

Mestrado em Matemática e Aplicações

Dissertação para obtenção do Grau de
Mestre em Matemática e Aplicações
Especialização em Finanças

*Estimação dos Componentes da Variância em
modelos Lineares mistos - O Método Sub-D
Mestrado em Matemática e Aplicações*

António Moreira Monteiro

Praia, Janeiro de 2021

Mestrado em Matemática e Aplicações

Dissertação para obtenção do Grau de
Mestre em Matemática e Aplicações
Especialização em Finanças

Estimação dos Componentes da Variância em modelos Lineares mistos - O Método Sub-D Mestrado em Matemática e Aplicações

António Moreira Monteiro

Dissertação apresentada à Universidade de Cabo Verde para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Matemática e Aplicações, realizada sob a orientação científica de Adilson de Jesus Martins da Silva, Prof. e Investigador da Faculdade de Ciências e Tecnologia, Universidade de Cabo Verde.

Praia, Janeiro de 2021

Estimação dos Componentes da Variância em
modelos Lineares mistos - *O Método Sub-D*

Mestrado em Matemática e Aplicações

Copyright © António Moreira Monteiro, Faculdade de Ciências e Tecnologia,
Universidade de Cabo Verde

A Faculdade de Ciências e Tecnologia e a Universidade de Cabo Verde têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel e/ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objetivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

Sem dúvida, a avaliação que defendo é uma utopia, vale dizer, algo que não existe mas precisa ser criado, um contínuo convite à lucidez, à retomada de nossas ideias e de nosso fazer, a delinear com clareza e rigor esse conceito e a realizar, tornar real, o "in existente", no duplo sentido de não existir (ainda) e de algo que já se dá como possível no que hoje temos, isto é, no existente.

(COELHO, I.,2000)

Dedicatória

Aos pilares da minha vida...

Aos meus pais, Joaquim Monteiro e Domingas Moreira, por me terem preparado para os desafios da vida e por me proporcionarem o acesso à educação.

O júri :

Presidente

Arguente

Orientador

Doutor Adilson de Jesus Martins da Silva, Prof. e Investigador da Faculdade de Ciências e Tecnologia, Universidade de Cabo Verde.

Agradecimentos

Esta formatura é o colmatar de muito suor, noites sem dormir e outros sacrifícios, e sacrifícios de outras pessoas.

É por isso, que agradeço profundamente aos meus pais, por me oferecerem um amor incondicional e toda a minha família, por compreender as ausências e pelo estímulo permanente.

Gostaria de expressar a minha gratidão ao meu orientador Adilson de Jesus Martins da Silva, pela orientação, amizade, dedicação, compreensão e exigência.

Finalmente, devo grande parte desta jornada, aos meus colegas do programa de Mestrado em Matemática e Aplicações, que vivemos bons momentos, compartilhamos a mesma jornada na busca de um ideal sonhado e principalmente, a realização do que mais se esperou. O fim de uma caminhada começa a outra.

Resumo

Recentemente foi introduzido um novo método de estimação dos componentes da variância em modelos mistos lineares (MLM), denominado *Sub-D*. Simulações numéricas levadas a cabo em MLM com dois e MLM com três componentes da variância sugerem que *Sub-D* tem um desempenho similar ao estimador REML e melhor que o estimador ANOVA. Nesse sentido, e uma vez que *Sub-D* foi introduzido na sua forma generalizada, isto é, para um número arbitrário de componentes da variância, e o seu desempenho testado somente em modelos com dois e modelos com três componentes da variância, este trabalho tem por objeto deduzir a sua estrutura para modelos com quatro componentes da variância, bem como simular o seu desempenho nesses modelos.

Palavras-chave:

Modelos lineares mistos, Componentes da variância, Dedução do método *Sub-D*.

Abstract

Recently, a new method of estimation of components of variance was introduced in mixed linear models (MLM) called *Sub-D*. Numerical simulations carried out in MLM with two and MLM with three components of variance suggest that *Sub-D* has a performance similar to the REML estimator and better than the ANOVA estimator. In this sense, and since *Sub-D* was introduced in its generalized form, that is, for an arbitrary number of components of variance, and its performance tested only in models with two and models with three components of variance, this work has by object, deduce its structure for models with four components of variance, as well as simulate its performance in these models.

Keywords:

Mixed linear models, Components of variance, *Sub-D* method deduction.

Conteúdo

Lista de Figuras	iii
Lista de Tabelas	iv
Símbolos e Abreviaturas	v
1 INTRODUÇÃO	1
2 ÁLGEBRA MATRICIAL	2
2.1 Noções básicas e Notações	2
2.2 Matrizes ortogonais	12
2.3 Valores e Vetores próprios	13
2.4 Matrizes definidas e semi-definidas positivas	15
2.5 Decomposição em valores singulares	17
2.6 Diagonalização de Matrizes Simétricas	18
3 MODELO LINEAR MISTO	21
3.1 Modelo one-way	22
3.2 Modelo two-way	23
3.2.1 Modelos Cruzados	23
3.2.2 Modelos Aninhados	24
3.3 Estimação dos Componentes da Variância	24
3.3.1 Método da Análise da Variância	25
3.3.2 Método da Máxima Verossimilhança	26
3.3.3 Método da Máxima Verossimilhança Restrita	28
4 INFERÊNCIA EM MLM COM 4 COMPONENTES DA VARIÂNCIA	30
4.1 Introdução	30
4.2 Método Sub-D	30
4.2.1 Dedução do Método Sub-D para MLM com 4 Componentes da Variância	31
4.3 Resultados Numéricos	44
5 CONCLUSÃO	47

Anexo	49
Bibliografia	50

Lista de Figuras

Lista de Tabelas

4.1	Valores reais e estimados, usando o <i>Método Sub-D</i>	45
4.2	Variância dos valores reais e estimados, usando o <i>Método Sub-D</i>	45

Símbolos e Abreviaturas

MMA	Mestrado em Matemática e Aplicações
FCT	Faculdade de ciências e tecnologias
UniCV	Universidade de Cabo Verde
TFC	Trabalho Final de Curso
UNL	Universidade Nova de Lisboa
MLM	Modelo Linear Misto
ANOVA	Análise de Variância
ML	Máxima Verossimilhança
REML	Máxima Verossimilhança Restrita
VR	Valores Reais
VE	Valores Estimados

Capítulo 1

INTRODUÇÃO

Este trabalho tem como foco principal a dedução do método para a estimação dos componentes da variância em modelos mistos lineares (MLM), desenvolvido por Silva et al. [17]. Os Autores introduziram o método, deduzindo-o para modelos com dois componentes da variância, modelos com três componentes da variância e, finalmente, generalizaram-no, deduzindo-o para modelos com um número arbitrário de componentes da variância. Nessa perspectiva, e tendo em conta que o método foi particularizado somente para os casos de modelos com dois e três componentes da variância, nesse trabalho pretendemos deduzi-lo para o caso de modelos com quatro componentes da variância. O processo de dedução compreende a aplicação de uma sequência finita de transformações ortogonais, às quais denominaram de sub-diagonalizadoras, à estrutura da covariância do modelo, produzindo um conjunto de sub-modelos, que serão usados para criar o estimador ao qual denominaram de *Sub-D*.

O segundo capítulo é dedicado à introdução das notações e dos conceitos relacionados com a Álgebra matricial, que são indispensáveis para a dedução de *Sub-D*. O Terceiro capítulo destina-se à introdução e caracterização dos modelos mistos lineares, destacando os modelos “one-way” e “two-way”. A última parte deste capítulo dedica-se à introdução e descrição dos mais importantes métodos de estimação dos componentes da variância, destacando o método ANOVA e os métodos baseados em máxima verossimilhança. No quarto capítulo introduz-se o método *Sub-D* para o caso particular de modelos mistos lineares com 4 componentes da variância; ainda neste capítulo testes numéricos serão realizados, testando a validade do estimador *Sub-D* anteriormente introduzido.

O quinto e último capítulo integra as conclusões e/ou comentários finais.

ÁLGEBRA MATRICIAL

Neste capítulo apresentamos algumas noções, notações e propriedades fundamentais da álgebra matricial, exibindo-os em tópicos necessários para o desenvolvimento desta dissertação, começando com a introdução das noções básicas e algumas notações (ver Seção 2.1). Seguidamente, na Seção 2.2, introduzimos o conceito de matrizes ortogonais e, na Seção 2.3, o de valores e vetores próprios. Na Seção 2.4, introduzimos o conceito de matrizes definidas e semidefinidas.

Terminamos o capítulo com as duas últimas seções (ver Seções 2.5 e 2.6), subordinadas à decomposição de valores singulares e à diagonalização de matrizes simétricas, respectivamente.

Para uma revisão mais profunda sobre a Álgebra matricial, recomenda-se, por exemplo, Schott [15]; Horn e Johnson [4]; Rancher e Schaalje [13]; Albert [1]; Lay [5]; Rao e Rao [12]; Monteiro [8]; Magnus e Neudecker [6]; Burden e Faires [2]; entre outras.

2.1 Noções básicas e Notações

Nesta seção apresentaremos as noções básicas e notações sobre a teoria matricial que são fundamentais para o desenvolvimento deste trabalho.

Seja $\mathcal{A}_{m \times n}$, o conjunto das matrizes (definidas em baixo) com m linhas e n colunas.

Definição 2.1.1. *Uma matriz $A \in \mathcal{A}_{m \times n}$ é uma tabela de mn elementos algébricos dispostos em m linhas e n colunas, definida por*

$$A = (a_{ij}) = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix},$$

onde a_{ij} representa o elemento na i -ésima linha e na j -ésima coluna. Diz-se que A é uma *matriz quadrada* de ordem m se $m = n$.

Se $n = 1$, A diz-se um vetor, e neste caso usaremos uma letra minúscula para representá-lo, isto é, $a \in \mathcal{A}_{m \times 1}$.

Seja A uma *matriz quadrada*. Se os elementos fora da diagonal de A forem todos nulos, isto é, $a_{ij} = 0$, $i \neq j$, diz-se que A é uma *matriz diagonal* e denota-se (notações mais usuais) por,

$$A = \text{diag}(a_{11}, \dots, a_{mm}) = \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{mm} \end{bmatrix}. \quad (2.1)$$

Se $a_{ii} = 1$, $i = 1 \cdots m$, diz-se que A é uma *matriz identidade* de ordem m e denota - se por $A = I_m$, ou seja,

$$I_m = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}. \quad (2.2)$$

Seja $A \in \mathcal{A}_{m \times m}$, se A satisfizer a equação $AA = A$, ou seja, $AA = A^2 = A$, diz-se que é uma *matriz idempotente*.

Seja $A \in \mathcal{A}_{m \times n}$, se todos os elementos de A forem nulos, isto é, $a_{ij} = 0$, para qualquer $i = 1, \dots, m$ e $j = 1, \dots, n$, diz-se que A é uma *matriz nula*.

Definição 2.1.2. *Seja A^\top a matriz que se obtém de A , trocando as linhas pelas colunas ou vice-versa, tal que cada elemento na linha i e coluna j de A é exatamente o elemento na linha j e coluna i de A^\top , à matriz A^\top diz-se transposta da matriz A .*

Teorema 2.1.1. *Sejam A, B e $C \in \mathcal{A}_{m \times m}$, α e β escalares. Seja ainda $A_0 \in \mathcal{A}_{m \times m}$ uma matriz nula. As propriedades seguintes são válidas para as operações matriciais:*

- (i) $A + B = B + A$;
- (ii) $(A + B) + C = A + (B + C)$;
- (iii) $\alpha(A + B) = \alpha A + \alpha B$;
- (iv) $(\alpha + \beta)A = \alpha A + \beta A$;
- (v) $A - A = A + (-A) = A_0$;
- (vi) $A(B + C) = AB + AC$ e $(A + B)C = AC + BC$;
- (vii) $(AB)C = A(BC)$;
- (viii) $(A^\top)^\top = A$;
- (ix) $(\alpha A)^\top = \alpha A^\top$;

$$(x) (AB)^T = B^T A^T.$$

Demonstração. (i) $[A + B]_{ij} = a_{ij} + b_{ij} = b_{ij} + a_{ij} = [B + A]_{ij}$. Recorde que, $[A + B]_{ij}$ representa a soma dos elementos das matrizes A e B na i -ésima linha e j -ésima coluna.

$$(ii) [A+(B+C)]_{ij} = a_{ij}+[B+C]_{ij} = a_{ij}+(b_{ij}+c_{ij}) = (a_{ij}+b_{ij})+c_{ij} = [A+B]_{ij}+c_{ij} = [(A+B)+C]_{ij}.$$

$$(iii) [\alpha(A + B)]_{ij} = \alpha[A + B]_{ij} = \alpha(a_{ij} + b_{ij}) = \alpha a_{ij} + \alpha b_{ij} = [\alpha A]_{ij} + [\alpha B]_{ij} = [\alpha A + \alpha B]_{ij}.$$

$$(iv) [(\alpha + \beta)A]_{ij} = (\alpha + \beta)a_{ij} = (\alpha a_{ij}) + (\beta a_{ij}) = [\alpha A]_{ij} + [\beta A]_{ij} = [\alpha A + \beta A]_{ij}.$$

(v) Dado A e $X \in \mathcal{A}_{m \times n}$, tal que

$$A + X = 0. \tag{2.3}$$

Comparando os elementos correspondentes, temos

$$a_{ij} + x_{ij} = 0,$$

ou seja, $x_{ij} = -a_{ij}$, para $i = 1, \dots, m$ e $j = 1, \dots, n$. A matriz que satisfaz (2.3) é a matriz em que todos os seus elementos são iguais aos simétricos dos elementos de A . Denotamos a matriz X por $-A$.

(vi)

$$\begin{aligned} [A(B + C)]_{ij} &= \sum_{k=1}^p a_{ik}[B + C]_{kj} = \sum_{k=1}^p a_{ik}(b_{kj} + c_{kj}) = \sum_{k=1}^p (a_{ik}b_{kj} + a_{ik}c_{kj}) \\ &= \sum_{k=1}^p a_{ik}b_{kj} + \sum_{k=1}^p a_{ik}c_{kj} = [AB]_{ij} + [AC]_{ij} = [AB + AC]_{ij}. \end{aligned}$$

(vii) Sejam $A \in \mathcal{A}_{m \times p}$, $B \in \mathcal{A}_{p \times q}$ e $C \in \mathcal{A}_{q \times n}$. Então,

$$\begin{aligned} [A(BC)]_{ij} &= \sum_{k=1}^p a_{ik}[BC]_{kj} = \sum_{k=1}^p a_{ik} \left(\sum_{l=1}^q b_{kl}c_{lj} \right) = \sum_{k=1}^p \sum_{l=1}^q a_{ik}(b_{kl}c_{lj}) \\ &= \sum_{k=1}^p \sum_{l=1}^q (a_{ik}b_{kl})c_{lj} = \sum_{l=1}^q \sum_{k=1}^p (a_{ik}b_{kl})c_{lj} = \sum_{l=1}^q \left(\sum_{k=1}^p a_{ik}b_{kl} \right) c_{lj} \\ &= \sum_{l=1}^q [AB]_{il}c_{lj} = [(AB)C]_{ij}. \end{aligned}$$

$$(viii) [(A^T)^T]_{ij} = [A^T]_{ji} = a_{ij}$$

$$(ix) [(\alpha A)^\top]_{ij} = [\alpha A]_{ji} = \alpha a_{ji} = \alpha [A^\top]_{ij} = [\alpha A^\top]_{ij}$$

$$(x) [(AB)^\top]_{ij} = [AB]_{ji} = \sum_{k=1}^p a_{jk} b_{ki} = \sum_{k=1}^p [A^\top]_{kj} [B^\top]_{ik} = \sum_{k=1}^p [B^\top]_{ik} [A^\top]_{kj} = [B^\top A^\top]_{ij}$$

□

Definição 2.1.3. *Seja $A \in \mathcal{A}_{m \times m}$, isto é, uma matriz quadrada.*

(a) *Se $A = A^\top$, diz-se que A é uma matriz simétrica;*

(b) *Se $A = -A^\top$, diz-se que A é uma matriz anti - simétrica.*

Definição 2.1.4. *Ao número de colunas (linhas) da matriz A linearmente independentes diz-se rank de A e denota-se usualmente por $r(A)$. Se $A \in \mathcal{A}_{m \times n}$ de $r(A) = n$, onde $n < m$, diz-se que A tem o maior rank possível.*

Recorde-se que, um conjunto de vetores x_1, x_2, \dots, x_n são linearmente dependentes se existirem escalares $\alpha_1, \alpha_2, \dots, \alpha_n$, não todos nulos tais que

$$\alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_n x_n = 0. \quad (2.4)$$

Se não existirem $\alpha_1, \alpha_2, \dots, \alpha_n$ que satisfaçam a equação acima, diz-se que os vetores x_1, x_2, \dots, x_n são linearmente independentes.

Definição 2.1.5. *Seja $A \in \mathcal{A}_{m \times m}$, à matriz quadrada $B \in \mathcal{A}_{m \times m}$ tal que $AB = BA = I_m$, onde I_m é a matriz identidade, diz-se matriz inversa de A . A matriz inversa de A é usualmente denotada por A^{-1} . Se A admitir matriz inversa, diz-se que é uma matriz invertível ou não singular, caso contrário, ela é chamada de matriz não invertível ou singular.*

Teorema 2.1.2. *Seja $A \in \mathcal{A}_{m \times n}$ e A^- uma inversa generalizada de A . Então:*

(i) $(A^-)^\top$ é uma inversa generalizada de A^\top ;

(ii) Se α é um escalar diferente de zero, $\alpha^- A^{-1}$ é uma inversa generalizada de αA ;

(iii) Se $A \in \mathcal{A}_{m \times m}$ e não singular, $A^- = A^{-1}$ é única;

(iv) Se B e C são não singulares, $C^{-1} A^- B^{-1}$ é uma inversa generalizada de BAC ;

(v) $r(A) = r(AA^-) = r(A^-A) \leq r(A^-)$;

(vi) $r(A) = m$ se e somente se $AA^- = I_m$;

(vii) $r(A) = n$ se e somente se $A^-A = I_n$.

