



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

Anderson Ferreira Arruda de Araujo

Experimentos em Quadrados Latinos: Aspectos Teóricos com Aplicação

Campina Grande
Dezembro, 2014

Anderson Ferreira Arruda de Araujo

Experimentos em Quadrados Latinos: Aspectos Teóricos com Aplicação

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística 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 Bacharel em Estatística.

Orientador:
João Gil de Luna

Campina Grande
Dezembro, 2014

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.

A663e Araujo, Anderson Ferreira Arruda de.
Experimentos em Quadrados Latinos [manuscrito] : Aspectos Teóricos com Aplicação / Anderson Ferreira Arruda de Araujo. - 2014.

53 p. nao

Digitado.

Trabalho de Conclusão de Curso (Graduação em Estatística) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2014.

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

1. Quadrados latinos. 2. Análise de variância. 3. Estatística experimental. I. Título.

21. ed. CDD 519.5

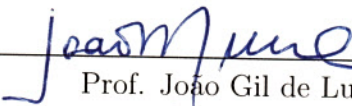
Anderson Ferreira Arruda de Araujo


Experimentos em Quadrados Latinos: Aspectos Teóricos com Aplicação

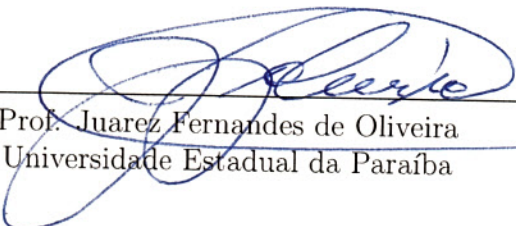
Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de Bacharel em Estatística.

Aprovado em: 17/12 /2014

Banca Examinadora:


Prof. João Gil de Luna
Orientador


Prof. Ricardo Alves de Olinda
Universidade Estadual da Paraíba


Prof. Juarez Fernandes de Oliveira
Universidade Estadual da Paraíba

Dedicatória

A Deus que me abençoa todos os dias de minha vida.

Ao meus pais Antonio Almeida de Araujo e Eliane Ferreira Arruda de Araujo.

Aos meus demais familiares que sempre me ajudaram de alguma forma.

Ao meu orientador João Gil de Luna.

Agradecimentos

Agradecer primeiramente a DEUS que me fortalece todos os dias.

Aos meus pais, Antonio Almeida de Araujo e Eliane Ferreira Arruda de Araújo, que sempre me incentivaram a estudar, investiram e acreditaram em mim.

A minha avó, Alzenir Ferreira de Arruda que sempre me incentivou a estudar e que eu sei que a qualquer momento está ali para me ajudar.

Aos meus tios e tias, que de alguma forma me ajudaram a chegar onde cheguei.

A minha amiga, Severina Francielly Silva Tomaz, pelo apoio nos momentos em que precisei.

A minha prima, Pamela Yonally Gomes Araújo, que tem sido uma grande amiga e irmã.

Ao meu amigo, Jucenio Gomes de Araujo, que também me deu muito incentivo para chegar onde cheguei.

Ao Prof., Dr. João Gil de Luna, que não só é um professor ou orientador, mas um amigo que em todos os momentos me ajudou.

Aos outros professores que estiveram comigo nesta caminhada. Não vou citar nomes para não pecar em esquecer algum. Mas vai aqui o meu agradecimento.

Resumo

A pesquisa experimental é utilizada em várias áreas do conhecimento, em que o pesquisador aplica na amostra diferentes tipos de estímulos, também chamados de tratamentos, para verificar se os efeitos produzidos geram diferentes resultados na sua resposta. Muitas vezes esta resposta pode ser interferida por outro fator, neste caso a área na qual foi alocado os tratamentos é heterogênea sendo necessário a utilização de blocos para retirar o efeito do solo, para que não venha interferir na resposta. Neste trabalho, será abordado o delineamento em quadrado latino que é utilizado quando se tem dois fatores que podem influenciar na resposta. Aqui apresenta-se um possível desenho do experimento no campo, a organização dos dados em uma tabela e o modelo estatístico do experimento. Na seqüência encontra-se as estimativas dos parâmetros do modelo, suas somas de quadrados e quadrados médios. Encontra-se também suas esperanças e variâncias, denota-se suas distribuições de probabilidade, apresenta-se o quadro da análise de variância e, por fim, aplica-se toda a teoria desenvolvida ao longo do trabalho em um exemplo prático.

Palavras chaves: Quadrados latinos, Análise de variância, Estatística experimental.

Abstract

Experimental research is used in various areas of knowledge, in which the researcher is involved in the sample different types of stimuli that are called treatments to see if the effects produced generate different results in your reply. Often this response may be interfered with by another factor. For example, by verifying the effect of insecticides on plants, it was found that the location in which the treatments were allocated are clay and sandy soil, in which case the area in which the treatments were allocated is heterogeneous blocks will be necessary to consider how to remove the effect of soil that will not affect the response. In this paper we address the Latin square design, which is used when it is found that two factors may influence the response. Here we present a possible experiment design in the field, the organization of data in a table, the statistical model of the experiment, find the parameter estimates of the model of their mean squares and square sums, find their hopes and variances and denote their probability distributions we present the framework of analysis of variance and finally, we will apply all the theory developed over work on a practical example.

Key words: Latin squares, variance analysis, experimental Statistics.

Sumário

1	Introdução	p. 9
2	Fundamentação Teórica	p. 11
2.1	Um possível desenho do experimento no campo	p. 11
2.2	Organização dos dados experimentais em uma tabela	p. 12
2.3	Modelo estatístico do experimento	p. 13
2.4	Estimação dos parâmetros do modelo	p. 14
2.5	Decomposição da variabilidade total	p. 18
2.6	Distribuição de probabilidade dos estimadores	p. 22
2.6.1	Distribuição de probabilidade de $\bar{y}_{..}$ estimador de μ	p. 22
2.6.2	Distribuição de probabilidade da correção para média, C	p. 23
2.6.3	Distribuição de probabilidade de a_i o estimador de α_i	p. 24
2.6.4	Distribuição de probabilidade da SQLinhas	p. 25
2.6.5	Distribuição de probabilidade de b_j o estimador de β_j	p. 25
2.6.6	Distribuição de probabilidade da SQColunas	p. 26
2.6.7	Distribuição de probabilidade de t_k o estimador de τ_k	p. 27
2.6.8	Distribuição de probabilidade da SQTratamentos	p. 28
2.6.9	Distribuição de probabilidade de m_i , estimador de $\mu_i = \mu + \alpha_i$, de m_j , estimador de $\mu_j = \mu + \beta_j$ e de m_k , estimador de $\mu_k = \mu + \tau_k$.	p. 28
2.6.10	Distribuição de probabilidade de e_{ijk} , o estimador de ϵ_{ijk}	p. 29
2.6.11	Distribuição de probabilidade de contrastes de interesse	p. 30
2.6.12	Distribuição de probabilidade da soma de quadrados dos contrastes.	p. 32

2.7	Esperança das somas de quadrados e dos quadrados médios	p. 33
2.7.1	Esperança das somas de quadrado total	p. 33
2.7.2	Esperança das somas de quadrado e do quadrado médio das linhas	p. 35
2.7.3	Esperança das somas de quadrado e do quadrado médio das colunas	p. 36
2.7.4	Esperança das somas de quadrado e do quadrado médio dos tra- tamentos	p. 37
2.7.5	Esperança das somas de quadrado e do quadrado médio dos con- trastes de interesse	p. 38
2.7.6	Esperança das somas de quadrado e do quadrado médio dos resíduos	p. 38
2.8	Análises estatísticas	p. 39
2.9	Estruturação da tabela da análise de variância - ANOVA	p. 42
2.10	Comparação de médias duas a duas dos tratamentos, das linhas e das colunas	p. 42
3	Aplicação em um exemplo real	p. 44
3.1	Um exemplo de aplicação	p. 44
3.1.1	Cálculo das somas de quadrado e análise de variância	p. 45
3.1.2	Comparação de médias duas a duas pelo teste de tukey	p. 48
3.1.3	Comprovação da idoneidade do modelo	p. 49
4	Conclusão	p. 51
	Referências	p. 52

1 Introdução

No início do século XX, Ronald Aylmer Fisher (1890-1962), um jovem matemático do colégio *Caius de Camdridge*, iniciou o desenvolvimento do ramo da estatística relacionado ao planejamento e a análise de experimentos. Fisher lançou os fundamentos modernos da pesquisa experimental na estação de *Rothamsted* e em outras instituições de pesquisa, introduzindo diversas técnicas de análise de dados. Entre elas, a análise de variância que passou a ser amplamente utilizada na análise estatística de dados de experimentos (SANTOS, 2010).

Os delineamentos são planejados de forma que a variação do acaso seja reduzida ao máximo possível. Os principais delineamentos são: Inteiramente Casualizado, Blocos Completos Casualizados e Quadrado Latino (SANTOS, 2010). Nesse trabalho nos detemos a teoria do delineamento em quadrado latino.

Em alguns campos de aplicação é bastante comum ter dois diferentes critérios qualitativos para agrupar unidades em blocos sendo assim em vez das unidades serem conceitualmente agrupadas em conjuntos ou blocos pode ser razoável considerar as unidades como dispostas de uma forma bidimensional (COX; REID, 2002). Segundo Cox e Reid (2002), a primeira vez de que se tem registro de que alguém pensou em quadrado latino foi em 1639, em um jogo de baralho. O primeiro matemático que publicou algo sobre quadrado latino foi Leonhard Euler em 1783, texto que se referia à aplicações à estatística. O seu nomeo deve-se ao fato de Euler usar letras latinas para os quadrados nesta obra (ALEGRI, 2006).

Na literatura especializada, não é comum encontrar materiais relacionados aos desdobramentos dos graus de liberdade e correspondentes somas de quadrados, quadrado médio, esperanças e variâncias dos estimadores do parâmetro do modelo. Desta forma, torna-se difícil a aplicabilidade da técnica em pesquisas envolvendo outras distribuições de probabilidade para o erro experimental associado ao modelo. No entanto, este delineamento é de extrema importância para estatística experimental. Pois ele é usado quando verifica-se a existencia de dois tipos de fatores que podem interferir na resposta (linhas e colunas).

Para isso, deve-se utilizar um modelo estatístico que descreva as observações adequadamente, com o objetivo de encontrar os estimadores dos parâmetros do modelo e suas somas de quadrado. É necessário mostrar que o modelo segue uma distribuição normal e as somas de quadrados seguem distribuição de qui-quadrado, definir os valores esperados das somas de quadrados, as estatísticas de testes utilizadas na contrastação das hipóteses de interesse para tratamento, linhas e colunas. A análise de variância deverá ser utilizada como análises complementares, testes de comparações múltiplas das médias dos tratamentos, linhas e colunas ou contrastes de média de interesse, ou ainda desdobramento dos graus de liberdade das somas de quadrados em componentes polinomiais.

2 Fundamentação Teórica

Os delineamentos em quadrado latino levam em consideração os princípios básicos da repetição, casualização e controle local em duas direções (linhas e colunas).

