



Universidade Estadual da Paraíba  
Centro de Ciências e Tecnologia  
Departamento de Estatística

**Erivan Barbosa da Silva**

**Experimentos em blocos completos ao acaso  
com medidas repetidas: Aspectos teóricos  
com aplicação**

Campina Grande  
Abril, 2016

Erivan Barbosa da Silva

# **Experimentos em blocos completos ao acaso com medidas repetidas: Aspectos teóricos com aplicação**

Trabalho de Conclusão de Curso apresentado ao curso de Especialização em Estatística Aplicada do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de Especialista em Estatística Aplicada.

Orientador:  
João Gil de Luna

Campina Grande  
Abril, 2016

É expressamente proibida a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano da dissertação.

S586e Silva, Erivan Barbosa da  
Experimentos em blocos completos ao acaso com medidas repetidas: aspectos teóricos com aplicação [manuscrito] / Erivan Barbosa da Silva. - 2016.  
86 p. : il. color.

Digitado.

Monografia (Estatística Aplicada) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2016.

"Orientação: Prof. Drº João Gil de Luna, Departamento de Estatística".

1. Análise de variância. 2. Delineamento em Blocos. 3. Estatística experimental. 4. Análise de Perfil. I. Título.

21. ed. CDD 519.538

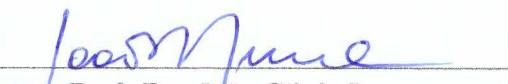
Erivan Barbosa da Silva


## Experimentos em blocos completos ao acaso com medidas repetidas: Aspectos teóricos com aplicação


Trabalho de Conclusão de Curso apresentado ao curso de Especialização em Estatística Aplicada do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de Especialista em Estatística Aplicada.

Aprovado em: 22/04/2016

### Banca Examinadora:

  
Prof. Dr. João Gil de Luna  
Orientador

  
Prof. Dr. Tiago Almeida de Oliveira  
Universidade Estadual da Paraíba

  
Prof. Dra. Ana Patrícia Bastos Peixoto  
Universidade Estadual da Paraíba

# Dedicatória

A Deus,

porque Dele, por Ele e para Ele são todas as coisas.

# Agradecimentos

A Deus por ter me concedido força e dedicação para mais uma vitória em minha vida.

A minha esposa Rosiane por ter me incentivado e apoiado sempre, dando-me força, coragem e companhia, fazendo com que eu sempre acreditasse nos meus sonhos e lutasse por meus objetivos, a quem sou muito grato pelo amor e carinho.

Ao meus pais Marluce e Nerivaldo, meus irmãos Nerivaldo Filho, Erivaldo e Erivania e minha sobrinha Marília Gabriela, minhas bases, na qual me fundamentou e fundamenta, com amor, dedicação e carinho, agradeço.

Ao meus avós paternos e maternos, em especial aos meus avós Irene e Zé Vaqueiro (In Memoriam), a quem estimo-os e sou grato pela educação e força que me foi concedido para me tornar a pessoa que sou hoje.

Aos meus tios e tias, primos e primas, amigos e amigas, que contribuíram direto e indiretamente e torceram para mais essa realização na minha vida.

A turma da especialização, Isaac, Luzidark, Aline, Raquel, Tathiana, Rafaela, Alana, Conceição e Adrícia que foram como irmãos na batalha em que o foco foi a vitória coletiva.

Aos professores da Especialização, João Gil, Ana Patrícia, Tiago, Ricardo, Gustavo, Diana, Edwird, Silvio, Jomar, Kleber e Elias, a quem agradeço de coração pelos conhecimentos adquiridos para a vida profissional e pessoal e pela amizade de cada um de vocês.

Ao meu orientador João Gil de Luna pela sua dedicação, paciência, respeito e amizade que sempre teve comigo, sou grato por todo conhecimento repassado e por sua disponibilidade.

Aos membros da banca examinadora Tiago Almeida de Oliveira e Ana Patrícia Bastos Peixoto, por aceitarem o convite e contribuírem para o enriquecimento desse trabalho.

# Resumo

A cultura do girassol é de grande importância no cenário mundial por ser uma planta de múltiplas finalidades. No Brasil, o interesse do produtor por esta cultura vem crescendo a cada ano, pois além do destacado teor de óleo nos aquênios, a planta pode ser utilizada para alimentação humana, na alimentação animal na forma de silagem e especialmente de aves. O objetivo principal deste trabalho é apresentar a metodologia estatística utilizada na análise de um conjunto de dados obtido de um experimento em blocos ao acaso com medidas repetidas em que foi analisado o diâmetro médio do caule do girassol em cinco épocas diferentes de avaliação, o experimento foi conduzido no Campus da Universidade Federal de Campina Grande-PB no ano de 2010 e 2011. Em seguida foi feita uma análise Multivariada de Perfil para o experimento de 2010, e para o experimento de 2011 foi feita uma análise univariada Split-Plot. Verificou-se que não houve efeito significativo da interação aumento do diâmetro médio do caule e o tratamento T4 (Testemunha), no entanto houve um aumento significativo do diâmetro médio do caule para o tratamento T2 (Leirão) em comparação com os demais tratamentos nas cinco épocas de avaliação.

**Palavras chaves:** Blocos ao acaso, Análise de variância, Estatística experimental, Análise de Perfil.

# Abstract

The sunflower crop is of great importance on the world stage to be a multi-purpose plant. In Brazil, the interest of the producer of this crop is growing every year, because besides the outstanding oil content in the achenes, the plant can be used for human consumption in animal feed in the form of silage and especially birds. The main objective of this paper is to present the statistical methodology used in the analysis of a set of data obtained from an experiment in a randomized block design with repeated measures which analyzed the average diameter of the sunflower stem in five different evaluation times, the experiment was conducted at the campus of the Federal University of Campina Grande-PB in 2010 and 2011. then it made a profile Multivariate analysis for the 2010 experiment, and for the 2011 experiment was performed a univariate analysis Split-Plot. It was found that there was no significant interaction effect increased stem diameter and the T4 treatment (control), however there was a significant increase in stem diameter for T2 (Dormouse) treatment compared to the other treatments in the five evaluation periods.

**Key words:** Randomized block, Analysis of variance , experimental Statistics Profile Analysis.



# Sumário

## Lista de Figuras

## Lista de Tabelas

<b>1</b>	<b>Introdução</b>	p. 13
<b>2</b>	<b>Fundamentação Teórica</b>	p. 15
2.1	Características de um experimento em blocos completos ao acaso . . .	p. 15
2.2	A organização das observações experimentais e o modelo estatístico associado . . . . .	p. 16
2.3	O modelo estatístico e suposições associadas às observações . . . . .	p. 17
2.4	Estimação dos parâmetros, das observações e do erro experimental . . .	p. 18
2.5	A Decomposição da Variabilidade Total . . . . .	p. 21
2.6	Distribuição de Probabilidade dos Estimadores e das Somas de Quadrados	p. 23
2.7	Valores esperados dos quadrados médios . . . . .	p. 25
2.8	As estatísticas dos testes para o modelo de blocos completos ao acaso .	p. 26
2.8.1	Desdobramento dos graus de liberdade e somas de quadrados em contrastes de interesse . . . . .	p. 27
2.8.2	Desdobramento dos graus de liberdade e somas de quadrados em componentes polinomiais . . . . .	p. 30
2.8.3	Comparações múltiplas das médias pelo método de Tukey . . .	p. 32
2.9	Experimentos em blocos ao acaso com medidas repetidas . . . . .	p. 33
2.10	O teste de esfericidade . . . . .	p. 37
2.11	Teste de hipótese sobre a uniformidade de $\Sigma$ . . . . .	p. 39

2.12	A análise de variância univariada ( <i>Split-Plot</i> ) . . . . .	p. 41
2.12.1	A decomposição da variabilidade total . . . . .	p. 43
2.12.2	Distribuições de probabilidade dos estimadores, dos parâmetros e das somas de quadrados . . . . .	p. 44
2.13	Valores esperados dos quadrados médios para o modelo <i>split – plot</i> . .	p. 45
2.13.1	As estatísticas dos testes no modelo <i>Split-plot</i> . . . . .	p. 46
2.14	O Modelo multivariado . . . . .	p. 51
2.14.1	As estatísticas dos testes para o modelo MANOVA . . . . .	p. 55
2.14.2	Contrastes multivariados . . . . .	p. 56
2.14.3	Intervalos de confiança simultâneos de Bonferroni . . . . .	p. 57
2.15	Análise de perfil . . . . .	p. 58
<b>3</b>	<b>Material e métodos</b>	p. 62
<b>4</b>	<b>Resultados e discussão</b>	p. 64
4.1	Teste de uniformidade de $\Sigma$ . . . . .	p. 65
4.2	A análise de variância multivariada MANOVA . . . . .	p. 68
4.3	Intervalos de confiança simultâneos . . . . .	p. 70
4.4	A análise de perfil de médias . . . . .	p. 72
4.4.1	A hipótese de paralelismo . . . . .	p. 72
4.4.2	A hipótese de coincidência . . . . .	p. 73
4.4.3	A hipótese de horizontalidade . . . . .	p. 74
4.5	A análise <i>Split-plot</i> com correção dos graus de liberdade . . . . .	p. 74
4.6	Análise do experimento em blocos ao acaso com parcelas subdivididas .	p. 74
<b>5</b>	<b>Considerações finais</b>	p. 75
	<b>Referências</b>	p. 76

Apêndice A - Coeficientes dos polinômios ortogonais

p. 78

Apêndice B - Programas SAS

p. 79

# Lista de Figuras

- 1 Um possível desenho de um experimento em blocos completos ao acaso com 6 tratamentos e 4 repetições . . . . . p. 16
- 2 Gráfico dos perfis de médias dos tratamentos em função das *cinco* épocas de avaliação . . . . . p. 73

# Lista de Tabelas

1	Organização das observações obtidas de um experimento em blocos completos ao acaso . . . . .	p. 17
2	Estimadores das características no modelo estatístico . . . . .	p. 20
3	Distribuições de probabilidade dos estimadores dos parâmetros e das somas de quadrados associadas, sob a hipótese nula . . . . .	p. 25
4	Valores esperados dos quadrados médios . . . . .	p. 26
5	Análise da variância para testar a significância dos efeitos de blocos e tratamentos sobre a variável resposta . . . . .	p. 27
6	Análise de variância para testar a significância dos contrastes de interesse, os efeitos dos tratamentos e dos blocos . . . . .	p. 29
7	Análise de variância para testar a nulidade dos coeficientes do polinômio, dos efeitos dos tratamentos e dos blocos sobre a variável resposta . . . . .	p. 31
8	Organização das observações de um experimento em blocos ao acaso com medidas repetidas . . . . .	p. 34
9	Organização das observações de um experimento em blocos ao acaso com medidas repetidas . . . . .	p. 41
10	Os parâmetros do modelo e seus estimadores . . . . .	p. 42
11	Distribuição de probabilidade dos estimadores dos parâmetros e das somas de quadrados sob $H_0^{(.)}$ . . . . .	p. 45
12	Análise da variância com os valores esperados dos quadrados médios . . . . .	p. 46
13	ANOVA univariada para o modelo <i>Split-plot</i> baseada nos dados da Tabela 8 . . . . .	p. 47
14	ANOVA univariada para o modelo <i>Split-plot</i> para testar os componentes polinomiais para o fator $C$ . . . . .	p. 50

15	ANOVA para o desdobramento dos graus de liberdade e somas de quadrados em componentes polinomiais para o fator medidas repetidas dentro de cada nível de $A$ . . . . .	p. 51
16	Diâmetro do caule de girassol avaliado em diferentes épocas após a emergência num experimento em blocos completos ao acaso, Campina Grande, em 2010 e 2011 . . . . .	p. 63
17	Análise de variância para testar a hipótese de igualdade das médias dos tratamentos e contrastes de interesse do ano de 2010 . . . . .	p. 65
18	Resíduos obtidos em cada época de avaliação para os anos de 2010 e 2011	p. 66
19	Autovalores das matrizes $E^{-1}H_A$ , $E^{-1}H_B$ , $E^{-1}H_1$ , $E^{-1}H_2$ e $E^{-1}H_3$	p. 70
20	Análise de variância multivariada (MANOVA) para testar a nulidade dos efeitos dos tratamentos, dos blocos e dos contrastes de interesse sobre os tratamentos . . . . .	p. 70
21	Estimativas das médias do diâmetro do caule em relação aos tratamentos em cada época de avaliação . . . . .	p. 71
22	Intervalos com 95% de confiança simultâneos de Bonferroni para os contrastes de médias dos tratamento em cada época de avaliação (variável)	p. 72
23	Coeficientes dos polinômios ortogonais para um fator com 3 a 9 níveis.	p. 78

# 1 Introdução

Dentre os desenhos experimentais, o de blocos completos ao acaso é o mais empregado nas pesquisas científicas. Este delineamento se caracteriza por levar em consideração os três princípios básicos da experimentação, quais sejam: *repetição*, *casualização* e o *controle local*. O terceiro princípio viabiliza, por exemplo, a instalação de um experimento agrícola num local com diferentes tipos de solo ou diferentes gradientes de fertilidade, sendo que cada bloco constitui um ambiente homogêneo no qual é alocado, de modo aleatório, um (ou mais de um) conjunto completo de tratamentos. Com o objetivo de avaliar o comportamento de algumas variáveis resposta em relação ao tempo, em muitas situações, os pesquisadores costumam fazer medições nas unidades experimentais, em momentos pontuais, caracterizando desse modo, um desenho de blocos ao acaso com medidas repetidas no tempo. Esta prática, leva a que as análises estatísticas subsequentes sejam mais complexas, tendo em vista que demandam a verificação de condições especiais para uma análise estatística univariada a partir de um modelo de parcelas sub-divididas (*split-plot*) ou a utilização de técnicas estatísticas multivariadas.

A utilização do desenho experimental em blocos completos ao acaso com medidas repetidas no planejamento das pesquisas de muitos estudiosos é muito comum e as técnicas estatísticas apropriadas às análises dos dados oriundos desse tipo de experimento, são bastante difundidas no meio científico, embora ainda sejam objetos de estudos na ciência estatística. Neste sentido, podem ser citadas como referências os textos de Rencher e Christensen (2012), Johnson e Wichern (2007), Santos e Luna (2014), dentre muitas outras. É importante frisar que os textos disponíveis em estatística experimental se preocupam apenas em apresentar as condições de aplicação das técnicas, os procedimentos e as fórmulas das estatísticas dos testes das hipóteses de interesse, bem como, orientações para implementação das análises através de um *software* estatístico. No entanto, a teoria estatística que permeia as técnicas e dá origem às fórmulas não é encontrada facilmente na literatura experimental e isso dificulta o avanço da estatística enquanto ciência. Como se sabe, a compreensão da teoria estatística por aqueles que adotam a estatística como

profissão tem importância fundamental, pois, é esperado desses profissionais a criação de novos métodos que, sem dúvida, serão facilitados com a produção de textos apropriados a uma boa formação teórica.

O objetivo principal deste trabalho é apresentar a metodologia estatística utilizada na análise de um conjunto de dados obtido de um experimento em blocos ao acaso com medidas repetidas, levando-se em consideração os pressupostos que levam a uma abordagem univariada ou multivariada e como objetivos específicos pretende-se: desenvolver a teoria estatística para um experimento em blocos completos ao acaso, desde a descrição das observações até a análise de variância, desdobramento dos graus de liberdade e somas dos quadrados em contrastes de interesse, gestados no planejamento da pesquisa, ou em componentes polinomiais, bem como, comparações das médias duas a duas; discutir os pressupostos que levam a uma análise univariada quando são feitas medidas repetidas e sob que condição a análise multivariada de perfil das médias deve ser implementada. Por fim, dois conjuntos de dados reais serão utilizados para ilustrar toda teoria desenvolvida. Após as análises estatísticas dos dados, em cada etapa, os resultados obtidos serão interpretado e comentados convenientemente.



## 2 Fundamentação Teórica

Neste capítulo é abordada a teoria estatística básica envolvida na análise dos dados de um experimento em blocos completos ao acaso quando a variável resposta segue uma distribuição normal e os desdobramentos, quando são feitas medições repetidas nas unidades experimentais. Neste sentido, serão discutidas desde as características dos modelos estatísticos usados para a descrição das observações, até às inferências sobre a nulidade dos efeitos dos níveis dos fatores envolvidos nos modelos, do ponto de vista univariado e multivariado, conforme sejam os objetivos dos pesquisadores ao planejarem suas pesquisas. Com este propósito, serão discutidas em detalhe, a teoria inferencial univariada para os experimentos individuais, a teoria estatística multivariada a ser utilizada quando são feitas, em momentos planejados, medições repetidas nas unidades experimentais (ou indivíduos) ao longo do tempo; as condições impostas à matriz de covariância dos erros que permitirão analisar os dados através de um modelo de blocos ao acaso com parcelas sub-divididas (*split-plot*) e finalmente, a análise multivariada dos perfis médios da variável resposta relativos aos tratamentos, conforme recomendam Rencher e Christensen (2012) e Johnson e Wichern (2007).

### 2.1 Características de um experimento em blocos completos ao acaso

Os experimentos em blocos completos ao acaso são aqueles que levam em consideração os três princípios básicos da experimentação, quais sejam: *repetição*, *casualização* e *controle local*. Para Campos (1984) e Gomes (1985), dentre outros, o delineamento em blocos ao acaso é o mais empregado na experimentação, pela versatilidade de poder ser utilizado quando se tem dúvida da homogeneidade do local onde o experimento vai ser instalado ou certeza de sua heterogeneidade. Cada bloco representa uma repetição e deve constituir um ambiente homogêneo contendo as unidades experimentais para as quais serão designadas aleatoriamente um conjunto completo de tratamentos. Um possível desenho

desse tipo de experimento instalado no campo, com 6 tratamentos e 4 blocos, pode ser apreciado na Figura 1.

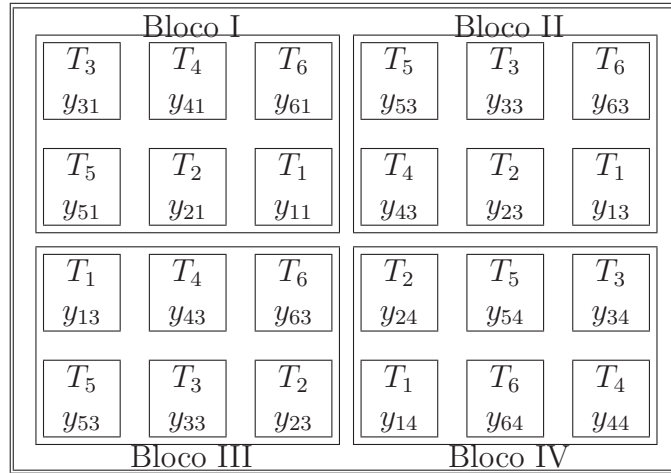


Figura 1: Um possível desenho de um experimento em blocos completos ao acaso com 6 tratamentos e 4 repetições

Observe na Figura 1 que cada bloco contém seis unidades experimentais para as quais foram designados, de forma aleatória, os 6 tratamentos. Em cada unidade experimental  $T_i$ ,  $i = 1, 2, \dots, 6$ , indica que aquela unidade experimental foi submetida ao  $i$ -ésimo tratamento e  $y_{ij}$ ,  $i = 1, 2, \dots, 6$  e  $j = 1, 2, 3, 4$ , representa o valor observado de uma variável resposta na unidade experimental que recebeu o tratamento  $i$  no  $j$ -ésimo bloco.

## 2.2 A organização das observações experimentais e o modelo estatístico associado

As medições da variável resposta relativas às unidades experimentais (ou indivíduos), correspondente aos tratamentos e blocos, identificadas na Figura 1, deverão ser convenientemente organizadas em tabelas para facilitar o desenvolvimento da teoria e as análises estatísticas subsequentes. Uma sugestão para esta finalidade pode ser observada na Tabela 1, na qual  $y_{ij}$  representa o valor da variável resposta obtido na unidade experimental que recebeu o  $i$ -ésimo tratamento no  $j$ -ésimo bloco,  $i = 1, 2, \dots, I$  e  $j = 1, 2, \dots, J$ ,  $y_{i.} = \sum_{j=1}^J y_{ij}$  e  $\bar{y}_{i.} = \frac{y_{i.}}{J}$  são respectivamente, a soma e a média das observações que receberam o  $i$ -ésimo tratamento,  $y_{.j} = \sum_{i=1}^I y_{ij}$  e  $\bar{y}_{.j} = \frac{y_{.j}}{I}$  são o total e a média das observações no  $j$ -ésimo bloco,  $y_{..} = \sum_{i=1}^I \sum_{j=1}^J y_{ij}$  e  $\bar{y}_{..} = \frac{y_{..}}{IJ}$  são o total geral e a média geral das observações

experimentais.

Tabela 1: Organização das observações obtidas de um experimento em blocos completos ao acaso

Tratamento	Bloco						Total ( $y_{i.}$ )	Média ( $\bar{y}_i$ )
	1	2	...	$j$	...	$J$		
1	$y_{11}$	$y_{12}$	...	$y_{1j}$	...	$y_{1J}$	$y_{1.}$	$\bar{y}_1$
2	$y_{21}$	$y_{22}$	...	$y_{2j}$	...	$y_{2J}$	$y_{2.}$	$\bar{y}_2$
...	...	...	...	...	...	...	...	...
$i$	$y_{i1}$	$y_{i2}$	...	$y_{ij}$	...	$y_{iJ}$	$y_{i.}$	$\bar{y}_i$
...	...	...	...	...	...	...	...	...
$I$	$y_{I1}$	$y_{I2}$	...	$y_{Ij}$	...	$y_{IJ}$	$y_{I.}$	$\bar{y}_I$
Soma	$y_{.1}$	$y_{.2}$	...	$y_{.j}$	...	$y_{.J}$	$y_{..}$	$\bar{y}_{..}$

## 2.3 O modelo estatístico e suposições associadas às observações

Quando o experimento é instalado num desenho em blocos completos ao acaso, admite-se que as observações são descritas pelo modelo aditivo

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}, \quad \begin{cases} i = 1, 2, \dots, I \\ j = 1, 2, \dots, J \end{cases} \quad (2.1)$$

em que  $\mu$  é a média geral,  $\tau_i$  e  $\beta_j$  são respectivamente, os efeitos fixos do  $i$ -ésimo tratamento e do  $j$ -ésimo bloco sobre a variável resposta e  $\epsilon_{ij}$  é o erro experimental associado a observação  $y_{ij}$ . Além disso, supomos que os erros  $\epsilon_{ij}$  são variáveis aleatórias independentes e identicamente distribuídas como uma normal de média zero e variância  $\sigma^2$  e denotaremos por  $\epsilon_{ij} \stackrel{iid}{\sim} N(0, \sigma^2)$ . Consequentemente, teremos:

$$\begin{aligned} E(\mu) &= \mu, & E(\mu^2) &= \mu^2, & E(\tau_i) &= \tau_i, & E(\tau_i^2) &= \tau_i^2, & E(\beta_j) &= \beta_j, \\ E(\beta_j) &= \beta_j, & E(\beta_j^2) &= \beta_j^2, & E(\epsilon_{ij}) &= 0, & E(\epsilon_{ij}^2) &= \sigma^2, & E(\mu\tau_i) &= \mu\tau_i, \\ E(\mu\beta_j) &= \mu\beta_j, & E(\mu\epsilon_{ij}) &= 0, & E(\tau_i\beta_j) &= \tau_i\beta_j, & E(\tau_i\epsilon_{ij}) &= 0, & E(\beta_j\epsilon_{ij}) &= 0. \end{aligned} \quad (2.2)$$

Os resultados dos valores esperados em (2.2) são de grande importância teórica e ajudam a entender melhor as demonstrações relativas aos estimadores dos parâmetros, dos valores esperados das somas de quadrados e respectivas distribuições de probabilidade associadas.

## 2.4 Estimação dos parâmetros, das observações e do erro experimental

Para estimação dos parâmetros do modelo, utilizou-se o método da máxima verossimilhança que consiste de encontrar estimativas para os parâmetros de modo que maximize a função de verossimilhança denotada por  $L(\boldsymbol{\theta}|\mathbf{y})$ . Como pode ser observado na seção anterior, ao definir o modelo (2.1), admitiu-se que os erros experimentais  $\epsilon_{ij}$  são variáveis aleatórias independentes e identicamente distribuídas como uma normal de média *zero* e variância comum  $\sigma^2$  e foi denotado por  $\epsilon_{ij} \stackrel{iid}{\sim} N(0, \sigma^2)$ . Dessa forma, sendo  $y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}$ , o erro experimental pode ser escrito como  $\epsilon_{ij} = y_{ij} - \mu - \tau_i - \beta_j$  e a função densidade de probabilidade de  $\epsilon_{ij}$  será dada por

$$f(\epsilon_{ij}) = \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2}\left(\frac{\epsilon_{ij}-0}{\sigma}\right)^2}, \quad -\infty < \epsilon_{ij} < \infty, \quad \sigma^2 > 0. \quad (2.3)$$

e substituindo-se  $\epsilon_{ij}$  por  $y_{ij} - \mu - \tau_i - \beta_j$  em (2.3), a função de verossimilhança fica

$$L(\boldsymbol{\theta}|\mathbf{y}) = \prod_{i=1}^I \prod_{j=1}^J f(\epsilon_{ij}) = (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^I \sum_{j=1}^J (y_{ij} - \mu - \tau_i - \beta_j)^2}$$

em que  $\boldsymbol{\theta}$  é um vetor  $(I + J + 2) \times 1$  dos parâmetros, isto é,

$$\boldsymbol{\theta} = \left( \mu \quad : \quad \tau_1 \quad \cdots \quad \tau_I \quad : \quad \beta_1 \quad \cdots \quad \beta_J \quad : \quad \sigma^2 \right)'$$

e  $\mathbf{y}$  é um vetor  $IJ \times 1$  das observações, ou seja,

$$\mathbf{y} = \left( y_{11} \quad \cdots \quad y_{1J} \quad y_{21} \quad \cdots \quad y_{2J} \quad \cdots \quad \cdots \quad y_{I1} \quad \cdots \quad y_{IJ} \right)'$$

Observe que se  $\hat{\boldsymbol{\theta}}$  maximiza  $L(\boldsymbol{\theta}|\mathbf{y})$ , então  $\hat{\boldsymbol{\theta}}$  também maximiza  $\ln L(\boldsymbol{\theta}|\mathbf{y})$ , uma vez que o logaritmo é uma função monótona crescente. Assim, definiu-se a logverossimilhança como,

$$\ell(\boldsymbol{\theta}) = \ln L(\boldsymbol{\theta}|\mathbf{y}) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^I \sum_{j=1}^J (y_{ij} - \mu - \tau_i - \beta_j)^2.$$

Para encontrar o vetor  $\hat{\boldsymbol{\theta}}$  que maximiza  $L(\boldsymbol{\theta}|\mathbf{y})$ , deriva-se parcialmente  $\ell(\boldsymbol{\theta})$  em relação a cada componente de  $\boldsymbol{\theta}$ , iguala-se a *zero* e explicita-se o estimador de cada componente, isto é,

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \mu} = -\frac{2}{2\sigma^2} \sum_{i=1}^I \sum_{j=1}^J (y_{ij} - \hat{\mu} - \hat{\tau}_i - \hat{\beta}_j)(-1) = 0$$

portanto,

$$IJ\hat{\mu} + J \sum_{i=1}^I \hat{\tau}_i + I \sum_{j=1}^J \hat{\beta}_j = y.. \quad (I),$$

$$\frac{\partial \ell(\theta)}{\partial \tau_i} = -\frac{2}{2\sigma^2} \sum_{j=1}^J (y_{ij} - \hat{\mu} - \hat{\tau}_i - \hat{\beta}_j)(-1) = 0$$

e segue que

$$J\hat{\mu} + J\hat{\tau}_i + \sum_{j=1}^J \hat{\beta}_j = y_i. \quad (II),$$

$$\frac{\partial \ell(\theta)}{\partial \beta_j} = -\frac{2}{2\sigma^2} \sum_{i=1}^I (y_{ij} - \hat{\mu} - \hat{\tau}_i - \hat{\beta}_j)(-1) = 0$$

e portanto,

$$I\hat{\mu} + \sum_{i=1}^I \hat{\tau}_i + I\hat{\beta}_j = y.j \quad (III)$$

finalmente,

$$\frac{\partial \ell(\theta)}{\partial \sigma^2} = -\frac{n}{2} \frac{1}{2\pi\hat{\sigma}^2} 2\pi + \frac{1}{2} \frac{1}{(\hat{\sigma}^2)^2} \sum_{i=1}^I \sum_{j=1}^J (y_{ij} - \hat{\mu} - \hat{\tau}_i - \hat{\beta}_j)^2 = 0$$

e portanto

$$\begin{aligned} \hat{\sigma}^2 &= \frac{\sum_{i=1}^I \sum_{j=1}^J (y_{ij} - \hat{\mu} - \hat{\tau}_i - \hat{\beta}_j)^2}{n} = \frac{1}{n} \sum_{i=1}^I \sum_{j=1}^J (y_{ij} - \hat{y}_{ij})^2 \\ &= \frac{SQRes}{n} = \frac{(I-1)(J-1)}{n} QMRes. \quad (IV) \end{aligned}$$

As equações (I), (II) e (III) formam o sistema de equações normais, isto é,

$$\left\{ \begin{array}{l} IJ\hat{\mu} + J \sum_{i=1}^I \hat{\tau}_i + I \sum_{j=1}^J \hat{\beta}_j = y.. \quad (I) \\ J\hat{\mu} + J\hat{\tau}_i + \sum_{j=1}^J \hat{\beta}_j = y_i. \quad (II) \\ I\hat{\mu} + \sum_{i=1}^I \hat{\tau}_i + I\hat{\beta}_j = y.j \quad (III) \end{array} \right. \quad (2.4)$$

Observe que o sistema em (2.4) é inconsistente e, portanto indeterminado, pois é constituído de três equações e no mínimo 5 incógnitas, uma vez que num experimento em blocos ao acaso envolve pelo menos dois tratamentos e dois blocos, isto é,  $I \geq 2$  e  $J \geq 2$ . Como os efeitos dos tratamentos e dos blocos são expressos como desvios entre a média do tratamento  $i$  ou do bloco  $j$  e a média geral, isto é,  $\tau_i = \mu_i - \mu$  e  $\beta_j = \mu_j - \mu$  implicando

em  $\sum_{i=1}^I \tau_i = 0$  e  $\sum_{j=1}^J \beta_j = 0$ , geralmente impõe-se esta mesma restrição para tornar o sistema consistente, ou seja,  $\sum_{i=1}^I \hat{\tau}_i = \sum_{j=1}^J \hat{\beta}_j = 0$ . Isto posto, o sistema de equações normais pode ser resolvidos e fornece os seguintes estimadores de máxima verossimilhança:

$$\hat{\mu} = \bar{y}_{..}, \quad \hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..}, \quad \hat{\beta}_j = \bar{y}_{.j} - \bar{y}_{..} \quad (2.5)$$

e a partir da equação (IV), obtém-se o estimador de máxima verossimilhança  $\hat{\sigma}^2$  de  $\sigma^2$ , isto é,

$$\hat{\sigma}^2 = \frac{(I-1)(J-1)}{n} QMRes \quad (2.6)$$

o qual é um estimador viesado para  $\sigma^2$ , a variância do erro experimental. A Tabela 2 exhibe os estimadores encontrados nesta seção para os parâmetros e para algumas funções lineares importantes dos mesmos.