Demonstração. Ver Schott [15], Teorema 5.22. □

Teorema 2.1.3. *Sejam A e B elementos não singulares (invertíveis) de $\mathcal{A}_{m \times m}$. Então:*

- (i) $(A^{-1})^{-1} = A$;
- (ii) $(AB)^{-1} = B^{-1}A^{-1}$;
- (iii) $(A^{\top})^{-1} = (A^{-1})^{\top}$.

Demonstração. (i) Recorde-se, que uma matriz B é inversa de A^{-1} se $A^{-1}B = BA^{-1} = I_m$. Como A^{-1} é inversa de A , então $AA^{-1} = A^{-1}A = I_m$. A inversa é única, então $B = A$ é a inversa de A^{-1} , ou seja, $(A^{-1})^{-1} = A$.

(ii) Neste caso, temos que mostrar que a inversa de AB é $B^{-1}A^{-1}$, ou seja, mostrar que os produtos $(AB)(B^{-1}A^{-1})$ e $(B^{-1}A^{-1})(AB)$ são iguais à matriz identidade. Mas,

$$\begin{aligned} (AB)(B^{-1}A^{-1}) &= A(BB^{-1})A^{-1} \\ &= AI_mA^{-1} \\ &= AA^{-1} \\ &= I_m \end{aligned}$$

e

$$\begin{aligned} (B^{-1}A^{-1})(AB) &= B^{-1}(AA^{-1})B \\ &= B^{-1}I_mB \\ &= B^{-1}B \\ &= I_m. \end{aligned}$$

Então, AB é não singular, visto que a inversa da matriz é única, conclui-se que $(AB)^{-1} = B^{-1}A^{-1}$.

(iii) Por último, tem-se que $AA^{-1} = I_m$ e $A^{-1}A = I_m$. Tomando a transposta, $(AA^{-1})^{\top} = (I_m)^{\top} = I_m$ e $(A^{-1}A)^{\top} = (I_m)^{\top} = I_m$. Então, $(A^{-1})^{\top}A^{\top} = (I_m)^{\top} = I_m$ e $A^{\top}(A^{-1})^{\top} = I_m$, isto implica que, $(A^{\top})^{-1} = (A^{-1})^{\top}$. □

A *inversa generalizada*, também chamada *inversa condicional* de $A \in \mathcal{A}_{n \times p}$ é qualquer matriz A^{-} que satisfaz

$$AA^{-}A = A \tag{2.5}$$

Teorema 2.1.4. *Seja $A \in \mathcal{A}_{n \times p}$, onde $r(A) = r$, A^{-} a inversa generalizada de A e seja $(A^{\top}A)^{-}$ a inversa generalizada de $A^{\top}A$. Então*

- (i) $r(A^{-}A) = r(AA^{-}) = r(A) = r$;
- (ii) $(A^{-})^{\top}$ é a inversa generalizada de A^{\top} , isto é, $(A^{\top})^{-} = (A^{-})^{\top}$;

- (iii) $A = A(A^\top A)^- A^\top A$ e $A^\top = A^\top A(A^\top A)^- A^\top$;
- (iv) $(A^\top A)^- A^\top$ é uma inversa generalizada de A , isto é, $A^- = (A^\top A)^- A^\top$;
- (v) $A(A^\top A)^- A^\top$ é simétrico, tem $rank = r$ e é invariante para a escolha de $(A^\top A)^-$, isto é, $A(A^\top A)^- A^\top$ continua a ser o mesmo, não importa qual é o valor de $(A^\top A)^-$ é usado.

Demonstração.

- (i) Sabendo que as matrizes A e B são conformais para multiplicação, então $r(AB) \leq r(A)$ e $r(AB) \leq r(B)$. Neste caso, podemos concluir que, $r(A^-A) \leq r(A)$ e $r(A) = r(AA^-A) \leq r(A^-A)$.

Consequentemente,

$$r(A^-A) = r(A).$$

- (ii) $(A^-AA)^\top = A^\top(A^-)^\top A^\top$.

- (iii) Seja $W^\top W = A[I - (A^\top A)^- A^\top A]$. Mostra que

$$\begin{aligned} W &= [I - (A^\top A)^- A^\top A][A^\top A - A^\top A(A^\top A)^- A^\top A] \\ &= [I - (A^\top A)^- A^\top A]O = O. \end{aligned}$$

- (iv) $A[(A^\top A)^- A^\top A]A = A(A^\top A)^- A^\top A = A$, pela parte (iii).

- (v) (Searle 1982, p.222) mostra que $A(A^\top A)^- A^\top$ é invariante para as escolhas de $(A^\top A)^-$, seja B e C dois valores de $(A^\top A)^-$. Então, pela parte (iii), $A = ABA^\top A$ e $A = ACA^\top A$, de modo que $ABA^\top A = ACA^\top A$.

Para demonstrar que isto implica $ABA^\top = ACA^\top$, mostra - se que

$$\begin{aligned} (ABA^\top A - ACA^\top A)(B^\top A^\top - C^\top A^\top) &= (ABA^\top - ACA^\top) \\ &\times (ABA^\top - ACA^\top)^\top. \end{aligned}$$

O lado esquerdo é O , porque $ABA^\top A = ACA^\top A$. Então, o lado direito, também é O , e recorde que para matrizes quadradas os elementos de AA^\top são os produtos das linhas de A , isso implica que $ABA^\top - ACA^\top = O$. Para mostrar simetria, seja S um inverso generalizado de $A^\top A$. Então, ASA^\top é simétrico e $ASA^\top = ABA^\top$ desde que ABA^\top é invariante para $(A^\top A)^-$. Portanto, ABA^\top é também simétrico. Para mostrar que $r[A(A^\top A)^- A^\top] = r$, usando as partes (i) e (iv). \square

A inversa generalizada de uma matriz simétrica não é necessariamente simétrica. Mais ainda, qualquer matriz admite uma inversa generalizada.

Definição 2.1.6. *Seja $A \in \mathcal{A}_{m \times m}$. A soma dos elementos da diagonal de A diz-se traço de A e denota-se por $tr(A)$, isto é*

$$\text{tr}(A) = \sum_{i=1}^m a_{ii}.$$

Teorema 2.1.5. Se $A \in \mathcal{A}_{m \times n}$ e $B \in \mathcal{A}_{n \times m}$, então $AB \in \mathcal{A}_{m \times m}$ e

$$\text{tr}(AB) = \text{tr}(BA).$$

Demonstração. Sabendo que $(AB)_{ii}$ representa o produto dos elementos das matrizes A e B na i -ésima linha e $(A)_i$ representa a i -ésima linha da matriz A . Então

$$\begin{aligned} \text{tr}(AB) &= \sum_{i=1}^m (AB)_{ii} = \sum_{i=1}^m (A)_i \cdot (B)_i \\ &= \sum_{i=1}^m \sum_{j=1}^n a_{ij} b_{ji} = \sum_{j=1}^n \sum_{i=1}^m b_{ji} a_{ij} \\ &= \sum_{j=1}^n (B)_j \cdot (A)_j = \sum_{j=1}^n (BA)_{jj} = \text{tr}(BA). \end{aligned}$$

□

Teorema 2.1.6.

(i) Se A e $B \in \mathcal{A}_{n \times n}$, então

$$\text{tr}(A \pm B) = \text{tr}(A) \pm \text{tr}(B).$$

(ii) Se $A \in \mathcal{A}_{n \times p}$ e $B \in \mathcal{A}_{p \times n}$, então

$$\text{tr}(AB) = \text{tr}(BA).$$

(iii) Se $A \in \mathcal{A}_{n \times p}$, então

$$\text{tr}(A^T A) = \sum_{i=1}^p a_i^T a_i.$$

(iv) Se $A \in \mathcal{A}_{n \times p}$, então

$$\text{tr}(AA^T) = \sum_{i=1}^p a_i^T a_i,$$

onde a_i^T , é o número de linhas de A .

(v) Se $A = (a_{ij}) \in \mathcal{A}_{n \times p}$ for uma matriz com elemento representativo a_{ij} , então

$$\text{tr}(A^T A) = \text{tr}(AA^T) = \sum_{i=1}^n \sum_{j=1}^p a_{ij}^2.$$

(vi) Se $A \in \mathcal{A}_{n \times n}$ e $P \in \mathcal{A}_{n \times n}$ não singular, então

$$\text{tr}(P^{-1}AP) = \text{tr}(A).$$

(vii) Se $A, C \in \mathcal{A}_{n \times n}$ e C é uma matriz ortogonal, então

$$\text{tr}(C^T AC) = \text{tr}(A).$$

(viii) Se $A \in \mathcal{A}_{n \times p}$ de $r(r)$ e A^- for a inversa generalizada de A , então

$$\text{tr}(A^- A) = \text{tr}(AA^-) = r.$$

Demonstração. Ver Rencher e Schaalje [13], Teorema 2.11. □

Definição 2.1.7. *Seja $A \in \mathcal{A}_{m \times m}$. O determinante de A , denotado por $\det(A)$ ou $|A|$, é definido por*

$$\begin{aligned} |A| &= \sum (-1)^{f(i_1, \dots, i_m)} a_{1i_1} a_{2i_2} \dots a_{mi_m} \\ &= \sum (-1)^{f(i_1, \dots, i_m)} a_{i_1 1} a_{i_2 2} \dots a_{i_m m}, \end{aligned}$$

onde a soma de todas as permutações (i_1, \dots, i_m) do conjunto dos números inteiros $(1, \dots, m)$ e da função $f(i_1, \dots, i_m)$ é igual ao número de transposição necessária para trocar (i_1, \dots, i_m) para $(1, \dots, m)$. A transposição é a troca de dois inteiros.

O cofator correspondente de a_{ij} é denominado por $A_{ij} = (-1)^{i+j} m_{ij}$. Para qualquer i, \dots, m o $|A|$ pode ser obtido expandindo o número de linhas,

$$|A| = \sum_{j=1}^m a_{ij} A_{ij} \tag{2.6}$$

ou o número de colunas,

$$|A| = \sum_{i=1}^m a_{ij} A_{ij} \tag{2.7}$$

Teorema 2.1.7. *Seja α um escalar e $A \in \mathcal{A}_{m \times m}$. Então*

(i) $|A| = |A^T|$;

(ii) $|\alpha A| = \alpha^m |A|$;

(iii) Se A é uma matriz diagonal, então $|A| = a_{11} \dots a_{mm} = \prod_{i=1}^m a_{ii}$;

(iv) Se todos os elementos da linha ou coluna de A são zeros, $|A| = 0$;

- (v) Se duas linhas ou colunas de A são proporcionais a um outro, $|A| = 0$;
- (vi) A troca de duas linhas ou colunas muda o sinal de $|A|$;
- (vii) Se todos os elementos de uma linha ou coluna de A são multiplicados por α , então o determinante é multiplicado por α ;
- (viii) O determinante de A não muda quando um múltiplo de uma linha ou coluna é adicionado a outra linha ou coluna.

Demonstração. Ver Schott [15], Teorema 1.4. □

Teorema 2.1.8. *Se $A \in \mathcal{A}_{m \times m}$ for idempotente, $A^2 = A$, e $P \in \mathcal{A}_{m \times m}$ for não singular e $C \in \mathcal{A}_{m \times m}$ for ortogonal, então*

- (i) $I - A$ é idempotente;
- (ii) $A(I - A) = 0$ e $(I - A)A = 0$;
- (iii) $P^{-1}AP$ é idempotente;
- (iv) C^TAC é idempotente.

Demonstração.

(i)

$$\begin{aligned} (I - A)^2 &= I - 2A + A^2 \\ &= I - 2A + A \\ &= I - A. \end{aligned}$$

(ii)

$$\begin{aligned} A(I - A) &= A - A^2 \\ &= A - A \\ &= 0. \end{aligned}$$

(iii)

$$\begin{aligned} (P^{-1}AP)^2 &= P^{-1}APP^{-1}AP \\ &= P^{-1}A^2P \\ &= P^{-1}AP. \end{aligned}$$

2.2 Matrizes ortogonais

A relação entre uma matriz P e sua transposta P^T , fornece espécies importantes na caracterização de uma matriz ortogonal, ou seja, matriz real é ortogonal se for quadrada e invertível ou não singular.

Um vetor $p \in \mathcal{A}_{m \times 1}$ é denominado de *vetor normalizado* ou *vetor unitário* se

$$p^T p = 1 \quad (2.9)$$

Definição 2.2.1. Os vetores $p_1, \dots, p_n \in \mathcal{A}_{m \times 1}$, onde $n \leq m$ são chamados ortogonais se

$$p_i^T p_j = 0,$$

qualquer que seja $i \neq j$. Se em adição, cada p_i é um vetor normalizado, então os vetores dizem - se ortonormais.

Definição 2.2.2. $P \in \mathcal{A}_{m \times m}$, cujas colunas formam um conjunto de vetores ortonormais é denominado matriz ortogonal se

$$P^T P = I.$$

Aplicando o determinante a ambos os lados, obtemos

$$|P^T P| = |P^T| |P| = |P|^2 = |I| = 1.$$

Assim, $|P| = \pm 1$ de modo que P seja não singular, $P^{-1} = P^T$, e $PP^T = I$. Na adição para $P^T P = I$, isto é, as linhas de P formam um conjunto de vetores ortonormais $\in \mathcal{A}_{m \times 1}$.

Teorema 2.2.1. Sejam P e $Q \in \mathcal{A}_{m \times m}$ forem ortogonais e $A \in \mathcal{A}_{m \times m}$ qualquer, então

- (i) $|P| = \pm 1$;
- (ii) $|P^T A P| = |A|$;
- (iii) PQ é uma matriz ortogonal.

Demonstração. Ver Schott [15], Teorema 1.10. □

2.3 Valores e Vetores próprios

Valores e vetores próprios são explicitamente especiais, definidos como funções dos elementos de uma matrix quadrada. Em muitas aplicações envolvendo as análises de uma matrix quadrada, a informação chave das análises é munido de alguns ou todos esses valores e vetores próprios.

Definição 2.3.1. *Seja $A \in \mathcal{A}_{m \times m}$. Qualquer número real λ é denominado de valor próprio de A se existir um vetor não nulo $x \in \mathbb{R}^n$, tal que*

$$Ax = \lambda x.$$

Qualquer vetor x não nulo que satisfaz a definição acima é chamado de vetor próprio (vetor característico) de A associado ao valor próprio λ . Os valores próprios também são chamados de *autovalores* ou ainda *valores característicos*.

A equação da definição 2.3.1 é equivalente a

$$(\lambda I - A)x = 0. \tag{2.10}$$

Se $|\lambda I - A| \neq 0$, então $(\lambda I - A)^{-1}$ existiria e a multiplicação da equação acima pela inversa conduziria a uma contradição assumido que $x \neq 0$.

Definição 2.3.2. *Seja $A \in \mathcal{A}_{m \times m}$, o determinante*

$$|\lambda I_m - A| = \begin{vmatrix} \lambda - a_{11} & -a_{12} & \cdots & -a_{1m} \\ -a_{21} & \lambda - a_{22} & \cdots & -a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{m1} & -a_{m2} & \cdots & \lambda - a_{mm} \end{vmatrix}$$

é denominado de *polinómio característico* de A . E a equação

$$|\lambda I_m - A| = 0 \tag{2.11}$$

é chamada equação característica de A .

Teorema 2.3.1. *Os valores próprios de A são as raízes do polinómio característico de A .*

Demonstração. Se λ for um valor próprio de A associado ao vetor próprio x , então (2.10) pode ser escrito

$$Ax = (\lambda I_m)x \quad \text{ou} \quad (\lambda I_m - A)x = 0,$$

um sistema homogénio de m equações e m incógnitas. Tem uma solução não trivial se e somente se o determinante da sua matriz de coeficientes se anular, ou seja, se e somente se $|\lambda I_m - A| = 0$. Se λ for uma raiz do polinómio característico de A , então $|\lambda I_m - A| = 0$, logo o sistema homogénio $(\lambda I_m - A)x = 0$ tem solução não trivial x , portanto λ é um valor próprio de A . \square

Teorema 2.3.2. *Seja $A \in \mathcal{A}_{m \times m}$, então*

- (i) O valor próprio de A^T é o mesmo de A ;
- (ii) A é singular se e somente se pelo menos um valor próprio de A é igual a zero;
- (iii) Os elementos diagonais de A são os valores próprios de A , se A é uma matriz triangular;
- (iv) Os valores próprios de BAB^{-1} são os mesmos que os valores próprios de A , se $B \in \mathcal{A}_{m \times m}$ e não singular;
- (v) Cada valor próprio de A é ± 1 , se A é uma matriz ortogonal.

Demonstração. ver Schott [15], Teorema 3.2. □

Teorema 2.3.3. *Seja λ um valor próprio de $A \in \mathcal{A}_{m \times m}$ e x o vetor próprio correspondente, então*

- (i) Se n é um inteiro ≥ 1 , λ^n é um valor próprio de A^n correspondendo ao vetor próprio de x ;
- (ii) Se A é não singular, λ^{-1} é um valor próprio de A^{-1} correspondente ao vetor próprio de x ;

Demonstração. Parte (i) é demonstrado usando a relação $Ax = \lambda x$, ou seja, temos

$$A^n x = A^{n-1}(Ax) = A^{n-1}(\lambda x) = \lambda A^{n-1}x = \dots = \lambda^n x.$$

Para demonstrar (ii), multiplicamos a equação $Ax = \lambda x$ por A^{-1} , originando a equação

$$x = \lambda A^{-1}x. \tag{2.12}$$

Desde que A for não singular e $\lambda \neq 0$, dividindo ambos os lados da equação (2.12) por λ produz

$$A^{-1}x = \lambda^{-1}x,$$

que é a equação do valor - vetor próprio para A^{-1} , com valor próprio λ^{-1} e vetor próprio x . □

Teorema 2.3.4. *Seja $A \in \mathcal{A}_{m \times m}$ com valores próprios, $\lambda_1, \dots, \lambda_m$, então*

- (i) $tr(A) = \sum_{i=1}^m \lambda_i$;
- (ii) $|A| = \prod_{i=1}^m \lambda_i$.