É um delineamento pouco utilizado na prática por ser mais restritivo que os experimentos em blocos ao acaso, pois o número de linhas e colunas devem ser os mesmos, devido ao número de repetições serem iguais ao número de tratamentos. Em geral, não se usam os quadrados latinos com mais de oito tratamentos, pois o número de unidades experimentais cresce exageradamente à medida que aumentamos o número de tratamentos dificultando a condução do experimento a nível de campo. Por outro lado os quadrados latinos 2x2, 3x3 e 4x4 encerram tão poucas unidades experimentais, que só se justificam se os experimentos incluírem vários quadrados latinos. (LUNA, 2010)

2.1 Um possível desenho do experimento no campo

Vamos considerar um experimento em quadrado latino 5x5, isso é, 5 linhas e 5 colunas em que serão alocados 5 tratamentos. Cada linha e cada coluna devem conter todos os tratamentos sem repetições. Assim, o experimento poderá ter um desenho no campo conforme apresentado na **Figura 1**.

	<i>Coluna₁</i>	<i>Coluna₂</i>	<i>Coluna₃</i>	<i>Coluna₄</i>	<i>Coluna₅</i>
<i>Linha₁</i>	<i>T₂</i>	<i>T₁</i>	<i>T₅</i>	<i>T₃</i>	<i>T₄</i>
<i>Linha₂</i>	<i>T₄</i>	<i>T₅</i>	<i>T₂</i>	<i>T₁</i>	<i>T₃</i>
<i>Linha₃</i>	<i>T₁</i>	<i>T₄</i>	<i>T₃</i>	<i>T₂</i>	<i>T₅</i>
<i>Linha₄</i>	<i>T₃</i>	<i>T₂</i>	<i>T₄</i>	<i>T₅</i>	<i>T₁</i>
<i>Linha₅</i>	<i>T₅</i>	<i>T₃</i>	<i>T₁</i>	<i>T₄</i>	<i>T₂</i>

Figura 1. Croqui do experimento

2.2 Organização dos dados experimentais em uma tabela

A organização dos dados em uma tabela apropriada, facilitará o estudo do experimento posteriormente no encontro de suas estimativas e no teste de comparação de médias. A tabela será organizada por linhas e colunas com suas respectivas somas e médias. Considerando que a variável resposta em estudo seja y , da forma apresentada na Tabela 1.

Tabela 1: Organização dos dados coletados no experimento.

Linhas (L_i)	Colunas (C_j)					Total $y_{i..}$	Media $\bar{y}_{i..}$
	C_1	C_2	C_3	C_4	C_5		
L_1	y_{112}	y_{121}	y_{135}	y_{143}	y_{154}	$y_{1..}$	$\bar{y}_{1..}$
L_2	y_{214}	y_{225}	y_{232}	y_{241}	y_{253}	$y_{2..}$	$\bar{y}_{2..}$
L_3	y_{311}	y_{324}	y_{333}	y_{342}	y_{355}	$y_{3..}$	$\bar{y}_{3..}$
L_4	y_{413}	y_{422}	y_{434}	y_{445}	y_{451}	$y_{4..}$	$\bar{y}_{4..}$
L_5	y_{515}	y_{523}	y_{531}	y_{544}	y_{552}	$y_{5..}$	$\bar{y}_{5..}$
Total $y_{.j}$	$y_{.1}$	$y_{.2}$	$y_{.3}$	$y_{.4}$	$y_{.5}$	$y_{...}$	-
Media $\bar{y}_{.j}$	$\bar{y}_{.1}$	$\bar{y}_{.2}$	$\bar{y}_{.3}$	$\bar{y}_{.4}$	$\bar{y}_{.5}$	-	$\bar{y}_{...}$

em que

y_{ijk} é a observação obtida da unidade experimental localizada na linha i e na coluna j , que recebeu o tratamento k .

$y_{...} = \sum_i \sum_{j,k} y_{ijk}$ é o total geral das observações no experimento.

$y_{i..} = \sum_{j,k} y_{ijk}$ é a soma dos valores observados na linha i .

$y_{.j} = \sum_{i,k} y_{ijk}$ é a soma dos valores observados na coluna j .

$y_{..k} = \sum_{i,j} y_{ijk}$ é a soma dos valores observados que receberam o tratamento k , isto é,

$$y_{121} + y_{241} + y_{311} + y_{451} + y_{531} = y_{.1}$$

$$y_{112} + y_{232} + y_{342} + y_{422} + y_{552} = y_{.2}$$

⋮

$$y_{135} + y_{225} + y_{355} + y_{445} + y_{515} = y_{..5}.$$

$\bar{y}_{...}, \bar{y}_{i..}, \bar{y}_{.j.}$ e $\bar{y}_{..k}$, são respectivamente, a média geral, a média da i -ésima linha, a média da j -ésima coluna e a média da k -ésimo tratamento.

2.3 Modelo estatístico do experimento

Um delineamento em quadrado latino contém os seguintes fatores: o efeito do tratamento (τ_k) e dos fatores secundários linhas (α_i) e colunas (β_j). Supõe-se que não existe interação entre os três fatores. Considerando-se o modelo empregado é um modelo aditivo. Se considerarmos que os três fatores são de efeito fixo, o modelo estatístico para esse delineamento será:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \tau_k + \epsilon_{ijk} \begin{cases} i = 1, 2, \dots, R, \\ j = 1, 2, \dots, R, \\ k = 1, 2, \dots, R. \end{cases} \quad (2.1)$$

Para o qual,

y_{ijk} representa a observação correspondente a i -ésima linha, j -ésima coluna e que recebeu o k -ésimo tratamento,

μ é a média geral,

α_i é o efeito fixo produzido pela i -ésima linha sobre a variável resposta,

β_j é o efeito fixo produzido pela j -ésima coluna sobre a variável resposta,

τ_k é o efeito fixo produzido pelo k -ésimo tratamento sobre a variável resposta,

ϵ_{ijk} representa o erro experimental aleatório relacionado a cada observação, supondo que é distribuída normalmente com média *zero* e variância comum σ^2 e denotaremos por $\epsilon_{ijk} \sim N(0, \sigma^2)$.

Neste texto, utiliza-se a notação μ_i , μ_j e μ_k para denotar as médias das linhas, das colunas e dos tratamentos, respectivamente, ou seja,

$$\mu_i = \mu + \alpha_i, \quad \mu_j = \mu + \beta_j \quad \text{e} \quad \mu_k = \mu + \tau_k.$$

Como conseqüências, tem-se.

1. Sendo ϵ_{ijk} variáveis aleatórias independentes, α_i , β_j e τ_k de efeitos fixos, então,

$$\begin{aligned} E(\mu) &= \mu, & E(\mu^2) &= \mu^2, & E(\alpha_i) &= \alpha_i, & E(\alpha_i^2) &= \alpha_i^2, & E(\beta_j) &= \beta_j, \\ E(\beta_j^2) &= \beta_j^2, & E(\tau_k) &= \tau_k, & E(\tau_k^2) &= \tau_k^2, & E(\epsilon_{ijk}) &= 0, & E(\epsilon_{ijk}^2) &= \sigma^2 \end{aligned}$$

2. Devido a hipótese de independência dos erros, tem-se que:

$$E(\epsilon_{ijk}\epsilon_{dfg}) = E(\epsilon_{ijk})E(\epsilon_{dfg}) = 0, \quad \forall i \neq d, j \neq f \text{ e } k \neq g.$$

3. Os efeitos das linhas, colunas e tratamentos são definidos como desvios das médias da linha, da coluna e do tratamento em relação à média geral, isto é, $\alpha_i = \mu_i - \mu$, $\beta_j = \mu_j - \mu$ e $\tau_k = \mu_k - \mu$ e, sendo assim, teremos $\sum_i \alpha_i = \sum_j \beta_j = \sum_k \tau_k = 0$.
4. As observações y_{ijk} são variáveis aleatórias independentes e identicamente distribuídas como uma normal de média $E(y_{ijk}) = \mu + \alpha_i + \beta_j + \tau_k = \mu_i + \mu_j + \mu_k - 2\mu$ e variância comum σ^2 , denotada por $y_{ijk} \sim N(\mu_i + \mu_j + \mu_k - 2\mu, \sigma^2)$.

As relações citadas nos itens de 1 a 4 são de grande utilidade no processo de estimação dos parâmetros e no estudo das distribuições de probabilidade associadas.

2.4 Estimação dos parâmetros do modelo

O conjunto de dados amostrais em um delineamento em quadrado latino pode ser representado da seguinte forma:

$$y_{ijk} = m + a_i + b_j + t_k + e_{ijk}, \quad \begin{cases} i = 1, 2, \dots, R, \\ j = 1, 2, \dots, R, \\ k = 1, 2, \dots, R. \end{cases} \quad (2.2)$$

Em que,

y_{ijk} representa a observação correspondente a i -ésima linha, j -ésima coluna e o k -ésimo tratamento,

m é o estimador de μ , a média geral,

a_i é o estimador de α_i , o efeito da linha i sobre a variável resposta,

b_j é o estimador de β_j , o efeito da coluna j sobre a variável resposta,

t_k é o estimador de τ_k , o efeito do tratamento k sobre a variável resposta,

e_{ijk} é o estimador de ϵ_{ijk} .

Sendo $y_{ijk} = m + a_i + b_j + t_k + e_{ijk}$, então $e_{ijk} = y_{ijk} - m - a_i - b_j - t_k$.

