1, \dots, \mathbf{Q}_v$  são idempotentes e mutuamente ortogonais, a matriz inversa de Moore-Penrose<sup>4</sup> de  $\mathbf{M}$  é da forma

$$\mathbf{M}^+ = \sum_{j=1}^v a_j^+ \mathbf{Q}_j \in \mathcal{A}$$

com  $a_j^+ = a_j^{-1}$  se  $a_j \neq 0$ , e  $a_j^+ = 0$  se  $a_j = 0$  para  $j = 1, \dots, v$ .

A inversa de Moore-Penrose de uma matriz  $\mathbf{M}$  é a matriz,  $\mathbf{M}^+$ , que observa

---

<sup>4</sup>Também designada por pseudo-inversa, sendo uma generalização da inversa de uma matriz. Deve o seu nome a Eliakim Hastings Moore (1862–1932) e Roger Penrose (1931– ) que independentemente a descreveram em 1920 e 1955 respectivamente. Moore colocou a questão sobre a sua existência e Penrose que então é única. No entanto, o conceito de pseudo-inversa de operadores integrais, terá sido introduzido por Erik Ivar Fredholm (1866–1927) em 1903.

as seguintes condições

$$\left\{ \begin{array}{l} \mathbf{M}\mathbf{M}^+\mathbf{M} = \mathbf{M} \\ \mathbf{M}^+\mathbf{M}\mathbf{M}^+ = \mathbf{M}^+ \\ (\mathbf{M}\mathbf{M}^+)^\top = \mathbf{M}\mathbf{M}^\top \\ (\mathbf{M}^+\mathbf{M})^\top = \mathbf{M}^+\mathbf{M}. \end{array} \right.$$

Prova-se que essa matriz existe sempre e que é única, ver por exemplo os trabalhos de Pollock (1979) [37].

8. A álgebra de Jordan comutativa  $\mathcal{A}$  contém matrizes regulares se e só se contiver a matriz  $\mathbf{I}_n$ . Com efeito, se a matriz  $\mathbf{M} \in \mathcal{A}$  for regular, ter-se-á  $\mathbf{M}^{-1} = \mathbf{M}^+ \in \mathcal{A}$  dado  $\mathbf{M}$  regular, e  $\mathbf{I}_n = \mathbf{M}\mathbf{M}^{-1} \in \mathcal{A}$ .
9. A álgebra de Jordan comutativa  $\mathcal{A}$  contém matrizes regulares se e só se,

$$\mathbf{I}_n = \mathbf{U}$$

designado-se então  $\mathcal{A}$  como completa. Se  $\mathcal{A}$  não for completa pode juntar-se  $\mathbf{Q}^\perp = \mathbf{I}_n - \mathbf{U}$  à  $bp(\mathcal{A})$  obtendo-se a base principal de uma nova álgebra de Jordan comutativa, a completada de  $\mathcal{A}$ .

10. Se a matriz  $\mathbf{M} = \sum_{j=1}^v a_j \mathbf{Q}_j$  pertencer a  $\mathcal{A}$ , então  $a_1, \dots, a_v$  serão os valores próprios de  $\mathbf{M}$  com multiplicidades  $g_1, \dots, g_v$  tendo-se, caso  $\mathcal{A}$  seja completa, ver Malley (1994) [26],

$$\det \left( \sum_{j=1}^v a_j \mathbf{Q}_j \right) = \prod_{j=1}^v a_j^{g_j}.$$

11. Dada uma matriz ortogonal  $\mathbf{P}$ , a família  $\mathcal{A}(\mathbf{P})$  das matrizes diagonalizadas por  $\mathbf{P}$  é uma álgebra de Jordan comutativa.
12. Intersectando-se álgebras de Jordan comutativas, obtêm-se álgebras de Jordan comutativas.

13. Caso os vectores linha de uma matriz  $\mathbf{A}_j$  constituam uma base ortonormada para o espaço imagem de  $\mathbf{Q}_j$ ,  $j = 1, \dots, u$ , tem-se

$$\mathbf{Q}_j = \mathbf{A}_j^\top \mathbf{A}_j \quad j = 1, \dots, u$$

e

$$\mathbf{I}_{g_j} = \mathbf{A}_j \mathbf{A}_j^\top \quad j = 1, \dots, u$$

com  $g_j = \text{car}(\mathbf{Q}_j)$ ,  $j = 1, \dots, u$ .

**Definição 2.2.** Uma matriz  $\mathbf{P} = [\mathbf{A}_1^\top, \dots, \mathbf{A}_u^\top]$  tais que  $\mathbf{I}_{g_j} = \mathbf{A}_j \mathbf{A}_j^\top$  e  $\mathbf{Q}_j = \mathbf{A}_j^\top \mathbf{A}_j$ , com  $j = 1, \dots, u$ , é uma matriz associada à AJC com base principal constituída pelas matrizes  $\{\mathbf{Q}_1, \dots, \mathbf{Q}_u\}$ .

### 2.1.2 Álgebras geradas

Seja  $\mathcal{M} = \{\mathbf{M}_1, \dots, \mathbf{M}_w\}$  uma família de matrizes simétricas  $n \times n$  que comutam, o que equivale, a que as matrizes sejam diagonalizadas por uma matriz ortogonal  $\mathbf{P}$  tendo-se pois  $\mathcal{M} \subset \mathcal{A}(\mathbf{P})$ , onde  $\mathcal{A}(\mathbf{P})$  é a álgebra de Jordan diagonalizadas por  $\mathbf{P}$ , ver Schott (1997) [39].

Existe pois, pelo menos uma álgebra de Jordan comutativa, que contém  $\mathcal{M}$ . Ora, intersectando-se álgebras de Jordan comutativas que contenham  $\mathcal{M}$ , obtêm-se álgebras de Jordan comutativas que contêm  $\mathcal{M}$ . Existirá pois uma álgebra de Jordan comutativa mínima, a álgebra de Jordan comutativa gerada por  $\mathcal{M}$ , que contém  $\mathcal{M}$ , e que representaremos por  $\mathcal{A}(\mathcal{M})$ .

Os vectores linha  $\alpha_1, \dots, \alpha_n$  de  $\mathbf{P}$  são vectores próprios das matrizes de  $\mathcal{M}$ . Pondo  $\alpha_i \tau \alpha_l$  caso

$$\alpha_i^\top \mathbf{M}_j \alpha_i = \alpha_l^\top \mathbf{M}_j \alpha_l, \quad j = 1, \dots, w$$

isto é, se  $\alpha_i$  e  $\alpha_l$  estiverem associados a valores próprios idênticos para todas as matrizes de  $\mathcal{M}$ , definimos uma relação de equivalência,  $\tau$ , em  $\{\alpha_1, \dots, \alpha_n\}$ . Podem ser definidos dois tipos de classe de equivalência.

**Definição 2.3.** *Uma classe de equivalência  $\tau$  é de primeiro tipo se houver pelo menos uma matriz de  $\mathcal{M}$  para a qual os vectores da classe têm valor próprio não nulo. Além das classes de primeiro tipo pode haver uma classe de segundo tipo constituída pelos vectores linha de  $\mathbf{P}$  que têm valores próprios nulos para todas as matrizes de  $\mathcal{M}$ .*

O número de classes de primeiro tipo é o índice próprio de  $\mathcal{M}$ .

Seja  $v$  o índice próprio de  $\mathcal{M}$  e  $\mathcal{C}_1, \dots, \mathcal{C}_v$  os conjuntos de índices dos vectores das várias classes de primeiro tipo. Ponhamos

$$\mathbf{Q}_j = \sum_{i \in \mathcal{C}_j} \boldsymbol{\alpha}_i \boldsymbol{\alpha}_i^\top, \quad j = 1, \dots, v,$$

verificando-se facilmente que estas são matrizes de projecção ortogonal mutuamente ortogonais. Assim  $\mathcal{Q} = \{\mathbf{Q}_1, \dots, \mathbf{Q}_v\}$  será a base principal de uma álgebra de Jordan comutativa que representamos por  $bp(\mathcal{A})$  já que se trata da intersecção de álgebras de Jordan comutativas que contêm  $\mathcal{Q}$ . Por outro lado temos

$$\mathbf{M}_i = \sum_{j=1}^v b_{ij} \mathbf{Q}_j, \quad i = 1, \dots, w,$$

com  $b_{ij}$  o valor próprio de  $\mathbf{M}_i$  para os vectores próprios  $\boldsymbol{\alpha}_i$  com  $l \in \mathcal{C}_j$ ,  $j = 1, \dots, v$ ,  $i = 1, \dots, w$ . Consideremos ainda que  $\mathbf{B} = [b_{i,j}]$ .

Estabeleçamos agora o seguinte lema

**Lemma 2.2.** *Se as matrizes de  $\mathcal{M}$  forem linearmente independentes, os vectores linha de  $\mathbf{B}$  serão linearmente independentes.*

*Demonstração.* Se as matrizes de  $\mathcal{M}$  forem linearmente independentes, então, se  $\mathbf{c}^\top \mathcal{M} = \sum_{i=1}^w c_i \mathbf{M}_i = \mathbf{0}_{n \times n}$  implica  $\mathbf{c} = \mathbf{0}_w$ .

Tem-se  $\mathbf{c}^\top \mathcal{M} = \sum_{i=1}^w c_i \left( \sum_{j=1}^v b_{i,j} \mathbf{Q}_j \right) = \sum_{j=1}^v \left( \sum_{i=1}^w b_{i,j} c_i \right) \mathbf{Q}_j$  e como as matrizes  $\mathbf{Q}_1, \dots, \mathbf{Q}_w$  são mutuamente ortogonais, sendo pois linearmente independentes, tem-se  $\mathbf{c}^\top \mathcal{M} = \mathbf{0}_{n \times n}$  se e só se  $\mathbf{B}^\top \mathbf{c} = \mathbf{0}_v$ . Assim, se as matrizes de  $\mathcal{M}$  forem linearmente

independentes, tem-se  $\mathbf{B}^\top \mathbf{c} = \mathbf{0}_v$  se, e só se,  $\mathbf{c} = \mathbf{0}_v$ , o que equivale a que os vectores linha de  $\mathbf{B}$  sejam linearmente independentes.  $\square$

Tem-se agora o seguinte corolário:

**Corolário 1.** *Para que as matrizes de  $\mathcal{M}$  sejam linearmente independentes, o cardinal  $\#(\mathcal{M})$  não pode exceder o índice próprio de  $\mathcal{M}$ .*

Por outro lado vê-se que, dado  $\mathcal{M} \subset \mathcal{A}(\mathcal{Q})$ , se tem  $\mathcal{A}(\mathcal{M}) \subseteq \mathcal{A}(\mathcal{Q})$ . Assim as matrizes de  $bp(\mathcal{A}(\mathcal{M}))$  pertencem a  $\mathcal{A}(\mathcal{Q})$  sendo portanto somas de uma ou mais das matrizes de  $\mathcal{Q}$ . Se uma das matrizes  $\mathbf{Q}_i^\circ$  da  $bp(\mathcal{A}(\mathcal{M}))$  fosse a soma de duas ou mais dessas matrizes, tendo-se

$$\mathbf{Q}_i^\circ = \sum_{j \in \mathcal{C}_i'} \alpha_j \alpha_j^\top + \sum_{j \in \mathcal{C}_i''} \alpha_j \alpha_j^\top + \dots,$$

os vectores próprios com índices em  $\mathcal{C}_i' \cup \mathcal{C}_i'' \cup \dots$ , teriam vectores próprios idênticos para todas as matrizes da  $\mathcal{M}$  o que é impossível. Assim as matrizes da  $bp(\mathcal{A}(\mathcal{M}))$  serão as matrizes de  $\mathcal{Q}$ , o que permite estabelecer a seguinte proposição:

**Proposição 2.1.** *A álgebra gerada pela família  $\mathcal{M}$  coincide com a álgebra gerada pela família  $\mathcal{Q}$ . Tem-se  $\mathcal{A}(\mathcal{M}) = \mathcal{A}(\mathcal{Q})$ .*

Sendo  $\mathcal{M}$  uma família de matrizes simétricas que comutam, linearmente independentes, podemos completá-la por forma a ter-se uma base  $\mathcal{M}^\circ$  para  $\mathcal{A}(\mathcal{M})$ , dado que qualquer conjunto de elementos linearmente independentes num espaço vectorial pode ser completada por forma a obter-se uma base. Ter-se-á então

$$\mathbf{M}_i^\circ = \sum_{j=1}^v b_{i,j}^\circ \mathbf{Q}_j, \quad i = 1, \dots, v,$$

com  $\mathbf{B}^\circ = [b_{i,j}^\circ]$  uma matriz regular com sub-matriz superior  $\mathbf{B}$ .

Estamos agora em condições de enunciar a seguinte proposição:

**Proposição 2.2.**  *$\mathcal{M}^\circ$  é base de  $\mathcal{A}(\mathcal{M}^\circ)$  se e só se for constituída por matrizes linearmente independentes e tiver cardinal igual ao índice próprio.*

*Demonstração.* Basta observar que a condição equivale a que a matriz  $\mathbf{B}^\circ$  seja regular, o que faz com que  $\mathcal{M}^\circ$  seja, atendendo ao Lemma 2.2, uma família de matrizes de  $\mathcal{A}(\mathcal{M}^\circ)$  linearmente independentes e em número igual a  $\dim(\mathcal{A}(\mathcal{M}^\circ))$ , pelo que  $\mathcal{M}^\circ$  terá de ser base de  $\mathcal{A}(\mathcal{M}^\circ)$ .<sup>5</sup>  $\square$

Caso  $\mathcal{M}$ , família de matrizes simétricas que comutam, constitua uma base para  $\mathcal{A}(\mathbf{M})$ , diremos que é uma família perfeita.

Consideremos agora o caso em que  $\mathcal{M}$  é constituída por uma única matriz, a matriz  $\mathbf{M}$ . Sendo  $a_1, \dots, a_v$  os valores próprios não nulos de  $\mathbf{M}$  e  $\mathbf{Q}_1, \dots, \mathbf{Q}_v$  as matrizes de projecção ortogonal sobre os correspondentes espaços próprios, ter-se-á  $\mathbf{M} = \sum_{j=1}^v a_j \mathbf{Q}_j$  vendo-se ainda que  $\mathcal{Q} = \{\mathbf{Q}_1, \dots, \mathbf{Q}_v\}$  é a base principal que se obtém para  $\mathcal{A}(\mathcal{M})$ .

Com efeito, se se tiver  $\mathbf{Q}_j = \sum_{l \in \mathcal{C}_j} \alpha_j \alpha_j^\top$ ,  $j = 1, \dots, v$  é fácil verificar que  $\mathcal{C}_1, \dots, \mathcal{C}_v$  são os conjuntos dos índices dos vectores próprios pertencentes às várias classes de equivalência  $\tau$ , de primeiro tipo.

A concluir interessa-nos considerar o caso em que  $\mathcal{M} = \mathcal{A}_1 \cup \mathcal{A}_2$  com  $\mathcal{A}_1$  e  $\mathcal{A}_2$  álgebras de Jordan comutativas completas com bases principais  $\mathcal{Q}_1$  e  $\mathcal{Q}_2$ .

Estabeleçamos agora a seguinte proposição:

**Proposição 2.3.** *Seendo  $\mathcal{Q}_l = \{\mathbf{Q}_{l,1}, \dots, \mathbf{Q}_{l,w_l}\} = bp(\mathcal{A}_l)$ ,  $l = 1, 2$  com  $\mathcal{A}_1$  e  $\mathcal{A}_2$  completas, a  $bp(\mathcal{A}(\mathcal{A}_1 \cup \mathcal{A}_2))$ , é constituída pelas matrizes não nulas da forma  $\mathbf{Q}_{1,j_1} \mathbf{Q}_{2,j_2}$ ,  $j_1 = 1, \dots, w_1$ ,  $j_2 = 1, \dots, w_2$ .*

*Demonstração.* Começemos por observar que as matrizes de  $\mathcal{A}_1 \cup \mathcal{A}_2$  comutam gerando pois uma álgebra de Jordan comutativa  $\mathcal{A}(\mathcal{A}_1 \cup \mathcal{A}_2)$  que contém  $\mathcal{Q}_1 \cup \mathcal{Q}_2$  e que as matrizes não nulas da forma  $\mathbf{Q}_{1,j_1} \mathbf{Q}_{2,j_2}$  são matrizes de projecção ortogonal

---

<sup>5</sup>A extensão desta equivalência a espaços vectoriais com dimensão infinita pode ser feita utilizando o lemma de Zorn, em honra do matemático alemão Max August Zorn (1906–1993), apesar da sua primeira formulação se dever ao matemático polaco Kazimierz Kuratowski (1896–1980). Também pode ser consultado o axioma da Escolha e o Teorema de Zermelo (Ernest Zermelo (1871–1953)).

mutuamente ortogonais, constituindo a base principal duma AJC  $\mathcal{A}^\circ$ , que estará contida em todas as álgebras de Jordan comutativas que contêm  $\mathcal{A}_1 \cup \mathcal{A}_2$ . Basta-nos pois mostrar que  $\mathcal{A}_1 \cup \mathcal{A}_2 \subset \mathcal{A}^\circ$  para o que basta mostrar que  $\mathcal{Q}_1 \cup \mathcal{Q}_2 \subset \mathcal{A}^\circ$ . Como  $\mathcal{A}_1$  e  $\mathcal{A}_2$  são completas tem-se

$$\mathbf{Q}_{1,j} = \sum_{j'=1}^{w_2} \mathbf{Q}_{1,j} \mathbf{Q}_{2,j'} \in \mathcal{A}^\circ, \quad j = 1, \dots, w_1,$$

vindo  $\mathbf{Q}_1 \subset \mathcal{A}^\circ$ . De igual forma se mostra que  $\mathbf{Q}_2 \subset \mathcal{A}^\circ$ , o que completa a demonstração.  $\square$

### 2.1.3 Sub-famílias

Seja  $\mathcal{M}$  uma base para  $\mathcal{A}$ , sendo  $\mathcal{Q}$  a base principal de  $\mathcal{A}$  ter-se-á, com  $w = \dim(\mathcal{A})$

$$\mathbf{M}_i = \sum_{j=1}^w b_{i,j} \mathbf{Q}_j, \quad i = 1, \dots, w,$$

sendo  $\mathbf{B} = [b_{i,j}]$  matriz regular  $w \times w$ .

Dada uma sub-família  $\mathcal{M}'$  de  $\mathcal{M}$ , podemos dar às  $m$  matrizes de  $\mathcal{M}'$  os  $m < w$  primeiros índices de  $\mathcal{M}$ . Diremos que  $\mathcal{M}'$  é uma sub-família segregada se se tiver

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_{1,1} & \mathbf{0} \\ \mathbf{B}_{2,1} & \mathbf{B}_{2,2} \end{bmatrix}$$

com  $\mathbf{B}_{2,1}$  sub-matriz  $m \times m$ . Assim, sendo  $\mathcal{Q}' = \{\mathbf{Q}_1, \dots, \mathbf{Q}_m\}$  temos

$$\mathbf{M}' = \mathbf{B}_{1,1} \mathbf{Q}'$$

e

$$\mathbf{M} = \mathbf{B} \mathbf{Q}$$

bem como, dado as matrizes  $\mathbf{B}_{1,1}$  e  $\mathbf{B}$  serem regulares,

$$\mathbf{Q}' = \mathbf{B}_{1,1}^{-1} \mathbf{M}'$$

e

$$Q = B^{-1}M .$$

Enquanto  $\mathcal{Q}$  é a base principal da álgebra  $\mathcal{A}(\mathcal{M})$ ,  $\mathcal{Q}'$  será a base principal da álgebra  $\mathcal{A}(\mathcal{M}')$ .

Até agora temos considerado  $\mathcal{M}$  como base para  $\mathcal{A}$ . Suponhamos agora que  $\mathcal{M}$  não é base para  $\mathcal{A} = \mathcal{A}(\mathcal{M})$ , embora as matrizes de  $\mathcal{M}$  continuem a ser linearmente independentes<sup>6</sup>. Temos  $l = \#(\mathcal{M}) < w$ , pelo que a matriz  $B$  será do tipo  $l \times w$ . Suponhamos que as  $h < l$  primeiras matrizes de  $\mathcal{M}$  constituem a sub-família  $\mathcal{M}'$ . A situação de segregação continua a conduzir a uma matriz de transição de forma triangular inferior, com  $M' = B_{1,1}Q'$  e  $M = BQ$  e as matrizes  $B_{1,1}$  e  $B$  são dos tipos  $h \times m$  e  $l \times w$ . Podemos completar  $\mathcal{M}'$  para obter uma base  $\mathcal{M}^\circ$  para  $\mathcal{A}(\mathcal{Q}')$  e em seguida completar a família  $\mathcal{M} \setminus \mathcal{M}'$  das últimas  $l - h$  matrizes de  $\mathcal{M}$ , de forma a ter-se uma base para  $\mathcal{A}(\mathcal{Q} \setminus \mathcal{Q}')$ .

Interessa-nos também considerar um outro aspecto do completamento de famílias, a de famílias de matrizes simétricas que comutam. Duas destas famílias dizem-se  $\mathcal{G}$ -equivalentes, se geram a mesma AJC. As bases da AJC serão  $\mathcal{G}$ -equivalentes.

Interessa-nos ainda considerar os pares  $(\mathcal{M}', \mathcal{M})$  de famílias de matrizes simétricas que comutam. Um par deste tipo de matrizes diz-se estruturado se  $\mathcal{M}' \subset \mathcal{M}$  e  $\mathcal{A}(\mathcal{M}') \subset \mathcal{A}(\mathcal{M})$ . Dois pares estruturados serão  $\mathcal{G}$ -equivalentes, se as AJC geradas pelas famílias e sub-famílias forem idênticas. Sendo  $\mathcal{A}' = \mathcal{A}(\mathcal{M}')$  e  $\mathcal{A} = \mathcal{A}(\mathcal{M})$ , representaremos por  $U'$  e  $U$  as respectivas matrizes unidade e por  $c'$  e  $c$  as características das mesmas. Se  $c' < c$ ,  $U'$  terá de ser soma de apenas parte das matrizes da base principal de  $\mathcal{A}$ , observando-se a segregação. Assim, o par estruturado

---

<sup>6</sup>A título de exemplo, considerem-se as matrizes  $M_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$  e  $M_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ ,

com vectores próprios  $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ ,  $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$  e  $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$

$(\mathcal{M}', \mathcal{M})$  de famílias de matrizes simétricas que comutam é segregado, se e só se  $c' < c$ .

### 2.1.4 Produto de Kronecker

Neste trabalho iremos usar, para além do produto usual de matrizes, o produto de Kronecker<sup>7</sup>, que representaremos por  $\otimes$ . Diversos são os trabalhos onde está desenvolvida esta operação, veja-se por exemplo Graham (1981) [19]. Apresentamos de seguida alguns resultados do produto de Kronecker que serão necessários para o desenvolvimento do nosso trabalho.

Considere-se uma matriz  $\mathbf{A} = [a_{i,j}]$  do tipo  $r \times s$ , define-se então

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{1,1}\mathbf{B} & \cdots & a_{1,s}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{r,1}\mathbf{B} & \cdots & a_{r,s}\mathbf{B} \end{bmatrix}$$

Se os produtos  $\mathbf{A}_j\mathbf{B}_j$ , com  $j = 1, 2$ , estiverem definidos, poder-se-á então verificar que

$$\begin{aligned} (\mathbf{A}_1 \otimes \mathbf{A}_2)(\mathbf{B}_1 \otimes \mathbf{B}_2) &= \begin{bmatrix} a_{1,1}\mathbf{A}_2 & \cdots & a_{1,s}\mathbf{A}_2 \\ \vdots & \ddots & \vdots \\ a_{r,1}\mathbf{A}_2 & \cdots & a_{r,s}\mathbf{A}_2 \end{bmatrix} \begin{bmatrix} b_{1,1}\mathbf{B}_2 & \cdots & b_{1,t}\mathbf{B}_2 \\ \vdots & \ddots & \vdots \\ b_{s,1}\mathbf{B}_2 & \cdots & b_{s,t}\mathbf{B}_2 \end{bmatrix} = \\ &= \begin{bmatrix} \left( \sum_{j=1}^s a_{1,j}b_{j,1} \right) & \cdots & \left( \sum_{j=1}^s a_{1,j}b_{j,t} \right) \\ \vdots & \ddots & \vdots \\ \left( \sum_{j=1}^s a_{r,j}b_{j,1} \right) & \cdots & \left( \sum_{j=1}^s a_{r,j}b_{j,t} \right) \end{bmatrix} \otimes (\mathbf{A}_2\mathbf{B}_2) = \\ &= (\mathbf{A}_1\mathbf{B}_1) \otimes (\mathbf{A}_2\mathbf{B}_2). \end{aligned}$$

---

<sup>7</sup>Leopold Kronecker (1823–1891). Matemático alemão nascido em Liegnitz (Prússia, agora Legnica - Polónia) tendo morrido em Berlin.

Igualmente se verifica que

$$(\mathbf{A} \otimes \mathbf{B})^\top = \mathbf{A}^\top \otimes \mathbf{B}^\top.$$

Veamos ainda algumas outras propriedades que serão importantes no desenvolvimento do nosso trabalho.

Considere-se a matriz identidade de ordem  $r$ ,  $\mathbf{I}_r$ , e  $\mathbf{J}_u = \mathbf{1}^u \mathbf{1}^{u^\top}$ , verifica-se que

$$\begin{cases} \mathbf{I}_u \otimes \mathbf{I}_v = \mathbf{I}_{uv} \\ \mathbf{1}^u \otimes \mathbf{1}^v = \mathbf{1}^{uv} \\ \mathbf{J}_u \otimes \mathbf{J}_v = \mathbf{J}_{uv} \end{cases}.$$

Interessa-nos também ver o que se passa com matrizes de projecção ortogonal. Uma matriz é MPO se e só se for simétrica e idempotente. Verifica-se facilmente que o produto de Kronecker de matrizes ortogonais [simétricas, idempotentes] origina matrizes ortogonais [simétricas, idempotentes], pelo que o produto de Kronecker para matrizes de projecção ortogonal é também uma matriz de projecção ortogonal. É ainda fácil de verificar que se  $\mathbf{Q}_1 \mathbf{Q}_1^\top = \mathbf{0}_{n_1 \times n_1}$  ou  $\mathbf{Q}_2 \mathbf{Q}_2^\top = \mathbf{0}_{n_2 \times n_2}$ , se tem  $(\mathbf{Q}_1 \otimes \mathbf{Q}_2) (\mathbf{Q}_1^\top \otimes \mathbf{Q}_2^\top) = \mathbf{0}_{n_1 n_2 \times n_1 n_2}$ .

Sendo  $\mathbf{D}(r^l)$ , a matriz diagonal, cujos elementos principais são as componentes de  $r^l$ , ter-se-á

$$\mathbf{D}(r^l) \otimes \mathbf{D}(s^h) = \mathbf{D}(r^l \otimes s^h).$$

Uma matriz ortogonal de ordem  $s$  é ortogonal estandardizada se os elementos da sua primeira linha forem iguais a  $s^{-\frac{1}{2}}$ . O produto de Kronecker de matrizes ortogonais estandardizadas  $\mathbf{P}_1$  e  $\mathbf{P}_2$ , de ordens  $u$  e  $v$ , será uma matriz ortogonal estandardizada de ordem  $uv$ .

Considere-se agora uma matriz simétrica  $\mathbf{M}$  e o vector dos seus valores próprios  $\mathbf{r}(\mathbf{M})$ . Se  $\mathbf{D}(\mathbf{M}) = \mathbf{D}(\mathbf{r}(\mathbf{M}))$ , então, com  $\mathbf{P}$  a matriz diagonalizadora ortogonal de  $\mathbf{M}$ , verifica-se que

$$\mathbf{M} = \mathbf{P}^\top \mathbf{D}(\mathbf{M}) \mathbf{P},$$

e como tal, se  $\mathbf{P}_1$  e  $\mathbf{P}_2$  forem as matrizes diagonalizadoras ortogonais das matrizes simétricas  $\mathbf{M}_1$  e  $\mathbf{M}_2$ , então

$$\begin{aligned} & (\mathbf{P}_1 \otimes \mathbf{P}_2)(\mathbf{M}_1 \otimes \mathbf{M}_2)(\mathbf{P}_1 \otimes \mathbf{P}_2)^\top = \\ &= (\mathbf{P}_1 \otimes \mathbf{P}_2) [(\mathbf{P}_1^\top \mathbf{D}(\mathbf{M}_1) \mathbf{P}_1) (\mathbf{P}_2^\top \mathbf{D}(\mathbf{M}_2) \mathbf{P}_2)] (\mathbf{P}_1^\top \otimes \mathbf{P}_2^\top) \\ &= [\mathbf{P}_1 \mathbf{P}_1^\top \mathbf{D}(\mathbf{M}_1) \mathbf{P}_1 \mathbf{P}_1^\top] \otimes [\mathbf{P}_2 \mathbf{P}_2^\top \mathbf{D}(\mathbf{M}_2) \mathbf{P}_2 \mathbf{P}_2^\top] \\ &= \mathbf{D}(\mathbf{M}_1) \otimes \mathbf{D}(\mathbf{M}_2). \end{aligned}$$

Da expressão anterior demonstra-se que  $\mathbf{P}_1 \otimes \mathbf{P}_2$  é diagonalizadora ortogonal de  $\mathbf{M}_1 \otimes \mathbf{M}_2$  e que

$$\mathbf{D}(\mathbf{M}_1 \otimes \mathbf{M}_2) = \mathbf{D}(\mathbf{M}_1) \otimes \mathbf{D}(\mathbf{M}_2).$$

Um outro resultado que se pode deduzir, relacionado com as características de matrizes, é

$$\text{car}(\mathbf{M}_1 \otimes \mathbf{M}_2) = \text{car}(\mathbf{M}_1) \text{car}(\mathbf{M}_2),$$

com  $\text{car}(\mathbf{M})$  a característica da matriz  $\mathbf{M}$ . Essa característica é igual ao número dos seus valores próprios não nulos.

Este resultado pode ainda estender-se a matrizes não simétricas, ver por exemplo Silvey (1975)[45]. Assim,

$$\text{car}(\mathbf{U}\mathbf{U}^\top) = \text{car}(\mathbf{U})$$

tendo-se

$$\begin{aligned} \text{car}(\mathbf{M}_1 \otimes \mathbf{M}_2) &= \text{car} [(\mathbf{M}_1 \otimes \mathbf{M}_2)(\mathbf{M}_1 \otimes \mathbf{M}_2)^\top] \\ &= \text{car} [\mathbf{M}_1 \mathbf{M}_1^\top \otimes \mathbf{M}_2 \mathbf{M}_2^\top] \\ &= \text{car}(\mathbf{M}_1) \text{car}(\mathbf{M}_2). \end{aligned}$$

### 2.1.5 Operações Binárias

No que se segue, utilizaremos duas operações binárias sobre as AJC.

#### Produto de Kronecker

Consideremos a seguinte proposição.

**Proposição 2.4.** Sendo  $\{\mathcal{Q}_{j,1}, \dots, \mathcal{Q}_{j,u_j}\}$  MPO mutuamente ortogonais que constituem a base principal da AJC  $\mathcal{A}_j$ ,  $j = 1, 2$ ,

$$\mathcal{A}_1 \otimes \mathcal{A}_2 = \left\{ \sum_{i=1}^{u_1} \sum_{j=1}^{u_2} a_{i,j} (\mathcal{Q}_{1,i} \otimes \mathcal{Q}_{2,j}) \right\},$$

será uma AJC com base principal  $\{\mathcal{Q}_{1,i} \otimes \mathcal{Q}_{2,j}\}$ ,  $i = 1, \dots, u_1; j = 1, \dots, u_2$ .

*Demonstração.* É fácil verificar que  $\mathcal{A}_1 \otimes \mathcal{A}_2$  é um espaço vectorial que contém o quadrado das suas matrizes as quais comutam, sendo portanto uma AJC. Verifica-se ainda que  $\{\mathcal{Q}_{1,i} \otimes \mathcal{Q}_{2,j}\}$ ,  $i = 1, \dots, u_1; j = 1, \dots, u_2$ , é um sistema de geradores de  $\mathcal{A}_1 \otimes \mathcal{A}_2$  constituído por MPO. Como estas MPO são simétricas, idempotentes e mutuamente ortogonais, são linearmente independentes constituindo portanto a base principal de  $\mathcal{A}_1 \otimes \mathcal{A}_2$ .  $\square$

Define-se assim uma primeira operação binária entre AJCs. Prova-se que a mesma é associativa, ver Fonseca *et al* (2006) [15].

Suponhamos agora que

$$\mathbf{M}_{l,i} = \sum_{j=1}^{u_l} b_{i,j}(l) \mathcal{Q}_{l,j} \quad l = 1, \dots, v_l; \quad l = 1, 2$$

ter-se-á então

$$\mathbf{M}_{1,l} \otimes \mathbf{M}_{2,h} = \sum_{j=1}^{u_1} \sum_{k=1}^{u_2} b_{i,j}(1) b_{h,k}(2) \mathcal{Q}_{1,j} \otimes \mathcal{Q}_{2,k} \quad i = 1, \dots, v_1; h = 1, \dots, v_2.$$

Em particular se  $u_1 = u_2$  e  $\{\mathbf{M}_{l,1}, \dots, \mathbf{M}_{l,u_l}\}$  for uma base para  $\mathcal{A}_l$ ,  $l = 1, 2$  as matrizes

$$\mathbf{B}(l) = [b_{i,j}(l)],$$

sendo matrizes de transição, então estas bases são regulares.