Demonstração. ver Schott [15], Teorema 3.5. □

Teorema 2.3.5. *Suponha que x_1, \dots, x_r são os vetores próprios da matriz $A \in \mathcal{A}_{m \times m}$, onde $r \leq m$. Se o valor próprio correspondente $\lambda_1, \dots, \lambda_r$ são tais que $\lambda_i \neq \lambda_j$, para qualquer $i \neq j$, então os vetores x_1, \dots, x_r são linearmente independentes.*

Demonstração. ver Schott [15], Teorema 3.6. □

2.4 Matrizes definidas e semi-definidas positivas

Para falar de matrizes definidas e semidefinidas positivas há necessidade de introduzir a *forma quadrática*.

Definição 2.4.1. *Dada a matriz real e simétrica $A \in \mathcal{A}_{n \times n}$, se define a forma quadrática associado a ela como aplicação de \mathbb{R} em \mathbb{R} que a cada vetor x , o número real atribuído $x^T Ax$. Se $n = 2$ e consideramos a matriz $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix}$ e o vetor $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, então a expressão na forma quadrática é*

$$x^T Ax = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2.$$

De forma análoga, se $n = 3$ e consideramos a matriz $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{bmatrix}$ e o vetor $x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$, então a expressão na forma quadrática é

$$\begin{aligned} x^T Ax &= \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \\ &= a_{11}x_1^2 + a_{22}x_2^2 + 2a_{12}x_1x_2 + 2a_{13}x_1x_3 + 2a_{23}x_2x_3 + a_{33}x_3^2. \end{aligned}$$

Observa que uma forma quadrática é composto por todos os monómios possíveis de grau dois de um polinómio de duas ou três variáveis. Também se observa que o valor de uma forma quadrática na origem, isto é, quando x é o vetor nulo é zero.

Definição 2.4.2. *Seja $A \in \mathcal{A}_{n \times n}$, simétrica, de modo que $x^T Ax > 0$, \forall não nulo, então, A é denotado de matriz definida positiva.*

Definição 2.4.3. Seja $A \in \mathcal{A}_{n \times n}$, simétrica, de modo que $x^T A x \geq 0$, \forall não nulo, então, A é denotado de matriz semi-definida positiva.

Os valores próprios $\lambda_1, \lambda_2, \dots, \lambda_n$, de matrizes definidas e semi-definidas positivas são positivas e não negativas, respectivamente, ou seja, $\lambda_i > 0$, $\forall i = 1, 2, \dots, n$.

Teorema 2.4.1. Seja $A \in \mathcal{A}_{n \times n}$ com valores próprios, $\lambda_1, \lambda_2, \dots, \lambda_n$.

- (i) Se A é definida positiva, então $\lambda_i > 0$, para $i = 1, 2, \dots, n$;
- (ii) Se A é semi-definida positiva, então $\lambda_i \geq 0$, para $i = 1, 2, \dots, n$. O número de valores próprios λ_i para cada $\lambda_i > 0$ é $\text{rank } A$.

Demonstração. (i) Para qualquer λ_i , temos $Ax_i = \lambda_i x_i$. Multiplicando por x_i^T , obtemos

$$\begin{aligned} x_i^T A x_i &= \lambda_i x_i^T x_i \\ \lambda_i &= \frac{x_i^T A x_i}{x_i^T x_i} > 0. \end{aligned}$$

Na segunda expressão, $x_i^T A x_i$ é positiva porque A é definida positiva, e $x_i^T x_i$ é positiva porque $x_i \neq 0$. \square

Se uma matriz A é definida positiva, podemos encontrar uma matriz de raiz quadrada \sqrt{A} . Desde que os valores próprios de A são positivos, podemos substituir a raiz quadrada $\sqrt{\lambda_i}$ na expressão,

$$\begin{aligned} A &= XDX^T \\ &= \sum_{i=1}^n \lambda_i x_i x_i^T, \end{aligned} \tag{2.13}$$

obtemos,

$$\sqrt{A} = X \sqrt{D} X^T \tag{2.14}$$

onde, $\sqrt{D} = \text{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \dots, \sqrt{\lambda_n})$ e $X = (x_1, x_2, \dots, x_n)$ é uma matriz ortogonal. A matriz \sqrt{A} é simétrico e tem propriedade

$$\sqrt{A} \sqrt{A} = (\sqrt{A})^2 = A.$$

2.5 Decomposição em valores singulares

A decomposição em valores singulares é uma fatorização para uma matriz de qualquer dimensão, enquanto que as outras fatorizações só se aplicam a matrizes quadradas.

Teorema 2.5.1. *Se $A \in \mathcal{A}_{m \times n}$ de rank $r > 0$, existem matrizes ortogonais $P \in \mathcal{A}_{m \times m}$ e $Q \in \mathcal{A}_{n \times n}$ de modo que $A = PDQ^T$ e $D = P^T A Q$, onde a matriz $D \in \mathcal{A}_{m \times n}$ é dado por*

(i) Δ se $r = m = n$;

(ii) $\begin{bmatrix} \Delta & (0) \end{bmatrix}$ se $r = m < n$;

(iii) $\begin{bmatrix} \Delta \\ (0) \end{bmatrix}$ se $r = n < m$;

(iv) $\begin{bmatrix} \Delta & (0) \\ (0) & (0) \end{bmatrix}$ se $r < m, r < n$, e $\Delta \in \mathcal{A}_{r \times r}$, uma matriz diagonal com elementos diagonais positivos. Os elementos diagonais de Δ^2 são os vetores próprios de $A^T A$ e AA^T .

Demonstração. ver Schott [15], Teorema 4.1. □

A decomposição em valores singulares pode ser expressa numa relação matricial que depende do rank da matriz.

Considere $A \in \mathcal{A}_{m \times n}$ e seja $r \leq \min(m, n)$, rank A . Então, existem r constantes positivas, ou valores singulares, $\sigma_1 = \sqrt{\lambda_1}, \sigma_2 = \sqrt{\lambda_2}, \dots, \sigma_r = \sqrt{\lambda_r}$, em que $\lambda_i > 0, i = 1, 2, \dots, r$ são os r valores próprios positivos de $A^T A$.

Existem, ainda, r vetores próprios q_1, q_2, \dots, q_r e r vetores próprios p_1, p_2, \dots, p_r , tal que

$$A = \sum_{i=1}^r \sigma_i p_i q_i^T = P_r \Delta Q_r, \quad (2.15)$$

em que, $P_r = [p_1 | p_2 | \dots | p_r]$ e $Q_r = [q_1 | q_2 | \dots | q_r]$, são matrizes ortogonais e Δ é uma matriz diagonal do tipo

$$\Delta = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_r \end{bmatrix}.$$

Nesta situação, $\lambda_1, \lambda_2, \dots, \lambda_r > 0$ e $q_1, q_2, \dots, q_r > 0$, são os r primeiros pares de valores próprios e vetores próprios de $A^T A$, obtidos de

$$A^T A v_i = \lambda_i v_i$$

em que, $\lambda_1 > \lambda_2 > \dots > \lambda_r > 0$ são os valores estritamente positivos.

Os vetores próprios p_i , por sua vez, estão associados aos vetores próprios $q_i, i = 1, 2, \dots, r$, pela relação

$$p_i = \frac{1}{\sigma_i} A q_i.$$

Desta forma, a decomposição em valores singulares pode ser escrita pela expressão

$$A = P_r \Delta Q_r^T. \quad (2.16)$$

Alternativamente, p_i , $i = 1, 2, \dots, r$, são os r valores próprios associados aos mesmos valores próprios positivos $\lambda_1 \geq \lambda_2 \geq \dots \lambda_r > 0$ de AA^T , em que $\sigma_i = \sqrt{\lambda_i}$, $i = 1, 2, \dots, r$ são os respetivos valores próprios singulares.

Os vetores próprios q_i , por sua vez, estão relacionados aos vetores próprios p_i , $i = 1, 2, \dots, r$ pela relação

$$q_i = \frac{1}{\sigma_i} A^T p_i.$$

2.6 Diagonalização de Matrizes Simétricas

A matriz A é diagonalizável se for semelhante a uma matriz diagonal, ou seja, se existe uma matriz invertível (ver Definição 2.1.5) P tal que $D = P^{-1}AP$. Então

- (i) os determinantes são iguais, $|D| = |A|$;
- (ii) os polinómios característicos são iguais, $|D - \lambda I| = |A - \lambda I|$;
- (iii) os valores próprios são os mesmos e com as mesmas multiplicidades algébricas e geométricas;
- (iv) os vetores próprios correspondem - se, mas não são necessariamente os mesmos, se x é o vetor próprio de A associado ao valor próprio λ , então Px é vetor próprio de D associado ao valor próprio λ .

Suponha que A é diagonalizável e que $D = P^{-1}AP$, onde D é uma matriz diagonal

$$D = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}.$$

Conclui-se que $AP = DP$. Se as colunas de P são os vetores x_1, x_2, \dots, x_n , então

$$AP = A(x_1 \dots x_n) = (Ax_1 \dots Ax_n) \quad (2.17)$$

e

$$DP = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} (x_1 \dots x_n) = (\lambda_1 x_1 \dots \lambda_n x_n)$$

Então, isto significa que

$$Ax_1 = \lambda_1 x_1, Ax_2 = \lambda_2 x_2, \dots, Ax_n = \lambda_n x_n.$$

A existência de P^{-1} , significa que nenhum dos vetores x_i é o vetor zero. λ_i é um valor próprio de A e x_i é seu vetor próprio correspondente. Como P tem uma inversa, esses vetores próprios são linearmente independentes, portanto A possui n vetores próprios linearmente independentes. Caso contrário, se A tem n vetores próprios linearmente independentes, então a matriz P cujas colunas são esses vetores próprios será invertível, e teremos $P^{-1}AP = D$, onde D é uma matriz diagonal com entradas iguais aos valores próprios de A .

Uma matriz A é diagonalizável se ela for semelhante a uma matriz diagonal, neste caso, diz - se que A pode ser diagonalizada.

Teorema 2.6.1. *Matrizes semelhantes têm os mesmos valores próprios.*

Demonstração. Sejam A e B semelhantes, então $B = P^{-1}AP$ para alguma matriz P invertível. Provaremos que A e B têm os mesmos polinômios característico, $f_A(\lambda)$ e $f_B(\lambda)$, respectivamente.

Assim,

$$\begin{aligned} f_B(\lambda) = |\lambda I_n - B| &= |\lambda I_n - P^{-1}AP| \\ &= |P^{-1}\lambda I_n P - P^{-1}AP| \\ &= |P^{-1}(\lambda I_n - A)P| \\ &= |P^{-1} \parallel P \parallel \lambda I_n - A| \\ &= |\lambda I_n - A| = f_A(\lambda) \end{aligned} \tag{2.18}$$

Como $f_A(\lambda) = f_B(\lambda)$, então A e B têm os mesmos valores próprios. □

Os valores próprios de uma matriz quadrada nem sempre são reais. Por exemplo, a matriz $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ tem valores próprios complexos, $-i$ e i , respectivamente. Por isso, nem sempre é possível diagonalizar uma matriz. Essa situação muda se as matrizes são simétricas.

Teorema 2.6.2. *Toda matriz simétrica de elementos reais tem valores reais.*

Demonstração. Considere uma matriz simétrica

$$A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$$

onde a , b e c são reais. O polinómio característico dessa matriz é

$$\begin{aligned} p(\lambda) &= |A - \lambda I| \\ &= \begin{vmatrix} a - \lambda & b \\ b & c - \lambda \end{vmatrix} \\ &= (a - \lambda)(c - \lambda) - b^2 \\ &= ac - a\lambda - c\lambda + \lambda^2 - b^2 \\ &= \lambda^2 - (a + c)\lambda + ac - b^2. \end{aligned}$$

Calculando o discriminante Δ , tem - se

$$\begin{aligned} \Delta &= [-(a + c)]^2 - 4.1(ac - b^2) = a^2 + 2ac + c^2 - 4ac + 4b^2 \\ &= a^2 - 2ac + c^2 + 4b^2 = (a - c)^2 + 4b^2 \geq 0. \end{aligned}$$

Como $\Delta \geq 0$, as raízes de $p(\lambda)$ são reais. Portanto, os valores próprios de uma matriz simétrica são reais. \square

Se uma matriz A é diagonalizável através de uma matriz ortogonal P , então ao invés de escrevermos $A = PDP^{-1}$, simplesmente podemos escrever $A = PDP^T$, isso ocorre para qualquer matriz simétrica.

Teorema 2.6.3. *Se uma matriz A é simétrica ($A^T = A$), então os vetores próprios correspondentes a diferentes valores próprios são ortogonais*

Demonstração. Suponha que λ e μ são dois quaisquer valores próprios de A e que x e y são os vetores próprios correspondentes. Então, $Ax = \lambda x$ e $Ay = \mu y$. O objetivo dessa demonstração é encontrar duas expressões diferentes para o produto $x^T Ay$. O produto da matriz $x^T Ay \in \mathcal{A}_{1 \times 1}$ ou equivalente a um número. Desde que $Ay = \mu y$, temos

$$x^T Ay = x^T(\mu y) = \mu x^T y.$$

Mas também, desde que $Ax = \lambda x$, temos $(Ax)^T = (\lambda x)^T = \lambda x^T$. Para quaisquer matrizes M, N , $(MN)^T = N^T M^T$, assim $(Ax)^T x^T A^T$. Mas $A^T = A$, assim $x^T A = \lambda x^T$, portanto

$$x^T Ay = \lambda x^T y.$$

Portanto, temos duas expressões diferentes para $x^T Ay$ que são $\mu x^T y$ e $\lambda x^T y$. Se pode concluir que

$$\mu x^T y = \lambda x^T y \tag{2.19}$$

ou

$$(\mu - \lambda)x^T y = 0.$$

Desde que $\lambda \neq \mu$, tem-se que $\mu - \lambda \neq 0$. Deduzimos, entretanto que, $x^T y = 0$, diz precisamente que x e y são ortogonais, que é exatamente o que queríamos provar. \square

MODELO LINEAR MISTO

A seleção de modelos é uma parte importante de toda pesquisa em modelagem estatística e envolve a procura de um modelo que seja o mais simples possível e que descreva bem os dados observados que surgem nas áreas da agricultura, ecologia, economia, engenharia, medicina, ciência política, sociologia, zootecnia, bem como o controle da qualidade industrial e melhoria, etc.

Um modelo linear que apresenta somente fatores de efeitos fixos, além do erro experimental, que é sempre aleatório, é denominado *modelo fixo*, enquanto que, os modelos que apresentam apenas fatores de efeitos aleatórios, exceto a constante μ , que é sempre fixa, é denominado *modelo aleatório*.

Um *modelo linear misto (MLM)* é aquele que apresenta tanto fatores de efeitos fixos como aleatórios, além do erro experimental e da constante μ . Para a análise dos MLM, alguns pontos são relevantes, como por exemplo, *testes e estimação dos componentes da variância e testes de hipóteses para os efeitos definidos como fixos*.

Na análise dos MLM não equilibrados, o problema não é tão simples, em razão da opção pela estrutura da matriz de variância-covariância e da escolha das técnicas de estimação dos componentes da variância. Por exemplo, a técnica de ANOVA adapta facilmente aos MLM com dados equilibrados, mas não se adapta em situações com dados não equilibrados, porque utiliza cálculos baseados em modelos de efeitos fixos, em vez de mistos. Por conseguinte, a ML e REML, fornecem estimadores com várias propriedades estatísticas ótimas tanto para modelos com dados equilibrados como para dados não equilibrados.

Os modelos estatísticos têm estado associados aos modelos de efeitos fixos envolvendo um fator com k níveis definindo grupos, denominado preditor, e n_i observações (independentes) de cada grupo i , denominado erros residuais, que podem ser escritos na seguinte equação escalar:

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \quad i = 1, \dots, k; \quad j = 1, \dots, n_i, \quad (3.1)$$

onde y representa o vetor das observações, μ e α_i são constantes fixas finitas e desconhecidas que caracterizam a média do modelo e as variáveis aleatórias, ε_{ij} representam os erros residuais aleatórios e independentes com média zero e variância σ^2 , ou seja, $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$.

O modelo (3.1) pode ser representado na seguinte forma matricial:

$$y = X\beta + \varepsilon, \quad (3.2)$$

ou seja,

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix},$$

onde o vetor ε tem uma distribuição com média 0_n , vetor nulo cujos componentes são zeros e matriz variância - covariância, $\sigma^2 I_n$.

Seja $\Sigma(z)$ a matriz da variância-covariância do vetor aleatório z . O modelo (3.2) tem uma distribuição com média $X\beta$ e matriz variância - covariância, $\Sigma(y) = \sigma^2 I$.