Tabela 2: Estimadores das características no modelo estatístico

Característica	Notação na população	Estimador da característica
Média geral	$\mu$	$\hat{\mu} = \bar{y}_{..}$
Efeito do $i$ -ésimo tratamento	$\tau_i$	$\hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..}$
Efeito do $j$ -ésimo bloco	$\beta_j$	$\hat{\beta}_j = \bar{y}_{.j} - \bar{y}_{..}$
Média do tratamento $i$	$\mu_i = \mu + \tau_i$	$\hat{\mu}_i = \bar{y}_{i.}$
Média do bloco $j$	$\mu_j = \mu + \beta_j$	$\hat{\mu}_j = \bar{y}_{.j}$
Média de uma observação	$E(y_{ij}) = \mu + \tau_i + \beta_j$	$\hat{y}_{ij} = \hat{\mu} + \hat{\tau}_i + \hat{\beta}_j$
Erro experimental	$\epsilon_{ij}$	$\hat{\epsilon}_{ij} = y_{ij} - \hat{y}_{ij}$

Quando considera-se que as observações são provenientes de um experimento em blocos completos ao acaso, onde os efeitos dos tratamentos e blocos são fixos, estaremos admitindo que as conclusões acerca dos efeitos dos tratamentos e dos blocos, sobre a variável resposta, serão válidas apenas para os tratamentos e blocos envolvidos no experimento, não podendo ser estendidas para a população de todos os possíveis tratamentos ou blocos. Nestas condições, estaremos interessados em contrastar a hipótese de que as médias da variável resposta relativas aos tratamentos (ou blocos) são todas iguais a  $\mu$ , contra a hipótese de que existe pelo menos um par de médias da variável resposta em

relação aos tratamentos (ou blocos)  $(\mu_i, \mu_{i'})$  ou  $(\mu_j, \mu_{j'})$  que diferem entre si. Formalmente, denota-se estas hipóteses por:

$$\left( \begin{array}{c} \text{Hipóteses sobre} \\ \text{as médias dos} \\ \text{tratamentos} \end{array} \right) \longrightarrow \left\{ \begin{array}{l} H_0^{(\tau)} : \mu_1 = \mu_2 = \cdots = \mu_I = \mu \\ vs \\ H_1^{(\tau)} : \mu_i \neq \mu_{i'}, \text{ para pelo menos um par } (\mu_i, \mu_{i'}) \end{array} \right. \quad (2.7)$$

e

$$\left( \begin{array}{c} \text{Hipóteses sobre} \\ \text{as médias dos} \\ \text{blocos} \end{array} \right) \longrightarrow \left\{ \begin{array}{l} H_0^{(\beta)} : \mu_1 = \mu_2 = \cdots = \mu_J = \mu \\ vs \\ H_1^{(\beta)} : \mu_j \neq \mu_{j'}, \text{ para pelo menos um par } (\mu_j, \mu_{j'}) \end{array} \right. \quad (2.8)$$

ou alternativamente

$$\left( \begin{array}{c} \text{Hipóteses sobre} \\ \text{os efeitos dos} \\ \text{tratamentos} \end{array} \right) \longrightarrow \left\{ \begin{array}{l} H_0^{(\tau)} : \tau_i = 0, \forall i \\ vs \\ H_1^{(\tau)} : \tau_i \neq 0, \text{ para pelo menos um } i \end{array} \right. \quad (2.9)$$

e

$$\left( \begin{array}{c} \text{Hipóteses sobre} \\ \text{os efeitos dos} \\ \text{blocos} \end{array} \right) \longrightarrow \left\{ \begin{array}{l} H_0^{(\beta)} : \beta_j = 0, \forall j \\ vs \\ H_1^{(\beta)} : \beta_j \neq 0, \text{ para pelo menos um } j \end{array} \right. \quad (2.10)$$

Tendo em vista que os argumentos teóricos para contrastação das hipótese em (2.7) e (2.8) se baseiam nas partes componentes da variabilidade total, expomos na seção seguinte, como a variabilidade total é decomposta.

## 2.5 A Decomposição da Variabilidade Total

Dado o conjunto das obsevações de um experimento em blocos completos ao acaso,  $y_{ij}$ ,  $i = 1, 2, \dots, I$  e  $j = 1, 2, \dots, J$  e os estimadores na Tabela 2, cada observação pode ser escrita como

$$y_{ij} = \hat{\mu} + \hat{\tau}_i + \hat{\beta}_j + \hat{\epsilon}_{ij}, \quad \left\{ \begin{array}{l} i = 1, 2, \dots, I \\ j = 1, 2, \dots, J \end{array} \right.$$

e os estimadores dos  $\epsilon_{ij}$  (os resíduos) como

$$\hat{\epsilon}_{ij} = y_{ij} - \hat{y}_{ij} = y_{ij} - \hat{\mu} - \hat{\tau}_i - \hat{\beta}_j$$

ou

$$\hat{\epsilon}_{ij} = (y_{ij} - \bar{y}_{..}) - (\bar{y}_{i.} - \bar{y}_{..}) - (\bar{y}_{.j} - \bar{y}_{..}). \quad (2.11)$$

Elevando-se os dois lados da expressão (2.11) ao quadrado e somando-se para todas as observações, obtém-se

$$\underbrace{\sum_{i=1}^I \sum_{j=1}^J \hat{\epsilon}_{ij}^2}_{SQ_{Res.}} = \underbrace{\sum_{i=1}^I \sum_{j=1}^J (y_{ij} - \bar{y}_{..})^2}_{SQ_{Total}} - \underbrace{J \sum_{i=1}^I (\bar{y}_{i.} - \bar{y}_{..})^2}_{SQ_{Trat.}} - \underbrace{I \sum_{j=1}^J (\bar{y}_{.j} - \bar{y}_{..})^2}_{SQ_{Blocos}} \quad (2.12)$$

ou

$$\underbrace{\sum_{i=1}^I \sum_{j=1}^J \hat{\epsilon}_{ij}^2}_{SQ_{Res.}} = \underbrace{\left( \sum_{i=1}^I \sum_{j=1}^J y_{ij}^2 - C \right)}_{SQ_{Total}} - \underbrace{\left( \frac{1}{J} \sum_{i=1}^I y_{i.}^2 - C \right)}_{SQ_{Trat.}} - \underbrace{\left( \frac{1}{I} \sum_{j=1}^J y_{.j}^2 - C \right)}_{SQ_{Blocos}}. \quad (2.13)$$

em que  $C = \frac{(y_{..})^2}{IJ}$ . Isto é, a soma de quadrados dos resíduos,  $SQ_{Res.}$ , é igual a soma de quadrados total,  $SQ_{Total}$ , menos a soma de quadrados dos tratamentos,  $SQ_{Trat.}$ , menos a soma de quadrados dos blocos,  $SQ_{Blocos}$  e, segue que, a variabilidade total é decomposta em três partes, ou seja,

$$SQ_{Total} = SQ_{Trat.} + SQ_{Blocos} + SQ_{Res.} \quad (2.14)$$

Na prática, a soma de quadrados dos resíduos é obtido por diferença como em (2.13), ou seja,

$$SQ_{Res.} = SQ_{Total} - SQ_{Trat.} - SQ_{Blocos} \quad (2.15)$$

Para definir o que seja os quadrados médio é conveniente, primeiro, definir o que são os *graus de liberdade*. De acordo com Montgomery e Runger (1994, p. 28), o termo *graus de liberdade* resulta do fato de que os  $I$  desvios da média dos tratamentos em relação à média geral,  $(\bar{y}_{1.} - \bar{y}_{..})$ ,  $(\bar{y}_{2.} - \bar{y}_{..})$ ,  $\dots$ ,  $(\bar{y}_{I.} - \bar{y}_{..})$  têm sempre soma *zero*, de modo que a especificação de  $(I - 1)$  de quaisquer dessas quantidades determina, automaticamente, a restante. Assim, apenas  $(I - 1)$  dos  $I$  desvios  $(\bar{y}_{i.} - \bar{y}_{..})$  são independentes. Assim sendo, a soma de quadrados dos tratamentos tem  $\nu_t = (I - 1)$  graus de liberdade, a soma de quadrados de blocos tem  $\nu_b = (J - 1)$  graus de liberdade e a soma de quadrados dos resíduos tem  $\nu = (I - 1)(J - 1)$  graus de liberdade. Dessa forma, define-se os quadrados médios como a razão entre as somas de quadrados e seus respectivos graus de liberdade. Isto é,

$$QM_{Trat} = \frac{SQ_{Trat.}}{\nu_t}, \quad QM_{Bloco} = \frac{SQ_{Bloco}}{\nu_b} \quad \text{e} \quad QM_{Res} = \frac{SQ_{Res.}}{\nu}.$$

Os argumentos teóricos utilizados para contrastação das hipóteses em (2.7) e (2.8) são baseados nas suposições impostas aos termos do modelo em (2.1) e por consequência, nas



distribuições de probabilidade dos estimadores dos parâmetros e das somas dos quadrados. Logo, é preciso conhecer estas distribuições de probabilidade.

## 2.6 Distribuição de Probabilidade dos Estimadores e das Somas de Quadrados

Para o modelo estatístico em (2.1), supõe-se que os  $\epsilon_{ij}$ ,  $i = 1, 2, \dots, I$  e  $j = 1, 2, \dots, J$  são variáveis aleatórias independentes e identicamente distribuídas como uma normal de média zero e variância comum  $\sigma^2$ . Isto é,  $\epsilon_{ij} \stackrel{iid}{\sim} N(0, \sigma^2)$ . Supõe-se ainda, que os parâmetros  $\mu$ ,  $\tau_i$  e  $\beta_j$  são de efeitos fixos, então segue que as observações dos  $y_{ij}$ , também são variáveis aleatórias independentemente distribuídas como uma normal de média  $\mu_{ij} = \mu + \tau_i + \beta_j$  e variância comum  $\sigma^2$ , isto é,  $y_{ij} \stackrel{id}{\sim} N(\mu_{ij}, \sigma^2)$ . Isto posto, e considerando-se que sendo os estimadores dos parâmetros na Tabela 2, combinações lineares dos  $y_{ij}$  e que combinações lineares de variáveis aleatórias normais independentes, também seguem uma distribuição normal (MOOD *et al.*, 1974), então, a partir destes e outros argumentos que serão citados em momento oportuno, podem-se ser demonstradas as distribuições de probabilidade dos estimadores,  $\hat{\mu}$ ,  $\hat{\tau}_i$ ,  $\hat{\beta}_j$  e do estimador da variância do erro,  $\hat{\epsilon}_{ij}$ . Será demonstrado primeiro as distribuições de probabilidade dos estimadores dos efeitos dos tratamentos e da correspondente soma de quadrados devida aos tratamentos.

Da Tabela 2, tem-se que

$$\hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..} = \frac{1}{J} \sum_{j=1}^J y_{ij} - \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J y_{ij}, \quad (2.16)$$

que corresponde a uma combinação linear dos  $y_{ij}$ . Consequentemente,  $\hat{\tau}_i$ ,  $i = 1, 2, \dots, I$ , seguem uma distribuição normal cujas médias e variância são obtidas como:

Mas,

$$\begin{aligned} \hat{\tau}_i &= \frac{1}{J} \sum_{j=1}^J y_{ij} - \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J y_{ij} \\ &= \frac{1}{J} \sum_{j=1}^J (\mu + \tau_i + \beta_j + \epsilon_{ij}) - \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J (\mu + \tau_i + \beta_j + \epsilon_{ij}) \\ &= \frac{1}{J} \left( J\mu + J\tau_i + \sum_{j=1}^J \beta_j + \sum_{j=1}^J \epsilon_{ij} \right) - \frac{1}{IJ} \left( IJ\mu + j \sum_{i=1}^I \tau_i + I \sum_{j=1}^J \beta_j + \sum_{i=1}^I \sum_{j=1}^J \epsilon_{ij} \right) \\ &= \tau_i + \frac{1}{J} \sum_{j=1}^J \epsilon_{ij} - \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J \epsilon_{ij}. \end{aligned}$$

Daí, empregando a definição e propriedades do valor esperado, (MOOD *et al.*, 1974), obtém-se

$$E(\hat{\tau}_i) = E(\tau_i) + \frac{1}{J} \sum_{j=1}^J E(\epsilon_{ij}) - \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J E(\epsilon_{ij}) = \tau_i$$

Por outro lado,

$$\begin{aligned} Var(\hat{\tau}_i) &= E[\hat{\tau}_i - E(\hat{\tau}_i)]^2 = E\left[\tau_i + \frac{1}{J} \sum_{j=1}^J \epsilon_{ij} - \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J \epsilon_{ij} - \tau_i\right]^2 \\ &= E\left[\frac{1}{J^2} \left(\sum_{j=1}^J \epsilon_{ij}\right)^2 + \frac{1}{I^2 J^2} \left(\sum_{i=1}^I \sum_{j=1}^J \epsilon_{ij}\right)^2 - \frac{2}{IJ^2} \left(\sum_{j=1}^J \epsilon_{ij}\right) \left(\sum_{i=1}^I \sum_{j=1}^J \epsilon_{ij}\right)\right] \\ &= E\left[\frac{1}{J^2} \underbrace{(\epsilon_{i1}^2 + \dots + \epsilon_{iJ}^2)}_{J \text{ termos}} + 2\epsilon_{i1}\epsilon_{i2} + odp\right] + \frac{1}{I^2 J^2} \underbrace{(\epsilon_{11}^2 + \dots + \epsilon_{IJ}^2)}_{IJ \text{ termos}} + 2\epsilon_{11}\epsilon_{12} + odp \\ &\quad - \frac{2}{IJ^2} (\epsilon_{i1}\epsilon_{11} + op + \dots + \underbrace{\epsilon_{i1}^2 + \dots + \epsilon_{iJ}^2}_{J \text{ termos}} + \dots + \epsilon_{iJ}\epsilon_{IJ} + op) \\ &= \frac{\sigma^2}{J} + \frac{\sigma^2}{IJ} - \frac{2\sigma^2}{IJ} = \frac{(I-1)\sigma^2}{I} \frac{1}{J}. \end{aligned}$$

em que *odp*, significa outros duplos produtos e *op*, outros produtos.

Portanto,

$$\hat{\tau}_i \stackrel{id}{\sim} N\left(\tau_i, \frac{(I-1)\sigma^2}{I} \frac{1}{J}\right) \quad (2.17)$$

Além disso, sabe-se que

$$\frac{\hat{\tau}_i - E(\hat{\tau}_i)}{\sqrt{\frac{(I-1)\sigma^2}{I} \frac{1}{J}}} = \frac{\bar{y}_{i.} - \bar{y}_{..} - \tau_i}{\sqrt{\frac{(I-1)\sigma^2}{I} \frac{1}{J}}} \sim N(0, 1), \quad i = 1, 2, \dots, I$$

e que

$$\left(\frac{\bar{y}_{i.} - \bar{y}_{..} - \tau_i}{\sqrt{\frac{(I-1)\sigma^2}{I} \frac{1}{J}}}\right)^2 = \frac{J(\bar{y}_{i.} - \bar{y}_{..} - \tau_i)^2}{\frac{(I-1)\sigma^2}{I}} \sim \chi_{(1)}^2$$

e sob a hipótese  $H_0 : \tau_i = 0$  a estatística

$$\frac{J(\bar{y}_{i.} - \bar{y}_{..})^2}{\sigma^2} \sim \frac{(I-1)}{I} \chi_{(1)}^2$$

Como os termos  $\frac{(\bar{y}_{i.} - \bar{y}_{..})^2}{\sigma^2}$ ,  $i = 1, 2, \dots, I$  são independentemente distribuídos e a soma de variáveis aleatórias independentes com distribuição de qui-quadrado, também segue uma distribuição de qui-quadrado cujos graus de liberdade é a soma dos graus de liberdade

das variáveis individuais, então

$$\frac{J \sum_{i=1}^I (\bar{y}_{i.} - \bar{y}_{..})^2}{\sigma^2} = \frac{SQ_{Trat}}{\sigma^2} \sim \frac{I(I-1)}{I} \chi_{(1)}^2 = \chi_{(I-1)}^2.$$

Em palavras, a soma de quadrados de tratamentos dividida por  $\sigma^2$ , segue uma distribuição de qui-quadrado com  $(I-1)$  graus de liberdade.

Por raciocínio semelhante ao que foi usado para demonstrar as distribuições de probabilidade dos estimadores dos efeitos dos tratamentos e da respectiva soma de quadrados, demonstra-se facilmente as distribuições de probabilidade dos estimadores,  $\hat{\mu}$ ,  $\hat{\beta}_j$ ,  $\hat{\epsilon}_{ij}$  e respectivas somas de quadrados. Estes resultados estão apresentados na Tabela 3.

Tabela 3: Distribuições de probabilidade dos estimadores dos parâmetros e das somas de quadrados associadas, sob a hipótese nula

Estimador do parâmetro e sua distribuição de probabilidade	Hipótese nula	Distribuição da soma de quadrados sob $H_0^{(\cdot)}$
$\hat{\mu} = \bar{y}_{..} \sim N\left(\mu, \frac{\sigma^2}{IJ}\right)$	$H_0^{(\mu)} : \mu = 0$	$\frac{(y_{..})^2/IJ}{\sigma^2} \sim \chi_{(1)}^2$
$\hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..} \stackrel{id}{\sim} N\left(\tau_i, \frac{(I-1)\sigma^2}{I}\right)$	$\begin{cases} H_0^{(\mu_i)} : \mu_i = \mu, \forall i \\ ou \\ H_0^{(\tau_i)} : \tau_i = 0, \forall i \end{cases}$	$\frac{SQ_{Trat}}{\sigma^2} \sim \chi_{(I-1)}^2$
$\hat{\beta}_j = \bar{y}_{.j} - \bar{y}_{..} \stackrel{id}{\sim} N\left(\beta_j, \frac{J-1}{J} \frac{\sigma^2}{I}\right)$	$\begin{cases} H_0^{(\mu_j)} : \mu_j = \mu, \forall j \\ ou \\ H_0^{(\beta_j)} : \beta_j = 0, \forall j \end{cases}$	$\frac{SQ_{Blocos}}{\sigma^2} \sim \chi_{(J-1)}^2$
$\hat{\epsilon}_{ij} = y_{ij} - \hat{y}_{ij} \stackrel{iid}{\sim} N\left(0, \frac{(I-1)(J-1)}{IJ} \sigma^2\right)$	independente de $H_0^{(\cdot)}$	$\frac{SQ_{Res}}{\sigma^2} \sim \chi_{[(I-1)(J-1)]}^2$

para  $i = 1, 2, \dots, I$ ,  $j = 1, 2, \dots, J$ ,

## 2.7 Valores esperados dos quadrados médios

Os valores esperados das somas de quadrados e dos quadrados médios são tão importantes para a escolha das estatísticas dos teste das hipóteses de nulidade dos efeitos dos tratamentos e blocos, quanto as suas distribuições de probabilidade. A Tabela 4 exhibe os valores esperados dos quadrados médios para o modelo de blocos completos ao acaso, cujas demonstrações podem ser vistas em Barbin (1993, p. 65-70).

Tabela 4: Valores esperados dos quadrados médios

F. Variação	$GL$	$SQ$	$QM$	$E(QM)$
Tratamento	$\nu_t = (I - 1)$	$SQ_{Trat}$	$QM_{Trat}$	$\sigma^2 + \frac{J}{I-1} \sum_{i=1}^I (\mu_i - \mu)^2$
Bloco	$\nu_b = (J - 1)$	$SQ_{Bloco}$	$QM_{Bloco}$	$\sigma^2 + \frac{I}{J-1} \sum_{j=1}^J (\mu_j - \mu)^2$
Resíduo	$\nu = (I - 1)(J - 1)$	$SQ_{Res}$	$QM_{Res}$	$\sigma^2$

## 2.8 As estatísticas dos testes para o modelo de blocos completos ao acaso

Com base nas informações contidas na Tabela 2, sobre as distribuições de probabilidade das somas de quadrados e Tabela 3, relativas aos valores esperados dos quadrados médios, temos o seguinte:

Independentemente de que  $H_0^{(\tau)}$  e/ou  $H_0^{(\beta)}$  se verifiquem,  $\frac{SQ_{Res}}{\sigma^2}$  segue uma distribuição de qui-quadrado central com  $\nu$  graus de liberdade, sob  $H_0^{(\tau)}$ ,  $\frac{SQ_{Trat}}{\sigma^2}$  é distribuída como uma qui-quadrado com  $\nu_t$  graus de liberdade e, sob  $H_0^{(\beta)}$ ,  $\frac{SQ_{Bloco}}{\sigma^2}$  se distribui como uma qui-quadrado com  $\nu_b$  graus de liberdade. Além disso, pelo teorema de Fisher-Cochran, estas somas de quadrados se distribuem independentemente uma da outra (LUNA; OLINDA, 2014, p. 87). Observemos ainda, que o  $QM_{Res}$  é um estimador não viciado de  $\sigma^2$ , a variância do erro experimental e que sob as hipóteses  $H_0^{(\tau)}$  e  $H_0^{(\beta)}$ , tanto o  $QM_{Trat}$  quanto o  $QM_{Bloco}$  também são estimadores não viciados de  $\sigma^2$ . Sendo assim, podemos deduzir que as estatísticas de teste apropriadas para testa  $H_0^{(\tau)}$  e  $H_0^{(\beta)}$  são aquela que comparam  $QM_{Trat}$  com  $QM_{Res}$  e  $QM_{Bloco}$  com  $QM_{Res}$ , com base nos dados provenientes de um experimento (amostra). Portanto, de acordo com Rohatgi (1976, p. 317) e Mood *et al.* (1974, p. 247) dentre outros, as estatísticas

$$F_\tau = \frac{\frac{SQ_{Trat}}{\sigma^2}}{\frac{SQ_{Res}}{\sigma^2}} = \frac{\frac{(I-1)QM_{Trat}}{\sigma^2}}{\frac{(I-1)(J-1)QM_{Res}}{\sigma^2}} = \frac{QM_{Trat}}{QM_{Res}} \sim F_{[(I-1), (I-1)(J-1)]} \quad (2.18)$$

e

$$F_\beta = \frac{\frac{SQ_{Bloco}}{\sigma^2}}{\frac{SQ_{Res}}{\sigma^2}} = \frac{\frac{(J-1)QM_{Bloco}}{\sigma^2}}{\frac{(I-1)(J-1)QM_{Res}}{\sigma^2}} = \frac{QM_{Bloco}}{QM_{Res}} \sim F_{[(J-1), (I-1)(J-1)]} \quad (2.19)$$

serão as estatísticas de teste para testar as hipótese  $H_0^{(\tau)}$  e  $H_0^{(\beta)}$  ao nível de significância  $\alpha$ . Dessa forma, as regras de decisão serão: Rejeitar  $H_0^{(\tau)}$  em favor de  $H_1^{(\tau)}$  e/ou  $H_0^{(\beta)}$

em favor de  $H_1^{(\beta)}$  se observarmos um valor para  $F_\tau$  e/ou  $F_\beta$  maior que o correspondente valor do percentil superior da distribuição  $F$  com  $\nu_t$  e  $\nu$  e/ou  $\nu_b$  e  $\nu$  graus de liberdade ao nível de significância  $\alpha$ . Na prática, estas hipóteses serão verificadas organizando-se os resultados dos cálculos na Tabela 5, conhecida por tabela da análise de variância.

Tabela 5: Análise da variância para testar a significância dos efeitos de blocos e tratamentos sobre a variável resposta

Fonte de variação	$GL$	$SQ$	$QM$	$F$
Tratamento	$\nu_t$	$SQ_{Trat}$	$QM_{Trat}$	$QM_{Trat}/QM_{Res}$
Bloco	$\nu_b$	$SQ_{Bloco}$	$QM_{Bloco}$	$QM_{Bloco}/QM_{Res}$
Resíduo	$\nu$	$SQ_{Res}$	$QM_{Res}$	-
Total	$IJ - 1$	$SQ_{Total}$	-	-

A rejeição de  $H_0^{(\tau)}$  e/ou  $H_0^{(\beta)}$  sugere apenas que existe pelo menos um par de médias dos tratamentos  $(\mu_i, \mu_{i'})$  e/ou dos blocos  $(\mu_j, \mu_{j'})$  que diferem entre si. Isto nos leva aos seguintes questionamentos:

- a) Quais contrastes de médias estabelecidos no planejamento da pesquisa são estatisticamente significativos?
- b) Se os tratamentos e/ou os níveis do fator bloco forem quantitativos, qual o comportamento da variável resposta em relação aos tratamentos ou aos níveis do fator bloco?
- c) Quais os pares de médias que diferem entre si?

### 2.8.1 Desdobramento dos graus de liberdade e somas de quadrados em contrastes de interesse

Com respeito ao item a) da seção 2.8, é comum, em muitas situações práticas, o pesquisador planeja um experimento de tal forma que lhe interessa testar a não significância de  $\nu_t$  contrastes, ortogonais entre si, sobre as médias dos tratamentos e/ou sobre as  $\nu_b$  médias dos blocos. Neste caso, é ilustrativo e mais indicado que desdobremos os graus de liberdade e respectivas somas de quadrados, em contrastes de interesse, dentro da própria análise de variância, levando-se em conta os seguintes argumentos teóricos:

Seja o contraste entre as médias dos tratamentos,

$$\Psi_h = \sum_{i=1}^I c_{hi}\mu_i = c_{h1}\mu_1 + c_{h2}\mu_2 + \cdots + c_{hI}\mu_I, \quad h = 1, 2, \dots, (I - 1)$$

em que  $c_{hi}$  são constantes reais, com  $\sum_{i=1}^I c_{hi} = 0$ , e cujo estimador do contraste é dado por

$$\hat{\Psi}_h = \sum_{i=1}^I c_{hi} \bar{y}_i. = c_{h1} \bar{y}_1. + c_{h2} \bar{y}_2. + \cdots + c_{hI} \bar{y}_I., \quad h = 1, 2, \dots, (I - 1).$$

Nestas condições, considerando-se que  $y_{ij} \stackrel{iid}{\sim} N(\mu_i, \sigma^2)$ , que os tratamentos são igualmente repetidos e que a soma de variáveis aleatórias normais, também segue uma distribuição normal, é simples demonstrar que,

$$E(\hat{\Psi}_h) = \sum_{i=1}^I c_{hi} \mu_i, \quad Var(\hat{\Psi}_h) = \sum_{i=1}^I \frac{c_{hi}^2}{J} \sigma^2, \quad \hat{\Psi}_h \sim N \left( \sum_{i=1}^I c_{hi} \mu_i, \sum_{i=1}^I \frac{c_{hi}^2}{J} \sigma^2 \right)$$

e segue que

$$\frac{\hat{\Psi}_h - E(\hat{\Psi}_h)}{\sqrt{Var(\hat{\Psi}_h)}} = \frac{\sum_{i=1}^I c_{hi} \bar{y}_i. - \sum_{i=1}^I c_{hi} \mu_i}{\sqrt{\sum_{i=1}^I \frac{c_{hi}^2}{J} \sigma^2}} \sim N(0, 1).$$

Além disso, sabendo-se que o quadrado de uma variável aleatória com distribuição normal padrão segue uma distribuição de qui-quadrado com *um* grau de liberdade, podemos escrever, sob a hipótese nula,  $H_0 : \Psi = \sum_{i=1}^I c_{hi} \mu_i = 0$ , o seguinte,

$$\frac{(\sum_{i=1}^I c_{hi} \bar{y}_i.)^2}{\sum_{i=1}^I \frac{c_{hi}^2}{J} \sigma^2} = \frac{(\sum_{i=1}^I c_{hi} y_i.)^2}{JK_h \sigma^2} = \frac{SQ(\hat{\Psi}_h)}{\sigma^2} \sim \chi_{(1)}^2. \quad (2.20)$$

Em palavras, a soma de quadrados do contraste  $\hat{\Psi}_h$ , isto é,  $SQ(\hat{\Psi}_h) = (\sum_{i=1}^I c_{hi} y_i.)^2 / JK_h$ , em que  $K_h = \sum_{i=1}^I c_{hi}^2$ , dividido por  $\sigma^2$  é distribuída como uma chi-quadrado com *um* grau de liberdade.