Verifica-se facilmente que

$$\mathbf{B} = \mathbf{B}(1) \otimes \mathbf{B}(2)$$

também é regular, pelo que  $\{\mathbf{M}_{1,i} \otimes \mathbf{M}_{2,h}\}$ ,  $i = 1, \dots, u_1, h = 1, \dots, u_2$  também será base de  $\mathcal{A}_1 \otimes \mathcal{A}_2$ .

Admitamos agora que  $\mathcal{A}_l$  é completa e regular sendo constituída por matrizes  $n_l \times n_l$ ,  $l = 1, 2$ . Então

$$\left\{ \mathbf{Q}_{1,i} \otimes \frac{1}{n_2} \mathbf{J}_{n_2} \right\} \cup \{ \mathbf{I}_{n_1} \otimes \mathbf{Q}_{2,j} \}$$

com  $i = 1, \dots, u_1$  e  $j = 2, \dots, u_2$ , será uma família de MPO mutuamente ortogonais a qual será a base principal duma AJC,  $\mathcal{A}_1 \star \mathcal{A}_2$ . Define-se assim uma segunda operação binária sobre AJC, que designaremos por Produto de Kronecker Restrito. Também esta operação é associativa, como veremos adiante.

### Produto de Kronecker Restrito

Consideremos as AJC  $\mathcal{A}_1$  e  $\mathcal{A}_2$  com base principal  $\mathcal{Q}_1 = \{\mathbf{Q}_{1j}, j = 1, \dots, w_1\}$  e  $\mathcal{Q}_2 = \{\mathbf{Q}_{2j}, j = 1, \dots, w_2\}$  constituída por MPO  $n_i \times n_i$ ,  $i = 1, 2$ . O elemento identidade de  $\mathcal{A}_i$  é dado por

$$\mathbf{U}_i = \sum_{j=1}^{w_i} \mathbf{Q}_{ij}.$$

Estamos em condições de enunciar a seguinte proposição.

**Proposição 2.5.** *Dado  $k = 1, \dots, w_2 - 1$ , a família*

$$\mathcal{Q}_k = \{ \mathbf{Q}_{1h} \otimes \mathbf{Q}_{2h'} \} \cup \{ \mathbf{U}_1 \otimes \mathbf{Q}_{2h''} \}$$

com  $h = 1, \dots, w_1$ ,  $h' = 1, \dots, k$  e  $h'' = k + 1, \dots, w_2$ , então é uma família de matrizes ortogonais mutuamente ortogonais.

*Demonstração.* Uma vez que o produto de Kronecker de MPO é uma MPO,  $\mathcal{Q}_k$  será uma família de MPO. Para além disso, dadas duas matrizes distintas  $\mathbf{Q}_{11}^\circ \otimes \mathbf{Q}_{21}^\circ$  e  $\mathbf{Q}_{12}^\circ \otimes \mathbf{Q}_{22}^\circ$  pertencentes a  $\mathcal{Q}_k$ , por forma a que  $\mathbf{Q}_{11}^\circ \neq \mathbf{Q}_{12}^\circ$  ou  $\mathbf{Q}_{21}^\circ \neq \mathbf{Q}_{22}^\circ$ , necessariamente

$$(\mathbf{Q}_{11}^\circ \otimes \mathbf{Q}_{21}^\circ) (\mathbf{Q}_{12}^\circ \otimes \mathbf{Q}_{22}^\circ) = (\mathbf{Q}_{11}^\circ \mathbf{Q}_{12}^\circ) \otimes (\mathbf{Q}_{21}^\circ \mathbf{Q}_{22}^\circ) = \mathbf{0},$$

o que completa a demonstração. □

A AJC com base principal  $\mathcal{Q}_k$  será o Produto de Kronecker Restrito de  $\mathcal{A}_1$  e  $\mathcal{A}_2$ . Representaremos esta AJC por  $\mathcal{A}_1 \star_k \mathcal{A}_2$ . Quando  $k = 1$ , escreveremos  $\mathcal{A}_1 \star \mathcal{A}_2$ .

De referir que  $\mathcal{A}_1 \star_{w_2} \mathcal{A}_2 = \mathcal{A}_1 \otimes \mathcal{A}_2$ .

Admitamos que os vetores linha de  $\mathbf{A}_{lh}$  constitui uma base ortonormal do espaço imagem de  $\mathbf{Q}_{lh}$ ,  $h = 1, \dots, w_1$ ,  $l = 1, 2$ . Se  $\mathcal{A}_1$  e  $\mathcal{A}_2$  forem completas, as matrizes

$$\mathbf{G}_l = \left[ \mathbf{A}_{l1}^\top \quad \dots \quad \mathbf{A}_{lw_l}^\top \right]^\top, \quad l = 1, 2$$

serão matrizes ortogonais associadas a  $\mathcal{A}_1$  e  $\mathcal{A}_2$ .

Diremos que a matriz é sub-ortogonal se for obtida a partir de uma matriz ortogonal à qual se retira parte das linhas. Por exemplo, se as matrizes  $\mathbf{Q}_j = \mathbf{A}_j^\top \mathbf{A}_j$ ,  $j = 1, \dots, w$ , constituem a base principal de uma AJC não completa,  $\mathcal{A}$ , então  $\mathbf{A} = \left[ \mathbf{A}_1^\top \quad \dots \quad \mathbf{A}_w^\top \right]^\top$  é a sub-matriz ortogonal associada a  $\mathcal{A}$ , i.e.,  $\mathbf{A}$  será a sub-matriz de uma matriz ortogonal associada a  $\mathcal{A}_c$ . Então

$$\mathbf{G} = \mathbf{G}_1 \star_k \mathbf{G}_2 = \left[ \mathbf{A}_{1h}^\top \otimes \mathbf{A}_{2h'}^\top; \mathbf{G}_1 \otimes \mathbf{A}_{2h''}^\top \right]$$

onde  $h = 1, \dots, w_1$ ,  $h' = 1, \dots, k$  e  $h'' = k + 1, \dots, w_2$ , será a matriz sub-ortogonal associada a  $\mathcal{A}_1 \star_k \mathcal{A}_2$  se uma das  $\mathcal{A}_1$  e  $\mathcal{A}_2$  não for completa. É imediato que se  $\mathcal{A}_1$  e  $\mathcal{A}_2$  forem ambas completas, então  $\mathbf{G}_1 \star_k \mathbf{G}_2$  é ortogonal.

Do exposto ressalta que se  $\mathbf{G}$  for uma matriz ortogonal ou sub-ortogonal associada à AJC,

$$\mathbf{U} = \mathbf{G}\mathbf{G}^+,$$

sendo a soma das matrizes da base principal da AJC, a unidade da mesma.

Se considerarmos ainda as matrizes sub-ortogonais  $\mathbf{G}_1$ ,  $\mathbf{G}_2$  e  $\mathbf{G}_3$  associadas às AJCs  $\mathcal{A}_1$ ,  $\mathcal{A}_2$  e  $\mathcal{A}_3$ , podemos verificar que

$$\mathbf{G}_1 \star_{k_1} \left( \mathbf{G}_2 \star_{k_2} \mathbf{G}_3 \right) = \left( \mathbf{G}_1 \star_{k_1} \mathbf{G}_2 \right) \star_{k_2} \mathbf{G}_3$$

pelo que podemos constatar que estas operações binárias gozam de associatividade generalizada.

Utilizando a associatividade das duas operações, podem-se construir álgebras complexas a partir de álgebras completas e regulares simples. Recorde-se que enquanto definimos  $\mathcal{A}_1 \otimes \mathcal{A}_2$  para quaisquer AJC, só definimos  $\mathcal{A}_3 \star \mathcal{A}_4$  para  $\mathcal{A}_3$  e  $\mathcal{A}_4$  completas e regulares. Adiante encontraremos, no tratamento de uma aplicação, uma álgebra da forma

$$(\mathcal{A}_1 \otimes \mathcal{A}_2 \otimes \mathcal{A}_3) \star \mathcal{A}_4 \star \mathcal{A}_5.$$

Regressando ao produto  $\mathcal{A} \star \mathcal{A}_2$ , admitamos que  $\{\mathbf{M}_{l,1}, \dots, \mathbf{M}_{l,u_2}\}$  constituem uma base para  $\mathcal{A}_l$  e que se tem a matriz de transição  $\mathbf{B}(l)$ ,  $l = 1, 2$ . Ter-se-á então,

$$\mathbf{M}_{1,i} \otimes \mathbf{J}_{n_2} = n_1 \sum_{j=1}^{u_1} b_{i,j}(1) \otimes \left( \frac{1}{n_1} \mathbf{J}_{n_1} \right) \quad i = 1, \dots, u_1$$

bem como, dado  $\sum_{j=1}^{u_1} \mathbf{Q}_{1,j} = \mathbf{I}_{n_1}$  e  $\mathbf{Q}_{2,1} = \frac{1}{n_2} \mathbf{J}_{n_2}$ ,

$$\begin{aligned} \mathbf{I}_{n_1} \otimes \mathbf{M}_{2,h} &= \sum_{k=1}^{u_1} b_{h,k}(2) (\mathbf{I}_{n_1} \otimes \mathbf{Q}_{2,k}) \\ &= b_{h,1}(2) \left( \sum_{j=1}^{u_1} \mathbf{Q}_{1,j} \otimes \frac{1}{n_2} \mathbf{J}_{n_2} \right) + \sum_{k=2}^{u_2} b_{h,k} (\mathbf{I}_{n_1} \otimes \mathbf{Q}_{2,k}) \end{aligned}$$

$h = 2, \dots, u_2$ .

Assim, teremos agora a matriz de transição

$$\mathbf{B}(1) \star \mathbf{B}(2) = \begin{bmatrix} n_1 \mathbf{B}_1 & \mathbf{0} \\ \mathbf{b}(2) \mathbf{1}^\top & \mathbf{B}^\circ(2) \end{bmatrix},$$

onde  $\mathbf{b}(2)$  é obtido retirando a primeira componente ao primeiro vector coluna de  $\mathbf{B}(2)$  e  $\mathbf{B}^\circ(2)$  retirando a primeira linha e coluna a  $\mathbf{B}(2)$ .

Caso  $\mathbf{B}^\circ(2)$  seja regular,  $\mathbf{B}(1) \star \mathbf{B}(2)$  será regular e

$$\{\mathbf{M}_{1,i} \otimes \mathbf{I}_{n_2}\} \cup \{\mathbf{I}_{n_1} \otimes \mathbf{M}_{2,h}\},$$

$i = 1, \dots, u_1$ ,  $h = 2, \dots, u_2$ , constituirá uma base para  $\mathcal{A}_1 \star \mathcal{A}_2$ . Esta situação verifica-se em particular quando

$$\mathbf{M}_{2,1} = \frac{a}{n_2} \otimes \mathbf{J}_{n_2}.$$

Considere-se a base principal da álgebra  $\mathcal{A}$

$$bp(\mathcal{A}) = \{\mathbf{Q}_1, \dots, \mathbf{Q}_w\}$$

bem como  $\bar{w} = \{1, \dots, w\}$ , tomando-se

$$\begin{cases} \mathcal{C} \subseteq \bar{w} \\ \mathcal{C}^c = \bar{w} \setminus \mathcal{C} \end{cases}.$$

Podemos então definir a base principal para  $\mathcal{A}(\mathcal{C})$ , como

$$bp(\mathcal{A}(\mathcal{C})) = \{\mathbf{Q}_j; j \in \mathcal{C}\}.$$

Consideremos ainda uma outra álgebra  $\mathcal{A}^\circ$  com dimensão  $n^\circ \times n^\circ$ . A sua base principal será

$$bp(\mathcal{A}^\circ) = \{\mathbf{Q}_1^\circ, \dots, \mathbf{Q}_w^\circ\}$$

tomandos-se

$$bp(\overline{\mathcal{A}}(\mathcal{C})) = \{\mathbf{Q}_i^\circ \otimes \mathbf{Q}_j, i = 1, \dots, w^\circ, j \in \mathcal{C}\} \cup \{\mathbf{I}_{n^\circ} \otimes \mathbf{Q}_j; j \notin \mathcal{C}\}.$$

Finalmente se verifica que

$$\overline{\mathcal{A}}(\mathcal{C}) = (\mathcal{A}^\circ \otimes \mathcal{A}(\mathcal{C})) \boxplus (\mathcal{A}(\mathbf{I}_{n^\circ}) \otimes \mathcal{A}(\mathcal{C}^c)),$$

com  $bp(\mathcal{A}(\mathbf{I}_{n^\circ})) = \{\mathbf{I}_{n^\circ}\}$ , onde a primeira parcela é ortogonal à segunda parcela.

Igualmente se verifica que  $\mathcal{A}(\mathbf{I}_{n^\circ}) = \{c\mathbf{I}_{n^\circ}\}$ .

Facilmente se verifica que

$$\left( \sum_{i=1}^{n^\circ} \mathbf{Q}_i^\circ \right) \otimes \left( \sum_{j \in \mathcal{C}} \mathbf{Q}_j \right) = \mathbf{I}_{n^\circ} \otimes \left( \sum_{j \in \mathcal{C}} \mathbf{Q}_j \right)$$

e que

$$\mathbf{I}_{n^\circ} \otimes \left( \sum_{j \notin \mathcal{C}} \mathbf{Q}_j \right) = \mathbf{I}_{n^\circ} \otimes \mathbf{I}_{n^\circ}$$

O caso particular em que  $\mathcal{C} = \{1, \dots, k\}$  foi já considerado por Covas (2007)[7].

Para efeitos deste estudo, interessa-nos o caso em que operaremos com famílias,  $\mathcal{M}_1$  e  $\mathcal{M}_2$  de matrizes simétricas que comutam, com  $\mathcal{Q}_1 = \{\mathbf{Q}_{1,1}, \dots, \mathbf{Q}_{1,w_1}\}$ ,  $\mathcal{Q}_2 = \{\mathbf{Q}_{2,1}, \dots, \mathbf{Q}_{2,w_2}\}$  as bases principais de  $\mathcal{A}_1 = \mathcal{A}(\mathcal{M}_1)$  e  $\mathcal{A}_2 = \mathcal{A}(\mathcal{M}_2)$ .

Como, com  $\mathbf{U}_1 = \sum_{j=1}^{w_1} \mathbf{Q}_{1,j}$ , a base principal de  $\mathcal{A}_1 \star_k \mathcal{A}_2$  é

$$\begin{aligned} bp(\mathcal{A}_1 \star_k \mathcal{A}_2) &= \left( \bigcup_{j=1}^k \{\mathbf{Q}_{1,1} \otimes \mathbf{Q}_{2,j} \dots \mathbf{Q}_{1,w_1} \otimes \mathbf{Q}_{2,j}\} \right) \cup \\ &\cup \{\mathbf{U}_1 \otimes \mathbf{Q}_{2,j}; j = k+1, \dots, w_2\}, \end{aligned}$$

então, qualquer matriz de  $\mathcal{A}_1 \star_k \mathcal{A}_2$  será da forma

$$\mathcal{M}_1 \star_k \mathcal{M}_2 = \{\mathbf{M}_{ih} \otimes \mathbf{M}_{2h'}\} \cup \{\mathbf{U}_1 \otimes \mathbf{M}_{2h''}\},$$

com  $h = 1, \dots, w_1$ ,  $h' = 1, \dots, k$  e  $h'' = k+1, \dots, w_2$ , então, qualquer matriz, por exemplo  $\mathbf{M}$ , da  $\mathcal{A}(\mathcal{M}_1 \star_k \mathcal{M}_2)$  será da forma

$$\mathbf{M} = \sum_{i_1=1}^{w_1} \sum_{i_2=1}^k a_{1i_1 i_2} \mathbf{M}_{1i_1} \otimes \mathbf{M}_{2i_2} + \sum_{i_3=k+1}^{w_2} a_{2i_3} \mathbf{U}_1 \otimes \mathbf{M}_{2i_3}. \quad (2.1)$$

## 2.2 Mínimos Quadrados

Consideremos um modelo linear

$$\mathbf{y} = \sum_{i=1}^w \mathbf{X}_i \beta_i + \boldsymbol{\varepsilon}_i$$

onde  $\mathbf{X}_i = [\mathbf{X}_1, \dots, \mathbf{X}_m]$ ,  $\boldsymbol{\beta} = [\beta_1, \dots, \beta_m]^\top$  com vector médio  $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$  e matriz de covariância  $\mathbf{V}$ . Seja ainda  $\mathbf{T}$ , a MPO sobre o espaço-imagem de  $\mathbf{X}$ . Seguindo Zmyslony (1986) [52], vamos mostrar que, quando  $\mathbf{V}$  e  $\mathbf{T}$  comutam, os estimadores



dos mínimos quadrados têm variância mínima na classe dos estimadores lineares e centrados, sendo pois BLUE.

Ora, ver Mexia (1995) [30], tem-se

$$\mathbf{T} = \mathbf{X} (\mathbf{X}^\top \mathbf{X})^+ \mathbf{X}^\top$$

pelo que o estimador de mínimos quadrados de  $\boldsymbol{\beta}$  será

$$\boldsymbol{\beta}^* = (\mathbf{X}^\top \mathbf{X})^+ \mathbf{X}^\top \mathbf{Y}.$$

Por outro lado  $\boldsymbol{\Psi} = \mathbf{A}\boldsymbol{\beta}$  é, ver Mexia (1990) [29], estimável se e só se  $R(\mathbf{A}^\top) \subseteq R(\mathbf{X}^\top)$  o que equivale a existir  $\mathbf{K}$  tal que

$$\mathbf{A} = \mathbf{K}\mathbf{X},$$

tendo-se para  $\boldsymbol{\Psi}$  o estimador de mínimos quadrados

$$\boldsymbol{\Psi}^* = \mathbf{A}\boldsymbol{\beta}^* = \mathbf{K}\mathbf{X}\boldsymbol{\beta}^* = \mathbf{K}\mathbf{T}\mathbf{Y}.$$

Por outro lado,

$$\mathbb{E}(\mathbf{M}\mathbf{Y}) = \mathbf{M}\mathbf{X}\boldsymbol{\beta} = \mathbf{M}\mathbf{T}\mathbf{X}\boldsymbol{\beta} = \mathbb{E}(\mathbf{M}\mathbf{T}\mathbf{Y})$$

pelo que,  $\mathbf{M}\mathbf{Y}$  e  $\mathbf{M}\mathbf{T}\mathbf{Y}$  terão o mesmo vector médio.

Caso  $\mathbf{T}$  comute com  $\mathbf{V}$ ,  $\mathbf{I}_n - \mathbf{T}$  também comutará com  $\mathbf{V}$ , vindo

$$\begin{aligned} \mathbf{V}(\mathbf{M}\mathbf{Y}) &= \mathbf{M}\mathbf{V}\mathbf{M}^\top = \mathbf{M}[\mathbf{T} + (\mathbf{I} - \mathbf{T})]\mathbf{V}[\mathbf{T} + (\mathbf{I} - \mathbf{T})]\mathbf{M}^\top \\ &= \mathbf{M}\mathbf{T}\mathbf{V}\mathbf{T}\mathbf{M}^\top + \mathbf{M}\mathbf{T}\mathbf{V}(\mathbf{I} - \mathbf{T})\mathbf{M}^\top + \mathbf{M}(\mathbf{I} - \mathbf{T})\mathbf{V}\mathbf{T}\mathbf{M}^\top + \\ &\quad + \mathbf{M}(\mathbf{I} - \mathbf{T})\mathbf{V}(\mathbf{I} - \mathbf{T})\mathbf{M}^\top \\ &= \mathbf{M}\mathbf{T}\mathbf{V}\mathbf{T}\mathbf{M}^\top + \mathbf{M}(\mathbf{I} - \mathbf{T})\mathbf{V}(\mathbf{I} - \mathbf{T})^\top\mathbf{M}^\top \\ &= \mathbf{V}(\mathbf{M}\mathbf{T}\mathbf{Y}) + \mathbf{V}(\mathbf{M}(\mathbf{I} - \mathbf{T})\mathbf{Y}). \end{aligned}$$

Interessa-nos estabelecer a seguinte proposição:

**Proposição 2.6.** *Tem-se  $\mathbb{E}(\mathbf{M}_1\mathbf{Y}) = \mathbb{E}(\mathbf{M}_2\mathbf{Y})$  se e só se  $\mathbf{M}_1\mathbf{T} = \mathbf{M}_2\mathbf{T}$ .*

*Demonstração.* Se  $\mathbb{E}(\mathbf{M}_1\mathbf{Y}) = \mathbb{E}(\mathbf{M}_2\mathbf{Y})$ , tem-se,  $\forall\beta$ ,  $\mathbf{M}_1\mathbf{X}\beta = \mathbf{M}_2\mathbf{X}\beta$ , vindo  $\mathbf{M}_1\mathbf{X} = \mathbf{M}_2\mathbf{X}$  bem como  $\mathbf{M}_1\mathbf{T}\mathbf{X} = \mathbf{M}_2\mathbf{T}\mathbf{X}$  pelo que  $(\mathbf{M}_1 - \mathbf{M}_2)\mathbf{T}\mathbf{X} = \mathbf{0}$  pelo que os vectores linha de  $(\mathbf{M}_1 - \mathbf{M}_2)\mathbf{T}$  terão de ser ortogonais a  $\Omega = R(\mathbf{X})$ .

No entanto como  $\mathbf{T}$  é a MPO sobre  $\Omega$ , esses vectores linha pertencem a  $\Omega$ , sendo pois nulos, logo  $(\mathbf{M}_1 - \mathbf{M}_2)\mathbf{T} = \mathbf{0}$ , ou seja,  $\mathbf{M}_1\mathbf{T} = \mathbf{M}_2\mathbf{T}$ .

Inversamente, se  $\mathbf{M}_1\mathbf{T} = \mathbf{M}_2\mathbf{T}$  tem-se,

$$\mathbb{E}(\mathbf{M}_1\mathbf{Y}) = \mathbf{M}_1\mathbf{X}\beta = \mathbf{M}_1\mathbf{T}\mathbf{X}\beta = \mathbf{M}_2\mathbf{T}\mathbf{X}\beta = \mathbf{M}_2\beta = \mathbb{E}(\mathbf{M}_2\mathbf{Y}).$$

□

**Corolário 2.** *O BLUE de  $\mathbb{E}(\mathbf{M}\mathbf{Y})$  é  $\mathbf{M}\mathbf{T}\mathbf{Y}$ .*

*Demonstração.* Sendo  $\mathbf{M}^\circ\mathbf{Y}$  outro estimador centrado de  $\mathbb{E}(\mathbf{M}\mathbf{Y})$  ter-se-á  $\mathbf{M}^\circ\mathbf{T} = \mathbf{M}\mathbf{T}$  bem como

$$\begin{aligned} \mathbf{V}(\mathbf{M}^\circ\mathbf{Y}) &= \mathbf{V}(\mathbf{M}^\circ\mathbf{T}\mathbf{Y}) + \mathbf{V}(\mathbf{M}^\circ(\mathbf{I} - \mathbf{T})\mathbf{Y}) \\ &= \mathbf{V}(\mathbf{M}\mathbf{T}\mathbf{Y}) + \mathbf{V}(\mathbf{M}^\circ(\mathbf{I} - \mathbf{T})\mathbf{Y}), \end{aligned}$$

logo  $\mathbf{V}(\mathbf{M}^\circ\mathbf{Y}) - \mathbf{V}(\mathbf{M}\mathbf{T}\mathbf{Y})$  será semi-definida positiva, uma vez que matrizes de covariância são semi-definidas positivas, o que estabelece a tese. □

Estamos em condições de enunciar o seguinte teorema:

**Teorema 2.1.** *Quando  $\mathbf{V}$  e  $\mathbf{T}$  comutam os estimadores de mínimos quadrados dos vectores estimáveis são BLUE.*

*Demonstração.* Sendo  $\Psi = \mathbf{A}\beta = \mathbf{K}\mathbf{X}\beta$  um vector estimável, o respectivo estimador de mínimos quadrados é, como vimos,  $\Psi^\circ = \mathbf{K}\mathbf{T}\mathbf{Y}$  tendo-se pois  $\mathbf{M} = \mathbf{K}\mathbf{T}$ , pelo que  $\mathbf{M}\mathbf{T} = \mathbf{K}\mathbf{T}\mathbf{T} = \mathbf{K}\mathbf{T} = \mathbf{M}$  o que estabelece a tese. □

Observe-se que este teorema foi primeiro estabelecido por Zmysłony (1978) [51] e que representa uma versão mais potente do Teorema de Gauss-Markov<sup>8</sup>.

---

<sup>8</sup>Em honra do matemático alemão Carl Friedrich Gauss (1777–1855) e do matemático russo Andrei Andreyevich Markov (1856–1922).

## 2.3 Estatísticas Suficientes e Completas

Nesta secção iremos apresentar algumas definições e resultados que serão utilizados no desenvolvimento deste trabalho, em particular aqueles que dizem respeito às estatísticas suficientes e completas. Para um estudo mais detalhado poderá consultar-se, por exemplo, Mood (1974)[34].

Considerem-se todas as amostras de dimensão  $n$ , representadas por  $\mathbf{x}^n$ . O espaço constituído por essas amostras será  $\mathcal{A}_n$ . Estas amostras, são igualmente realizações do vector aleatório  $\mathbf{X}^n$ , verificando-se que  $pr(\mathbf{X}^n \in \mathcal{A}_n) = 1$ .

Podemos verificar que as probabilidades associadas a  $\mathbf{X}^n$  podem ser especificadas a partir da respectiva distribuição  $F(x^n; \xi^s) = P(X^n \leq x^n; \xi^s)$ , e o parâmetro vectorial  $\xi^s$  varia num espaço de parâmetros  $\Omega$ , ver Williams (1997) [49].

Representemos por  $f(x^n; \xi^n)$  a função densidade [função probabilidade] para  $\mathbf{X}^n$  no caso contínuo [discreto]. Consideremos ainda uma partição de  $\mathcal{A}_n$ , uma família de conjuntos  $D_i$ ,  $i = 1, \dots$ , disjuntos dois a dois, para os quais se verifica  $\bigcup_i D_i = \mathcal{A}_n$ . Podemos pois definir distribuições condicionadas sob a forma  $F(x^n | D_i; \xi^s) = P(X^n \leq x^n | X^n \in D_i; \xi^s)$ .

Com base na função densidade [probabilidade] e na partição sobre  $\mathcal{A}_n$ , podemos definir

$$F(x^n; \xi^s) = \sum_i P(X^n \in D_i; \xi^s) F(x^n | D_i; \xi^s).$$

Uma partição diz-se suficiente quando as distribuições condicionadas não dependem de  $\xi^s$ , i.e.,

$$F(x^n; \xi^s) = \sum_i P(X^n \in D_i; \xi^s) F(x^n | D_i).$$

Uma estatística é qualquer função escalar ou vectorial da amostra. As imagens inversas da estatística constituem uma partição de  $\mathcal{A}_n$ . A estatística diz-se suficiente quando quando gera uma partição suficiente. Estamos em condições de formalmente estabelecer a seguinte definição,

**Definição 2.4 (Estatística Suficiente).** *Uma estatística  $T$  é suficiente, se e só se*

a distribuição  $\mathbf{Y}$  condicionada pelo valor da estatística não depender do parâmetro  $\theta$ , i.e., se  $F_{\mathbf{Y}}(\cdot|T = t)$  não depende de  $\theta$ .

Esta definição é particularmente importante quando se pretende mostrar que uma dada estatística *não* é suficiente. Para o efeito, caso se pretenda mostrar que uma determinada estatística  $T$  não é suficiente, basta encontrar uma outra estatística  $T'$  para a qual a distribuição condicionada de  $T'$  dado  $T$  depender de  $\theta$ .

São vários os exemplos onde não existe uma única estatística suficiente, no entanto existirá sempre estatísticas suficientes conjuntas que se definem como,

**Definição 2.5 (Estatísticas Conjuntamente Suficientes).** *Seja  $x_1, \dots, x_n$  uma amostra aleatória da densidade  $f(\cdot; \theta)$ . As estatísticas  $S_1, \dots, S_r$  são definidas como conjuntamente suficientes se e só se a distribuição conjunta de  $X_1, \dots, X_n$  dado  $S_1 = s_1, \dots, s_r = s_r$  não depende de  $\theta$ .*

Estamos em condições de enunciar o seguinte teorema.

**Teorema 2.2.** *Seja  $Y$  uma variável aleatória cuja função densidade, ou função probabilidade, depende de um parâmetro  $\theta$ . As estatísticas  $T_1, \dots, T_r$  são conjuntamente suficientes se, e só se,*

$$f(y|\theta) = g(t_1, \dots, t_r|\theta)h(y)$$

onde a função  $h$  é não negativa e independente de  $\theta$ , e a função  $g$  depende de  $y$  apenas através das estatísticas  $T_1, \dots, T_r$ .

Do exposto ressalta a possibilidade da existência de vários estimadores centrados (apenas focaremos estes por razões óbvias) e função de estatísticas suficientes. Coloca-se então o problema da escolha de estimadores para o parâmetro de interesse. Com este objectivo em mente, iremos introduzir alguns conceitos que interessa reter. Por serem por demais conhecidos e as suas demonstrações poderem ser encontradas na mais diversa bibliografia, apenas iremos demonstrar aqueles cujo interesse seja pertinente.

**Teorema 2.3 (Teorema da Factorização).** *Seja  $X_1, \dots, X_n$  uma amostra aleatória de tamanho  $n$  da densidade  $f(\cdot; \theta)$ , onde o parâmetro  $\theta$  pode ser um vector. A estatística  $S = s(X_1, \dots, X_n)$  é suficiente se e só se a densidade conjunta de  $X_1, \dots, X_n$ , dada por  $\prod_{i=1}^n f(x_i; \theta)$ , se factoriza como*

$$\begin{aligned} f_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta) &= g(s(x_1, \dots, x_n); \theta) h(x_1, \dots, x_n) \\ &= g(s; \theta) h(x_1, \dots, x_n) \end{aligned}$$

onde a função  $h(x_1, \dots, x_n)$  é não negativa e não depende do parâmetro  $\theta$  e a função  $g(s(x_1, \dots, x_n); \theta)$  é não negativa e depende de  $x_1, \dots, x_n$  apenas pela função  $s(\cdot, \dots, \cdot)$ .

Podemos ainda enunciar o anterior teorema distinguindo estatísticas suficientes (simples) e para o caso de estatísticas conjuntamente suficientes. Ver, por exemplo, Mood (1974) [34].

**Teorema 2.4 (Teorema da Factorização para Estatísticas Simplesmente Suficientes).** *Considere-se  $X_1, \dots, X_n$  uma amostra aleatória, de dimensão  $n$ , de uma densidade  $f(\cdot; \theta)$ , onde o parâmetro  $\theta$  poderá ser um vector. Uma estatística  $S = s(X_1, \dots, X_n)$  é suficiente (simplesmente) se e só se a densidade conjunta de  $X_1, \dots, X_n$  dada por  $\prod_{i=1}^n f(x_i; \theta)$ , poder ser factorizada como*

$$\begin{aligned} f_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta) &= g(s(x_1, \dots, x_n); \theta) h(x_1, \dots, x_n) \\ &= g(S; \theta) h(x_1, \dots, x_n) \end{aligned}$$

onde a função  $h(x_1, \dots, x_n)$  é não negativa e não depende do parâmetro  $\theta$  e a função  $g(s(x_1, \dots, x_n); \theta)$  é não negativa e depende de  $x_1, \dots, x_n$  apenas através da função  $s(\cdot, \dots, \cdot)$ .

**Teorema 2.5 (Teorema da Factorização para Estatísticas Conjuntamente Suficientes).** *Considere-se  $X_1, \dots, X_n$  uma amostra aleatória, de dimensão  $n$ , da*

densidade  $f(\cdot; \theta)$ , onde o parâmetro  $\theta$  pode ser um vector. Um conjunto de estatísticas  $S_1 = s_1(X_1, \dots, X_n), \dots, S_r = s_r(X_1, \dots, X_n)$  é conjuntamente suficiente se e só se a densidade conjunta de  $X_1, \dots, X_n$  pode ser factorizada como

$$\begin{aligned} f_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta) &= g(s_1(x_1, \dots, x_n), \dots, s_r(x_1, \dots, x_n); \theta) h(x_1, \dots, x_n) \\ &= g(s_1, \dots, s_r; \theta) h(x_1, \dots, x_n) \end{aligned}$$

onde a função  $h(x_1, \dots, x_n)$  é não negativa e não depende do parâmetro  $\theta$  e a função  $g(s_1, \dots, s_r; \theta)$  é não negativa, e depende de  $x_1, \dots, x_n$  apenas através das funções  $s_1(\cdot, \dots, \cdot), \dots, s_r(\cdot, \dots, \cdot)$ .

Uma outra definição importante que iremos usar amiúde é a definição de *estimador centrado com variância mínima*.

**Definição 2.6 (UMVUE).** Dado  $Y$  com função densidade  $f(\cdot|\theta)$ , um estimador  $T$  para o parâmetro  $\theta$  diz-se *Estimador Centrado com Variância Uniformemente Mínima (Uniformly Minimum Variance Unbiased Estimator – UMVUE)* de  $\theta$  se

- $T$  é centrado;
- $V(T) \leq V(T^*)$ , para todo o estimador  $T^*$  centrado de  $\theta$ .