Definição 3.0.1. *Seja $y \in \mathbb{R}^n$ (vetor de respostas). O modelo linear misto é definido por*

$$y = X\beta + \sum_{i=1}^{n+1} X_i \beta_i, \quad (3.3)$$

onde $y \sim (X\beta, \Sigma)$, com $\Sigma = \sum_{i=1}^{n+1} \sigma_i^2 M_i$, sendo $M_i = X_i X_i^T$ matriz dos efeitos fixos e $\beta \in \mathbb{R}^{p_i}$ vetor de valores esperados (fixos). A matriz $X \in \mathcal{A}_{n \times p}$ representa os efeitos fixos do modelo, enquanto que $X_i, i = 1, \dots, n+1$, matrizes conhecidas (fixas), determinam a estrutura da variância - covariância do modelo. Os vetores $\beta_i, i = 1, \dots, n+1$, representam os efeitos aleatórios e independentes não observáveis do modelo. Assim, tem-se que $E(\beta_i) = 0_{p_i}$ e $\Sigma(\beta_i) = \sigma_i^2 I_{p_i}$, de modo que

$$E(y) = 0_n \quad \text{e} \quad \Sigma(y) = \sum_{i=1}^{n+1} \sigma_i^2 M_i,$$

onde $M_{n+1} = I_n$ e $\sigma_i^2, \dots, \gamma_{n+1}^2$, são parâmetros positivos, desconhecidos e fixos, designados de *componentes da variância*.

3.1 Modelo one-way

O modelo definido em (3.1), conhecido como modelo *one-way*, aplica-se a dados equilibrados e não equilibrados. Os parâmetros μ (constante fixa e desconhecida caracterizando

a média) e α_i são os efeitos independentes devido à observação y , tendo a distribuição com média $\mathbf{0}$ e variância σ_α^2 , enquanto que, ε_{ij} são erros aleatórios e independentes com média $\mathbf{0}$ e variância σ_ε^2 .

O modelo *one-way* de efeitos aleatórios é (matricialmente) representado por:

$$y = Z\mu + Z_1\alpha + \varepsilon, \tag{3.4}$$

ou

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 1 & 1 & 0 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 1 & 0 & 0 & \dots & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_k \end{bmatrix} + \begin{bmatrix} 1 & 0 & \dots & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 1 & 0 & \dots & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & 1 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix},$$

onde y e ε é definido em (3.2), $Z_1 \in \mathcal{A}_{(\sum_{s=1}^k \times k)n_s}$, α tem uma distribuição com média 0 e matriz variância-covariância $\sigma^2 I$.

O modelo em (3.4) tem uma distribuição com média $X\mu$ e matriz variância-covariância dado por $\Sigma(y) = \sigma_\alpha^2 Z_1 Z_1^T + \sigma_\varepsilon^2$, onde σ_α^2 e σ_ε^2 são denominados de componentes da variância.

3.2 Modelo two-way

Nesta Seção apresentamos o modelo two-way para dois fatores aleatórios, pois temos dois modelos que podem ser utilizados e são eles: modelos cruzados, Seção 3.2.1 e modelos aninhados, Seção 3.2.2.

3.2.1 Modelos Cruzados

O modelo com dois fatores equilibrados e com efeitos cruzados e interação é dado por:

$$y_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}, \tag{3.5}$$

com $i = 1, \dots, a$; $j = 1, \dots, b$; $k = 1, \dots, n$.

Para este modelo μ é um parâmetro comum a todos os tratamentos e representa a média geral dos dados, α_i e β_j são os efeitos devido ao i -ésimo e ao j -ésimo nível do fator P e O e são variáveis aleatórias independentes com média zero e variâncias σ_P^2 e σ_O^2 , respectivamente e

γ_{ij} é a interação entre os fatores P e O , que também tem distribuição normal com média zero e variância σ_γ^2 . A variável aleatória ε_{ijk} corresponde ao erro aleatório experimental, ou seja, a variabilidade não explicada pelo modelo devido as variações presentes em diversas fontes não consideradas no estudo. Este tem distribuição normal com média zero e variância σ^2 .

Resumindo, podemos ter μ é a média geral dos dados, α_i o efeito do nível i do fator P , β_j o efeito de nível j do fator O , γ_{ij} é o efeito do nível ij da interação entre P e O e ε_{ijk} componente aleatório do erro. O erro, os efeitos α_i , β_j , γ_{ij} têm as distribuições, $\varepsilon_{ijk} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$, $\alpha_i \sim \mathcal{N}(0, \sigma_P^2)$, $\beta_j \sim \mathcal{N}(0, \sigma_O^2)$ e $\gamma_{ij} \sim (0, \sigma_\gamma^2)$, respectivamente. Além disso, α_i , β_j , γ_{ij} e ε_{ijk} são independentes para todo, i, j, k .

3.2.2 Modelos Aninhados

A análise de uma variância *hierarquizada* ou *aninhada* é uma extensão da ANOVA, em que cada fator é dividido em subgrupos destes fatores. Estes subgrupos são escolhidos aleatoriamente a partir de um conjunto maior de subgrupos possíveis.

O modelo de dois fatores aleatórios para classificação aninhada com dados não equilibrados é dada por:

$$y_{ijk} = \mu + \alpha_i + \beta_{ij} + \varepsilon_{ijk}, \quad (3.6)$$

onde $i = 1, \dots, a$; $j = 1, \dots, b$; $k = 0, \dots, n$; y_{ijk} é a k -ésima observação de j -ésima nível do fator B dentro do i -ésimo nível do fator A , μ é a média total, α_i é o efeito devido ao i -ésimo nível do fator A , β_{ij} é o efeito devido ao j -ésimo nível do fator B hierarquizado sob o i -ésimo nível do fator A e ε_{ijk} é o erro residual.

Supõe-se que, α_i s, β_{ij} s e ε_{ijk} são mutuamente e completamente variáveis aleatórias não correlacionadas com média zero e variâncias σ_α^2 , σ_β^2 e σ_ε^2 , respectivamente. Os parâmetros σ_α^2 , σ_β^2 e σ_ε^2 são conhecidos como componentes de variâncias. O modelo descrito acima implica que o número dos níveis do fator A é a e existem níveis b do fator B dentro de cada nível de A . Seja b denominado o número total de cada subgrupos, sabendo que $b = \sum_{i=1}^a b_i$. O número das observações n subgrupo j do grupo i é n_{ij} .

Matricialmente o modelo em (3.6) é representado por:

$$y = X\beta + X_1\alpha + X_2\beta + X_3\gamma + eI, \quad (3.7)$$

onde α, β , tem distribuição normal com média zero e variância σ^2 .

3.3 Estimação dos Componentes da Variância

As Componentes da variância são as variâncias associadas aos efeitos aleatórios de um modelo, pois a população e o método de melhoramento a serem utilizados dependem de algumas informações que podem ser obtidas a partir dessas componentes da variância.

Diversos métodos têm sido desenvolvidos para estimar as componentes da variância, destacando-se o método da análise da variância (ANOVA), métodos de Henderson, estimador quadrático não - viesado de norma mínima (MINQUE), estimador quadrático não - viesado da variância mínima (MIVQUE), estimador quadrático não - viesado de norma mínima iterativo (I-MINQUE), máxima verossimilhança (ML) e máxima verossimilhança restrita (REML).

Dos vários métodos de estimação das componentes da variância , destacamos a nossa atenção no estudo da ANOVA, ML e REML.

Para melhor análise e um conhecimento mais profundo sobre Modelo Linear Misto: Modelo One-Way, Modelos Two-Way e Estimação dos Componentes da Variância, por exemplo, consulte: Hocking [3]; McCulloch and Searle [7]; Moser [9]; Muller and Stewart [10]; Rencher and Schaalje [13]; Sahai and Ojeda [14]; Searle et al. [16]; Stapleton [19]; Myers et al. [11], e muitas outras.

3.3.1 Método da Análise da Variância

O método da análise da variância(ANOVA), em geral é adequado para modelos simples, que envolvem dados balanceados. Os estimadores ANOVA apresentam muitas propriedades, por exemplo, não-viesados e têm variância mínima, são funções de estatísticas suficientes, para as quais podem ser obtidas estimativas dos erros padrões associados, e uma aproximação dos números de graus de liberdade. Quando os dados não são balanceados, não existe um único modo de se obter a tabela da análise da variância, levando a diferentes estimativas para um mesmo componente. Como uma desvantagem pode-se citar o fato de que esse método não exclui a ocorrência de estimativas negativas para as componentes da variância, fato que torna a propriedade de estimador não viesado pouco interessado.

Do MLM (3.3), podemos obter os erros quadráticos das diferentes fontes, dado por,

$$S_i^2 = \mathbf{y}^\top P_i \mathbf{y}, \quad i = 1, \dots, n + 1 \tag{3.8}$$

de modo que, $X^\top P_i X = 0_{p,p}$, onde $P_i \in \mathcal{L}$, $X \in \mathcal{A}_{n \times p}$ dos efeitos fixos conhecidos.

Segundo Schott [15], Teorema 9.18(a), o valor esperado de S_i^2 que dependerá apenas das componentes da variância é dado por

$$\begin{aligned} E(S_i^2) &= \text{tr}\{P_i E(\mathbf{y}\mathbf{y}^\top)\} = \text{tr}(P_i) + (X\beta)^\top P_i (X\beta) \\ &= \text{tr}\left(\sum_{j=1}^{n+1} \gamma_j P_i M_j\right) \\ &= \sum_{j=1}^{n+1} \text{tr}(X_j^\top P_i X_j). \end{aligned} \tag{3.9}$$

Desde que, $S = \begin{bmatrix} S_1^2 \\ \cdots \\ S_{n+1}^2 \end{bmatrix}$, $\gamma = \begin{bmatrix} \gamma_1 \\ \cdots \\ \gamma_{n+1}^2 \end{bmatrix}$ e $C = \begin{bmatrix} tr(X_1^T P_1 X_1) & \cdots & tr(X_{n+1}^T P_1 X_{n+1}) \\ \vdots & \cdots & \vdots \\ tr(X_{n+1}^T P_1 X_1) & \cdots & tr(X_{n+1}^T P_{n+1} X_{n+1}) \end{bmatrix}$,
obtém-se

$$E(S) = C\gamma. \quad (3.10)$$

Igualando $S = C\gamma$, obtém-se

$$\hat{\gamma} = C^{-1}S, \quad (3.11)$$

onde C é uma matriz quadrada não singular e de rank completo, uma vez que o número de fontes da variação é igual ao número das componentes da variância. Para situações em que há mais variações de fontes do que componentes da variância, pode-se usar uma das variações do estimador baseada em ANOVA:

$$\tilde{\gamma} = (C^T C)^{-1} C^T S. \quad (3.12)$$

As equações (3.10) e (3.11) são imparciais, isso implica que:

$$E(\hat{\gamma}) = C^{-1}E(S) = C^{-1}C\gamma = \gamma;$$

$$E(\tilde{\gamma}) = (C^T C)^{-1} C^T E(S) = (C^T C)^{-1} C^T C\gamma = \gamma.$$

3.3.2 Método da Máxima Verossimilhança

O método da máxima verossimilhança (ML) foi desenvolvida por Fisher em 1922, mas Hartley e Rao em 1967 apresentaram a especificação matricial de um modelo misto e a derivação de equações ML para várias classes de modelos.

Um das vantagens do ML é a geração de estimativas não negativas dos componentes da variância. Para a estimação ML dos componentes da variância os efeitos fixos devem ser conhecidos, caso contrário, são substituídos por suas estimativas obtidas por ML. Porém, na estimação dos componentes da variância, o método ML não considera a perda de número de graus de liberdade devido à estimação desses efeitos fixos, causando então o vício, que conduz a subestimativas dos parâmetros da variância e, portanto podem conduzir a inferências incorretas.

O método da máxima verossimilhança consiste em maximizar a função densidade de probabilidade em relação a efeitos fixos e aos componentes da variância.

Seja o MLM dado em (3.3), sabendo que y terá distribuição normal multivariada, com média $X\beta$ e matriz variâncias-covariância Σ , ou seja, $y \sim \mathcal{N}(X\beta, \Sigma)$.

A função de verossimilhança é dado por

$$L(\beta, \gamma) = (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{X}\beta)^\top \Sigma^{-1}(\mathbf{y} - \mathbf{X}\beta)\right], \quad (3.13)$$

sendo $|\Sigma|$ o determinante da matriz Σ .

Maximizando o logarítmo da função de verossimilhança e diferenciando-as em relação aos componentes da variâncias(ver Silva et al. [17], Teorema A.1.3 e A.1.4), obtemos respectivamente,

$$\begin{aligned} L(\beta, \gamma) &= \log[L(\beta, \gamma)] \\ &= -\frac{n \log(2\pi)}{2} - \frac{\log |\Sigma|}{2} - \frac{(\mathbf{y} - \mathbf{X}\beta)^\top \Sigma^{-1}(\mathbf{y} - \mathbf{X}\beta)}{2} \end{aligned} \quad (3.14)$$

e

$$\begin{aligned} \frac{\partial L(\beta, \gamma)}{\partial \gamma_i} &= \frac{\partial \left[-\frac{n \log(2\pi)}{2} - \frac{\log |\Sigma|}{2} - \frac{(\mathbf{y} - \mathbf{X}\beta)^\top \Sigma^{-1}(\mathbf{y} - \mathbf{X}\beta)}{2} \right]}{\partial \gamma_i} \\ &= 0, \quad i = 1, \dots, n+1. \end{aligned} \quad (3.15)$$

De (3.14), obtemos

$$\text{tr}(\Sigma^{-1}M_i) = (\mathbf{y} - \mathbf{X}\beta)^\top \Sigma^{-1}M_i \Sigma^{-1}(\mathbf{y} - \mathbf{X}\beta). \quad (3.16)$$

Por conveniência, escreve-se

$$P = \Sigma^{-1} - \Sigma^{-1}X(X^\top \Sigma^{-1}X)^\top \Sigma^{-1}, \quad (3.17)$$

e notando que $\Sigma^{-1}(\mathbf{y} - \mathbf{X}\beta^*) = P\mathbf{y}$, onde β^* é a solução da equação geral normal, $X^\top \Sigma^{-1}X\beta = X^\top \Sigma^{-1}\mathbf{y}$ em β , teremos

$$\text{tr}(\Sigma^{-1}M_i) = \mathbf{y}^\top P M_i P \mathbf{y}. \quad (3.18)$$

Quando $(X^\top \Sigma^{-1}X)$ é uma matriz singular, $(X^\top \Sigma^{-1}X)^{-1}$ deve ser substituído por $(X^\top \Sigma^{-1}X)^-$.

Notando que

$$\begin{aligned} \text{tr}(\Sigma^{-1}M_i) &= \text{tr}(\Sigma^{-1}M_i \Sigma^{-1}\Sigma) \\ &= \sum_{j=1}^{n+1} \gamma_j \text{tr}(\Sigma^{-1}M_i \Sigma^{-1}M_j). \end{aligned} \quad (3.19)$$

Matricialmente, o sistema de equações em (3.17) dado por

$$\begin{bmatrix} \mathbf{y}^\top P M_1 P \mathbf{y} \\ \mathbf{y}^\top P M_2 P \mathbf{y} \\ \vdots \\ \mathbf{y}^\top P M_{n+1} P \mathbf{y} \end{bmatrix} = \begin{bmatrix} \text{tr}(\Sigma^{-1}M_1 \Sigma^{-1}M_1) & \cdots & \text{tr}(\Sigma^{-1}M_1 \Sigma^{-1}M_{n+1}) \\ \text{tr}(\Sigma^{-1}M_2 \Sigma^{-1}M_1) & \cdots & \text{tr}(\Sigma^{-1}M_2 \Sigma^{-1}M_{n+1}) \\ \vdots & \ddots & \vdots \\ \text{tr}(\Sigma^{-1}M_{n+1} \Sigma^{-1}M_1) & \cdots & \text{tr}(\Sigma^{-1}M_{n+1} \Sigma^{-1}M_{n+1}) \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_{n+1} \end{bmatrix}, \quad (3.20)$$

que é resolvido em $\gamma^\top = [\gamma_1 \ \gamma_2 \ \gamma_{n+1}]$ dá-se a estimativa desejada $\hat{\gamma}^\top = [\hat{\gamma}_1 \ \hat{\gamma}_2 \ \hat{\gamma}_{n+1}]$, onde $\hat{\gamma}$ é chamado estimador de ML para γ .

3.3.3 Método da Máxima Verossimilhança Restrita

Os estimadores das componentes da variância, pelo método da máxima verossimilhança restrita(REML), tem sido amplamente adotados, porque eliminam o primeiro dos problemas encontrados no método ML, ou seja, leva em consiração os graus de liberdade envolvidos na estimação dos parâmetros fixos do modelo. Sendo assim, estimativas REML das componentes da variância tendem a ser menos viesadas que as estimativas de ML, permite também a imposição de restrições de não negatividade e ainda o método REML difere de ML pelo fato da verossimilhança dos dados a ser maximizada somente para efeitos aleatórias, sendo assim, REML é chamado de solução restrita. Dessa forma, o método REML é o procedimento ideal de estimação dos componentes de variância com dados desbalanceados.

O método REML, pode ser aplicado a dados desbalanceados, permite ajustar modelos que não podem ser acomodados pela ANOVA e também ajustar vários modelos alternativos, escolhendo o que se ajusta melhor os dados.

O REML é uma variante de ML, para modelos mistos e foi utilizada por Patterson e Thompson em 1971 para delineamento em blocos. Os estimadores REML são obtidos maximizando a parte da função de verossilhança de uma matriz de combinações lineares das observações que são invariantes para o parâmetro de locação, ou seja, em termos de modelos mistos $y = X\beta + \sum_{i=1}^{n+1} X_i\beta_i$ é invariante para $X\beta$. Seja L essa matriz, então, a distribuição $z = Ly$, se e somente se, $LX = \mathbf{0}_{n,k}$, onde n é o número de linhas de L .

Em z , a dependência dos efeitos fixos β é eliminado, portanto z terá menos graus de liberdade do que y , consequentemente os estimadores baseados viés, por isso, que o método de REML é preferível em vez de ML.