Por outro lado, mostramos que o valor esperado da  $SQ(\hat{\Psi}_h)$ ,  $h = 1, 2, \dots, (I - 1)$ , é

$$E[SQ(\hat{\Psi}_h)] = \sigma^2 + \frac{J}{K_h} \sum_{i=1}^I (c_{hi} \tau_i)^2 = \sigma^2 + \frac{J}{K_h} \sum_{i=1}^I (c_{hi} \mu_i)^2 \quad (2.21)$$

e que o valor esperado do  $QM(\hat{\Psi}_h)$ ,  $h = 1, 2, \dots, (I - 1)$ , é

$$E[QM(\hat{\Psi}_h)] = E \left[ \frac{SQ(\hat{\Psi}_h)}{1} \right] = \sigma^2 + \frac{J}{K_h} \sum_{i=1}^I (c_{hi} \tau_i)^2 = \sigma^2 + \frac{J}{K_h} \sum_{i=1}^I (c_{hi} \mu_i)^2. \quad (2.22)$$

Observe que, sendo ortogonais entre si e estabelecidos no planejamento do experimento, as somas de quadrados dos contrastes de interesse são independentemente distribuídas e obedecem ao que postula o teorema de Fisher-Cochran, (LUNA; OLINDA, 2014, p. 87). Além disso, sob  $H_0^{(\Psi_h)} : \Psi_h = \sum_{i=1}^I c_{hi}\mu_i = 0$ ,  $h = 1, 2, \dots, (I - 1)$ , o quadrado médio do contraste  $h$ , isto é,  $QM(\hat{\Psi}_h)$ , é um estimador não viciado para  $\sigma^2$  e segue que a estatística

$$F_{\Psi_h} = \frac{QM(\hat{\Psi}_h)}{QM_{Res}} \sim F_{[1, \nu]}. \quad (2.23)$$

Portanto, rejeitamos  $H_0^{(\Psi)} : \Psi_h = 0$  em favor de  $H_1^{(\Psi_h)} : \Psi_h \neq 0$ , ao nível de significância  $\alpha$ , se para um determinado conjunto de dados experimentais, observarmos que  $F_{\Psi_h} > F_{[1, \nu, \alpha]}$ .

A Tabela 6 a seguir, apresenta uma reestruturação da análise de variância.

Tabela 6: Análise de variância para testar a significância dos contrastes de interesse, os efeitos dos tratamentos e dos blocos

Fonte de variação	$GL$	$SQ$	$QM$	$F$
Contraste $\hat{\Psi}_1$	1	$SQ(\hat{\Psi}_1)$	$QM(\hat{\Psi}_1)$	$QM(\hat{\Psi}_1)/QM_{Res}$
Contraste $\hat{\Psi}_2$	1	$SQ(\hat{\Psi}_2)$	$QM(\hat{\Psi}_2)$	$QM(\hat{\Psi}_2)/QM_{Res}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
Contraste $\hat{\Psi}_{(I-1)}$	1	$SQ(\hat{\Psi}_{I-1})$	$QM(\hat{\Psi}_{I-1})$	$SQ(\hat{\Psi}_{I-1})/QM_{Res}$
Tratamento	$(\nu_t)$	$(SQ_{Trat})$	$QM_{Trat}$	$QM_{Trat}/QM_{Res}$
Bloco	$\nu_b$	$SQ_{Bloco}$	$QM_{Bloco}$	$QM_{Bloco}/QM_{Res}$
Resíduo	$\nu$	$SQ_{Res}$	$QM_{Res}$	
Total	$IJ - 1$	$SQ_{Total}$		

Se a hipótese de igualdade das médias, na análise de variância,  $H_0^{(\tau)}$  e/ou  $H_0^{(\beta)}$ , for rejeitada e ao pesquisador interessar apenas alguns poucos contrastes sobre as médias dos tratamentos e/ou blocos, recomendamos a utilização do procedimento de Scheffé (CAMPOS, 1984, p. 39-40). Para tanto, deveremos calcular os valores absolutos das estimativas dos contrastes, isto é,

$$|\hat{\psi}_h| = \left| \sum_{i=1}^I c_{hi}\bar{y}_i \right| = |c_{h1}\bar{y}_1 + c_{h2}\bar{y}_2 + \dots + c_{hI}\bar{y}_I|, \quad h = 1, 2, \dots$$

e comparamos, ao nível de significância  $\alpha$ , com o valor da estatística,

$$S = \sqrt{\nu_t \hat{Var}(\hat{\psi}_h) F_{[1, \nu, \alpha]}} \quad (2.24)$$

em que  $\hat{Var}(\hat{\psi}_h) = \sum_{i=1}^I c_{hi}^2 \frac{s^2}{J}$  e  $s^2 = QM_{Res}$  e, finalmente, rejeitamos a hipótese  $H_0^{(\Psi_h)}$  :

$\Psi_h = 0$  em favor de  $H_1^{(\Psi_h)} : \Psi_h \neq 0$ , ao nível de significância  $\alpha$ , se observarmos que  $|\hat{\psi}_h| > S$ , para os dados de um experimento em particular.

## 2.8.2 Desdobramento dos graus de liberdade e somas de quadrados em componentes polinomiais

Com relação ao item b) da seção 2.8, existem situações nas quais queremos estudar o comportamento de uma variável resposta (por exemplo, a produção) em relação aos níveis de um fator (por exemplo, quantidades de  $P_2O_5$  aplicados no solo: 0, 20, 40, 60, 80kg/ha) ou o comportamento do crescimento de uma cultivar em relação ao tempo (por exemplo, 30, 40, 50, 60, 70 dias após a emergência). Nestas condições, a comparação simples das médias da variável resposta em relação aos níveis do fator (os tratamentos) não é recomendada. Uma alternativa para esta situação, será estudar o comportamento das respostas em relação aos níveis do fator através dos polinômios ortogonais, desdobrando-se os graus de liberdade dos tratamentos e/ou blocos e respectivas somas de quadrados em componentes polinomiais do primeiro até o  $(I - 1)$ -ésimo grau.

O respauço inferencial utilizado no desdobramento dos graus de liberdade e respectivas somas de quadrados dos tratamentos e/ou bloco em componentes polinomiais é semelhante a aquele empregado no desdobramento em contrastes de interesse, sendo que os coeficientes dos contrastes  $c_{hi}$  são substituídos pelos coeficientes dos polinômios ortogonais  $c_{gi}$ , com  $g = 1, 2, \dots, (I - 1)$ , onde o índice  $g$  indica o grau do polinômio. Assim, considerando-se os tratamentos, a soma de quadrados do componente de  $g$ -ésimo grau,

$$SQ(C_g) = SQ(\text{Comp. do } g^o \text{ grau}) = \frac{\left(\sum_{i=1} c_{gi} y_i\right)^2}{JK_g}, \quad g = 1, 2, \dots, (I - 1) \quad (2.25)$$

tem um grau de liberdade, a razão  $\frac{SQ(C_g)}{\sigma^2}$  segue uma distribuição de qui-quadrado com  $um$  grau de liberdade,  $F_g = QM(C_g)/QM_{Res}$  segue uma distribuição  $F$  de Snedecor e Fisher com  $um$  grau de liberdade do numerador e  $\nu$  graus de liberdade do denominador e será a estatística para contrastar as hipótese  $H_0^{(\theta_g)} : \theta_g = 0$  contra  $H_1^{(\theta_g)} : \theta_g \neq 0$ , onde  $\theta_g$  representa o coeficiente do  $g$ -ésimo grau do polinômio,  $g = 1, 2, \dots, (I - 1)$ . Os coeficientes  $c_{gi}$  estão disponíveis na Tabela 23 do Apênce A, para o número de níveis equidistantes do fator (os tratamentos). Nos casos em que os níveis não são equidistantes poderemos determiná-los facilmente usando um processo prático (CAMPOS, 1984, p. 251-271).

A técnica dos polinômios ortogonais consiste em assumir *à priori* que a variável resposta em relação aos níveis do fator (os tratamentos), por exemplo, pode ser descrita pelo



polinômio de grau  $(I - 1)$ , isto é,

$$y_{ij} = \mu + \underbrace{\theta_1 X_i + \theta_2 X_i^2 + \cdots + \theta_{I-1} X_i^{I-1}}_{\tau_i} + \beta_j + e_{ij}, \quad \begin{cases} i = 1, 2, \dots, I \\ j = 1, 2, \dots, J \end{cases}$$

em que  $\theta_i$  são os parâmetros de um polinômio de grau  $(I - 1)$ ,  $X_i$  são os níveis do fator (os tratamentos). Assim, a nossa tarefa em termos de análises estatísticas é contrastar as hipóteses,

$$H_0^{(\theta_g)} : \theta_g = 0 \quad vs \quad H_1^{(\theta_g)} : \theta_g \neq 0, \quad g = 1, 2, \dots, (I - 1)$$

e em seguida, selecionar o modelo que descreve o comportamento da variável resposta em função das rejeições das hipóteses  $H_0^{(\theta_g)} : \theta_g = 0, g = 1, 2, \dots, (I - 1)$ . A estatística de teste em (2.25) deve ser usada para esta finalidade. A Tabela 7 corresponde a uma reestruturação da análise da variância quando os graus de liberdade e somas de quadrados devidos aos tratamentos são desdobrados em componentes polinomiais.

Tabela 7: Análise de variância para testar a nulidade dos coeficientes do polinômio, dos efeitos dos tratamentos e dos blocos sobre a variável resposta

Fonte de variação	$GL$	$SQ$	$QM$	$F$
$C_1$ =Comp. do 1 <sup>o</sup> grau	1	$SQ(C_1)$	$QM(C_1)$	$\frac{QM(C_1)}{QM_{Res}}$
$C_2$ =Comp. do 2 <sup>o</sup> grau	1	$SQ(C_2)$	$QM(C_2)$	$\frac{QM(C_2)}{QM_{Res}}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$C_{\nu_t}$ =Comp. do $\nu_t^0$ grau	1	$SQ(C_{(I-1)})$	$QM(C_{(I-1)})$	$\frac{SQ(C_{(I-1)})}{QM_{Res}}$
Tratamento	$(\nu_t)$	$(SQ_{Trat})$	$QM_{Trat}$	$\frac{QM_{Trat}}{QM_{Res}}$
Bloco	$\nu_b$	$SQ_{Bloco}$	$QM_{Bloco}$	$\frac{QM_{Bloco}}{QM_{Res}}$
Resíduo	$\nu$	$SQ_{Res}$	$QM_{Res}$	
Total	$IJ - 1$	$SQ_{Total}$		

O valor médio da variável resposta para um dado tratamento  $X_i$  estimado por um polinômio de grau  $g$  é dado pela expressão

$$\hat{Y}_i = \hat{E}(Y|x) = \hat{\mu} + T_1 M_1 P_1 + T_2 M_2 P_2 + \cdots + T_g M_g P_g, \quad i = 1, 2, \dots, I \quad (2.26)$$

em que

$$\hat{\mu} = \bar{y}_{..}, \quad T_1 = \frac{\sum_{i=1}^I c_{1i} y_i}{JK_1}, \quad T_2 = \frac{\sum_{i=1}^I c_{2i} y_i}{JK_2}, \quad \dots, \quad T_g = \frac{\sum_{i=1}^I c_{gi} y_i}{JK_g}, \quad (2.27)$$

as constantes  $M_g$  e  $K_g$  estão disponíveis na Tabela 23,  $\hat{Y}_i$  é o estimador da variável resposta em relação ao nível  $X_i$  e  $P_g$  são os polinômios de grau  $g$ , que são dados por

$$P_1 = x, P_2 = x^2 - \frac{I-1}{12}, P_3 = x^3 - \frac{3I^2-7}{20}x, \dots, \quad \text{com } x = \frac{X_i - \bar{X}}{q} \quad (2.28)$$

e  $q$  é a equidistância entre os níveis do fator (os tratamentos). Substituindo-se (2.27) e (2.28) em (2.26) obtemos o polinômio de grau  $g$  ajustado

$$\hat{Y}_i = \hat{\theta}_0 + \hat{\theta}_1 X_i + \hat{\theta}_2 X_i^2 + \dots + \hat{\theta}_g X_i^g, \quad i = 1, 2, \dots, I.$$

### 2.8.3 Comparações múltiplas das médias pelo método de Tukey

Embora tenha-se abordado primeiro as possibilidades dos desdobramentos dos graus de liberdade e respectivas somas de quadrados dos tratamentos e/ou blocos em contrastes de interesse e em componentes polinomiais (subseções 2.8.1 e 2.8.2), as comparações simples das médias dos tratamentos e/ou blocos por um método de comparações múltiplas, citada no item c) da seção 2.8, é o procedimento mais usado na prática. Para esta finalidade, utilizaremos neste trabalho, o método de Tukey para comparar as médias dos tratamentos (a comparação das médias dos blocos é semelhante) cujo procedimento é o seguinte:

1. Calcular o valor absoluto das estimativas de todos os possíveis contrastes entre duas médias dos tratamentos (existem  $\binom{I}{2} = I(I-1)/2$  contrastes entre duas médias dos tratamentos), isto é,

$$|\hat{\psi}_h| = |\bar{y}_i - \bar{y}_{i'}|, \quad \forall i \neq i', \quad i, i' = 1, 2, \dots, I \text{ e } h = 1, 2, \dots, I(I-1)/2,$$

2. Calcular a diferença mínima significativa,  $\Delta$ , ao nível de significância  $\alpha$ , (em geral  $\alpha = 0,05$ ), ou seja,

$$\Delta = q_{(I; \nu, \alpha)} \sqrt{\frac{QM_{Res}}{J}}$$

onde  $q_{(I; \nu, \alpha)}$  é o valor crítico da distribuição de probabilidade da amplitude total estudentizada de Tukey correspondente ao número de tratamentos  $I$ , número de graus de liberdade dos resíduos  $\nu$  e nível de significância  $\alpha$ ;

3. Comparar  $|\hat{\psi}_h|$ ,  $h = 1, 2, \dots, I(I-1)/2$ , com  $\Delta$ ;

4. **Usar a regra de decisão:** Se  $|\hat{\psi}_h| > \Delta$  rejeitar a hipótese  $H_0^{(h)} : \mu_i = \mu_{i'}$  e concluir, ao nível de significância  $\alpha$ , que há diferença significativa entre as médias  $\mu_i$  e  $\mu_{i'}$ .

Finalmente, as análises estatísticas posteriores à rejeição da hipótese de igualdade das médias dos tratamentos e/ou dos blocos pela ANOVA, pode envolver simultaneamente, contrastes de interesse, polinômios ortogonais e testes de comparações múltiplas, dependendo dos objetivos da pesquisa, do plano experimental e dos níveis do fator em estudo, se qualitativo, quantitativo ou ambos.

## 2.9 Experimentos em blocos ao acaso com medidas repetidas

Na prática, é muito comum o pesquisador instalar um experimento em blocos ao acaso e, em momentos pontuais ao longo do tempo, ele faz medições de uma (ou mais de uma) variável resposta nas unidades experimentais ou indivíduos. Esta atitude, conduz a um desenho experimental de blocos ao acaso com medidas repetidas, cuja disposição das observações pode ser vista na Tabela 8. Uma outra situação que caracteriza um desenho com medidas repetidas é aquela em que o pesquisador atribui às unidades experimentais (os indivíduos) uma sequência de tratamentos. Para um delineamento em blocos ao acaso com medidas repetidas temos *um* fator de medidas repetidas (intra-sujeitos), *C* e *dois* fatores de agrupamento (entre-sujeitos), *A* e *B*. O modelo estatístico univariado para descrever cada observação é

$$y_{ijk} = \mu + \tau_i + \beta_j + \epsilon_{ij} + \theta_k + \gamma_{ik} + \varepsilon_{ijk}, \quad \begin{cases} i = 1, 2, \dots, I \\ j = 1, 2, \dots, J \\ k = 1, 2, \dots, K \end{cases} \quad (2.29)$$

em que  $\mu$  é a média geral,  $\tau_i$  e  $\beta_j$  são, respectivamente, o efeito do tratamento  $i$  e do  $j$ -ésimo bloco sobre a variável resposta,  $\epsilon_{ij}$  é o erro experimental entre-sujeitos,  $\theta_k$  é o efeito da  $k$ -ésima época de avaliação (efeito do tempo),  $\gamma_{ij}$  é o efeito da interação do  $i$ -ésimo tratamento com a  $k$ -ésima época de avaliação e  $\varepsilon_{ijk}$  é o erro experimental intra-sujeitos.

Na Tabela 8, as observações  $y_{ijk}$  com  $k = 1, 2, \dots, K$  estão entre parênteses e denotadas por  $\mathbf{y}'_{ij}$  para indicar que essas  $K$  variáveis são medidas sobre um indivíduo constituindo um vetor de variáveis correlacionadas. As amplitudes dos subscritos podem ser vistos na Tabela 8:  $i = 1, 2, \dots, I$ ,  $j = 1, 2, \dots, J$  e  $k = 1, 2, \dots, K$  com os fatores *A*, *B* e *C* fixos e os indivíduos aleatórios, a ANOVA univariada é dada na Tabela 13.

Tabela 8: Organização das observações de um experimento em blocos ao acaso com medidas repetidas

Tratamento (Fator A)	Bloco (Fator B)	Época de Avaliação (Fator C)						Obs.
		1	2	...	$k$	...	$K$	
1	1	$(y_{111}$	$y_{112}$	$\cdots$	$y_{11k}$	$\cdots$	$y_{11K})$	$= \mathbf{y}'_{11}$
1	2	$(y_{121}$	$y_{122}$	$\cdots$	$y_{12k}$	$\cdots$	$y_{12K})$	$= \mathbf{y}'_{12}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
1	$j$	$(y_{1j1}$	$y_{1j2}$	$\cdots$	$y_{1jk}$	$\cdots$	$y_{1jK})$	$= \mathbf{y}'_{1j}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
1	$J$	$(y_{1J1}$	$y_{1J2}$	$\cdots$	$y_{1Jk}$	$\cdots$	$y_{1JK})$	$= \mathbf{y}'_{1J}$
2	1	$(y_{211}$	$y_{212}$	$\cdots$	$y_{21k}$	$\cdots$	$y_{21K})$	$= \mathbf{y}'_{21}$
2	2	$(y_{221}$	$y_{222}$	$\cdots$	$y_{22k}$	$\cdots$	$y_{22K})$	$= \mathbf{y}'_{22}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
2	$j$	$(y_{2j1}$	$y_{2j2}$	$\cdots$	$y_{2jk}$	$\cdots$	$y_{2jK})$	$= \mathbf{y}'_{2j}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
2	$J$	$(y_{2J1}$	$y_{2J2}$	$\cdots$	$y_{2Jk}$	$\cdots$	$y_{2JK})$	$= \mathbf{y}'_{2J}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$i$	1	$(y_{i11}$	$y_{i12}$	$\cdots$	$y_{i1k}$	$\cdots$	$y_{i1K})$	$= \mathbf{y}'_{i1}$
$i$	2	$(y_{i21}$	$y_{i22}$	$\cdots$	$y_{i2k}$	$\cdots$	$y_{i2K})$	$= \mathbf{y}'_{i2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$i$	$j$	$(y_{ij1}$	$y_{ij2}$	$\cdots$	$y_{ijk}$	$\cdots$	$y_{ijK})$	$= \mathbf{y}'_{ij}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$i$	$J$	$(y_{iJ1}$	$y_{iJ2}$	$\cdots$	$y_{iJk}$	$\cdots$	$y_{iJK})$	$= \mathbf{y}'_{iJ}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$I$	1	$(y_{I11}$	$y_{I12}$	$\cdots$	$y_{I1k}$	$\cdots$	$y_{I1K})$	$\mathbf{y}'_{I1}$
$I$	2	$(y_{I21}$	$y_{I22}$	$\cdots$	$y_{I2k}$	$\cdots$	$y_{I2K})$	$\mathbf{y}'_{I2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$I$	$j$	$(y_{Ij1}$	$y_{Ij2}$	$\cdots$	$y_{Ijk}$	$\cdots$	$y_{IjK})$	$\mathbf{y}'_{Ij}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$I$	$J$	$(y_{IJ1}$	$y_{IJ2}$	$\cdots$	$y_{IJk}$	$\cdots$	$y_{IJK})$	$\mathbf{y}'_{IJ}$

No entanto, de imediato poderia-se pensar em uma ANOVA univariada porque os  $y$ 's em cada linha são correlacionados e a suposição de independência é, de acordo com Rencher e Christensen (2012), mais importante do que as suposições de normalidade e homogeneidade das variâncias. Discutiremos a seguir algumas suposições exigidas para a realização de uma análise univariada.

Na abordagem multivariada, as respostas  $y_{ij1}, y_{ij2}, \dots, y_{ijK}$  (as medidas repetidas) para a unidade experimental que recebeu o tratamento  $i$  no bloco  $j$  constituem um vetor  $\mathbf{y}_{ij}$ , conforme pode ser visto na Tabela 8.

O modelo multivariado é um modelo MANOVA para dois fatores sem interação,

$$\mathbf{y}_{ij} = \boldsymbol{\mu} + \boldsymbol{\tau}_i + \boldsymbol{\beta}_j + \boldsymbol{\epsilon}_{ij} \quad (2.30)$$

em que  $\boldsymbol{\mu}$  é o vetor média geral correspondente às  $K$  variáveis em  $\mathbf{y}_{ij}$ ,  $\boldsymbol{\tau}_i$  é um vetor de  $K$  efeitos principais (correspondentes às  $K$  variáveis em  $\mathbf{y}_{ij}$ ) do fator  $A$ ,  $\boldsymbol{\beta}_j$  é um vetor de  $K$  efeitos principais (correspondentes às  $K$  variáveis em  $\mathbf{y}_{ij}$ ) do fator  $B$ , e  $\boldsymbol{\epsilon}_{ij}$  é um vetor de erros da unidade experimental que recebeu o tratamento  $i$  no bloco  $j$ . Este modelo inclui apenas os fatores  $A$  e  $B$  mas, veremos posteriormente como usar uma abordagem similar àquela usada na análise de perfil para obter testes sobre o fator  $C$  e para a interação  $A \times C$ . A MANOVA exige que  $Cov(\mathbf{y}_{ij}) = \boldsymbol{\Sigma}$  para todo  $i$  e  $j$  permitindo que as  $K$  medidas repetidas sejam correlacionadas em qualquer padrão, uma vez que  $\boldsymbol{\Sigma}$  é completamente geral. Por outro lado, as suposições para a MANOVA de independência e homogeneidade das variâncias podem ser expressas como  $Cov(\mathbf{y}_{ij}) = \sigma^2 \mathbf{I}$ . Será uma surpresa se as medidas repetidas sobre o mesmo sujeito forem independentes.

A abordagem pela análise de variância univariada é apropriado em condições menos rigorosas do que  $\boldsymbol{\Sigma} = \mathbf{I}\sigma^2$ . Wilks (1946) mostrou que o teste  $F$  normal da ANOVA é válido quando a estrutura de covariância é do tipo

$$Cov(\mathbf{y}_{ij}) = \boldsymbol{\Sigma} = \sigma^2 \begin{pmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \cdots & \rho \\ \vdots & \vdots & \vdots & & \vdots \\ \rho & \rho & \rho & \cdots & 1 \end{pmatrix} = \sigma^2[(1 - \rho)\mathbf{I} + \rho\mathbf{J}], \quad (2.31)$$

em que  $\mathbf{J}$  é uma matriz quadrada de 1's. O padrão de covariância em (2.31) é também conhecido por *padrão de uniformidade*, *simetria composta*, ou ainda o modelo de *correlação intraclasse*. Este padrão de covariância, permite que as variáveis sejam correlacionadas, mas restringe a que todas as variáveis tenham a mesma variância e cada par de variáveis tenha a mesma covariância. Num experimento cuidadosamente planejado com uma aleatorização apropriada, essa suposição pode ser assegurada sob a hipótese de que  $C$  não tem efeito. Alternativamente, poderemos usar um teste da hipótese de que  $\boldsymbol{\Sigma}$  tem o padrão (2.31). Se esta hipótese é aceita poderemos prosseguir com o teste  $F$  da ANOVA habitual.

Bock (1963) e Huynh e Feldt (1970) mostraram que a condição mais geral sob a qual

o teste  $F$  univariado continua válido é que

$$\mathbf{C}\Sigma\mathbf{C}' = \sigma^2\mathbf{I}, \quad (2.32)$$

em que  $\mathbf{C}$  é uma matriz  $(K - 1) \times K$  cujas linhas são contrastes ortogonais (contrastos ortogonais normalizados para ter comprimento  $um$ ). Podemos construir  $\mathbf{C}$ , escolhendo quaisquer  $(K - 1)$  contrastes ortogonais entre as médias  $\mu_1, \mu_2, \dots, \mu_K$  do fator medidas repetidas e dividindo cada contraste  $\sum_{k=1}^K c_k \mu_k$  por  $\sqrt{\sum_{k=1}^K c_k^2}$ . Pode ser mostrado que a expressão (2.31) é um caso especial de (2.32). A condição (2.32) é conhecida como *condição de esfericidade*, embora este termo também possa se referir ao padrão de covariância  $\Sigma = \sigma^2\mathbf{I}$  quando  $\mathbf{y}_{ij}$  não é transformado.

Uma maneira simples de testar a hipótese de que (2.32) se verifique é transformar os dados em  $\mathbf{z}_{ij} = \mathbf{C}\mathbf{y}_{ij}$  e testar  $H_0 : \Sigma_z = \sigma^2\mathbf{I}$ , usando  $\mathbf{C}\mathbf{S}_{pl}\mathbf{C}'$  em que  $\mathbf{S}_{pl} = \mathbf{E}/\nu$ . Assim, um procedimento para análise dos dados de um experimento com medidas repetidas é fazer preliminarmente o teste (2.31) ou (2.32) e, se a hipótese for aceita, usar o teste  $F$  univariado da ANOVA de acordo com a Tabela 13.

Se o teste univariado sobre o fator  $C$  das medidas repetidas, for apropriado, ele é mais poderoso porque tem mais graus de liberdade para o erro do que o teste multivariado correspondente. No entanto, mesmo suaves afastamentos de (2.32) aumenta seriamente o Erro do Tipo I do teste univariado para o fator  $C$  (BOX, 1954); (DAVIDSON, 1972); (BOIK, 1981 apud RENCHER; CHRISTENSEN, 2012, p. 218). Como esses desvios podem ser facilmente perdidos num teste preliminar, Boik (1981, p. 248 e 254) conclui que “no geral, o teste  $F$  normal não pode ser recomendado” e enfatizou que “não há justificativa para o emprego normal do teste  $F$  univariado para contrastes dos tratamentos com medidas repetidas”.

Outra abordagem para análise de experimentos com medidas repetidas é ajustar o teste  $F$  univariado para a quantidade de afastamento da esfericidade. Box (1954) e Greenhouse e Geisser (1959) mostraram que quando  $\Sigma \neq \sigma^2\mathbf{I}$ , uma aproximação do teste  $F$  para efeitos envolvendo as medidas repetidas é obtida reduzindo-se os graus de liberdade do numerador e do denominador multiplicando-se pelo fator

$$\varepsilon = \frac{[\text{tr}(\Sigma - \mathbf{J}\Sigma/K)]^2}{(K - 1)\text{tr}(\Sigma - \mathbf{J}\Sigma/K)^2}, \quad (2.33)$$

em que  $\mathbf{J}$  é uma matriz quadrada  $(K \times K)$  de 1's. Por exemplo, na Tabela 9 o valor de  $F_{A \times C}$ , para interação  $A \times C$ , será comparado com o valor de  $F_\alpha$  com  $\varepsilon(I - 1)(K - 1)$  e  $\varepsilon I(J - 1)(K - 1)$  graus de liberdade, isto é, com o percentil  $F_{[\varepsilon\nu_g, \varepsilon\nu, \alpha]}$ . Uma estimativa  $\hat{\varepsilon}$

pode ser obtida substituindo-se  $\Sigma$  em (2.33) por  $\hat{\Sigma} = \mathbf{E}/\nu$ . Greenhouse e Geisser (1959) mostraram que  $\varepsilon$  e  $\hat{\varepsilon}$  variam de  $1/(K-1)$  a 1, com  $\varepsilon = 1$  quando a esfericidade se verifica e  $\varepsilon \geq 1/(K-1)$  para outros valores de  $\Sigma$ . Assim,  $\varepsilon$  é uma medida de não esfericidade. Para um teste conservador, Greenhouse e Geisser (1959) recomendam dividir os graus de liberdade do numerador e do denominador por  $(K-1)$ .

Conforme Rencher e Christensen (2012, p. 219), o comportamento do teste  $F$  univariado aproximado com os graus de liberdade ajustados por  $\hat{\varepsilon}$  tem sido investigado por Collier *et al.* (1967). Huynh (1978), Davidson (1972), Rogan *et al.* (1979) e Maxwell e Avery (1982). Nestes estudos, o nível  $\alpha$  revelou-se próximo do  $\alpha$  nominal e o poder foi próximo ao do teste multivariado. No entanto, uma vez que o teste  $F$  com  $\varepsilon$  ajustado é apenas aproximado e não tem nenhuma vantagem sobre o poder do teste multivariado exato, não parece haver nenhuma razão para usá-lo. O único caso em que precisamos voltar a cair num teste univariada é quando os graus de liberdade são insuficientes para executar uma análise multivariada, isto é, quando  $K > \nu$ .