Portanto, considerando-se o caso de um quadrado latino 5x5, tem-se que:

$$y_{311} = m + a_3 + b_1 + t_1 + e_{311} \Rightarrow e_{311} = y_{311} - m - a_3 - b_1 - t_1,$$

$$y_{121} = m + a_1 + b_2 + t_1 + e_{121} \Rightarrow e_{121} = y_{121} - m - a_1 - b_2 - t_1,$$

$$y_{531} = m + a_5 + b_3 + t_1 + e_{531} \Rightarrow e_{531} = y_{531} - m - a_5 - b_3 - t_1,$$

$$y_{241} = m + a_2 + b_4 + t_1 + e_{241} \Rightarrow e_{241} = y_{241} - m - a_2 - b_4 - t_1,$$

$$y_{451} = m + a_4 + b_5 + t_1 + e_{451} \Rightarrow e_{451} = y_{451} - m - a_4 - b_5 - t_1,$$

$$y_{112} = m + a_1 + b_1 + t_2 + e_{112} \Rightarrow e_{112} = y_{112} - m - a_1 - b_1 - t_2,$$

$$y_{422} = m + a_4 + b_2 + t_2 + e_{422} \Rightarrow e_{422} = y_{422} - m - a_4 - b_2 - t_2,$$

$$y_{232} = m + a_2 + b_3 + t_2 + e_{232} \Rightarrow e_{232} = y_{232} - m - a_2 - b_3 - t_2,$$

$$y_{342} = m + a_3 + b_4 + t_2 + e_{342} \Rightarrow e_{342} = y_{342} - m - a_3 - b_4 - t_2,$$

$$y_{552} = m + a_5 + b_5 + t_2 + e_{552} \Rightarrow e_{552} = y_{552} - m - a_5 - b_5 - t_2,$$

$$y_{413} = m + a_4 + b_1 + t_3 + e_{413} \Rightarrow e_{413} = y_{413} - m - a_4 - b_1 - t_3,$$

$$y_{523} = m + a_5 + b_2 + t_3 + e_{523} \Rightarrow e_{523} = y_{523} - m - a_5 - b_2 - t_3,$$

$$y_{333} = m + a_3 + b_3 + t_3 + e_{333} \Rightarrow e_{333} = y_{333} - m - a_3 - b_3 - t_3,$$

$$y_{143} = m + a_1 + b_4 + t_3 + e_{143} \Rightarrow e_{143} = y_{143} - m - a_1 - b_4 - t_3,$$

$$y_{253} = m + a_2 + b_5 + t_3 + e_{253} \Rightarrow e_{253} = y_{253} - m - a_2 - b_5 - t_3,$$

$$y_{214} = m + a_2 + b_1 + t_4 + e_{214} \Rightarrow e_{214} = y_{214} - m - a_2 - b_1 - t_4,$$

$$y_{324} = m + a_3 + b_2 + t_4 + e_{324} \Rightarrow e_{324} = y_{324} - m - a_3 - b_2 - t_4,$$

$$y_{434} = m + a_4 + b_3 + t_4 + e_{434} \Rightarrow e_{434} = y_{434} - m - a_4 - b_3 - t_4,$$

$$y_{544} = m + a_5 + b_4 + t_4 + e_{544} \Rightarrow e_{544} = y_{544} - m - a_5 - b_4 - t_4,$$

$$y_{154} = m + a_1 + b_5 + t_4 + e_{154} \Rightarrow e_{154} = y_{154} - m - a_1 - b_5 - t_4,$$

$$y_{515} = m + a_5 + b_1 + t_5 + e_{515} \Rightarrow e_{515} = y_{515} - m - a_5 - b_1 - t_5,$$

$$y_{225} = m + a_2 + b_2 + t_5 + e_{225} \Rightarrow e_{225} = y_{225} - m - a_2 - b_2 - t_5,$$

$$y_{135} = m + a_1 + b_3 + t_5 + e_{135} \Rightarrow e_{135} = y_{135} - m - a_1 - b_3 - t_5,$$

$$y_{445} = m + a_4 + b_4 + t_5 + e_{445} \Rightarrow e_{445} = y_{445} - m - a_4 - b_4 - t_5,$$

$$y_{355} = m + a_3 + b_5 + t_5 + e_{355} \Rightarrow e_{355} = y_{355} - m - a_3 - b_5 - t_5.$$

Usando o método dos mínimos quadrados para encontrar os estimadores dos parâmetros μ , α_i , β_j e τ_k . Segundo Mood *et al.* (1974), o método consiste em minimizar a soma dos quadrados dos resíduos, e_{ijk} .

Seja z a soma de quadrado dos erros, isso é,

$$z = \sum_i \sum_{j,k} e_{ijk}^2 = e_{311}^2 + \dots + e_{451}^2 + e_{112}^2 + \dots + e_{552}^2 + e_{413}^2 + \dots + e_{253}^2 + e_{214}^2 + \dots + e_{154}^2 + e_{515}^2 + \dots + e_{355}^2$$

Derivando-se z parcialmente em relação a cada um dos parâmetros e igualando a zero, encontra-se:

$$\begin{aligned} \frac{\partial z}{\partial m} &= \sum_i \sum_{j,k} \frac{\partial}{\partial m} (y_{ijk} - m - a_i - b_j - t_k)^2 \\ &= 2 \sum_i \sum_{j,k} (y_{ijk} - m - a_i - b_j - t_k)(-1) \\ &= -2 \left(y_{\dots} - R^2 m - R \sum_i a_i - R \sum_j b_j - R \sum_k t_k \right) \end{aligned}$$

Sendo $\frac{\partial z}{\partial m} = 0$, tem-se que:

$$R^2 m + R \sum_i a_i + R \sum_j b_j + R \sum_k t_k = y_{\dots} \quad (2.3)$$

$$\begin{aligned} \frac{\partial z}{\partial a_1} &= -2 (y_{121} - m - a_1 - b_2 - t_1 + y_{112} - m - a_1 - b_1 - t_2 + y_{143} - m - a_1 - b_4 - t_3 \\ &\quad + y_{154} - m - a_1 - b_5 - t_4 + y_{135} - m - a_1 - b_3 - t_5) \\ &= -2 \left(y_{1..} - 5m - 5a_1 - \sum_j b_j - \sum_k t_k \right) \end{aligned}$$

Genericamente,

$$\frac{\partial z}{\partial a_i} = -2 \left(y_{i..} - Rm - Ra_i - \sum_j b_j - \sum_k t_k \right)$$

Sendo $\frac{\partial z}{\partial a_i} = 0$, tem-se que:

$$Rm + Ra_i + \sum_j b_j + \sum_k t_k = y_{i..} \quad (2.4)$$

De modo similar,

$$\begin{aligned}\frac{\partial z}{\partial b_1} &= -2(y_{311} - m - a_3 - b_1 - t_1 + y_{112} - m - a_1 - b_1 - t_2 + y_{413} - m - a_4 - b_1 - t_3 \\ &\quad + y_{214} - m - a_3 - b_1 - t_4] + y_{515} - m - a_5 - b_1 - t_5) \\ &= -2\left(y_{..1} - 5m - \sum_i a_i - 5b_j - \sum_k t_k\right)\end{aligned}$$

Genericamente,

$$\frac{\partial z}{\partial b_j} = -2\left(y_{.j} - Rm - \sum_i a_i - Rb_j - \sum_k t_k\right)$$

Sendo $\frac{\partial z}{\partial b_j} = 0$, tem-se que:

$$Rm + \sum_i a_i + Rb_j + \sum_k t_k = y_{.j}. \quad (2.5)$$

$$\begin{aligned}\frac{\partial z}{\partial t_1} &= -2(y_{241} - m - a_2 - b_4 - t_1 + y_{531} - m - a_5 - b_3 - t_1 + y_{121} - m - a_1 - b_2 - t_1 \\ &\quad + y_{311} - m - a_3 - b_1 - t_1 + y_{451} - m - a_4 - b_5 - t_1) \\ &= -2\left(y_{..1} - 5m - \sum_i a_i - \sum_j b_j - 5t_1\right)\end{aligned}$$

Genericamente

$$\frac{\partial z}{\partial t_k} = -2\left(y_{..k} - Rm - \sum_i a_i - \sum_j b_j - Rt_k\right).$$

Sendo $\frac{\partial z}{\partial t_k} = 0$, tem-se que:

$$Rm + \sum_i a_i + \sum_j b_j + Rt_k = y_{..k}. \quad (2.6)$$

As equações (2.3) (2.4) (2.5) e (2.6) compõem o Sistema de Equações Normais (S.E.N), isto é:

$$\begin{cases} R^2m + R\sum_i a_i + R\sum_j b_j + R\sum_k t_k = y_{..} \\ Rm + Ra_i + \sum_i b_j + \sum_k t_k = y_{i..} \\ Rm + \sum_i a_i + Rb_j + \sum_k t_k = y_{.j}. \\ Rm + \sum_i a_i + \sum_j b_j + Rt_K = y_{..k} \end{cases} \quad (2.7)$$

O sistema em (2.7) é inconsistente, uma vez que é composto por 4 equações e $3R + 1 > 4$

incógnitas. Ou seja, tem menos equações do que incógnitas. Para resolvê-lo, impomos convenientemente as restrições $\sum_i a_i = \sum_j b_j = \sum_k t_k = 0$. Sendo assim, teremos:

$$\begin{aligned} m &= \frac{y_{...}}{R^2} = \bar{y}_{...}, \text{ é o estimador de } \mu, \text{ a média geral,} \\ a_i &= \frac{y_{i..}}{R} - \frac{y_{...}}{R^2} = \bar{y}_{i..} - \bar{y}_{...}, \text{ é o estimador de } \alpha_i, \text{ o efeito da } i\text{-ésima linha,} \\ b_j &= \frac{y_{.j.}}{R} - \frac{y_{...}}{R^2} = \bar{y}_{.j.} - \bar{y}_{...}, \text{ é o estimador de } \beta_j, \text{ o efeito da } j\text{-ésima coluna,} \\ t_k &= \frac{y_{..k}}{R} - \frac{y_{...}}{R^2} = \bar{y}_{..k} - \bar{y}_{...}, \text{ é o estimador de } \tau_k, \text{ o efeito do } k\text{-ésimo tratamento,} \end{aligned} \quad (2.8)$$

Denotaremos como m_i , m_j e m_k os estimadores de μ_i , μ_j e μ_k , respectivamente. Daí, teremos:

$$\begin{aligned} m_i &= m + a_i = \bar{y}_{...} + \bar{y}_{i..} - \bar{y}_{...} = \bar{y}_{i..}, \text{ é o estimador da média da } i\text{-ésima linha,} \\ m_j &= m + b_j = \bar{y}_{...} + \bar{y}_{.j.} - \bar{y}_{...} = \bar{y}_{.j.}, \text{ é o estimador da média da } j\text{-ésima coluna,} \\ m_k &= m + t_k = \bar{y}_{...} + \bar{y}_{..k} - \bar{y}_{...} = \bar{y}_{..k}, \text{ é o estimador da média do } k\text{-ésimo tratamento.} \\ \hat{y}_{ijk} &= \bar{y}_{i..} + \bar{y}_{.j.} + \bar{y}_{..k} - 2\bar{y}_{...}, \text{ é o estimador de uma observação } y_{ijk}, \\ e_{ijk} &= y_{ijk} - \hat{y}_{ijk}, \text{ é o estimador do erro } \epsilon_{ijk}. \end{aligned} \quad (2.9)$$

2.5 Decomposição da variabilidade total

Para obter as partes componentes da variabilidade total, parte-se da soma dos quadrados dos resíduos, ou seja,

$$\begin{aligned} \sum_i \sum_{j,k} e_{ijk}^2 &= \sum_i \sum_{j,k} (y_{ijk} - m - a_i - b_j - t_k)^2 \\ &= \sum_i \sum_{j,k} [(y_{ijk} - \bar{y}_{...}) - (\bar{y}_{i..} - \bar{y}_{...}) - (\bar{y}_{.j.} - \bar{y}_{...}) - (\bar{y}_{..k} - \bar{y}_{...})]^2 \\ &= \underbrace{\sum_i \sum_{j,k} (y_{ijk} - \bar{y}_{...})^2}_{\text{(I)}} + R \underbrace{\sum_i (\bar{y}_{i..} - \bar{y}_{...})^2}_{\text{(II)}} + R \underbrace{\sum_j (\bar{y}_{.j.} - \bar{y}_{...})^2}_{\text{(III)}} + R \underbrace{\sum_k (\bar{y}_{..k} - \bar{y}_{...})^2}_{\text{(IV)}} \\ &\quad - 2 \underbrace{\sum_i \sum_{j,k} (y_{ijk} - \bar{y}_{...})(\bar{y}_{i..} - \bar{y}_{...})}_{\text{(V)}} - 2 \underbrace{\sum_j \sum_{i,k} (y_{ijk} - \bar{y}_{...})(\bar{y}_{.j.} - \bar{y}_{...})}_{\text{(VI)}} \\ &\quad - 2 \underbrace{\sum_k \sum_{i,j} (y_{ijk} - \bar{y}_{...})(\bar{y}_{..k} - \bar{y}_{...})}_{\text{(VII)}} - 2 R \underbrace{\sum_i \sum_j (\bar{y}_{i..} - \bar{y}_{...})(\bar{y}_{.j.} - \bar{y}_{...})}_{\text{(VIII)}} \end{aligned}$$

$$-2R \underbrace{\sum_i \sum_k (\bar{y}_{i..} - \bar{y}_{...})(\bar{y}_{..k} - \bar{y}_{...})}_{(IX)} - 2R \underbrace{\sum_j \sum_k (\bar{y}_{.j.} - \bar{y}_{...})(\bar{y}_{..k} - \bar{y}_{...})}_{(X)}.$$