Recordando que $y \sim \mathcal{N}(X\beta, \Sigma)$, e usando o teorema A.1.8 (ver Silva et al. [17]), obtemos

$$z = Ly \sim \mathcal{N}(0_n, \Sigma^\bullet), \quad (3.21)$$

onde, $\Sigma^\bullet = \sum_{i=1}^{n+1} \gamma_i LM_i L^\top$ e L deve ser de rank completo com o número máximo de linhas e um elemento de $\mathcal{A}_{(m-r) \times m}$, onde $r = r(X)$ e da forma $L = B(I - X(X^\top X)^{-1}X^\top)$, onde B especifica um rank completo de transformações das linhas de $X(X^\top X)^{-1}X^\top$ e $(X^\top X)^{-1}$, deve ser substituído por (X^\top) quando $X^\top X$ é singular.

O logaritmo da função de probabilidades do modelo (3.20) será

$$l^\bullet(\gamma) = \frac{(n-r)\log(2\pi)}{2} - \frac{\log |\Sigma^\bullet|}{2} - \frac{z^\top (\Sigma^\bullet)^{-1} z}{2} \quad (3.22)$$

Derivando l^\bullet em relação aos componentes da variância, obtém-se (ver Silva et al. [17], Teoremas A.1.3 e A.1.4)

$$\begin{aligned}
 \frac{\partial l^\bullet(\gamma)}{\partial \gamma_i} &= -\frac{\text{tr}((\Sigma^\bullet)^{-1} \frac{\partial \Sigma^\bullet}{\partial \gamma_i})}{2} + \frac{z^\top ((\Sigma^\bullet)^{-1} \frac{\partial \Sigma^\bullet}{\partial \gamma_i} (\Sigma^\bullet)) z}{2} \\
 &= -\frac{\text{tr}((\Sigma^\bullet)^{-1} \sum_{j=1}^{n+1} \frac{\partial}{\partial \gamma_i} \gamma_j LM_j L^\top)}{2} \\
 &\quad + \frac{z^\top ((\Sigma^\bullet)^{-1} \frac{\partial}{\partial \gamma_i} (\sum_{j=1}^{n+1} \gamma_j LM_j L^\top) (\Sigma^\bullet)^{-1}) z}{2} \\
 &= \frac{\text{tr}((\Sigma^\bullet)^{-1} LM_i L^\top)}{2} + \frac{z^\top ((\Sigma^\bullet)^{-1} LM_i L^\top (\Sigma^\bullet)^{-1}) z}{2} \\
 &= 0, \quad i = 1, \dots, n+1,
 \end{aligned} \tag{3.23}$$

De acordo com (3.22), concluimos que (ver Silva et al. [17], Proposição 2.1.1)

$$\begin{aligned}
 \text{tr}((\Sigma^\bullet)^{-1} LM_i L^\top) &= z^\top ((\Sigma^\bullet)^{-1} LM_i L^\top (\Sigma^\bullet)^{-1}) z \\
 &= \text{tr}((\Sigma^\bullet)^{-1} LM_i L^\top (\Sigma^\bullet)^{-1} \Sigma^\bullet) \\
 &= \text{tr}((\Sigma^\bullet)^{-1} LM_i L^\top (\Sigma^\bullet)^{-1} \sum_{j=1}^{n+1} \gamma_j LM_j L^\top) \\
 &= \sum_{j=1}^{n+1} \gamma_j \text{tr}((\Sigma^\bullet)^{-1} LM_i L^\top (\Sigma^\bullet)^{-1} LM_j L^\top) \\
 &= \sum_{j=1}^{n+1} \gamma_j \text{tr}((L^\top \Sigma^\bullet)^{-1} LM_i L^\top (\Sigma^\bullet)^{-1} LM_j).
 \end{aligned} \tag{3.24}$$

Assim, a equação acima se torna

$$\psi = M\gamma \tag{3.25}$$

$$\text{onde } \psi = \begin{bmatrix} z^\top ((\Sigma^\bullet)^{-1} LM_1 L^\top (\Sigma^\bullet)^{-1}) z \\ z^\top ((\Sigma^\bullet)^{-1} LM_2 L^\top (\Sigma^\bullet)^{-1}) z \\ \vdots \\ z^\top ((\Sigma^\bullet)^{-1} LM_{n+1} L^\top (\Sigma^\bullet)^{-1}) z \end{bmatrix}, \quad \gamma = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_{n+1} \end{bmatrix} \text{ e}$$

$$M = \begin{bmatrix} \text{tr}(L^\top \Sigma^{-1} LM_1 L^\top \Sigma^{-1} LM_1) & \cdots & \text{tr}(L^\top \Sigma^{-1} LM_1 L^\top \Sigma^{-1} LM_{n+1}) \\ \text{tr}(L^\top \Sigma^{-1} LM_2 L^\top \Sigma^{-1} LM_1) & \cdots & \text{tr}(L^\top \Sigma^{-1} LM_2 L^\top \Sigma^{-1} LM_{n+1}) \\ \vdots & \ddots & \vdots \\ \text{tr}(L^\top \Sigma^{-1} LM_{n+1} L^\top \Sigma^{-1} LM_1) & \cdots & \text{tr}(L^\top \Sigma^{-1} LM_{n+1} L^\top \Sigma^{-1} LM_{n+1}) \end{bmatrix}.$$

INFERÊNCIA EM MLM COM 4 COMPONENTES DA VARIÂNCIA

4.1 Introdução

Neste capítulo introduzimos e discutimos o método Sub-D, para MLM com 4 componentes da variância, que é um método desenvolvido por Silva et al. [17]. O desempenho do método Sub-D será comparado ao desempenho dos métodos baseados em ANOVA e REML.

Na subseção (4.2.1), deduziremos o método *Sub-D* para MLM com 4 componentes da variância. Finalmente, na seção (4.3), realizaremos testes numéricos utilizando o método *Sub-D* para MLM com 4 componentes da variância.

4.2 Método Sub-D

O método *Sub-D* (ver Silva et al. [17]), é um método desenvolvido para estimar os componentes da variância em MLM com um número arbitrário de componentes da variância, tendo-se revelado um método extremamente eficiente, uma vez que produzimos estimativas centradas em todos os modelos simulados: *One way* e *two way cruzados* e *aninhados*, ao contrário dos métodos de ANOVA e REML.

Segundo Silva o método ANOVA revelou-se eficiente somente no *modelo one way equilibrados*, enquanto que o REML, apesar do desempenho eficiente os *modelos one way* e *two way cruzados* simulados, mantém um menor erro quadrático médio (e.q.m), produzindo estimativas com baixa eficiência nos modelos aninhados.

O método *Sub-D* mantém um desempenho constante para todos os modelos que foram aplicados, fornecendo estimativas sempre precisas, ao contrário de REML, que por exemplo, em modelos não equilibrados *two way aninhados* (ver Silva et al. [17], seção 5.4) fornece

estimativas não centradas, embora parece ter melhor desempenho para componentes da variância com valores maiores que 1. No caso de ANOVA, as estimativas são extremamente viesadas, uma vez que fornece estimativas não centradas tanto para *modelos aninhados* e *cruzados*.

Quanto à sua eficiência e o tempo de execução, calcularam as estimativas e o desvio padrão correspondente ao modelo *two way aninhado*, γ_1 , γ_2 e γ_3 , tomando valores 0.25, 0.5, 0.75, 1, 2, 5, para 1000 observações, concluíram que o tempo de execução para ANOVA e *Sub-D* são respectivamente 1.2471 e 2.06 segundos, enquanto que o tempo de execução do estimador REML é cerca de 6.2618 minutos, o que significa que ANOVA e *Sub-D* são mais de 187 vezes mais rápido que o do REML.

4.2.1 Dedução do Método Sub-D para MLM com 4 Componentes da Variância

Nesta subseção, começamos a nossa abordagem deduzindo o *Método Sub-D* para modelos mistos lineares com 4 componentes da variância:

$$y = X\beta + X_1\beta_1 + X_2\beta_2 + X_3\beta_3 + \varepsilon, \quad (4.1)$$

onde $y \in \mathcal{A}_{m \times 1}$, denotado vetor das observações, com $E(y) = X\beta$ e $\Sigma(y) = \gamma_1 M_1 + \gamma_2 M_2 + \gamma_3 M_3 + \gamma_4 I$, $\beta_1 \sim (0, \gamma_1 I)$, $\beta_2 \sim (0, \gamma_2 I)$, $\beta_3 \sim (0, \gamma_3 I)$, com $\beta_i \in \mathcal{A}_{q_i \times 1}$, vetor dos parâmetros de efeitos aleatórios desconhecidos, enquanto que $\beta \in \mathcal{A}_{p \times 1}$, vetor de efeitos fixos desconhecidos, $X_i \in \mathcal{A}_{m \times q_i}$, matriz de efeito aleatória conhecida, $M_d \in \mathcal{A}_{m \times m}$ e $\varepsilon \in \mathcal{A}_{m \times 1}$, vetor de erros aleatórios desconhecidos.

O MLM, pode ser escrito como:

$$y \sim (X\beta, \sum_{d=1}^4 \gamma_d M_d), \quad (4.2)$$

onde $\gamma_d > 0$, $d = 1, \dots, r$ são os parâmetros desconhecidos, chamados de componentes da variância, $M_d = X_d X_d^T \in \mathcal{L}^m$, com $\mathcal{L}^m = \{A : A \in \mathcal{A}_{m \times m}, A = A^T\}$, assumidos como matrizes semidefinidas positivas e representam o conjunto das matrizes simétricas, sendo que $X_d \in \mathcal{A}_{m \times m_d}$ matrizes conhecidas, denominadas matrizes dos efeitos aleatórios e $M_{r+1} = I_m$.

Seja $P_\circ = P_{R(X)}$, a matriz de projeção ortogonal sobre o subespaço gerado pelas colunas da matriz X e $P^* = P_{R(X)^\perp} = I_m - P_\circ$, a matriz de projeção ortogonal sobre o complemento ortogonal do espaço das colunas de X . Seja B_\circ , a matriz cujas colunas são os vetores próprios associados aos valores próprios nulos de P_\circ , pelo que $B_\circ^T B_\circ = I_{m-r(P_\circ)}$ e $B_\circ B_\circ^T = P^*$.

Assim, produzimos o novo modelo, $z = B_\circ^T y$ (*modelo misto restrito*) que denotamos por,

$$z = B_\circ^T y \sim (0_n, \sum_{d=1}^4 \gamma_d N_d), \quad (4.3)$$

onde $N_d = B_o^T M_d B_o$, $n = m - r(P_o)$, e $0_n \in \mathcal{A}_{n \times 1}$, vetor de zeros.

Para prosseguir com a dedução do *método Sub-D*, introduzimos a noção, que pensamos ser indispensável para a mesma.

Definição 4.2.1. *Seja $A \in \mathcal{A}_{n \times n}$, uma matriz em bloco. Se TA produz uma matriz em bloco, cujas matrizes na diagonal são todas matrizes diagonais, dizemos que a matriz T sub - diagonaliza A , isto é, T diagonaliza as matrizes A_{11}, \dots, A_{nn} na diagonal de A .*

A dedução do *método Sub-D* passa fundamentalmente pela sub-diagonalização da matriz variância-covariância, $\sum_{d=1}^4 \gamma_d N_d$ (modelo misto restrito em (4.3)), ou seja,

$$z \sim \mathcal{N}_n(0_n, \gamma_1 N_1 + \gamma_2 N_2 + \gamma_3 N_3 + \gamma_4 I_n), \quad (4.4)$$

onde N_1 é uma matriz simétrica. Nessas condições existe uma matriz ortogonal

$$P_1 = \begin{bmatrix} A_{11} \\ A_{12} \\ \vdots \\ A_{1h_1} \end{bmatrix} \in \mathcal{A}_{(\sum_{i=1}^{h_1} g_i) \times n'} \quad (4.5)$$

(ver Schott [15], capítulo 4, seção 3 e 4), com $A_{1i} \in \mathcal{A}_{g_i \times n} (\sum_{i=1}^{h_1} g_i = n)$, de modo que $N_1 = P_1^T D_1 P_1$ equivalente a $P_1 N_1 P_1^T = D_1$, com

$$D_1 = \begin{bmatrix} \theta_{11} I_{g_1} & 0 & \dots & 0 \\ 0 & \theta_{12} I_{g_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \theta_{1h_1} I_{gh_1} \end{bmatrix}, \quad (4.6)$$

onde D_1 é uma matriz diagonal, em que as entradas na diagonal $\theta_{1i}, i = 1, \dots, h_1$ são os valores próprios da matriz N_1 , com multiplicidade $g_i = r(A_{1i}^T)$. O conjunto das colunas de cada matriz A_{1i}^T formam um conjunto de vetores ortonormais g_i associados ao valor próprio θ_{1i} da matriz N_1 (ver Silva et al. [17], Teorema 2.1.8), de modo que, $A_{1i}^T A_{1i} = P_{R(A_{1i}^T)}$ e $A_{1i} A_{1i}^T = I_{g_i}$.

Por essa razão, $P_1 P_1^T = I_n$, e

$$\begin{aligned} P_1^T P_1 &= A_{11}^T A_{11} + \dots + A_{1h_1}^T A_{1h_1} \\ &= P_{R(A_{11}^T)} + \dots + P_{R(A_{1h_1}^T)} \\ &= I_n. \end{aligned} \quad (4.7)$$

Considere-se,

$$A_{1i} N_2 A_{1s}^T = \begin{cases} N_{ii}^2 & i = s \\ W_{is}^2 & i \neq s \end{cases} \quad (4.8)$$

e

$$A_{1i}N_3A_{1s}^\top = \begin{cases} O_{ii}^2 & i = s \\ U_{is}^2 & i \neq s. \end{cases} \quad (4.9)$$

Assim, teremos

$$\begin{aligned} \text{cov}(P_1z) &= \gamma_1 P_1 N_1 P_1^\top + \gamma_2 P_1 N_2 P_1^\top + \gamma_3 P_1 N_3 P_1^\top + \gamma_4 P_1 P_1^\top \\ &= \gamma_1 \begin{bmatrix} \theta_{11} I_{g1} & 0 & \dots & 0 \\ 0 & \theta_{12} I_{g2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \theta_{1h_1} I_{gh_1} \end{bmatrix} + \gamma_2 \begin{bmatrix} N_{11}^2 & W_{12}^2 & \dots & W_{1h_1}^2 \\ W_{21}^2 & N_{22}^2 & \dots & W_{2h_1}^2 \\ \vdots & \vdots & \ddots & \vdots \\ W_{h_11}^2 & W_{h_12}^2 & \dots & N_{h_1h_1}^2 \end{bmatrix} \\ &+ \gamma_3 \begin{bmatrix} O_{11}^2 & U_{12}^2 & \dots & U_{1h_1}^2 \\ U_{21}^2 & O_{22}^2 & \dots & U_{2h_1}^2 \\ \vdots & \vdots & \ddots & \vdots \\ U_{h_11}^2 & U_{h_12}^2 & \dots & O_{h_1h_1}^2 \end{bmatrix} + \gamma_4 \begin{bmatrix} I_{g1} & 0 & \dots & 0 \\ 0 & I_{g2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & I_{gh_1} \end{bmatrix} \\ &= \gamma_1 D(\theta_1 I_{g1} \dots \theta_{h_1} I_{gh_1}) + \gamma_2 \Gamma_1 + \gamma_3 \Gamma_2 + \gamma_4 D(I_{g1} \dots I_{gh_1}), \end{aligned} \quad (4.10)$$

onde

$$\Gamma_1 = \begin{bmatrix} N_{11}^2 & W_{12}^2 & \dots & W_{1h_1}^2 \\ W_{21}^2 & N_{22}^2 & \dots & W_{2h_1}^2 \\ \vdots & \vdots & \ddots & \vdots \\ W_{h_11}^2 & W_{h_12}^2 & \dots & N_{h_1h_1}^2 \end{bmatrix} \quad \text{e} \quad \Gamma_2 = \begin{bmatrix} O_{11}^2 & U_{12}^2 & \dots & U_{1h_1}^2 \\ U_{21}^2 & O_{22}^2 & \dots & U_{2h_1}^2 \\ \vdots & \vdots & \ddots & \vdots \\ U_{h_11}^2 & U_{h_12}^2 & \dots & O_{h_1h_1}^2 \end{bmatrix}.$$

Agora, diagonalizamos as matrizes simétricas N_{ii}^2 , ou seja, sub-diagonalizaremos Γ_1 .

Sabendo que, N_{ii}^2 são matrizes simétricas (ver Schott [15], capítulo 4, seção 3 e 4),

existe uma matriz ortogonal $P_{2i} = \begin{bmatrix} A_{2i1} \\ A_{2i2} \\ \vdots \\ A_{2ih_{2i}} \end{bmatrix} \in \mathcal{A}_{(\sum_{j=1}^{h_{2i}} g_{ij}) \times g_i}$, onde $A_{2ij} \in \mathcal{A}_{g_{ij} \times g_i} (\sum_{j=1}^{h_{2i}} g_{ij} = g_i)$,

de modo que,

$$P_{2i} N_{ii}^2 P_{2i}^\top = \begin{bmatrix} \theta_{2i1} I_{g_{i1}} & 0 & \dots & 0 \\ 0 & \theta_{2i2} I_{g_{i2}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \theta_{2ih_{2i}} I_{g_{ih_{2i}}} \end{bmatrix} = D_{ii}^2, \quad i = 1, \dots, h_1. \quad (4.11)$$

g_{ij} é a multiplicidade de valores próprios θ_{2ij} , e $A_{2ij}^\top A_{2ij} = P_{R(A_{2ij}^\top)}$ e $A_{2ij} A_{2ij}^\top = I_{g_{ij}}$, onde A_{2ij}^\top , $i = 1, \dots, h_1$, $j = 1, \dots, h_{2i}$, é uma matriz ortogonal.