## 2.10 O teste de esfericidade

Para Johnson e Wichern (2007) e Rencher e Christensen (2012) dentre outros, a hipótese de que as variáveis  $y_1, y_2, \dots, y_K$  em  $\mathbf{y}$  são independentes e têm a mesma variância pode ser expressa por

$$H_0 : \Sigma = \sigma^2 \mathbf{I} \quad vs \quad H_1 : \Sigma \neq \sigma^2 \mathbf{I},$$

em que  $\sigma^2$  é uma variância comum desconhecida. Esta hipótese é de interesse em *medidas repetidas*. Sob  $H_0$ , o elipsóide  $(\mathbf{y} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{y} - \boldsymbol{\mu}) = c^2$  reduz-se a  $(\mathbf{y} - \boldsymbol{\mu})' (\mathbf{y} - \boldsymbol{\mu}) = \sigma^2 c^2$ , a equação de uma esfera; portanto, o termo *esfericidade* é aplicado à estrutura de covariância  $\Sigma = \sigma^2 \mathbf{I}$ . Uma outra hipótese de *esfericidade* de interesse em *medidas repetidas* é  $H_0 : \mathbf{C} \Sigma \mathbf{C}' = \sigma^2 \mathbf{I}$ , onde  $\mathbf{C}$  é qualquer matriz de posto linha completo  $(K-1) \times K$  de contrastes ortogonais com comprimento *um*.

Para uma amostra aleatória de vetores de observações  $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n$  de uma  $N_K(\boldsymbol{\mu}, \Sigma)$ , a razão de verossimilhanças para testar  $H_0 : \Sigma = \sigma^2 \mathbf{I}$  é

$$LR = \left[ \frac{|\mathbf{S}|}{(\text{tr} \mathbf{S}/K)^K} \right]^{n/2}. \quad (2.34)$$

em que  $n = IJ$  é o número de observações experimentais para um experimento em blocos

ao acaso com  $I$  níveis do fator  $A$  e  $J$  blocos.

Neste caso, a razão de verossimilhanças é uma função simples de uma estatística de teste como  $F$ ,  $T^2$ ,  $\Lambda$  de Wilks. No entanto, a  $LR$  em (2.34) não se reduz a uma estatística padrão e recorreremos a uma aproximação para sua distribuição. Foi mostrado que para a razão de verossimilhanças generalizada  $LR$ ,

$$-2 \ln(LR) \sim \chi^2_{[\nu]} \quad (2.35)$$

para  $n$  grande, onde  $\nu$  é o número total de parâmetros menos o número estimado sob as restrições imposta por  $H_0$ .

Para a estatística da razão de verossimilhanças em (2.35), obtemos

$$-2 \ln(LR) = -n \ln \left[ \frac{|\mathbf{S}|}{(\text{tr} \mathbf{S}/K)^K} \right] = -n \ln u,$$

onde

$$u = (LR)^{2/n} = \frac{K^K |\mathbf{S}|}{(\text{tr} \mathbf{S}/K)^K}. \quad (2.36)$$

Como  $\text{tr}(\mathbf{S}) = \sum_{k=1}^K \lambda_k$  e  $|\mathbf{S}| = \prod_{k=1}^K \lambda_k$ , assim  $u$  fica

$$u = \frac{K^K \prod_{k=1}^K \lambda_k}{\left( \sum_{k=1}^K \lambda_k \right)^K} \quad (2.37)$$

em que  $\lambda_1, \lambda_2, \dots, \lambda_K$  são os autovalores de  $\mathbf{S}$ . Um aprimoramento a mais para  $-n \ln u$  é dado por

$$u' = - \left( \nu - \frac{2K^2 + K + 2}{6K} \right) \ln u, \quad (2.38)$$

em que  $\nu$  é o número de graus de liberdade de  $\mathbf{S}$ . A estatística  $u'$  tem distribuição aproximada de  $\chi^2$  com  $\frac{1}{2}(K(K+1) - 1)$  graus de liberdade. Rejeitamos  $H_0$ , ao nível de significância  $\alpha$  se, e somente se,  $u' \geq \chi^2_{[\frac{1}{2}K(K+1)-1; \alpha]}$ . Como podemos observar, os graus de liberdade na  $\chi^2$  aproximada é igual ao número total de parâmetros menos o número de parâmetros estimados sob  $H_0$ . O número de parâmetros em  $\Sigma$  é  $K + \binom{K}{2} = \frac{1}{2}K(K+1)$ , além de  $um$  grau de liberdade que é perdido na estimação de  $\sigma^2$ .

Pode-se observar que se os  $\lambda_k$  amostrais forem todos iguais, então pelas expressões em (2.37) e (2.38), terem  $u = 1$  e  $u' = 0$ . Portanto, esta estatística também testa a hipótese de igualdade dos autovalores da população.



Para testar  $H_0 : \mathbf{C}\Sigma\mathbf{C}' = \sigma^2\mathbf{I}$ , use  $\mathbf{C}\mathbf{S}_{pl}\mathbf{C}'$ , com  $\mathbf{S}_{pl} = \mathbf{E}/\nu$ , no lugar de  $\mathbf{S}$  em (2.36) e use  $K - 1$  no lugar de  $K$  nas expressões de (2.36) a (2.38) e nos graus de liberdade da  $\chi^2$ . A razão de verossimilhanças em (2.35) foi primeiro dada por Mauchly (1940), e seu nome está normalmente associado a este teste.

## 2.11 Teste de hipótese sobre a uniformidade de $\Sigma$

Na Seção 2.9, vimos que o teste  $F$  normal da ANOVA é válido quando a estrutura de covariância é do tipo

$$\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \cdots & \rho \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \rho & \cdots & 1 \end{pmatrix} = \sigma^2[(1 - \rho)\mathbf{I} + \rho\mathbf{J}], \quad (2.39)$$

em que  $\mathbf{I}$  é uma matriz identidade e  $\mathbf{J}$  é uma matriz quadrada de  $un$ 's ambas  $K \times K$  e  $\rho$  é a correlação populacional entre duas variáveis quaisquer. Este padrão de variâncias e covariâncias iguais em  $\Sigma$  é conhecido na literatura estatística por *uniformidade, composto simétrico* ou *modelo de correlação intraclasse*.

Consideraremos agora a hipótese

$$H_0 : \Sigma = \Sigma_0 \text{ vs } H_1 : \Sigma \neq \Sigma_0 \quad (2.40)$$

em que

$$\Sigma_0 = \begin{pmatrix} \sigma^2 & \sigma^2\rho & \cdots & \sigma^2\rho \\ \sigma^2\rho & \sigma^2 & \cdots & \sigma^2\rho \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^2\rho & \sigma^2\rho & \cdots & \sigma^2 \end{pmatrix}.$$

A partir de uma amostra, obtêm-se a matriz de covariância amostral  $\mathbf{S}$ . As estimativas de  $\sigma^2$  e  $\sigma^2\rho$  sob  $H_0$  são dadas por

$$s^2 = \frac{1}{K} \sum_{k=1}^K s_{kk} \quad \text{e} \quad s^2r = \frac{1}{K(K-1)} \sum_{k \neq k'} s_{kk'}, \quad (2.41)$$

respectivamente, onde  $s_{kk}$  e  $s_{kk'}$  são elementos de  $\mathbf{S}$ . Assim,  $s^2$  é uma média das variâncias na diagonal de  $\mathbf{S}$  e  $s^2r$  é uma média das covariâncias fora da diagonal de  $\mathbf{S}$ . Uma

estimativa de  $\rho$  pode ser obtida por  $r = s^2 r / s^2$ . Usando  $s^2$  e  $s^2 r$  em (2.41), a estimativa de  $\Sigma$  sob  $H_0$  é obtida como

$$\mathbf{S}_0 = \begin{pmatrix} s^2 & s^2 r & \cdots & s^2 r \\ s^2 r & s^2 & \cdots & s^2 r \\ \vdots & \vdots & \ddots & \vdots \\ s^2 r & s^2 r & \cdots & s^2 \end{pmatrix} = s^2 [(1-r)\mathbf{I} + r\mathbf{J}]. \quad (2.42)$$

Para comparar  $\mathbf{S}$  com  $\mathbf{S}_0$ , usamos a seguinte função da razão de verossimilhanças:

$$u = \frac{|\mathbf{S}|}{|\mathbf{S}_0|}, \quad (2.43)$$

a qual pode ser expressa alternativamente como

$$u = \frac{|\mathbf{S}|}{(s^2)^K (1-r)^{K-1} [1 + (K-1)r]}. \quad (2.44)$$

Por analogia com (2.38), a estatística de teste é dada por

$$u' = - \left[ \nu - \frac{K(K+1)^2(2K-3)}{6(K-1)(K^2+K-4)} \right] \ln u, \quad (2.45)$$

onde  $\nu$  é o número de graus de liberdade associado a  $\mathbf{S}$ . A estatística  $u'$  é aproximadamente distribuída como uma  $\chi^2_{[\frac{1}{2}K(K+1)-2]}$  e rejeitamos  $H_0$ , ao nível de significância  $\alpha$  se, e somente se,  $u' > \chi^2_{[\frac{1}{2}K(K+1)-2; \alpha]}$ . Observe que são perdidos 2 graus de liberdade devido a estimação de  $\sigma^2$  e  $\rho$ .

Um teste aproximado alternativo mais preciso, quando  $K$  é grande e  $\nu$  é relativamente pequeno é dado por

$$F = \frac{-(\nu_2 - \nu_2 c_1 - \nu_1)\nu}{\nu_1 \nu_2} \ln u,$$

em que

$$c_1 = \frac{K(K+1)^2(2K-3)}{6\nu(K-1)(K^2+K-4)}, \quad c_2 = \frac{K(K^2-1)(K+2)}{6\nu^2(K^2+K-4)},$$

$$\nu_1 = \frac{1}{2}K(K+1) - 2, \quad \nu_2 = \frac{\nu_1 + 2}{c_2 - c_1^2}.$$

Rejeitamos  $H_0 : \Sigma = \sigma^2[(1-\rho)\mathbf{I} + \rho\mathbf{J}]$ , ao nível de significância  $\alpha$  se, e somente se,  $F > F_{[\nu_1; \nu_2; \alpha]}$ .

## 2.12 A análise de variância univariada (*Split-Plot*)

Quando as hipóteses  $H_0 : \Sigma = \sigma^2 \mathbf{I}$  ou  $H_0 : \mathbf{C}\Sigma\mathbf{C}' = \sigma^2 \mathbf{I}$  não são rejeitadas, poderemos admitir que as observações são provenientes de um experimento em blocos ao acaso com parcelas sub-divididas no tempo, cujas observações na Tabela 8 estão postas adequadamente na Tabela 9 para facilitar a notação e obtenção dos estimadores.

Tabela 9: Organização das observações de um experimento em blocos ao acaso com medidas repetidas

Tratamento (Fator A)	Bloco (Fator B)	Época de Avaliação (Fator C)						Soma ( $y_{ij}$ )
		1	2	...	$k$	...	$K$	
1	1	$y_{111}$	$y_{112}$	...	$y_{11k}$	...	$y_{11K}$	$y_{11}$
1	2	$y_{121}$	$y_{122}$	...	$y_{12k}$	...	$y_{12K}$	$y_{12}$
⋮	⋮	⋮	⋮		⋮		⋮	⋮
1	$j$	$y_{1j1}$	$y_{1j2}$	...	$y_{1jk}$	...	$y_{1jK}$	$y_{1j}$
⋮	⋮	⋮	⋮		⋮		⋮	⋮
1	$J$	$y_{1J1}$	$y_{1J2}$	...	$y_{1Jk}$	...	$y_{1JK}$	$y_{1J}$
-	Soma	$y_{1.1}$	$y_{1.2}$	...	$y_{1.k}$	...	$y_{1.K}$	$y_{1..}$
2	1	$y_{211}$	$y_{212}$	...	$y_{21k}$	...	$y_{21K}$	$y_{21}$
2	2	$y_{221}$	$y_{222}$	...	$y_{22k}$	...	$y_{22K}$	$y_{22}$
⋮	⋮	⋮	⋮		⋮		⋮	⋮
2	$j$	$y_{2j1}$	$y_{2j2}$	...	$y_{2jk}$	...	$y_{2jK}$	$y_{2j}$
⋮	⋮	⋮	⋮		⋮		⋮	⋮
2	$J$	$y_{2J1}$	$y_{2J2}$	...	$y_{2Jk}$	...	$y_{2JK}$	$y_{2J}$
-	Soma	$y_{2.1}$	$y_{2.2}$	...	$y_{2.k}$	...	$y_{2.K}$	$y_{2..}$
⋮	⋮	⋮	⋮		⋮		⋮	⋮
⋮	⋮	⋮	⋮		⋮		⋮	⋮
$i$	1	$y_{i11}$	$y_{i12}$	...	$y_{i1k}$	...	$y_{i1K}$	$y_{i1}$
$i$	2	$y_{i21}$	$y_{i22}$	...	$y_{i2k}$	...	$y_{i2K}$	$y_{i2}$
⋮	⋮	⋮	⋮		⋮		⋮	⋮
$i$	$j$	$y_{ij1}$	$y_{ij2}$	...	$y_{ijk}$	...	$y_{ijK}$	$y_{ij}$
⋮	⋮	⋮	⋮		⋮		⋮	⋮
$i$	$J$	$y_{iJ1}$	$y_{iJ2}$	...	$y_{iJk}$	...	$y_{iJK}$	$y_{iJ}$
-	Soma	$y_{i.1}$	$y_{i.2}$	...	$y_{i.k}$	...	$y_{i.K}$	$y_{i..}$
⋮	⋮	⋮	⋮		⋮		⋮	⋮

$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
$I$	1	$y_{I11}$	$y_{I12}$	$\cdots$	$y_{I1k}$	$\cdots$	$y_{I1K}$	$y_{I1.}$
$I$	2	$y_{I21}$	$y_{I22}$	$\cdots$	$y_{I2k}$	$\cdots$	$y_{I2K}$	$y_{I2.}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$I$	$j$	$y_{Ij1}$	$y_{Ij2}$	$\cdots$	$y_{Ijk}$	$\cdots$	$y_{IjK}$	$y_{Ij.}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$I$	$J$	$y_{IJ1}$	$y_{IJ2}$	$\cdots$	$y_{IJk}$	$\cdots$	$y_{IJK}$	$y_{IJ.}$
-	Soma	$y_{I.1}$	$y_{I.2}$	$\cdots$	$y_{I.k}$	$\cdots$	$y_{I.K}$	$y_{I..}$
-	Soma	$y_{.1}$	$y_{.2}$	$\cdots$	$y_{.k}$	$\cdots$	$y_{.K}$	$y_{...}$

Empregando o método dos mínimos quadrados e impondo convenientemente as restrições habituais na solução do sistema de equações normais:  $\sum_{i=1}^I \hat{\tau}_i = 0$ ,  $\sum_{j=1}^J \hat{\beta}_j = 0$ ,  $\sum_{k=1}^K \hat{\theta}_k = 0$ ,  $\sum_{i=1}^I \hat{\gamma}_{ik} = \sum_{k=1}^K \hat{\gamma}_{ik} = \sum_{i=1}^I \sum_{k=1}^K \hat{\gamma}_{ik} = 0$ , encontramos os estimadores dos parâmetros do modelo em (2.29), os quais encontram-se disponíveis na Tabela 10 a seguir:

Tabela 10: Os parâmetros do modelo e seus estimadores

Características na população	Estimador da característica
$\mu$	$\hat{\mu} = \bar{y}_{...}$
$\tau_i$	$\hat{\tau}_i = \bar{y}_{i..} - \bar{y}_{...}$
$\mu_i$	$\hat{\mu}_i = \bar{y}_{i..}$
$\beta_j$	$\hat{\beta}_j = \bar{y}_{.j.} - \bar{y}_{...}$
$\mu_j$	$\hat{\mu}_j = \bar{y}_{.j.}$
$\epsilon_{ij}$	$\hat{\epsilon}_{ij} = \bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}$
$\theta_k$	$\hat{\theta}_k = \bar{y}_{..k} - \bar{y}_{...}$
$\mu_k$	$\hat{\mu}_k = \bar{y}_{..k}$
$\gamma_{ik}$	$\hat{\gamma}_{ik} = \bar{y}_{i.k} - \bar{y}_{i..} - \bar{y}_{..k} + \bar{y}_{...}$
$\mu_{ik}$	$\hat{\mu}_{ik} = \bar{y}_{i.k}$
$y_{ijk}$	$\hat{y}_{ijk} = \hat{\mu} + \hat{\tau}_i + \hat{\beta}_j + \hat{\epsilon}_{ij} + \hat{\theta}_k + \hat{\gamma}_{ik}$
$\epsilon_{ijk}$	$\hat{\epsilon}_{ijk} = y_{ijk} - \hat{y}_{ijk}$

Veremos na Seção 2.12.1 como se decompõe a variabilidade total das observações experimentais a partir dos estimadores.

### 2.12.1 A decomposição da variabilidade total

As observações na Tabela 8 podem ser reescritas a partir dos estimadores dos parâmetros disponíveis na Tabela 10, como

$$y_{ijk} = \hat{\mu} + \hat{\tau}_i + \hat{\beta}_j + \hat{\epsilon}_{ij} + \hat{\theta}_k + \hat{\gamma}_{ik} + \hat{\epsilon}_{ijk}$$

e segue que

$$\begin{aligned} \hat{\epsilon}_{ijk} &= y_{ijk} - \hat{\mu} - \hat{\tau}_i - \hat{\beta}_j - \hat{\epsilon}_{ij} - \hat{\theta}_k - \hat{\gamma}_{ik} \\ &= y_{ijk} - \bar{y}_{...} - (\bar{y}_{i..} - \bar{y}_{...}) - (\bar{y}_{.j.} - \bar{y}_{...}) - (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}) - (\bar{y}_{..k} - \bar{y}_{...}) - \\ &\quad (\bar{y}_{i.k} - \bar{y}_{i..} - \bar{y}_{..k} + \bar{y}_{...}) \end{aligned}$$

Elevando-se ao quadrado os dois lados da expressão acima e somando-se para todas as observações, encontra-se

$$\begin{aligned} \underbrace{\sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K \hat{\epsilon}_{ijk}^2}_{SQ_{Res(b)}} &= \underbrace{\sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (y_{ijk} - \bar{y}_{...})^2}_{SQ_{Total}} - \underbrace{JK \sum_{i=1}^I (\bar{y}_{i..} - \bar{y}_{...})^2}_{SQ_A} - \underbrace{IK \sum_{j=1}^J (\bar{y}_{.j.} - \bar{y}_{...})^2}_{SQ_B} \\ &\quad - \left[ \underbrace{K \sum_{i=1}^I \sum_{j=1}^J (\bar{y}_{ij.} - \bar{y}_{...})^2}_{SQ_{A,B}} - \underbrace{JK \sum_{i=1}^I (\bar{y}_{i..} - \bar{y}_{...})^2}_{SQ_A} - \underbrace{IK \sum_{j=1}^J (\bar{y}_{.j.} - \bar{y}_{...})^2}_{SQ_B} \right] \\ &\quad - \underbrace{IJ \sum_{k=1}^K (\bar{y}_{..k} - \bar{y}_{...})^2}_{SQ_C} - \left[ \underbrace{J \sum_{i=1}^I \sum_{k=1}^K (\bar{y}_{i.k} - \bar{y}_{...})^2}_{SQ_{A,C}} - \underbrace{JK \sum_{i=1}^I (\bar{y}_{i..} - \bar{y}_{...})^2}_{SQ_A} \right] \\ &\quad - \underbrace{IJ \sum_{k=1}^K (\bar{y}_{..k} - \bar{y}_{...})^2}_{SQ_C} \end{aligned}$$

Uma forma equivalente e mais habitual de expressar estas quantidades é como segue

$$\underbrace{\sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K \hat{\epsilon}_{ijk}^2}_{SQ_{Res(b)}} = \underbrace{\left( \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K y_{ijk}^2 - C \right)}_{SQ_{Total}} - \underbrace{\left( \frac{1}{JK} \sum_{i=1}^I y_{i..}^2 - C \right)}_{SQ_A} - \underbrace{\left( \frac{1}{IK} \sum_{j=1}^J y_{.j.}^2 - C \right)}_{SQ_B}$$

$$\begin{aligned}
& - \left[ \underbrace{\left( \frac{1}{K} \sum_{i=1}^I \sum_{j=1}^J y_{ij.}^2 - C \right)}_{SQ_{A,B}} - \underbrace{\left( \frac{1}{JK} \sum_{i=1}^I y_{i..}^2 - C \right)}_{SQ_A} - \underbrace{\left( \frac{1}{IK} \sum_{j=1}^J y_{.j.}^2 - C \right)}_{SQ_B} \right] \\
& \underbrace{\hspace{10em}}_{SQ_{Res(a)}} \\
& - \underbrace{\left( \frac{1}{IJ} \sum_{k=1}^K y_{..k}^2 - C \right)}_{SQ_C} - \left[ \underbrace{\left( \frac{1}{J} \sum_{i=1}^I \sum_{k=1}^K y_{i.k}^2 - C \right)}_{SQ_{A,C}} - \underbrace{\left( \frac{1}{JK} \sum_{i=1}^I y_{i..}^2 - C \right)}_{SQ_A} \right] \\
& - \underbrace{\left( \frac{1}{IJ} \sum_{k=1}^K y_{..K}^2 - C \right)}_{SQ_C} \Big]
\end{aligned}$$

O último termo entre colchetes, nas duas expressões acima, corresponde à soma de quadrados da interação  $A \times C$ . Isto é,

$$\begin{aligned}
SQ_{A \times C} &= \underbrace{J \sum_{i=1}^I \sum_{k=1}^K (\bar{y}_{i.k} - \bar{y}_{..})^2}_{SQ_{A,C}} - \underbrace{JK \sum_{i=1}^I (\bar{y}_{i..} - \bar{y}_{..})^2}_{SQ_A} - \underbrace{IJ \sum_{k=1}^K (\bar{y}_{..k} - \bar{y}_{..})^2}_{SQ_C} \\
&= \underbrace{\left( \frac{1}{J} \sum_{i=1}^I \sum_{k=1}^K y_{i.k}^2 - C \right)}_{SQ_{A,C}} - \underbrace{\left( \frac{1}{JK} \sum_{i=1}^I y_{i..}^2 - C \right)}_{SQ_A} - \underbrace{\left( \frac{1}{IJ} \sum_{k=1}^K y_{..k}^2 - C \right)}_{SQ_C}
\end{aligned}$$

Finalmente, podemos observar que a variabilidade total é decomposta como

$$SQ_{Total} = SQ_A + SQ_B + SQ_{Res(a)} + SQ_C + SQ_{A \times C} + SQ_{Res(b)}.$$

Em palavras, a variabilidade total se decompõe em *seis* partes: a variabilidade devida ao Fator  $A$  mais a variabilidade devida ao fator Bloco, mais a variabilidade entre-sujeitos, mais a variabilidade devida ao fator  $C$ , mais a variabilidade devida a interação  $A \times C$ , mais a variabilidade intra-sujeitos.

## 2.12.2 Distribuições de probabilidade dos estimadores, dos parâmetros e das somas de quadrados

Para se fazer as inferências habituais sobre os parâmetros e combinações lineares deles deveremos associar distribuições de probabilidade aos estimadores, bem como às somas de quadrados. Para tanto, precisamos descobrir as características (média e variância) e as respectivas distribuições de probabilidade correspondentes. A Tabela 11 exhibe os resultados encontrados.

Tabela 11: Distribuição de probabilidade dos estimadores dos parâmetros e das somas de quadrados sob  $H_0^{(\cdot)}$ 

Estimador do parâmetro e sua distribuição de probabilidade	Hipótese nula $H_0^{(\cdot)}$	Distribuição da soma de quadrados sob $H_0^{(\cdot)}$
$\begin{cases} \hat{\tau}_i = \bar{y}_{i..} - \bar{y}_{...} \\ \hat{\tau}_i \stackrel{id}{\sim} N\left(\tau_i, \frac{(I-1)\sigma^2 + K\sigma_a^2}{IJK}\right) \end{cases}$	$\begin{cases} H_0^{(\mu_i)} : \mu_i = \mu, \forall i \\ ou \\ H_0^{(\tau_i)} : \tau_i = 0, \forall i \end{cases}$	$\frac{SQ_A}{\sigma^2 + K\sigma_a^2} \sim \chi_{(I-1)}^2$
$\begin{cases} \hat{\beta}_j = \bar{y}_{.j.} - \bar{y}_{...} \\ \hat{\beta}_j \stackrel{id}{\sim} N\left(\beta_j, \frac{(J-1)\sigma^2 + K\sigma_a^2}{IK}\right) \end{cases}$	$\begin{cases} H_0^{(\mu_j)} : \mu_j = \mu, \forall j \\ ou \\ H_0^{(\beta_j)} : \beta_j = 0, \forall j \end{cases}$	$\frac{SQ_B}{\sigma^2 + K\sigma_a^2} \sim \chi_{(J-1)}^2$
$\begin{cases} \hat{\epsilon}_{ij} = \bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...} \\ \hat{\epsilon}_{ij} \stackrel{iid}{\sim} N\left(0, \frac{(I-1)(J-1)\sigma^2 + K\sigma_a^2}{IK}\right) \end{cases}$	Independente de $H_0^{(\cdot)}$	$\frac{SQ_{Res(a)}}{\sigma^2 + K\sigma_a^2} \sim \chi_{[(I-1)(J-1)]}^2$
$\begin{cases} \hat{\theta}_k = \bar{y}_{..k} - \bar{y}_{...} \\ \hat{\theta}_k \stackrel{id}{\sim} N\left(\theta_k, \frac{(K-1)\sigma^2}{IJ}\right) \end{cases}$	$\begin{cases} H_0^{(\mu_k)} : \mu_k = \mu, \forall k \\ ou \\ H_0^{(\theta_k)} : \theta_k = 0, \forall k \end{cases}$	$\frac{SQ_C}{\sigma^2} \sim \chi_{(K-1)}^2$
$\begin{cases} \hat{\gamma}_{ik} = \bar{y}_{i.k} - \bar{y}_{i..} - \bar{y}_{.k.} + \bar{y}_{...} \\ \hat{\gamma}_{ik} \stackrel{id}{\sim} N\left(\gamma_{ik}, \frac{(I-1)(K-1)\sigma^2}{IKJ}\right) \end{cases}$	$\begin{cases} H_0^{(\mu_{ik})} : \mu_{ik} = \mu, \forall i, k \\ ou \\ H_0^{(\gamma_{ik})} : \gamma_{ik} = 0, \forall i, k \end{cases}$	$\frac{SQ_{A \times C}}{\sigma^2} \sim \chi_{[(I-1)(K-1)]}^2$
$\begin{cases} \hat{\epsilon}_{ijk} = y_{ijk} - \hat{y}_{ijk} \\ \hat{\epsilon}_{ijk} \stackrel{iid}{\sim} N(0, \sigma^2) \end{cases}$	Independente de $H_0^{(\cdot)}$	$\frac{SQ_{Res(b)}}{\sigma^2} \sim \chi_{[I(J-1)(K-1)]}^2$

em que  $i = 1, 2, \dots, I$ ,  $j = 1, 2, \dots, J$ ,  $k = 1, 2, \dots, K$ ,  $\sigma_a^2 = Var(\epsilon_{ij})$  é a variância do erro experimental a nível de parcela,  $\sigma^2 = Var(\epsilon_{ijk})$  é a variância do erro experimental a nível de sub-parcela. Os termos *id*, indica que o estimador é independentemente distribuído e *iid*, que é independente e identicamente distribuído.

## 2.13 Valores esperados dos quadrados médios para o modelo *split – plot*

Outros resultados importantes na definição das estatísticas dos testes associadas às hipóteses sobre a não existência de efeitos dos tratamentos, dos bolcos, do fator medida repetida (fator C) e interação  $A \times C$  são os valores esperados dos quadrados médios. Estes resultados foram obtidos aplicando-se a definição de esperança matemática conforme Barbin (1993, p. 87) e estão apresentados na Tabela 12.

Tabela 12: Análise da variância com os valores esperados dos quadrados médios

F. Variação	$GL$	$SQ$	$QM$	$E(QM)$
Fator A	$\nu_t$	$SQ_A$	$V_1 = QM_A$	$\sigma^2 + K\sigma_a^2 + \frac{JK}{\nu_t} \sum_{i=1}^I \tau_i^2$
Bloco	$\nu_b$	$SQ_B$	$V_2 = QM_B$	$\sigma^2 + K\sigma_a^2 + \frac{IK}{\nu_b} \sum_{j=1}^J \beta_j^2$
Resíduo(a)	$\nu_a$	$SQ_{Res(a)}$	$V_3 = QM_{Res(a)}$	$\sigma^2 + K\sigma_a^2$
Fator C	$\nu_c$	$SQ_C$	$V_4 = QM_C$	$\sigma^2 + \frac{IJ}{\nu_c} \sum_{k=1}^K \theta_k^2$
Interação $A \times C$	$\nu_g$	$SQ_{A \times C}$	$V_5 = QM_{A \times C}$	$\sigma^2 + \frac{J}{\nu_g} \sum_{i=1}^I \sum_{k=1}^K \gamma_{ik}^2$
Resíduo(b)	$\nu$	$SQ_{Res(b)}$	$V_6 = QM_{Res(b)}$	$\sigma^2$
Total				

em que os graus de liberdade são denotados como,

$$\begin{aligned} \nu_t &= (I - 1), & \nu_b &= (J - 1), & \nu_a &= (I - 1)(J - 1), \\ \nu_c &= (K - 1), & \nu_g &= (I - 1)(K - 1), & \nu &= I(J - 1)(K - 1). \end{aligned}$$

Também, visto que os efeitos são expressos como desvios em relação a média geral, tem-se,  $\sum_{i=1}^I \tau_i^2 = \sum_{i=1}^I (\mu_i - \mu)^2$ ,  $\sum_{j=1}^J \beta_j^2 = \sum_{j=1}^J (\mu_j - \mu)^2$ ,  $\sum_{k=1}^K \theta_k^2 = \sum_{k=1}^K (\mu_k - \mu)^2$  e  $\sum_{i=1}^I \sum_{k=1}^K \gamma_{ik}^2 = \sum_{i=1}^I \sum_{k=1}^K (\mu_{ik} - \mu)^2$ . Estas expressões ajudam-nos a compreender a equivalência entre as hipóteses quando são formuladas em termos dos efeitos dos níveis dos fatores e as hipóteses formuladas em termos das médias dos níveis dos fatores.