Desenvolvendo algebricamente os termos de (I) a (X) obteremos os seguintes resultados

$$\begin{aligned} (I) &= \sum_i \sum_{j,k} (y_{ijk} - \bar{y}_{...})^2 = \sum_i \sum_{j,k} y_{ijk}^2 + \sum_i \sum_{j,k} \bar{y}_{...}^2 - 2 \sum_i \sum_{j,k} y_{ijk} \bar{y}_{...} \\ &= \sum_i \sum_{j,k} y_{ijk}^2 + R^2 \bar{y}_{...}^2 - 2 \bar{y}_{...} \sum_i \sum_{j,k} y_{ijk} \\ &= \sum_i \sum_{j,k} y_{ijk}^2 + \frac{y_{...}^2}{R^2} - 2 \frac{y_{...}^2}{R^2} \\ &= \sum_i \sum_{j,k} y_{ijk}^2 - \frac{y_{...}^2}{R^2} = \sum_i \sum_{j,k} y_{ijk}^2 - C, \end{aligned}$$

em que $C = \frac{(y_{...})^2}{R^2}$.

Em seguida, tomamos

$$\begin{aligned} (II) &= R \sum_i (\bar{y}_{i..} - \bar{y}_{...})^2 = R \sum_i (\bar{y}_{i..}^2 + \bar{y}_{...}^2 - 2\bar{y}_{i..}\bar{y}_{...}) \\ &= R \sum_i \left(\frac{y_{i..}^2}{R^2} + \frac{y_{...}^2}{R^4} - 2 \frac{y_{...}}{R^2} \frac{y_{i..}}{R} \right) \\ &= \sum_i \frac{y_{i..}^2}{R} + \frac{R^2 y_{...}^2}{R^4} - 2 \frac{R y_{...}^2}{R^3} \\ &= \frac{1}{R} \sum_i y_{i..}^2 - \frac{(y_{...})^2}{R^2} = \frac{1}{R} \sum_i y_{i..}^2 - C. \end{aligned}$$

De maneira similar,

$$\begin{aligned} (III) &= R \sum_j (\bar{y}_{.j.} - \bar{y}_{...})^2 = R \sum_j (\bar{y}_{.j.}^2 + \bar{y}_{...}^2 - 2\bar{y}_{.j.}\bar{y}_{...}) \\ &= R \sum_j \left(\frac{y_{.j.}^2}{R^2} + \frac{y_{...}^2}{R^4} - 2 \frac{y_{...}}{R^2} \frac{y_{.j.}}{R} \right) \\ &= \frac{1}{R} \sum_j y_{.j.}^2 + \frac{y_{...}^2}{R^2} - 2 \frac{y_{...}^2}{R^2} \\ &= \frac{1}{R} \sum_j y_{.j.}^2 - \frac{(y_{...})^2}{R^2} = \frac{1}{R} \sum_j y_{.j.}^2 - C. \end{aligned}$$

Em seguida,

$$\begin{aligned}
\text{(IV)} &= R \sum_k (\bar{y}_{..k} - \bar{y}_{...})^2 = R \sum_k (\bar{y}_{..k}^2 + \bar{y}_{...}^2 - 2\bar{y}_{..k}\bar{y}_{...}) \\
&= R \sum_k \left(\frac{y_{..k}^2}{R^2} + \frac{y_{...}^2}{R^4} - 2\frac{y_{..k}}{R^2} \frac{y_{...}}{R} \right) \\
&= \sum_k \frac{y_{..k}^2}{R} + R^2 \frac{y_{...}^2}{R^4} - 2R \frac{y_{...}^2}{R^3} \\
&= \frac{1}{R} \sum_k y_{..k}^2 - \frac{y_{...}^2}{R^2} = \frac{1}{R} \sum_k y_{..k}^2 - C.
\end{aligned}$$

De maneira semelhante,

$$\begin{aligned}
\text{(V)} &= \sum_i \sum_{j,k} (y_{ijk} - \bar{y}_{...})(\bar{y}_{i..} - \bar{y}_{...}) = \sum_i (y_{i..} - R\bar{y}_{...})(\bar{y}_{i..} - \bar{y}_{...}) \\
&= \sum_i \left(y_{i..} \frac{y_{i..}}{R} - y_{i..} \frac{y_{...}}{R^2} - R \frac{y_{i..}}{R} \frac{y_{...}}{R^2} + R \frac{y_{...}^2}{R^4} \right) \\
&= \sum_i \frac{y_{i..}^2}{R} - \frac{y_{...}^2}{R^2} - \frac{y_{...}^2}{R^2} + \frac{y_{...}^2}{R^2} \\
&= \frac{1}{R} \sum_i y_{i..}^2 - \frac{y_{...}^2}{R^2} = \frac{1}{R} \sum_i y_{i..}^2 - C.
\end{aligned}$$

Mais uma vez,

$$\begin{aligned}
\text{(VI)} &= \sum_j \sum_{i,k} (y_{ijk} - \bar{y}_{...})(\bar{y}_{.j.} - \bar{y}_{...}) = \sum_j (y_{.j.} - R\bar{y}_{...})(\bar{y}_{.j.} - \bar{y}_{...}) \\
&= \sum_j \left(y_{.j.} \frac{y_{.j.}}{R} - y_{.j.} \frac{y_{...}}{R^2} - R \frac{y_{.j.}}{R} \frac{y_{...}}{R^2} + R \frac{y_{...}^2}{R^4} \right) \\
&= \sum_j \frac{y_{.j.}^2}{R} - \frac{y_{...}^2}{R^2} - \frac{y_{...}^2}{R^2} + \frac{y_{...}^2}{R^2} \\
&= \frac{1}{R} \sum_j y_{.j.}^2 - \frac{y_{...}^2}{R^2} = \frac{1}{R} \sum_j y_{.j.}^2 - C.
\end{aligned}$$

Como anteriormente,

$$\begin{aligned}
\text{(VII)} &= \sum_k \sum_{i,j} (y_{ijk} - \bar{y}_{...})(\bar{y}_{..k} - \bar{y}_{...}) = \sum_k (y_{..k} - R\bar{y}_{...})(\bar{y}_{..k} - \bar{y}_{...}) \\
&= \sum_k \left(y_{..k} \frac{y_{..k}}{R} - y_{..k} \frac{y_{...}}{R^2} - R \frac{y_{..k}}{R} \frac{y_{...}}{R^2} + R \frac{y_{...}^2}{R^4} \right) \\
&= \sum_k \frac{y_{..k}^2}{R} - \frac{y_{...}^2}{R^2} - \frac{y_{...}^2}{R^2} + \frac{y_{...}^2}{R^2}
\end{aligned}$$

$$= \frac{1}{R} \sum_k y_{..k}^2 - \frac{y_{...}^2}{R^2} = \frac{1}{R} \sum_k y_{..k}^2 - C.$$

De maneira análoga,

$$\begin{aligned} \text{(VIII)} &= \sum_i \sum_j (\bar{y}_{i..} - \bar{y}_{...})(\bar{y}_{.j.} - \bar{y}_{...}) = \sum_i (\bar{y}_{i..} - \bar{y}_{...}) \sum_j (\bar{y}_{.j.} - \bar{y}_{...}) \\ &= \sum_i \left(\frac{y_{i..}}{R} - \frac{y_{...}}{R^2} \right) \sum_j \left(\frac{y_{.j.}}{R} - \frac{y_{...}}{R^2} \right) \\ &= \left(\frac{y_{...}}{R} - \frac{y_{...}}{R} \right) \left(\frac{y_{...}}{R} - \frac{y_{...}}{R} \right) = 0. \end{aligned}$$

Da mesma forma,

$$\text{(IX)} = \sum_i \sum_k (\bar{y}_{i..} - \bar{y}_{...})(\bar{y}_{..k} - \bar{y}_{...}) = 0,$$

e

$$\text{(X)} = \sum_j \sum_k (\bar{y}_{.j.} - \bar{y}_{...})(\bar{y}_{..(k)} - \bar{y}_{...}) = 0.$$

Substituindo as expressões de (I) a (X) na equação da soma de quadrados dos resíduos, vem,

$$\begin{aligned} \sum_i \sum_{j,k} e_{ijk}^2 &= \left(\sum_i \sum_{j,k} y_{ijk}^2 - C \right) + \left(\frac{1}{R} \sum_i y_{i..} - C \right) + \left(\frac{1}{R} \sum_j y_{.j.} - C \right) \\ &+ \left(\frac{1}{R} \sum_k y_{..k} - C \right) - 2 \left(\frac{1}{R} \sum_i y_{i..}^2 - C \right) - 2 \left(\frac{1}{R} \sum_j y_{.j.}^2 - C \right) \\ &- 2 \left(\frac{1}{R} \sum_k y_{..k}^2 - C \right) + 2 \times 0 + 2 \times 0 + 2 \times 0 \end{aligned}$$

E segue que

$$\underbrace{\sum_i \sum_{j,k} e_{ijk}^2}_{\text{(SQResiduo)}} = \underbrace{\left(\sum_i \sum_{j,k} y_{ijk}^2 - C \right)}_{\text{(SQTotal)}} - \underbrace{\left(\frac{1}{R} \sum_i y_{i..} - C \right)}_{\text{(SQLinhas)}} - \underbrace{\left(\frac{1}{R} \sum_j y_{.j.} - C \right)}_{\text{(SQColunas)}} - \underbrace{\left(\frac{1}{R} \sum_k y_{..k} - C \right)}_{\text{(SQTratamentos)}} \quad (2.10)$$

Ou

$$\underbrace{\left(\sum_i \sum_{j,k} y_{ijk}^2 - C \right)}_{(\text{SQTotal})} = \underbrace{\left(\frac{1}{R} \sum_i y_{i..} - C \right)}_{(\text{SQLinhas})} + \underbrace{\left(\frac{1}{R} \sum_j y_{.j.} - C \right)}_{(\text{SQColunas})} + \underbrace{\left(\frac{1}{R} \sum_k y_{..k} - C \right)}_{(\text{SQTratamentos})} + \underbrace{\sum_i \sum_{j,k} e_{ijk}^2}_{(\text{SQResiduo})}. \quad (2.11)$$

Portanto, de acordo com a expressão (2.11), a soma de quadrados total é decomposta em quatro partes componentes. Na prática, primeiro calcula-se as demais somas de quadrados e a soma de quadrados dos resíduos é calculada por diferença usando-se a expressão (2.10). A seguir, mostraremos as distribuições de probabilidade associadas aos estimadores dos parâmetros, bem como, às somas de quadrados.