Interessa-nos também introduzir o conceito de estatística ancilar.

**Definição 2.7. Estatística Ancilar** Uma estatística  $S(\mathbf{X})$  cuja distribuição não depende do parâmetro  $\theta$  é denominada de *estatística ancilar para o parâmetro*.

Só por si, uma estatística ancilar não contém informação sobre  $\theta$ . Uma estatística ancilar é uma observação sobre uma variável aleatória cuja distribuição é fixa e conhecida, não relacionada com  $\theta$ . No entanto, uma estatística ancilar, quando usada em conjunção com outras estatísticas, por vezes contém informação importante para inferência sobre  $\theta$ .

Uma estatística  $T$  é muito útil na redução de dados se nenhuma função não constante de  $T$  é ancilar ou ancilar de primeira ordem, isto é, se  $\mathbb{E}_\theta[f(T)] = c$  para qualquer  $\theta \in \Omega$  implica  $f(t) = c$ . Esta condição é equivalente a  $\mathbb{E}_\theta[f(T)] = 0$ , para

qualquer  $\theta \in \Omega$  implica  $f(t) = 0$ . Uma estatística que satisfaça esta condição diz-se completa.

O que nos leva à seguinte definição,

**Definição 2.8. Estatística Completa.** *Seja  $f(t|\theta)$  uma família de funções densidade [de probabilidade] para a estatística  $T(\mathbf{X})$ . A família das distribuições de probabilidade é denominada de completa se  $\mathbb{E}_\theta[g(t)] = 0$  para qualquer  $\theta$ , implica  $P_\theta(g(T) = 0) = 1$  para qualquer  $\theta$ . Por equivalência,  $T(\mathbf{X})$  é denominada estatística completa.*

De referir que o facto de ser completa é uma propriedade para uma família de distribuições de probabilidade, não de uma distribuição particular.

Considere-se o exemplo: Se  $X$  tem distribuição normal  $N(0; 1)$ , então definindo  $g(x) = x$ , tem-se que  $\mathbb{E}[g(X)] = \mathbb{E}[X] = 0$ . Mas a função  $g(x) = x$  satisfaz  $P(g(X) = 0) = P(X = 0) = 0$  e não igual a 1. Contudo, esta é uma distribuição particular, e não uma família de distribuições. Se  $X$  tiver distribuição normal  $N(\theta; 1)$  com  $-\infty < \theta < +\infty$ , vê-se que nenhuma função de  $x$ , excepto uma, ou seja 0 com probabilidade 1 para qualquer  $\theta$ , satisfaz  $\mathbb{E}_\theta(g(X)) = 0$  para todo o  $\theta$ . Então, a família de distribuições com  $N(\theta; 1)$  e  $-\infty < \theta < +\infty$ , é completa.

Apesar de podermos comparar dois ou mais estimadores centrados e de entre estes definir qual o melhor, leia-se o de menor variância, nada nos garante que efectivamente este é o de menor variância de entre todos os estimadores possíveis em idênticas circunstâncias.

Vejamos algumas condições de regularidade dos estimadores:

1.  $\frac{\partial}{\partial \theta} \log f(x; \theta)$  existe para qualquer  $x$  e todo o  $\theta$ ;
2.  $\frac{\partial}{\partial \theta} \int \dots \int \prod_{i=1}^n f(x_i; \theta) dx_1, \dots, dx_n = \int \dots \int \frac{\partial}{\partial \theta} \prod_{i=1}^n f(x_i; \theta) dx_1, \dots, dx_n$ ;
3.  $\frac{\partial}{\partial \theta} \int \dots \int t(x_1, \dots, x_n) \prod_{i=1}^n f(x_i; \theta) dx_1, \dots, dx_n =$

$$= \int \dots \int t(x_1, \dots, x_n) \frac{\partial}{\partial \theta} \prod_{i=1}^n f(x_i; \theta) dx_1, \dots, dx_n;$$

$$4. 0 < E_{\theta} \left[ \left[ \frac{\partial}{\partial \theta} \log f(X; \theta) \right]^2 \right] < \infty \text{ para qualquer } \theta \in \Theta.$$

Considerando que poderão existir vários estimadores centrados não enviesados, interessa estudar aquele que terá variância mínima. Para esse efeito, podemos estabelecer o seguinte teorema.

**Teorema 2.6. Desigualdade de Cramér-Rao.**<sup>9</sup> *Sob as anteriores condições de regularidade, e para o caso unidimensional, tem-se*

$$V_{\theta}(T) \geq \frac{[\tau'(\theta)]^2}{n E_{\theta} \left[ \left[ \frac{\partial}{\partial \theta} \log f(X; \theta) \right]^2 \right]}$$

onde  $T = t(X_1, \dots, X_n)$  é um estimador não enviesado de  $\tau(\theta)$ .

Esta desigualdade apenas existe se existir uma função  $K(\theta, n)$  tal que

$$\sum_{i=1}^n \frac{\partial}{\partial \theta} \log f(x_i; \theta) = K(\theta, n)[t(x_1, \dots, x_n) - \tau(\theta)].$$

A desigualdade de Cramér-Rao, tem duas grandes e óbvias utilidades. Se por um lado a variância de um estimador utilizado se encontra próximo deste limite inferior, há garantia que este estimador é um bom estimador. Por outro lado, se a variância coincidir com o limite inferior, então poder-se-á afirmar que o estimador utilizado é UMVUE.

Uma vez que iremos trabalhar com estatísticas suficientes e completas, é ainda de interesse o seguinte teorema:

---

<sup>9</sup>Expressão que contém o Limite Inferior de Cramér-Rao [Cramér-Rao Lower Bound]. Em honra de Harald Cramér (1893–1985), matemático, actuário e estatístico sueco; e Calyampudi Radhakrishna Rao (1920– ), estatístico indiano e uma lenda viva na área da estatística.

**Teorema 2.7 (Teorema de Basu).** *Se  $T$  é uma estatística suficiente e completa para a família  $\mathcal{P} \{P_\theta, \theta \in \Omega\}$ , então qualquer estatística ancilar  $V$  é independente de  $T$ .*

*Demonstração.* Se  $V$  é ancilar, a probabilidade  $P_A = P(V \in A)$  é independente de  $\theta$  para qualquer  $A$ . Tomando  $\eta_A(t) = P(V \in A | T = t)$ , então,  $\mathbb{E}_\theta[\eta_A(T)] = p_A$  e, de acordo com a definição de estatística completa,

$$\eta_A(t) = p_A,$$

o que estabelece a independência de  $V$  e  $T$ . □

No desenvolvimento deste trabalho, uma das condições que iremos colocar é a existência de normalidade. Uma vez que a distribuição normal pertence a uma maior família, a das exponenciais, o seguinte teorema terá particular interesse.

**Teorema 2.8 (Teorema da Factorização).** *Considerem-se  $X_1, \dots, X_n$  observações independentes e identicamente distribuídas de uma família exponencial com função densidade [de probabilidade] dada por*

$$f(x|\theta) = h(x)c(\theta)e^{w(\theta)t(x)},$$

*e onde o espaço paramétrico contém abertos.*

*Então a estatística*

$$T(\mathbf{X}) = \sum_{i=1}^n t(x_i)$$

*é completa.*

Um resultado importante adentro deste capítulo, prende-se com a denominada desigualdade de Jensen.

**Definição 2.9. Desigualdade de Jensen.** *Dada uma variável aleatória  $X$  e a função convexa  $g(\cdot)$ , teremos*

$$\mathbb{E}[g(X)] \geq g(\mathbb{E}[X]).$$

*Demonstração.* Admitamos que  $r(X) = a + bX$ , com  $a, b \in \mathbb{R}$ , a recta tangente ao gráfico de  $g(\cdot)$  no ponto  $\mathbb{E}[X]$ . Uma vez que  $g(x)$  é convexa,  $g(x) \geq a + bx$ , e como tal,  $\mathbb{E}[g(X)] \geq \mathbb{E}[a + bX] = a + b\mathbb{E}[X] = r(\mathbb{E}[X]) = g(\mathbb{E}[X])$ , uma vez que  $a + bx$  é a recta tangente.  $\square$

A demonstração refere-se ao caso em que  $g(\cdot)$  é derivável. Para o caso geral poder-se-á consultar, por exemplo, Frazer (1957) [18].

Nas secções seguintes iremos utilizar igualmente um conjunto de teoremas, cujo enunciado aqui deixamos para que não haja dúvidas sobre a sua aplicação e entendimento. Nesse sentido iremos enunciar o seguinte teorema,

**Teorema 2.9. Rao-Blackwell.** *Seja  $Y$  com função densidade  $f$  e  $T_1, \dots, T_r$  uma família de estatísticas conjuntamente suficientes. Seja  $T$  um estimador centrado de  $\theta$ , então, com  $T' = \mathbb{E}[T|T_1, \dots, T_r]$ ,*

1.  $T'$  é uma função das estatísticas  $T_1, \dots, T_r$ ;
2.  $T'$  é um estimador centrado de  $\theta$ ;
3.  $V[T'] \leq V[T]$  para qualquer  $\theta$ , e existir desigualdade para algum  $\theta$ , a não ser que  $P[T' = T] = 1$ .

De interesse é igualmente o

**Teorema 2.10. Blackwell-Lehmann-Scheffé.** *Seja  $Y$  uma variável aleatória com função densidade de probabilidade  $f$ . Se  $T_1, \dots, T_r$  são estatísticas suficientes, completas e  $g(T_1, \dots, T_r)$  é um estimador centrado desse parâmetro, então  $g(t_1, \dots, T_r)$  será UMVUE.*

**Teorema 2.11. Lehmann-Scheffé.** *Seja  $T$  uma estatística, tal que,  $\mathbb{E}[T] = \theta$  e  $V[T] < \infty$ , onde  $\theta$  é um parâmetro real. Então, uma condição necessária e suficiente para  $T$  ser UMVUE para  $\theta$  é:  $cov(t, T) = 0$  para  $\forall t$ , tal que  $\mathbb{E}[t] = 0$  e  $V[t] < \infty$ .*

## 2.4 Variáveis Pivot Indutoras

### 2.4.1 Medidas Induzidas

Variáveis pivot são função das estatísticas e dos parâmetros com distribuições conhecidas. Por exemplo, se  $S \sim \gamma\chi_g^2$ , isto é, se  $S$  é o produto de  $\gamma$  por um qui-quadrado central com  $g$  graus de liberdade, então, com

$$Z = \frac{S}{\gamma}$$

teremos  $Z \sim \chi_g^2$  sendo por isso uma variável pivot.

Admitamos que o espaço paramétrico da distribuição de  $\mathbf{Y}^n$  pertence à  $\sigma$ -álgebra  $\mathcal{B}_r$  dos borelianos de  $\mathbb{R}^n$ .

**Definição 2.10.** *A variável pivot  $\mathbf{Z}^r = \mathbf{g}^r(\mathbf{Y}^n, \boldsymbol{\theta}^r)$  será, ver Covas (2003) [5], indutora, se a função  $L(\boldsymbol{\theta}|\mathbf{y}^n) = g^r(\mathbf{y}^n, \boldsymbol{\theta}^r)$  tiver função inversa  $h(\mathbf{z}^r|\mathbf{y}^n)$  mensurável  $\mathcal{B}_r$ .*

Com efeito, sendo  $P^\circ$  a medida de probabilidade associada à distribuição de  $\mathbf{Y}$ , teremos as medidas de probabilidade induzidas em  $\mathcal{B}_r$ , dadas por

$$P_{\mathbf{y}^n}(C) = P^\circ(l(C \cap \Theta) | \mathbf{y}^n)$$

onde  $\mathbf{y}^n$  é o valor tomado por  $\mathbf{Y}$ .

Note-se que para qualquer  $\mathbf{y}^n$

$$P_{\mathbf{y}^n}(\Theta) = 1.$$

Interessa-nos considerar um caso particular, onde as componentes  $Z_1, \dots, Z_r$  da variável pivot induzida  $\mathbf{Z}^r$  são independentes e dadas por

$$Z_i = g_i(\mathbf{y}^n, \theta_i) \quad i = 1, \dots, r$$

com  $\theta_1, \dots, \theta_r$  as componentes de  $\Theta^r$ . Se para além disso tivermos

$$\Theta = \times_{i=1}^r \Theta_i,$$

com  $\Theta_i \in \mathcal{B}_i$ ,  $i = 1, \dots, r$  e  $\times_{i=1}^r$  o produto cartesiano, e as funções

$$l_i(\theta|\mathbf{y}^n) = g_i(\mathbf{y}^n, \theta_i) \quad i = 1, \dots, r$$

tiverem inversas mensuráveis  $h_i(z_i|\mathbf{y}^n) \in \mathcal{B}$ ,  $i = 1, \dots, r$ , podemos induzir em  $(\mathbb{R}, \mathcal{B})$  as medidas de probabilidade

$$P_{\mathbf{y}^n, i}(C) = P_i^\circ(l_i(C \cap \Theta_i)|\mathbf{y}^n) \quad i = 1, \dots, r$$

onde  $P_i^\circ$ ,  $i = 1, \dots, r$  é a medida de probabilidade associada à distribuição de  $Z_i$ ,  $i = 1, \dots, r$ . Uma vez que as componentes são independentes, tomando  $C_1, \dots, C_r \in \mathcal{B}$ , teremos, ver Ferreira (2006) [11],

$$P^\circ\left(\times_{i=1}^r l_i(C_i \cap \Theta_i)|\mathbf{y}^n\right) = \prod_{i=1}^r P_i^\circ(l_i(C_i \cap \Theta_i)|\mathbf{y}^n)$$

pelo que, com  $\bar{P}_{\mathbf{y}^n}$  a medida produto das  $P_{\mathbf{y}^n, i}$ ,  $i = 1, \dots, r$ , teremos, ver Williams (1997) [49], a medida produto dada por,

$$\begin{aligned} \bar{P}_{\mathbf{y}^n}\left(\times_{i=1}^r C_i\right) &= \prod_{i=1}^r P_{\mathbf{y}^n, i}(C_i) = \prod_{i=1}^r P_i^\circ(l_i(C_i \cap \Theta_i)|\mathbf{y}^n) = \\ &= P^\circ\left(\times_{i=1}^r l_i(C_i \cap \Theta_i)|\mathbf{y}^n\right) = \\ &= P\left(\mathcal{I}^r\left(\left(\times_{i=1}^r C_i\right) \cap \Theta\right)\right), \end{aligned}$$

uma vez que

$$\begin{aligned} \times_{i=1}^r (C_i \cap \Theta_i) &= \left(\times_{i=1}^r C_i\right) \cap \left(\times_{i=1}^r \Theta_i\right) = \\ &= \left(\times_{i=1}^r C_i\right) \cap \Theta. \end{aligned}$$

Assim a medida produto  $\bar{P}_{\mathbf{y}^n}$  das medidas induzidas pelas componentes é idêntica a  $P_{\mathbf{y}^n}$ .

Em particular, admitindo que  $Z \sim \chi_g^2$ , com  $x_{g,p}$  o quantil de  $\chi_g^2$  para a probabilidade  $p$ , teremos, caso  $p_1 < p_2$ ,

$$pr \left( x_{g,p_1} \leq \frac{S}{\gamma} \leq x_{g,p_2} \right) = p_2 - p_1$$

de forma a que

$$pr \left( \frac{S}{x_{g,p_2}} \leq \gamma \leq \frac{S}{x_{g,p_1}} \right) = p_2 - p_1,$$

pelo que é possível utilizar  $Z \sim \chi_g^2$  para induzir uma medida de probabilidade em  $(\mathbb{R}, \mathcal{B})$  para  $\gamma$ .

Consideremos agora que  $S_l \sim \gamma_l \chi_{g_l}^2$ ,  $l = 1, \dots, t$ , o que nos permite obter a variável pivot indutora  $Z_l = \frac{S_l}{\gamma_l} \sim \chi_{g_l}^2$ ,  $l = 1, \dots, t$ . Se os  $S_l$ ,  $l = 1, \dots, t$  forem independentes, então  $Z_l$ ,  $l = 1, \dots, t$ , também serão independentes.

A partir destas medidas de probabilidade, podemos obter a correspondente medida de probabilidade produto.

Quando  $S_l = s_l$ ,  $l = 1, \dots, t$ , é possível gerar amostras com distribuição  $F_{\mathbf{Z}}$  associadas ao produto das medidas de probabilidade. Se, com  $X_i = (X_{i,1}, \dots, X_{i,t})$ , teremos  $X_{i,l} \sim \chi_{g_l}^2$ ,  $l = 1, \dots, t$ ,  $i = 1, \dots, N$ , a amostra constituída pelos

$$\mathbf{Z}_i = (Z_{i,1}, \dots, Z_{i,t}) = \left( \frac{s_1}{X_{i,1}}, \dots, \frac{s_t}{X_{i,t}} \right) \quad i = 1, \dots, N$$

terá distribuição  $F_{\mathbf{Z}}$ .

Consideremos agora o parâmetro

$$\boldsymbol{\theta} = \mathbf{d}^\top \boldsymbol{\gamma}.$$

A amostra constituída pelos  $U_i = \mathbf{d}^\top \mathbf{Z}_i$ ,  $i = 1, \dots, N$  estará associada à distribuição  $F_{\mathbf{d}^\top \mathbf{Z}}$  do produto interno  $\mathbf{d}^\top \mathbf{Z}$ . Com  $F_{N, \mathbf{d}^\top \mathbf{Z}}$ , a distribuição empírica da amostra, atendendo ao teorema de Glivenko-Cantelli<sup>10</sup> sabemos que

$$\sup \left\{ |F_{N, \mathbf{d}^\top \mathbf{Z}}(u) - F_{\mathbf{d}^\top \mathbf{Z}}(u)| \right\} \xrightarrow[N \rightarrow \infty]{q.c.} 0, \quad (2.2)$$

onde  $\xrightarrow[N \rightarrow \infty]{q.c.} 0$  representa convergência quase certa.

Estabeleçamos agora a seguinte proposição

---

<sup>10</sup>Ver a este respeito, por exemplo Loève (1961) [25].

**Proposição 2.7.** Com  $z_{N,q}$  e  $z_q$  os quantis para a probabilidade  $q$  de  $F_{N,\mathbf{d}^\top \mathbf{Z}}$  e  $F_{\mathbf{d}^\top \mathbf{Z}}$ , para qualquer que seja  $\varepsilon > 0$ , temos

$$\sup \{ |z_{N,q} - z_q|; \varepsilon < q < 1 - \varepsilon \} \xrightarrow[N \rightarrow \infty]{q.c.} 0.$$

*Demonstração.* A densidade de  $\chi_g^2$  é dada por  $f(x|g) = \frac{1}{2\Gamma(\frac{g}{2})} \left(\frac{x}{2}\right)^{\frac{g}{2}-1} e^{-\frac{x}{2}}$ ,  $x > 0$ , pelo que a densidade de  $Z = \frac{s}{\chi_g^2}$ ,  $s \neq 0$ , será (ver Ferreira, 2006, [9]),

$$\dot{f}(z|g, d) = \frac{s}{z^2} f\left(\frac{s}{z}|g\right) = \frac{s^{\frac{g}{2}}}{2^{\frac{g}{2}}\Gamma(\frac{g}{2})} \frac{e^{-\frac{s}{2z}}}{z^{\frac{g}{2}+1}}$$

com  $z > 0$  se  $d > 0$  e  $z < 0$  se  $d < 0$ . Então  $\dot{f}(z|g, d)$  é contínua e positiva no interior do suporte. Considerando que a densidade  $f_{\mathbf{d}^\top \mathbf{Z}}$  correspondente a  $F_{\mathbf{d}^\top \mathbf{Z}}$  é obtida através da convolução de densidades  $\dot{f}(|g_l, d_l)$ ,  $l = 1, \dots, t$ , esta será contínua e não nula no interior do suporte. De acordo com o teorema de Weierstrass<sup>11</sup>,  $f_{\mathbf{d}^\top \mathbf{Z}}$  terá um mínimo  $a > 0$  no intervalo  $[z_{\frac{\varepsilon}{2}}; z_{1-\frac{\varepsilon}{2}}]$ . Quando  $\sup |F_{N,\mathbf{d}^\top \mathbf{Z}}(z) - F_{\mathbf{d}^\top \mathbf{Z}}(z)| < \frac{\delta}{4}$ , temos  $q' = F_{\mathbf{d}^\top \mathbf{Z}}(z_{N,q}) \geq F_{N,\mathbf{d}^\top \mathbf{Z}}(z_{N,q}) - \frac{\delta}{4} = q - \frac{\delta}{4}$ , de forma que  $z_{q-\frac{\delta}{4}} \leq z_{q'} = z_{N,q}$ , e observa-se que da mesma forma  $z_{q+\frac{\delta}{4}} \geq z_{N,q}$ , pelo que  $z_{q-\frac{\delta}{4}} \leq z_{N,q} \leq z_{q+\frac{\delta}{4}}$ , e dado que  $z_{q-\frac{\delta}{4}} \leq z_q \leq z_{q+\frac{\delta}{4}}$ , obtem-se  $|z_{N,q} - z_q| \leq z_{q+\frac{\delta}{4}} - z_{q-\frac{\delta}{4}}$ . Então, se  $\sup |F_{N,\mathbf{d}^\top \mathbf{Z}}(z) - F_{\mathbf{d}^\top \mathbf{Z}}(z)| < \frac{\delta}{4}$  e  $\frac{\varepsilon}{2} \leq q - \frac{\delta}{4} < q + \frac{\delta}{4} \leq 1 - \frac{\varepsilon}{2}$ , obtemos  $|z_{N,q} - z_q| < a \frac{\varepsilon}{2}$ .

De acordo com a expressão (2.2), observa-se que, qualquer que seja  $q < 1$  e  $\varepsilon > 0$ , existe  $\dot{N}$  tal que  $\text{pr} \left[ \bigcap_{N > \dot{N}} (\sup |z_{N,q} - z_q|; \varepsilon < q < 1 - \varepsilon) \right]$ , o que estabelece a tese.  $\square$

Podemos utilizar os  $z_{N,q}$  para estimar os  $z_q$  e conseqüentemente obter os extremos dos intervalos de confiança  $] - \infty; z_{N,1-q}]$ ,  $[z_{N,\frac{q}{2}}; z_{N,1-\frac{q}{2}}]$  e  $[z_{N,q}; +\infty[$  para  $\theta$ , com significância  $1 - q$ .

<sup>11</sup>Karl Wilhelm Theodor Weierstraß (1815–1897). Matemático alemão e pioneiro da moderna análise matemática, tendo lançado as bases da teoria das funções analíticas.

Os intervalos assim definidos podem ser utilizados para, por dualidade, construir testes de hipóteses para

$$H_0 : \theta = \theta_0$$

contra as hipóteses alternativas

$$H_1 : \theta < \theta_0 \quad ; \quad H_1 : \theta \neq \theta_0 \quad ; \quad H_1 : \theta > \theta_0.$$

Em qualquer um dos casos considerados, rejeita-se  $H_0$  caso o intervalo de confiança correspondente não contenha  $\theta_0$ . Para um teste a nível de significância  $q$ , utilizaremos um intervalo de confiança com um nível de  $1 - q$ .

Estes testes são semelhantes aos testes generalizados apresentados por Weerahandi (1996) [48]. No entanto, julgamos que a variante aqui apresentada é de mais fácil aplicação.

## 2.4.2 Densidades

As estatísticas suficientes que usaremos para realizar inferência para as componentes de variância, serão as componentes independentes escritas como  $S_i \sim \gamma_i \chi_{g_i}^2$ ,  $i = 1, \dots, r$ , das quais podemos obter as variáveis pivot independentes

$$Z_i = \frac{S_i}{\gamma_i} \sim \chi_{g_i}^2 \quad i = 1, \dots, r$$

e como as funções inversas,

$$h_i(z_i, s_i) = \frac{s_i}{z_i} \quad i = 1, \dots, r$$

são mensuráveis em  $\mathcal{B}$ , as  $Z_i$ ,  $i = 1, \dots, r$ , são indutoras independentes (ver Ferreira (2006) [11]).

Partindo da densidade do  $\chi_g^2$ ,

$$f(x|g) = \frac{1}{2\Gamma(\frac{g}{2})} \left(\frac{x}{2}\right)^{\frac{g}{2}-1} e^{-\frac{x}{2}} \quad x > 0,$$

fazendo a transformação  $x = \frac{s}{z}$ , obtem-se a densidade

$$\begin{aligned} f_0(z|g) &= \frac{1}{2\Gamma(\frac{g}{2})} \left(\frac{2}{2z}\right)^{\frac{g}{2}-1} e^{-\frac{s}{2z}} \frac{s}{z^2} = \\ &= k_g(s) z^{-\frac{g+2}{2}} e^{-\frac{s}{2z}} \quad z > 0 \end{aligned}$$

onde  $k_g(s) = \frac{s^{\frac{g}{2}}}{2^{\frac{g}{2}}\Gamma(\frac{g}{2})}$ . Ora,  $z_0 = \frac{s}{g+2} > 0$  será o único extremo local, a moda, de  $f_0(z|g)$  (ver Ferreira (2006) [11]).

Os pontos de inflexão de  $f_0(z|g)$  são dados por

$$z_i = \frac{s \left[ (g+4) + (-1)^i \sqrt{2(g+4)} \right]}{(g+2)(g+4)} \quad i = 1, 2.$$

A função densidade  $f_0(z|g)$  chama-se gama invesa, tendo sido estudada por Witkovský (2001) [50].

A distribuição correspondente a  $f_0(z|g)$  é dada por,

$$F_0(z|g) = 1 - \frac{1}{\Gamma(\frac{g}{2})} \int_0^{\frac{s}{2z}} u^{\frac{g}{2}-1} e^{-u} du$$

uma vez que

$$F(z|g) = k_g(s) \int_0^z x^{-\frac{g+2}{2}} e^{-\frac{s}{2x}} dx.$$

Este resultado pode encontrar-se em Ferreira (2006) [11].

Usando a transformação,  $x = \frac{s}{2y}$ , podemos obter o momento de ordem  $l$  relativamente à origem da variável pivot, vindo, com  $l < \frac{g}{2}$ ,

$$\begin{aligned} \mu'_l(\cdot|g) &= k_g(s) \int_0^{+\infty} x^{-\frac{g+2}{2}+l} e^{-\frac{s}{2x}} dx = \\ &= k_g(s) \int_0^{+\infty} \left(\frac{s}{2y}\right)^{-\frac{g+2}{2}+l} e^{-y} \frac{s}{2y^2} dy = \\ &= k_g(s) \left(\frac{s}{2}\right)^{l-\frac{g}{2}} \int_0^{+\infty} y^{\frac{g}{2}-l-1} e^{-y} dy = \\ &= \frac{s^l \Gamma\left(\frac{g}{2}-l\right)}{s^l \Gamma\left(\frac{g}{2}\right)}, \end{aligned}$$

em particular,

$$\begin{cases} \mu(\cdot|g) = \frac{s}{2\left(\frac{g}{2}-1\right)} = \frac{s}{g-2}, & g > 2 \\ \mu'_2(\cdot|g) = \frac{s^2}{4\left(\frac{g}{2}-1\right)\left(\frac{g}{2}-2\right)} = \frac{s^2}{(g-2)(g-4)}, & g > 4. \end{cases}$$

Interessa-nos agora estabelecer o seguinte resultado sobre modas de densidades conjuntas.

**Proposição 2.8.** *Se as variáveis  $X_1, \dots, X_m$  são independentes e com funções densidades unimodais  $f_1(x_1), \dots, f_m(x_m)$ , e modas  $\gamma_1, \dots, \gamma_m$ , as suas densidades conjuntas  $f(\mathbf{x}^m)$  serão unimodais com moda  $\gamma^m$ .*

*Demonstração.* Temos para qualquer  $\mathbf{x}^m$

$$f(\mathbf{x}^m) = \prod_{j=1}^m f_j(x_j) \leq \prod_{j=1}^m f_j(\gamma_j) = f(\gamma^m)$$

o que estabelece a tese. □

Da proposição anterior, resulta que as densidades conjuntas de variáveis inversas de gama independentes são unimodais.

Tomemos

$$m(z|g) = \ln [f_0(z|g)]$$

vindo,

$$\begin{cases} m'(z|g) = \frac{s - (g+2)z}{2z^2} & z > 0 \\ m''(z|g) = \frac{(g+2)z - 2s}{2z^3} & z > 0. \end{cases}$$

Dado o vector  $\mathbf{Z}$ , cujos componentes são as variáveis pivot indutoras independentes  $Z_i$ ,  $i = 1, \dots, r$ , com graus de liberdade  $g_1, \dots, g_r$ , ponhamos

$$m(\mathbf{z}^r|\mathbf{g}^r) = \ln \left[ \prod_{j=1}^r f(z_j|g_j) \right] = \sum_{j=1}^r m(z_j|g_j)$$

vindo,

$$\left\{ \begin{array}{l} \frac{\partial m(\mathbf{z}^r | \mathbf{g}^r)}{\partial z_j} = m'(z_j | g_j) = \frac{s_j - (g_j + 2)z_j}{2z_j^2} \quad j = 1, \dots, r \\ \frac{\partial^2 m(\mathbf{z}^r | \mathbf{g}^r)}{\partial z_j^2} = m''(z_j | g_j) = \frac{(g_j + 2)z_j - 2s_j}{2z_j^3} \quad j = 1, \dots, r \\ \frac{\partial^2 m(\mathbf{z}_j | \mathbf{g}_j)}{\partial z_j \partial z_l} = 0 \quad j \neq l. \end{array} \right.$$

Podemos pois obter o gradiente, dado por

$$\mathbf{grad}^r (m(\mathbf{z} | \mathbf{g}^r)) = \left[ \frac{s_1 - (g_1 + 2)z_1}{2z_1^2}, \dots, \frac{s_r - (g_r + 2)z_r}{2z_r^2} \right]^\top,$$

e se representarmos por  $D(a_1, \dots, a_r)$ , a matriz diagonal com elementos principais,  $a_1, \dots, a_r$ , teremos a matriz Hessiana<sup>12</sup>,

$$\mathbf{Hess}^{r \times r} (m(\mathbf{z}^r | \mathbf{g}^r)) = D \left( \frac{(g_1 + 2)z_1 - 2s_1}{2z_1^3}, \dots, \frac{(g_r + 2)z_r - 2s_r}{2z_r^3} \right).$$

O único zero do gradiente ( $z_0^r$ ), terá componentes  $z_{0,j} = \frac{s_j}{g_j + 2}$ ,  $j = 1, \dots, r$ . Verifica-se igualmente que caso  $\mathbf{z}_0^r < 2\mathbf{z}_0^r$ , a matriz hessiana será definida negativa e definida positiva caso  $\mathbf{z}_0^r > 2\mathbf{z}_0^r$ . No entanto, julgamos que a variante aqui apresentada é de mais fácil aplicação.

Podemos então concluir que,  $m(\mathbf{z}^r | \mathbf{g}^r)$  será estritamente côncava ou convexa, caso a matriz hessiana seja definida negativa ou positiva respectivamente, ver Bazarra (1992) [1].

Uma vez que

$$\nabla(a) = \{\mathbf{z}^r : m(\mathbf{z}^r | \mathbf{g}^r) > a\} = \{\mathbf{z}^r : f(\mathbf{z}^r | \mathbf{g}^r) > e^a\},$$

quando  $\mathbf{z}^r < 2\mathbf{z}_0^r$ , observa-se que se  $\nabla(a) \subset \prod_{j=1}^r [0; 2z_{0,j}]$ ,  $\nabla(a)$  será um conjunto convexo.

Para  $r = 2$ ,  $f(\mathbf{z}^2 | \mathbf{g}^2)$  terá as curvas de nível representadas na Figura 2.1

<sup>12</sup>Em honra de James Joseph Sylvester (1814–1897)

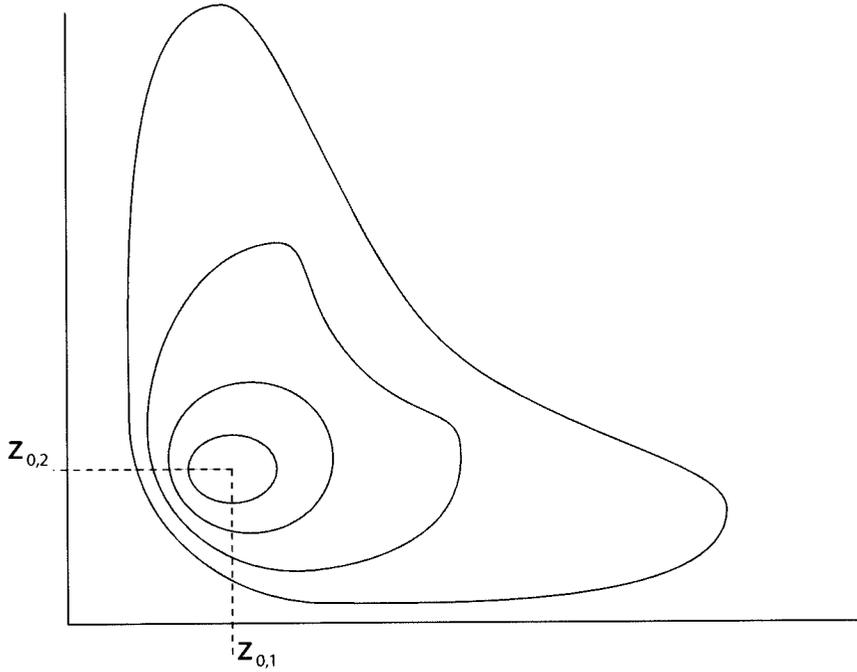


Figura 2.1: Curvas de nível

### 2.4.3 Indução para vectores estimáveis

Esta indução vai-nos permitir obter regiões de confiança e, por dualidade, testar hipóteses para os vectores estimáveis.