### 2.13.1 As estatísticas dos testes no modelo *Split-plot*

No modelo *split-plot* temos interesse em testar as seguintes hipóteses

$$a) \left( \begin{array}{l} \text{Hipótese sobre as} \\ \text{médias do fator A} \end{array} \right) \longrightarrow \left\{ \begin{array}{l} H_0^{(\mu_i)} : \mu_i = \mu, \forall i \\ vs \\ H_1^{(\mu_i)} : \mu_i \neq \mu, \text{ para pelo menos um } i, \end{array} \right. \quad (2.46)$$

$$b) \left( \begin{array}{l} \text{Hipótese sobre as} \\ \text{médias dos blocos} \end{array} \right) \longrightarrow \left\{ \begin{array}{l} H_0^{(\mu_j)} : \mu_j = \mu, \forall j \\ vs \\ H_1^{(\mu_j)} : \mu_j \neq \mu, \text{ para pelo menos um } j, \end{array} \right. \quad (2.47)$$

$$c) \left( \begin{array}{l} \text{Hipótese sobre as} \\ \text{médias do Fator C} \end{array} \right) \longrightarrow \left\{ \begin{array}{l} H_0^{(\mu_k)} : \mu_k = \mu, \forall k \\ vs \\ H_1^{(\mu_k)} : \mu_k \neq \mu, \text{ para pelo menos um } k, \end{array} \right. \quad (2.48)$$



$$d) \left( \begin{array}{c} \text{Hipótese sobre as mé-} \\ \text{dias da interação } A \times C \end{array} \right) \longrightarrow \begin{cases} H_0^{(\mu_{ik})} : \mu_{ik} = \mu, \forall i, k \\ \nu_s \\ H_1^{(\mu_{ik})} : \mu_{ik} \neq \mu, \text{ para pelo menos um } i, k \end{cases} \quad (2.49)$$

Observando estas hipóteses, os resultados da Tabela 12 para os valores esperados dos quadrados médios e as distribuições de probabilidade das somas de quadrados na Tabela 11 podemos, por analogia com (2.18) e (2.19), mostrar que as expressões

$$F_A = \frac{QM_A}{QM_{Res(a)}} \sim F_{(\nu_t, \nu_a)}, \quad F_B = \frac{QM_B}{QM_{Res(a)}} \sim F_{(\nu_b, \nu_a)}, \quad (2.50)$$

$$F_C = \frac{QM_C}{QM_{Res(b)}} \sim F_{(\nu_c, \nu)} \quad e \quad F_{A \times C} = \frac{QM_{A \times C}}{QM_{Res(b)}} \sim F_{(\nu_g, \nu)}$$

são, respectivamente, as estatísticas de teste de  $H_0^{(\mu_i)}$ ,  $H_0^{(\mu_j)}$ ,  $H_0^{(\mu_k)}$  e  $H_0^{(\mu_{ik})}$ . Portanto, rejeitamos estas hipóteses, ao nível de significância  $\alpha$ , se para os dados de um experimento em particular, observarmos  $F_A > F_{(\nu_t, \nu_a, \alpha)}$ ,  $F_B > F_{(\nu_b, \nu_a, \alpha)}$ ,  $F_C > F_{(\nu_c, \nu, \alpha)}$  e  $F_{A \times C} > F_{(\nu_g, \nu, \alpha)}$ , respectivamente. A partir destes resultados podemos estruturar a tabela da ANOVA para o modelo *Split-plot*. ou seja,

Tabela 13: ANOVA univariada para o modelo *Split-plot* baseada nos dados da Tabela 8

F. Variação	$GL$	$SQ$	$QM$	$F$
Fator $A$	$\nu_t$	$SQ_A$	$QM_A$	$F_A = QM_A/QM_{Res(a)}$
Fator $B$	$\nu_b$	$SQ_B$	$QM_B$	$F_B = QM_B/QM_{Res(a)}$
Resíduo(a)	$\nu_a$	$SQ_{Res(a)}$	$QM_{Res(a)}$	
Parcela (A,B)	$(IJ - 1)$	$(SQ_{A,B})$		
Fator $C$	$\nu_c$	$SQ_C$	$QM_C$	$F_C = QM_C/QM_{Res(b)}$
Interação $A \times C$	$\nu_g$	$SQ_{A \times C}$	$QM_{A \times C}$	$F_{A \times C} = QM_{A \times C}/QM_{Res(b)}$
Resíduo(b)	$\nu$	$SQ_{Res(b)}$	$QM_{Res(b)}$	
Total	$IJK - 1$	$SQ_{Total}$		

Quando uma (ou mais de uma) hipótese é rejeitada, interessa-nos averiguar os efeitos que levaram à rejeição da(s) mesma(s). Para isto, podemos aplicar um teste de comparações múltiplas, por exemplo o teste de Tukey, para comparar as médias duas-a-duas. Neste sentido, expomos a seguir, os procedimentos para se fazer estes quatro tipos de comparações (CAMPOS, 1984, p. 151-165) e (BARBIN, 1993, p. 85-91):

1. Entre dois níveis do fator  $A$ :

- (a) Fixar  $\alpha = 0,05$  e estabelecer a hipótese nula  $H_0^{(1)} : \mu_i - \mu_{i'} = 0, \forall i \neq i', i, i' = 1, 2, \dots, I,$

- (b) Calcular as estimativas das possíveis diferenças,  $\hat{\psi}_h = \bar{y}_i - \bar{y}_{i'}$ ,  $i \neq i'$ ,  $h = 1, 2, \dots, I(I-1)/2$ ,
- (c) calcular a diferença mínima significativa,  $\Delta = q_{(I;\nu_a;0,05)} \sqrt{\frac{QM_{Res(a)}}{JK}}$ ,
- (d) Usar a regra de decisão: rejeitar  $H_0^{(1)}$ , se  $\hat{\psi}_h > \Delta$  e concluir, ao nível de significância  $\alpha = 0,05$ , que  $\mu_i$  difere de  $\mu_{i'}$ .

2. Entre dois níveis do fator  $C$

- (a) Fixar  $\alpha = 0,05$  e estabelecer a hipótese nula  $H_0^{(2)} : \mu_{.k} - \mu_{.k'} = 0$ ,  $\forall k \neq k'$ ,  $k, k' = 1, 2, \dots, K$ ,
- (b) Calcular as estimativas das possíveis diferenças,  $\hat{\psi}_h = \bar{y}_{.k} - \bar{y}_{.k'}$ ,  $k \neq k'$ ,  $h = k, 2, \dots, K(K-1)/2$ ,
- (c) calcular a diferença mínima significativa,  $\Delta = q_{(K;\nu_b;0,05)} \sqrt{\frac{QM_{Res(b)}}{IJ}}$ ,
- (d) Usar a regra de decisão: rejeitar  $H_0^{(2)}$ , se  $\hat{\psi}_h > \Delta$  e concluir, ao nível de significância  $\alpha = 0,05$ , que  $\mu_{.k}$  difere de  $\mu_{.k'}$ .

3. Entre dois níveis do fator  $C$  num mesmo nível do fator  $A$

- (a) Fixar  $\alpha = 0,05$  e estabelecer a hipótese nula  $H_0^{(3)} : \mu_{ik} - \mu_{ik'} = 0$ ,  $\forall k \neq k'$ ,  $k, k' = 1, 2, \dots, K$  e  $i = 1, 2, \dots, I$
- (b) Calcular as estimativas das possíveis diferenças,  $\hat{\psi}_h = \bar{y}_{ik} - \bar{y}_{ik'}$ ,  $k \neq k'$ ,  $h = 1, 2, \dots, K(K-1)/2$  e  $i = 1, 2, \dots, I$
- (c) calcular a diferença mínima significativa,  $\Delta = q_{(K;\nu_b;0,05)} \sqrt{\frac{QM_{Res(b)}}{J}}$ ,
- (d) Usar a regra de decisão: rejeitar  $H_0^{(3)}$ , se  $\hat{\psi}_h > \Delta$  e concluir, ao nível de significância  $\alpha = 0,05$ , que  $\mu_{ik}$  difere de  $\mu_{ik'}$ .

4. Entre dois níveis do fator  $A$  num mesmo nível do fator  $C$

- (a) Fixar  $\alpha = 0,05$  e estabelecer a hipótese nula  $H_0^{(4)} : \mu_{ik} - \mu_{i'k} = 0$ ,  $\forall i \neq i'$ ,  $i, i' = 1, 2, \dots, I$  e  $k = 1, 2, \dots, K$
- (b) Calcular as estimativas das possíveis diferenças,  $\hat{\psi}_h = \bar{y}_{ik} - \bar{y}_{i'k}$ ,  $i \neq i'$ ,  $h = 1, 2, \dots, I(I-1)/2$  e  $k = 1, 2, \dots, K$
- (c) calcular a diferença mínima significativa,  $\Delta = q_{(I;\nu';0,05)} \sqrt{\frac{QM_{Res}}{J}}$ , em que

$$QM_{Res} = \frac{QM_{Res(a)} + (K-1)QM_{Res(b)}}{K} \quad \text{e} \quad (2.51)$$

com graus de liberdade associados

$$\nu' = \frac{[QM_{Res(a)} + (K-1)QM_{Res(b)}]^2}{\frac{[QM_{Res(a)}]^2}{\nu_t} + \frac{[(K-1)QM_{Res(b)}]^2}{\nu}}, \quad (2.52)$$

conforme (SATTERTHWAITE, 1946 apud CAMPOS, 1984, p. 164).

- (d) Usar a regra de decisão: rejeitar  $H_0^{(4)}$ , se  $\hat{\psi}_h > \Delta$  e concluir, ao nível de significância  $\alpha = 0,05$ , que  $\mu_{ik}$  difere de  $\mu_{i'k}$ .

Quando a hipótese da não existência de interação em (2.49) não é rejeitada, significa que o comportamento da variável resposta em relação aos níveis do fator  $C$ , é o mesmo para cada nível do fator  $A$  e vice-versa. Ao contrário, quando a hipótese de interação é rejeitada, significa que o comportamento da variável resposta em relação aos níveis de um fator é diferente em pelo menos um nível do outro fator. Isto sugere que as comparações entre as médias dos níveis de um fator dentro de cada nível do outro, só devem ser feitas, quando a interação for significativa e as comparações entre as médias dos níveis dos fatores  $A$  e  $C$  individualmente devem ser feitas quando a interação não for significativa.

Em muitas situações práticas, os níveis do fator  $C$  são épocas de avaliação e interessa ao pesquisador estudar o comportamento da variável resposta em relação ao tempo (fator  $C$ ). Nestes casos, sugerimos estudar o comportamento da variável resposta em relação ao tempo através dos polinômios ortogonais. Se a hipótese de interação  $H_0^{(\mu_{ik})} : \mu_{ik} = \mu$  em (2.49) não for rejeitada, deveremos proceder ao desdobramento dos graus de liberdade e soma de quadrados do fator  $C$  em componentes polinomiais de grau  $um$  a grau  $(K-1)$  semelhante ao que foi feito na subseção 2.8.2. Assim, teremos

$$SQ(C_g) = SQ(Comp. de g^o grau Fator C) = \frac{\left(\sum_{k=1}^K c_{gk} y_{..k}\right)^2}{IJK_g}, \quad (2.53)$$

com  $g = 1, 2, \dots, K-1$ .

Para cada componente, teremos

$$F_{C_g} = \frac{QM(C_g)}{QM_{Res(b)}} \sim F_{(1, \nu_b)}$$

que será a estatística de teste de  $H_0^{(\theta_g)} : \theta_g = 0$ ,  $g = 1, 2, \dots, (K-1)$ , e segue que, rejeitamos  $H_0^{(\theta_g)}$ , ao nível de significância  $\alpha$  se, observarmos para um particular experimento, que  $F_{C_g} > F_{(1, \nu_b, \alpha)}$ . Assim sendo, a reestruturação da Tabela da ANOVA contemplando os componentes polinomiais para o fator medidas repetidas (o fator  $C$ ), fica:

Tabela 14: ANOVA univariada para o modelo *Split-plot* para testar os componentes polinomiais para o fator  $C$ .

F. Variação	$GL$	$SQ$	$QM$	$F$
Fator $A$	$\nu_t$	$SQ_A$	$QM_A$	$F_A = QM_A/QM_{Res(a)}$
Fator $B$	$\nu_b$	$SQ_B$	$QM_B$	$F_B = QM_B/QM_{Res(a)}$
Resíduo(a)	$\nu_a$	$SQ_{Res(a)}$	$QM_{Res(a)}$	
Parcela (A,B)	$(IJ - 1)$	$(SQ_{A,B})$	-	-
$C_1$	1	$SQ(C_1)$	$QM(C_1)$	$QM(C_1)/QM_{Res(b)}$
$C_2$	1	$SQ(C_1)$	$QM(C_1)$	$QM(C_2)/QM_{Res(b)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	
$C_{\nu_c}$	1	$SQ(C_{\nu_c})$	$QM(C_{\nu_c})$	$QM(C_{\nu_c})/QM_{Res(b)}$
Fator $C$	$\nu_c$	$SQ_C$	$QM_C$	$F_C = QM_C/QM_{Res(b)}$
Interação $A \times C$	$\nu_g$	$SQ_{A \times C}$	$QM_{A \times C}$	$F_{A \times C} = QM_{A \times C}/QM_{Res(b)}$
Resíduo(b)	$\nu$	$SQ_{Res(b)}$	$QM_{Res(b)}$	-
Total	$IJK - 1$	$SQ_{Total}$	-	-

Para estimar o polinômio de grau  $g$  que descreve o comportamento médio da variável resposta em relação ao tempo (as  $K$  medidas repetidas), use as expressões (2.26) e (2.28) e substituir (2.27) por

$$\hat{\mu} = \bar{y}_{...}, \quad T_1 = \frac{\sum_{k=1}^K c_{1k} y_{..k}}{IJK_1}, \quad T_2 = \frac{\sum_{k=1}^K c_{2k} y_{..k}}{IJK_2}, \quad \dots, \quad T_g = \frac{\sum_{k=1}^K c_{gk} y_{..k}}{IJK_g}.$$

No entanto, se a interação for significativa, deveremos estudar o comportamento da variável resposta em relação ao tempo dentro de cada nível do fator  $A$ . Para tal, calculamos

$$SQ(C_g \text{ d. } A_i) = SQ(\text{Comp. de } g^{\text{O}} \text{ grau para } C \text{ d. } A_i) = \frac{\left( \sum_{k=1}^K c_{gk} y_{i..k} \right)^2}{JK_g}, \quad (2.54)$$

com  $g = 1, 2, \dots, K - 1$  para cada  $i = 1, 2, \dots, I$ , que sob  $H_0^{(\theta_g \text{ d. } A_i)} : \theta_g = 0$ , a estatística  $F_{C_g \text{ d. } A_i} = \frac{QM(C_g \text{ d. } A_i)}{QM_{Res(b)}} \sim F_{(1, \nu)}$  e, segue que,  $H_0^{(\theta_g \text{ d. } A_i)} : \theta_g = 0$  será rejeitada, ao nível de significância  $\alpha$  se observarmos, para um experimento em particular, que  $F_{C_g \text{ d. } A_i} = \frac{QM(C_g \text{ d. } A_i)}{QM_{Res(b)}} < F_{(1, \nu, \alpha)}$ . Uma sugestão para a organização desse tipo de desdobramento é apresentada na Tabela 15.

Para estimar o polinômio de grau  $g$  que descreve o comportamento médio da variável resposta em relação ao tempo (as  $K$  medidas repetidas) dentro de cada nível do fator  $A$ ,

Tabela 15: ANOVA para o desdobramento dos graus de liberdade e somas de quadrados em componentes polinomiais para o fator medidas repetidas dentro de cada nível de  $A$

F. Variação	$GL$	$SQ$	$QM$	$F$
$C_1 d. A_1$	1	$SQ(C_1 d. A_1)$	$QM(C_1 d. A_1)$	$F_{C_1 d. A_1} = QM(C_1 d. A_1)/QM_{Res(b)}$
$C_2 d. A_1$	1	$SQ(C_2 d. A_1)$	$QM(C_2 d. A_1)$	$F_{C_2 d. A_1} = QM(C_2 d. A_1)/QM_{Res(b)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$C_{\nu_c} d. A_1$	1	$SQ(C_{\nu_c} d. A_1)$	$QM(C_{\nu_c} d. A_1)$	$F_{C_{\nu_c} d. A_1} = QM(C_{\nu_c} d. A_1)/QM_{Res(b)}$
Fator $C d. A_1$	$(\nu_c)$	$SQ(C d. A_1)$	$QM(C d. A_1)$	$F_C d. A_1 = QM(C d. A_1)/QM_{Res(b)}$
$C_1 d. A_2$	1	$SQ(C_1 d. A_2)$	$QM(C_1 d. A_2)$	$F_{C_1 d. A_2} = QM(C_1 d. A_2)/QM_{Res(b)}$
$C_2 d. A_2$	1	$SQ(C_2 d. A_2)$	$QM(C_2 d. A_2)$	$F_{C_2 d. A_2} = QM(C_2 d. A_2)/QM_{Res(b)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$C_{\nu_c} d. A_2$	1	$SQ(C_{\nu_c} d. A_2)$	$QM(C_{\nu_c} d. A_2)$	$F_{C_{\nu_c} d. A_2} = QM(C_{\nu_c} d. A_2)/QM_{Res(b)}$
Fator $C d. A_2$	$(\nu_c)$	$SQ(C d. A_2)$	$QM(C d. A_2)$	$F_C d. A_2 = QM(C d. A_2)/QM_{Res(b)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$C_1 d. A_I$	1	$SQ(C_1 d. A_I)$	$QM(C_1 d. A_I)$	$F_{C_1 d. A_I} = QM(C_1 d. A_I)/QM_{Res(b)}$
$C_2 d. A_I$	1	$SQ(C_2 d. A_I)$	$QM(C_2 d. A_I)$	$F_{C_2 d. A_I} = QM(C_2 d. A_I)/QM_{Res(b)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$C_{\nu_c} d. A_I$	1	$SQ(C_{\nu_c} d. A_I)$	$QM(C_{\nu_c} d. A_I)$	$F_{C_{\nu_c} d. A_I} = QM(C_{\nu_c} d. A_I)/QM_{Res(b)}$
Fator $C d. A_I$	$(\nu_c)$	$SQ(C d. A_I)$	$QM(C d. A_I)$	$F_C d. A_I = QM(C d. A_I)/QM_{Res(b)}$
Reíduo(b)	$\nu$	$SQ_{Res(b)}$	$QM_{Res(b)}$	-

em que  $SQ(C d. A_i) = \frac{1}{j} \sum_{k=1}^K y_{i.k}^2 - \frac{(y_{i..})^2}{JK}$ , para  $i = 1, 2, \dots, I$ .

basta usar as expressões (2.26) e (2.28) e substituir (2.27) por

$$\hat{\mu}_i = \bar{y}_{i..}, \quad T_1 = \frac{\sum_{k=1}^K c_{1k} y_{i.k}}{JK_1}, \quad T_2 = \frac{\sum_{k=1}^K c_{2k} y_{i.k}}{JK_2}, \quad \dots, \quad T_g = \frac{\sum_{k=1}^K c_{gk} y_{i.k}}{JK_g}, \quad \text{com } i = 1, 2, \dots, I.$$

De modo análogo, o desdobramento dos graus de liberdade e somas de quadrados do fator  $A$  dentro de cada nível do fator  $C$  em contrastes de interesse ou componentes polinomiais também podem ser feitos adaptando-se convenientemente as somas de quadrados e graus de liberdade (veja por exemplo Campos (1984, p. 158)).

## 2.14 O Modelo multivariado

Na Seção 2.9 foi dito que a análise multivariada dos dados de um experimento em blocos ao acaso, permite que as medidas repetidas sejam correlacionadas em qualquer padrão. Quando for o caso, o modelo estatístico para descrever as observações é definido como

$$\mathbf{y}_{ij} = \boldsymbol{\mu} + \boldsymbol{\tau}_i + \boldsymbol{\beta}_j + \boldsymbol{\epsilon}_{ij} \begin{cases} i = 1, 2, \dots, I \\ j = 1, 2, \dots, J \end{cases} \quad (2.55)$$

Em cada vetor de observação  $\mathbf{y}_{ij}$  temos  $K$  variáveis, de modo que o modelo em ?? fica

$$\begin{pmatrix} y_{ij1} \\ y_{ij2} \\ \vdots \\ y_{ijK} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_K \end{pmatrix} + \begin{pmatrix} \tau_{i1} \\ \tau_{i2} \\ \vdots \\ \tau_{iK} \end{pmatrix} + \begin{pmatrix} \beta_{j1} \\ \beta_{j2} \\ \vdots \\ \beta_{jK} \end{pmatrix} + \begin{pmatrix} \epsilon_{ij1} \\ \epsilon_{ij2} \\ \vdots \\ \epsilon_{ijK} \end{pmatrix}$$

uma vez que o modelo para a  $k$ -ésima variável ( $k = 1, 2, \dots, K$ ) em cada vetor  $\mathbf{y}_{ij}$  é

$$y_{ijk} = \mu_k + \tau_{ik} + \beta_{jk} + \epsilon_{ijk}.$$

Temos interesse em comparar os vetores médios relativos aos  $I$  tratamentos (fator A) e os vetores médios em relação aos  $J$  blocos. Estas hipóteses serão denotadas por

$$\begin{cases} H_0^{(\tau)} : \boldsymbol{\mu}_{.1} = \boldsymbol{\mu}_{.2} = \dots = \boldsymbol{\mu}_{.I} \\ vs \\ H_1^{(\tau)} : \text{pelo menos dois } \boldsymbol{\mu}'s \text{ não são iguais} \end{cases}$$

e

$$\begin{cases} H_0^{(\beta)} : \boldsymbol{\mu}_{.1} = \boldsymbol{\mu}_{.2} = \dots = \boldsymbol{\mu}_{.J} \\ vs \\ H_1^{(\beta)} : \text{pelo menos dois } \boldsymbol{\mu}'s \text{ não são iguais} \end{cases}$$

em que cada vetor  $\boldsymbol{\mu}_{.i}$  e  $\boldsymbol{\mu}_{.j}$ ,  $i = 1, 2, \dots, I$  e  $j = 1, 2, \dots, J$  têm  $K$  componentes.

A igualdade dos vetores médios implica que as  $I$  médias relativas aos tratamentos ou as  $J$  médias relativas aos blocos são iguais para cada variável, isto é,  $\mu_{1.k} = \mu_{2.k} = \dots = \mu_{I.k}$  ou  $\mu_{.1k} = \mu_{.2k} = \dots = \mu_{.Jk}$ , para  $k = 1, 2, \dots, K$ . Se *duas* médias diferem para pelo menos uma variável, por exemplo,  $\mu_{2,3} \neq \mu_{4,3}$ , então  $H_0^{(\tau)}$  será falsa e nós a rejeitamos. Podemos ver isto examinando os elementos dos vetores médios populacionais:

$$H_0^{(\tau)} : \begin{pmatrix} \mu_{1.1} \\ \mu_{1.2} \\ \vdots \\ \mu_{1.K} \end{pmatrix} = \begin{pmatrix} \mu_{2.1} \\ \mu_{2.2} \\ \vdots \\ \mu_{2.K} \end{pmatrix} = \dots = \begin{pmatrix} \mu_{I.1} \\ \mu_{I.2} \\ \vdots \\ \mu_{I.K} \end{pmatrix}$$

e

$$H_0^{(\beta)} : \begin{pmatrix} \mu_{.11} \\ \mu_{.12} \\ \vdots \\ \mu_{.1K} \end{pmatrix} = \begin{pmatrix} \mu_{.21} \\ \mu_{.22} \\ \vdots \\ \mu_{.2K} \end{pmatrix} = \dots = \begin{pmatrix} \mu_{.J1} \\ \mu_{.J2} \\ \vdots \\ \mu_{.JK} \end{pmatrix}$$

Assim, tando  $H_0^{(\tau)}$  como  $H_0^{(\beta)}$  implica em  $K$  conjuntos de igualdades, isto é,

$$\left\{ \begin{array}{l} \mu_{1.1} = \mu_{2.1} = \dots = \mu_{I.1} \\ \mu_{1.2} = \mu_{2.2} = \dots = \mu_{I.2} \\ \vdots \\ \mu_{1.K} = \mu_{2.K} = \dots = \mu_{I.K} \end{array} \right\} \text{ ou } \left\{ \begin{array}{l} \mu_{.11} = \mu_{.21} = \dots = \mu_{.J1} \\ \mu_{.12} = \mu_{.22} = \dots = \mu_{.J2} \\ \vdots \\ \mu_{.1K} = \mu_{.2K} = \dots = \mu_{.JK} \end{array} \right\}$$

Todas as  $K(I-1)$  ou  $K(J-1)$  igualdades devem prevalecer para que  $H_0^{(\tau)}$  ou  $H_0^{(\beta)}$  seja verdadeira. A falha de uma única igualdade vai falsificar a hipótese  $H_0^{(\tau)}$  ou  $H_0^{(\beta)}$ .

No caso univariado tem-se *duas* somas de quadrados *entre*  $SQ_A$ ,  $SQ_B$  e *uma* soma de quadrados *dentro*  $SQ_E$  as quais são dadas por

$$SQ_A = J \sum_{i=1}^I (\bar{y}_i - \bar{y}_{..})^2 = \frac{1}{J} \sum_{i=1}^I y_i^2 - \frac{y_{..}^2}{IJ}, \quad SQ_B = I \sum_{j=1}^J (\bar{y}_{.j} - \bar{y}_{..})^2 = \frac{1}{I} \sum_{j=1}^J y_{.j}^2 - \frac{y_{..}^2}{IJ} \text{ e}$$

$$SQ_E = \sum_{i=1}^I \sum_{j=1}^J (y_{ij} - \bar{y}_i - \bar{y}_{.j} + \bar{y}_{..})^2 = \left( \sum_{i=1}^I \sum_{j=1}^J y_{ij}^2 - \frac{y_{..}^2}{IJ} \right) - \left( \frac{1}{J} \sum_{i=1}^I y_i^2 - \frac{y_{..}^2}{IJ} \right) - \left( \frac{1}{I} \sum_{j=1}^J y_{.j}^2 - \frac{y_{..}^2}{IJ} \right)$$

Para o caso multivariado, a partir das observações na Tabela 8, definimos os vetores

$$\mathbf{y}_i = \mathbf{y}_{i1} + \mathbf{y}_{i2} + \dots + \mathbf{y}_{iJ}, \quad i = 1, 2, \dots, I$$

$$\mathbf{y}_{.j} = \mathbf{y}_{1j} + \mathbf{y}_{2j} + \dots + \mathbf{y}_{Ij}, \quad j = 1, 2, \dots, J$$

e, por analogia com as somas de quadrados do caso univariado, as matrizes correspondentes  $\mathbf{H}_A$ ,  $\mathbf{H}_B$  e  $\mathbf{E}$ , como

$$\mathbf{H}_A = J \sum_{i=1}^I (\bar{\mathbf{y}}_i - \bar{\mathbf{y}}_{..})(\bar{\mathbf{y}}_i - \bar{\mathbf{y}}_{..})' = \frac{1}{J} \sum_{i=1}^I \mathbf{y}_i \mathbf{y}_i' - \frac{1}{IJ} \mathbf{y}_{..} \mathbf{y}_{..}', \quad (2.56)$$

$$\mathbf{H}_B = I \sum_{j=1}^J (\bar{\mathbf{y}}_{.j} - \bar{\mathbf{y}}_{..})(\bar{\mathbf{y}}_{.j} - \bar{\mathbf{y}}_{..})' = \frac{1}{I} \sum_{j=1}^J \mathbf{y}_{.j} \mathbf{y}_{.j}' - \frac{1}{IJ} \mathbf{y}_{..} \mathbf{y}_{..}' \quad (2.57)$$

e

$$\mathbf{E} = \sum_{i=1}^I \sum_{j=1}^J (\mathbf{y}_{ij} - \bar{\mathbf{y}}_i - \bar{\mathbf{y}}_{.j} + \bar{\mathbf{y}}_{..})(\mathbf{y}_{ij} - \bar{\mathbf{y}}_i - \bar{\mathbf{y}}_{.j} + \bar{\mathbf{y}}_{..})'$$

$$= \sum_{i=1}^I \sum_{j=1}^J \mathbf{y}_{ij} \mathbf{y}'_{ij} - \frac{1}{J} \sum_{i=1}^I \mathbf{y}_{i.} \mathbf{y}'_{i.} - \frac{1}{I} \sum_{j=1}^J \mathbf{y}_{.j} \mathbf{y}'_{.j} + \frac{1}{IJ} \mathbf{y}_{..} \mathbf{y}'_{..} \quad (2.58)$$

As matrizes hipóteses  $\mathbf{H}_A$  e  $\mathbf{H}_B$  são  $K \times K$  e têm somas de quadrados *entre* na diagonal principal para cada variável. Os elementos fora da diagonal principal dessas matrizes são somas de produtos para cada par de variáveis. Assumindo que não existe dependência linear nas variáveis, o posto de  $\mathbf{H}_A$ , isto é,  $r(\mathbf{H}_A) = \min(K, \nu_t)$  e o posto de  $\mathbf{H}_B$  é  $r(\mathbf{H}_B) = \min(K, \nu_b)$ , em que, para este modelo,  $\nu_t = (I - 1)$  e  $\nu_b = (J - 1)$ . Assim,  $\mathbf{H}_A$  e  $\mathbf{H}_B$  podem ser singulares. A matriz  $\mathbf{E}$  é  $K \times K$  e tem soma de quadrados *dentro* para cada variável na diagonal principal e soma de produtos *dentro* fora da diagonal. O posto de  $\mathbf{E}$ , isto é,  $r(\mathbf{E}) = K$  a menos que  $\nu = (I - 1)(J - 1) < K$ .