Seja,

$$P_2 = \begin{bmatrix} P_{21} & 0 & \dots & 0 \\ 0 & P_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & P_{2h_1} \end{bmatrix} \in \mathcal{A}_{(\sum_{i=1}^{h_1} \sum_{j=1}^{h_{2i}} g_{ij}) \times (\sum_{i=1}^{h_1} g_i)}. \quad (4.12)$$

Prova-se facilmente que P_2 , é ortogonal, uma vez que P_{2i} é ortogonal (ver Silva et al. [17], Proposição 4.2.1).

O novo modelo $w_2 = P_2 P_1 z$ terá a matriz variância-covariância dado por,

$$\begin{aligned} \text{cov}(w_2) = \Sigma(P_2 P_1 z) &= \gamma_1 P_2 D(\theta_1 I_{g_1} \dots \theta_{h_1} I_{g_{h_1}}) P_2^\top + \gamma_2 P_2 \Gamma_1 P_2^\top + \gamma_3 P_2 \Gamma_2 P_2^\top + \gamma_4 P_2 D(I_{g_1} \dots I_{g_{h_1}}) P_2^\top \\ &= \gamma_1 \begin{bmatrix} \theta_{11} P_{21} P_{21}^\top & 0 & \dots & 0 \\ 0 & \theta_{12} P_{22} P_{22}^\top & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \theta_{1h_1} P_{2h_1} P_{2h_1}^\top \end{bmatrix} \\ &+ \gamma_2 \begin{bmatrix} D_{11}^2 & P_{21} W_{12}^2 P_{22}^\top & \dots & P_{21} W_{1h_1}^2 P_{2h_1}^\top \\ P_{22} W_{21}^2 P_{21}^\top & D_{22}^2 & \dots & P_{22} W_{2h_1}^2 P_{2h_1}^\top \\ \vdots & \vdots & \ddots & \vdots \\ P_{2h_1} W_{h_1 1}^2 P_{21}^\top & P_{2h_1} W_{h_1 2}^2 P_{22}^\top & \dots & D_{h_1 h_1}^2 \end{bmatrix} \\ &+ \gamma_3 \begin{bmatrix} P_{21} O_{11}^2 P_{21}^\top & P_{21} U_{12}^2 P_{22}^\top & \dots & P_{21} U_{1h_1}^2 P_{2h_1}^\top \\ P_{22} U_{21}^2 P_{21}^\top & P_{22} O_{22}^2 P_{22}^\top & \dots & P_{22} U_{2h_1}^2 P_{2h_1}^\top \\ \vdots & \vdots & \ddots & \vdots \\ P_{2h_1} U_{h_1 1}^2 P_{21}^\top & P_{2h_1} U_{h_1 2}^2 P_{22}^\top & \dots & P_{2h_1} O_{h_1 h_1}^2 P_{2h_1}^\top \end{bmatrix} \\ &+ \gamma_4 \begin{bmatrix} P_{21} P_{21}^\top & 0 & \dots & 0 \\ 0 & P_{22} P_{22}^\top & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & P_{2h_1} P_{2h_1}^\top \end{bmatrix}. \end{aligned} \quad (4.13)$$

Note-se que,

$$P_{2i} P_{2i}^\top = \begin{bmatrix} A_{2i1} A_{2i1}^\top & 0 & \dots & 0 \\ 0 & A_{2i2} A_{2i2}^\top & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & A_{2ih_{2i}} A_{2ih_{2i}}^\top \end{bmatrix} = \begin{bmatrix} I_{g_{i1}} & 0 & \dots & 0 \\ 0 & I_{g_{i2}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & I_{g_{ih_{2i}}} \end{bmatrix}.$$

Para $i \neq s$,

$$P_{2i} W_{is}^2 P_{2s}^\top = \begin{bmatrix} A_{2i1} W_{is}^2 A_{2s1}^\top & A_{2i1} W_{is}^2 A_{2s2}^\top & \dots & A_{2i1} W_{is}^2 A_{2sh_{2s}}^\top \\ A_{2i2} W_{is}^2 A_{2s1}^\top & A_{2i2} W_{is}^2 A_{2s2}^\top & \dots & A_{2i2} W_{is}^2 A_{2sh_{2s}}^\top \\ \vdots & \vdots & \ddots & \vdots \\ A_{2ih_{2i}} W_{is}^2 A_{2s1}^\top & A_{2ih_{2i}} W_{is}^2 A_{2s2}^\top & \dots & A_{2ih_{2i}} W_{is}^2 A_{2sh_{2s}}^\top \end{bmatrix}$$

e

$$P_{2i}U_{is}^2P_{2s}^\top = \begin{bmatrix} A_{2i1}U_{is}^2A_{2s1}^\top & A_{2i1}U_{is}^2A_{2s2}^\top & \dots & A_{2i1}U_{is}^2A_{2sh_{2s}}^\top \\ A_{2i2}U_{is}^2A_{2s1}^\top & A_{2i2}U_{is}^2A_{2s2}^\top & \dots & A_{2i2}U_{is}^2A_{2sh_{2s}}^\top \\ \vdots & \vdots & \ddots & \vdots \\ A_{2ih_{2i}}U_{is}^2A_{2s1}^\top & A_{2ih_{2i}}U_{is}^2A_{2s2}^\top & \dots & A_{2ih_{2i}}U_{is}^2A_{2sh_{2s}}^\top \end{bmatrix}.$$

A matriz D_{ii}^2 , $i = 1, \dots, h_1$, que figura na diagonal da matriz do lado direito de (4.13), está definida em (4.11).

A seguir sub-diagonalizamos a matriz,

$$\begin{bmatrix} P_{21}O_{11}^2P_{21}^\top & P_{21}U_{12}^2P_{22}^\top & \dots & P_{21}U_{1h_1}^2P_{2h_1}^\top \\ P_{22}U_{21}^2P_{21}^\top & P_{22}O_{22}^2P_{22}^\top & \dots & P_{22}U_{2h_1}^2P_{2h_1}^\top \\ \vdots & \vdots & \ddots & \vdots \\ P_{2h_1}U_{h_11}^2P_{21}^\top & P_{2h_1}U_{h_12}^2P_{22}^\top & \dots & P_{2h_1}O_{h_1h_1}^2P_{2h_1}^\top \end{bmatrix}, \quad (4.14)$$

ou seja, diagonalizaremos as matrizes $P_{2i}O_{ii}^2P_{2i}^\top$, $i = 1, \dots, h_1$, que figuram em (4.13).

Existe uma matriz ortogonal, $P_{3ij} = \begin{bmatrix} A_{3ij1} \\ A_{3ij2} \\ \vdots \\ A_{3ijh_{2ij}} \end{bmatrix} \in \mathcal{A}_{(\sum_{j=1}^{h_{2i}} \sum_{k=1}^{h_{2ij}} g_{ijk}) \times (\sum_{i=1}^{h_1} \sum_{j=1}^{h_{2i}} g_{ij})}$, tal que

$$P_{3ij}P_{2i}O_{ii}^2P_{2i}^\top P_{3ij}^\top = \begin{bmatrix} \theta_{3ij1}I_{g_{ij1}} & 0 & \dots & 0 \\ 0 & \theta_{3ij2}I_{g_{ij2}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \theta_{3ijh_{2ij}}I_{g_{ijh_{2ij}}} \end{bmatrix} = D_{iii}^3, \quad (4.15)$$

onde $i = 1, \dots, h_1$, $j = 1, \dots, h_{2i}$, uma vez que, O_{ii}^2 são matrizes simétricas.

As colunas da matriz A_{2ijk}^\top , $i = 1, \dots, h_1$, $j = 1, \dots, h_{2i}$, $k = 1, \dots, h_{2ij}$, são vetores próprios ortonormais associados ao valor próprio θ_{2ijk} da matriz $P_{3ij}P_{2i}O_{ii}^2P_{2i}^\top P_{3ij}^\top$ e g_{ijk} é a multiplicidade de valores próprios θ_{2ijk} , $A_{2ijk}^\top A_{2ijk} = P_{R(A_{2ijk}^\top)}$ e $A_{2ijk}A_{2ijk}^\top = I_{g_{ijk}}$.

Seja

$$P_3 = \begin{bmatrix} P_{31} & 0 & \dots & 0 \\ 0 & P_{32} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & P_{3h_1} \end{bmatrix} \in \mathcal{A}_{(\sum_{i=1}^{h_1} \sum_{j=1}^{h_{2i}} g_{ij} \sum_{k=1}^{h_{2ij}} g_{ijk}) \times (\sum_{i=1}^{h_1} g_i \sum_{j=1}^{h_{2i}} g_{ij})}. \quad (4.16)$$

Proposição 4.2.1. P_3 é uma matriz ortogonal.

Demonstração. Tem-se que,

$$P_3 P_3^T = \begin{bmatrix} P_{31} P_{31}^T & 0 & \dots & 0 \\ 0 & P_{32} P_{32}^T & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & P_{3h_1} P_{3h_1}^T \end{bmatrix},$$

onde

$$P_{3ij} P_{3ij}^T = \begin{bmatrix} A_{3ij1} A_{3ij1}^T & 0 & \dots & 0 \\ 0 & A_{3ij2} A_{3ij2}^T & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & A_{3ijh_{2ij}} A_{3ijh_{2ij}}^T \end{bmatrix}.$$

Recorde-se que, $A_{3ijk} A_{3ijk}^T = I_{g_{ijk}}$ e $P_3 P_3^T = I_{\sum_{i=1}^{h_1} g_i}$, então,

$$P_3^T P_3 = \begin{bmatrix} P_{31}^T P_{31} & 0 & \dots & 0 \\ 0 & P_{32}^T P_{32} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & P_{3h_1}^T P_{3h_1} \end{bmatrix},$$

e pelo Teorema 2.2.8 (ver Silva et al. [17]), temos

$$\begin{aligned} P_{3ij}^T P_{3ij} &= A_{3ij1}^T A_{3ij1} + \dots + A_{3ijh_{2ij}}^T A_{3ijh_{2ij}} \\ &= P_{R(A_{3ij1}^T)} + \dots + P_{R(A_{3ijh_{2ij}}^T)} \\ &= I_{g_{ij}}. \end{aligned} \tag{4.17}$$

□

O novo modelo $w_3 = P_3 P_2 P_1 z$ terá a matriz variância-covariância dada por,

$$\begin{aligned}
 cov(w_3) = \Sigma(P_3 P_2 P_1 z) &= \gamma_1 P_3 P_2 D(\theta_1 I_{g_1} \dots \theta_{h_1} I_{g_{h_1}}) P_2^\top P_3^\top + \gamma_2 P_3 P_2 \Gamma_1 P_2^\top P_3^\top \\
 &+ \gamma_3 P_3 P_2 \Gamma_2 P_2^\top P_3^\top + \gamma_4 P_3 P_2 D(I_{g_1} \dots I_{g_{h_1}}) P_2^\top P_3^\top \\
 &= \gamma_1 \begin{bmatrix} \theta_{11} P_{31} P_{21} P_{21}^\top P_{31}^\top & 0 & \dots & 0 \\ 0 & \theta_{12} P_{32} P_{22} P_{22}^\top P_{32}^\top & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \theta_{1h_1} P_{3h_1} P_{2h_1} P_{2h_1}^\top P_{3h_1}^\top \end{bmatrix} \\
 &+ \gamma_2 \begin{bmatrix} P_{31} D_{11}^2 P_{31}^\top & P_{31} P_{21} W_{12}^2 P_{22}^\top P_{32}^\top & \dots & P_{31} P_{21} W_{1h_1}^2 P_{2h_1}^\top P_{3h_1}^\top \\ P_{32} P_{22} W_{21}^2 P_{21}^\top P_{31}^\top & P_{32} D_{22}^2 P_{32}^\top & \dots & P_{32} P_{22} W_{2h_1}^2 P_{2h_1}^\top P_{3h_1}^\top \\ \vdots & \vdots & \ddots & \vdots \\ P_{3h_1} P_{2h_1} W_{h_1 1}^2 P_{21}^\top P_{31}^\top & P_{3h_1} P_{2h_1} W_{h_1 2}^2 P_{22}^\top P_{32}^\top & \dots & P_{3h_1} D_{h_1 h_1}^2 P_{3h_1}^\top \end{bmatrix} \\
 &+ \gamma_3 \begin{bmatrix} D_{11}^3 & P_{31} P_{21} U_{12}^2 P_{22}^\top P_{32}^\top & \dots & P_{31} P_{21} U_{1h_1}^2 P_{2h_1}^\top P_{3h_1}^\top \\ P_{32} P_{22} U_{21}^2 P_{21}^\top P_{31}^\top & D_{22}^3 & \dots & P_{32} P_{22} U_{2h_1}^2 P_{2h_1}^\top P_{3h_1}^\top \\ \vdots & \vdots & \ddots & \vdots \\ P_{3h_1} P_{2h_1} U_{h_1 1}^2 P_{21}^\top P_{31}^\top & P_{3h_1} P_{2h_1} U_{h_1 2}^2 P_{22}^\top P_{32}^\top & \dots & D_{h_1 h_1}^3 \end{bmatrix} \\
 &+ \gamma_4 \begin{bmatrix} P_{31} P_{21} P_{21}^\top P_{31}^\top & 0 & \dots & 0 \\ 0 & P_{32} P_{22} P_{22}^\top P_{32}^\top & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & P_{3h_1} P_{2h_1} P_{2h_1}^\top P_{3h_1}^\top \end{bmatrix}. \tag{4.18}
 \end{aligned}$$

Note-se que,

$$\begin{aligned}
 P_{3i} P_{2i} P_{2i}^\top P_{3i}^\top &= \begin{bmatrix} A_{3i1} A_{2i1} A_{2i1}^\top A_{3i1}^\top & 0 & \dots & 0 \\ 0 & A_{3i2} A_{2i2} A_{2i2}^\top A_{3i2}^\top & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & A_{3ih_1} A_{2ih_1} A_{2ih_1}^\top A_{3ih_1}^\top \end{bmatrix} \\
 &= \begin{bmatrix} I_{gi1} & 0 & \dots & 0 \\ 0 & I_{gi2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & I_{gih_1} \end{bmatrix}.
 \end{aligned}$$

Repare que,

$$w_3 = P_3 P_2 P_1 z = \begin{bmatrix} A_{211} A_{11} z \\ \vdots \\ A_{21h_{21}} A_{11} z \\ A_{221} A_{12} z \\ \vdots \\ A_{22h_{22}} A_{12} z \\ A_{231} A_{13} z \\ \vdots \\ A_{23h_{23}} A_{13} z \\ \vdots \\ \vdots \\ A_{2h_1 1} A_{1h_1} z \\ \vdots \\ A_{2h_1 h_{2h_1}} A_{1h_1} z \end{bmatrix},$$

onde a matriz $P_3 P_2 P_1$ sub-diagonaliza a matriz variância-covariância, $\sum_{d=1}^4 \gamma_d N_d$.

A distribuição dos sub-modelos

$$z_{ijk} = A_{3ijk} A_{2ij} A_{1i} z, \quad i = 1, \dots, h_1, \quad j = 1, \dots, h_{2i}, \quad k = 1, \dots, h_{2ij},$$

resume-se na proposição seguinte.

Proposição 4.2.2. *Seja $z_{ijk} \sim \mathcal{N}_{g_{ijk}}(\mathbf{0}_{g_{ijk}}, \lambda_{ijk} I_{g_{ijk}})$, $i = 1, \dots, h_1$; $j = 1, \dots, h_{2i}$; $k = 1, \dots, h_{2ij}$,*

onde $\lambda_{ijk} = \gamma_1 \theta_{1i} + \gamma_2 \theta_{2ij} + \gamma_3 \theta_{3ijk} + \gamma_4$.

Demonstração. Recorde que, $A_{3ijk} A_{2ij} A_{1i} \in \mathcal{A}_{g_{ijk} \times n}$ e $g_{ijk} \leq n$, de acordo com o Teorema A.1.8(c) (ver Silva et al. [17]), teremos que,

$$z_{ijk} \sim (\mathbf{0}_{g_{ijk}}, \sum_{d=1}^3 \gamma_d A_{3ijk} A_{2ij} A_{1i} N_d A_{1i}^\top A_{2ij}^\top A_{3ijk}^\top + \gamma_4 A_{3ijk} A_{2ij} A_{1i} A_{1i}^\top A_{2ij}^\top A_{3ijk}^\top).$$

As porções, $\sum_{d=1}^3 \gamma_d A_{3ijk} A_{2ij} A_{1i} N_d A_{1i}^\top A_{2ij}^\top A_{3ijk}^\top$ e $\gamma_4 A_{3ijk} A_{2ij} A_{1i} A_{1i}^\top A_{2ij}^\top A_{3ijk}^\top$ da matriz variância-covariância produz respectivamente:

$$\begin{aligned} \sum_{d=1}^3 \gamma_d A_{3ijk} A_{2ij} A_{1i} N_d A_{1i}^\top A_{2ij}^\top A_{3ijk}^\top &= \gamma_1 A_{3ijk} A_{2ij} (\theta_{1i} I_{g_i}) A_{2ij}^\top A_{3ijk}^\top + \gamma_2 A_{3ijk} A_{2ij} N_{ii}^2 A_{2ij}^\top A_{3ijk}^\top + \gamma_3 A_{3ijk} N_{ii}^2 A_{3ijk}^\top \\ &= \gamma_1 \theta_{1i} I_{g_{ijk}} + \gamma_2 \theta_{2ij} I_{g_{ijk}} + \gamma_3 \theta_{3ijk} I_{g_{ijk}} \end{aligned}$$

e

$$\begin{aligned}
 \gamma_4 A_{3ijk} A_{2ij} A_{1i} A_{1i}^\top A_{2ij}^\top A_{3ijk}^\top &= \gamma_4 A_{3ijk} A_{2ij} I_{g_i} A_{2ij}^\top A_{3ijk}^\top \\
 &= \gamma_4 A_{3ijk} I_{g_{ij}} A_{3ijk}^\top \\
 &= \gamma_4 I_{g_{ijk}}
 \end{aligned}$$

□

Proposição 4.2.3. *Seja $i \leq s$; $j \leq r$ e $k \leq x$ (aplica-se a simetria). Então*

$$\begin{aligned}
 cov(z_{ijk}, z_{srx}) &= \gamma_3 A_{3ijk} A_{2ij} A_{1i} N_3 A_{1s}^\top A_{2sr}^\top A_{3srx}^\top \\
 &= \begin{cases} 0 & i = s; j \neq r; k \neq x \\ \lambda_{ijk} I_{g_{ijk}} & i = s; j = r; k = x \\ \gamma_3 A_{3ijk} A_{2ij} A_{1i} N_3 A_{1s}^\top A_{2sr}^\top A_{3srx}^\top & i \neq s. \end{cases} \quad (4.19)
 \end{aligned}$$

Demonstração.

$$\begin{aligned}
 cov(z_{ijk}, z_{srx}) &= A_{3ijk} A_{2ij} A_{1i} cov(z) A_{1s}^\top A_{2sr}^\top A_{3srx}^\top \\
 &= A_{3ijk} A_{2ij} A_{1i} (\gamma_1 N_1 + \gamma_2 N_2 + \gamma_3 N_3 + \gamma_4 N_4) A_{1s}^\top A_{2sr}^\top A_{3srx}^\top \\
 &= \gamma_1 C_1 + \gamma_2 C_2 + \gamma_3 C_3 + \gamma_4 C_4, \quad (4.20)
 \end{aligned}$$

onde $C_d = A_{3ijk} A_{2ij} A_{1i} N_d A_{1s}^\top A_{2sr}^\top A_{3srx}^\top$, com $N_4 = I_n$.