2.6 Distribuição de probabilidade dos estimadores

Para se fazer inferência sobre os parâmetros ou combinações lineares deles, é preciso conhecer as distribuições de probabilidade associadas aos correspondentes estimadores. Nesta seção, serão estudadas as distribuições de probabilidade dos estimadores dos parâmetros do modelo, bem como algumas propriedades destes. Para isso, utiliza-se os elementos a seguir para ajudar nas demonstrações posteriores a serem desenvolvidas.

Já sabemos que:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \tau_k + \epsilon_{ijk}$$

$$E(\epsilon_{ijk}) = 0, \quad E(\epsilon_{ijk}^2) = \sigma^2 \quad \text{e} \quad \epsilon_{ijk} \sim N(0; \sigma^2) \quad (2.12)$$

Conseqüentemente

$$E(y_{ijk}) = \mu + \alpha_i + \beta_j + \tau_k, \quad Var(y_{ijk}) = \sigma^2 \quad \text{e} \quad y_{ijk} \sim N(\mu + \alpha_i + \beta_j + \tau_k; \sigma^2). \quad (2.13)$$

Estes fatos ajudarão a elucidar as demonstrações seguintes.

2.6.1 Distribuição de probabilidade de $\bar{y}_{..}$ estimador de μ

Sabe-se que $\bar{y}_{..} = \frac{1}{R^2} y_{..}$ é uma combinação linear dos y_{ijk} os quais, por (2.12), seguem uma distribuição normal. De acordo com Mood *et al.* (1974), Rohatgi (1976), Roussas (1997), dentre outros, combinações lineares de variáveis aleatórias normais, também seguem uma distribuição normal, então $\bar{y}_{..}$ é normal. Para encontrar as características da distribuição de $\bar{y}_{..}$, procede-se da seguinte maneira:

$$\begin{aligned}
\bar{y}_{...} &= \frac{1}{R^2} \sum_i \sum_{j,k} y_{ijk} = \frac{1}{R^2} \sum_i \sum_{j,k} (\mu + \alpha_i + \beta_j + \tau_k + \epsilon_{ijk}) \\
&= \frac{1}{R^2} \left(R^2 \mu + R \sum_i \alpha_i + R \sum_j \beta_j + R \sum_k \tau_k + \sum_i \sum_{j,k} \epsilon_{ijk} \right) \\
&= \mu + \frac{1}{R^2} \sum_i \sum_{j,k} \epsilon_{ijk}.
\end{aligned}$$

Logo,

$$\bar{y}_{...} = \mu + \frac{1}{R^2} \sum_i \sum_{j,k} \epsilon_{ijk} \quad \text{e} \quad \bar{y}_{...} - \mu = \frac{1}{R^2} \sum_i \sum_{j,k} \epsilon_{ijk} \quad (2.14)$$

e segue que

$$E(\bar{y}_{...}) = \mu + \frac{1}{R^2} \sum_i \sum_{j,k} E(\epsilon_{ijk}) = \mu$$

e

$$Var(\bar{y}_{...}) = E[\bar{y}_{...} - E(\bar{y}_{...})]^2 = E\left(\frac{1}{R^2} \sum_i \sum_{j,k} \epsilon_{ijk}\right)^2 = \frac{1}{R^4} R^2 \sigma^2 = \frac{\sigma^2}{R^2}.$$

Portanto,

$$\bar{y}_{...} \sim N\left(\mu; \frac{\sigma^2}{R^2}\right). \quad (2.15)$$

2.6.2 Distribuição de probabilidade da correção para média, C

De (2.15), tem-se que

$$\bar{y}_{...} \sim N\left(\mu; \frac{\sigma^2}{R^2}\right) \Rightarrow \frac{\bar{y}_{...} - \mu}{\sqrt{\frac{\sigma^2}{R^2}}} \sim N(0, 1)$$

Assim, sob $H_0 : \mu = 0$, as estatísticas

$$\frac{\bar{y}_{...} - 0}{\sqrt{\frac{\sigma^2}{R^2}}} \sim N(0, 1) \quad \text{e} \quad \left(\frac{\bar{y}_{...}}{\sqrt{\frac{\sigma^2}{R^2}}}\right)^2 = \frac{(\bar{y}_{...})^2}{\frac{\sigma^2}{R^2}} = \frac{C}{\sigma^2} \sim \chi_{(1)}^2.$$

Em palavras, a estatística $\frac{(\bar{y}_{...})^2}{\sigma^2} = \frac{C}{\sigma^2}$ segue uma distribuição de qui-quadrado com *um* grau de liberdade.

2.6.3 Distribuição de probabilidade de a_i o estimador de α_i

Vimos em (2.8) que $a_i = \frac{1}{R}y_{i..} - \frac{1}{R^2}y_{...} = \bar{y}_{i..} - \bar{y}_{...}$, que é uma combinação linear dos y_{ijk} os quais seguem uma distribuição normal. Logo, a_i é normal.

Sendo assim,

$$\begin{aligned}
a_i &= \bar{y}_{i..} - \bar{y}_{...} = \frac{1}{R} \sum_{j,k} y_{ijk} - \frac{1}{R^2} \sum_i \sum_{j,k} y_{ijk} \\
&= \frac{1}{R} \sum_{j,k} (\mu + \alpha_i + \beta_j + \tau_k + \epsilon_{ijk}) - \frac{1}{R^2} \sum_i \sum_{j,k} (\mu + \alpha_i + \beta_j + \tau_k + \epsilon_{ijk}) \\
&= \frac{1}{R} (R\mu + R\alpha_i + \sum_j \beta_j + \sum_k \tau_k + \sum_{j,k} \epsilon_{ijk}) - \frac{1}{R^2} (R^2\mu + R \sum_i \alpha_i + R \sum_j \beta_j \\
&\quad + R \sum_k \tau_k + \sum_i \sum_{j,k} \epsilon_{ijk}) \\
&= \mu + \alpha_i + \frac{1}{R} \sum_{j,k} \epsilon_{ijk} - \mu - \frac{1}{R^2} \sum_i \sum_{j,k} \epsilon_{ijk} \\
&= \alpha_i + \frac{1}{R} \sum_{j,k} \epsilon_{ijk} - \frac{1}{R^2} \sum_i \sum_{j,k} \epsilon_{ijk}
\end{aligned}$$

Portanto,

$$a_i = \alpha_i + \frac{1}{R} \sum_{j,k} \epsilon_{ijk} - \frac{1}{R^2} \sum_i \sum_{j,k} \epsilon_{ijk} \quad \text{e} \quad a_i - \alpha_i = \frac{1}{R} \sum_{j,k} \epsilon_{ijk} - \frac{1}{R^2} \sum_i \sum_{j,k} \epsilon_{ijk}$$

e segue que

$$E(a_i) = \alpha_i + \frac{1}{R} \sum_{j,k} E(\epsilon_{ijk}) - \frac{1}{R^2} \sum_i \sum_{j,k} E(\epsilon_{ijk}) = \alpha_i$$

e

$$\begin{aligned}
\text{Var}(a_i) &= E[a_i - E(a_i)]^2 = E(a_i - \alpha_i)^2 \\
&= E\left(\frac{1}{R} \sum_{j,k} \epsilon_{ijk} - \frac{1}{R^2} \sum_i \sum_{j,k} \epsilon_{ijk}\right)^2 \\
&= \frac{\sigma^2}{R} + \frac{\sigma^2}{R^2} - \frac{2\sigma^2}{R^2} = \frac{(R-1)}{R^2} \sigma^2.
\end{aligned}$$

Portanto,

$$a_i \sim N\left(\alpha_i; \frac{(R-1)}{R^2} \sigma^2\right). \quad (2.16)$$

2.6.4 Distribuição de probabilidade da SQLinhas

Partindo de (2.16), vem

$$a_i \sim N\left(\alpha_i; \frac{(R-1)}{R^2}\sigma^2\right) \Rightarrow \frac{a_i - \alpha_i}{\sqrt{\frac{(R-1)}{R^2}\sigma^2}} \sim N(0; 1) \Rightarrow \frac{\bar{y}_{i..} - \bar{y}_{...} - \alpha_i}{\sqrt{\frac{(R-1)}{R^2}}} \sim N(0; 1)$$

Assim, sob a hipótese de que $H_0 : \alpha_i = 0, \forall i$, as estatísticas

$$\frac{\bar{y}_{i..} - \bar{y}_{...}}{\sqrt{\frac{(R-1)}{R^2}\sigma^2}} \sim N(0, 1) \quad \text{e} \quad \frac{(\bar{y}_{i..} - \bar{y}_{...})^2}{\frac{(R-1)}{R^2}\sigma^2} \sim \chi_{(1)}^2.$$

Além disso, como

$$\frac{(\bar{y}_{i..} - \bar{y}_{...})^2}{\frac{(R-1)}{R^2}\sigma^2} \sim \chi_{(1)}^2 \quad \text{então,} \quad \frac{R(\bar{y}_{i..} - \bar{y}_{...})^2}{\frac{(R-1)}{R}\sigma^2} \sim \chi_{(1)}^2 \quad \text{e} \quad \frac{R(\bar{y}_{i..} - \bar{y}_{...})^2}{\sigma^2} \sim \frac{R-1}{R}\chi_{(1)}^2$$

e segue que

$$\frac{R \sum_i (\bar{y}_{i..} - \bar{y}_{...})^2}{\sigma^2} \sim (R-1)\chi_{(1)}^2 \Rightarrow \frac{R \sum_i (\bar{y}_{i..} - \bar{y}_{...})^2}{\sigma^2} = \frac{SQLinhas}{\sigma^2} \sim \chi_{(R-1)}^2. \quad (2.17)$$

Portanto, a soma de quadrados das linhas dividida por σ^2 , segue uma distribuição de qui-quadrado com $(R-1)$ graus de liberdade.

2.6.5 Distribuição de probabilidade de b_j o estimador de β_j

De (2.8), tem-se que $b_j = \frac{1}{R}y_{.j} - \frac{1}{R^2}y_{...} = \bar{y}_{.j} - \bar{y}_{...}$, que é uma combinação linear dos y_{ijk} . Assim, b_j também segue uma distribuição normal.