Assim, as matrizes  $\mathbf{H}_A$  e  $\mathbf{H}_B$  são dadas por

$$\mathbf{H}_A = \begin{pmatrix} SQA_{11} & SPA_{12} & \cdots & SPA_{1K} \\ SPA_{12} & SQA_{22} & \cdots & SPA_{2K} \\ \vdots & \vdots & & \vdots \\ SPA_{1K} & SPA_{2K} & \cdots & SQA_{KK} \end{pmatrix} \quad (2.59)$$

e

$$\mathbf{H}_B = \begin{pmatrix} SQB_{11} & SPB_{12} & \cdots & SPB_{1K} \\ SPB_{12} & SQB_{22} & \cdots & SPB_{2K} \\ \vdots & \vdots & & \vdots \\ SPB_{1K} & SPB_{2K} & \cdots & SQB_{KK} \end{pmatrix} \quad (2.60)$$

em que

$$\begin{aligned} SQA_{kk} &= J \sum_{i=1}^I (\bar{y}_{i.k} - \bar{y}_{..k})^2 = \frac{1}{J} \sum_{i=1}^I y_{i.k}^2 - \frac{y_{..k}^2}{IJ}, \quad k = 1, 2, \dots, K \\ SPA_{kk'} &= J \sum_{i=1}^I (\bar{y}_{i.k} - \bar{y}_{..k})(\bar{y}_{i.k'} - \bar{y}_{..k'}) = \frac{1}{J} \sum_{i=1}^I y_{i.k} y_{i.k'} - \frac{y_{..k} y_{..k'}}{IJ}, \quad k < k' = 2, \dots, K \\ SQB_{kk} &= I \sum_{j=1}^J (\bar{y}_{.jk} - \bar{y}_{..k})^2 = \frac{1}{I} \sum_{j=1}^J y_{.jk}^2 - \frac{y_{..k}^2}{IJ}, \quad k = 1, 2, \dots, K \\ SPB_{kk'} &= I \sum_{j=1}^J (\bar{y}_{.jk} - \bar{y}_{..k})(\bar{y}_{.jk'} - \bar{y}_{..k'}) = \frac{1}{I} \sum_{j=1}^J y_{.jk} y_{.jk'} - \frac{y_{..k} y_{..k'}}{IJ}, \quad k < k' = 2, \dots, K. \end{aligned}$$

Nestas expressões, os subscritos  $k$  e  $k'$  indicam as variáveis  $k$  e  $k'$ . Assim, por exemplo,  $\bar{y}_{i.1}$  é o primeiro elemento do vetor  $\bar{\mathbf{y}}_i$ , isto é,  $\bar{\mathbf{y}}_i = \left( \bar{y}_{i.1} \quad \bar{y}_{i.2} \quad \cdots \quad \bar{y}_{i.K} \right)'$ .



A matriz  $\mathbf{E}$  pode ser expressa de forma similar a (2.53) e (2.54), ou seja

$$\mathbf{E} = \begin{pmatrix} SQE_{11} & SPE_{12} & \cdots & SPE_{1K} \\ SPE_{12} & SQE_{22} & \cdots & SPE_{2K} \\ \vdots & \vdots & & \vdots \\ SPE_{1K} & SPE_{2K} & \cdots & SQE_{KK} \end{pmatrix} \quad (2.61)$$

em que

$$\begin{aligned} SQE_{kk} &= \sum_{i=1}^I \sum_{j=1}^J (y_{ijk} - \bar{y}_{i.k} - \bar{y}_{.jk} + \bar{y}_{..k})^2 \\ &= \sum_{i=1}^I \sum_{j=1}^J y_{ijk}^2 - \frac{1}{J} \sum_{i=1}^I y_{i.k}^2 - \frac{1}{I} \sum_{j=1}^J y_{.jk}^2 + \frac{y_{..k}^2}{IJ}, \quad k = 1, 2, \dots, K \end{aligned}$$

e

$$\begin{aligned} SPE_{kk'} &= \sum_{i=1}^I \sum_{j=1}^J (y_{ijk} - \bar{y}_{i.k} - \bar{y}_{.jk} + \bar{y}_{..k})(y_{ijk'} - \bar{y}_{i.k'} - \bar{y}_{.jk'} + \bar{y}_{..k'}) \\ &= \sum_{i=1}^I \sum_{j=1}^J y_{ijk} y_{ijk'} - \frac{1}{J} \sum_{i=1}^I y_{i.k} y_{i.k'} - \frac{1}{I} \sum_{j=1}^J y_{.jk} y_{.jk'} + \frac{y_{..k} y_{..k'}}{IJ} \end{aligned}$$

observe que os elementos de  $\mathbf{E}$  são somas de quadrados e produtos, não variâncias e covariâncias. Para estimar  $\Sigma$ , usamos  $\mathbf{S}_{pl} = \frac{\mathbf{E}}{(I-1)(J-1)}$ , uma vez que

$$E \left[ \frac{\mathbf{E}}{(I-1)(J-1)} \right] = \Sigma.$$

### 2.14.1 As estatísticas dos testes para o modelo MANOVA

Para Rencher e Christensen (2012) e Johnson e Wichern (2007), dentre outros, as estatísticas dos testes da razão de verossimilhança de  $H_0^{(\tau)} : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_I$ , e de  $H_0^{(\beta)} : \boldsymbol{\mu}_{.1} = \boldsymbol{\mu}_{.2} = \cdots = \boldsymbol{\mu}_{.J}$  são conhecidas por estatísticas  $\Lambda$  de Wilks e são denotadas, respectivamente, por

$$\Lambda_A = \frac{|\mathbf{E}|}{|\mathbf{E} + \mathbf{H}_A|} = \prod_{i=1}^s \frac{1}{1 + \lambda_i} \sim \Lambda_{(K, \nu_t, \nu)} \quad (2.62)$$

e

$$\Lambda_B = \frac{|\mathbf{E}|}{|\mathbf{E} + \mathbf{H}_B|} = \prod_{j=1}^s \frac{1}{1 + \lambda_j} \sim \Lambda_{(K, \nu_b, \nu)} \quad (2.63)$$

em que  $\lambda_i$  e  $\lambda_j$  são as raízes características de  $\mathbf{E}^{-1}\mathbf{H}_A$  e  $\mathbf{E}^{-1}\mathbf{H}_B$ , respectivamente. Portanto, rejeitaremos as hipóteses  $H_0^{(\tau)}$  e  $H_0^{(\beta)}$ , ao nível de significância  $\alpha$ , se observarmos

que  $\Lambda_A < \Lambda_{(K, \nu_t, \nu, \alpha)}$  e  $\Lambda_B < \Lambda_{(K, \nu_b, \nu, \alpha)}$ , respectivamente.

Os valores críticos da estatística de Wilks para  $\alpha = 0,05$  podem ser vistos em Rencher e Christensen (2012, p. 613-620), Tabela A.9.

Outras três estatísticas de teste de  $H_0^{(\tau)}$  e  $H_0^{(\beta)}$  que podem ser usadas são:

$$\begin{aligned} \text{Traço de Pillai: } V^{(s)} &= \sum_{k=1}^K \frac{\lambda_k}{1+\lambda_k} \quad \text{e} \quad V^{(s)} = \sum_{k=1}^K \frac{\lambda_j}{1+\lambda_j}, \\ \text{Traço de Lawley-Hotelling: } U^{(s)} &= \sum_{i=1}^K \lambda_i \quad \text{e} \quad U^{(s)} = \sum_{j=1}^K \lambda_j \\ \text{Maior raiz característica de Roy: } \theta &= \frac{\lambda_1}{1+\lambda_1} \end{aligned} \quad (2.64)$$

Os valores críticos de  $V^{(s)}$ ,  $U^{(s)}$  e  $\theta$  são dados nas Tabelas A11, A12 e A10 em (RENCHER; CHRISTENSEN, 2012, p. 621-631).

## 2.14.2 Contrastes multivariados

Quando a hipótese sobre a igualdade dos efeitos dos tratamentos é rejeitada, os efeitos que levaram  $H_0^{(\tau)}$  a ser rejeitada é de interesse. Uma alternativa seria estudar os efeitos de contrastes que, em geral, são estabelecidos pelo pesquisador no planejamento do experimento. Neste contexto definiremos o contraste multivariado

$$\Psi = \mathbf{C}\boldsymbol{\mu} = c_1\boldsymbol{\mu}_1 + c_2\boldsymbol{\mu}_2 + \cdots + c_I\boldsymbol{\mu}_I, \quad (2.65)$$

com  $\mathbf{C}\mathbf{j} = 0$ , cuja estimativa é

$$\hat{\Psi} = c_1\bar{\mathbf{y}}_1 + c_2\bar{\mathbf{y}}_2 + \cdots + c_I\bar{\mathbf{y}}_I, \quad (2.66)$$

e, como assumimos que os vetores  $\bar{\mathbf{y}}_1, \bar{\mathbf{y}}_2, \dots, \bar{\mathbf{y}}_I$  são independentes com matriz de covariância comum,  $Cov(\bar{\mathbf{y}}_i) = \boldsymbol{\Sigma}/J$ , então

$$Cov(\hat{\Psi}) = c_1^2 \frac{\boldsymbol{\Sigma}}{J} + c_2^2 \frac{\boldsymbol{\Sigma}}{J} + \cdots + c_I^2 \frac{\boldsymbol{\Sigma}}{J} = \frac{\boldsymbol{\Sigma}}{J} \sum_{i=1}^I c_i^2, \quad (2.67)$$

que pode ser estimado por

$$\frac{\mathbf{S}_{pl}}{J} \sum_{i=1}^I c_i^2 = \left( \frac{\mathbf{E}}{\nu_E} \right) \left( \frac{\sum_{i=1}^I c_i^2}{J} \right). \quad (2.68)$$

em que  $\mathbf{S}_{pl} = \mathbf{E}/\nu_E$  é um estimador não viciado de  $\boldsymbol{\Sigma}$ .

A hipótese  $H_0^{(\tau)} : \Psi_h = \mathbf{0}$  ou  $H_0^{(\tau)} : c_1\boldsymbol{\mu}_1 + c_2\boldsymbol{\mu}_2 + \cdots + c_I\boldsymbol{\mu}_I = \mathbf{0}$  faz comparações

entre os vetores médios populacionais, que sob suposições de normalidade multivariada apropriadas,  $H_0^{(\tau)}$  pode ser testada com a estatística  $T^2$ , ou seja,

$$\begin{aligned} T^2 &= \hat{\Psi}' \left( \frac{\mathbf{S}_{pl}}{J} \sum_{i=1}^I c_i^2 \right)^{-1} \hat{\Psi}_h \\ &= \frac{J}{\sum_{i=1}^I c_i^2} \left( \sum_{i=1}^I c_i \bar{\mathbf{y}}_i \right)' \left( \frac{\mathbf{E}}{\nu_E} \right)^{-1} \left( \sum_{i=1}^I c_i \bar{\mathbf{y}}_i \right) \end{aligned} \quad (2.69)$$

a qual é distribuída como uma  $T_{(p, \nu_E)}^2$ . No caso dos experimentos em blocos completos casualizados  $\nu_E = (I - 1)(J - 1)$ . Um teste equivalente de  $H_0^{(\tau)}$  pode ser feito com a  $\Lambda$  de Wilks. De acordo com Rencher e Christensen (2012, p. 193) a matriz hipótese para o contraste é dada por

$$\mathbf{H}_1 = \frac{J}{\sum_{i=1}^I c_i^2} \left( \sum_{i=1}^I c_i \bar{\mathbf{y}}_i \right) \left( \sum_{i=1}^I c_i \bar{\mathbf{y}}_i \right)' \quad (2.70)$$

O posto de  $\mathbf{H}_1$  é 1 e a estatística de teste é

$$\Lambda = \frac{|\mathbf{E}|}{|\mathbf{E} + \mathbf{H}_1|} \quad (2.71)$$

que é distribuída como uma  $\Lambda_{(p, 1, \nu_E)}$ . Os outros *três* testes MANOVA também podem ser aplicados aqui usando os autovalores não nulos de  $\mathbf{E}^{-1}\mathbf{H}_1$ . Devido a  $\nu_H = 1$  neste caso, todas as *quatro* estatísticas MANOVA e  $T^2$  levam ao mesmo resultado. Se  $(I - 1)$  contrastes ortogonais forem usados, eles particionam a matriz  $\mathbf{H}$  em  $(I - 1)$  matrizes independentes  $\mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_{I-1}$ . Cada matriz  $\mathbf{H}_i$  tem um grau de liberdade, uma vez que  $r(\mathbf{H}_i) = 1$ .

### 2.14.3 Intervalos de confiança simultâneos de Bonferroni

Para Johnson e Wichern (2007, p. 317-318) uma outra alternativa usada para a verificação dos efeitos que levaram  $H_0^{(\tau)}$  a ser rejeitada é a construção de intervalos com  $(1 - \alpha)100\%$  de confiança simultâneos de Bonferroni para cada componente (variável) de  $\boldsymbol{\mu}_i - \boldsymbol{\mu}_{i'}$ , usando a declaração

$$\mu_{ik} - \mu_{i'k} \quad \text{está contida no intervalo} \quad \bar{y}_{ik} - \bar{y}_{i'k} \pm t_{(\nu, \frac{\alpha}{2m})} \sqrt{\frac{2w_{kk}}{J\nu}} \quad (2.72)$$

para todo  $i < i' = 1, 2, \dots, I$  e  $k = 1, 2, \dots, K$ . Aqui,  $w_{kk}$  é o  $k$ -ésimo elemento da diagonal de  $\mathbf{E}$  e  $m = \frac{IK(I-1)}{2}$  é o número de declarações de confiança simultâneas. Procedimento análogo pode ser usado quando  $H_0^{(\beta)}$  for rejeitada e houver interesse em averiguar os

efeitos que levaram à sua rejeição. Isto é,

$$\mu_{jk} - \mu_{j'k} \quad \text{está contida no intervalo} \quad \bar{y}_{jk} - \bar{y}_{j'k} \pm t_{(\nu, \frac{\alpha}{2m'})} \sqrt{\frac{2w_{kk}}{I\nu}} \quad (2.73)$$

para todo  $j < j' = 1, 2, \dots, J$  e  $k = 1, 2, \dots, K$ . Aqui,  $w_{kk}$  é o  $k$ -ésimo elemento da diagonal de  $\mathbf{E}$  e  $m' = \frac{JK(J-1)}{2}$  é o número de declarações de confiança simultâneas.

Para um grupo de contrastes, ortogonais entre si, sobre as médias (ou efeitos) dos tratamentos, teremos  $(I - 1)$  contrastes ortogonais entre si envolvendo  $K$  variáveis e, por consequência, teremos  $K(I - 1)$  contrastes do tipo  $\ell' \boldsymbol{\mu}_{ik}$ . Assim, o intervalo com  $100(1 - \alpha)\%$  de confiança simultâneo para  $\ell' \boldsymbol{\mu}_{ik}$  são

$$\ell' \boldsymbol{\mu}_{ik} \quad \text{está contida no intervalo} \quad \ell' \bar{\mathbf{x}}_{ik} \pm t_{(\nu, \frac{\alpha}{2K(I-1)})} \sqrt{\frac{\ell' w_{kk} \ell}{J\nu}}, \quad k = 1, 2, \dots, K. \quad (2.74)$$

em que  $\ell$  é um vetor  $I \times 1$  dos coeficientes do contraste entre as médias dos tratamentos com  $\ell' \mathbf{j} = 0$ , onde  $\mathbf{j}$  é um vetor  $I \times 1$  de *uns*.

## 2.15 Análise de perfil

Assumir que as variáveis são comensuráveis, isto é, quando a cada indivíduo é dada uma bateria de testes ou quando instalamos um experimento e em cada unidade experimental fazemos medições em momentos pontuais ao longo do tempo de uma ou mais variáveis.

O modelo MANOVA básico para um experimento em blocos completos ao acaso é dado por

$$\mathbf{y}_{ij} = \boldsymbol{\mu} + \boldsymbol{\tau}_i + \boldsymbol{\beta}_j + \boldsymbol{\epsilon}_{ij}, \quad i = 1, 2, \dots, I, \quad j = 1, 2, \dots, J.$$

Para testar  $H_0^{(\boldsymbol{\tau})} : \boldsymbol{\mu}_{.1} = \boldsymbol{\mu}_{.2} = \dots = \boldsymbol{\mu}_{.I}$  usamos normalmente as matrizes  $\mathbf{H}_A$  e  $\mathbf{E}$  dadas em (2.59) e (2.61) e para testar  $H_0^{(\boldsymbol{\beta})} : \boldsymbol{\mu}_{.1} = \boldsymbol{\mu}_{.2} = \dots = \boldsymbol{\mu}_{.J}$  usamos  $\mathbf{H}_B$  e  $\mathbf{E}$  em (2.60) e (2.61). Se as variáveis são comensuráveis, podemos ser mais específicos e estender  $H_0$  a um exame dos  $I$  perfis obtidos através da representação gráfica dos  $K$  valores

$\mu_{i1}, \mu_{i2}, \dots, \mu_{iK}$  em cada  $\mu_i$ . Neste sentido, estaremos interessados em três hipóteses:

$H_0^{(1)}$  : Os  $I$  perfis são paralelos.

$H_0^{(2)}$  : Os  $I$  perfis estão no mesmo nível (são coincidentes).

$H_0^{(3)}$  : Os  $I$  perfis são horizontais.

A hipótese de paralelismo para  $I$  grupos pode ser expressa como  $H_0^{(1)} : \mathbf{C}\mu_1 = \mathbf{C}\mu_2 = \dots = \mathbf{C}\mu_I$ , em que  $\mathbf{C}$  é qualquer matriz  $(K-1) \times K$  de posto  $K-1$ , tal que  $\mathbf{C}\mathbf{j} = \mathbf{0}$ , em que  $\mathbf{j}$  é um vetor  $K \times 1$  de *uns*. Por exemplo,

$$\mathbf{C} = \begin{pmatrix} 1 & -1 & 0 & \dots & 0 & 0 \\ 0 & 1 & -1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & -1 \end{pmatrix}.$$

Para  $I$  grupos, a hipótese de paralelismo é

$$H_0^{(1)} : \mathbf{C}\mu_1 = \mathbf{C}\mu_2 = \dots = \mathbf{C}\mu_I. \quad (2.75)$$

A hipótese em (2.75) é equivalente a hipótese  $H_0 : \mu_{z1} = \mu_{z2} = \dots = \mu_{zI}$  na MANOVA para um experimento em blocos completos ao acaso sobre as variáveis transformadas  $\mathbf{z}_{ij} = \mathbf{C}\mathbf{y}_{ij}$ . Como  $\mathbf{C}$  tem  $K-1$  linhas, então  $\mathbf{C}\mathbf{y}_{ij}$  é  $(K-1) \times 1$  e  $\mathbf{C}\Sigma\mathbf{C}'$  é  $(K-1) \times (K-1)$ . Por propriedade da distribuição normal,  $\mathbf{z}_{ij}$  é distribuída como  $N_{K-1}(\mathbf{C}\mu_i, \mathbf{C}\Sigma\mathbf{C}')$  (RENCHER; CHRISTENSEN, 2012, p. 211).

Conforme Rencher e Christensen (2012, p. 212), a matriz hipótese  $\mathbf{H}_A$  e a matriz erro  $\mathbf{E}$  para testar  $H_0^{(1)}$  em (2.75) são:

$$\mathbf{H}_z = \mathbf{C}\mathbf{H}_A\mathbf{C}' \quad \text{e} \quad \mathbf{E}_z = \mathbf{C}\mathbf{E}\mathbf{C}'.$$

Assim, tem-se

$$\Lambda = \frac{|\mathbf{C}\mathbf{E}\mathbf{C}'|}{|\mathbf{C}\mathbf{E}\mathbf{C}' + \mathbf{C}\mathbf{H}_A\mathbf{C}'|} = \frac{|\mathbf{C}\mathbf{E}\mathbf{C}'|}{|\mathbf{C}(\mathbf{E} + \mathbf{H}_A)\mathbf{C}'|}, \quad (2.76)$$

que é distribuída como uma  $\Lambda_{[\nu_c, \nu_H, \nu_E]}$ , em que  $\nu_H = I-1$ ,  $\nu_E = (I-1)(J-1)$ . As outras *três* estatísticas de teste MANOVA podem ser obtidas dos autovalores de  $(\mathbf{C}\mathbf{E}\mathbf{C}')^{-1}(\mathbf{C}\mathbf{H}_A\mathbf{C}')$ .

A hipótese de que dois perfis estão no mesmo nível (são coincidentes) pode ser expressa

como

$$H_0^{(2)} : \mathbf{j}'\boldsymbol{\mu}_1 = \mathbf{j}'\boldsymbol{\mu}_2 = \cdots = \mathbf{j}'\boldsymbol{\mu}_k. \quad (2.77)$$

Para testar  $H_0^{(2)}$ , para  $I$  grupos podemos empregar o teste  $F$  da ANOVA para um experimento em blocos completos ao acaso, comparando  $I$  grupos a partir das observações transformadas  $\mathbf{j}'\mathbf{y}_{ij}$ . Alternativamente, podemos utilizar (2.76) com  $\mathbf{C} = \mathbf{j}'$ ,

$$\Lambda = \frac{\mathbf{j}'\mathbf{E}\mathbf{j}}{\mathbf{j}'\mathbf{E}\mathbf{j} + \mathbf{j}'\mathbf{H}_{Aj}}, \quad (2.78)$$

é distribuída como uma  $\Lambda_{(1, \nu_H, \nu_E)}$  ( $K = 1$  porque  $\mathbf{j}'\mathbf{y}_{ij}$  é um escalar). Isto é, de fato, equivalente ao teste  $F$  sobre  $\mathbf{j}'\mathbf{y}_{ij}$ . Pois, conforme Rencher e Christensen (2012, p. 212),

$$F = \left( \frac{1 - \Lambda}{\Lambda} \right) \frac{\nu_E}{\nu_H} \quad (2.79)$$

é distribuída como uma  $F_{(\nu_H, \nu_E)}$ .

A terceira hipótese, a de horizontalidade, essencialmente, afirma que a média dos  $I$  grupos de médias é a mesma para cada variável:

$$H_0^{(3)} : \frac{\mu_{11} + \mu_{21} + \cdots + \mu_{I1}}{I} = \frac{\mu_{12} + \mu_{22} + \cdots + \mu_{I2}}{I} = \cdots = \frac{\mu_{1K} + \mu_{2K} + \cdots + \mu_{IK}}{I},$$

ou

$$H_0^{(3)} : \frac{\mathbf{C}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2 + \cdots + \boldsymbol{\mu}_I)}{I} = \mathbf{0}, \quad (2.80)$$

em que  $\mathbf{C}$ , é uma matriz  $(K - 1) \times K$  de posto  $(K - 1)$ , tal que  $\mathbf{C}\mathbf{j} = \mathbf{0}$ . A hipótese de horizontalidade também pode ser indicada como, as médias das  $K$  variáveis em cada grupo são as mesmas, ou  $\mu_{i1} = \mu_{i2} = \cdots = \mu_{iK}$ ,  $i = 1, 2, \cdots, I$ . Esta pode ser expressa como  $H_0^{(3)} : \mathbf{C}\boldsymbol{\mu}_1 = \mathbf{C}\boldsymbol{\mu}_2 = \cdots = \mathbf{C}\boldsymbol{\mu}_I = \mathbf{0}$ .

Para testar  $H_0^{(3)}$  como dada por (2.80), podemos estender a estatística de teste  $T^2$ . O vetor de médias geral  $(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2 + \cdots + \boldsymbol{\mu}_I)/I$  pode ser estimado por

$$\bar{\mathbf{y}}_{..} = \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J \mathbf{y}_{ij}.$$

Sob  $H_0^{(3)}$  (e  $H_0^{(1)}$ ),  $\mathbf{C}\bar{\mathbf{y}}_{..} \sim N_{K-1}(\mathbf{0}, \mathbf{C}\boldsymbol{\Sigma}\mathbf{C}'/IJ)$ , e  $H_0^{(3)}$  pode ser testada por

$$T^2 = IJ(\mathbf{C}\bar{\mathbf{y}}_{..})'(\mathbf{C}\mathbf{E}\mathbf{C}'/\nu_E)^{-1}(\mathbf{C}\bar{\mathbf{y}}_{..}), \quad (2.81)$$

em que  $\mathbf{E}/\nu_E$  é um estimador de  $\boldsymbol{\Sigma}$ . A rigor,  $H_0^{(3)}$  não é afetada pelo *status* de  $H_0^{(2)}$ . Quando  $H_0^{(1)}$  e  $H_0^{(3)}$  são verdadeiras e  $T^2$  em (2.81) é distribuída como uma  $T^2_{[(K-1), \nu_E]}$

(RENCHEK; CHRISTENSEN, 2012).

### 3 Material e métodos

Para ilustrar a teoria desenvolvida foram considerados os dados sobre o diâmetro do caule de girassol obtidos de dois experimento conduzidos, em 2010 e 2011, pelo Professor Doutor Pedro Dantas Fernandes da Universidade Federal de Campina Grande. Os tratamentos planejados foram:  $T_1 =$  Bacia,  $T_2 =$  Leirão,  $T_3 =$  Sulco e  $T_4 =$  Testemunha, distribuídos aleatoriamente em 6 blocos e avaliados em 5 épocas após a emergência:  $E_1 = 30$  dias,  $E_2 = 45$  dias,  $E_3 = 60$  dias,  $E_4 = 70$  dias e  $E_5 = 80$  dias. Os valores do diâmetro do caule em cada unidade experimental estão apresentados nas Tabelas 16. Os métodos empregados foram as análises multivariada e univariada visando obter o melhor tratamento para a obtenção de uma planta com maior diâmetro médio do caule.