Se $i = s$, $j = r$ e $k = x$, $cov(z_{ijk}) = \lambda_{ijk}$ e se $i = s$, $j = r$ e $k \neq x$, temos o seguinte:

$$\begin{cases} C_1 = \theta_{1i} A_{3ijk} A_{2ij} A_{2ir}^\top A_{3irx}^\top = 0_{g_{ijk} \cdot g_{irx}}; \\ C_2 = A_{3ijk} A_{2ij} N_{ii}^2 A_{2ir}^\top A_{3irx}^\top = 0_{g_{ijk} \cdot g_{irx}}; \\ C_3 = A_{3ijk} (I_{g_{ijk} \cdot g_{irx}}) A_{3irx} = 0_{g_{ijk} \cdot g_{irx}}; \\ C_4 = A_{3ijk} (I_{g_{ijk} \cdot g_{irx}}) A_{3irx} = 0_{g_{ijk} \cdot g_{irx}}. \end{cases}$$

E, quando $i \neq s$, teremos

$$\begin{cases} C_1 = A_{3ijk} A_{2ij} (0_{g_i \cdot g_s}) A_{2ir}^\top A_{3irx}^\top = 0_{g_{ijk} \cdot g_{irx}}; \\ C_2 = A_{3ijk} A_{2ij} W_{is}^2 A_{2sr} A_{3srx}; \\ C_3 = A_{3ijk} A_{2ij} U_{is}^2 A_{2sr} A_{3srx}; \\ C_4 = A_{3ijk} A_{2ij} (I_{g_i \cdot g_s}) A_{2sr} A_{3irx} = 0_{g_{ijk} \cdot g_{srx}}. \end{cases}$$

□

Recorde-se que, $w_3 = P_3 P_2 P_1 z$, escreve-se o sub-modelo linear fixo dado por

$$z_{ijk} \sim \mathcal{N}_{g_{ijk}}(0_{g_{ijk}}, \lambda_{ijk} I_{g_{ijk}}), \quad i = 1, \dots, h_1, j = 1, \dots, h_{2i}, k = 1, \dots, h_{2ij}, \quad (4.21)$$

do modelo $z \sim \mathcal{N}_n(0_n, \gamma_1 N_1 + \gamma_2 N_2 + \gamma_3 N_3 + \gamma_4 I_n)$, onde $\lambda_{ijk} = \gamma_1 \theta_{1i} + \gamma_2 \theta_{2ij} + \gamma_3 \theta_{3ijk} + \gamma_4$.

Um estimador imparcial para λ_{ijk} do modelo z_{ijk} é dado por (com base no estimador da máxima verossimilhança $\hat{\lambda}_{ijk}$)

$$S_{ijk}^2 = \frac{z_{ijk}^\top z_{ijk}}{g_{ijk}}, \quad i = 1, \dots, h_1, \quad j = 1, \dots, h_{2i}, \quad k = 1, \dots, h_{2ij}.$$

Pelo Teorema A.1.8(ver Silva et al. [17]),

$$E(S_{ijk}^2) = \frac{1}{g_{ijk}} \text{tr}\{\lambda_{ijk} I_{g_{ijk}}\} = \lambda_{ijk}. \quad (4.22)$$

Assim, $E(S_{ijk}^2) = \lambda_{ijk} = \gamma_1 \theta_{1i} + \gamma_2 \theta_{2ij} + \gamma_3 \theta_{3ijk} + \gamma_4$, onde $i = 1, \dots, h_1$, $j = 1, \dots, h_{2i}$, $k = 1, \dots, h_{2ij}$, tal que, com

$$S = \begin{bmatrix} S_{11}^2 \\ \dots \\ S_{1h_{21}}^2 \\ S_{21}^2 \\ \dots \\ S_{2h_{22}}^2 \\ S_{31}^2 \\ \dots \\ S_{3h_{23}}^2 \\ \dots \\ \dots \\ S_{h_1 1}^2 \\ \dots \\ S_{h_1 h_{2h_1}}^2 \end{bmatrix}, \quad \Theta = \begin{bmatrix} \theta_{11} & \theta_{211} & \theta_{3111} & 1 \\ \dots & \dots & \dots & \dots \\ \theta_{11} & \theta_{21h_{21}} & \theta_{311h_{21}} & 1 \\ \theta_{12} & \theta_{221} & \theta_{3211} & 1 \\ \dots & \dots & \dots & \dots \\ \theta_{12} & \theta_{22h_{22}} & \theta_{321h_{22}} & 1 \\ \theta_{13} & \theta_{23h_{21}} & \theta_{3311} & 1 \\ \dots & \dots & \dots & \dots \\ \theta_{13} & \theta_{23h_{23}} & \theta_{331h_{23}} & 1 \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \theta_{1h_1} & \theta_{2h_1 1} & \theta_{3h_1 h_1 1} & 1 \\ \dots & \dots & \dots & \dots \\ \theta_{1h_1} & \theta_{2h_1 h_{2h_1}} & \theta_{3h_1 h_1 h_{2h_1}} & 1 \end{bmatrix} \quad \text{e} \quad \gamma = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \end{bmatrix},$$

teremos,

$$E(S) = \Theta \gamma. \quad (4.23)$$

Igualando as variâncias λ_{ijk} , aos estimadores correspondentes, S_{ijk}^2 , produz o seguinte sistema de equações:

$$\begin{aligned}
 S_{11}^2 &= \gamma_1 \theta_{11} + \gamma_2 \theta_{211} + \gamma_3 \theta_{311} + \gamma_4; \\
 \dots &\dots \dots \dots; \\
 S_{1h_{21}}^2 &= \gamma_1 \theta_{11} + \gamma_2 \theta_{21h_{21}} + \gamma_3 \theta_{31h_{21}} + \gamma_4; \\
 S_{21}^2 &= \gamma_1 \theta_{12} + \gamma_2 \theta_{221} + \gamma_3 \theta_{321} + \gamma_4; \\
 \dots &\dots \dots \dots; \\
 S_{2h_{22}}^2 &= \gamma_1 \theta_{12} + \gamma_2 \theta_{22h_{22}} + \gamma_3 \theta_{32h_{22}} + \gamma_4; \\
 S_{31}^2 &= \gamma_1 \theta_{13} + \gamma_2 \theta_{231} + \gamma_3 \theta_{331} + \gamma_4; \\
 \dots &\dots \dots \dots; \\
 S_{3h_{23}}^2 &= \gamma_1 \theta_{13} + \gamma_2 \theta_{23h_{23}} + \gamma_3 \theta_{33h_{23}} + \gamma_4; \\
 \dots &\dots \dots \dots; \\
 \dots &\dots \dots \dots; \\
 S_{h_1 1}^2 &= \gamma_1 \theta_{1h_1} + \gamma_2 \theta_{2h_1} + \gamma_3 \theta_{3h_1} + \gamma_4; \\
 \dots &\dots \dots \dots; \\
 S_{h_1 h_{2h_1}}^2 &= \gamma_1 \theta_{1h_1} + \gamma_2 \theta_{2h_1 h_{2h_1}} + \gamma_3 \theta_{3h_1 h_{2h_1}} + \gamma_4;
 \end{aligned}$$

cuja a notação matricial é dada por,

$$S = \Theta \gamma. \tag{4.24}$$

Desde que, $\theta_{1i} \neq \theta_{1i'}, i \neq i' = 1, \dots, h_1$ (são os diferentes valores próprios de N_1), $\theta_{2ij} \neq \theta_{2ij'}, j \neq j' = 1, \dots, h_{2i}$ (são os valores próprios distintos de $N_{ii}^2 = A_{1i} N_2 A_{1i}^T$) e $\theta_{3ijk} \neq \theta_{3ijk'}, k \neq k' = 1, \dots, h_{2ij}$ (são os valores próprios distintos de $N_{iii}^3 = A_{2ij} N_3 A_{2ij}^T$), concluí-se que a matriz Θ é um rank completo, isto é, $r(\Theta) = 4$.

Pelo Teorema A.1.5(ver Silva et al. [17]),

$$\Theta^T \Theta = \begin{bmatrix}
 \sum_i^{h_1} \sum_j^{h_{2i}} \theta_{1i}^2 & \sum_i^{h_1} \sum_j^{h_{2i}} \theta_{1i} \theta_{2ij} & \sum_i^{h_1} \sum_k^{h_{2ij}} \theta_{1i} \theta_{3ijk} & \sum_i^{h_1} \sum_j^{h_{2i}} \theta_{1i} \\
 \sum_i^{h_1} \sum_j^{h_{2i}} \theta_{1i} \theta_{2ij} & \sum_i^{h_1} \sum_j^{h_{2i}} \theta_{2ij}^2 & \sum_j^{h_{2i}} \sum_k^{h_{2ij}} \theta_{2ij} \theta_{2ijk} & \sum_i^{h_1} \sum_j^{h_{2i}} \theta_{2ij} \\
 \sum_i^{h_1} \sum_k^{h_{2ij}} \theta_{3ijk} & \sum_i^{h_1} \sum_k^{h_{2ij}} \theta_{2ij} \theta_{3ijk} & \sum_i^{h_1} \sum_k^{h_{2ij}} \theta_{3ijk}^2 & \sum_i^{h_1} \sum_k^{h_{2ij}} \theta_{3ijk} \\
 \sum_i^{h_1} \sum_j^{h_{2i}} \theta_{1i} & \sum_i^{h_1} \sum_j^{h_{2i}} \theta_{2ij} & \sum_j^{h_{2i}} \sum_k^{h_{2ij}} \theta_{3ijk} & \sum_j^{h_{2i}} \sum_k^{h_{2ij}}
 \end{bmatrix}$$

é definida positiva, e pelo Teorema A.1.6(ver Silva et al. [17]), $\Theta^T \Theta$ é uma matriz não singular, e a sua inversa é denotada por $(\Theta^T \Theta)^{-1}$.

Multiplicando o sistema (4.24) em ambos os lados por Θ^T resulta o sistema de equações,

$$\Theta^T S = \Theta^T \Theta \gamma, \tag{4.25}$$

cuja, a solução única é (um estimador para γ)

$$\hat{\gamma} = (\Theta^\top \Theta)^{-1} \Theta^\top S. \quad (4.26)$$

$\hat{\gamma} = \begin{bmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \\ \hat{\gamma}_3 \\ \hat{\gamma}_4 \end{bmatrix}$ denominado o *estimador Sub-D* e é referido como o *método Sub-D*.

Proposição 4.2.4. $\hat{\gamma}$ é um estimador imparcial de γ , com $\gamma = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \end{bmatrix}$.

Demonstração. (ver Silva et al. [17], Proposição 4.2.5) □

Proposição 4.2.5. Seja $i \leq i^*$; $j \leq j^*$; $k \leq k^*$ (aplica-se a simetria).

$$\text{cov}(S_{ijk}^2, S_{i^*j^*k^*}^2) = \begin{cases} (a) & i = i^*; j \neq j^*; k \neq k^* : & 0, \\ (b) & i = i^*; j = j^*; k = k^* : & 2 \frac{\lambda_{ijk}^2}{g_{ijk}}, \\ (c) & i \neq i^* : & 2\gamma_3^2 \text{tr}(\Omega N_3), \end{cases}$$

onde $\Omega = \nabla_{ijk} N_3 \nabla_{i^*j^*k^*}$, com $\nabla_{ijk} = \frac{A_{1i}^\top A_{2ij}^\top A_{3ijk}^\top A_{3ijk} A_{2ij} A_{1i}}{g_{ijk}}$.

Demonstração. Tem - se que,

$$\begin{aligned} \text{cov}(S_{ijk}^2, S_{i^*j^*k^*}^2) &= \text{cov}\left(\frac{Z_{ijk}^\top Z_{ijk}}{g_{ijk}}, \frac{Z_{i^*j^*k^*}^\top Z_{i^*j^*k^*}}{g_{i^*j^*k^*}}\right) \\ &= \text{cov}\left(Z^\top \left(\frac{A_{1i}^\top A_{2ij}^\top A_{3ijk}^\top A_{3ijk} A_{2ij} A_{1i}}{g_{ijk}}\right) Z, Z^\top \left(\frac{A_{1i^*}^\top A_{2i^*j^*}^\top A_{3i^*j^*k^*}^\top A_{3i^*j^*k^*} A_{2i^*j^*} A_{1i^*}}{g_{i^*j^*k^*}}\right) Z\right) \\ &= \text{cov}(Z^\top \nabla_{ijk} Z, Z^\top \nabla_{i^*j^*k^*} Z) = 2 \text{tr}(\nabla_{ijk} V \nabla_{i^*j^*k^*} V) \\ &= 2\gamma_1^2 \text{tr}(\nabla_{ijk} N_1 \nabla_{i^*j^*k^*} N_1) + 2\gamma_1 \gamma_2 \text{tr}(\nabla_{ijk} N_1 \nabla_{i^*j^*k^*} N_2) + 2\gamma_1 \gamma_3 \text{tr}(\nabla_{ijk} N_1 \nabla_{i^*j^*k^*} N_3) \\ &+ 2\gamma_1 \gamma_4 \text{tr}(\nabla_{ijk} N_1 \nabla_{i^*j^*k^*}) + 2\gamma_2 \gamma_4 \text{tr}(\nabla_{ijk} N_2 \nabla_{i^*j^*k^*} N_1) + 2\gamma_2^2 \text{tr}(\nabla_{ijk} N_2 \nabla_{i^*j^*k^*} N_2) \\ &+ 2\gamma_2 \gamma_3 \text{tr}(\nabla_{ijk} N_2 \nabla_{i^*j^*k^*} N_3) + 2\gamma_2 \gamma_4 \text{tr}(\nabla_{ijk} N_2 \nabla_{i^*j^*k^*}) + 2\gamma_3 \gamma_1 \text{tr}(\nabla_{ijk} N_3 \nabla_{i^*j^*k^*} N_1) \\ &+ 2\gamma_3 \gamma_2 \text{tr}(\nabla_{ijk} N_3 \nabla_{i^*j^*k^*} N_2) + 2\gamma_3^2 \text{tr}(\nabla_{ijk} N_3 \nabla_{i^*j^*k^*} N_3) + 2\gamma_3 \gamma_4 \text{tr}(\nabla_{ijk} N_3 \nabla_{i^*j^*k^*}) \\ &+ 2\gamma_4 \gamma_1 \text{tr}(\nabla_{ijk} \nabla_{i^*j^*k^*} N_1) + 2\gamma_4 \gamma_2 \text{tr}(\nabla_{ijk} \nabla_{i^*j^*k^*} N_2) + 2\gamma_4 \gamma_3 \text{tr}(\nabla_{ijk} \nabla_{i^*j^*k^*} N_3) + 2\gamma_4^2 \text{tr}(\nabla_{ijk} \nabla_{i^*j^*k^*}) \\ &= \begin{cases} i = i^*, j \neq j^*, k \neq k^* : & 0, \\ i = i^*, j = j^*, k = k^* : & 2 \frac{\lambda_{ijk}^2}{g_{ijk}}, \\ i \neq i^* : & 2\gamma_3^2 \text{tr}(\nabla_{ijk} M_3 \nabla_{i^*j^*k^*} M_3). \end{cases} \end{aligned}$$