Além disso,

$$\begin{aligned} b_j &= \bar{y}_{.j} - \bar{y}_{...} = \frac{1}{R} \sum_{i,k} y_{ijk} - \frac{1}{R^2} \sum_j \sum_{i,k} y_{ijk} \\ &= \frac{1}{R} \sum_{i,k} (\mu + \alpha_i + \beta_j + \tau_k + \epsilon_{ijk}) - \frac{1}{R^2} \sum_j \sum_{i,k} (\mu + \alpha_i + \beta_j + \tau_k + \epsilon_{ijk}) \\ &= \frac{1}{R} (R\mu + \sum_i \alpha_i + R\beta_j + \sum_k \tau_k + \sum_{i,k} \epsilon_{ijk}) - \frac{1}{R^2} (R^2\mu + R \sum_i \alpha_i + R \sum_j \beta_j \\ &\quad + R \sum_k \tau_k + \sum_j \sum_{i,k} \epsilon_{ijk}) \\ &= \beta_j + \frac{1}{R} \sum_{i,k} \epsilon_{ijk} - \frac{1}{R^2} \sum_j \sum_{i,k} \epsilon_{ijk}. \end{aligned}$$

Portanto,

$$b_j = \beta_j + \frac{1}{R} \sum_{i,k} \epsilon_{ijk} - \frac{1}{R^2} \sum_j \sum_{i,k} \epsilon_{ijk} \quad \text{e} \quad \beta_j - \beta_j = \frac{1}{R} \sum_{i,k} \epsilon_{ijk} - \frac{1}{R^2} \sum_j \sum_{i,k} \epsilon_{ijk}. \quad (2.18)$$

Usando (2.18), encontra-se

$$E(b_j) = \beta_j + \frac{1}{R} \sum_{i,k} E(\epsilon_{ijk}) - \frac{1}{R^2} \sum_j \sum_{i,k} E(\epsilon_{ijk}) = \beta_j$$

e

$$\begin{aligned} \text{Var}(b_j) &= E[b_j - E(b_j)]^2 = E(b_j - \beta_j)^2 \\ &= E\left(\frac{1}{R} \sum_{i,k} \epsilon_{ijk} - \frac{1}{R^2} \sum_j \sum_{i,k} \epsilon_{ijk}\right)^2 \\ &= E\left(\frac{1}{R} \sum_{i,k} \epsilon_{ijk}\right)^2 + E\left(\frac{1}{R^2} \sum_j \sum_{i,k} \epsilon_{ijk}\right)^2 - \frac{2}{R^3} E\left(\sum_{i,k} \epsilon_{ijk}\right) \left(\sum_j \sum_{i,k} \epsilon_{ijk}\right) \\ &= \frac{\sigma^2}{R} + \frac{\sigma^2}{R^2} - \frac{2\sigma^2}{R^2} = \frac{(R-1)}{R} \sigma^2. \end{aligned}$$

Logo, concluí-se que

$$b_j \sim N\left(\beta_j; \frac{(R-1)}{R} \sigma^2\right). \quad (2.19)$$

Isto é, b_j segue uma distribuição normal com média β_j e variância $\frac{(R-1)}{R} \sigma^2$.

2.6.6 Distribuição de probabilidade da SQColunas

Partindo do resultado (2.19), obtêm-se

$$\frac{b_j - \beta_j}{\sqrt{\frac{(R-1)}{R^2} \sigma^2}} \sim N(0, 1) \Rightarrow \frac{\bar{y}_{.j} - \bar{y}_{...} - \beta_j}{\sqrt{\frac{(R-1)}{R^2} \sigma^2}} \sim N(0, 1)$$

e sob a hipótese $H_0 : \beta_j = 0, \forall j$, então as estatísticas

$$\frac{\bar{y}_{.j} - \bar{y}_{...}}{\sqrt{\frac{(R-1)}{R^2} \sigma^2}} \sim N(0, 1) \quad \text{e} \quad \frac{(\bar{y}_{.j} - \bar{y}_{...})^2}{\frac{(R-1)}{R^2} \sigma^2} \sim \chi_{(1)}^2$$

consequentemente,

$$\frac{R(\bar{y}_{.j} - \bar{y}_{...})^2}{\frac{(R-1)}{R} \sigma^2} \sim \chi_{(1)}^2 \Rightarrow \frac{R(\bar{y}_{.j} - \bar{y}_{...})^2}{\sigma^2} \sim \frac{R-1}{R} \chi_{(1)}^2$$

e

$$\frac{R \sum_j (\bar{y}_{.j} - \bar{y}_{...})^2}{\sigma^2} \sim (R-1)\chi_{(1)}^2 \Rightarrow \frac{R \sum_j (\bar{y}_{.j} - \bar{y}_{...})^2}{\sigma^2} = \frac{SQ_{colunas}}{\sigma^2} \sim \chi_{(R-1)}^2. \quad (2.20)$$

Logo, a soma de quadrados das colunas dividida por σ^2 , segue uma distribuição de qui-quadrado com $(R-1)$ graus de liberdade.

2.6.7 Distribuição de probabilidade de t_k o estimador de τ_k

Considerando-se ainda os resultados (2.8), no qual $t_k = \frac{1}{R}y_{..k} - \frac{1}{R^2}y_{...} = \bar{y}_{..k} - \bar{y}_{...}$, percebe-se que t_k é uma combinação linear dos y_{ijk} que seguem distribuição normal. Logo, t_k é normal.

Além disso, tem-se que:

$$\begin{aligned} t_k &= \bar{y}_{..k} - \bar{y}_{...} = \frac{1}{R} \sum_{i,j} y_{ijk} - \frac{1}{R^2} \sum_k \sum_{i,j} y_{ijk} \\ &= \frac{1}{R} (R\mu + \sum_i \alpha_i + \sum_j \beta_j + R\tau_k + \sum_{i,j} \epsilon_{ijk}) - \frac{1}{R^2} (R^2\mu + R \sum_i \alpha_i + R \sum_j \beta_j \\ &\quad + R \sum_k \tau_k + \sum_k \sum_{i,j} \epsilon_{ijk}) \\ &= \tau_k + \frac{1}{R} \sum_{i,j} \epsilon_{ijk} - \frac{1}{R^2} \sum_k \sum_{i,j} \epsilon_{ijk}. \end{aligned}$$

Logo,

$$t_k = \tau_k + \frac{1}{R} \sum_{i,j} \epsilon_{ijk} - \frac{1}{R^2} \sum_k \sum_{i,j} \epsilon_{ijk} \quad \text{e} \quad t_k - \tau_k = \frac{1}{R} \sum_{i,j} \epsilon_{ijk} - \frac{1}{R^2} \sum_k \sum_{i,j} \epsilon_{ijk}.$$

Portanto,

$$E(t_k) = \tau_k + \frac{1}{R} \sum_{i,j} E(\epsilon_{ijk}) - \frac{1}{R^2} \sum_k \sum_{i,j} E(\epsilon_{ijk}) = \tau_k$$

e

$$\begin{aligned} Var(t_k) &= E[t_k - E(t_k)]^2 = E(t_k - \tau_k)^2 \\ &= E\left(\frac{1}{R} \sum_{i,j} \epsilon_{ijk} - \frac{1}{R^2} \sum_k \sum_{i,j} \epsilon_{ijk}\right)^2 \\ &= \frac{1}{R^2} E\left(\sum_{i,j} \epsilon_{ijk}\right)^2 + \frac{1}{R^4} E\left(\sum_k \sum_{i,j} \epsilon_{ijk}\right)^2 - \frac{2}{R^3} E\left(\sum_{i,j} \epsilon_{ijk}\right) \left(\sum_k \sum_{i,j} \epsilon_{ijk}\right) \\ &= \frac{\sigma^2}{R} + \frac{\sigma^2}{R^2} - \frac{2\sigma^2}{R^2} = \frac{(R-1)}{R^2} \sigma^2. \end{aligned}$$

Portanto,

$$t_k \sim N\left(\tau_k; \frac{(R-1)}{R}\sigma^2\right). \quad (2.21)$$

Isto é, t_k segue uma distribuição normal com média τ_k e variância $\frac{(R-1)}{R^2}$.

2.6.8 Distribuição de probabilidade da SQTratamentos

Usando (2.21), tem-se

$$\frac{t_k - \tau_k}{\sqrt{\frac{(R-1)}{R^2}\sigma^2}} \sim N(0, 1) \Rightarrow \frac{\bar{y}_{..k} - \bar{y}_{...} - \tau_k}{\sqrt{\frac{(R-1)}{R^2}\sigma^2}} \sim N(0, 1)$$

que sob a hipótese $H_0 : \tau_k = 0, \forall k$ as estatísticas

$$\frac{\bar{y}_{..k} - \bar{y}_{...}}{\sqrt{\frac{(R-1)}{R^2}\sigma^2}} \sim N(0, 1) \quad \text{e} \quad \frac{(\bar{y}_{..k} - \bar{y}_{...})^2}{\frac{(R-1)}{R^2}\sigma^2} \sim \chi_{(1)}^2.$$

Como,

$$\frac{R(\bar{y}_{..k} - \bar{y}_{...})^2}{\frac{(R-1)}{R}\sigma^2} \sim \chi_{(1)}^2 \Rightarrow \frac{R(\bar{y}_{..k} - \bar{y}_{...})^2}{\sigma^2} \sim \frac{R-1}{R}\chi_{(1)}^2$$

e

$$\frac{R \sum_k (\bar{y}_{..k} - \bar{y}_{...})^2}{\sigma^2} \sim (R-1)\chi_{(1)}^2 \Rightarrow \frac{R \sum_k (\bar{y}_{..k} - \bar{y}_{...})^2}{\sigma^2} = \frac{SQTratamento}{\sigma^2} \sim \chi_{(R-1)}^2 \quad (2.22)$$

Portanto, a soma de quadrado dos tratamentos dividida por σ^2 segue distribuição de qui-quadrado com $(R-1)$ graus de liberdade.

2.6.9 Distribuição de probabilidade de m_i , estimador de $\mu_i = \mu + \alpha_i$, de m_j , estimador de $\mu_j = \mu + \beta_j$ e de m_k , estimador de $\mu_k = \mu + \tau_k$.