Tabela 16: Diâmetro do caule de girassol avaliado em diferentes épocas após a emergência num experimento em blocos completos ao acaso, Campina Grande, em 2010 e 2011

Tratamento	Bloco	Ano 2010					Ano 2011				
		Época de avaliação (em dias após a emergência)									
		30	45	60	70	80	30	45	60	70	80
1	1	8,45	9,88	18,75	20,62	20,86	8,05	14,06	18,52	24,03	24,02
1	2	8,12	11,24	18,12	20,28	19,60	7,03	15,20	21,20	22,84	23,64
1	3	8,10	9,26	18,55	19,94	24,05	9,45	15,23	20,12	22,64	23,90
1	4	7,62	9,48	21,07	23,40	21,82	7,70	15,20	21,05	23,94	24,32
1	5	7,50	10,30	20,80	21,80	22,45	8,20	16,08	20,30	21,52	23,92
1	6	7,84	10,43	20,54	22,05	20,33	7,80	14,92	19,88	24,33	22,30
2	1	8,48	11,92	20,71	26,30	23,54	10,02	16,94	24,55	26,06	25,60
2	2	9,05	12,35	22,76	27,04	27,80	9,30	16,92	24,82	27,20	25,24
2	3	8,42	11,05	20,20	26,33	27,74	8,45	19,24	24,20	24,82	27,52
2	4	9,12	13,06	21,37	24,08	26,33	9,40	21,02	25,05	26,46	24,65
2	5	9,24	12,30	20,90	23,36	25,05	8,30	15,88	26,50	26,04	27,20
2	6	9,20	13,04	21,97	28,23	26,08	8,88	18,20	23,84	26,30	26,74
3	1	7,50	11,49	19,64	22,33	24,10	8,10	16,25	21,10	23,00	26,03
3	2	8,25	12,12	21,43	24,82	25,12	8,02	15,10	24,33	26,38	24,85
3	3	8,76	12,26	21,52	24,06	21,52	7,39	16,03	20,82	27,20	26,45
3	4	8,12	10,41	22,76	25,30	19,98	7,02	15,33	23,22	25,39	25,42
3	5	9,05	12,08	22,41	26,20	22,30	6,98	15,02	21,33	24,98	26,20
3	6	7,98	11,05	20,77	21,03	26,60	7,72	14,48	23,66	26,66	25,02
4	1	6,85	8,02	14,02	16,10	15,20	7,20	12,35	13,60	17,20	16,48
4	2	7,01	7,82	13,68	15,85	14,88	7,33	14,12	16,01	16,33	16,80
4	3	6,23	8,04	11,75	15,20	14,36	7,10	13,05	13,92	15,62	15,02
4	4	7,03	7,20	14,41	16,07	15,45	6,32	12,40	16,25	16,33	16,10
4	5	7,20	7,06	14,04	14,55	16,04	6,30	12,93	13,85	15,55	15,80
4	6	6,82	7,33	13,86	14,90	15,82	6,50	12,84	14,20	16,30	17,03

## 4 Resultados e discussão

Inicialmente os dados foram inspecionados com o objetivo de identificar algumas observações atípicas. No entanto, não foi identificada nenhuma observação suspeita. Em seguida, procedemos às análises estatísticas dos dados utilizando *dois* programa SAS: PROGRAMA SAS #1 e PROGRAMA SAS #2 que se encontra no Apêndice A. Os procedimentos analíticos utilizados estão baseados nas expressões em (2.13) e Tabela 6 da análise de variância com o desdobramento dos graus de liberdade e somas de quadrados de tratamentos em contrastes de interesse. Conforme podemos observar os resultados dos quadrados médios da análise de variância na Tabela 17, a hipótese de igualdade das médias do diâmetro do caule do girassol em relação aos tratamentos foi rejeitada em todas as épocas de avaliação, indicando que existe pelo menos um par de médias relativo aos tratamentos, dentro de cada época, que difere estatisticamente entre si, ao nível de 1% de significância. Também, estão apresentas na Tabela 17, os resultados dos quadrados médios relativos a *três* contrastes de interesse:

$\Psi_1 = \frac{\mu_1 + \mu_2 + \mu_3}{3} - \mu_4 = 0$  (o diâmetro médio do caule do girassol plantado usando um dos *três* sistemas de captação de água é igual ao diâmetro do caule do girassol que foi plantado sem nenhum sistema de captação),

$\Psi_2 = \frac{\mu_1 + \mu_3}{2} - \mu_2 = 0$  (o diâmetro médio do caule do girassol plantado em *bacias* ou *sulcos* é igual ao diâmetro médio do caule quando o girassol é plantado em *leirões*),

$\Psi_3 = \mu_1 - \mu_3 = 0$  (o diâmetro médio do caule do girassol quando plantado em *bacias* é igual ao diâmetro médio do caule quando plantado em *sulcos*).

de onde podem observar que o contraste  $\Psi_1$  apresentou significância estatística, ao nível de 1% pelo teste  $F$ , para todas as épocas de avaliação, isto quer dizer que usar algum sistema de captação de água no plantio, induz o girassol a produzir maior diâmetro. Com relação ao contraste  $\Psi_2$ , observou-se que apenas nas épocas de avaliação  $E_1 = 30 \text{ dias}$  e  $E_3 = 60 \text{ dias}$ , não foi significativo, sendo que nas demais épocas de avaliação ele apresentou

significância estatística ao nível de 1%. Além disso, o contraste  $\Psi_3$  apresentou significância estatística, ao nível de 5% ou 1% nas épocas  $E_2 = 45 \text{ dias}$ ,  $E_3 = 60 \text{ dias}$ ,  $E_4 = 70 \text{ dias}$ ,  $E_5 = 80 \text{ dias}$  e não foi estatisticamente significativo nas épocas  $E_1 = 30 \text{ dias}$  e  $E_5 = 80 \text{ dias}$ . Também estão apresentadas na Tabel 17 as estimativas das médias, as diferenças mínimas significativas, (DMS), calculadas pelo método de Tukey a 5% de significância, e os valores da estatística  $W$  de Shapiro-Wilk para testar a normalidade dos erro em cada época de avaliação. Como podemos observar na Tabela 17, a hipótese de normalidade dos

Tabela 17: Análise de variância para testar a hipótese de igualdade das médias dos tratamentos e contrastes de interesse do ano de 2010

F. Variação	GL	Quadrados médios em cada época de avaliação				
		30 dias	45 dias	60 dias	70 dias	80 dias
$\hat{\Psi}_1 : \frac{\bar{y}_1 + \bar{y}_2 + \bar{y}_3}{3} - \bar{y}_4$	1	10,41**	62,92**	231,09**	309,01**	312,58**
$\hat{\Psi}_2 : \frac{\bar{y}_1 + \bar{y}_3}{2} - \bar{y}_2$	1	2,63	8,45*	2,49	41,93**	54,64**
$\hat{\Psi}_3 : \bar{y}_1 - \bar{y}_3$	1	0,34	6,48*	9,54**	20,41**	9,20**
Tratamento	(3)	4,46**	25,95**	81,04**	123,78**	125,47**
Bloco	5	0,10	0,35	2,15	0,52	1,19
Resíduo	15	0,20	0,49	0,86	2,92	3,48
	$\bar{y}_1$	7,94b	10,10b	19,64b	21,35b	21,52b
Estimativas das médias dos tratamentos	$\bar{y}_2$	8,92a	12,29a	21,32a	25,90a	26,09a
	$\bar{y}_3$	8,28ab	11,57a	21,42a	23,96a	23,27ab
	$\bar{y}_4$	6,86c	7,58c	13,63c	15,45c	15,29c
Teste de Tukey a 5%	DMS	0,74	1,16	1,54	2,84	3,10
Teste de Shapiro-Wilk	W	0,983	0,976	0,965	0,954	0,953

\* estatisticamente significativo ao nível de 5%,

\*\* estatisticamente significativo ao nível de 1%,

Médias seguidas das mesmas letras não diferem estatisticamente entre si (Tukey a 5%).

erros dentro de cada época de avaliação não foi rejeitada.

## 4.1 Teste de uniformidade de $\Sigma$

Para verificação da hipótese de uniformidade da matriz de covariância dos erros entre as épocas de avaliação, obtivemos através do programa SAS os resíduos dos experimentos em blocos, dentro de cada época, os quais estão apresentados na Tabela 19.

A matriz das somas de quadrados e produtos residuais  $\mathbf{E}$  para cada ano, foi obtida aqui, usando as linhas de comando do SAS/IML, conforme PROGRAMA SAS #3 no Apêndice A,

Tabela 18: Resíduos obtidos em cada época de avaliação para os anos de 2010 e 2011

Ano	Trat.	Bloco	Época de avaliação				
			$E_1$	$E_2$	$E_3$	$E_4$	$E_5$
2010	1	1	0,68917	-0,16292	-0,16708	-0,40583	-0,04083
	1	2	0,07167	0,64208	-1,51458	-1,40583	-2,22583
	1	3	0,28167	-0,60792	-0,09208	-1,13083	2,15667
	1	4	-0,29333	-0,27292	0,53042	1,49917	0,94917
	1	5	-0,68833	0,14958	0,62542	0,63417	1,01417
	1	6	-0,06083	0,25208	0,61792	0,80917	-1,85333
	2	1	-0,26083	-0,31125	0,11292	0,73250	-1,93250
	2	2	0,02167	-0,43625	1,44542	0,81250	1,40250
	2	3	-0,37833	-1,00625	-0,12208	0,71750	1,27500
	2	4	0,22667	1,11875	-0,84958	-2,36250	0,88750
	2	5	0,07167	-0,03875	-0,95458	-2,34750	-0,95750
	2	6	0,31917	0,67375	0,36792	2,44750	-0,67500
	3	1	-0,59917	-0,02292	-1,06042	-1,30417	1,44750
	3	2	-0,13667	0,05208	0,01208	0,52583	1,54250
	3	3	0,60333	0,92208	1,09458	0,38083	-2,12500
	3	4	-0,13167	-0,81292	0,43708	0,79083	-2,64250
	3	5	0,52333	0,45958	0,45208	2,42583	-0,88750
	3	6	-0,25917	-0,59792	-0,93542	-2,81917	2,66500
	4	1	0,17083	0,49708	1,11458	0,97750	0,52583
	4	2	0,04333	-0,25792	0,05708	0,06750	-0,71917
4	3	-0,50667	0,69208	-0,88042	0,03250	-1,30667	
4	4	0,19833	-0,03292	-0,11792	0,07250	0,80583	
4	5	0,09333	-0,57042	-0,12292	-0,71250	0,83083	
4	6	0,00083	-0,32792	-0,05042	-0,43750	-0,13667	
2011	1	1	-0,47417	-0,58875	-0,58750	1,03750	0,23125
	1	2	-1,07167	0,11625	-0,05500	-0,76750	0,25125
	1	3	1,17083	-0,40625	0,69000	-0,35000	-0,07875
	1	4	-0,09167	-0,53625	-0,00750	0,49000	0,94125
	1	5	0,57333	1,35375	0,14000	-0,92250	-0,11625
	1	6	-0,10667	0,06125	-0,18000	0,51250	-1,22875
	2	1	0,47583	-0,62708	0,79417	0,13750	-0,66375
	2	2	0,17833	-1,08208	-1,08333	0,66250	-0,62375
	2	3	-0,84917	0,68542	0,12167	-1,10000	1,06625
	2	4	0,58833	2,36542	-0,65583	0,08000	-1,20375
	2	5	-0,34667	-1,76458	1,69167	0,66750	0,68875
	2	6	-0,04667	0,42292	-0,86833	-0,44750	0,73625
	3	1	0,07583	1,34792	-0,23917	-2,37750	0,26292
	3	2	0,41833	-0,23708	0,84333	0,38750	-0,51708
	3	3	-0,38917	0,14042	-0,84167	1,82500	0,49292
	3	4	-0,27167	-0,65958	-0,06917	-0,44500	0,06292
	3	5	-0,14667	0,04042	-1,06167	0,15250	0,18542
	3	6	0,31333	-0,63208	1,36833	0,45750	-0,48708
	4	1	-0,07750	-0,13208	0,03250	1,20250	0,16958
	4	2	0,47500	1,20292	0,29500	-0,28250	0,88958
4	3	0,06750	-0,41958	0,03000	-0,37500	-1,48042	
4	4	-0,22500	-1,16958	0,73250	-0,12500	0,19958	
4	5	-0,08000	0,37042	-0,77000	0,10250	-0,75792	
4	6	-0,16000	0,14792	-0,32000	-0,52250	0,97958	

$$\mathbf{E} = \begin{pmatrix} 2,9288833 & 1,3710333 & 1,3508333 & 0,9443667 & -2,6842080 \\ 1,3710333 & 7,2851625 & -0,5266790 & 1,1376250 & -6,7047750 \\ 1,3508333 & -0,5266790 & 12,9153460 & 17,3624420 & -1,6492580 \\ 0,9443667 & 1,1376250 & 17,3624420 & 43,7638170 & -13,2814200 \\ -2,6842080 & -6,7047750 & -1,6492580 & -13,2814200 & 52,1810670 \end{pmatrix} \quad (4.1)$$

Este resultado coincide com aquela da PROC MANOVA.

Obviamente, como o pesquisador observou o diâmetro do caule do girassol em diferentes épocas do ciclo da cultura, depreende-se que o mesmo tem interesse em verificar o comportamento médio do diâmetro ao longo do tempo. Para esta finalidade, poderemos adotar, pelo menos *três* abordagens: (a) considerar que as observação são provenientes de um experimento com parcelas sub-divididas no tempo (split-plot), cuja aplicabilidade está restrita à condição de esfericidade da matriz de covariâncias dos erros,  $\Sigma$  (WILKS, 1946), (b) considerar a proposta do item (a) e fazer a correção dos graus de liberdade (BOCK, 1963); (HUYNH; FELDT, 1970) ou (c) considerar a abordagem multivariada por análise de perfil que não exige nenhuma estrutura especial para  $\Sigma$ , (RENCHE; CHRISTENSEN, 2012); (JOHNSON; WICHERN, 2007); etc. Neste sentido, para testar a hipótese de uniformidade da matriz  $\Sigma$ , isto é,  $H_0 : \Sigma = \sigma^2[(1 - \rho)\mathbf{I} + \rho\mathbf{J}]$  em (2.40), consideramos a matriz de contrastes

$$\mathbf{C} = \begin{pmatrix} \frac{1}{\sqrt{1 \times 2}} & \frac{-1}{\sqrt{1 \times 2}} & 0 & 0 & 0 \\ \frac{1}{\sqrt{2 \times 3}} & \frac{1}{\sqrt{2 \times 3}} & \frac{-2}{\sqrt{2 \times 3}} & 0 & 0 \\ \frac{1}{\sqrt{3 \times 4}} & \frac{1}{\sqrt{3 \times 4}} & \frac{1}{\sqrt{3 \times 4}} & \frac{-3}{\sqrt{3 \times 4}} & 0 \\ \frac{1}{\sqrt{4 \times 5}} & \frac{1}{\sqrt{4 \times 5}} & \frac{1}{\sqrt{4 \times 5}} & \frac{1}{\sqrt{4 \times 5}} & \frac{-4}{\sqrt{4 \times 5}} \end{pmatrix} \quad (4.2)$$

Daí, calculamos

$$\mathbf{S}_{pl} = \frac{\mathbf{E}}{\nu} = \begin{pmatrix} 0,1952589 & 0,0914022 & 0,0900556 & 0,0629578 & -0,178947 \\ 0,0914022 & 0,4856775 & -0,035112 & 0,0758417 & -0,446985 \\ 0,0900556 & -0,035112 & 0,8610231 & 1,1574961 & -0,109951 \\ 0,0629578 & 0,0758417 & 1,1574961 & 2,9175878 & -0,885428 \\ -0,178947 & -0,446985 & -0,109951 & -0,885428 & 3,4787378 \end{pmatrix}$$

e

$$\mathbf{C}\mathbf{S}_{pl}\mathbf{C}' = \begin{pmatrix} 0,249066 & -0,156102 & -0,025842 & -0,197688 \\ -0,156102 & 0,6813431 & 0,6617729 & -0,133765 \\ -0,025842 & 0,6617729 & 1,6929302 & -1,109759 \\ -0,197688 & -0,133765 & -1,109759 & 3,7987561 \end{pmatrix}.$$

Substituindo-se  $\mathbf{C}\mathbf{S}_{pl}\mathbf{C}'$  no lugar de  $\mathbf{S}$  e  $(K-1)$  no lugar de  $K$  em (2.34), obtivemos

$$LR = \left[ \frac{|\mathbf{C}\mathbf{S}_{pl}\mathbf{C}'|}{\left(\frac{\text{tr}(\mathbf{S})}{K-1}\right)^{K-1}} \right]^{n/2} = 2,046 \times 10^{-15}$$

em que  $n = 24$  para os dados em consideração.

Agora, substituindo o resultado  $LR$  nas equações (2.36) e (2.38) encontramos

$$u = (LR)^{2/n} = 0,0596901 \quad \text{e} \quad u' = -\left(\nu - \frac{2(K-1)^2 + (K-1) + 2}{6(K-1)}\right) \ln u = 37.81607.$$

Como  $u' > \chi^2_{[\frac{1}{2}K(K+1)-1, 0,05]} = \chi^2_{[9, 0,05]} = 16,919$ , então rejeitamos a hipótese de uniformidade de  $\Sigma$  e concluímos ao nível de 5% que a matriz  $\Sigma$  não é uniforme. Portanto, deveremos analisar os dados usando um procedimento multivariado tipo análise de perfil.

## 4.2 A análise de variância multivariada MANOVA

Para verificar a nulidade dos efeitos dos tratamentos, contrastes de interesse sobre os tratamentos e blocos, inicialmente obtivemos as matrizes  $\mathbf{H}_A$ ,  $\mathbf{H}_B$  e  $\mathbf{E}$ , a partir das expressões em (2.56), (2.57) e (2.58) respectivamente, cujos resultados foram

$$\mathbf{H}_A = \begin{pmatrix} 13,385216667 & 31,802191667 & 53,418941667 & 69,869983333 & 70,814033333 \\ 31,802191667 & 77,8569125 & 133,03507917 & 169,7673 & 169,457825 \\ 53,418941667 & 133,03507917 & 243,11587917 & 291,38778333 & 289,79060833 \\ 69,869983333 & 169,7673 & 291,38778333 & 371,34563333 & 372,35926667 \\ 70,814033333 & 169,457825 & 289,79060833 & 372,35926667 & 376,42508333 \end{pmatrix},$$

$$\mathbf{H}_B = \begin{pmatrix} 0,49015 & 0,44445 & 1,3922 & 0,28905 & 0,276225 \\ 0,44445 & 1,7364208333 & 0,0274125 & 0,166075 & 1,494875 \\ 1,3922 & 0,0274125 & 10,7722375 & 3,509475 & -1,473925 \\ 0,28905 & 0,166075 & 3,509475 & 2,58015 & -0,86125 \\ 0,276225 & 1,494875 & -1,473925 & -0,86125 & 5,9391 \end{pmatrix}$$

e

$$\mathbf{E} = \begin{pmatrix} 2,9288833333 & 1,3710333333 & 1,3508333333 & 0,9443666667 & -2,6842083333 \\ 1,3710333333 & 7,2851625 & -0,526679167 & 1,137625 & -6,704775 \\ 1,3508333333 & -0,526679167 & 12,915345833 & 17,362441667 & -1,649258333 \\ 0,9443666667 & 1,137625 & 17,362441667 & 43,763816667 & -13,28141667 \\ -2,684208333 & -6,704775 & -1,649258333 & -13,28141667 & 52,181066667 \end{pmatrix}.$$

Além disso, obtivemos  $\mathbf{H}_1$ ,  $\mathbf{H}_2$  e  $\mathbf{H}_3$  para os respectivos contrastes multivariados de interesse sobre os tratamentos de,  $\Psi_1 = \frac{\mu_1 + \mu_2 + \mu_3}{3} - \mu_4$ ,  $\Psi_2 = \frac{\mu_1 + \mu_3}{2} - \mu_2$  e  $\Psi_3 = \mu_1 - \mu_3$ , quais sejam,

$$\mathbf{H}_1 = \begin{pmatrix} 10,412005556 & 25,596497222 & 49,052030556 & 56,722233333 & 57,049272222 \\ 25,596497222 & 62,925501389 & 120,58773472 & 139,44388333 & 140,24786389 \\ 49,052030556 & 120,58773472 & 231,08916806 & 267,22428333 & 268,76499722 \\ 56,722233333 & 139,44388333 & 267,22428333 & 309,0098 & 310,79143333 \\ 57,049272222 & 140,24786389 & 268,76499722 & 310,79143333 & 312,58333889 \end{pmatrix},$$

$$\mathbf{H}_2 = \begin{pmatrix} 2,6298027778 & 4,7136444444 & 2,5568277778 & 10,500291667 & 11,986819444 \\ 4,7136444444 & 8,4487111111 & 4,5828444444 & 18,820666667 & 21,485111111 \\ 2,5568277778 & 4,5828444444 & 2,4858777778 & 10,208916667 & 11,654194444 \\ 10,500291667 & 18,820666667 & 10,208916667 & 41,925625 & 47,861041667 \\ 11,986819444 & 21,485111111 & 11,654194444 & 47,861041667 & 54,636736111 \end{pmatrix},$$

e

$$\mathbf{H}_3 = \begin{pmatrix} 0,3434083333 & 1,49205 & 1,8100833333 & 2,6474583333 & 1,7779416667 \\ 1,49205 & 6,4827 & 7,8645 & 11,50275 & 7,72485 \\ 1,8100833333 & 7,8645 & 9,5408333333 & 13,954583333 & 9,3714166667 \\ 2,6474583333 & 11,50275 & 13,954583333 & 20,410208333 & 13,706791667 \\ 1,7779416667 & 7,72485 & 9,3714166667 & 13,706791667 & 9,2050083333 \end{pmatrix}.$$

Observe que, sendo  $\mathbf{H}_1$ ,  $\mathbf{H}_2$  e  $\mathbf{H}_3$  independentes, então  $\mathbf{H}_A = \mathbf{H}_1 + \mathbf{H}_2 + \mathbf{H}_3$ .

Para calcular os diferentes valores das estatísticas dos testes de

$$H_0^{(\tau)} : \mu_1 = \mu_2 = \dots = \mu_4 = \mu, \quad H_0^{(\beta)} : \mu_{.1} = \mu_{.2} = \dots = \mu_{.6} = \mu, \\ H_0^{(1)} : \Psi_1 = 0, \quad H_0^{(2)} : \Psi_2 = 0, \quad H_0^{(3)} : \Psi_3 = 0,$$

calculamos os autovalores de  $\mathbf{E}^{-1}\mathbf{H}_A$ ,  $\mathbf{E}^{-1}\mathbf{H}_B$ ,  $\mathbf{E}^{-1}\mathbf{H}_1$ ,  $\mathbf{E}^{-1}\mathbf{H}_2$  e  $\mathbf{E}^{-1}\mathbf{H}_3$  e apresentamos na Tabela 19

A seguir apresentamos na Tabela 20 os resultados da MANOVA com base nos resultados da Tabela 19

Tabela 19: Autovalores das matrizes  $E^{-1}H_A$ ,  $E^{-1}H_B$ ,  $E^{-1}H_1$ ,  $E^{-1}H_2$  e  $E^{-1}H_3$ 

Autovalor	Matriz				
	$E^{-1}H_A$	$E^{-1}H_B$	$E^{-1}H_1$	$E^{-1}H_2$	$E^{-1}H_3$
$\hat{\lambda}_1$	50,9256188	1,51687271	44,8834225	6,67380619	2,53439851
$\hat{\lambda}_2$	3,0533952	0,42224590	0,0000000	0,0000000	0,0000000
$\hat{\lambda}_3$	0,1126132	0,08066287	0,0000000	0,0000000	0,0000000
$\hat{\lambda}_4$	0,0000000	0,05750453	0,0000000	0,0000000	0,0000000
$\hat{\lambda}_5$	0,0000000	0,03535154	0,0000000	0,0000000	0,0000000

Tabela 20: Análise de variância multivariada (MANOVA) para testar a nulidade dos efeitos dos tratamentos, dos blocos e dos contrastes de interesse sobre os tratamentos

Hipótese	Estatística	Valor $\Lambda_{(.)}$	Valor de $F$	gl Num	gl Den	$Pr > F$
$H_0^{(\tau)}$	Wilks' Lambda	0,00427	12,75	15	30,768	< .0001
	Pillai's Trace	1,83525	4,10	15	39,000	0.0002
	Hotelling-Lawley Trace	54,09163	37,04	15	16,143	< .0001
	Roy's Greatest Root	50,92562	132,41	5	13,000	< .0001
$H_0^{(\beta)}$	Wilks' Lambda	0,23610	0,80	25	42,365	0.7149
	Pillai's Trace	1,06273	0,81	25	75,000	0.7178
	Hotelling-Lawley Trace	2,11264	0,85	25	19,000	0.6532
	Roy's Greatest Root	1,51687	4,55	5	15,000	0.0100
$H_0^{(1)}$	Wilks' Lambda	0,02179	98,74	5	11,000	< .0001
	Pillai's Trace	0,97821	98,74	5	11,000	< .0001
	Hotelling-Lawley Trace	44,88342	98,74	5	11,000	< .0001
	Roy's Greatest Root	44,88342	98,74	5	11,000	< .0001
$H_0^{(2)}$	Wilks' Lambda	0,13031	14,68	5	11,000	0.0001
	Pillai's Trace	0,86969	14,68	5	11,000	0.0001
	Hotelling-Lawley Trace	6,67381	14,68	5	11,000	0.0001
	Roy's Greatest Root	6,67381	14,68	5	11,000	0.0001
$H_0^{(3)}$	Wilks' Lambda	0,13031	14,68	5	11,000	0.0001
	Pillai's Trace	0,86969	14,68	5	11,000	0.0001
	Hotelling-Lawley Trace	6,67381	14,68	5	11,000	0.0001
	Roy's Greatest Root	6,67381	14,68	5	11,000	0.0001

### 4.3 Intervalos de confiança simultâneos

Para verificação dos efeitos que levaram  $H_0^{(\tau)}$  a ser rejeitada foram construídos intervalos com 95% de confiança simultâneos de Bonferroni para cada componente (variável) de  $\mu_i - \mu_{i'}$ , conforme expressão (2.72), isto é,

$$\mu_{ik} - \mu_{i'k} \quad \text{está contida no intervalo} \quad \bar{y}_{ik} - \bar{y}_{i'k} \pm t_{(\nu, \frac{\alpha}{2m})} \sqrt{\frac{2w_{kk}}{J\nu}}$$



para todo  $i < i' = 1, 2, 3, 4$  e  $k = 1, 2, \dots, 5$ . Aqui,  $w_{kk}$  é o  $k$ -ésimo elemento da diagonal de  $\mathbf{E}$  e  $m = \frac{IK(I-1)}{2} = 30$  é o número de declarações de confiança simultâneas. Estes intervalos estão apresentados na Tabela 22. Procedimento análogo poderia ter sido usado se  $H_0^{(\beta)}$  tivesse sido rejeitada e houvesse interesse em averiguar os efeitos que levaram à sua rejeição. Isto é,

$$\mu_{jk} - \mu_{j'k} \quad \text{está contida no intervalo} \quad \bar{y}_{jk} - \bar{y}_{j'k} \pm t_{(\nu, \frac{\alpha}{2m'})} \sqrt{\frac{2w_{kk}}{I\nu}}$$

para todo  $j < j' = 1, 2, \dots, 6$  e  $k = 1, 2, \dots, 5$ . Aqui,  $w_{kk}$  é o  $k$ -ésimo elemento da diagonal de  $\mathbf{E}$  e  $m' = \frac{JK(J-1)}{2} = 75$  é o número de declarações de confiança simultâneas.

Para os *três* contrastes, ortogonais entre si, sobre as médias (ou efeitos) dos tratamentos, teremos  $(I - 1) = 3$  contrastes ortogonais entre si envolvendo  $K = 5$  variáveis e, por consequência, teremos  $K(I - 1) = 15$  contrastes do tipo  $\ell' \boldsymbol{\mu}_{ik}$ . Assim, o intervalo com 95% de confiança simultâneo para  $\ell' \boldsymbol{\mu}_{ik}$  são

$$\ell' \boldsymbol{\mu}_{ik} \quad \text{está contida no intervalo} \quad \ell' \bar{\mathbf{x}}_{ik} \pm t_{[\nu, \frac{\alpha}{2K(I-1)}}] \sqrt{\frac{\ell' w_{kk} \ell}{J\nu}}, \quad k = 1, 2, \dots, 5.$$

em que  $\ell$  é um vetor  $4 \times 1$  dos coeficientes do contraste entre as médias dos tratamentos com  $\ell' \mathbf{j} = 0$ , onde  $\mathbf{j}$  é um vetor  $4 \times 1$  de *uns*. Estes intervalos também estão apresentados na Tabel 22.

Para facilitar os cálculos na construção dos intervalos de confiança, organizamos as médias da interação Tratamento por Época, conforme Tabela 21

Tabela 21: Estimativas das médias do diâmetro do caule em relação aos tratamentos em cada época de avaliação

Tratamento	Época de avaliação					Média
	$E_1$	$E_2$	$E_3$	$E_4$	$E_5$	
$T_1$	7,94	10,10	19,64	21,35	21,52	16,11
$T_2$	8,92	12,29	21,32	25,90	26,09	18,90
$T_3$	8,28	11,57	21,42	23,96	23,27	17,70
$T_4$	6,86	7,58	13,63	15,45	15,29	11,76
Média	8,00	10,39	19,00	21,67	21,54	16,12

Observe que os intervalos de confiança para  $\Psi_1$ , na Tabela 22, não contêm a origem, o que significa dizer que este contraste não é estatisticamente nulo (ou é estatisticamente significativo) nas 5 épocas de avaliação. Já o intervalo de confiança para o contraste  $\Psi_2$ , não contém a origem apenas na época 2, o que quer dizer que apenas na época 2

Tabela 22: Intervalos com 95% de confiança simultâneos de Bonferroni para os contrastes de médias dos tratamentos em cada época de avaliação (variável)

Contraste entre as média dos tratamentos	Época de avaliação (variável)									
	$E_1$		$E_2$		$E_3$		$E_4$		$E_5$	
	$LI$	$LS$	$LI$	$LS$	$LI$	$LS$	$LI$	$LS$	$LI$	$LS$
$\Psi_1 : \frac{\mu_1 + \mu_2 + \mu_3}{3} - \mu_4$	0,73	2,31	2,49	4,99	5,50	8,83	4,10	10,23	3,82	10,51
$\Psi_2 : \frac{\mu_1 + \mu_3}{2} - \mu_2$	-1,65	0,03	-2,78	-0,13	-2,56	0,98	-4,04	2,46	-4,34	2,76
$\Psi_3 : \mu_1 - \mu_3$	-1,31	0,63	-3,00	0,06	-3,82	0,26	-5,53	1,97	-5,88	2,32
$\Psi_4 : \mu_1 - \mu_2$	-2,06	0,11	-3,90	-0,48	-3,96	0,60	-5,86	2,52	-6,26	2,90
$\Psi_5 : \mu_1 - \mu_3$	-1,43	0,75	-3,18	0,24	-4,06	0,50	-5,98	2,42	-6,36	2,80
$\Psi_6 : \mu_1 - \mu_4$	-0,01	2,17	0,81	4,23	3,73	8,29	1,81	10,21	1,43	10,59
$\Psi_7 : \mu_2 - \mu_3$	-0,45	1,73	-0,99	2,43	-2,38	2,18	-4,30	4,10	-4,68	4,48
$\Psi_8 : \mu_2 - \mu_4$	0,97	3,15	3,00	6,42	5,41	9,97	3,49	11,89	3,11	12,27
$\Psi_9 : \mu_3 - \mu_4$	0,33	2,51	2,28	5,70	5,51	10,07	3,59	11,99	3,21	12,37
$DMS(\alpha = 0,05)$	$\pm 1,09$		$\pm 1,71$		$\pm 2,28$		$\pm 4,20$		$\pm 4,58$	

$LI$  denota o limite inferior do intervalo e  $LS$  o limite superior.

o contraste é significativo. Por outro lado, o intervalo de confiança para  $\Psi_3$ , contém a origem em todas as épocas de avaliação, sendo portanto, não significativo em todas as épocas.

Observamos ainda que os intervalos de confiança para  $\Psi_5$  e  $\Psi_7$  contém a origem nas 5 épocas de avaliação, logo os contrastes não são significativos, ao contrário dos intervalos de confiança para  $\Psi_8$  e  $\Psi_9$ . Para os intervalos de confiança de  $\Psi_5$  e  $\Psi_6$ , temos que, não contém a origem apenas na época 2 e contém a origem apenas na época 1, respectivamente, seguindo as mesmas interpretações já vistas anteriormente.