Admitamos que  $\mathbf{Z}^n$  tem distribuição normal com vector médio  $\boldsymbol{\lambda}^n$  e matriz de covariância  $\sigma^2 \mathbf{M}$ , ou seja  $\mathbf{Z}^n \sim N(\boldsymbol{\lambda}^n, \sigma^2 \mathbf{M})$ , independente de  $\mathbf{S}$  com componentes  $S_1, \dots, S_t$ , também independentes entre si.

Uma vez que, ver Mexia (1989) [28],  $(\mathbf{Z}^n - \boldsymbol{\lambda}^n)^\top \mathbf{M}^+ (\mathbf{Z}^n - \boldsymbol{\lambda}^n) \sim \sigma^2 \chi_k^2$ , com  $k = \text{car}(\mathbf{M})$ , tem-se,

$$\mathcal{F} = \frac{\chi_k^2}{\sum_j \left( \frac{d_j \gamma_j}{\sigma^2} \right) \chi_{g_j}^2} = \frac{(\mathbf{Z}^n - \boldsymbol{\lambda}^n)^\top \mathbf{M}^+ (\mathbf{Z}^n - \boldsymbol{\lambda}^n)}{\mathbf{d}^\top \mathbf{S}}$$

tem distribuição  $F(z|1, \boldsymbol{\rho}^t, k, \mathbf{g}^t)$ , com  $\boldsymbol{\rho}^t = \left( \frac{d_1 \gamma_1}{\sigma^2}, \dots, \frac{d_t \gamma_t}{\sigma^2} \right)$  e  $\mathbf{g}^t = (g_1, \dots, g_t)$ , uma vez que podemos dividir quer o numerador quer o denominador de  $\mathcal{F}$  por  $\sigma^2$ . Com

$f_{1-q}$  o quantil para esta distribuição para a probabilidade  $1 - q$ , teremos

$$pr \left[ (\boldsymbol{\lambda}^n - \mathbf{Z}^n)^\top \mathbf{W}^+ (\boldsymbol{\lambda}^n - \mathbf{Z}^n) \leq f_{1-q} \mathbf{d}^\top \mathbf{S} \right] = pr(\mathcal{F} \leq f_{1-q}) = 1 - q.$$

Obteremos assim um elipsóide de confiança, possivelmente degenerado, para  $\boldsymbol{\lambda}^n$  com nível limite de  $1 - q$ . Por outro lado, ver Mexia (1989) [28],  $(\boldsymbol{\lambda}^n - \mathbf{Z}^n)^\top \mathbf{W}^+ (\boldsymbol{\lambda}^n - \mathbf{Z}^n) \leq c$  se e só se

$$\bigcap_{\mathbf{a}^n} \left( \left| \mathbf{a}^{n^\top} \boldsymbol{\lambda}^n - \mathbf{a}^{n^\top} \mathbf{Z}^n \right| \leq \sqrt{c (\mathbf{a}^{n^\top} \mathbf{W} \mathbf{a}^n)} \right)$$

onde  $\bigcap_{\mathbf{a}^n}$  indica que são considerados todos os vectores  $\mathbf{a}^n \in \mathbb{R}$ . Isto permite-nos re-escrever o teorema de Scheffé, que pode ser enunciado da seguinte forma:

**Teorema 2.12.**

$$pr \left[ \bigcap_{\mathbf{a}^n} \left( \left| \mathbf{a}^{n^\top} \boldsymbol{\lambda}^n - \mathbf{a}^{n^\top} \mathbf{Z}^n \right| \leq \sqrt{f_{1-q} (\mathbf{a}^{n^\top} \mathbf{W} \mathbf{a}^n) (\mathbf{d}^\top \mathbf{S})} \right) \right] = 1 - q$$

Estes resultados permite-nos testar a hipótese

$$H_0 : \sigma^2 = \sigma_0^2$$

e construir intervalos de confiança simultâneos para  $\mathbf{a}^{n^\top} \boldsymbol{\varepsilon}$ .

No entanto subsiste o problema dos parâmetros perturbadores  $\left( \frac{\gamma_1}{\sigma^2}, \dots, \frac{\gamma_t}{\sigma^2} \right)$ . Este problema poderá ser resolvido com uma abordagem semelhante à já exposta, *i.e.*, gerar amostras constituída pelo  $\mathbf{d}^\top \mathbf{S}_i$ ,  $i = 1, \dots, N$ . Somos conduzidos a usar uma abordagem semelhante para

$$Z_i^+ = X_i \mathbf{d}^\top \mathbf{S}_i \quad i = 1, \dots, N,$$

onde  $X_i \sim \chi_k^2$ ,  $i = 1, \dots, N$  são independentes de  $i = 1, \dots, N$ . Com  $z_{N,1-q}^+$  o quantil empírico da última amostra de probabilidade  $1 - q$ , obtemos

$$pr \left[ (\boldsymbol{\lambda}^n - \mathbf{Z}^n)^\top \mathbf{W}^+ (\boldsymbol{\lambda}^n - \mathbf{Z}^n) \leq z_{N,1-q}^+ \right] \xrightarrow{N \rightarrow \infty} 1 - q.$$

Teremos agora  $(\boldsymbol{\lambda}^n - \mathbf{Z}^n)^\top \mathbf{W}^+ (\boldsymbol{\lambda} - bs\mathbf{Z}^n) \leq z_{N,1-q}^+$ , se e só se

$$\bigcap_{\mathbf{a}^n} \left( \left| \mathbf{a}^{n^\top} \boldsymbol{\lambda}^n - \mathbf{a}^{n^\top} \mathbf{Z}^n \right| \leq \sqrt{z_{N,1-q}^+ (\mathbf{a}^{n^\top} \mathbf{W} \mathbf{a}^n)} \right)$$

tal que

$$pr \left[ \bigcap_{\mathbf{a}^n} \left( \left| \mathbf{a}^{n^\top} \boldsymbol{\lambda}^n - \mathbf{a}^{n^\top} \mathbf{Z}^n \right| \leq \sqrt{z_{N,1-q}^+ (\mathbf{a}^{n^\top} \mathbf{W} \mathbf{a}^n)} \right) \right] \xrightarrow{N \rightarrow \infty} 1 - q.$$

# Parte II

# Modelos



# Capítulo 3

## Modelos Mistos

Os modelos mistos são amplamente conhecidos e utilizados. Na vasta literatura que existe sobre este tema, vários métodos e técnicas de estimação têm sido consideradas para a inferência em modelos mistos, ver por exemplo Khuri *et al* (1998) [23]. A nossa abordagem considerará a utilização de AJC no estudo destes modelos.

### 3.1 Estrutura Algébrica

Consideremos o modelo misto escrito da forma

$$\mathbf{Y} = \sum_{i=1}^m \mathbf{X}_i \boldsymbol{\beta}_i + \sum_{i=m+1}^w \mathbf{X}_i \tilde{\boldsymbol{\beta}}_i + \boldsymbol{\varepsilon}$$

onde as matrizes  $\mathbf{X}_i$ ,  $i = 1, \dots, w$  são conhecidas, os vectores  $\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_m$  são fixos e os  $\tilde{\boldsymbol{\beta}}_{m+1}, \dots, \tilde{\boldsymbol{\beta}}_w$  e  $\boldsymbol{\varepsilon}$  são aleatórios, mutuamente independentes, com vectores médios nulos e matrizes de variância-covariância  $\sigma_{m+1}^2 \mathbf{I}_{c_{w+1}}, \dots, \sigma_w \mathbf{I}_{c_w}$  e  $\sigma^2 \mathbf{I}_n$ , respectivamente.

O vector médio de  $\mathbf{y}$

$$\boldsymbol{\mu} = \sum_{i=1}^m \mathbf{X}_i \boldsymbol{\beta}_i$$

varre o espaço  $\Omega = R(\mathbf{X})$ . De forma equivalente com  $\mathbf{X} = [X_1, \dots, X_m]$  e  $\boldsymbol{\beta} =$

$[\boldsymbol{\beta}_1^\top \dots \boldsymbol{\beta}_m^\top]^\top$ , tem-se

$$\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$$

Observe-se que, com  $\mathbf{M}_i = \mathbf{X}_i\mathbf{X}_i^\top$ ,  $i = 1, \dots, w$ , se tem

$$\mathbf{X}\mathbf{X}^\top = \sum_{i=1}^m \mathbf{X}_i\mathbf{X}_i^\top = \sum_{i=1}^m \mathbf{M}_i$$

A matriz de covariância de  $\mathbf{y}$  será

$$\mathbf{V} = \sum_{i=m+1}^w \sigma_i^2 \mathbf{M}_i + \sigma^2 \mathbf{I}_n.$$

Estamos em condições de enunciar a seguinte proposição

**Proposição 3.1.** *Caso as matrizes  $\mathbf{M}_{m+1}, \dots, \mathbf{M}_w$  comutem, o modelo terá Orthogonal Block Structure (OBS).*

*Demonstração.* Se as matrizes  $\mathbf{M}_{m+1}, \dots, \mathbf{M}_w$  comutarem, essas matrizes e  $\mathbf{I}_n$  comutam, pertencendo a uma AJC,  $\mathcal{A}^\circ$ , completa com base principal  $bp(\mathcal{A}^\circ) = \{\mathbf{Q}_1^\circ, \dots, \mathbf{Q}_u^\circ\}$  tendo-se  $\mathbf{M}_i = \sum_{j=1}^u a_{i-m,j}^\circ \mathbf{Q}_j^\circ$  e  $\mathbf{I}_n = \sum_{j=1}^u \mathbf{Q}_j^\circ$ . Logo

$$\mathbf{V} = \sum_{j=1}^u (\gamma_j^\circ + \sigma^2) \mathbf{Q}_j^\circ$$

com  $\gamma_j^\circ = \sum_{i=1}^w a_{i-m,j}^\circ \sigma_i^2$ ,  $j = 1, \dots, u$ .

Basta agora admitir que  $\mathcal{A}^\circ = \mathcal{A}(\mathbf{M}^\circ)$ , com  $\mathbf{M}^\circ = \{\mathbf{M}_{m+1}, \dots, \mathbf{M}_w\}$  para que as matrizes  $\mathbf{Q}_1^\circ, \dots, \mathbf{Q}_u^\circ$  sejam conhecidas, o que estabelece a tese.  $\square$

Uma vez que a matriz  $\mathbf{T}$  de projecção ortogonal sobre  $\Omega$  é, ver Mexia (1995) [30], dada por

$$\mathbf{T} = \mathbf{X}\mathbf{X}^\top = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^+ \mathbf{X}^\top,$$

podemos então estabelecer a seguinte proposição.

**Proposição 3.2.** *Caso as matrizes  $\mathbf{M}_{m+1}, \dots, \mathbf{M}_w$  e  $\mathbf{T}$  comutem, o modelo terá Commutative Orthogonal Block Structure (COBS).*

*Demonstração.* Se as matrizes  $\mathbf{M}_{m+1}, \dots, \mathbf{M}_w$  e  $\mathbf{T}$  comutarem, essas matrizes, que comutam também com  $\mathbf{I}_n$ , pertencem a uma AJC completa  $\mathcal{A}$  com base principal  $bp(\mathcal{A}) = \{\mathbf{Q}_1, \dots, \mathbf{Q}_k\}$ . Observe-se que  $\mathbf{T}$  sendo MPO de  $\mathcal{A}$  é soma de matrizes da  $bp(\mathcal{A})$  por forma a ter-se  $\mathbf{T} = \sum_{k=1}^h \mathbf{Q}_j$ . Por outro lado ter-se-á  $\mathbf{M}_i = \sum_{j=1}^k c_{i,j} \mathbf{Q}_j$ ,

$i = m + 1, \dots, w$  e  $\mathbf{I}_n = \sum_{j=1}^k \mathbf{Q}_j$ , logo

$$\mathbf{V} = \sum_{j=1}^k (\gamma_j + \sigma^2) \mathbf{Q}_j,$$

com  $\gamma_j = \sum_{i=1}^{w-m} c_{i,j} \sigma_j^2$ ,  $j = 1, \dots, k$ , pelo que  $\mathbf{T}$  e  $\mathbf{V}$  comutam. Como, e atendendo à proposição anterior, o modelo terá COBS.  $\square$

**Corolário 3.** *Caso as matrizes  $\mathbf{M}_{m+1}, \dots, \mathbf{M}_w$  e  $\mathbf{T}$  comutem, os estimadores dos mínimos quadrados dos vectores estimáveis serão BLUE.*

Como vimos, nas condições da preposição anterior, tem-se

$$\mathbf{M}_i = \sum_{j=1}^k c_{i-m,j} \mathbf{Q}_j = \sum_{j \in \mathcal{C}_i} c_{i-m,j} \mathbf{Q}_j \quad i = m + 1, \dots, w$$

com  $\mathcal{C}_i = \{j : c_{i-m,j} \neq 0\}$ . Ora as matrizes  $\mathbf{Q}_j$  são mutuamente ortogonais logo, dado  $\mathbf{v} \in R(\mathbf{Q}_j)$ ,  $j = 1, \dots, k$ , tem-se

$$\mathbf{v}^\top \mathbf{M}_i \mathbf{v} = c_{i-m,j} \|\mathbf{v}\|^2, \quad j = 1, \dots, k, i = m + 1, \dots, w$$

e como as mtrizes  $\mathbf{M}_i$ ,  $i = m + 1, \dots, w$ , são semi-defnidas positivas, tem-se

$$c_{i-m,j} \geq 0 \quad j = 1, \dots, k, i = m + 1, \dots, w$$

vindo

$$\mathbf{T} + \sum_{i=m+1}^w \mathbf{M}_i = \sum_{j=1}^h \mathbf{Q}_j + \sum_{j=1}^k \left( \sum_{i=m+1}^w c_{i-m,j} \right) \mathbf{Q}_j = \sum_{j=1}^h d_j \mathbf{Q}_j$$

com

$$d_j = 1 + \sum_{i=m+1}^w c_{i-m,j} \geq 1 \quad j = 1, \dots, k$$

e, caso  $h < k$

$$d_j = \sum_{i=m+1}^w c_{i-m,j} \geq 0 \quad j = h+1, \dots, k.$$

A MPO sobre  $\bar{\Omega} = R \left( \mathbf{T} + \sum_{i=m+1}^w \mathbf{M}_i \right)$  será

$$\bar{\mathbf{T}} = \sum_{j \in \mathcal{D}} \mathbf{Q}_j,$$

com  $\mathcal{D} = \{j : d_j > 0\}$ . Podemos ordenar as matrizes da  $bp(\mathcal{A})$  por forma a ter-se  $\mathcal{D} = 1, \dots, l$ , com  $h \leq l$ , bem como

$$\left\{ \begin{array}{l} \mathbf{T} = \sum_{j=1}^h \mathbf{Q}_j \\ \mathbf{M}_i = \sum_{j=1}^l c_{i-m,j} \mathbf{Q}_j \quad i = m+1, \dots, w \\ \mathbf{I} = \sum_{j=1}^k \mathbf{Q}_j. \end{array} \right.$$

Vê-se que  $\Omega \subset \bar{\Omega} \subset R^n$ , se e só se  $h < l < k$ .

No que se segue admitimos que estas desigualdades se verificam, tendo-se  $\gamma_j = 0$ ,  $j = l+1, \dots, k$  e pondo-se

$$\left\{ \begin{array}{l} \mathbf{C}_1 = [c_{i,j}; i = 1, \dots, w-m, j = 1, \dots, h] \\ \mathbf{C}_2 = [c_{i,j}; i = 1, \dots, w-m, j = h+1, \dots, l] \end{array} \right.$$

com

$$\left\{ \begin{array}{l} \boldsymbol{\sigma}^2 = (\sigma_{m+1}^2, \dots, \sigma_w^2) \\ \boldsymbol{\gamma}(1) = (\gamma_1, \dots, \gamma_h) \\ \boldsymbol{\gamma}(2) = (\gamma_{h+1}, \dots, \gamma_k). \end{array} \right.$$

tem-se

$$\boldsymbol{\gamma}(1) = \mathbf{C}_1^\top \boldsymbol{\sigma}^2$$

e

$$\boldsymbol{\gamma}(2) = \mathbf{C}_2^\top \boldsymbol{\sigma}^2$$

bem como, caso os vectores linha de  $\mathbf{C}_2$  sejam linearmente independentes

$$\boldsymbol{\sigma}^2 = (\mathbf{C}_2^\top)^+ \boldsymbol{\gamma}(2)$$

e

$$\boldsymbol{\gamma}(1) = \mathbf{C}_1^\top (\mathbf{C}_2^\top)^+ \boldsymbol{\gamma}(2).$$

Quando se verifica esta condiço, os parmetros  $\boldsymbol{\sigma}^2$  e  $\boldsymbol{\gamma}(2)$  da parte de efeitos aleatrios, determinam-se mutuamente. Essa parte segrega-se ento como um sub-modelo que designaremos por COBS segregados (Seggregated Commutative Orthogonal Block Structure – S-COBS).

## 3.2 Projeces Ortogonais e Estimadores

Nesta seco utilizaremos projeces ortogonais para obter estimadores. O nosso tratamento assentará na estrutura algbrica do modelo, no se admitindo a normalidade do mesmo.

Continuaremos a admitir que  $h < l < k$  e punhamos

$$\mathbf{Q}^\perp = \sum_{j=l+1}^k \mathbf{Q}_j = \mathbf{A}^{\perp\top} \mathbf{A}^\perp$$

bem como

$$\mathbf{Q}_j = \mathbf{A}_j^\top \mathbf{A}_j \quad j = 1, \dots, l,$$

constituindo os vectores linha das matrizes  $\mathbf{A}_j$ ,  $j = 1, \dots, l$  e  $\mathbf{A}^\perp$  bases ortonormadas para os  $R(\mathbf{Q}_j)$ ,  $j = 1, \dots, l$  e  $R(\mathbf{Q}^\perp)$ . Tomando-se

$$\begin{cases} \boldsymbol{\eta}_j = \mathbf{A}_j \boldsymbol{\mu} & j = 1, \dots, l \\ \tilde{\boldsymbol{\eta}}_j = \mathbf{A}_j \mathbf{Y} & j = 1, \dots, l, \end{cases}$$

vê-se que  $\boldsymbol{\eta}_j$  é o vector médio de  $\tilde{\boldsymbol{\eta}}_j$ ,  $j = 1, \dots, l$  e que  $\boldsymbol{\eta}_j = \mathbf{0}$ ,  $j = h + 1, \dots, l$ . Assim, os  $\tilde{\boldsymbol{\eta}}_j$ ,  $j = 1, \dots, h$  serão estimadores centrados dos  $\boldsymbol{\eta}_j$ ,  $j = 1, \dots, h$ .

Por outro lado, como as matrizes  $\mathbf{Q}_1, \dots, \mathbf{Q}_l$  e  $\mathbf{Q}^\perp$  são mutuamente ortogonais, as matrizes de covariância das matrizes  $\mathbf{X}_i$ ,  $i = 1, \dots, w$  dos  $\tilde{\boldsymbol{\eta}}_j$ ,  $j = 1, \dots, l$ , são dadas por

$$\mathbf{V}_j = (\gamma_j + \sigma^2) \mathbf{I}_{g_j} \quad j = 1, \dots, l$$

com  $g_j = \text{car}(\mathbf{Q}_j)$ ,  $j = 1, \dots, l$ .

Então

$$S_j = \|\tilde{\boldsymbol{\eta}}_j\|^2 \quad j = h + 1, \dots, l$$

terá valor médio  $g_j(\gamma_j + \sigma^2)$ ,  $j = h + 1, \dots, l$ .

Da mesma forma se vê que o valor médio de

$$S = \|\mathbf{A}^\perp \mathbf{Y}\|^2$$

será  $g\sigma^2$  com  $g = \text{car}(\mathbf{Q}^\perp) = \sum_{j=l+1}^k \text{car}(\mathbf{Q}_j)$ .

Temos pois os estimadores centrados

$$\tilde{\sigma}^2 = \frac{S}{g}$$

e

$$\tilde{\gamma}_j = \frac{S_j}{g_j} - \tilde{\sigma}^2, \quad j = h + 1, \dots, h$$

sendo estes últimos as componentes de  $\tilde{\boldsymbol{\gamma}}(2)$ . Se o modelo for S-COBS, temos ainda os estimadores centrados

$$\tilde{\boldsymbol{\sigma}}^2 = (\mathbf{C}_2^\top) + \tilde{\boldsymbol{\gamma}}(2)$$

e

$$\tilde{\boldsymbol{\gamma}}(1) = \mathbf{C}_1^\top (\mathbf{C}_2^\top) + \tilde{\boldsymbol{\gamma}}(2).$$

Passando agora aos vectores estimáveis,

$$\boldsymbol{\Psi} = \mathbf{G}\boldsymbol{\beta}$$

será, estimável se e só se  $R(\mathbf{G}^\top) \subseteq R(\mathbf{X}^\top)$ , tendo-se o estimador de mínimos quadrados, ver por exemplo Mexia (1990) [29],

$$\tilde{\Psi} = \mathbf{G}\tilde{\beta}$$

com

$$\tilde{\beta} = (\mathbf{X}^\top \mathbf{X})^+ \mathbf{X}^\top \mathbf{Y}.$$

Recorde-se que estes estimadores são BLUE.

Por outro lado, como  $R(\mathbf{G}^\top) \subseteq R(\mathbf{X}^\top)$ , equivale a ter-se  $\mathbf{G}^\top = \mathbf{X}^\top \mathbf{W}^\top$  e a  $\mathbf{G} = \mathbf{W}\mathbf{X}$ , vindo por sua vez,

$$\Psi = \mu = \mathbf{W} \sum_{j=1}^h \mathbf{A}_j^\top \eta_j$$

visto ter-se

$$\mu = \mathbf{T}\mu = \left( \sum_{j=1}^h \mathbf{Q}_j \mu \right) = \sum_{j=1}^h \mathbf{A}_j^\top \eta_j.$$

Tem-se então o estimador centrado para  $\Psi$  dado por

$$\Psi^* = \mathbf{W} \sum_{j=1}^h \mathbf{A}_j^\top \tilde{\eta}_j.$$

Estabeleçamos agora a seguinte proposição.

**Proposição 3.3.** *Caso  $\Psi$  seja estimável, tem-se  $\Psi^* = \tilde{\Psi}$ , pelo que  $\Psi^*$  é BLUE.*

*Demonstração.* Como  $\sum_{j=1}^h \mathbf{A}_j^\top \tilde{\eta}_j = \sum_{j=1}^h \mathbf{Q}_j \mathbf{Y} = \mathbf{T}\mathbf{Y} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^+ \mathbf{X}^\top \mathbf{Y}$  ter-se-á

$$\Psi^* = \mathbf{W}\mathbf{X}(\mathbf{X}^\top \mathbf{X})^+ \mathbf{X}^\top \mathbf{Y} = \mathbf{G}\tilde{\beta} = \tilde{\Psi}$$

o que estabelece a tese. □

### 3.3 Normalidade

Admitamos que  $h < l < k$  e que  $\mathbf{y}$  é normal com vector médio  $\mu$  e matriz de covariância  $\mathbf{V}$ .

Como

$$\mathbf{V}^{-1} = \sum_{j=1}^l (\gamma_j + \sigma^2)^{-1} \mathbf{Q}_j + \frac{1}{\sigma^2} \mathbf{Q}^\perp$$

e

$$\det(\mathbf{V}) = \prod_{j=1}^l (\gamma_j + \sigma^2)^{g_j} \sigma^{2g},$$

a densidade de  $\mathbf{y}$  será dada por,

$$n(\mathbf{y}|\boldsymbol{\mu}, \mathbf{V}) = \frac{e^{-\frac{1}{2} \left( \sum_{j=1}^h \frac{\|\tilde{\boldsymbol{\eta}}_j - \boldsymbol{\eta}_j\|^2}{\gamma_j + \sigma^2} + \sum_{j=h+1}^l \frac{S_j}{\gamma_j + \sigma^2} + \frac{\mathbf{S}}{\sigma^2} \right)}}{(2\pi)^{n/2} \prod_{j=1}^l (\gamma_j + \sigma^2)^{g_j/2} \sigma^g},$$

já que

$$\begin{aligned} (\mathbf{y} - \boldsymbol{\mu})^\top \mathbf{V}^{-1} (\mathbf{y} - \boldsymbol{\mu}) &= \sum_{j=1}^l \frac{(\mathbf{y} - \boldsymbol{\mu})^\top \mathbf{A}_j^\top \mathbf{A}_j (\mathbf{y} - \boldsymbol{\mu})}{\gamma_j + \sigma^2} + \\ &+ \sum_{j=h+1}^l \frac{(\mathbf{y} - \boldsymbol{\mu})^\top \mathbf{A}_j^\top \mathbf{A}_j (\mathbf{y} - \boldsymbol{\mu})}{\gamma_j + \sigma^2} + \frac{(\mathbf{y} - \boldsymbol{\mu})^\top \mathbf{A}_j^\top \mathbf{A}_j (\mathbf{y} - \boldsymbol{\mu})}{\sigma^2} \\ &= \sum_{j=1}^h \frac{\|\tilde{\boldsymbol{\eta}}_j - \boldsymbol{\eta}_j\|^2}{\gamma_j + \sigma^2} + \sum_{j=h+1}^l \frac{S_j}{\gamma_j + \sigma^2} + \frac{\mathbf{S}}{\sigma^2}. \end{aligned}$$

Estabeleçamos agora a seguinte proposição.

**Proposição 3.4.** Quando  $\mathbf{y}$  é normal, os  $\tilde{\boldsymbol{\eta}}_j$ ,  $j = 1, \dots, h$ , as  $S_j$ ,  $j = h + 1, \dots, l$  e  $\mathbf{S}$  são estatísticas suficientes e completas.

*Demonstração.* Basta aplicar o teorema da factorização para verificar que as estatísticas indicadas são suficientes. Como a distribuição normal pertence à família exponencial e, como para estes modelos o espaço paramétrico contém abertos, as estatísticas serão completas.  $\square$

Para além do teorema enunciado anteriormente, é de especial importância para este trabalho os corolários que de seguida são enunciados.

**Corolário 4.** *Se o modelo for COBS, os  $\tilde{\sigma}^2$ ,  $\tilde{\gamma}(2)$  e  $\tilde{\Psi}$  serão UMVUE.*

*Demonstração.* Basta utilizar a proposição anterior, juntamente com o teorema de Blackwell-Lehman-Scheffé.  $\square$

**Corolário 5.** *Se o modelo for S-COBS os  $\tilde{\sigma}^2$ ,  $\tilde{\gamma}(2)$ ,  $\tilde{\Psi}$ ,  $\sigma^2$  e  $\tilde{\Psi}(1)$  serão UMVUE.*

*Demonstração.* Idêntica à do corolário anterior.  $\square$

### 3.3.1 Estimadores de Máxima Verosimilhança

No âmbito da estimação pontual, vamos, nesta sub-secção estudar como poderão ser obtidos os estimadores para estes modelos. Vamos assumir que temos um modelo normal com COBS.

Uma vez que o modelo tem COBS, a matriz de variância-covariância é dada por

$$\mathbf{V}(\boldsymbol{\sigma}^2) = \sum_{i=1}^m \sigma_i^2 \mathbf{P}_i,$$

onde  $\mathbf{P}_1, \dots, \mathbf{P}_m$  são MPO mutuamente ortogonais conhecidas que comutam com  $\mathbf{T}$ , a MPO sobre  $\Omega$ , o espaço-imagem de  $\mathbf{X}$ ,  $R(\mathbf{X})$ .

Uma vez que  $\mathbf{V}$  e  $\mathbf{T}$  comutam, mesmo não assumindo a normalidade, o estimador dos mínimos quadrados dos vectores estimáveis é, como vimos, o BLUE.

Recordemos que  $\boldsymbol{\Psi} = \mathbf{A}\boldsymbol{\beta}$  é estimável se  $R(\mathbf{A}^\top) \subseteq R(\mathbf{X}^\top)$  e que o seu estimador dos mínimos quadrados é  $\tilde{\boldsymbol{\Psi}} = \mathbf{A}\tilde{\boldsymbol{\beta}}$  com

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^+ \mathbf{X}^\top \mathbf{y}.$$

Vamos agora assumir que o modelo é normal. De Fonseca *et al* (2006) [15], sabemos que

$$\det(\mathbf{V}(\boldsymbol{\sigma}^2)) = \prod_{i=1}^m \sigma_i^{2g_i},$$

com  $g_i = \text{car}(\mathbf{P}_i)$ , e que

$$\mathbf{V}^{-1}(\boldsymbol{\sigma}^2) = \sum_{i=1}^m \frac{1}{\sigma_i^2} \mathbf{P}_i.$$

Podemos escrever facilmente a função de máxima verosimilhança como

$$l(\boldsymbol{\beta}, \boldsymbol{\sigma}^2 | \mathbf{y}) = -\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{V}^{-1}(\boldsymbol{\sigma}^2)(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) - \sum_{i=1}^m \frac{g_i}{2} \log \sigma_i^2 - \frac{n}{2} \log 2\pi.$$

Para se proceder à maximização da função de máxima verosimilhança, iremos usar uma técnica com duas fases. Na primeira fase, maximizamos

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{V}^{-1}(\boldsymbol{\sigma}^2)(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

como uma função de  $\boldsymbol{\beta}$  para um dado  $\boldsymbol{\sigma}^2$ . Para o efeito, sabendo que

$$\mathbf{T}^\perp = \mathbf{I} - \mathbf{T},$$

temos

$$\begin{aligned} & (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{T} + \mathbf{T}^\perp) \mathbf{V}^{-1}(\boldsymbol{\sigma}^2) (\mathbf{T} + \mathbf{T}^\perp) (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \\ & = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{T} \mathbf{V}^{-1}(\boldsymbol{\sigma}^2) \mathbf{T} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{T}^\perp \mathbf{V}^{-1}(\boldsymbol{\sigma}^2) \mathbf{T}^\perp (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ & = (\mathbf{T}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{V}^{-1}(\boldsymbol{\sigma}^2) (\mathbf{T}\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + (\mathbf{T}^\perp \mathbf{y})^\top \mathbf{V}^{-1}(\boldsymbol{\sigma}^2) (\mathbf{T}^\perp \mathbf{y}), \end{aligned}$$

Uma vez que

$$\mathbf{T} \mathbf{V}^{-1}(\boldsymbol{\sigma}^2) \mathbf{T}^\perp = \mathbf{V}^{-1}(\boldsymbol{\sigma}^2) \mathbf{T} \mathbf{T}^\perp = \mathbf{0},$$

porque  $\mathbf{T} \mathbf{T}^\perp = \mathbf{0}$ , só teremos de minimizar o primeiro termo. O mínimo é obtido para  $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$  com

$$\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{T}\mathbf{y},$$

pois a forma é quadrática e  $\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{T}\mathbf{y}$  anula a forma.

Este é para o estimador de mínimos quadrados para  $\boldsymbol{\beta}$ .

No segundo passo temos, de maximizar

$$\begin{aligned} & -\frac{1}{2}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top \mathbf{V}^{-1}(\boldsymbol{\sigma}^2)(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) - \sum_{i=1}^m \frac{g_i}{2} \log \sigma_i^2 - \frac{n}{2} \log 2\pi = \\ & = -\frac{1}{2}(\mathbf{T}^\perp \mathbf{y})^\top \mathbf{V}^{-1}(\boldsymbol{\sigma}^2)(\mathbf{T}^\perp \mathbf{y}) - \sum_{i=1}^m \frac{g_i}{2} \log \sigma_i^2 - \frac{n}{2} \log 2\pi \\ & = -\frac{1}{2} \sum_{i=1}^m \left( \frac{1}{\sigma_i^2} (\mathbf{T}^\perp \mathbf{y})^\top \mathbf{P}_i (\mathbf{T}^\perp \mathbf{y}) + g_i \log \sigma_i^2 \right) - \frac{n}{2} \log 2\pi, \end{aligned}$$

uma vez que  $\mathbf{T}^\perp \mathbf{X} = \mathbf{0}$  e  $\mathbf{TX} = \mathbf{X}$ .