As médias das linhas, das colunas e dos tratamentos foram definidas em (2.9) e usando os resultados (2.15), (2.16), (2.19) e (2.21), tem-se que:

$$\begin{aligned} m_i &= m + a_i = \bar{y}_{..} + \bar{y}_{i..} - \bar{y}_{...} = \bar{y}_{i..} \\ &= \frac{1}{R} \sum_{j,k} y_{ijk} = \frac{1}{R} \sum_{j,k} (\mu + \alpha_i + \beta_j + \tau_k + \epsilon_{ijk}) \\ &= \frac{1}{R} \left(R\mu + R\alpha_i + \sum_j \beta_j + \sum_k \tau_k + \sum_{j,k} \epsilon_{ijk} \right) \end{aligned}$$

$$= \mu + \alpha_i + \frac{1}{R} \sum_{j,k} \epsilon_{ijk} = \mu_i + \frac{1}{R} \sum_{j,k} \epsilon_{ijk}.$$

Logo,

$$m_i = \mu_i + \frac{1}{R} \sum_{j,k} \epsilon_{ijk} \quad \text{e} \quad m_i - \mu_i = \frac{1}{R} \sum_{j,k} \epsilon_{ijk}$$

consequentemente,

$$E(m_i) = \mu_i + \frac{1}{R} \sum_{j,k} E(\epsilon_{ijk}) = \mu_i$$

e

$$\text{Var}(m_i) = E[m_i - E(m_i)]^2 = E(m_i - \mu_i)^2 = E\left(\frac{1}{R} \sum_{j,k} \epsilon_{ijk}\right)^2 = \frac{\sigma^2}{R}.$$

Portanto,

$$m_i \sim N\left(\mu_i; \frac{\sigma^2}{R}\right) \quad (2.23)$$

e seguindo procedimento semelhante, encontra-se

$$m_j \sim N\left(\mu_j; \frac{\sigma^2}{R}\right) \quad \text{e} \quad m_k \sim N\left(\mu_k; \frac{\sigma^2}{R}\right). \quad (2.24)$$

2.6.10 Distribuição de probabilidade de e_{ijk} , o estimador de ϵ_{ijk}

Sabe-se que,

$$e_{ijk} = y_{ijk} - m - a_i - b_j - t_k \quad (2.25)$$

Substituindo as equações de (2.8) na equação (2.25), tem-se que:

$$\begin{aligned} e_{ijk} &= y_{ijk} - m - a_i - b_j - t_k \\ &= y_{ijk} - \bar{y}_{...} - (\bar{y}_{i..} - \bar{y}_{...}) - (\bar{y}_{.j.} - \bar{y}_{...}) - (\bar{y}_{..k} - \bar{y}_{...}) \\ &= y_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} - \bar{y}_{..k} + 2\bar{y}_{...} \end{aligned}$$

Substituindo as equações de (2.15), (2.16), (2.19) e (2.21) tem-se que,

$$\begin{aligned} e_{ijk} &= \mu + \alpha_i + \beta_j + \tau_k + \epsilon_{ijk} - \left(\mu + \alpha_i + \frac{1}{R} \sum_j \epsilon_{ijk}\right) - \left(\mu + \beta_j + \frac{1}{R} \sum_k \epsilon_{ijk}\right) \\ &\quad - \left(\mu + \tau_i + \frac{1}{R} \sum_i \epsilon_{ijk}\right) + 2 \left(\mu + \frac{1}{R^2} \sum_i \sum_{j,k} \epsilon_{ijk}\right) \\ &= \epsilon_{ijk} - \frac{1}{R} \sum_j \epsilon_{ijk} - \frac{1}{R} \sum_k \epsilon_{ijk} - \frac{1}{R} \sum_i \epsilon_{ijk} + 2 \frac{1}{R^2} \sum_i \sum_{j,k} \epsilon_{ijk} \end{aligned}$$

Assim é fácil verificar que $E(e_{ijk}) = 0$. Calculando-se $Var(e_{ijk})$.

$$\begin{aligned}
Var(e_{ijk}) &= E(e_{ijk} - E(e_{ijk}))^2 = E(e_{ijk})^2 \\
&= E\left(\epsilon_{ijk} - \frac{1}{R} \sum_j \epsilon_{ijk} - \frac{1}{R} \sum_k \epsilon_{ijk} - \frac{1}{R} \sum_i \epsilon_{ijk} + 2\frac{1}{R^2} \sum_i \sum_{j,k} \epsilon_{ijk}\right)^2 \\
&= E(\epsilon_{ijk})^2 + E\left(\frac{1}{R} \sum_j \epsilon_{ijk}\right)^2 + E\left(\frac{1}{R} \sum_k \epsilon_{ijk}\right)^2 + E\left(\frac{1}{R} \sum_i \epsilon_{ijk}\right)^2 \\
&\quad + E\left(2\frac{1}{R^2} \sum_i \sum_{j,k} \epsilon_{ijk}\right)^2 - 2E\left(\epsilon_{ijk} \frac{1}{R} \sum_j \epsilon_{ijk}\right) - 2E\left(\epsilon_{ijk} \frac{1}{R} \sum_k \epsilon_{ijk}\right) \\
&\quad - 2E\left(\epsilon_{ijk} \frac{1}{R} \sum_i \epsilon_{ijk}\right) + 4E\left(\epsilon_{ijk} \frac{1}{R^2} \sum_i \sum_{j,k} \epsilon_{ijk}\right) + 2E\left(\frac{1}{R} \sum_j \epsilon_{ijk} \frac{1}{R} \sum_k \epsilon_{ijk}\right) \\
&\quad + 2E\left(\frac{1}{R} \sum_j \epsilon_{ijk} \frac{1}{R} \sum_i \epsilon_{ijk}\right) - 4E\left(\frac{1}{R} \sum_j \epsilon_{ijk} \frac{1}{R^2} \sum_i \sum_{j,k} \epsilon_{ijk}\right) \\
&\quad + 2E\left(\frac{1}{R} \sum_k \epsilon_{ijk} \frac{1}{R} \sum_i \epsilon_{ijk}\right) - 4E\left(\frac{1}{R} \sum_k \epsilon_{ijk} \frac{1}{R^2} \sum_i \sum_{j,k} \epsilon_{ijk}\right) \\
&\quad - 4E\left(\frac{1}{R} \sum_i \epsilon_{ijk} \frac{1}{R^2} \sum_i \sum_{j,k} \epsilon_{ijk}\right) \\
&= \sigma^2 + \frac{1}{R}\sigma^2 + \frac{1}{R}\sigma^2 + \frac{1}{R}\sigma^2 + \frac{1}{R^2}\sigma^2 - \frac{2}{R}\sigma^2 - \frac{2}{R}\sigma^2 - \frac{2}{R}\sigma^2 + \frac{4}{R^2}\sigma^2 + \frac{2}{R^2}\sigma^2 + \frac{2}{R^2}\sigma^2 \\
&\quad - \frac{4}{R^2}\sigma^2 + \frac{2}{R^2}\sigma^2 - \frac{4}{R^2}\sigma^2 \\
&= \sigma^2 - \frac{3}{R}\sigma^2 + \frac{2}{R^2}\sigma^2 = \frac{R^2 - 3R + 2}{R^2}\sigma^2 \\
&= \frac{(R-2)(R-1)}{R^2}\sigma^2
\end{aligned}$$

Logo

$$e_{ijk} \sim N\left(0; \frac{(R-2)(R-1)}{R^2}\sigma^2\right) \quad (2.26)$$

2.6.11 Distribuição de probabilidade de contrastes de interesse

Muitas vezes, no planejamento de experimento, se tem interesse em comparar as médias dos tratamentos para verificar se a diferença entre essas médias são significativas ou não. Estas comparações de médias são chamadas de contrastes de interesse que são

combinações lineares das médias dos tratamentos. Aqui apresenta-se as demonstrações relativas a esses contrastes de média dos tratamentos.

Seja

$$\begin{aligned}
 \Psi_h = \sum_k C_{h_k} m_k &= \sum_k C_{h_k} \bar{y}_{..k} = \frac{1}{R} \sum_k C_{h_k} \sum_{i,j} y_{ijk}, \text{ com } \sum_k C_{h_k} = 0 \\
 &= \frac{1}{R} \sum_k C_{h_k} \sum_{i,j} (\mu + \alpha_i + \beta_j + \tau_k + \epsilon_{ijk}) \\
 &= \frac{1}{R} \sum_k C_{h_k} \left(R\mu + R\tau_k + \sum_j \epsilon_{ijk} \right) \\
 &= \sum_k C_{h_k} \mu_k + \frac{1}{R} \sum_k C_{h_k} \sum_{i,j} \epsilon_{ijk}
 \end{aligned}$$

Com isso teremos que:

$$\begin{aligned}
 E(\Psi_h) &= E \left(\sum_k C_{h_k} \mu_k + \frac{1}{R} \sum_k C_{h_k} \sum_{i,j} \epsilon_{ijk} \right) \\
 &= \sum_k C_{h_k} E \left(\mu_k + \frac{1}{R} \sum_{i,j} \epsilon_{ijk} \right) \\
 &= \sum_k C_{h_k} \mu_k
 \end{aligned}$$

Além disso, temos que:

$$\begin{aligned}
 Var(\Psi_h) &= E(\Psi_h - E(\Psi_h))^2 \\
 &= E \left(\sum_k C_{h_k} \mu_k + \frac{1}{R} \sum_k C_{h_k} \sum_{i,j} \epsilon_{ijk} - \sum_k C_{h_k} \mu_k \right)^2 \\
 &= E \left(\frac{1}{R} \sum_k C_{h_k} \sum_{i,j} \epsilon_{ijk} \right)^2 \\
 &= \frac{1}{R^2} E \left(\sum_k C_{h_k} \epsilon_{i1(k)} + \dots + \sum_k C_{h_k} \epsilon_{ijk} \right)^2 \\
 &= \frac{1}{R^2} \left(R\sigma^2 \sum_k C_{h_k}^2 \right) \\
 &= \frac{K_h}{R} \sigma^2, \text{ sendo } K_h = \sum_k C_{h_k}^2
 \end{aligned}$$

Logo,

$$\Psi_h \sim N \left(\sum_k C_{h_k} \mu_k; \frac{K_h}{R} \sigma^2 \right) \quad (2.27)$$

2.6.12 Distribuição de probabilidade da soma de quadrados dos contrastes.

A partir de (2.27), obtem-se:

$$\frac{\Psi_h - E(\Psi_h)}{\sqrt{\frac{K_h}{R} \sigma^2}} = \frac{\sum_k C_{h_k} m_k - \sum_k C_{h_k} \mu_k}{\sqrt{\frac{K_h}{R} \sigma^2}} \sim N(0, 1)$$

Sob a hipótese nula $H_0^{(\Psi)} : \sum_k C_{h_k} \mu_k = 0$ tem-se que:

$$\frac{\sum_k C_{h_k} m_k}{\sqrt{\frac{K_h}{R} \sigma^2}} \sim N(0, 1) \quad \text{e} \quad \frac{(\sum_k C_{h_k} m_k)^2}{\frac{K_h}{R} \sigma^2}$$

E segue-se que:

$$= \frac{R}{K \sigma^2} \frac{(\sum_k C_{h_k} y_{..k})^2}{R^2} = \frac{(\sum_k C_{h_k} y_{..k})^2}{RK_h \times \sigma^2} = \frac{SQ(\Psi_h)}{\sigma^2} \sim \chi_{(1)}^2 \quad (2.28)$$

Assim pode-se concluir que Ψ_h dividido por σ^2 segue uma qui-quadrado com um grau de liberdade.

2.7 Esperança das somas de quadrados e dos quadrados médios

Na subseção anterior mostrou-se que:

$$SQTotal = \sum_i \sum_{j,k} y_{ijk}^2 - C, \text{ e que } C = \frac{y_{...}^2}{R^2}$$

$$SQLinhas = R \sum_i (\bar{y}_{i..} - \bar{y}_{...})^2 = \frac{1}{R} \sum_i y_{i..} - C$$

$$SQColunas = R \sum_j (\bar{y}_{.j.} - \bar{y}_{...})^2 = \frac{1}{R} \sum_j y_{.j.} - C$$

$$SQTratamentos = R \sum_k (\bar{y}_{..k} - \bar{y}_{...})^2 = \frac{1}{R} \sum_k y_{..k} - C$$

$$SQ(\Psi_h) = \frac{(\sum_k C_{hk} y_{..k})^2}{RK_h}$$

$$SQResiduo = SQTotal - SQLinhas - SQColunas - SQTratamentos$$

(2.29)

Para melhor compreensão dos resultados e a escolha do teste, encontram-se os valores esperados de suas somas de quadrados e seus quadrados médios.