## 4.4 A análise de perfil de médias

Para analisar os perfis médios dos tratamentos ao longo do ciclo da cultura, usualmente verificamos *três* hipóteses: a de paralelismo, coincidência e horizontalidade dos perfis

### 4.4.1 A hipótese de paralelismo

Definimos a matriz  $\mathbf{C}$   $(K - 1) \times K$  de contrastes, como

$$\mathbf{C} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{pmatrix}$$

e de acordo com a expressão em (2.76) calculamos, para este experimento em particular, o valor da estatística lâmba de Wilks,

$$\Lambda = \frac{|CEC'|}{|C(E + H_A)C'|} = 0,0130893.$$

Como  $\Lambda = 0,0130893 < \Lambda_{(4; 3; 15; 0,05)} = 0,219$ , então rejeitamos a hipótese, e concluimos, ao nível de 5% de significância, que os perfis do diâmetro médio do caule de girassol ao longo do ciclo da cultura, dentro de cada tratamento, não são paralelos, conforme pode ser observado no Figura 2.

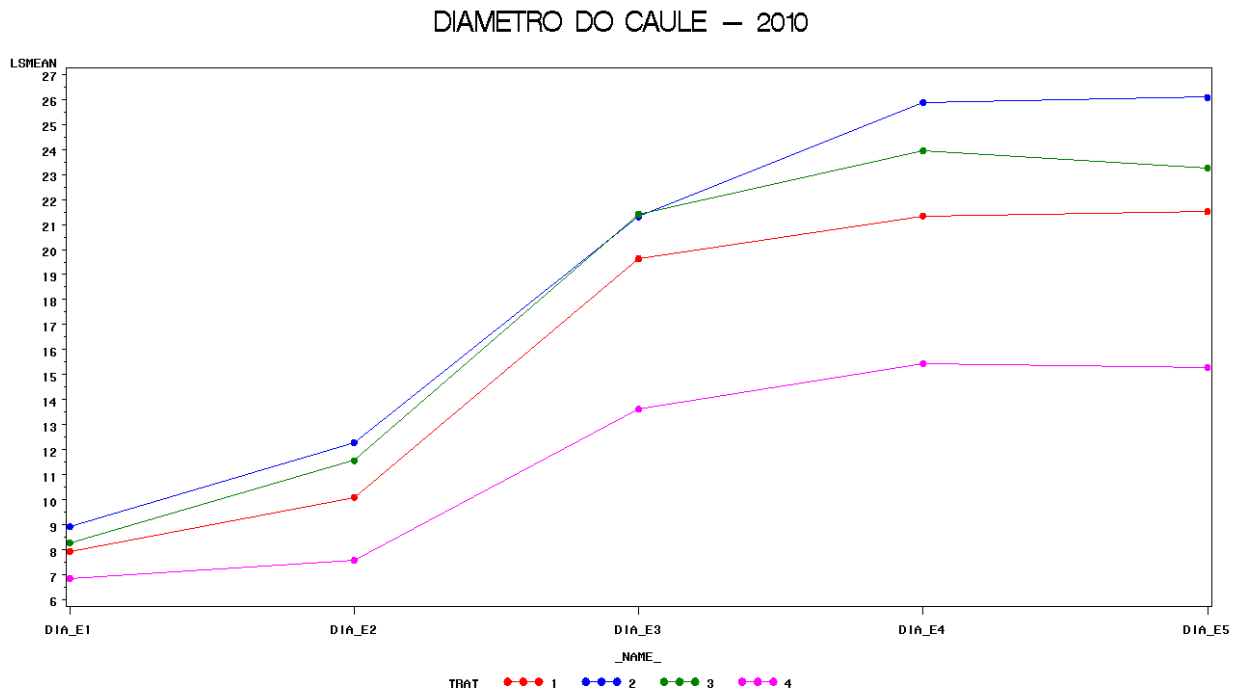


Figura 2: Gráfico dos perfis de médias dos tratamentos em função das *cinco* épocas de avaliação

#### 4.4.2 A hipótese de coincidência

Como a hipótese de paralelismo foi rejeitada, a hipótese de coincidência dos perfis, também o será. Embora não seja necessário o cálculo da estatística para o teste, o fizemos a título de ilustração, ou seja,

$$\Lambda_C = \frac{j'Ej}{j'Ej + j'H_Aj} = 0,025274.$$

Como  $\Lambda = 0,025274 < \Lambda_{(1; 3; 0,05)} = 0,603$ , rejeitamos a hipótese de coincidência dos perfis, ao nível de 5% de significância, e concluimos que pelo menos um dos perfis médios

dos diâmetros do caule de girassol difere dos demais.

### 4.4.3 A hipótese de horizontalidade

Dado que os perfis não são paralelos, obviamente não serão horizontais. Porém, a título de exercício, calculamos o valor da estatística  $T^2$ , isto é,

$$T^2 = IJ(\mathbf{C}\bar{\mathbf{y}}_{..})'(\mathbf{C}\mathbf{E}\mathbf{C}')^{-1}(\mathbf{C}\bar{\mathbf{y}}_{..}) = 3801,8547$$

Como  $T^2 = 3801,8547 > T^2_{(4;15;0,05)} = 16,296$ , então rejeitamos a hipótese de horizontalidade dos perfis das médias do diâmetro do caule do girassol em relação às épocas de avaliação dentro dos tratamentos, ao nível de 5% de significância.

## 4.5 A análise *Split-plot* com correção dos graus de liberdade

Para Box (1954) e Greenhouse e Geisser (1959), mesmo quando a hipótese de uniformidade de  $\Sigma$  é rejeitada podemos fazer uma análise univariada *Split-plot*, ajustando-se os graus de liberdade para o teste  $F$ , pelo fator  $\epsilon$ , que para este experimento encontramos o seguinte valor

$$\epsilon = \frac{[\text{tr}(\Sigma - \mathbf{J}\Sigma)]^2}{(K-1)\text{tr}(\Sigma - \mathbf{J}\Sigma/K)^2} = 0,4835$$

Assim, o valor  $F$  da estatística para testar a interação Tratamento *versus* Época,  $A \times C$ , deverá ser comparado com o percentil  $F_{[\epsilon\nu_g; \epsilon\nu; \alpha]} = F_{[5,8; 38,7; 0,05]}$ , por exemplo. Aqui, não foi realizada a análise do experimento usando esta aproximação, devemos fazê-la em outro momento.

## 4.6 Análise do experimento em blocos ao acaso com parcelas subdivididas

Ao submetermos os dados sobre o diâmetro do caule do experimento instalado em 2011 ao PROGRAMA SAS #2 do Apêndice A, observamos que o teste de Mouchly não rejeitou a hipótese de uniformidade. Sendo assim, deveremos proceder à análise de variância univariada *split-plot*, conforme pode ser vista na Seção 2.12 e subseções, nas quais encontra-se todo o embasamento inferencial para esta análise.

## 5 Considerações finais

Neste trabalho objetivou-se a fazer um estudo sobre experimentos em blocos completos ao acaso com medidas repetidas usando um experimento com a cultura do girassol nos anos de 2010 e 2011, nos quais os dados de 2010 foi feita uma análise multivariada e os dados de 2011 uma análise univariada. Na análise multivariada, chegamos a conclusão de que o diâmetro médio do caule de girassol ao longo da cinco épocas de avaliação é maior no tratamento 2 (Leirão), e menor no tratamento 4 (Testemunha) comparado com os outros três tratamentos. No entanto para os dados de 2011 daremos prosseguimento fazendo a análise univariada em um outro momento.

## Referências

- BARBIN, D. *Componentes de Variância - Teoria e Aplicações*. 2. ed. Piracicaba - SP: FEALQ - Fundação de Estudos Agrários Luiz de Queiroz, 1993. 117 p.
- BOCK, R. D. *Multivariate Analysis of Variance of Repeated Measurements*. Madison: C. W. Harris (ed.), 1963. 85-103 p.
- BOIK, R. J. A priori tests in repeated measures designs: Effects of non-sphericity. *Psychometrika*, v. 46, p. 241–255, 1981.
- BOX, G. E. P. Some theorems on quadratic forms applied in study of analysis of variance problems: II. the effect of inequality of variance and of correlation between errors in two-way classification. *Annals of Mathematical Statistics*, v. 25, p. 484–498, 1954.
- CAMPOS, H. de. *Estatística aplicada à experimentação com cana-de-açúcar*. Piracicaba - SP: FEALQ, 1984. 292 p.
- COLLIER, R. O.; MANDEVILLE, F. B.; HAYES, T. F. Estimates of test size for several procedures based on conventional ratios in repeated measure design. *Psychometrika*, v. 32, n. 32, p. 339–353, 1967.
- DAVIDSON, M. L. Univariate versus multivariate tests in repeated measures experiments. *Psychological Bulletin*, v. 77, p. 446–452, 1972.
- GOMES, F. P. *Curso de Estatística Experimental*. 4. ed. Piracicaba - SP: Nobel, 1985. 446 p.
- GREENHOUSE, S. W.; GEISSER, S. On methods in the analysis of profile data. *Psychometrika*, v. 24, p. 95–112, 1959.
- HUYNH, H. Some approximate tests in repeated measures designs. *Psychometrika*, v. 43, p. 1582–1589, 1978.
- HUYNH, H.; FELDT, L. S. Conditions under which mean square ratios in repeated measurement designs have exact  $f$  distributions. *American Statistical Association*, n. 65, p. 1582–1589, 1970.
- JOHNSON, R. A.; WICHERN, D. W. *Applied Multivariate Statistical Analysis*. 6. ed. New Jersey: Pearson Prentice Hall, 2007. 773 p.
- LUNA, J. G.; OLINDA, R. A. *Introdução a Modelos Lineares*. Campina Grande - PB: ADUEPB, 2014. 154 p. ISBN: 9788578791834.
- MAUCHLY, W. J. Significance test for a normal  $n$ -variate distribution. *Annals of Mathematical Statistics*, v. 11, n. 2, p. 204–209, 1940.

- MAXWELL, S. E.; AVERY, R. D. Small sample profile analysis with many variables. *Psychological Bulletin*, v. 92, p. 778–785, 1982.
- MONTGOMERY, D. C.; RUNGER, G. C. *Applied Statistics and Probability for Engineers*. New York: John Wiley & Sons, 1994. 895 p.
- MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. *Introduction to the theory of statistics*. 3. ed. Singapore: McGraw-Hill Book Company, Inc., 1974. 564 p.
- RENCHER, A. C.; CHRISTENSEN, W. F. *Methods of Multivariate Analysis*. Canada: John Wiley & sons, 2012. 758 p.
- ROGAN, J. C.; KESELMAN, H. J.; MENDOZA, J. L. Analysis of repeated measurements. *British Journal of Mathematical and Statistical Psychology*, v. 32, p. 269–286, 1979.
- ROHATGI, V. K. *An Introduction to Probability Theory and Mathematical Statistics*. New York: John Wiley, 1976. 684 p.
- SANTOS, M. P.; LUNA, J. G. *Análise de dados com medidas repetidas utilizando no estudo do desempenho escolar no ensino médio*. Dissertação (Trabalho de Conclusão de Curso de Graduação) — Universidade Estadual da Paraíba, Campina Grande - PB, 2014. Disponível em: <<http://dspace.bc.uepb.edu.br/jspui/handle/123456789/4180>>.
- SATTERTHWAITE, F. C. An approximate distribution of estimates of variance components. *Biometrics*, v. 2, p. 110–114, 1946.
- WILKS, S. S. Sample criteria for testing equality of means, equality of variances and equality covariances in normal multivariate distribution. *Annals of Mathematical Statistics*, v. 17, p. 257–281, 1946.

## APÊNDICE A – Coeficientes dos polinômios ortogonais

Tabela 23: Coeficientes dos polinômios ortogonais para um fator com 3 a 9 níveis.

Número de níveis	Grau	Coeficientes dos polinômios ortogonais									$K$	$M$
		$c_1$	$c_2$	$c_3$	$c_4$	$c_5$	$c_6$	$c_7$	$c_8$	$c_9$		
3	1 <sup>o</sup>	-1	0	1							2	1
	2 <sup>o</sup>	1	-2	1							6	3
4	1 <sup>o</sup>	-3	-1	1	3						20	2
	2 <sup>o</sup>	1	-1	-1	1						4	1
	3 <sup>o</sup>	-1	3	-3	1						20	10/3
5	1 <sup>o</sup>	-2	-1	0	1	2					10	1
	2 <sup>o</sup>	2	-1	-2	-1	2					14	1
	3 <sup>o</sup>	-1	2	0	-2	1					10	5/6
	4 <sup>o</sup>	1	-4	6	-4	1					70	35/12
6	1 <sup>o</sup>	-5	-3	-1	1	3	5				70	2
	2 <sup>o</sup>	5	-1	-4	-4	-1	5				84	3/2
	3 <sup>o</sup>	-5	7	4	-4	-7	5				180	5/3
	4 <sup>o</sup>	1	-3	2	2	-3	1				28	7/12
	5 <sup>o</sup>	-1	5	-10	10	-5	1				252	21/10
7	1 <sup>o</sup>	-3	-2	-1	0	1	2	3			28	1
	2 <sup>o</sup>	5	0	-3	-4	-3	0	5			84	1
	3 <sup>o</sup>	-1	1	1	0	-1	-1	1			6	1/6
	4 <sup>o</sup>	3	-7	1	6	1	-7	3			154	7/12
	5 <sup>o</sup>	-1	4	-5	0	5	-4	1			84	7/20
8	1 <sup>o</sup>	-7	-5	-3	-1	1	3	5	7		168	2
	2 <sup>o</sup>	7	1	-3	-5	-5	-3	1	7		168	1
	3 <sup>o</sup>	-7	5	7	3	-3	-7	-5	7		264	2/3
	4 <sup>o</sup>	7	-13	-3	9	9	-3	-13	7		616	7/12
	5 <sup>o</sup>	-7	23	-17	-15	15	17	-23	7		2184	7/10
9	1 <sup>o</sup>	-4	-3	-2	-1	0	1	2	3	4	60	1
	2 <sup>o</sup>	28	7	-8	-17	-20	-17	-8	7	28	2772	3
	3 <sup>o</sup>	-14	7	13	9	0	-9	-13	-7	14	990	5/6
	4 <sup>o</sup>	14	-21	-11	9	18	9	-11	-21	14	2002	7/12
	5 <sup>o</sup>	-4	11	-4	-9	0	9	4	-11	4	468	3/20



## APÊNDICE B – Programas SAS

### PROGRAS SAS #1

```

/* ===== */
/*          AS ANÁLISES INDIVIDUAIS: DIA - 2010          */
/* ----- */
/*          2010 - CAPTAÇÃO DE ÁGUA IN SITU          */
/*          */
/* TRATAMENTOS: T1- BACIA, T2- LEIRÃO, T3- SULCO, T4- TESTEMUNHA. */
/* TESTEMUNHA: sem estrutura especial para captação de água. */
/* DELINEAMENTO: Blocos Completos Casualizados (6 blocos) */
/* AVALIAÇÕES: Foram feitas 5 avaliações ao longo do ciclo da cultura. */
/*          (Avaliados aos 30, 45, 60, 70 e 80 dias após a emergência) */
/* VARIÁVEIS: DIA - diâmetro do caule da planta (mm). */
/* ===== */
OPTIONS NODATE PS=500;
DATA CA_DIA_2010;
    INPUT TRAT BLOCO DIA_E1 DIA_E2 DIA_E3 DIA_E4 DIA_E5;
DATALINES;
1 1 8.45 9.88 18.75 20.62 20.86
1 2 8.12 11.24 18.12 20.28 19.60
1 3 8.10 9.26 18.55 19.94 24.05
1 4 7.62 9.48 21.07 23.40 21.82
1 5 7.50 10.30 20.80 21.80 22.45
1 6 7.84 10.43 20.54 22.05 20.33
2 1 8.48 11.92 20.71 26.30 23.54
2 2 9.05 12.35 22.76 27.04 27.80
2 3 8.42 11.05 20.20 26.33 27.74

```

```

2 4 9.12 13.06 21.37 24.08 26.33
2 5 9.24 12.30 20.90 23.36 25.05
2 6 9.20 13.04 21.97 28.23 26.08
3 1 7.50 11.49 19.64 22.33 24.10
3 2 8.25 12.12 21.43 24.82 25.12
3 3 8.76 12.26 21.52 24.06 21.52
3 4 8.12 10.41 22.76 25.30 19.98
3 5 9.05 12.08 22.41 26.20 22.30
3 6 7.98 11.05 20.77 21.03 26.60
4 1 6.85 8.02 14.02 16.10 15.20
4 2 7.01 7.82 13.68 15.85 14.88
4 3 6.23 8.04 11.75 15.20 14.36
4 4 7.03 7.20 14.41 16.07 15.45
4 5 7.20 7.06 14.04 14.55 16.04
4 6 6.82 7.33 13.86 14.90 15.82

```

```
;
```

```
ODS HTML; ODS GRAPHICS ON;
```

```
TITLE 'DIÂMETRO DO CAULE - 2010';
```

```
PROC PRINT DATA=CA_DIA_2010;
```

```
RUN;
```

```
PROC GLM DATA=CA_DIA_2010;
```

```
CLASS TRAT BLOCO;
```

```
MODEL DIA_E1 = TRAT BLOCO / SS3; /*Só uma variável para resíduo*/
```

```
OUTPUT OUT=DIAGNOSTICO R=RES P=PRED;
```

```
RUN;
```

```
CONTRAST '(M1+M2+M3)/3 vs M4' TRAT 1 1 1 -3;
```

```
CONTRAST 'M2 vs (M1+M3)/2' TRAT -1 2 -1 0;
```

```
CONTRAST 'M3 vs M1' TRAT -1 0 1 0;
```

```
ESTIMATE '(M1+M2+M3)/3 vs M4' TRAT 1 1 1 -3 / DIVISOR=3;
```

```
ESTIMATE 'M2 vs (M1+M3)/2' TRAT -1 2 -1 0 / DIVISOR=2;
```

```
ESTIMATE 'M3 vs M1' TRAT -1 0 1 0;
```

```
MEANS TRAT / TUKEY HOVTEST=BARTLETT HOVTEST=LEVENE;
```

```
MANOVA H=TRAT H=BLOCO / PRINTH PRINTE;
```

```
REPEATED EPOCA 5 (30 45 60 70 80) POLYNOMIAL / SUMMARY PRINTE PRINTM;
```

```
LSMEANS TRAT / OUT=MEANS;
```

```
RUN;
PROC PRINT DATA=MEANS;
RUN;
PROC GPLOT DATA=MEANS;
PLOT LSMEAN*_NAME_=TRAT / OVERLAY HAXIS=1 TO 11 BY 2;
SYMBOL1 COLOR=RED INTERPOL=JOIN VALUE=DOT HEIGHT=0.8;
SYMBOL2 COLOR=BLUE INTERPOL=JOIN VALUE=DOT HEIGHT=0.8;
SYMBOL3 COLOR=GREEN INTERPOL=JOIN VALUE=DOT HEIGHT=0.8;
SYMBOL4 COLOR=MAGENTA INTERPOL=JOIN VALUE=DOT HEIGHT=0.8;
AXIS1 ORDER=(5 TO 35 BY 5) LABEL=(A=90 'MEANS');
RUN;
ODS GRAPHICS OFF;
ODS HTML CLOSE;
PROC PRINT DATA=DIAGNOSTICO;
VAR DIA_E1 PRED RES; /* Para cada variável DIA_Ei */
PROC UNIVARIATE DAT=DIAGNOSTICO NORMAL PLOT;
VAR RES;
RUN; QUIT;
```

PROGRAS SAS #2

```

/* ===== */
/*          AS ANÁLISES INDIVIDUAIS: DIA - 2011          */
/* ----- */
/*          2011 - CAPTAÇÃO DE ÁGUA IN SITU          */
/*          */
/* TRATAMENTOS: T1- BACIA, T2- LEIRÃO, T3- SULCO, T4- TESTEMUNHA. */
/* TESTEMUNHA: sem estrutura especial para captação de água. */
/* DELINEAMENTO: Blocos Completos Casualizados (6 blocos) */
/* AVALIAÇÕES: Foram feitas 5 avaliações ao longo do ciclo da cultura. */
/*          (Avaliados aos 30, 45, 60, 70 e 80 dias após a emergência) */
/* VARIÁVEIS: NF - Número de folhas; ALT - Altura das plantas (cm); */
/*          DIA - diâmetro do caule da planta (mm). */
/* ===== */
OPTIONS NODATE PS=500;
DATA CA_DIA_2011;
    INPUT TRAT BLOCO DIA_E1 DIA_E2 DIA_E3 DIA_E4 DIA_E5;
DATALINES;
1 1 8.05 14.06 18.52 24.03 24.02
1 2 7.03 15.20 21.20 22.84 23.64
1 3 9.45 15.23 20.12 22.64 23.90
1 4 7.70 15.20 21.05 23.94 24.32
1 5 8.20 16.08 20.30 21.52 23.92
1 6 7.80 14.92 19.88 24.33 22.30
2 1 10.02 16.94 24.55 26.06 25.60
2 2 9.30 16.92 24.82 27.20 25.24
2 3 8.45 19.24 24.20 24.82 27.52
2 4 9.40 21.02 25.05 26.46 24.65
2 5 8.30 15.88 26.50 26.04 27.20
2 6 8.88 18.20 23.84 26.30 26.74
3 1 8.10 16.25 21.10 23.00 26.03
3 2 8.02 15.10 24.33 26.38 24.85
3 3 7.39 16.03 20.82 27.20 26.45
3 4 7.02 15.33 23.22 25.39 25.42
3 5 6.98 15.02 21.33 24.98 26.20

```

```

3 6 7.72 14.48 23.66 26.66 25.02
4 1 7.20 12.35 13.60 17.20 16.48
4 2 7.33 14.12 16.01 16.33 16.80
4 3 7.10 13.05 13.92 15.62 15.02
4 4 6.32 12.40 16.25 16.33 16.10
4 5 6.30 12.93 13.85 15.55 15.80
4 6 6.50 12.84 14.20 16.30 17.03
;
ODS HTML; ODS GRAPHICS ON;
TITLE 'DIÂMETRO DO CAULE - 2011';
PROC PRINT DATA=CA_DIA_2011;
RUN;
PROC GLM DATA=CA_DIA_2011;
CLASS TRAT BLOCO;
MODEL DIA_E1 = TRAT BLOCO / SS3; /*Só uma variável para resíduo*/
OUTPUT OUT=DIAGNOSTICO R=RES P=PRED;
RUN;
CONTRAST '(M1+M2+M3)/3 vs M4' TRAT 1 1 1 -3;
CONTRAST 'M2 vs (M1+M3)/2' TRAT -1 2 -1 0;
CONTRAST 'M3 vs M1' TRAT -1 0 1 0;
ESTIMATE '(M1+M2+M3)/3 vs M4' TRAT 1 1 1 -3 / DIVISOR=3;
ESTIMATE 'M2 vs (M1+M3)/2' TRAT -1 2 -1 0 / DIVISOR=2;
ESTIMATE 'M3 vs M1' TRAT -1 0 1 0;
MEANS TRAT / TUKEY, /* HOVTEST=BARTLETT HOVTEST=LEVENE; SE ONE-WAY */
MANOVA H=TRAT H=BLOCO / PRINTH PRINTE;
REPEATED EPOCA 5 (30 45 60 70 80) POLYNOMIAL / SUMMARY PRINTE PRINTM;
LSMEANS TRAT / OUT=MEANS;
RUN;
PROC PRINT DATA=MEANS;
RUN;
PROC GPLOT DATA=MEANS;
PLOT LSMEAN*_NAME_=TRAT / OVERLAY HAXIS=1 TO 11 BY 2;
SYMBOL1 COLOR=RED INTERPOL=JOIN VALUE=DOT HEIGHT=0.8;
SYMBOL2 COLOR=BLUE INTERPOL=JOIN VALUE=DOT HEIGHT=0.8;
SYMBOL3 COLOR=GREEN INTERPOL=JOIN VALUE=DOT HEIGHT=0.8;

```

```
SYMBOL4 COLOR=MAGENTA INTERPOL=JOIN VALUE=DOT HEIGHT=0.8;  
AXIS1 ORDER=(5 TO 35 BY 5) LABEL=(A=90 'MEANS');  
RUN;  
ODS GRAPHICS OFF; ODS HTML CLOSE;  
PROC PRINT DATA=DIAGNOSTICO;  
VAR DIA_E1 PRED RES;  
PROC UNIVARIATE DAT=DIAGNOSTICO NORMAL PLOT;  
VAR RES;  
RUN; QUIT;
```

PROGRAS SAS #3

```

/* ===== */
/*          Cálculo da matriz E - 2010          */
/*  Aqui, a matriz X é a matriz dos resíduos e  */
/*  E = Soma de quadrados e produtos dos resíduos */
/* ===== */
OPTIONS NODATE; PS=500;
PROC IML; RESET PRINT;
X={0.68917 -0.16292 -0.16708 -0.40583 -0.04083,
   0.07167  0.64208 -1.51458 -1.40583 -2.22583,
   0.28167 -0.60792 -0.09208 -1.13083  2.15667,
  -0.29333 -0.27292  0.53042  1.49917  0.94917,
  -0.68833  0.14958  0.62542  0.63417  1.01417,
  -0.06083  0.25208  0.61792  0.80917 -1.85333,
  -0.26083 -0.31125  0.11292  0.73250 -1.93250,
   0.02167 -0.43625  1.44542  0.81250  1.40250,
  -0.37833 -1.00625 -0.12208  0.71750  1.27500,
   0.22667  1.11875 -0.84958 -2.36250  0.88750,
   0.07167 -0.03875 -0.95458 -2.34750 -0.95750,
   0.31917  0.67375  0.36792  2.44750 -0.67500,
  -0.59917 -0.02292 -1.06042 -1.30417  1.44750,
  -0.13667  0.05208  0.01208  0.52583  1.54250,
   0.60333  0.92208  1.09458  0.38083 -2.12500,
  -0.13167 -0.81292  0.43708  0.79083 -2.64250,
   0.52333  0.45958  0.45208  2.42583 -0.88750,
  -0.25917 -0.59792 -0.93542 -2.81917  2.66500,
   0.17083  0.49708  1.11458  0.97750  0.52583,
   0.04333 -0.25792  0.05708  0.06750 -0.71917,
  -0.50667  0.69208 -0.88042  0.03250 -1.30667,
   0.19833 -0.03292 -0.11792  0.07250  0.80583,
   0.09333 -0.57042 -0.12292 -0.71250  0.83083,
   0.00083 -0.32792 -0.05042 -0.43750 -0.13667};
N=NROW(X); P=NCOL(X); M=(1/N)*T(X)*J(N,1);
E=T(X)*(I(N)-(1/N)*J(N))*X;
A1=1/SQRT(2); A2=1/SQRT(6); A3=1/SQRT(12); A4=1/SQRT(20);

```

```

D={0.7071068 0 0 0, 0 0.4082483 0 0,0 0 0.2886751 0, 0 0 0 0.2236068};
M={1 -1 0 0 0, 1 1 -2 0 0, 1 1 1 -3 0, 1 1 1 1 -4}; C=D*M;
SPL=E/15; CSCL=C*SPL*T(C); LR=(DET(CSCL)/((TRACE(CSCL)/4)**4))**12;
U=LR**(2/24); U_LINHA= -(15-(2*(4**2)+4+2)/(6*4))*LOG(U);
QUI=CINV(0.95,9);
M={19.64, 21.32, 21.42, 13.63};
LC={0.3333 0.3333 0.3333 -1, 0.5 -1 0.5 0, 1 0 -1 0};
YHATC=LC*M; TC=TINV(0.9983,15); LC2=LC*T(LC);
DMSC=TC*SQRT(SPL[5,5]*LC2/5); /*VAI TROCANDO A VARIAVEL */
LINF_C=YHATC-VECDIAG(DMSC); LSUP_C=YHATC+VECDIAG(DMSC);
L={1 -1 0 0, 1 0 -1 0, 1 0 0 -1, 0 1 -1 0, 0 1 0 -1, 0 0 1 -1};
YHAT=L*M; T=TINV(0.9983,15); L2=L*T(L);
DMS=T*SQRT(SPL[5,5]*0.5); /* VAI TROCANDO A VARIAVEL */
LINF=YHAT-VECDIAG(DMS); LSUP=YHAT+VECDIAG(DMS);
EPSILON_G_G=(TRACE(SPL-(J(5)*SPL)/5)**2)/(4*(TRACE((SPL-(J(5)*SPL)/5)**2)));
/* analise de perfil */
C={1 -1 0 0 0,0 1 -1 0 0, 0 0 1 -1 0, 0 0 0 1 -1};
HA={13.385216667 31.802191667 53.418941667 69.869983333 70.814033333,
31.802191667 77.8569125 133.03507917 169.7673 169.457825,
53.418941667 133.03507917 243.11587917 291.38778333 289.79060833,
69.869983333 169.7673 291.38778333 371.34563333 372.35926667,
70.814033333 169.457825 289.79060833 372.35926667 376.42508333};
LAMB_PARALELO=DET(C*E*T(C))/DET(C*(E+HA)*T(C));
LAMB_COINCIDE=(T(J(5,1))*E*J(5,1))/((T(J(5,1))*E*J(5,1)+T(J(5,1))*HA*J(5,1)));
EPOCA={8.00, 10.39, 19.00, 21.67, 21.54};
LAMB_HORIZONTAL=4*5*T(C*EPOCA)*INV(C*E*T(C)/15)*(C*EPOCA);
QUIT;

```