Então, o máximo para os  $\sigma_i^2$ ,  $i = 1, \dots, m$ , será dado por

$$\hat{\sigma}_i^2 = \frac{S_i}{g_i}, \quad i = 1, \dots, m,$$

com

$$\begin{aligned} S_i &= (\mathbf{T}^\perp \mathbf{y})^\top \mathbf{P}_i (\mathbf{T}^\perp \mathbf{y}) \\ &= \mathbf{y}^\top \mathbf{T}^\perp \mathbf{P}_i \mathbf{T}^\perp \mathbf{y}, \quad i = 1, \dots, m. \end{aligned}$$

Estamos agora em condições para escrever a função densidade de probabilidade de  $\mathbf{y}$ . Esta pode ser escrita como

$$n(\mathbf{y} | \mathbf{X}\boldsymbol{\beta}, \forall(\boldsymbol{\sigma}^2)) = \frac{e^{-\frac{1}{2} \left( (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top \mathbf{X}^\top \mathbf{V}^{-1}(\boldsymbol{\sigma}^2) \mathbf{X} \sum_{i=1}^m \frac{S_i}{g_i} \right)}}{(2\pi)^{\frac{n}{2}} \prod_{i=1}^m \sigma_i^{2g_i}},$$

pelo que  $\hat{\boldsymbol{\beta}}$  e  $S_1, \dots, S_m$  constituem uma estatística suficiente e completa e, de acordo com o teorema de Blackwell-Lehman-Scheffé, o estimador de máxima verosimilhança de qualquer vector estimável será UMVUE, e as

$$\tilde{\sigma}_i^2 = \frac{S_i}{p_i},$$

com  $p_i = \text{car}(\mathbf{T}^\perp \mathbf{P}_i \mathbf{T}^\perp) = \text{car}(\mathbf{T}^\perp \mathbf{P}_i)$ ,  $i = 1, \dots, m$ , serão UMVUE para as componentes de variância.

Observemos que

$$\tilde{\sigma}_i^2 = \frac{S_i}{p_i} \hat{\sigma}_i^2 \quad i = 1, \dots, m,$$

pelo que temos os factores de correcção

$$\frac{g_i}{p_i} \quad i = 1, \dots, m,$$

que permitem obter UMVUE para as componentes de variância a partir dos Estimadores de Máxima Verosimilhança.

Esta observação generaliza o facto bem conhecido de, para amostras normais termos

$$\tilde{\sigma}^2 = \frac{n}{n-1} \hat{\sigma}^2.$$

### 3.4 Componentes de variância e vectores estimáveis

#### 3.4.1 Componentes variância

Vejamos agora como poderá ser feita inferência para as componentes de variância  $\sigma_i^2$ ,  $i = m + 1, \dots, w$  e para os  $\gamma_j$ ,  $j = 1, \dots, m$ .

Para o efeito iremos considerar parâmetros da forma

$$\theta = \sum_{j=m+1}^{w+1} d_j \gamma_j = \mathbf{d}^\top \boldsymbol{\gamma}_{(2)},$$

já que as componentes de variância são combinações lineares das componentes de  $\boldsymbol{\gamma}_{(2)}$ .

Tomando  $\mathbf{X}_i = (X_{i,m+1}, \dots, X_{i,w+1})$  com  $X_{i,j} \sim \chi_{g_j}^2$ ,  $j = m + 1, \dots, w + 1$ ,  $i = 1, \dots, N$  independentes, teremos uma amostra  $U_1, \dots, U_N$  com

$$U_i = \sum_{j=m+1}^{w+1} d_j \frac{s_j}{X_{i,j}} \quad i = 1, \dots, N,$$

a partir das quais se constroem intervalos de confiança para os parâmetros  $\theta$ , a partir da função de distribuição empírica. Estes intervalos poderão ser utilizados para testar  $H_0 : \theta = \theta_0$ , por dualidade.

No estudo das componentes de variância sucede amiúde observar-se que  $\sigma^2$  pode ser escrito como a combinação

$$\sigma^2 = \sigma^{2+} - \sigma^{2-}$$

onde

$$\begin{cases} \sigma^{2+} = \sum_{j \in \mathcal{C}^+} d_j \gamma_j \\ \sigma^{2-} = \sum_{j \in \mathcal{C}^-} d_j \gamma_j \end{cases},$$

com

$$\begin{cases} \mathcal{C}^+ = \{j : d_j > 0\} \\ \mathcal{C}^- = \{j : d_j < 0\} \end{cases}.$$

Consequentemente, utilizaremos as amostras  $\{U_1^+, \dots, U_N^+\}$  e  $\{U_1^-, \dots, U_N^-\}$ , sendo

$$\begin{cases} U_i^+ = \sum_{j \in \mathcal{C}^+} d_j \frac{s_j}{X_{i,j}} \\ U_i^- = \sum_{j \in \mathcal{C}^-} d_j \frac{s_j}{X_{i,j}} \end{cases} \quad i = 1, \dots, N.$$

Em Ferreira (2006) [11], as densidades conjuntas correspondentes aos pares  $(U^+, U^-)$  são unimodais. Tal facto, e por comparação com a localização da moda (ou centro da classe modal), obtém-se através da densidade empírica,

$$\begin{cases} \sigma^{2+} = \sum_{j \in \mathcal{C}^+} d_j \gamma_j \\ \sigma^{2-} = \sum_{j \in \mathcal{C}^-} d_j \gamma_j \end{cases}.$$

Como vimos, as amostras conjuntas  $\{(U_1^+, U_1^-), \dots, (U_N^+, U_N^-)\}$  correspondem a densidades unimodais com curvas de nível representadas na Figura 2.1.

Interessará considerar para além da moda, representada pelo centro de classe modal, o ponto cujas coordenadas são os estimadores UMVUE de  $\sigma^{2+}$  e  $\sigma^{2-}$ . Quando os dois pontos estão próximos, pode-se concluir pela boa qualidade do estimador. Na aplicação apresentada na secção 5.3 do capítulo 5, encontraremos uma situação em que tal se verifica.

### 3.4.2 Vectores Estimáveis

Como atrás foi estabelecido (ver corolário 3), os estimadores dos mínimos quadrados dos vectores estimáveis são BLUE.

Quando a normalidade é assumida, esses estimadores são UMVUE já que, sendo  $\Psi = \mu\beta$  um vector estimável se tem,  $L = W\mathbf{X}$ , ver Mexia 1990 [29]. Então o

estimador dos mínimos quadrados de  $\Psi$  será

$$\begin{aligned}\tilde{\Psi} &= L\tilde{\beta} = L(\mathbf{X}^\top \mathbf{X})^+ \mathbf{X}^\top \mathbf{y} \\ &= \mathbf{W}\mathbf{X}(\mathbf{X}^\top \mathbf{X})^+ \mathbf{X}^\top \mathbf{y} \\ &= \mathbf{W}\mathbf{T}\mathbf{y}\end{aligned}$$

e

$$\begin{aligned}\mathbf{T}\mathbf{y} &= \sum_{j=1}^m \mathbf{Q}_j \mathbf{y} = \sum_{j=1}^m \mathbf{A}_j^\top \mathbf{A}_j \mathbf{y} \\ &= \sum_{j=1}^m \mathbf{A}^\top \tilde{\eta}_j\end{aligned}$$

pelo que, atendendo ao teorema de Blackwell-Lehmann-Scheffé,  $\tilde{\Psi}$  será UMVUE.

# Capítulo 4

## Modelos Estritamente Associados

Estes modelos foram introduzidos por Jesus (2007)[21].

### 4.1 Estrutura Algébrica

Seja  $\mathbf{P}$  uma matriz ortogonal associada a uma álgebra de Jordan comutativa  $\mathcal{A}$ , constituída por matrizes  $m \times m$ , definida como

$$\mathbf{P} = [\mathbf{A}_1^\top, \dots, \mathbf{A}_w^\top]^\top.$$

Com  $\mathbf{K}_s$ , uma matriz obtida retirando a primeira linha a uma matriz ortogonal  $s \times s$ , multiplicada por  $\frac{1}{\sqrt{s}}\mathbf{1}_s^\top$ , obtém-se

$$\bar{\mathbf{P}} = \left[ \mathbf{A}_1^\top \otimes \frac{1}{\sqrt{s}}\mathbf{1}_s, \dots, \mathbf{A}_w^\top \otimes \frac{1}{\sqrt{s}}\mathbf{1}_s, \mathbf{A}^{\perp\top} \right]^\top.$$

Verifica-se então que  $\mathbf{A}^\perp = \mathbf{I}_m \otimes \mathbf{K}_s$  será a matriz ortogonal associada a  $\mathcal{A} \star \mathcal{A}(s)$ . Tem-se ainda que

$$\bar{\mathbf{J}}_s = \mathbf{I}_s - \frac{1}{s}\mathbf{J}_s = \mathbf{K}_s^\top \mathbf{K}_s.$$

Com estes resultados em mente, podemos definir modelos estritamente associados.

**Definição 4.1.** Um modelo com  $m = n \times s$  observações está estritamente associado a uma AJC,  $\mathcal{A}$ , se poder ser escrito sob a forma

$$\mathbf{y} = \sum_{i=1}^w \left( \mathbf{A}_j^\top \otimes \frac{1}{\sqrt{s}} \mathbf{1}_s \right) \tilde{\boldsymbol{\eta}}_j + \boldsymbol{\varepsilon} \quad (4.1)$$

com  $\tilde{\boldsymbol{\eta}}_j \sim N(\boldsymbol{\eta}_j, \gamma_j \mathbf{I}_{g_j})$ ,  $j = 1, \dots, w$ ,  $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{Q}^\perp)$ , e com  $\mathbf{Q}^\perp = \mathbf{A}^{\perp\top} \mathbf{A}^\perp$ , sendo estes vectores independentes.

Este modelo terá  $m$  termos de efeitos fixos, se

$$\boldsymbol{\eta}_j = \mathbf{0}, \quad j = m+1, \dots, w$$

e

$$\gamma_j = \sigma^2, \quad j = 1, \dots, m.$$

Estabeleçamos agora a seguinte proposição,

**Proposição 4.1.** Os modelos estritamente associados a uma AJC,  $\mathcal{A}$ , com  $m < w$  têm COBS.

*Demonstração.* Os modelos estritamente associados a uma AJC,  $\mathcal{A}$ , têm vector médio

$$\boldsymbol{\mu} = \sum_{j=1}^m \left( \mathbf{A}_j^\top \otimes \frac{1}{\sqrt{s}} \mathbf{1}_s \right) \boldsymbol{\eta}_j$$

e matriz de covariância

$$\mathbf{V} = \sum_{j=1}^w \gamma_j \left( \mathbf{Q}_j \otimes \frac{1}{s} \mathbf{J}_s \right) + \sigma^2 \mathbf{Q}^\perp,$$

com  $\mathbf{Q}_j = \mathbf{A}_j^\top \mathbf{A}_j$ ,  $j = 1, \dots, w$ . Logo a matriz  $\mathbf{T}$  de projecção ortogonal sobre o espaço  $\Omega$ , varrido pelo vector médio  $\boldsymbol{\mu}$ , é dado por

$$\mathbf{T} = \sum_{j=1}^m \mathbf{Q}_j \otimes \frac{1}{s} \mathbf{J}_s,$$

sendo evidente que  $\mathbf{T}$  comuta com as  $\mathbf{Q}_j \otimes \frac{1}{s} \mathbf{J}_s$ ,  $j = 1, \dots, w$  e com  $\mathbf{Q}^\perp$ , o que estabelece a tese.  $\square$

Consideremos agora os modelos de efeitos fixos. Particularizando podemos estabelecer a seguinte proposição.

**Proposição 4.2.** *Modelos estritamente associados a uma AJC,  $\mathcal{A}$ , com  $m = w$  têm COBS.*

*Demonstração.* Nestas condições estamos a considerar modelos com vector médio dado por,

$$\boldsymbol{\mu} = \left( \sum_{j=1}^w \mathbf{A}_j^\top \otimes \frac{1}{\sqrt{s}} \mathbf{1}_s \right) \boldsymbol{\eta}_j$$

e matriz de covariância

$$\mathbf{V} = \sum_{j=1}^w \gamma_j \left( \mathbf{Q}_j \otimes \frac{1}{s} \mathbf{J}_s \right) + \sigma^2 \mathbf{Q}^\perp,$$

com  $\mathbf{Q}_j = \mathbf{A}_j^\top \mathbf{A}_j$ ,  $j = 1, \dots, w$ . Logo a matriz  $\mathbf{T}$  de projecção ortogonal sobre o espaço  $\Omega$  varrido pelo vector médio  $\boldsymbol{\mu}$  é dada por

$$\mathbf{T} = \sum_{j=1}^w \left( \mathbf{Q}_j \otimes \frac{1}{s} \mathbf{J}_s \right),$$

sendo evidente que  $\mathbf{T}$  comuta com as  $\mathbf{Q}_j \otimes \frac{1}{s} \mathbf{J}_s$ ,  $j = 1, \dots, w$  e com  $\mathbf{Q}^\perp$ .  $\square$

Consideremos o modelo estritamente associado, como definido em 4.1,

$$\mathbf{y} = \sum_{i=1}^w \left( \mathbf{A}_i^\top \otimes \frac{1}{\sqrt{s}} \mathbf{1}_s \right) \tilde{\boldsymbol{\eta}}_i + \boldsymbol{\varepsilon},$$

com  $\tilde{\boldsymbol{\eta}}_j \sim N(\boldsymbol{\eta}_j, \gamma_j \mathbf{I}_{g_j})$ , onde

$$\left\{ \begin{array}{ll} \gamma_j = \sigma^2 & j = 1, \dots, m \\ \boldsymbol{\eta}_1 = \sqrt{n} \boldsymbol{\mu} & \\ \boldsymbol{\eta}_j = \mathbf{0} & j = m + 1, \dots, w \end{array} \right.$$

para o qual teremos os casos de efeitos fixos, aleatórios e misto, quando se verificarem as seguintes condições

$$\left\{ \begin{array}{llll} m = w & \text{efeitos fixos} & \gamma_j = \sigma^2 & j = 1, \dots, w \\ 1 < m < w & \text{modelos mistos} & \gamma_j = \sigma^2 & j = 1, \dots, m \\ & & \boldsymbol{\eta}_j = \mathbf{0} & j = m + 1, \dots, w \\ 1 = m & \text{efeitos aleatórios} & \boldsymbol{\eta}_j = \mathbf{0} & j = 2, \dots, w. \end{array} \right.$$

Nestes modelos temos uma única base relevante para a AJC, a base principal, já que se tem  $\mathbf{A}_j^\top = \mathbf{X}_j$ ,  $j = 1, \dots, w$ . Ter-se-á pois  $\mathbf{M}_j = \mathbf{Q}_j$ ,  $j = 1, \dots, w$ , pelo que a matriz de transição entre bases é a matriz identidade. Resulta daqui que, no caso do modelo ser misto,

$$\mathbf{T} = \sum_{j=1}^m \mathbf{Q}_j,$$

pelo que  $\mathbf{T}$  comutará com as  $\mathbf{Q}_1, \dots, \mathbf{Q}_w$  e o modelo tem COBS.

Por outro lado, dado ter-se

$$\mathbf{B} = \mathbf{I}_w = \begin{bmatrix} \mathbf{I}_m & \mathbf{0}_{m \times (w-m)} \\ \mathbf{0}_{(w-m) \times m} & \mathbf{I}_{w-m} \end{bmatrix},$$

estes modelos gozam de segregação, como vimos anteriormente.

## 4.2 Inferência

Para efeitos de inferência, vamos agora estudar o modelo para os casos de efeitos fixos, mistos e aleatórios.

Assumamos que  $g_j = \text{car}(\mathbf{Q}_j) = \text{car}(\mathbf{A}_j)$ ,  $j = 1, \dots, w$ , e  $g = \text{car}(\mathbf{Q}^\perp) = \dot{n}(s-1)$  com  $\dot{n}$  o número de tratamentos. Temos então

$$\mathbf{V}^{-1} = \sum_{j=1}^w \gamma_j^{-1} \left( \mathbf{Q}_j \otimes \frac{1}{s} \mathbf{J}_s \right) + \frac{1}{\sigma^2} \mathbf{Q}^\perp$$

e

$$\det(\mathbf{V}) = \prod_{j=1}^w \gamma_j^{g_j} (\sigma^2)^g$$

bem como

$$\begin{aligned} (\mathbf{y} - \boldsymbol{\mu})^\top \mathbf{V}^{-1} (\mathbf{y} - \boldsymbol{\mu}) &= \sum_{j=1}^w \frac{1}{\gamma_j} \|\tilde{\boldsymbol{\eta}}_j - \boldsymbol{\eta}_j\|^2 + \frac{s}{\sigma^2} \\ &= \sum_{j=1}^m \frac{1}{\gamma_j} \|\tilde{\boldsymbol{\eta}}_j - \boldsymbol{\eta}_j\|^2 + \sum_{j=m+1}^w \frac{s_j}{\gamma_j} + \frac{s}{\sigma^2}, \end{aligned}$$

com  $s_j = \|\tilde{\boldsymbol{\eta}}_j\|^2$ ,  $j = 1, \dots, w$  e  $s = \|\boldsymbol{\varepsilon}^\perp\|^2 = \|\mathbf{Q}^\perp \mathbf{y}\|^2$ .

Podemos assim escrever a função densidade de probabilidade de  $\mathbf{y}$  como

$$n(\mathbf{y}|\sigma^2) = \frac{e^{-\frac{1}{2} \left( \sum_{j=1}^m \frac{1}{\sigma^2} \|\tilde{\boldsymbol{\eta}}_j - \boldsymbol{\eta}_j\|^2 + \sum_{j=m+1}^w \frac{s_j}{\gamma_j} + \frac{s}{\sigma^2} \right)}}{(2\pi)^{\frac{n}{2}} \prod_{j=1}^w \gamma_j^{\frac{g_j}{2}} (\sigma^2)^g},$$

pelo que, para os  $\tilde{\boldsymbol{\eta}}_j$ ,  $j = 1, \dots, m$ , os  $S_j$ ,  $j = m + 1, \dots, w$ , e  $S$  constituem estatísticas suficientes e completas, pelo teorema da factorização, já que a densidade normal pertence à família exponencial e o espaço paramétrico contém o produto de intervalos não degenerados, ver Lehman (1986, pág. 132) [24].

Como  $\tilde{\boldsymbol{\eta}}_j \sim N(\boldsymbol{\eta}_j, \gamma_j \mathbf{I}_{g_j})$ ,  $j = 1, \dots, w$ ,  $s_j \sim \gamma_j \chi_{g_j}^2$ ,  $j = m + 1, \dots, w$ , e  $s \sim \sigma^2 \chi_g^2$ , têm-se os UMVUE para os vectores estimáveis e componentes de variância.

### 4.2.1 Modelos de efeitos fixos

No caso em que o modelo tem só efeitos fixos, temos  $m = w$ , sendo o parâmetro que figura na formulação canónica  $\sigma^2$ , já que  $\gamma_j = \sigma^2$ ,  $j = 1, \dots, w$  e  $\tilde{\boldsymbol{\eta}}_j$ ,  $j = 1, \dots, w$ .

Como  $s \sim \sigma^2 \chi_g^2$ , para  $\sigma^2$  podemos construir intervalos de confiança de nível  $1 - \alpha$

dados por

$$\left\{ \begin{array}{l} \left[ \frac{s}{x_{g;1-\frac{\alpha}{2}}}; \frac{s}{x_{g;\frac{\alpha}{2}}} \right] \\ \left[ 0; \frac{s}{x_{g;\alpha}} \right] \\ \left[ \frac{s}{x_{g;1-\alpha}}; 0 \right] \end{array} \right. ,$$

que permitem, por dualidade, construir testes bilaterais, unilaterais à esquerda e unilaterais à direita para

$$H_0 : \sigma^2 = \sigma_0^2,$$

respectivamente . Estes testes terão nível de significância  $\alpha$  ao rejeitar  $H_0$  quando  $\sigma_0^2$  não pertencer ao intervalo correspondente.

É interessante observar que o teste unilateral à esquerda rejeita  $H_0$  quando  $s < \sigma_0^2 x_{g,\alpha}$ . Ora  $s \sim \sigma^2 \chi_g^2$ , pelo que este teste terá função potência decrescente dada por

$$Pot(\sigma^2) = pr(s < \sigma_0^2 x_{g,\alpha}) = pr\left(\chi_g^2 < \frac{\sigma_0^2}{\sigma^2} x_{g,\alpha}\right),$$

sendo este teste unilateral à esquerda, não distorcido.

Da mesma maneira se mostra que o teste unilateral à direita tem função potência crescente

$$Pot(\sigma^2) = pr(s > \sigma_0^2 x_{g,1-\alpha}) = pr\left(\chi_g^2 > \frac{\sigma_0^2}{\sigma^2} x_{g,1-\alpha}\right),$$

pelo que, também, este será não distorcido.

Passando aos parâmetros  $\boldsymbol{\eta}_j$ ,  $j = 1, \dots, w$ , como  $s_j = \|\boldsymbol{\eta}_j - \tilde{\boldsymbol{\eta}}_j\|^2 \sim \sigma^2 \chi_{g_j}^2$ ,  $j = 1, \dots, w$  são independentes de  $s \sim \sigma^2 \chi_g^2$ , teremos as variáveis pivot

$$\mathcal{F}'_j = \frac{g}{g_j} \frac{\|\boldsymbol{\eta}_j - \tilde{\boldsymbol{\eta}}_j\|^2}{s} \quad j = 1, \dots, w$$

com distribuição  $F$  centrais com  $g_j$ ,  $j = 1, \dots, w$  e  $g$  graus de liberdade. Obtêm-se assim para os  $\boldsymbol{\eta}_j$ ,  $j = 1, \dots, w$ , as hiper-esferas de confiança de nível  $1 - \alpha$  dadas por

$$pr\left(\|\boldsymbol{\eta}_j - \tilde{\boldsymbol{\eta}}_j\|^2 \leq g_j f_{g_j, g, 1-\alpha} \frac{s}{g}\right) = 1 - \alpha.$$

Utilizando-se o teorema de Scheffé, ver Scheffé (1999) [38], obtêm-se os intervalos de confiança simultâneos dados por

$$pr \left[ \bigcap_{\mathbf{d}} \left( \mathbf{d}^\top \boldsymbol{\eta} - \mathbf{d}^\top \tilde{\boldsymbol{\eta}} \leq \sqrt{g_j f_{g_j, g, 1-\alpha} \|\mathbf{d}\|^2 \frac{s}{g}} \right) \right] = 1 - \alpha,$$

onde  $\bigcap_{\mathbf{d}}$  indica que se consideram todos os vectores com dimensão adequada.

Por outro lado pode utilizar-se as hiper-esferas de confiança para testar as hipóteses

$$H_{0,j} : \boldsymbol{\eta}_j = \boldsymbol{\eta}_{0,j} \quad j = 1, \dots, w.$$

Obtêm-se assim testes de nível  $\alpha$  que rejeitam as  $H_{0,j}$ ,  $j = 1, \dots, w$ , quando os  $\boldsymbol{\eta}_{0,j}$  não estão contidos na hiper-esfera.

Ora, ver Mexia (1995) [30], as estatísticas

$$\mathcal{F} = \frac{g \|\tilde{\boldsymbol{\eta}}_j - \boldsymbol{\eta}_{j,0}\|^2}{g_j s} \quad j = 1, \dots, w$$

terão distribuição  $F$  com  $g_j$  e  $g$  graus de liberdade,  $j = 1, \dots, w$ , e parâmetros de não centralidade dados por

$$\delta_j = \frac{1}{\sigma^2} \|\boldsymbol{\eta}_j - \boldsymbol{\eta}_{j,0}\|^2 \quad j = 1, \dots, w.$$

Como as hipóteses podem ser reescritas na forma

$$H_{0,j} : \delta_j = 0 \quad j = 1, \dots, w,$$

vê-se que estes testes são não distorcidos. Por outro lado o teste  $F$  de nível  $q$ , rejeita  $H_{0,j}$ ,  $j = 1, \dots, w$ , se e só se  $\boldsymbol{\eta}_{0,j}$  não estiver contido na hiper-esfera de confiança de nível  $1 - q$ ,  $j = 1, \dots, w$ .

### 4.2.2 Modelos de efeitos mistos

Consideremos agora os efeitos mistos. O modelo (4.1) pode ser escrito, de forma a colocar em evidência estes efeitos, como

$$\mathbf{y} = \sum_{j=1}^m \left( \mathbf{A}_j^\top \otimes \frac{1}{\sqrt{s}} \mathbf{1}_s \right) \boldsymbol{\eta}_j + \sum_{j=m+1}^w \left( \mathbf{A}_j^\top \otimes \frac{1}{\sqrt{s}} \mathbf{1}_s \right) \tilde{\boldsymbol{\eta}}_j + \boldsymbol{\varepsilon}$$

onde os  $\boldsymbol{\eta}_j$ ,  $j = 1, \dots, m$  são fixos e  $\tilde{\boldsymbol{\eta}}_j \sim N(\mathbf{0}, \gamma_j \mathbf{I}_{g_j})$ ,  $j = m+1, \dots, w$ , bem como  $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{Q}^\top)$ , sendo estes últimos independentes.

A inferência relativa a  $\sigma^2$  e aos  $\boldsymbol{\eta}_j$ ,  $j = 1, \dots, m$ , faz-se como para os modelos de efeitos fixos.

Por outro lado temos

$$\left( \mathbf{A}_j \otimes \frac{1}{\sqrt{s}} \mathbf{1}_s^\top \right) \mathbf{y} = \tilde{\boldsymbol{\eta}}_j + \left( \mathbf{A}_j \otimes \frac{1}{\sqrt{s}} \mathbf{1}_s^\top \right) \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \gamma_j + \sigma^2 \mathbf{I}_{g_j}),$$

com  $j = m+1, \dots, w$ , pelo que

$$S_j = \left\| \left( \mathbf{A}_j \otimes \frac{1}{\sqrt{s}} \mathbf{1}_s^\top \right) \mathbf{y} \right\|^2 \sim (\gamma_j + \sigma^2) \chi_{g_j}^2, \quad j = m+1, \dots, w,$$

tendo-se pois os UMVUE

$$\tilde{\gamma}_j = \frac{S_j}{g_j} - \tilde{\sigma}^2, \quad j = m+1, \dots, w.$$

Ponhamos agora

$$\nu_j = \frac{\gamma_j}{\sigma^2}, \quad j = m+1, \dots, w,$$

o coeficiente de componente de variância, vindo  $\nu \geq 0$ , e, como  $S_j$  é independente de  $S$ ,

$$\mathcal{F}_j = \frac{g_j S_j}{g_j S}, \quad j = m+1, \dots, w,$$

é o produto por  $\nu_j + 1$  de uma variável com distribuição  $F$  central com  $g_j$ ,  $j = m+1, \dots, w$ , e  $g$  graus de liberdade.

Pode-se pois utilizar  $\mathcal{F}_j$ ,  $j = m+1, \dots, w$ , para testar

$$H_{0,j} : \gamma_j = 0, \quad j = m+1, \dots, w.$$

Sendo  $f_{p,r,s}$  o quantil para a probabilidade  $p$  da distribuição  $F$  central com  $r$  e  $s$  graus de liberdade, o teste de nível  $\alpha$  rejeita a hipótese  $H_{0,j}$  quando  $\mathcal{F}_j > f_{1-\alpha, g_j, g}$ ,

$j = m + 1, \dots, w$ . Assim, este teste terá função potência dada por

$$\begin{aligned} Pot_j(\gamma_j) &= 1 - F\left(\frac{f_{1-\alpha, g_j, g}}{\frac{\gamma_j}{\sigma^2} + 1} | g_j, g\right) \\ &= 1 - F\left(\frac{f_{1-\alpha, g_j, g}}{\nu_j + 1} | g_j, g\right), \quad j = m + 1, \dots, w. \end{aligned}$$

Como esta função é crescente em  $\nu_j$  e  $H_{0,j}$  se verifica se, e só se,  $\nu = 1$ , o teste é estritamente não distorcido.

Podemos igualmente construir os intervalos de confiança de nível  $1 - \alpha$ , para  $\nu_j$ ,  $j = m + 1, \dots, w$ , dados pelas expressões

$$\left\{ \begin{array}{l} \left[ \frac{\mathcal{F}_j}{f_{1-\frac{\alpha}{2}, g_j, g}} - 1; \frac{\mathcal{F}_j}{f_{\frac{\alpha}{2}, g_j, g}} - 1 \right], \quad j = m + 1, \dots, w \\ \left[ \frac{\mathcal{F}_j}{f_{1-\alpha, g_j, g}} - 1; +\infty \right], \quad j = m + 1, \dots, w \\ \left[ 0; \frac{\mathcal{F}_j}{f_{\alpha, g_j, g}} - 1 \right], \quad j = m + 1, \dots, w. \end{array} \right.$$

Estes intervalos de confiança permitem-nos testar as hipóteses

$$H_{0,j}^\circ(\nu_{0,j}) : \nu_j = \nu_{0,j}, \quad j = m + 1, \dots, w,$$

rejeitando-se  $H_{0,j}^\circ(\nu_{0,j})$ ,  $j = m + 1, \dots, w$ , quando  $\nu_{0,j}$  não está contido no intervalo de confiança. Obtêm-se assim testes, bilaterais, unilaterais direitos e unilaterais esquerdos de nível  $1 - \alpha$ .

Observe-se ainda que

$$H_{0,j} = H_{0,j}^\circ(1); \quad j = m + 1, \dots, w,$$

pelo que estas hipóteses generalizam as hipóteses anteriores.

Do ponto de vista prático, em geral  $g \geq 3$ , pelo que

$$\mathbb{E}(\mathcal{F}_j) = \frac{g}{g-2}(\nu_j + 1) \quad j = m + 1, \dots, w,$$

e como tal,

$$\tilde{\nu}_j = \frac{g-2}{g} \mathcal{F}_j - 1, \quad j = m+1, \dots, w,$$

será UMVUE de  $\nu_j$ ,  $j = m+1, \dots, w$ .

Podemos ainda considerar o teste para as seguintes hipóteses,

$$\begin{aligned} H_{0,j,j'}(\lambda_0) : \frac{\nu_j + 1}{\nu_{j'} + 1} = \lambda_0 & \quad ; \quad j = m+1, \dots, w, \\ & \quad \quad \quad j' = m+1, \dots, w, \\ & \quad \quad \quad j \neq j'. \end{aligned}$$

Para o efeito, observe-se que

$$\begin{aligned} \mathcal{F}_{j,j'} = \frac{g_{j'} S_j}{g_j S_{j'}} & \quad ; \quad j = m+1, \dots, w, \\ & \quad \quad \quad j' = m+1, \dots, w, \\ & \quad \quad \quad j \neq j', \end{aligned}$$

é o produto por

$$\begin{aligned} \lambda_{j,j'} = \frac{\nu_j + 1}{\nu_{j'} + 1} & \quad ; \quad j = m+1, \dots, w, \\ & \quad \quad \quad j' = m+1, \dots, w, \\ & \quad \quad \quad j \neq j', \end{aligned}$$

duma variável com distribuição  $F$  com  $g_j$  e  $g_{j'}$ , graus de liberdade, pelo que os testes unilaterais são não distorcidos.

### 4.2.3 Modelos de efeitos aleatórios

Para o tratamento de efeitos aleatórios, basta proceder como na secção anterior, para o caso de efeitos mistos, onde o tratamento para efeitos aleatórios, com  $j = m+1, \dots, w$  é apresentado.

## 4.3 Factoriais de Base Prima

Existem situações em que se deseja estudar o efeito de diversos factores sobre uma ou mais variáveis resposta. Uma forma de o fazer, seria investigar separadamente

o efeito destes factores, outra, bem mais eficiente, seria fazer uso de uma estrutura factorial de tratamentos. Assim, todos os factores são investigados conjuntamente. Este tipo de estrutura (factorial) poderá ser vista como uma aplicação de modelos estritamente associados. De entre os possíveis modelos, aqueles com maior aplicabilidade do ponto de vista prático, e como tal com maior interesse, são os delineamentos de factoriais com base prima.

Consideremos que temos  $N$  factores com  $p$  níveis, sendo  $p$  primo. Teremos um factorial  $p^N$  com  $n = p^N$  combinações de níveis, a que se dão o nome de tratamentos. Para os casos em que  $N = 2$  e  $p = 3$ , ver Montgomery (2005) [33].

Numeremos os níveis de 0 a  $p - 1$  o que, como veremos, nos permitirá utilizar os espaços vectoriais sobre corpos de Galois<sup>1</sup> no tratamento destes modelos. Ponhamos  $[p] = 0, \dots, p - 1$  e  $[p]_0 = 1, \dots, p - 1$ . Definindo-se em  $[p]$  a adição e multiplicação módulo  $p$ , obtém-se um corpo de Galois  $G[p]$  com suporte  $[p]$ .

O conjunto  $[p]^N$  de vectores de dimensão  $N$ , com componentes em  $[p]$  será o suporte de um espaço vectorial  $G[p]^N$  de dimensão  $N$  sobre  $G[p]$ .

O dual  $\mathcal{L} = \mathcal{L}[p]^N$  de  $G[p]^N$  é constituído pelas aplicações lineares,

$$l(\mathbf{x}|\mathbf{a}) = \left( \sum_{j=1}^N a_j x_j \right)_{(p)},$$

onde  $\mathbf{x}, \mathbf{a} \in G[p]^N$  e  $(p)$  indicam o uso da aritmética módulo  $p$ , de  $G[p]^N$  em  $[p]$ .

Com  $c_1, \dots, c_u \in [p]$ , podemos então escrever

$$\sum_{i=1}^u c_i l(\mathbf{x}|\mathbf{a}_i) = l\left(\mathbf{x} \left| \sum_{i=1}^u (c_i \mathbf{a}_i)_{(p)} \right.\right).$$

---

<sup>1</sup>Évariste Galois (1811 - 1832). Apesar da sua controversa e dramática história, a denominada *Teoria de Galois*, é considerada por muitos como uma das mais belas partes da matemática. Deve-se basicamente a dois factores: a surpreendente ligação entre a teoria de grupos e as raízes de polinómios, e a elegância da sua apresentação. Com a sua morte prematura com 20 anos de idade, muito do seu trabalho foi deixado inacabado, mas lançou as fundações para muitos dos actuais conceitos da álgebra moderna, onde se incluem grupos e corpos. Ver a este respeito Cox (2004) [8].