Para o primeiro caso, temos: $i = i^*$; $j \neq j^*$; $k \neq k^*$, assim

$$\begin{aligned} \nabla_{ijk} N_1 \nabla_{ij^*k^*} &= \frac{1}{g_{ijk} g_{ij^*k^*}} A_{1i}^\top A_{2ij}^\top A_{3ijk}^\top A_{3ijk} A_{2ij} A_{1i} N_1 A_{1i}^\top A_{2ij^*}^\top A_{3ij^*k^*}^\top A_{3ij^*k^*} A_{2ij^*} A_{1i} \\ &= \frac{1}{g_{ijk} g_{ij^*k^*}} A_{1i}^\top A_{2ij}^\top A_{3ijk}^\top A_{3ijk} A_{2ij} (\theta_{1i} I_{g_i}) A_{2ij^*}^\top A_{3ij^*k^*}^\top A_{3ij^*k^*} A_{2ij^*} A_{1i} \\ &= \mathbf{0}_{g_i \times g_i}; \end{aligned} \quad (4.27)$$

$$\begin{aligned} \nabla_{ijk} N_2 \nabla_{ij^*k^*} &= \frac{1}{g_{ijk} g_{ij^*k^*}} A_{1i}^\top A_{2ij}^\top A_{3ijk}^\top A_{3ijk} A_{2ij} A_{1i} N_2 A_{1i}^\top A_{2ij^*}^\top A_{3ij^*k^*}^\top A_{3ij^*k^*} A_{2ij^*} A_{1i} \\ &= \frac{1}{g_{ijk} g_{ij^*k^*}} A_{1i}^\top A_{2ij}^\top A_{3ijk}^\top A_{3ijk} A_{2ij} (N_{ii}^2) A_{2ij^*}^\top A_{3ij^*k^*}^\top A_{3ij^*k^*} A_{2ij^*} A_{1i} \\ &= \mathbf{0}_{g_i \times g_i}; \end{aligned} \quad (4.28)$$

$$\begin{aligned} \nabla_{ijk} N_3 \nabla_{ij^*k^*} &= \frac{1}{g_{ijk} g_{ij^*k^*}} A_{1i}^\top A_{2ij}^\top A_{3ijk}^\top A_{3ijk} A_{2ij} A_{1i} N_3 A_{1i}^\top A_{2ij^*}^\top A_{3ij^*k^*}^\top A_{3ij^*k^*} A_{2ij^*} A_{1i} \\ &= \frac{1}{g_{ijk} g_{ij^*k^*}} A_{1i}^\top A_{2ij}^\top A_{3ijk}^\top A_{3ijk} A_{2ij} (N_{iii}^3) A_{2ij^*}^\top A_{3ij^*k^*}^\top A_{3ij^*k^*} A_{2ij^*} A_{1i} \\ &= \mathbf{0}_{g_i \times g_i}. \end{aligned} \quad (4.29)$$

$$\begin{aligned} \nabla_{ijk} \nabla_{ij^*k^*} &= \frac{1}{g_{ijk} g_{ij^*k^*}} A_{1i}^\top A_{2ij}^\top A_{3ijk}^\top (\mathbf{0}_{g_{ij} \times g_{ij^*}}) A_{3ij^*k^*} A_{2ij^*} A_{1i} \\ &= \mathbf{0}_{g_i \times g_i}. \end{aligned} \quad (4.30)$$

Por conseguinte, (4.27), (4.28), (4.29) e (4.30) com a proposição 2.1.1 (c) (ver Silva et al. [17]) demonstra o primeiro caso, isto é, $i = i^*$; $j \neq j^*$; $k \neq k^*$.

Para o segundo caso, $i = i^*$; $j = j^*$ e $k = k^*$, com $y_{ijk} \sim \mathcal{N}_n(\mathbf{0}_{g_{ijk}}, \lambda_{ijk} I_{g_{ijk}})$, resulta

$$\begin{aligned} \text{cov}(S_{ijk}^2) &= \Sigma\left(\frac{y_{ijk}^\top y_{ijk}}{g_{ijk}}, \frac{y_{ijk}^\top y_{ijk}}{g_{ijk}}\right) = 2 \text{tr}\left\{\frac{\lambda_{ijk}}{g_{ijk}} I_{g_{ijk}} \frac{\lambda_{ijk}}{g_{ijk}} I_{g_{ijk}}\right\} \\ &= 2 \frac{\lambda_{ijk}^2}{g_{ijk}^2} \text{tr}\{I_{g_{ijk}}\} = 2 \frac{\lambda_{ijk}^2}{g_{ijk}}. \end{aligned} \quad (4.31)$$

Finalmente, para último caso, isto é, $i \neq i^*$, resultado desejado se torna claro usar o Teorema 1.3.(d) de Schott [15] e nota que

$$A_{1i} M_1 A_{1i^*} = A_{1i} A_{1i^*} = \mathbf{0}_{g_i \times g_{i^*}}$$

□

4.3 Resultados Numéricos

O autor dos métodos *Sub-D* e *Sub-DI* (versão melhorada de *Sub-D*) (ver Silva et al. [17]), utilizou o *Software R* para todas as simulações que foram feitas e testou o desempenho dos métodos para pequenas amostras.

Os testes foram feitos e aplicados para três componentes da variância nos modelos seguintes: *Balanced “One Way Design”*, *Unbalanced “One Way Design”*, *“Two Way Crossed Design”* e *“Two Way Nested Design”*.

Para comprovar a validade, a eficiência e o desempenho dos métodos *Sub-D* e *Sub-DI*, simularam para cada $\gamma_i \in \{0.1, 0.25, 0.5, 0.75, 1.0, 2.0, 5.0\}$ com $i = 1, 2, 3$, 10000 observações para os modelos *Balanced “One Way Design”*, *Unbalanced “One Way Design”* e *“Two Way Nested Design”*. Para o modelo *Unbalanced “Two Way Crossed Design”* simulou para cada $\gamma_i \in \{0.1, 0.25, 0.5, 0.75, 1.0, 2.0, 5.0, 10\}$, 1000 observações, comparando com os estimadores de REML e ANOVA, apresentando os resultados em tabelas com quatro casas decimais.

Para o modelo *Balanced “One Way Design”* conclui-se que os estimadores *Sub-D* e ANOVA, são os preferidos, porque o estimador REML forneceu estimativas baixas para valores de componentes da variância de $\gamma_1 \leq 0.5$. No modelo *Unbalanced “One Way Design”*, *Sub-D* fornece estimativas imparciais, embora com maior dispersão. Para γ_2 REML fornece estimativas precisas, embora não tão precisas quanto às fornecidas por *Sub-D*. Mas, para $\gamma_1 \leq 0.75$ as estimativas não são tão precisas para os valores de $\gamma_1 \geq 0.75$. Embora não haja precisão, ANOVA fornece estimativas aceitáveis para γ_1 , mas para γ_2 produz estimativas não reais.

No modelo *“Two Way Crossed Design”*, os métodos *Sub-D*, *Sub-DI*, REML e ANOVA são aplicados e γ_1 , γ_2 e o erro (γ_3) foram estimados. O estimador REML fornece estimativas precisas para γ_1 , γ_2 e $\gamma_3 \in \{1.0, 2.0, 5.0\}$ e com baixa precisão para γ_1 , γ_2 e $\gamma_3 \in \{0.1, 0.25, 0.5\}$. O estimador ANOVA fornece estimativas aceitáveis para γ_1 , mas para γ_2 e γ_3 as estimativas fornecidas são extremamente não reais, enquanto que, os estimadores *Sub-D* e *Sub-DI* fornecem estimativas precisas para todos os parâmetros.

No modelo *“Two Way Nested Design”*, as únicas estimativas precisas são as fornecidas por *Sub-D* e *Sub-DI*, enquanto que os métodos REML e ANOVA forneceram estimativas com baixa precisão, sendo que ANOVA produz estimativas não reais. É claro que o *Sub-DI* produz estimativas com erros quadráticos médios menores do que *Sub-D* para todos os parâmetros.

Conclui-se que, o *Sub-D* e *Sub-DI* mantiveram um desempenho constante e preciso em relação a todos os modelos abordados.

Neste trabalho o método *Sub-D* será testado e aplicado para quatro componentes da variância e em todos modelos que foram aplicados para três componentes da variância. Para cada $\gamma_i \in \{0.5, 0.75, 1.0, 2.0, 5.0\}$, $i = 1, 2, 3$, $\gamma_4 = 1.0$ fixo, os modelos foram observados 10000 vezes. Os resultados serão apresentados em tabelas com cinco casas decimais, comparando os valores reais (VR) e valores estimados (VE).

	VR	VE	VR	VE	VR	VE	VR	VE	VR	VE
γ_1	0.5	0.49286	0.75	0.69873	1.0	1.00831	2.0	2.01046	5.0	5.09781
γ_2	0.75	0.77917	1.0	1.01745	0.5	0.48237	5.0	5.07745	2.0	2.03174
γ_3	2.0	2.03379	5.0	5.05215	0.75	0.75576	1.0	1.05495	0.5	0.50468
γ_4	1.0	0.98761	1.0	0.95146	1.0	1.01320	1.0	0.91023	1.0	0.98092

Tab. 4.1: Valores reais e estimados, usando o *Método Sub-D*.

	VR	VE	VR	VE	VR	VE	VR	VE	VR	VE
γ_1	0.5	2.49379	0.75	4.70352	1.0	2.13731	2.0	6.23395	5.0	7.45713
γ_2	0.75	2.44144	1.0	4.18521	0.5	1.71370	5.0	7.35756	2.0	5.12966
γ_3	2.0	3.47275	5.0	7.31941	0.75	1.83923	1.0	3.34504	0.5	2.62364
γ_4	1.0	2.27873	1.0	4.15645	1.0	1.65292	1.0	4.87531	1.0	4.26792

Tab. 4.2: Variância dos valores reais e estimados, usando o *Método Sub-D*.

Da tabela 4.1 podemos ver que, o *Método Sub-D* fornece estimativas aceitáveis e um bom desempenho. Se os valores reais fossem $\gamma_1 = 0.5$, $\gamma_2 = 0.75$, $\gamma_3 = 2.0$ e $\gamma_4 = 1.0$, podemos reparar que só o valor estimado por γ_2 , afasta muito do valor real, com maior dispersão. Quando os dados reais forem $\gamma_1 = 0.75$, $\gamma_2 = \gamma_4 = 1.0$ e $\gamma_3 = 5.0$, o valor estimado por γ_1 apresenta maior dispersão. Para valores reais com $\gamma_1 = \gamma_4 = 1.0$, $\gamma_2 = 0.5$ e $\gamma_3 = 0.75$, os valores estimados são aceitáveis e próximos dos valores reais. E se os valores reais fossem $\gamma_1 = 2.0$, $\gamma_2 = 5.0$ e $\gamma_3 = \gamma_4 = 1.0$, só os valores estimados por γ_4 , apresenta uma dispersão muito grande. E para valores reais $\gamma_1 = 5.0$, $\gamma_2 = 2.0$, $\gamma_3 = 0.5$ e $\gamma_4 = 1.0$, os resultados estimados são próximos de reais e aceitáveis.

Apartir da tabela 4.2, observa-se que os valores das variâncias obtidas pelo *Método Sub-D*, fornece estimativas confiáveis e um desempenho razoável.

Se $\gamma_1 = 0.5$, $\gamma_2 = 0.75$, $\gamma_3 = 2.0$ e $\gamma_4 = 1.0$, os valores estimados das variâncias são respectivamente, 2.49379, 2.44144, 3.47275, 2.27873, onde estão distribuídas em torno da média global para todas as observações, que é 2.67167. Repara-se que só o valor apresentado por 3.47275 dispersa um pouco da média global.

Para $\gamma_1 = 0.75$, $\gamma_2 = \gamma_4 = 1.0$ e $\gamma_3 = 5.0$, os valores estimados das variâncias são, 4.70352, 4.18521, 7.31941 e 4.15645, respectivamente, distribuídas em torno da média global de 5.09114. Destas, só o valor apresentado por 7.31941 afasta muito da média global.

Quando $\gamma_1 = \gamma_4 = 1.0$, $\gamma_2 = 0.5$ e $\gamma_3 = 0.75$, os valores estimados das variâncias são, 2.13731, 1.71370, 1.83923 e 1.65292, na devida ordem, onde estão distribuídas em torno da média global de 1.83579. Repara-se que, só o valor apresentado por 2.13731 dispersa mais da média global.

Para $\gamma_1 = 2.0$, $\gamma_2 = 5.0$ e $\gamma_3 = \gamma_4 = 1.0$ os valores estimados das variâncias, respectivamente são, 6.23395, 7.35756, 3.34504, 4.87521, onde estão distribuídas em torno da média

global de 5.45297. Repara-se que, os valores 7.35756 e 3.34504 estão mais dispersas da média global.

E se $\gamma_1 = 5.0$, $\gamma_2 = 2.0$, $\gamma_3 = 0.5$ e $\gamma_4 = 1.0$, os valores estimados das variâncias são, 7.45713, 5.12966, 2.62364, 4.26792, respectivamente, onde estão distribuídas em torno da média global de 4.86958. Repara-se que, os valores das variâncias apresentados por, 7.45713 e 2.62364, estão mais dispersas da média global.

Este capítulo encontra-se publicado em revista internacional(ver Silva et al. [18]).

CONCLUSÃO

Para comprovar a validade, a eficiência e o tempo de execução do Método Sub-D em MLM com três componentes da variância o autor (ver Silva et al. [17], capítulo 6), conclui que:

- O *Sub-D* fornece estimativas a qualquer tipo de dados e a qualquer modelos aplicados, mesmo tendo células vazias, o que não acontece com os estimadores ANOVA e REML;
- O estimador *Sub-DI* produz estimativas não viesadas, mas com menos dispersão do que o *Sub-D*, ambas fornecem estimativas ligeiramente mais precisas (devido à sua imparcialidade) do que o estimador REML em todos os modelos abordados, mas comparáveis quando os modelos são equilibrados, tendo em alguns casos um pouco mais disperso quando os modelos são desequilibrados;
- O estimador *Sub-D* e *Sub-DI* em modelos *Balanced “One Way Design”* fornecem estimativas precisas, enquanto que estimador REML fornece estimativas precisas baixas e com pouco desempenho e o estimador ANOVA não fornece estimativas reais em nenhum dos modelos *Unbalanced “Two Way Crossed Design”* e *“Two Way Nested Design”*, apenas parece fornecer estimativas realistas em modelos *Balanced “One Way Design”*, desde que ANOVA usa técnicas de efeito fixo;
- O *Sub-D* e *Sub-DI* mantém um desempenho constante, fornecendo estimativas sempre precisas para todos os modelos que foram aplicados, enquanto que REML não mostra um desempenho constante em particular os desequilibrados (“modelos com dois fatores aninhados”). Para o estimador ANOVA, o cenário é ainda pior, uma vez que fornece estimativas não centralizadas tanto em modelos *Unbalanced “Two Way Crossed Design”* e *“Two Way Nested Design”*;
- Para $\gamma_i \in \{0.25, 0.5, 0.75, 1.0, 2.0, 5.0\}$, $i = 1, 2, 3$, com 10000 observações, os tempos de execução dos estimadores ANOVA, *Sub-D* e *Sub-DI* são 1.2471, 2.06 e 4.3338 segundos respectivamente, enquanto que para o estimador REML é cerca de 6.2618

minutos, isso significa que, o código para ANOVA e *Sub-D* é mais rápido que o de REML 187 vezes.

Assim, neste trabalho, pensamos ter cumprido, os objetivos definidos e, por outro lado, ter confirmado o método *Sub-D*, anteriormente desenvolvido e provado para modelos com dois componentes da variância, modelos com três componentes de variância e generalizado deduzindo-o para modelos com um número arbitrário de componentes da variância(ver Silva et al. [17]).

De acordo com os resultados numéricos(ver tab.4.1 e 4.2), comparando os valores reais com os valores estimados, para o caso particular de modelos mistos lineares com 4 componentes da variância, conclui-se mais uma vez a eficiência e a validade do método *Sub-D*.

Anexo

Bibliografia

- [1] Albert, A. (1946). On jordan algebras of linear transformations. *Transactions of the American Mathematical Society.*, 59(3):524–555.
- [2] Burden, R. L. and Faires, J. D. (2011). *Numerical Analysis.*
- [3] Hocking, R. (1985). *The analysis of linear models.* Brooks/Cole Pub Co.
- [4] Horn, R. and Johnson, C. (1985). *Matrix analysis. Corrected reprint of the 1985 original.* Cambridge University Press, Cambridge.
- [5] Lay, D. (1994). *Linear Algebra and Its Applications.* Addison-Wesley.
- [6] Magnus, J. R. and Neudecker, H. (1995). *Matrix Differential Calculus With Applications in Statistics and Econometrics.* John Wiley & Sons.
- [7] McCulloch, C. E. and Searle, S. R. (2001). *Generalized, Linear, and Mixed Models.* New York: Wiley.
- [8] Monteiro, A. (2001). *Álgebra Linear e Geometria Analítica.*
- [9] Moser, B. (1996). *Linear Models: A Mean Model Approach.* Elsevier.
- [10] Muller, K. and Stewart, P. (2006). *Linear Model Theory : Univariate, Multivariate, and Mixed Models.* John Wiley & Sons.
- [11] Myers, R., Vining, G., Moutgomery, D., and Robinson, T. (2010). *Generalized Linear Models: With Applications in Engineering and the Sciences, 2nd Edition.* New York: Wiley.
- [12] Rao, C. R. and Rao, M. B. (1998). *Matrix Algebra And Its Applications To Statistics And Econometrics.* World Scientific.
- [13] Rencher, A. C. and Schaalje, G. B. (2008). *Linear Models in Statisitcs.* John Wiley & Sons.
- [14] Sahai, H. and Ojeda, M. M. (2004). *Analysis of Variance for Random Models, Volume 2: Unbalanced Data,* volume 2. Springer Science & Business Media.

- [15] Schott, J. R. (1997). *Matrix Analysis for Statistics*. John Wiley & Sons.
- [16] Searle, S., Casella, G., and McCulloch, C. (1992). *Variance Components*. John Wiley & Sons.
- [17] Silva, A. (2016). *Variance Components Estimation in Mixed Linear Models*. PhD thesis, Universidade Nova de Lisboa.
- [18] Silva, A., Fonseca, M., and Monteiro, A. (2017). Inference in mixed models with four variance components - sub-d and sub-di. *International Journal of Data Analysis Techniques and Strategies*.
- [19] Stapleton, J. (2009). *Linear Statistical Models, 2nd Edition*. New York: Wiley.