Se escrevermos  $\phi(\mathbf{a}) = l(\mathbf{x}|\mathbf{a})$ , podemos definir um isomorfismo entre  $G[p]^N$  e  $\mathcal{L}[p]^N$ , pelo que as aplicações lineares  $l(\mathbf{x}|\mathbf{a}_i)$ ,  $i = 1, \dots, u$  serão linearmente independentes, se e só se, os  $\mathbf{a}_1, \dots, \mathbf{a}_u$  forem linearmente independentes.

Tanto  $G[p]^N$  como  $\mathcal{L}[p]^N$  têm dimensão  $N$ , bem como

$$\#(G[p]^N) = \#(\mathcal{L}[p]^N) = p^N,$$

onde os vectores nulos de  $G[p]^N$  e  $\mathcal{L}[p]^N$  são  $\mathbf{0}$  e  $l_0(\mathbf{x}) = l(\mathbf{x}|\mathbf{0})$ .

Podemos definir uma relação de equivalência,  $\rho$ , em  $\mathcal{L}[p]^p$  pondo  $l(\mathbf{x}|\mathbf{a}_1)\rho l(\mathbf{x}|\mathbf{a}_2)$ , quando  $\mathbf{a}_2 = (c\mathbf{a}_1)_{(p)}$

Verifica-se que  $l_0(\mathbf{x})$  está isolado na sua classe de equivalência  $\mathcal{G}$ . As aplicações não nulas estarão agrupadas em classes com  $p - 1$  elementos. Teremos então

$$k_N(p) = \frac{p^N - 1}{p - 1}$$

classes. Cada uma destas classes contém uma e uma só aplicação cujo primeiro coeficiente não nulo é 1. Tais combinações lineares são denominadas de reduzidas e a sua família é representada por  $\mathcal{L}[p]_r^N$ , tendo-se  $\#(\mathcal{L}[p]_r^N) = k_N(p)$ . A homóloga reduzida de uma aplicação linear é a aplicação equivalente  $\rho$  à anterior.

As aplicações  $l(\mathbf{x}|\mathbf{a}_1), \dots, l(\mathbf{x}|\mathbf{a}_u)$  são independentes se e só se as suas homólogas reduzidas forem linearmente independentes, pelo que podemos assumir que a base  $L = \{l_1, \dots, l_u\}$ , de um subespaço linear  $\mathcal{L}_1 = \mathcal{L}_1(L)$  de  $\mathcal{L}$ , é constituída por aplicações reduzidas. É facilmente verificável que o subespaço  $\mathcal{L}_1$  é  $\rho$ -saturado (visto ser uma reunião de classes de equivalência  $\rho$ ) e contém  $p^u$  aplicações, pelo que, com  $\mathcal{L}_{1,r}$ , a família de aplicações reduzidas pertencentes a  $\mathcal{L}_1$ , teremos  $\#(\mathcal{L}_{1,r}) = k_u(p)$ .

Podemos ainda definir uma relação de equivalência em  $G[p]^N$  pondo  $\mathbf{x}_1 \tau_{\mathcal{L}_1} \mathbf{x}_2$  sempre que, qualquer que seja  $l \in \mathcal{L}_1$ ,  $l(\mathbf{x}_1) = l(\mathbf{x}_2)$ . Estas classes de equivalência são os blocos

$$[L|\mathbf{b}] = \{\mathbf{x} : l_i(\mathbf{x}) = b_i; i = 1, \dots, u\}.$$

onde cada um destes  $p^u$  blocos têm cardinal  $p^{N-u}$ .

Ordenemos agora as aplicações reduzidas, atribuindo os primeiros índices às que pertencem a  $\mathcal{L}_{1,r}$ . Ordenando igualmente os vectores de  $G[p]^N$ , poderemos definir a matriz  $\mathbf{C}(l_h)$ , com elementos

$$c_{i,j}(l_h) = \begin{cases} 1 & \text{se } l_h(\mathbf{x}_j) = i - 1 \\ 0 & \text{se } l_h(\mathbf{x}_j) \neq i - 1 \end{cases}$$

com  $i = 1, \dots, p$  e  $j = 1, \dots, p^N$ .

Podemos então estabelecer o seguinte lemma.

**Lemma 4.1.**  $\mathbf{C}(l_h)\mathbf{C}(l_h)^\top = p^{N-1}\mathbf{I}_p$ , com  $h = 1, \dots, k_N(p)$  e  $\mathbf{C}(l_h)\mathbf{C}(l_k)^\top = p^{N-2}\mathbf{J}_p$  quando  $h \neq k$ .

*Demonstração.* Para cada coluna de  $\mathbf{C}(l_h)$  temos um elemento igual a 1 e os restantes são nulos, e, para cada linha de  $\mathbf{C}(l_h)$ , teremos  $p^{N-1}$  elementos iguais a 1 e os restantes serão nulos, pelo que  $\mathbf{C}(l_h)\mathbf{C}(l_h)^\top = p^{N-1}\mathbf{I}_p$ , com  $h = 1, \dots, k_N(p)$ . Por outro lado, dada uma linha de  $\mathbf{C}(l_h)$  e uma linha de  $\mathbf{C}(l_k)$ , haverá  $p^{N-2} = \#([l_h, l_k|b_1, b_2])$  correspondências entre elementos não nulos nas duas colunas  $b_1, b_2 = 1, \dots, p-1$ , pelo que  $\mathbf{C}(l_h)\mathbf{C}(l_k)^\top = p^{N-2}\mathbf{J}_p$  quando  $h \neq k$ .  $\square$

Estamos em condições para estabelecer a seguinte proposição:

**Proposição 4.3.** *A matriz*

$$\left[ \frac{1}{\sqrt{p^N}}\mathbf{1}^{p^N} \quad \mathbf{A}^\top(l_1) \quad \dots \quad \mathbf{A}^\top(l_{k_N(p)}) \right],$$

onde  $\mathbf{A}^\top(l_h) = \frac{1}{p^{(N-1)/2}}\mathbf{C}(l_h)^\top\mathbf{T}_p$ ,  $h = 1, \dots, k_N(p)$ , é uma matriz ortogonal associada à AJC  $\mathcal{A}(p^N)$  com base principal

$$bp(\mathcal{A}(p^N)) = \left\{ \frac{1}{p^N}\mathbf{J}_{p^N}, \mathbf{Q}(l_1), \dots, \mathbf{Q}(l_{k_N(p)}) \right\},$$

onde  $\mathbf{Q}(l_h) = \mathbf{A}^\top(l_h)\mathbf{A}(l_h)$ .

*Demonstração.* Considerando o enunciado do Lemma 4.1, obtemos  $\mathbf{C}(l_h)\mathbf{1}^{p^N} = p^{N-1}\mathbf{1}^p$ , pelo que  $\mathbf{A}(l_h)^\top \mathbf{1}^{p^N} = \sqrt{p^{N-1}}\mathbf{K}_p^\top \mathbf{1}^p = \mathbf{0}$ ,  $h = 1, \dots, k_N(P)$  e  $\mathbf{K}_r$  é obtida retirando a primeira linha a uma qualquer matriz normal estandardizada de ordem  $r$ . Por outro lado  $\mathbf{A}(l_h)\mathbf{A}(l_h)^\top = \frac{1}{p^{N-1}}\mathbf{K}_p^\top \mathbf{C}(l_h)\mathbf{C}(l_h)^\top \mathbf{K}_p = \mathbf{K}_p^\top \mathbf{K}_p = \mathbf{I}_{p-1}$ ,  $h = 1, \dots, k_N(p)$ . Por fim, com  $h \neq k$ ,  $\mathbf{A}(l_h)\mathbf{A}(l_k)^\top = \frac{1}{p}\mathbf{K}_p^\top \mathbf{J}_p \mathbf{K}_p = \mathbf{K}_p^\top \mathbf{1}^p \mathbf{1}^{p^\top} \mathbf{K}_p = \mathbf{I}_{p-1} = \mathbf{0}_{(p-1) \times (p-1)}$ . Então  $\mathbf{P}(p^N)$  é ortogonal, sendo de imediato o restante da demonstração.  $\square$

Podemos então escrever o modelo na forma canónica, como

$$\mathbf{y} = \frac{1}{\sqrt{rp^N}}\mathbf{1}_{rp^N}\boldsymbol{\mu} + \sum_{j=1}^{k_N(p)} \mathbf{A}(l_j)^\top \boldsymbol{\beta}(l_j) + \boldsymbol{\varepsilon}$$

onde, para aligeirar a escrita,  $\boldsymbol{\beta}(l_j)$  pode ser fixo ou aleatório. Os vectores aleatórios que entram na expressão do modelo serão normais, independentes, com matrizes de covariância  $\gamma(l)\mathbf{I}_{p-1}$  e  $\sigma^2\mathbf{I}_n$ .

### 4.3.1 Caso de efeitos fixos

Representando por  $\mathcal{L}_r$  o conjunto das aplicações reduzidas, admitindo que

$$\boldsymbol{\beta}(l), l \in \mathcal{L}_r, \text{ é fixo,}$$

e

$$\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I}_n),$$

e ainda

$$\mathbf{A}^\perp = \mathbf{I}_{p^N} \otimes \mathbf{K}_r,$$

onde  $\mathbf{K}_r$  é extraída de uma ortogonal estandardizada de ordem  $r$ , retirando a primeira linha.

Como vimos na secção 3.2,

$$S = \|\mathbf{A}^\perp \mathbf{y}\|^2,$$

ou ainda

$$S = \|\mathbf{A}^\perp \mathbf{y}\|^2 = \|\mathbf{A}^\perp \boldsymbol{\varepsilon}\|^2 \sim \sigma^2 \chi_g^2$$

onde  $g = p^N(r - 1)$ .

Facilmente se observa que se pode obter o estimador UMVUE para  $\sigma^2$  dado por

$$\tilde{\sigma}^2 = \frac{S}{g}.$$

Por outro lado, sabendo que

$$\tilde{\boldsymbol{\beta}}(l) = \left( \mathbf{A}(l)^\top \otimes \frac{1}{\sqrt{r}} \mathbf{1}_r^\top \right) \mathbf{y} \sim N(\boldsymbol{\beta}(l), \sigma^2 \mathbf{I}_{p-1}),$$

verifica-se que

$$\|\tilde{\boldsymbol{\beta}}(l) - \boldsymbol{\beta}(l)\|^2 \sim \sigma^2 \chi_{p-1}^2,$$

pelo que podemos construir a variável pivot

$$\mathcal{F}'(l) = \frac{g}{p-1} \frac{\|\boldsymbol{\beta}(l) - \tilde{\boldsymbol{\beta}}(l)\|^2}{S} \sim F(p-1, g),$$

a que se associa uma esfera de confiança, com nível de significância  $1 - \alpha$ , para  $\boldsymbol{\beta}(l)$ , definida como

$$pr \left( \|\boldsymbol{\beta}(l) - \tilde{\boldsymbol{\beta}}(l)\|^2 \leq (p-1) F_{1-\alpha, p-1, g \frac{S}{g}} \right) = p-1.$$

Por dualidade obtém-se um teste de nível  $\alpha$  que rejeita a hipótese nula

$$H_0(l_i) : \boldsymbol{\beta}(l) = \boldsymbol{\beta}_0(l)$$

quando  $\boldsymbol{\beta}_0(l)$  não está contido na esfera de confiança de nível  $1 - \alpha$ . Aliás, este teste é equivalente ao teste de nível  $\alpha$  com estatística

$$\mathcal{F}(l) = \frac{g}{p-1} \frac{\|\tilde{\boldsymbol{\beta}}(l) - \boldsymbol{\beta}_0(l)\|^2}{S},$$

que se verifica ter uma distribuição  $F(p-1, g, \delta(l))$ , com

$$\delta(l) = \frac{1}{\sigma^2} \|\boldsymbol{\beta} - \boldsymbol{\beta}_0(l)\|^2.$$

Assim este teste gozará de dualidade. Observe-se que a hipótese nula,  $H_0(l)$ , pode ser reescrita como

$$H_0(l) : \delta(l) = 0$$

pelo que este teste é estritamente não distorcido.

Assinale-se ainda que os  $\tilde{\beta}(l)$  são também UMVUE.

### 4.3.2 Caso de efeitos aleatórios

Consideremos agora o caso em que apenas se consideram efeitos aleatórios, e para o qual podemos definir

$$\beta(l) \sim N(\mathbf{0}, \gamma(l)\mathbf{I}_{p-1})$$

independentes de

$$\varepsilon \sim N(\mathbf{0}, \sigma^2\mathbf{I}_n),$$

bem como,

$$S_l = \|\mathbf{A}(l)\mathbf{y}\|^2 \sim (\gamma(l) + \sigma^2) \chi_{p-1}^2$$

independentes de

$$S = \|\mathbf{A}^\perp\mathbf{y}\|^2 \sim \sigma^2\chi_g^2.$$

Verificamos ainda que temos os estimadores UMVUE,

$$\tilde{\sigma}^2 = \frac{S}{g}$$

e

$$\tilde{\gamma}(l) = \frac{S_l}{p-1} - \frac{S}{g},$$

uma vez que

$$\begin{cases} \mathbb{E}\left(\frac{S_l}{p-1}\right) = \gamma(l) + \sigma^2 \\ \mathbb{E}\left(\frac{S_l}{p-1} - \frac{S}{g}\right) = \gamma(l) \end{cases}$$

Por outro lado, a estatística

$$\mathcal{F}(l) = \frac{g}{p-1} \frac{S_l}{S}$$

é o produto por  $\theta(l) = \frac{\sigma^2 + \gamma(l)}{\sigma^2}$ , de uma variável com distribuição  $F$  central com  $p-1$  e  $g$  graus de liberdade, podendo ser utilizada para testar as hipóteses definidas como

$$\left\{ \begin{array}{l} H_0(l) : \gamma(l) = 0 \\ vs \\ H_1(l) : \gamma(l) > 0 \end{array} \right.$$

ou

$$\left\{ \begin{array}{l} H_0(l) : \theta(l) = 1 \\ vs \\ H_1(l) : \theta(l) > 1 \end{array} \right. .$$

Como

$$\mathcal{F}(l) \sim F \left( \frac{z}{\theta(l)} \mid p-1, g \right)$$

este teste é estritamente não distorcido.

### 4.3.3 Caso de efeitos mistos

Neste caso  $\mathcal{L}^\circ = \mathcal{L}_f \cup \mathcal{L}_a$ . Tendo-se  $\beta(l)$  fixo caso  $l \in \mathcal{L}_f$  e  $\beta(l) \sim N(\mathbf{0}, \sigma^2(l)\mathbf{I}_{p-1})$  caso  $l \in \mathcal{L}_a$ . Consoante  $l \in \mathcal{L}_f$  ou  $l \in \mathcal{L}_a$  aplicam-se os resultados apresentados nas secções anteriores.

### 4.3.4 Repetições

Consideremos agora  $r$ , o número de repetições do modelo base, o qual está estritamente associado a uma AJC. Admite-se que as repetições correspondem a uma parte que não interagem com os factores do modelo base. Havendo  $n^\circ$  tratamentos

no modelo base, podemos, numa primeira fase, admitir que temos um modelo com dois factores que cruzam com  $n^\circ$  e  $r$  níveis. Este modelo estará associado à AJC

$$\mathcal{A}(n^\circ) \otimes \mathcal{A}(r)$$

tendo esta matriz a base principal

$$\left\{ \frac{1}{n^\circ} \mathbf{J}_{n^\circ} \otimes \frac{1}{r} \mathbf{J}_r; \frac{1}{n^\circ} \mathbf{J}_{n^\circ} \otimes \bar{\mathbf{J}}_r; \bar{\mathbf{J}}_{n^\circ} \otimes \frac{1}{r} \mathbf{J}_r; \bar{\mathbf{J}}_{n^\circ} \otimes \bar{\mathbf{J}}_r \right\}.$$

Por outro lado, o modelo base estará estritamente associado a uma AJC com base principal  $\{\mathbf{Q}_1, \dots, \mathbf{Q}_w\}$ , tendo-se

$$\mathbf{Q}_1 = \frac{1}{n^\circ} \mathbf{J}_{n^\circ}$$

e

$$\sum_{j=2}^w \mathbf{Q}_j = \bar{\mathbf{J}}_{n^\circ}.$$

É agora fácil de verificar que as matrizes pertencentes a

$$\left\{ \frac{1}{n^\circ} \mathbf{J}_{n^\circ} \otimes \frac{1}{r} \mathbf{J}_r; \frac{1}{n^\circ} \mathbf{J}_{n^\circ} \otimes \bar{\mathbf{J}}_r; \mathbf{Q}_2 \otimes \frac{1}{r} \mathbf{J}_r, \dots, \mathbf{Q}_w \otimes \frac{1}{r} \mathbf{J}_r; \bar{\mathbf{J}}_{n^\circ} \otimes \bar{\mathbf{J}}_r \right\}$$

são MPO mutuamente ortogonais constituindo a base principal duma AJC  $\mathcal{A}^\circ$ . O modelo com repetições estará estritamente associado à AJC,  $\mathcal{A}^\circ$ .

Definindo-se

$$\mathbf{Q}_j = \mathbf{A}_j^\top \mathbf{A}_j, \quad j = 1, \dots, w,$$

teremos

$$\left\{ \begin{array}{l} \frac{1}{n^\circ} \mathbf{J}_{n^\circ} \otimes \frac{1}{r} \mathbf{J}_r = \left( \frac{1}{\sqrt{n^\circ}} \mathbf{1}_{n^\circ} \otimes \frac{1}{\sqrt{r}} \mathbf{1}_r \right) \left( \frac{1}{\sqrt{n^\circ}} \mathbf{1}_{n^\circ}^\top \otimes \frac{1}{\sqrt{r}} \mathbf{1}_r^\top \right) \\ \frac{1}{n^\circ} \mathbf{J}_{n^\circ} \otimes \bar{\mathbf{J}}_r = \left( \frac{1}{\sqrt{n^\circ}} \mathbf{1}_{n^\circ} \otimes \mathbf{K}_r^\top \right) \left( \frac{1}{\sqrt{n^\circ}} \mathbf{1}_{n^\circ}^\top \otimes \mathbf{K}_r \right) \\ \mathbf{Q}_j \otimes \frac{1}{r} \mathbf{J}_r = \left( \mathbf{A}_j^\top \otimes \frac{1}{\sqrt{r}} \mathbf{1}_r \right) \left( \mathbf{A}_j \otimes \frac{1}{\sqrt{r}} \mathbf{1}_r^\top \right) \quad j = 2, \dots, w \\ \bar{\mathbf{J}}_{n^\circ} \otimes \bar{\mathbf{J}}_r = (\mathbf{T}_{n^\circ}^\top \otimes \mathbf{T}_r^\top) (\mathbf{K}_{n^\circ} \otimes \mathbf{K}_r) \end{array} \right.$$

podendo agora seguir-se o mesmo método ao atrás apresentado. Observe-se que se pode admitir que as repetições são consideradas como componente fixa, estando-lhe associado as matrizes  $\frac{1}{n^\circ} \mathbf{J}_{n^\circ} \otimes \bar{\mathbf{J}}_r$  e  $\frac{1}{\sqrt{n^\circ}} \mathbf{1}_{n^\circ}^\top \otimes \mathbf{K}_r$ . Ao valor médio geral estarão associadas as matrizes  $\frac{1}{n^\circ} \mathbf{J}_{n^\circ} \otimes \bar{\mathbf{J}}_r$  e  $\frac{1}{\sqrt{n^\circ}} \mathbf{1}_{n^\circ}^\top \otimes \mathbf{K}_r$ , aos efeitos das partes do modelo base as matrizes  $\mathbf{Q}_j \otimes \frac{1}{r} \mathbf{J}_r$  e  $\mathbf{A}_j \otimes \frac{1}{\sqrt{r}} \mathbf{1}_r^\top$ ,  $j = 2, \dots, w$  e finalmente ao erro, as matrizes  $\bar{\mathbf{J}}_{n^\circ} \otimes \bar{\mathbf{J}}_r$  e  $\mathbf{K}_{n^\circ} \otimes \mathbf{K}_r$ .

Por exemplo, se o modelo base for um factorial  $p^N$ , as  $\mathbf{Q}_j \otimes \frac{1}{r} \mathbf{J}_r$  serão os produtos de Kronecker das matrizes correspondentes às aplicações lineares reduzidas por  $\frac{1}{r} \mathbf{J}_r$ .

### 4.3.5 Aplicação: Estudo do efeito da lidocaína no músculo cardíaco de uma raça de cães

A aplicação que vamos considerar, corresponde a um delineamento do tipo  $3^3$ , cujo tratamento clássico pode ser visualizado na Figura 4.1.

O objectivo deste delineamento consiste em determinar o efeito, ao nível enzimático, da lidocaína (fármaco do grupo dos antiarrítmicos e dos anestésicos locais que é usado no tratamento da arritmia cardíaca), no músculo cardíaco de cães de raça *beagles*. Os dados aqui utilizados são apresentados em Montgomery (2005) [33],

pag. 388.

Para efeitos desta experiência considera-se a existência de 3 marcas de lidocaína diferentes no mercado e considera-se ainda a administração de 3 doses deste fármaco. Utilizaram-se 3 cães da raça *beagle* e 2 repetições da experiência. Estamos pois perante um delineamento factorial de base prima  $3^3$ . Na Tabela 4.1 encontram-se os resultados experimentais.

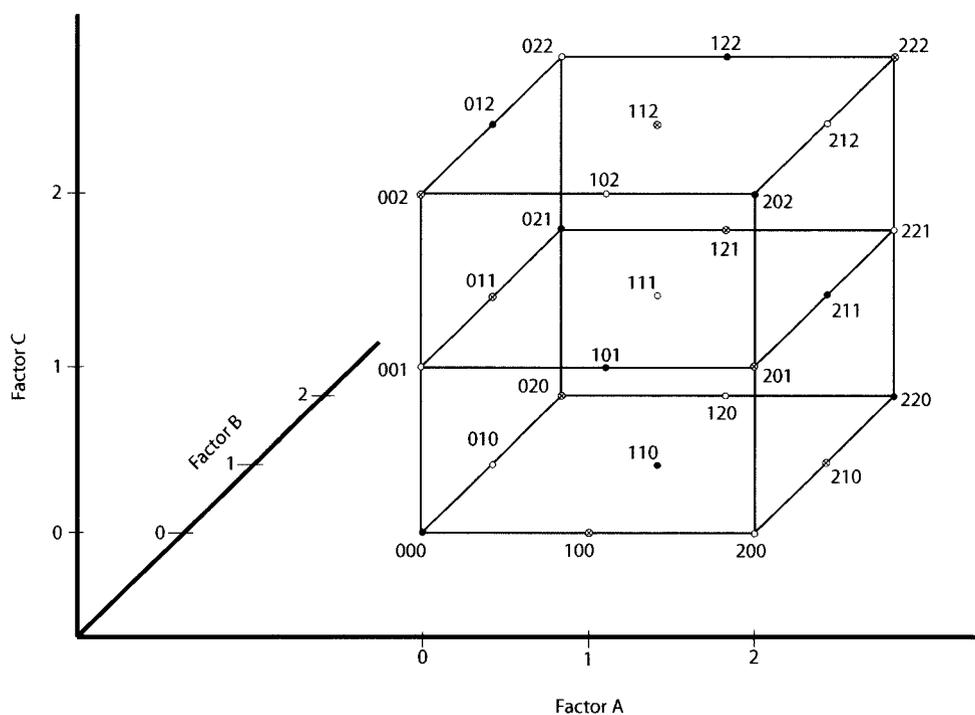


Figura 4.1: Combinações de tratamento num delineamento  $3^3$

Da estrutura dos dados, vamos considerar a existência de 2 factores fixos, marca e dose, e um efeito aleatório, o animal. De acordo com a teoria desenvolvida, atribuiremos o índice 1 à marca, o índice 2 à dose e o índice 3 ao animal.

Os resultados foram obtidos com o recurso ao *software* MatLab 7.0.4.

Na Tabela 4.2 apresentam-se os testes  $F$  e os respectivos  $p$  – values.

A análise da Tabela 4.2 mostra que apenas o factor *dose* ( $x_2$ ) e a respectiva interacção linear ( $x_1 + x_2$ ) com o factor *marca* são significativos.

Tabela 4.1: Níveis da enzima

Marca Dosagem		Repetição 1			Repetição 2		
		Cão			Cão		
		1	2	3	1	2	3
1	1	86	84	85	84	85	86
	2	94	99	98	95	97	90
	3	101	106	98	105	104	103
2	1	85	84	86	80	82	84
	2	95	98	97	93	99	95
	3	108	114	109	110	102	100
3	1	84	83	81	83	80	79
	2	95	97	93	92	96	93
	3	105	100	106	102	111	108

Tabela 4.2: Teste  $F$  e  $p$  - *values* da experiência

Aplicação Reduzida	Soma dos Quadrados	Teste $F$	$p$ - <i>value</i>
$x_1$	31,00	2,59	0,0936
$x_2$	4260,80	357,03	$\simeq 0$
$x_1 + x_2$	51,42	4,31	0,0242
$x_1 + 2x_2$	18,14	1,52	0,2375
$x_3$	28,00	2,35	0,1157
$x_1 + x_3$	0,83	0,07	0,9327
$x_1 + 2x_3$	2,50	0,21	0,8124
$x_2 + x_3$	25,50	2,14	0,1383
$x_2 + 2x_3$	11,39	0,95	0,3982
$x_1 + x_2 + x_3$	25,25	2,21	0,1408
$x_1 + x_2 + 2x_3$	4,42	0,37	0,6943
$x_1 + 2x_2 + x_3$	4,03	0,34	0,7166
$x_1 + 2x_2 + 2x_3$	27,08	2,27	0,1235

Centremos agora o estudo no factor dose. A variação atribuível ao factor dose é dominante, pelo que justifica a sua análise em particular. Os totais correspondentes aos três níveis de dose são:

$$T_0 = 1501$$

$$T_1 = 1716$$

$$T_2 = 1892$$

obtendo-se a partir destes os efeitos linear e quadrático dados por

$$EL = \frac{T_2 - T_0}{\sqrt{2 \times 3^2 \times 2}} = 65,17$$

e

$$EQ = \frac{T_2 - 2T_1 + T_0}{\sqrt{6 \times 3^2 \times 2}} = -3,75.$$

Os valores dos testes  $F$  para estes efeitos foram  $\mathcal{F}_{EL} = 714,02$  e  $\mathcal{F}_{EQ} = 2,36$ , onde apenas o primeiro dos quais é significativo ( $p - value \approx 0$ ). Assim pode concluir-se pelo aumento linear do teor de lidocaína com a dose.

A conclusão é consentânea com o que se obtém utilizando a metodologia clássica no tramento deste tipo de delineamento. A vantagem no uso da metodologia aqui desenvolvida, prende-se com o número de interações que é possível estudar em simultâneo. Para efeitos comparativos, apresentam-se de seguida, os resultados que se obtém usando a metodologia proposta em Montgomery (2005) [33] (Tabela 4.3).

A integração deste modelo nos COBS permitiu considerar o terceiro factor (*cães*) como tendo efeitos aleatórios, tornando o modelo mais realista.

Tabela 4.3: Teste  $F$  e  $p$  – values

Aplicação Reduzida	Soma dos Quadrados	Teste $F$	$p$ – value
$x_1$	31,00	2,59	0,0936
$x_2$	4260,80	357,03	$\simeq 0$
$x_1 + x_2$	51,42	4,31	0,0242
$x_3$	28,00	2,35	0,1157
$x_1 + x_3$	0,83	0,07	0,9327
$x_2 + x_3$	25,50	2,14	0,1383
$x_1 + x_2 + x_3$	25,25	2,21	0,1408

# Capítulo 5

## Modelos com Cross-Nesting Equilibrado

No delineamento experimental, é usual considerar-se modelos onde os factores aninham ou cruzam. Tais modelos podem ser tratados como sendo modelos com COBS, com vantagens para esta metodologia. Para melhor exemplificar a sua utilização, iremos aplicar a metodologia aqui desenvolvida a um exemplo.

Uma das questões elementares que servirão de base para a forma como vamos tratar os dados que servirão de base à aplicação, prende-se com a definição de factores, níveis e efeitos. Para o efeito usaremos como referência McCulloch *et al* (2008) [4].

O interesse destes modelos pode ser expresso na forma como podemos estudar, por exemplo, a variabilidade que poderá existir no seus factores. Neste sentido consideram-se factores as classificações que identificam a origem de cada variável a estudar. A classe individual de uma classificação denomina-se por nível. Interessa igualmente distinguir o que é o aninhamento (*nesting*) e o que é cruzamento (*crossing*). Para o efeito considere-se o exemplo apresentado em McCulloch *et al*, 2008, p. 3.

Neste exemplo, as classificações, sexo, idade e exposição solar, são denominados

Tabela 5.1: Exemplo 1

Sexo	Baixa exposição solar			Alta exposição solar		
	grupo etário			grupo etário		
	A	B	C	A	B	C
Masculino						
Feminino						

Tabela 5.2: Exemplo 2

Sexo	Inglês			Geologia		
	secção			secção		
	A	B	C	A	B	C
Masculino						
Feminino						

por factores e as classes individuais por níveis e a idade e exposição solar são cruzados (*crossing*), ver Tabela 5.1.

No segundo exemplo, a "secção" está aninhada (*nesting*) na disciplina, ver Tabela 5.2.

Ao classificar os dados em termos de factores e níveis, o interesse passa a ser a extensão para os quais diferentes níveis de um factor afectam a variável de interesse.

Modelos onde os únicos efeitos são fixos, denominam-se de modelos de efeitos fixos, e aqueles, que para além da média geral, comum a todas as observações, só tenham efeitos aleatórios, denominam-se modelos de efeitos aleatórios.

## 5.1 Estrutura Algébrica

Nesta secção, e para a apresentação da estrutura algébrica, seguiremos em Fonseca *et al* (2003)[14].

Começemos por considerar um grupo de  $w$  factores em que o primeiro com  $a(1)$  níveis aninha o segundo, havendo  $a(2)$  níveis do segundo factor para cada nível do primeiro e assim sucessivamente.

Ponhamos  $c(0) = 1$  e  $c(h) = \prod_{k=1}^h a(k)$ ,  $h = 1, \dots, w$ , bem como  $b(h) = \frac{c(w)}{c(h)}$ ,  $h = 1, \dots, w$ . Vê-se que há  $c(h)$  combinações de níveis dos  $h$  primeiros factores, cada uma das quais encaixa  $b(h)$  combinações de níveis dos restantes factores.

Consideremos ainda

$$\mathbf{X}(h) = \mathbf{I}_{c(h)} \otimes \mathbf{1}_{b(h)} \quad ; h = 0, \dots, w,$$

de onde obtemos as matrizes

$$\mathbf{M}(h) = \mathbf{X}(h)\mathbf{X}(h)^\top = \mathbf{I}_{c(h)} \otimes \mathbf{J}_{b(h)} \quad ; h = 0, \dots, w.$$

A MPO sobre  $R(\mathbf{X}(h)) = R(\mathbf{M}(h))$  é

$$\mathbf{Q}(h) = \frac{1}{b(h)} \mathbf{M}(h) \quad ; h = 0, \dots, w,$$

tendo-se  $c(h) = \text{car}(\mathbf{M}(h)) = \text{car}(\mathbf{Q}(h))$ .

É fácil de verificar que a base principal da AJC

$$\mathcal{A}^\circ = \bigstar_{k=1}^w \mathcal{A}(a_k)$$

é constituída pelas matrizes

$$\mathbf{Q}^\circ(0) = \mathbf{Q}(0)$$

e

$$\mathbf{Q}^\circ(h) = \mathbf{Q}(h) - \mathbf{Q}(h-1) \quad ; h = 1, \dots, w,$$

tendo-se

$$\mathbf{Q}(h) = \sum_{k=0}^h \mathbf{Q}^\circ(k),$$

pelo que, com  $g(k) = \text{car}(\mathbf{Q}(k))$ ,  $k = 0, \dots, w$ ,

$$c(h) = \sum_{k=0}^h g(k)$$

vindo

$$g(h) = c(h) - c(h-1), \quad h = 1, \dots, w,$$

e

$$g(0) = 1,$$

já que  $\mathbf{Q}^\circ(0) = \mathbf{Q}(0)$ .

Quando temos  $L$  grupos com  $w_1, \dots, w_L$  factores que aninham e não se considerem interacções entre factores do mesmo grupo, temos apenas os efeitos dos factores e as interacções entre conjuntos de factores pertencentes a grupos distintos. Esses conjuntos podem ser identificados pelos vectores  $\mathbf{h}$  com componentes  $h_l = 0, \dots, w_l$ ,  $l = 1, \dots, L$ . Se  $h_l = 0$  não se toma nenhum factor do grupo  $l$ , caso contrário  $h_l$  será o índice do factor desse grupo que pertence ao conjunto.

Sendo  $\Gamma$  o conjunto de vectores  $\mathbf{h}$ , caso se tenham  $r$  observações por tratamento, o modelo pode ser escrito como, ver Fonseca *et al* (2003)[14],

$$\mathbf{y} = \sum_{\mathbf{h} \in \Gamma} \mathbf{X}(\mathbf{h})\boldsymbol{\beta}(\mathbf{h}) + \boldsymbol{\varepsilon} \quad (5.1)$$

onde

$$\mathbf{X}(\mathbf{h}) = \bigotimes_{l=1}^L \mathbf{X}_l(h_l) \otimes \mathbf{1}_r,$$

com  $\mathbf{X}_l(h_l)$  obtida como atrás se indicou para um grupo de factores e  $\bigotimes$  representa os produtos de Kronecker.

Definamos agora

$$c(\mathbf{h}) = \prod_{l=1}^L c_l(h_l),$$

$$g(\mathbf{h}) = \prod_{l=1}^L g_l(h_l)$$

e

$$b(\mathbf{h}) = r \prod_{l=1}^L b_l(h_l).$$

Vê-se que  $c(\mathbf{h})$  é simultaneamente a característica de  $\mathbf{X}(\mathbf{h})$  e das matrizes  $\mathbf{M}(\mathbf{h}) = \mathbf{X}(\mathbf{h})\mathbf{X}(\mathbf{h})^\top$ .

Temos ainda

$$\mathbf{Q}(\mathbf{h}) = \frac{1}{b(\mathbf{h})} \mathbf{M}(\mathbf{h}) \quad ; \mathbf{h} \in \Gamma,$$

a MPO sobre  $R(\mathbf{M}(\mathbf{h}))$  e

$$\mathbf{Q}^\circ(\mathbf{h}) = \bigotimes_{l=1}^L \mathbf{Q}_l^\circ(h_l) \otimes \frac{1}{r} \mathbf{J}_r \quad ; \mathbf{h} \in \Gamma.$$

sendo  $c(\mathbf{h})$  e  $g(\mathbf{h})$  as características de  $\mathbf{Q}(\mathbf{h})$  e  $\mathbf{Q}^\circ(\mathbf{h})$ , respectivamente.

Estas últimas matrizes, juntamente com

$$\mathbf{Q}^\perp = \bigotimes_{l=1}^L \mathbf{I}_{c_l(w_l)} \otimes \left( \mathbf{I}_r - \frac{1}{r} \mathbf{J}_r \right)$$

constituem a base principal de

$$\mathcal{A}^{\circ\circ} = \left( \bigotimes_{l=1}^L \mathcal{A}_l \right) \star \mathcal{A}(r).$$

Dado que factores de efeitos aleatórios não encaixam factores de efeitos fixos, apenas os  $v_l$ ,  $l = 1, \dots, L$  primeiros factores dos vários grupos, podem ter efeitos fixos, tendo-se  $0 \leq v_l \leq w_l$ ,  $l = 1, \dots, L$ . Assim os efeitos e as interacções entre factores de efeitos fixos corresponderão aos vectores  $\mathbf{h} \leq \mathbf{v}$  com  $\mathbf{v}$  o vector com componentes  $v_l$ ,  $l = 1, \dots, L$ . Os  $\beta(\mathbf{h})$  correspondentes, com  $\mathbf{h} \leq \mathbf{v}$  serão pois fixos, admitindo-se, seguindo Fonseca *et al* (2003)[14], que os  $\beta(\mathbf{h})$  com  $\mathbf{h} \not\leq \mathbf{v}$  serão normais com vectores médios nulos e matrizes de variância-covariância  $\sigma^2(\mathbf{h})\mathbf{I}_{c(\mathbf{h})}$ . Assim  $\mathbf{y}$  será normal com vector médio

$$\boldsymbol{\mu} = \sum_{\mathbf{h} \leq \mathbf{v}} \mathbf{X}(\mathbf{h})\beta(\mathbf{h})$$

e matriz de variância-covariância

$$\mathbf{V} = \sum_{\mathbf{h} \neq \mathbf{v}} \sigma^2(\mathbf{h}) \mathbf{M}(\mathbf{h}) + \sigma^2 \mathbf{I}_n,$$

com

$$n = r \prod_{l=1}^L c_l(w_l).$$

Ora se  $\mathbf{h}_1 \leq \mathbf{h}_2$ , o espaço imagem de  $\mathbf{X}(\mathbf{h}_1)$  está contido no espaço imagem de  $\mathbf{X}(\mathbf{h}_2)$  pelo que o espaço varrido por  $\boldsymbol{\mu}$  será o espaço imagem de  $\mathbf{X}(\mathbf{v})$  que coincide com o espaço imagem de  $\mathbf{M}(\mathbf{v})$ . Como

$$\mathbf{M}(\mathbf{v}) = b(\mathbf{v}) \mathbf{Q}(\mathbf{v}) = b(\mathbf{v}) \sum_{\mathbf{h} \leq \mathbf{v}} \mathbf{Q}^\circ(\mathbf{h})$$

vê-se que

$$\mathbf{T} = \sum_{\mathbf{h} \leq \mathbf{v}} \mathbf{Q}^\circ(\mathbf{h}).$$

Por outro lado,

$$\mathbf{V} = \sum_{\mathbf{h} \in \Gamma} \gamma(\mathbf{h}) \mathbf{Q}^\circ(\mathbf{h}) + \sigma^2 \mathbf{Q}^\perp,$$

com

$$\gamma(\mathbf{h}) = \sigma^2 + \sum_{\mathbf{k}: \mathbf{h} \leq \mathbf{k}} b(\mathbf{k}) \sigma^2(\mathbf{k}),$$

pelo que estes modelos têm COBS, ver Fonseca *et al* (2003)[14].

Sendo  $\Theta(\mathbf{h})$  o conjunto dos  $\mathbf{k} \in \Gamma$  tais que  $h_l \leq k_l \leq h_l + 1$ , tem-se, ver Fonseca *et al* (2003)[14],

$$\sigma^2(\mathbf{h}) = \sum_{\mathbf{k} \in \Theta \mathbf{h}} (-1)^{n(\mathbf{h}, \mathbf{k})} \frac{\gamma(\mathbf{k})}{b(\mathbf{k})},$$

com  $n(\mathbf{h}, \mathbf{k})$  o número de componentes de  $\mathbf{h}$  inferiores às correspondentes componentes de  $\mathbf{k}$ .

Tem-se igualmente

$$\left\{ \begin{array}{l} u_1 = \prod_{l=1}^L (v_l + 1) \\ u_2 = 0 \\ u_3 = \prod_{l=1}^L (w_l + 1) - \prod_{l=1}^L (v_l + 1), \end{array} \right.$$

visto que, se  $\mathbf{k} \leq \mathbf{v}$  [ $\mathbf{k} \not\leq \mathbf{v}$ ], o espaço imagem de  $\mathbf{Q}(\mathbf{h})$  está contido em  $\Omega$  [ $\Omega^\perp$ ].

Por outro lado, com

$$\left\{ \begin{array}{l} \mathbf{A}_l(0) = \frac{1}{\sqrt{c_l(w_l)}} \mathbf{1}_{c_l(w_l)}^\top \quad ; l = 1, \dots, L \\ \mathbf{A}_l(h) = \mathbf{I}_{c_l(h_l-1)} \otimes \mathbf{K}_{a_l(h_l)} \otimes \frac{1}{\sqrt{b_l(h_l)}} \mathbf{1}_{b_l(h_l)}^\top \quad h = 1, \dots, w_l; l = 1, \dots, L, \end{array} \right.$$

onde  $\mathbf{K}_s$  é a matriz obtida eliminando a primeira linha numa matriz ortogonal estandarizada de ordem  $s$ .

Tomando-se

$$\left\{ \begin{array}{l} \mathbf{A}(\mathbf{h}) = \bigotimes_{l=1}^L \mathbf{A}_l(h_l) \otimes \frac{1}{\sqrt{r}} \mathbf{1}_r^\top \quad ; \mathbf{h} \in \Gamma \\ \mathbf{A}^\perp = \bigotimes_{l=1}^L \mathbf{I}_{c_l(w_l)} \otimes \mathbf{K}_r, \end{array} \right.$$

tem-se, ver Fonseca *et al* (2003)[14],

$$\left\{ \begin{array}{l} \mathbf{Q}(\mathbf{h}) = \mathbf{A}(\mathbf{h})^\top \mathbf{A}(\mathbf{h}) \quad \mathbf{h} \in \Gamma \\ \mathbf{Q}^\perp = \mathbf{A}^{\perp\top} \mathbf{A}^\perp, \end{array} \right.$$

como também as estatísticas suficientes e completas dadas por

$$\begin{cases} \tilde{\boldsymbol{\eta}}(\mathbf{h}) = \mathbf{A}(\mathbf{h})\mathbf{y} & \mathbf{h} \leq \mathbf{v} \\ S(\mathbf{h}) = \|\mathbf{A}(\mathbf{h})\mathbf{y}\|^2 & \mathbf{h} \not\leq \mathbf{v} \\ S = \|\mathbf{A}^{\perp}\mathbf{y}\|^2, \end{cases}$$

a partir das quais se obtêm os UMVUE  $\tilde{\boldsymbol{\eta}}$ ,  $\mathbf{h} \leq \mathbf{v}$ , para os  $\boldsymbol{\eta}(\mathbf{h}) = \mathbf{A}(\mathbf{h})\boldsymbol{\mu}$ ,  $\mathbf{h} \leq \mathbf{v}$ ,

$$\tilde{\boldsymbol{\eta}}(\mathbf{h}) = \mathbf{A}\mathbf{y}$$

e

$$\tilde{\gamma}(\mathbf{h}) = \frac{S(\mathbf{h})}{g(\mathbf{h})} \quad ; \mathbf{h} \in \Gamma$$

para os  $\gamma(\mathbf{h})$ ,  $\mathbf{h} \in \Gamma$ . A partir destes últimos, obtêm-se ainda os

$$\tilde{\sigma}^2(\mathbf{h}) = \sum_{\mathbf{k} \in \Theta(\mathbf{h})} (-1)^{n(\mathbf{h},\mathbf{k})} \frac{S(\mathbf{k})}{g(\mathbf{k})} \quad ; \mathbf{h} \not\leq \mathbf{v}.$$

Tomando-se  $\tilde{\sigma}^2(\mathbf{h}) = 0$  para  $\mathbf{h} \leq \mathbf{v}$ , temos por último

$$\begin{aligned} \tilde{\gamma}(\mathbf{h}) &= \tilde{\sigma}^2 + \sum_{\mathbf{k}:\mathbf{h} \leq \mathbf{v}} b(\mathbf{k})\tilde{\sigma}^2(\mathbf{k}) \\ &= \tilde{\sigma}^2 + \sum_{\mathbf{k}:\mathbf{h} \leq \mathbf{v}} b(\mathbf{k}) \sum_{\mathbf{l} \in \Theta(\mathbf{k})} (-1)^{n(\mathbf{k},\mathbf{l})} \tilde{\gamma}(\mathbf{l}); \quad \mathbf{h} \leq \mathbf{v} \end{aligned}$$

podendo reescrever-se estes estimadores sob a forma

$$\tilde{\gamma}(\mathbf{h}) = \tilde{\sigma}^2 + \sum_{\mathbf{l} \in \mathcal{D}^+(\mathbf{h})} c(\mathbf{l})\tilde{\gamma}(\mathbf{l}) - \sum_{\mathbf{l} \in \mathcal{D}^-(\mathbf{h})} c(\mathbf{l})\tilde{\gamma}(\mathbf{l}),$$

onde  $\mathcal{D}^+$  e  $\mathcal{D}^-$  são as famílias das  $\mathbf{l}$  parcelas que entram com coeficientes positivos e negativos, respectivamente, na expressão de  $\tilde{\gamma}(\mathbf{h})$ .

### 5.1.1 Testes

Começando pela parte de efeitos fixos do modelo 5.1, interessa-nos testar a hipótese nula,

$$H_0(\mathbf{h}) : \boldsymbol{\eta}(\mathbf{h}) = \boldsymbol{\eta}_0(\mathbf{h}); \quad \mathbf{h} \leq \mathbf{v}.$$

Como  $\tilde{\boldsymbol{\eta}}(\mathbf{h}) \sim N(\boldsymbol{\eta}(\mathbf{h}), \gamma(\mathbf{h})\mathbf{I}_{g(\mathbf{h})})$ , teremos

$$\tilde{\boldsymbol{\eta}}(\mathbf{h}) - \boldsymbol{\eta}_0(\mathbf{h}) \sim N(\boldsymbol{\eta}(\mathbf{h}) - \boldsymbol{\eta}_0(\mathbf{h}), \gamma(\mathbf{h})\mathbf{I}_{g(\mathbf{h})}),$$

pelo que,

$$\|\tilde{\boldsymbol{\eta}}(\mathbf{h}) - \boldsymbol{\eta}_0(\mathbf{h})\|^2 \sim \gamma(\mathbf{h}) \chi_{g(\mathbf{h}), \delta(\mathbf{h})}^2$$

com

$$\delta(\mathbf{h}) = \frac{1}{\gamma(\mathbf{h})} \|\boldsymbol{\eta}(\mathbf{h}) - \boldsymbol{\eta}_0(\mathbf{h})\|^2.$$

Ora a hipótese nula,  $H_0(\mathbf{h})$  para  $\mathbf{h} \leq \mathbf{v}$ , pode ser reescrita como

$$H_0(\mathbf{h}) : \gamma(\mathbf{h})\delta(\mathbf{h}) = 0; \quad \mathbf{h} \leq \mathbf{v}$$

o que nos permite utilizar os testes  $\mathcal{F}$  generalizados para testar a hipótese nula  $H_0(\mathbf{h}); \mathbf{h} \leq \mathbf{v}$  (ver Michalski & Zmyslony (1996)[31] e (1999)[32]).

Como

$$\begin{aligned} \mathbb{E}(\|\tilde{\boldsymbol{\eta}}(\mathbf{h}) - \boldsymbol{\eta}_0(\mathbf{h})\|^2) &= \gamma(\mathbf{h})(g(\mathbf{h}) + \delta(\mathbf{h})) \\ &= \gamma(\mathbf{h})g(\mathbf{h}) + \gamma(\mathbf{h})\delta(\mathbf{h}) \end{aligned}$$

para  $\lambda(\mathbf{h}) = \gamma(\mathbf{h})\delta(\mathbf{h})$  tem-se o estimador quadrático centrado

$$\begin{aligned} \tilde{\lambda}(\mathbf{h}) &= \|\tilde{\boldsymbol{\eta}}(\mathbf{h}) - \boldsymbol{\eta}_0(\mathbf{h})\|^2 - g(\mathbf{h})\tilde{\gamma}(\mathbf{h}) \\ &= \|\tilde{\boldsymbol{\eta}}(\mathbf{h}) - \boldsymbol{\eta}_0(\mathbf{h})\|^2 + g(\mathbf{h}) \sum_{\mathbf{l} \in \mathcal{D}^-(\mathbf{h})} c(\mathbf{l})\tilde{\gamma}(\mathbf{l}) - g(\mathbf{h}) \sum_{\mathbf{l} \in \mathcal{D}^+(\mathbf{h})} c(\mathbf{l})\tilde{\gamma}(\mathbf{l}), \end{aligned}$$

tendo-se a estatística  $\mathcal{F}$  generalizada

$$\mathcal{F}(\mathbf{h}) = \frac{\|\tilde{\boldsymbol{\eta}}(\mathbf{h}) - \boldsymbol{\eta}_0(\mathbf{h})\|^2 + g(\mathbf{h}) \sum_{\mathbf{l} \in \mathcal{D}^-(\mathbf{h})} c(\mathbf{l})\tilde{\gamma}(\mathbf{l})}{g(\mathbf{h}) \sum_{\mathbf{l} \in \mathcal{D}^+(\mathbf{h})} c(\mathbf{l})\tilde{\gamma}(\mathbf{l})}.$$

Outra abordagem é a seguida por Ferreira (2005)[9] baseada na utilização das variáveis pivot indutoras.

Considerando agora a parte de efeitos aleatórios do modelo, temos as hipóteses nulas,

$$H_0(\mathbf{h}) : \sigma^2(\mathbf{h}) = 0; \quad \mathbf{h} \not\leq \mathbf{v}$$

para as quais podemos utilizar testes F generalizados da mesma forma.

Sabendo que

$$\tilde{\sigma}(\mathbf{h}) = \sum_{\mathbf{k} \in \Theta^+(\mathbf{h})} \frac{S(\mathbf{k})}{g(\mathbf{k})} - \sum_{\mathbf{k} \in \Theta^-(\mathbf{h})} \frac{S(\mathbf{k})}{g(\mathbf{k})},$$

onde  $\Theta^+(\mathbf{h})[\Theta^-(\mathbf{h})]$  é a família dos  $\mathbf{k} \in \Theta(\mathbf{h})$  com  $n(\mathbf{h}, \mathbf{k})$  par [ímpar]. Logo, podemos então obter a estatística a usar nos testes definida como

$$\mathcal{F}(\mathbf{h}) = \frac{\sum_{\mathbf{k} \in \Theta^+(\mathbf{h})} \frac{S(\mathbf{k})}{g(\mathbf{k})}}{\sum_{\mathbf{k} \in \Theta^-(\mathbf{h})} \frac{S(\mathbf{k})}{g(\mathbf{k})}}.$$

As distribuições destas estatísticas foram obtidas, caso os graus de liberdade do numerador ou do denominador, sejam todos pares, ver Fonseca *et al* (2002)[13]. Por outro lado, também se pode utilizar variáveis pivot indutoras para realizar estes testes, ver Ferreira (2006) [11].

Na secção 5.3, e de forma a exemplificar a metodologia aqui desenvolvida, apresentaremos uma aplicação, na área da odontologia, onde as vantagens desta metodologia serão demonstradas.

## 5.2 Operações Binárias

Nos modelos com *cross-nesting* equilibrado podemos, através de cruzamento e aninhamento, construir modelos complexos a partir de modelos simples, o que se reveste de uma vantagem notória sobre outras metodologias.

Começando pelo caso em que há um único grupo de aninhamento de  $w$  factores com  $a(h)$ ,  $h = 1, \dots, w$ , níveis.

Consideremos assim a seguinte proposição.

**Proposição 5.1.** *As matrizes  $\mathbf{Q}^\circ(0), \dots, \mathbf{Q}^\circ(w)$  constituem a base principal da AJC  $\bigstar_{u=1}^w \mathcal{A}(a_u)$ .*

*Demonstração.* Considere-se que

$$\mathbf{Q}^\circ(0) = \frac{1}{c(w)} \mathbf{J}_{c(w)} = \bigotimes_{u=1}^w \frac{1}{a(u)} \mathbf{J}_{a(u)},$$

bem como com  $a(u) = a(w+1) = 1$ ,

$$\begin{aligned} \mathbf{Q}^\circ(h) &= \mathbf{I}_{c(h)} \otimes \frac{1}{b(h)} \mathbf{J}_{b(h)} - \mathbf{I}_{c(h-1)} \otimes \frac{1}{b(h-1)} \mathbf{J}_{b(h-1)} \\ &= \mathbf{I}_{c(h-1)} \otimes \left( \mathbf{I}_{a(h)} - \frac{1}{a(h)} \mathbf{J}_{b(h)} \right) \otimes \frac{1}{b(h)} \mathbf{J}_{b(h)} \\ &= \bigotimes_{u=0}^m h-1 \mathbf{I}_{a(u)} \otimes \bar{\mathbf{J}}_{b(h)} \otimes \bigotimes_{u=h+1}^{w+1} \frac{1}{a(u)} \mathbf{J}_{a(u)}. \end{aligned}$$

É agora fácil de verificar que as matrizes  $\mathbf{Q}^\circ(u)$ ,  $u = 0, \dots, w$ , vão sendo construídas pelo processo que dá a base principal de  $\bigstar_{u=1}^w \mathcal{A}(a(u))$ .  $\square$

Por outro lado as matrizes  $\mathbf{M}(h)$ ,  $h = 0, \dots, w$  constituem uma outra base para  $\bigstar_{u=1}^w \mathcal{A}(a(u))$ , tendo-se a matriz de transição  $\bigstar_{u=1}^w \mathbf{B}(a(u))$ , onde

$$\mathbf{B} = \begin{bmatrix} \mathbf{a} & \mathbf{0} \\ \mathbf{1} & \mathbf{1} \end{bmatrix}.$$

Consideremos agora a situação em que se cruzam  $L$  grupos de factores.

Para esta situação, podemos construir a álgebra dada por

$$\bigotimes_{l=1}^L \bigstar_{k=1}^w \mathcal{A}(a_l(k)),$$

a que corresponde a matriz de transição escrita como

$$\mathbf{B} = \bigotimes_{l=1}^L \mathbf{B}_l,$$

com  $\mathbf{B}_l$  triangular superior de ordem  $w_l$ ,  $l = 1, \dots, L$ .

Podemos ainda considerar modelos com mais um estrato. Em cada um desse estratos dá-se cruzamento e aninhamento de factores. Cada tratamento desse dado nível, aninha todos os tratamentos do nível seguinte.

Sendo  $\mathcal{A}_h^\circ$ ,  $h = 1, 2, \dots, \bar{h}$ , as álgebras e  $\mathbf{B}_l$ ,  $h = 1, 2, \dots, \bar{h}$  as matrizes de transição para os vários estratos, temos para o modelo final a álgebra e a matriz de transição dadas por

$$\mathcal{A}^\circ = \bigstar_{h=1}^{\bar{h}} \mathcal{A}_h^\circ$$

e

$$\mathbf{B} = \bigstar_{h=1}^{\bar{h}} \mathbf{B}_h.$$

Observe-se ainda que quando se toma repetições por tratamento, pode-se considerar para as mesmas, um último estrato com a álgebra  $\mathcal{A}(r)$  e a matriz de transição assume a forma

$$\mathbf{B}(r) = \begin{bmatrix} r & 0 \\ 1 & 1 \end{bmatrix}.$$

### 5.3 Aplicação: cimentos odontológicos

De forma a mostrar as potencialidades e aplicabilidade dos resultados obtidos nas secções anteriores, apresentamos uma aplicação na área da saúde, no âmbito do processo de solidificação de cimentos odontológicos utilizados em tratamentos dentários.

A experiência que a seguir apresentamos, visa avaliar as diferenças entre os cimentos odontológicos usados no tratamento de dentes, em especial no que diz respeito à velocidade de solidificação, sendo adaptado de Covas (2007) [7], pag. 93.

A velocidade de solidificação é um elemento importante a ter em consideração no tratamento de dentes. Quanto mais rápida for a solidificação, mais rapidamente se concluirá o tratamento e conseqüentemente, para o mesmo horizonte temporal, maior será o número de pontenciais doentes a serem tratados.

Nesta experiência, consideraram-se duas marcas de cimentos, designadas por  $C_1$  e  $C_2$ . A velocidade de solidificação, variável explicada, é expressa sob a forma de um índice. Este índice está ordenado por ordem inversa, uma vez que quanto mais rápida é a solidificação, melhor será o resultado esperado.

O processo de solidificação dos cimentos em si, é auxiliado através da incidência de uma luz, a mesma usada nesta experiência para ambos os cimentos, e de um fotopolimerizador. Uma vez que existem diversos fotopolimerizadores no mercado, a opção recaiu em seleccionar os três mais utilizados no processo. Designaremos esses fotopolimerizadores por  $F_1$ ,  $F_2$  e  $F_3$ .

No tratamento de dentes com estes cimentos, um outro aspecto a ter em consideração é o grau de solidificação. Nestes processos de tratamento, o tipo de solidificação não é constante. Torna-se pois interessante estudar as diferenças que poderão existir entre o grau de solidificação e o tempo. Não só o tempo e o grau de solidificação mas também, e para cada tipo de tratamento, qual o cimento e fotopolimerizador que melhor se adequam. Para estudar o efeito tempo no processo, a experiência foi repetida em dois períodos distintos, que designaremos por  $t_1$  e  $t_2$ .

O processo experimental foi realizado em cinco discos diferentes ( $d_1$ ,  $d_2$ ,  $d_3$ ,  $d_4$  e  $d_5$ ). Estes constituem as células, que são suficientemente grandes por forma a garantir três observações não correlacionadas, que representaremos por  $r_1$ ,  $r_2$  e  $r_3$ .

Os valores médios dos 5 discos encontram-se na Tabela 5.3.

### 5.3.1 Estrutura Algébrica

Todas as  $2 \times 3 \times 2$  combinações possíveis dos níveis de três factores foram consideradas. Como referido na secção anterior, vamos considerar três factores: cimentos,

Tabela 5.3: Médias dos discos  $d_1, d_2, d_3, d_4$  e  $d_5$ 

		$t_1$					$t_2$				
		$d_1$	$d_2$	$d_3$	$d_4$	$d_5$	$d_1$	$d_2$	$d_3$	$d_4$	$d_5$
$C_1$	$F_1$	26,03	28,43	27,40	26,10	26,77	29,37	30,53	30,27	29,80	29,63
	$F_2$	26,10	26,47	29,90	25,60	24,17	27,13	25,97	29,20	30,77	20,60
	$F_3$	9,83	10,00	10,17	10,67	11,37	9,83	10,00	10,17	10,67	11,37
$C_2$	$F_1$	26,00	26,93	25,17	26,50	25,17	29,97	29,53	29,00	29,03	25,47
	$F_2$	27,37	26,43	26,23	26,37	27,77	30,67	27,83	28,07	27,40	22,07
	$F_3$	6,07	6,40	6,63	6,60	6,43	6,07	6,40	6,63	6,60	6,43

fotopolimerizador e tempo, com 2, 3 e 2 níveis, respectivamente. Para cada um dos tratamentos correspondentes, foram utilizados cinco discos suficientemente grandes para fornecer três medidas independentes.

Para a construção do modelo iremos utilizar a metodologia desenvolvida por Covas (2007) [6]. Teremos então um modelo associado a

$$(\mathcal{A}(2) \otimes \mathcal{A}(3) \otimes \mathcal{A}(2)) \star \mathcal{A}(5) \star \mathcal{A}(3),$$

uma vez que poderemos considerar, numa primeira fase, um primeiro modelo com três factores [cimento, fotopolimerizador e duração], que aninham um outro modelo com um factor, cujos níveis são os discos.

Enquanto que os dois primeiros factores (cimento e fotopolimerizador), têm claramente efeitos fixos, começaremos por considerar que o tempo terá efeito aleatório. No que concerne aos discos, estes terão efeitos aleatórios. Uma vez que a interacção entre efeitos fixos e efeitos aleatórios pertence à parte dos efeitos aleatórios do modelo, teremos então  $m = 4$ , pelo que o modelo poderá ser escrito como,

Tabela 5.4: Índices utilizados no modelo

1	média geral	6	interacção entre cimentos e tempo
2	efeito do cimento	7	interacção entre fotopolimerizador e
3	efeitos do fotopolimerizador		duração
4	interacção entre cimentos e	8	interacção entre cimentos, fotopolimerizador
	fotopolimerizadores		e tempo
5	efeitos do tempo	9	efeitos dos discos

$$\mathbf{y} = \sum_{i=1}^4 \mathbf{X}_i \beta_i + \sum_{i=5}^9 \mathbf{X}_i \tilde{\beta}_i + \epsilon.$$

Na Tabela 5.4 está a codificação utilizada para cada um dos índices. No que se segue, o índice 10 está associado ao erro do modelo.

Para efeitos de construção do modelo, consideraremos as matrizes ortogonais standartizadas de ordem 2, 3 e 5, às quais de retirou a primeira linha ( $K_2$ ,  $K_3$  e  $K_5$ ), como defindo anteriormente. As que utilizaremos neste exemplo, são matrizes já conhecidas, mas qualquer uma outra nas mesmas condições é passível de ser utilizada. Uma outra abordagem seria a construção de tais matrizes pelo processo de ortogonalização de Graham-Shmidt.

$$\mathbf{K}_2 = \begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}, \quad \mathbf{K}_3 = \begin{bmatrix} -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{bmatrix}$$



e

$$\mathbf{K}_5 = \begin{bmatrix} -\frac{1}{\sqrt{2}} & 0 & 0 & 0 & \frac{1}{\sqrt{2}} \\ 0 & -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{6}} & 0 & -\frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{6}} \\ \sqrt{\frac{4}{30}} & -\sqrt{\frac{9}{30}} & \sqrt{\frac{4}{30}} & -\sqrt{\frac{9}{30}} & \sqrt{\frac{4}{30}} \end{bmatrix}$$

teremos

$$\mathbf{A}_1 = \frac{1}{\sqrt{2}} \mathbf{1}_2^\top \otimes \frac{1}{\sqrt{3}} \mathbf{1}_3^\top \otimes \frac{1}{\sqrt{2}} \mathbf{1}_2^\top \otimes \frac{1}{\sqrt{5}} \mathbf{1}_5^\top \otimes \frac{1}{\sqrt{3}} \mathbf{1}_3^\top \quad ; \quad g_1 = 1$$

$$\mathbf{A}_2 = \mathbf{K}_2 \otimes \frac{1}{\sqrt{3}} \mathbf{1}_3^\top \otimes \frac{1}{\sqrt{2}} \mathbf{1}_2^\top \otimes \frac{1}{\sqrt{5}} \mathbf{1}_5^\top \otimes \frac{1}{\sqrt{3}} \mathbf{1}_3^\top \quad ; \quad g_2 = 1$$

$$\mathbf{A}_3 = \frac{1}{\sqrt{2}} \mathbf{1}_2^\top \otimes \mathbf{K}_3 \otimes \frac{1}{\sqrt{2}} \mathbf{1}_2^\top \otimes \frac{1}{\sqrt{5}} \mathbf{1}_5^\top \otimes \frac{1}{\sqrt{3}} \mathbf{1}_3^\top \quad ; \quad g_3 = 2$$

$$\mathbf{A}_4 = \mathbf{K}_2 \otimes \mathbf{K}_3 \otimes \frac{1}{\sqrt{2}} \mathbf{1}_2^\top \otimes \frac{1}{\sqrt{5}} \mathbf{1}_5^\top \otimes \frac{1}{\sqrt{3}} \mathbf{1}_3^\top \quad ; \quad g_4 = 2$$

$$\mathbf{A}_5 = \frac{1}{\sqrt{2}} \mathbf{1}_2^\top \otimes \frac{1}{\sqrt{3}} \mathbf{1}_3^\top \otimes \mathbf{K}_2 \otimes \frac{1}{\sqrt{5}} \mathbf{1}_5^\top \otimes \frac{1}{\sqrt{3}} \mathbf{1}_3^\top \quad ; \quad g_5 = 1$$

$$\mathbf{A}_6 = \mathbf{K}_2 \otimes \frac{1}{\sqrt{3}} \mathbf{1}_3^\top \otimes \mathbf{K}_2 \otimes \frac{1}{\sqrt{5}} \mathbf{1}_5^\top \otimes \frac{1}{\sqrt{3}} \mathbf{1}_3^\top \quad ; \quad g_6 = 1$$

$$\mathbf{A}_7 = \frac{1}{\sqrt{2}} \mathbf{1}_2^\top \otimes \mathbf{K}_3 \otimes \mathbf{K}_2 \otimes \frac{1}{\sqrt{5}} \mathbf{1}_5^\top \otimes \frac{1}{\sqrt{3}} \mathbf{1}_3^\top \quad ; \quad g_7 = 2$$

$$\mathbf{A}_8 = \mathbf{K}_2 \otimes \mathbf{K}_3 \otimes \mathbf{K}_2 \otimes \frac{1}{\sqrt{5}} \mathbf{1}_5^\top \otimes \frac{1}{\sqrt{3}} \mathbf{1}_3^\top ; g_8 = 2$$

$$\mathbf{A}_9 = \mathbf{I}_2 \otimes \mathbf{I}_3 \otimes \mathbf{I}_2 \otimes \mathbf{K}_5 \otimes \frac{1}{\sqrt{3}} \mathbf{1}_3^\top ; g_9 = 48$$

$$\mathbf{A}_{10} = \mathbf{I}_2 \otimes \mathbf{I}_3 \otimes \mathbf{I}_2 \otimes \mathbf{I}_5 \otimes \mathbf{K}_3 ; g_{10} = 120.$$

A matriz de transção  $\mathbf{B}$ , poderá ser construída, como vimos anteriormente, da seguinte forma,

$$(\mathbf{B}(2) \otimes \mathbf{B}(3) \otimes \mathbf{B}(2)) \star \mathbf{B}(5) \star \mathbf{B}(3)$$

onde  $\mathbf{B}(r)$  é a matriz de transição para a AJC,  $\mathcal{A}(r)$ , e com a qual facilmente se obtém

$$\mathbf{B} = \begin{bmatrix} 180 & 0 & 0 & 0 & \vdots & 0 & 0 & 0 & 0 & 0 & 0 \\ 90 & 90 & 0 & 0 & \vdots & 0 & 0 & 0 & 0 & 0 & 0 \\ 60 & 0 & 60 & 0 & \vdots & 0 & 0 & 0 & 0 & 0 & 0 \\ 30 & 30 & 30 & 30 & \vdots & 0 & 0 & 0 & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \vdots & \dots & \dots & \dots & \dots & \dots & \dots \\ 90 & 0 & 0 & 0 & \vdots & 90 & 0 & 0 & 0 & 0 & 0 \\ 45 & 45 & 0 & 0 & \vdots & 45 & 45 & 0 & 0 & 0 & 0 \\ 30 & 0 & 30 & 0 & \vdots & 30 & 0 & 30 & 0 & 0 & 0 \\ 15 & 15 & 15 & 15 & \vdots & 15 & 15 & 15 & 15 & 0 & 0 \\ 3 & 3 & 3 & 3 & \vdots & 3 & 3 & 3 & 3 & 3 & 0 \\ 1 & 1 & 1 & 1 & \vdots & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

e onde, as sub-matrizes  $\mathbf{B}_{1,1}$ ,  $\mathbf{B}_{2,1}$  e  $\mathbf{B}_{2,2}$  estão indicadas.

Uma vez que  $\boldsymbol{\gamma}(2) = \mathbf{B}_{2,2}^\top \boldsymbol{\sigma}^2$ , teremos, com  $\sigma_{10}^2 = \sigma^2$ ,

$$\left\{ \begin{array}{l} \gamma_5 = 90\sigma_5^2 + 45\sigma_6^2 + 30\sigma_7^2 + 15\sigma_8^2 + 3\sigma_9^2 + \sigma_{10}^2 \\ \gamma_6 = 45\sigma_6^2 + 15\sigma_8^2 + 3\sigma_9^2 + \sigma_{10}^2 \\ \gamma_7 = 30\sigma_7^2 + 15\sigma_8^2 + 3\sigma_9^2 + \sigma_{10}^2 \\ \gamma_8 = 15\sigma_8^2 + 3\sigma_9^2 + \sigma_{10}^2 \\ \gamma_9 = 3\sigma_9^2 + \sigma_{10}^2 \\ \gamma_{10} = \sigma_{10}^2 \end{array} \right.$$

e como tal,

$$\left\{ \begin{array}{l} \sigma_5^2 = \frac{1}{90} (\gamma_5 - \gamma_6 - \gamma_7 + \gamma_8) \\ \sigma_6^2 = \frac{1}{45} (\gamma_6 - \gamma_8) \\ \sigma_7^2 = \frac{1}{30} (\gamma_7 - \gamma_8) \\ \sigma_8^2 = \frac{1}{15} (\gamma_8 - \gamma_9) \\ \sigma_9^2 = \frac{1}{3} (\gamma_9 - \gamma_{10}) \\ \sigma_{10}^2 = \gamma_{10} \end{array} \right.$$

Considerando ainda que  $\boldsymbol{\gamma}(1) = \mathbf{B}_{2,1}^\top \boldsymbol{\sigma}^2$ , podemos obter ainda  $\sigma_i^2$ ,  $i = 5, \dots, 10$ ,

substituindo pelos seus valores

$$\left\{ \begin{array}{l} \gamma_1 = 90\sigma_5^2 + 45\sigma_6^2 + 30\sigma_7^2 + 15\sigma_8^2 + 3\sigma_9^2 + \sigma_{10}^2 = \gamma_5 \\ \gamma_2 = 45\sigma_6^2 + 15\sigma_8^2 + 3\sigma_9^2 + \sigma_{10}^2 = \gamma_6 \\ \gamma_3 = 30\sigma_7^2 + 15\sigma_8^2 + 3\sigma_9^2 + \sigma_{10}^2 = \gamma_7 \\ \gamma_4 = 15\sigma_8^2 + 3\sigma_9^2 + \sigma_{10}^2 = \gamma_8 \end{array} \right.$$

De forma a aplicar os resultados desenvolvidos no âmbito do nosso trabalho, teremos de tratar os dados da Tabela 5.3 como elementos de uma matriz e trabalhar com os seus vectores.

Assim,

$$\tilde{\eta}_1 = 287,8763; \quad \tilde{\eta}_2 = -12,3133; \quad \tilde{\eta}_3 = \begin{bmatrix} -106,4681 \\ -57,3584 \end{bmatrix}; \quad \tilde{\eta}_4 = \begin{bmatrix} -7,7320 \\ -6,9517 \end{bmatrix};$$

$$\tilde{\eta}_5 = 8,8101; \quad \tilde{\eta}_6 = -2,0572; \quad \tilde{\eta}_7 = \begin{bmatrix} -7,6955 \\ 0,8696 \end{bmatrix}; \quad \tilde{\eta}_8 = \begin{bmatrix} 0,4473 \\ 2,1345 \end{bmatrix}$$

$$\tilde{\eta}_9 = \begin{bmatrix} 0,8982 \\ -2,8577 \\ -1,4142 \\ -1,0119 \\ 0,3266 \\ -0,8982 \\ -1,0842 \\ -0,7800 \\ -2,3678 \\ -1,0614 \\ -6,7411 \\ 1,3071 \\ 1,7963 \\ 5,8788 \\ -1,8856 \\ -0,1054 \\ 1,8779 \\ 0,8165 \\ 0,6128 \\ 0,2319 \\ 1,8779 \\ 0,8165 \\ \dots \end{bmatrix} ; \tilde{\eta}_{10} = \begin{bmatrix} 1,4142 \\ -0,0817 \\ -0,2828 \\ -2,7761 \\ -0,6364 \\ 1,5922 \\ 3,182 \\ -1,8371 \\ -0,4243 \\ 0,8165 \\ 3,5355 \\ 2,2862 \\ 2,2627 \\ 0,4083 \\ 0,7071 \\ 0,0817 \\ 1,2021 \\ -0,1225 \\ 3,5355 \\ 0,4083 \\ -0,7071 \\ 1,9596 \\ -2,7577 \\ 1,1839 \\ -5,0205 \\ 3,3068 \\ -0,0707 \\ 2,8169 \\ 1,8385 \\ 7,6751 \\ -5,2326 \\ -0,0817 \\ -3,1113 \\ 1,3064 \\ -0,9900 \\ -2,2045 \\ -2,4749 \\ -0,0408 \\ -5,0912 \\ 0,0000 \end{bmatrix} ; \begin{bmatrix} \dots \\ 0,1414 \\ -0,8165 \\ 3,8184 \\ 0,4899 \\ -0,7778 \\ 0,2041 \\ 1,6971 \\ -0,8982 \\ 0,3536 \\ 1,9188 \\ 0,1414 \\ -0,8165 \\ 3,8184 \\ 0,4899 \\ -0,7778 \\ 0,2041 \\ 1,6971 \\ -0,8982 \\ 0,3536 \\ 1,9188 \\ 0,0000 \\ -2,2045 \\ 0,4997 \\ 0,7757 \\ 0,9900 \\ 1,7963 \\ 1,4142 \\ 0,7349 \\ 2,6163 \\ 0,4491 \\ -0,9192 \\ 0,9390 \\ 1,9092 \\ -1,4289 \\ -4,879 \\ -1,1023 \\ 1,9092 \\ 4,9398 \\ 1,9799 \\ -3,3476 \\ \dots \end{bmatrix} ; \begin{bmatrix} \dots \\ 1,9092 \\ -4,205 \\ -0,4950 \\ 0,7757 \\ -3,4648 \\ -0,2041 \\ 0,4950 \\ -3,2252 \\ 0,4950 \\ -0,2858 \\ -4,3134 \\ -6,4095 \\ -3,9598 \\ 0,4083 \\ -3,3234 \\ -2,2454 \\ -4,5255 \\ 2,4495 \\ 6,1518 \\ -2,2454 \\ 0,42243 \\ -1,633 \\ 0,1414 \\ 0,4899 \\ -1,2021 \\ 0,0408 \\ 0,4243 \\ -0,2450 \\ -0,2121 \\ 1,0206 \\ 0,4243 \\ -1,633 \\ 0,1414 \\ 0,4899 \\ -1,2021 \\ 0,0408 \\ 0,4243 \\ -0,2450 \\ -0,2121 \\ 1,0206 \end{bmatrix} .$$

Para a parte dos efeitos fixos, podemos testar a ausência de efeitos e interações entre os dois primeiros factores. As estatísticas  $F$  para estas hipóteses são dadas por

$$\mathcal{F}_2 = \frac{g_6 \tilde{\eta}_2^2}{g_2 S_6} = \frac{\tilde{\eta}_2^2}{g_2 \tilde{\gamma}_6} = 35,8263 \quad ; \quad p - value = 0,1054$$

$$\mathcal{F}_3 = \frac{g_7 \|\tilde{\eta}_3\|^2}{g_3 S_7} = \frac{\|\tilde{\eta}_3\|^2}{g_3 \tilde{\gamma}_7} = 243,8511 \quad ; \quad p - value = 0,0041$$

$$\mathcal{F}_4 = \frac{g_8 \|\tilde{\eta}_4\|^2}{g_4 S_8} = \frac{\|\tilde{\eta}_4\|^2}{g_4 \tilde{\gamma}_8} = 22,7299 \quad ; \quad p - value = 0,0421$$

com 1 e 1, 2 e 2, e 2 e 2 graus de liberdade. O primeiro destes testes não é significativo. Devemos também salientar que obtivemos, para todos eles, graus de liberdade muito baixos no denominador. Veremos mais adiante que quando considerarmos o tempo como um factor de efeitos fixos, esta limitação será ultrapassada.

Por outro lado, considerando para os efeitos aleatórios,  $\tilde{\gamma}_j$ ,  $j = 5, \dots, 10$ , obtém-se

$$\tilde{\sigma}_5^2 = 0,50862; \quad \tilde{\sigma}_6^2 = 0,041196; \quad \tilde{\sigma}_7^2 = 0,92034; \quad \tilde{\sigma}_8^2 = -0,24558;$$

$$\tilde{\sigma}_9^2 = 0,2944; \quad \tilde{\sigma}_{10}^2 = 5,1787.$$

Um destes estimadores é negativo. Nomeadamente os associados ao tempo e aos discos. Estes resultados são compatíveis com o baixo número de graus de liberdade no denominador do teste  $F$ , o que nos leva a uma análise alternativa dos dados, onde

1. o factor tempo é considerado como tendo efeitos fixos;
2. usaremos então

$$S_9^\circ = S_9 + S_{10}$$

como soma dos quadrados do erro. Teremos agora,  $g^\circ = 168$  graus de liberdade para o erro.

O modelo para o vector de observações terá agora efeitos fixos e estará associado a uma AJC, que será construída da seguinte forma

$$(\mathcal{A}(2) \otimes \mathcal{A}(3) \otimes \mathcal{A}(2)) \star \mathcal{A}(15).$$

As matrizes  $\mathbf{A}_j$ ,  $j = 1, \dots, 8$ , manter-se-ão inalteráveis, uma vez que

$$\frac{1}{\sqrt{5}} \mathbf{1}_5^\top \otimes \frac{1}{\sqrt{3}} \mathbf{1}_3^\top = \frac{1}{\sqrt{15}} \mathbf{1}_{15}^\top,$$

assim como os vectores  $\tilde{\boldsymbol{\eta}}_j$ ,  $j = 1, \dots, 8$ , e as somas dos quadrados  $S_j$ ,  $j = 1, \dots, 8$ .

Para testar as hipóteses a que concerne os efeitos e interacções dos factores de efeitos fixos: cimento, fotopolimerizador e tempo, temos os testes  $F$  com estatísticas

$$\mathcal{F} = \frac{g^\circ S_j}{g_j S^\circ} = \frac{g^\circ \tilde{\gamma}_j}{S^\circ} \quad j = 2, \dots, 8$$

com  $g_j$  e  $g^\circ$  graus de liberdade. Os valores destas estatísticas e correspondentes  $p$ -values são apresentados na Tabela 5.5.

Estes resultados claramente apontam para a nova abordagem, um vez que agora temos resultados altamente significativos.

Os valores aqui apresentados foram obtidos por recurso ao *software* MatLab 7.0.4.

## 5.4 Simulações

Estando interessados em extrair mais informação sobre os dados, e partindo para uma abordagem diferente, vamos nesta secção, e por recurso às variáveis pivot, tal como foi proposto no início deste trabalho, induzir densidades e originar regiões de confiança para a qual, e por dualidade, se poderão testar hipóteses.

Se  $\tilde{\sigma}_5^2$  fosse não negativo, poderíamos estar interessados em testar a significância de  $\sigma_5^2$ . Para tal, poderíamos utilizar um teste  $F$  generalizado, ver Michalski & Zmysłony (1996)[31] (1999)[32].

Tabela 5.5:  $F$  tests

Efeitos e interações	Estatísticas	$p$ -values
Cimento	27,917	$\simeq 0$
Fotopolimerizador	1 346,5	$\simeq 0$
Cimento $\times$ Fotopolimerizador	9,9531	$\simeq 0$
Tempo	14,292	0,0002
Cimento $\times$ Tempo	0,7792	0,3786
Fotopolimerizador $\times$ Tempo	5,5217	0,0048
Cimento $\times$ Fotopolimerizador $\times$ Tempo	0,4379	0,6461

Dado um estimador quadrático não enviesado  $\tilde{\theta}$ , para o parâmetro  $\theta$ , a estatística de teste  $F$  generalizado para a hipótese

$$H_0 : \theta = 0$$

será dado por

$$\mathcal{F}_\theta = \frac{\tilde{\theta}_+}{\tilde{\theta}_-},$$

onde  $\tilde{\theta}_+$  e  $\tilde{\theta}_-$  são as componentes positivas e negativas de  $\tilde{\theta}$ . Obteríamos então uma estatística  $\mathcal{F}$  definida como

$$\mathcal{F}_{\sigma_5^2} = \frac{\tilde{\gamma}_5 + \tilde{\gamma}_8}{\tilde{\gamma}_6 + \tilde{\gamma}_7}.$$

No entanto, a distribuição de  $\mathcal{F}_{\sigma_5^2}$  dependeria de  $\gamma_5, \gamma_6, \gamma_7$  e  $\gamma_8$ , que teriam de ser substituídas pelos seus estimadores. Desta forma poder-se-ia optar uma abordagem alternativa para o teste de

$$H_0 : \sigma_5^2 = 0.$$

Seguindo Fonseca *et al* (2007) [16], assumam-se agora  $(X_{i,5}; X_{i,6}; X_{i,7}; X_{i,8})$  o conjunto de qui-quadrados com 1, 1, 2 e 2 graus de liberdade respectivamente, com

$i = 1, \dots, N$  e

$$\begin{cases} Z_{i,1} = \frac{S_5}{X_{i,5}} + \frac{S_8}{X_{i,8}} & i = 1, \dots, N \\ Z_{i,2} = \frac{S_6}{X_{i,6}} + \frac{S_7}{X_{i,7}} & i = 1, \dots, N \end{cases}$$

Tomando  $N = 10.000$ , obtemos a densidade conjunta para  $(\theta_+, \theta_-)$ , tal como descrita anteriormente, e que é apresentado na Figura 5.1.

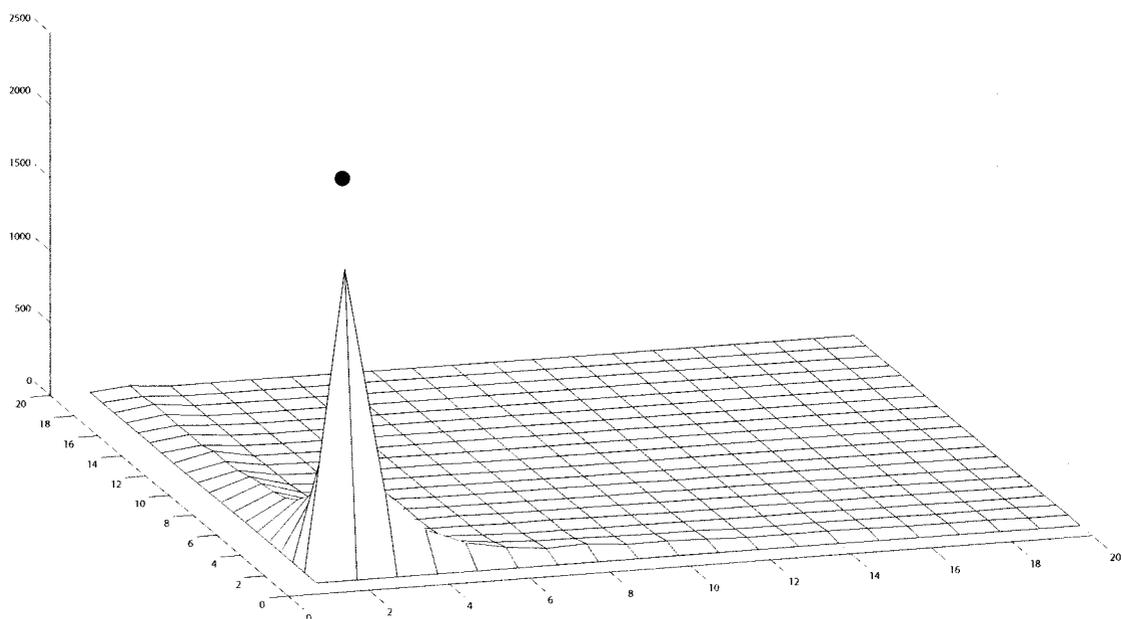


Figura 5.1: Função densidade

O ponto a vermelho representa o ponto de coordenadas

$$\begin{cases} \tilde{\theta}_+ = \tilde{\gamma}_5 + \tilde{\gamma}_8 \\ \tilde{\theta}_- = \tilde{\gamma}_6 + \tilde{\gamma}_7 \end{cases}$$

Como pode ser visto, a única moda desta densidade simulada está muito próxima de  $(\tilde{\theta}_+; \tilde{\theta}_-)$ . Uma vez que estes estimadores são UMVUE, podemos aceitar a robustez desta técnica indutora.

Aliás, na Figura 5.2 encontram-se representadas as curvas de nível para a densidade induzida. O ponto vermelho é  $(\tilde{\theta}_+; \tilde{\theta}_-)$  o que mostra mais uma vez a boa concordância da estimação pontual com a indução das densidades.

Da simulação da amostra de pares  $\{(Z_{i,1}, Z_{i,2}), i = 1, \dots, N\}$  obtemos uma amostra simulada  $\{Z_i, i = 1, \dots, N\}$ , com

$$Z_i = Z_{i,1} - Z_{i,2} \quad i = 1, \dots, N.$$

Podemos assim verificar que os quantis correspondentes às densidades empíricas, são estimadores fortemente consistentes dos quantis das distribuições exactas, ver Fonseca *et al* (2007) [16].

No nosso caso tenhamos os quantis empíricos dados por

$$\begin{cases} q_{N;0,05} = -10,625 \\ q_{N;0,95} = 697710, \end{cases}$$

que limitam um intervalo de confiança, de nível 90%, para  $\sigma_5^2$ . Observe-se que este intervalo contém a origem pelo que, mesmo ao nível de 10%, não se poderia rejeitar a hipótese nula

$$H_{0,5} : \sigma_5^2 = 0.$$

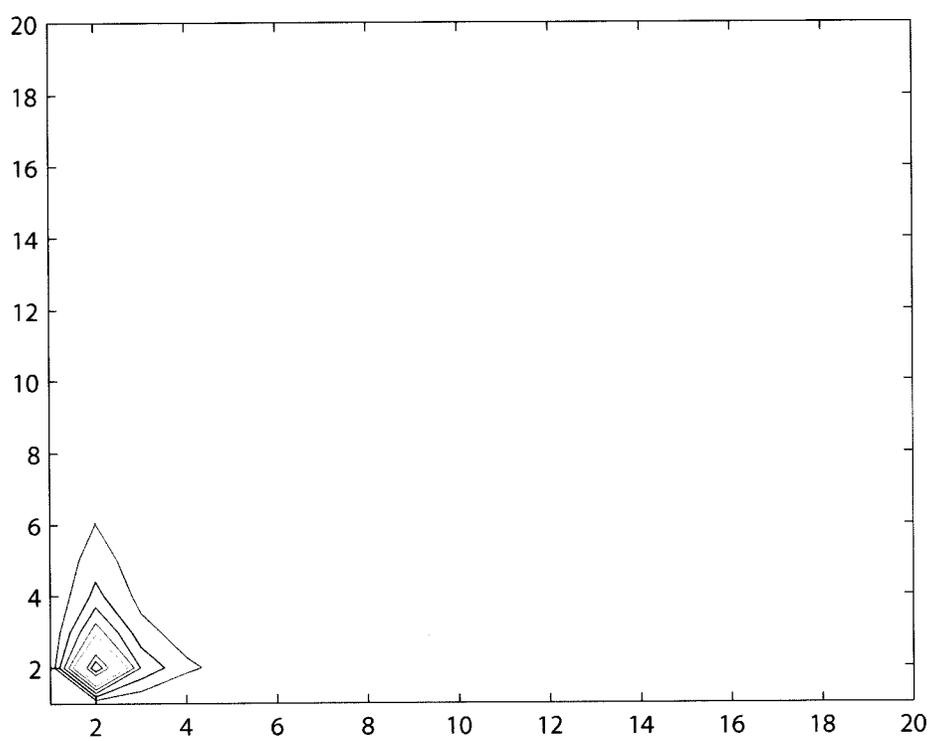


Figura 5.2: Curvas de nível

## **Parte III**

### **Considerações Finais**



# Capítulo 6

## Considerações finais e trabalho futuro

A inferência para modelos com OBS assenta na projecção ortogonal dos espaços das imagens das matrizes  $P_1, \dots, P_w$ . Este facto pode levar a problemas na combinação das estimativas obtidas a partir de diferentes projecções. Contudo, ao aplicarmos a metodologia aqui desenvolvida, com COBS, os potenciais problemas na combinação das estimativas dos diversos estratos deixa de existir, o que confere a esta abordagem uma generalização em termos de aplicabilidade, com enormes potencialidades em termos computacionais dada a sua facilidade na construção do modelo e da análise de dados no âmbito do delineamento experimental.

Como vimos, o uso de AJC permite uma generalização deste tipo de modelos, podendo ser construídos modelos complexos a partir de modelos mais simples.

Neste trabalho desenvolvemos numa forma matemática a inferência para uma classe especial de modelos mistos, aqueles com COBS, de que podemos salientar as principais conclusões. Assim:

- utilizámos AJC para estudar a respectiva estrutura algébrica;
- mostrámos que os estimadores dos mínimos quadrados dos respectivos vectores estimáveis são BLUE;

- apresentamos resultados para dois casos particularmente importantes:
  - modelos factoriais de base prima;
  - modelos com *cross-nesting* equilibrado;
- admitindo a normalidade:
  - obtivemos UMVUE para os parâmetros relevantes;
  - utilizámos variáveis pivot indutoras para construir regiões de confiança e testes de hipóteses para vectores estimáveis e componentes de variância.

Como vimos o tratamento aqui desenvolvido permite uma generalização no tratamento desta classe de modelos mistos.

As aplicações aqui apresentadas mostraram as vantagens e aplicabilidade dos modelos desenvolvidos.

Apesar de ser uma generalização no tratamento de modelos lineares mistos, esta temática não se encontra esgotada de forma alguma. Pretendemos completar o estudo dos modelos com COBS considerando extensões dos mesmos.

Nesta perspectiva, para além do abandonar de alguns pressupostos tomados no decorrer deste trabalho, uma outra área, com resultados que se afiguram numa fase prévia de bastante interesse, são as extensões dos COBS, nomeadamente os L-COBS.

Assim, considere-se um modelo misto escrito na forma

$$\begin{aligned} \mathbf{y} &= \sum_{i=1}^m (\mathbf{X}_i \otimes \mathbf{1}_r) \boldsymbol{\beta}_i + \sum_{i=m+1}^w (\mathbf{X}_i \otimes \mathbf{1}_r) \tilde{\boldsymbol{\beta}}_i + \boldsymbol{\varepsilon} \\ &= (\mathbf{I}_n \otimes \mathbf{1}_r) \left( \sum_{i=1}^m \mathbf{X}_i \boldsymbol{\beta}_i + \sum_{i=m+1}^w \mathbf{X}_i \tilde{\boldsymbol{\beta}}_i \right) + \boldsymbol{\varepsilon} \end{aligned}$$

com  $r$  observações por tratamento.

Se bem que o segundo termo é claramente um modelo com COBS, a matriz  $\mathbf{1}_r$  poderá ser substituída por uma outra matriz  $\mathbf{L}$ , pelo que o modelo poderá ser escrito

como

$$\mathbf{y} = \mathbf{L} \left( \sum_{i=1}^m \mathbf{X}_i \boldsymbol{\beta}_i + \sum_{m+1}^w \mathbf{X}_i \tilde{\boldsymbol{\beta}}_i \right) + \boldsymbol{\varepsilon},$$

onde  $\mathbf{L}$  é uma matriz com vectores coluna linearmente independentes. Estes modelos serão extensões dos anteriores, os L-COBS.

As características de tais modelos permitem uma nova abordagem onde poderão ser obtidos estimadores não enviesados para os parâmetros de interesse.

Um outro desenvolvimento a seguir, prende-se a com a ortogonalidade e o estudo de modelos não ortogonais. Será possível obter uma metodologia que permita de forma sistemática o tratamento de modelos não ortogonais? Obter-se-ão estatísticas suficientes e completas?

A não ortogonalidade abre todo um novo campo de estudo para este tipo de modelos, cuja aplicabilidade se antevê com grande relevo nas mais diversas áreas.



# Bibliografia

- [1] M. S. BAZARAA, H. D. SHERALI, AND C. M. SHETTY, *Nonlinear Programming: Theory and Algorithms*, John Wiley & Sons, New York, 1992.
- [2] T. CALIŃSKI AND S. KAGEYAMA, *Block Designs: A Randomization Approach. Vol. I: Analysis*, no. 150, Springer, Lecture Notes in Statistics, 2000.
- [3] ———, *Block Designs: A Randomization Approach. Vol. II: Design*, no. 170, Springer, Lecture Notes in Statistics, 2003.
- [4] S. R. S. CHARLES E. MCCULLOCH AND J. M. NEUHAUS, *Generalized, Linear, and Mixed Models*, Wiley Series in Probability and Statistics, Wiley, 2nd ed., 2008.
- [5] R. COVAS, *Inferência semi-bayesiana e modelos de componentes de variância*, Master's thesis, FCT - UNL, Lisboa, 2003.
- [6] ———, *Linear model genealogical tree application to an odontology experiment*, *Discussiones Mathematicae - Probability and Statistics*, 27 (2007), pp. 47–77.
- [7] ———, *Orthogonal Mixed Models and Commutative Jordan Algebras*, PhD thesis, Faculdade de Ciências e Tecnologia - Universidade Nova de Lisboa, 2007.
- [8] D. A. COX, *Galois Theory*, Wiley, 2004.

- [9] D. J. DA CONCEIÇÃO FERREIRA, *Variáveis pivot indutoras e componentes de variância em modelos normais ortogonais*, PhD thesis, Universidade da Beira Interior, 2006.
- [10] H. DRYGAS AND R. ZMYŚLONY, *Jordan algebras and bayesian quadratic estimation of variance components*, *Linear Algebra and its Applications*, (1992), pp. 259–275.
- [11] D. J. FERREIRA, *Inducing pivot variables and variance components in orthogonal normal models*, PhD thesis, Universidade da Beira Interior, 2006.
- [12] M. FONSECA, *Estimation and Hypothesis Testing in Mixed Linear Models*, PhD thesis, Universidade Nova de Lisboa, 2007.
- [13] M. FONSECA, J. T. MEXIA, AND R. ZMYŚLONY, *Exact distribution for the generalized  $f$  tests*, *Discussiones Mathematicae - Prob. Stat.*, (2002), pp. 37–51.
- [14] —, *Estimators and tests for variance components in cross nested orthogonal designs*, *Discussiones Mathematicae*, 23 (2003), pp. 175–201.
- [15] —, *Binary operation on jordan algebras and orthogonal normal models*, *Linear Algebra And Its Applications*, (2006), pp. 75–86.
- [16] —, *Jordan algebras. generating pivot variables and orthogonal normal models*, *Journal of Interdisciplinary Mathematics*, 1 (2007), pp. 305–326.
- [17] —, *Inference in normal models with commutative orthogonal block structure*, *Acta te Commentationes Universitatis Tartunensis de Mathematica*, (2008), pp. 3–16.
- [18] D. A. S. FRAZER, *Nonparametric Methods in Statistics*, John Wiley & Sons, New York, 1957.
- [19] A. GRAHAM, *Kronecker products and matrix calculus with applications*, Ellis Horwood, Ltd, New York, 1981.

- [20] V. JESUS, J. T. MEXIA, M. FONSECA, AND R. SMYŚLONY, *Binay operations and canonical forms for fractional and related models*, *Linear Algebra Appl.*, 430 (2009), pp. 2781–2797.
- [21] V. JESUS, J. T. MEXIA, AND P. C. RODRIGUES, *Inference for random effects in prime basis factorials using commutative jordan algebras*, *Discussiones Mathematicae – Probability and Statistics*, 27 (2007), pp. 15–25.
- [22] P. JORDAN, J. VON NEUMANN, AND E. WIGNER, *On an agebric generalization of the quantum mechanical formalism*, *Ann. of Math.*, 36 (1934), pp. 29–64.
- [23] A. I. KHURI, T. MATHEW, AND B. K. SINHA, *Statistical Tests for Mixed Linear Models*, Wiley, 1998.
- [24] E. L. LEHMAN, *Testing Statistical Hypothesis*, Jonh Willey & Sons, 2nd ed., 1986.
- [25] M. LOÈVE, *Probability Theorem*, D. Van Nostrand Co., Inc., Princeton, New Jersy, 1963.
- [26] J. D. MALLEY, *Statistical Applications of Jordan Algebras*, no. 91 in *Lecture Notes in Statistics*, Springer–Verlag, 1994.
- [27] K. MCCRIMMON, *Jordan algebras and their applications*, *Bulletin of the American Mathematical Society*, 84 (1978), pp. 612–627.
- [28] J. T. MEXIA, *Simultaneous confidence intervals: Generalization of the scheffé theorem*, Tech. Rep. 2, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 1989.
- [29] —, *Best linear unbiased estimates, duality of  $f$  tests and the scheffé multiple comparison method in presence of controled heterocedasticity*, *Comp. Stat y Data Analysis*, 10 (1990).

- [30] —, *Introdução à Inferência Estatística Linear*, Edições Universitárias Lusófonas, 1995.
- [31] A. MICHALSKI AND R. ZMYŚLONY, *Testing hypotheses for variance components in mixed linear models*, *Statistics*, 1 (1996), pp. 297–310.
- [32] —, *Testing hypotheses for linear functions of parameters in mixed linear models*, *Tatra Mountains Mathematical Publications*, (1999), pp. 103–110.
- [33] D. C. MONTGOMERY, *Design and analysis of experiments. 6th ed*, John Wiley and Sons, 2005.
- [34] A. M. MOOD, F. A. GRAYBILL, AND D. C. BOES, *Introduction to the Theory of Statistics*, *Statistics Series*, McGraw-Hill International Editions, Singapore, 3rd ed., 1974.
- [35] J. A. NELDER, *The analysis of randomized experiments with orthogonal block structure. i. block structure and the null analysis of variance*, in *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, vol. 283, The Royal Society, 1965, pp. 142–162.
- [36] —, *The analysis of randomized experiments with orthogonal block structure. ii. treatment structure and the general analysis of variance*, in *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, vol. 283, The Royal Society, 1965, pp. 163–178.
- [37] D. S. G. POLLOCK, *The Algebra of Econometrics*, John Wiley & Sons, New York, 1979.
- [38] H. SCHEFFÉ, *The Analysis of Variance*, Wiley – Interscience, 1999.
- [39] J. R. SCHOTT, *Matrix Analysis for Statistics*, *Wiley Series in Probability and Statistics*, 1997.

- [40] J. SEELY, *Linear spaces and unbiased estimators*, The Annals of Mathematical Statistics, 41 (1970), pp. 1735–1745.
- [41] —, *Linear spaces and unbiased estimators. application to a mixed linear model*, The Annals of Mathematical Statistics, 41 (1970), pp. 1735–1745.
- [42] —, *Quadratic subspaces and completeness*, The Annals of Mathematical Statistics, 42 (1971), pp. 710–721.
- [43] —, *Variance Components and Animal Breeding*, Cornell University, 1979, ch. Large-Sample Properties of Invariant Quadratic Unbiased Estimators in the Random One-Way Model.
- [44] J. SEELY AND G. ZYSKIND, *Linear spaces and minimum variance estimators*, The Annals of Mathematical Statistics, 42 (1971), pp. 691–703.
- [45] S. D. SILVEY, *Statistical inference*, Chapman & Hall, London, 1975.
- [46] D. M. VANLEEUEWEN, D. S. BIRKS, AND J. F. SEELY, *Balance and orthogonality in designs for mixed classification models*, The Annals of Statistics, 27 (1999), pp. 1927–1947.
- [47] D. M. VANLEEUEWEN, D. S. BIRKS, J. F. SEELY, J. MILLS, J. A. GREENWOOD, AND C. W. JONES, *Sufficient conditions for orthogonal designs in mixed linear models*, Journal of Statistical Planning and Inference, 73 (1998), pp. 373–389.
- [48] S. WEERAHADI, *Exact Statistical Methods for Data Analysis*, Springer, 2nd ed., 1996.
- [49] WILLIAMS, *Probability with martingales*, Cambridge University Press, 1997.
- [50] V. WITKOVSKÝ, *Computing the distribution of a linear combination of inverted gamma variables*, Kybernetika, 37 (2001), pp. 79–90.

- [51] R. ZMYŚLONY, *A characterization of best linear unbiased estimators in the general linear model*, Lecture Notes Statistics, 2 (1978), pp. 365–373.
- [52] —, *Completeness for a family of normal distributions*, Mathematical Statistics, Banach Center Publications, 6 (1980), pp. 355–357.