2.7.1 Esperança das somas de quadrado total

Sabe-se que:

$$SQTotal = \sum_i \sum_{j,k} y_{ijk}^2 - C,$$

assim:

$$E(SQTotal) = E\left(\sum_i \sum_{j,k} y_{ijk}^2 - C\right) = E\left(\sum_i \sum_{j,k} y_{ijk}^2\right) - E(C) \quad (2.30)$$

Calculando-se separadamente, tem-se que:

$$E(C) = E\left(\frac{y_{...}^2}{R^2}\right) = \frac{1}{R^2} E(y_{...}^2) = \frac{1}{R^2} E\left(R^2 \mu + \sum_i \sum_{j,k} \epsilon_{ijk}^2\right)^2$$

$$\begin{aligned}
&= \frac{1}{R^2} E \left(R^4 \mu^2 + \left(\sum_i \sum_{j,k} \epsilon_{ijk}^2 \right)^2 + odp \right) \\
&= R^2 \mu^2 + \sigma^2
\end{aligned}$$

Onde *odp* significa (Outros duplos produtos)

e que

$$\begin{aligned}
E \left(\sum_i \sum_{j,k} y_{ijk}^2 \right) &= E \left(\sum_i \sum_{j,k} (\mu + \alpha_i + \beta_j + \tau_k + \epsilon_{ijk})^2 \right) \\
&= \sum_i \sum_{j,k} E(\mu^2 + \alpha_i^2 + \beta_j^2 + \tau_k^2 + \epsilon_{ijk}^2 + 2\mu\alpha_i + 2\mu\beta_j + 2\mu\tau_k + 2\alpha_i\beta_j + 2\alpha_i\tau_k \\
&\quad + 2\beta_j\tau_k + odp) \\
&= \sum_i \sum_{j,k} (\mu^2 + \alpha_i^2 + \beta_j^2 + \tau_k^2 + \sigma^2 + 2\mu\alpha_i + 2\mu\beta_j + 2\mu\tau_k + 2\alpha_i\beta_j \\
&\quad + 2\alpha_i\tau_k + 2\beta_j\tau_k) \\
&= R^2 \mu^2 + R \sum_i \alpha_i^2 + R \sum_j \beta_j^2 + R \sum_k \tau_k^2 + R^2 \sigma^2 + 2R\mu \sum_i \alpha_i + 2R\mu \sum_j \beta_j \\
&\quad + 2R\mu \sum_k \tau_k + 2 \sum_i \alpha_i \sum_j \beta_j + 2 \sum_i \alpha_i \sum_k \tau_k + 2 \sum_j \beta_j \sum_k \tau_k) \\
&= R^2 \mu^2 + R \sum_i \alpha_i^2 + R \sum_j \beta_j^2 + R \sum_k \tau_k^2 + R^2 \sigma^2
\end{aligned}$$

Assim, substituindo-se em (2.30), tem-se que:

$$\begin{aligned}
E(SQTotal) &= E \left(\sum_i \sum_{j,k} y_{ijk}^2 - C \right) = E \left(\sum_i \sum_{j,k} y_{ijk}^2 \right) - E(C) \\
&= R^2 \mu^2 + R \sum_i \alpha_i^2 + R \sum_j \beta_j^2 + R \sum_k \tau_k^2 + R^2 \sigma^2 - R^2 \mu^2 - \sigma^2 \\
&= R \sum_i \alpha_i^2 + R \sum_j \beta_j^2 + R \sum_k \tau_k^2 + (R^2 - 1) \sigma^2
\end{aligned}$$

2.7.2 Esperança das somas de quadrado e do quadrado médio das linhas

Sabe-se que:

$$SQLinhas = \frac{1}{R} \sum_i y_{i..}^2 - C$$

Calcula-se a esperança.

$$E(SQLinhas) = E\left(\frac{1}{R} \sum_i y_{i..}^2 - C\right) = E\left(\frac{1}{R} \sum_i y_{i..}^2\right) - E(C) \quad (2.31)$$

Encontra-se na subseção anterior a esperança de C , conseqüentemente, encontra-se apenas a esperança de $\frac{1}{R} \sum_i y_{i..}^2$ e substitui-se na equação (2.31), assim:

$$\begin{aligned} E\left(\frac{1}{R} \sum_i y_{i..}^2\right) &= \frac{1}{R} E\left(\sum_i (R\mu + R\alpha_i + \sum_{j,k} \epsilon_{ijk})^2\right) \\ &= \frac{1}{R} E\left(R^3\mu^2 + R^2 \sum_i \alpha_i^2 + \sum_i \left(\sum_{j,k} \epsilon_{ijk}\right)^2 + 2R^2\mu \sum_i \alpha_i + odp\right) \\ &= R^2\mu^2 + R \sum_i \alpha_i^2 + R\sigma^2 \end{aligned}$$

Substituindo-se, tem-se que:

$$\begin{aligned} E(SQLinhas) &= E\left(\frac{1}{R} \sum_i y_{i..}^2\right) - E(C) \\ &= R^2\mu^2 + R \sum_i \alpha_i^2 + R\sigma^2 - R^2\mu^2 - \sigma^2 \\ &= R \sum_i \alpha_i^2 + (R-1)\sigma^2 \end{aligned}$$

Sabendo-se que:

$$QMLinhas = \frac{SQLinhas}{R-1}$$

tem-se:

$$\begin{aligned} E(QMLinhas) &= E\left(\frac{SQLinhas}{R-1}\right) = \frac{1}{R-1} E(SQLinhas) = \frac{1}{R-1} \left(R \sum_i \alpha_i^2 + (R-1)\sigma^2\right) \\ &= \sigma^2 + \frac{R}{R-1} \sum_i \alpha_i^2 \end{aligned}$$

2.7.3 Esperança das somas de quadrado e do quadrado médio das colunas

Sabe-se que:

$$SQColunas = \frac{1}{R} \sum_j y_{.j}^2 - C$$

Calculando-se sua esperança.

$$E(SQColunas) = E\left(\frac{1}{R} \sum_j y_{.j}^2 - C\right) = E\left(\frac{1}{R} \sum_j y_{.j}^2\right) - E(C) \quad (2.32)$$

Assim

$$\begin{aligned} E\left(\frac{1}{R} \sum_j y_{.j}^2\right) &= \frac{1}{R} E\left(\sum_j (R\mu + R\beta_j + \sum_{j,k} \epsilon_{ijk})^2\right) \\ &= \frac{1}{R} E\left(R^3\mu^2 + R^2 \sum_j \beta_j^2 + \sum_j \left(\sum_{i,k} \epsilon_{ijk}\right)^2 + 2R^2\mu \sum_j \beta_j + odp\right) \\ &= R^2\mu^2 + R \sum_j \beta_j^2 + R\sigma^2 \end{aligned}$$

Substituindo em (2.32) tem-se que:

$$\begin{aligned} E(SQColunas) &= E\left(\frac{1}{R} \sum_j y_{.j}^2\right) - E(C) \\ &= R^2\mu^2 + R \sum_j \beta_j^2 + R\sigma^2 - R^2\mu^2 - \sigma^2 \\ &= R \sum_j \beta_j^2 + (R-1)\sigma^2 \end{aligned}$$

Sabendo-se que,

$$QMColunas = \frac{SQColunas}{R-1}$$

segue-se que:

$$\begin{aligned} E(QMColunas) &= E\left(\frac{SQColunas}{R-1}\right) = \frac{1}{R-1} E(SQColunas) \\ &= \frac{1}{R-1} \left(R \sum_j \beta_j^2 + (R-1)\sigma^2\right) \end{aligned}$$

$$= \sigma^2 + \frac{R}{R-1} \sum_j \beta_j^2$$

2.7.4 Esperança das somas de quadrado e do quadrado médio dos tratamentos

Sabe-se que:

$$SQTratamento = \frac{1}{R} \sum_k y_{i..}^2 - C$$

Calculando-se sua esperança.

$$E(SQTratamento) = E\left(\frac{1}{R} \sum_k y_{i..}^2 - C\right) = E\left(\frac{1}{R} \sum_k y_{i..}^2\right) - E(C) \quad (2.33)$$

Novamente

$$\begin{aligned} E\left(\frac{1}{R} \sum_k y_{i..}^2\right) &= \frac{1}{R} E\left(\sum_k (R\mu + R\tau_k + \sum_{i,j} \epsilon_{ijk})^2\right) \\ &= \frac{1}{R} E\left((R^3\mu^2 + R^2 \sum_k \tau_k^2 + \sum_k \left(\sum_{i,j} \epsilon_{ijk}\right)^2 + 2R^2\mu \sum_k \tau_k + odp)\right) \\ &= R^2\mu^2 + R \sum_k \tau_k^2 + R\sigma^2 \end{aligned}$$

Substituindo-se em (2.33), tem-se que:

$$\begin{aligned} E(SQTratamento) &= E\left(\frac{1}{R} \sum_k y_{i..}^2\right) - E(C) \\ &= R^2\mu^2 + R \sum_k \tau_k^2 + R\sigma^2 - R^2\mu^2 - \sigma^2 \\ &= R \sum_k \tau_k^2 + (R-1)\sigma^2 \end{aligned}$$

Sabendo que:

$$QMTreatamento = \frac{SQTratamento}{R-1}$$

tem-se que:

$$\begin{aligned}
 E(QMTratamento) &= E\left(\frac{SQTratamento}{R-1}\right) = \frac{1}{R-1}E(SQTratamento) \\
 &= \frac{1}{R-1}\left(R\sum_k \tau_k^2 + (R-1)\sigma^2\right) \\
 &= \sigma^2 + \frac{R}{R-1}\sum_k \tau_k^2
 \end{aligned}$$

2.7.5 Esperança das somas de quadrado e do quadrado médio dos contrastes de interesse

Como foi visto.

$$SQ(\Psi_h) = \frac{(\sum_k C_{h_k} y_{..k})^2}{RK_h}$$

Desenvolvendo-se algebricamente a expressão acima teremos:

$$\begin{aligned}
 SQ(\Psi_h) &= \frac{(\sum_k C_{h_k} y_{..k})^2}{RK_h} \\
 &= \frac{1}{RK_h} \left[\sum_k C_{h_k} (\mu_k + \frac{1}{R} \sum_j \epsilon_{ijk}) \right]^2
 \end{aligned}$$

Calculando-se algebricamente o valor esperado da expressão acima tem-se que:

$$E[SQ(\Psi_h)] = E[QM(\Psi_h)] = \sigma^2 + \frac{R}{K_h} \sum_k c_{h_k} \mu_k \quad (2.34)$$

2.7.6 Esperança das somas de quadrado e do quadrado médio dos resíduos

Por definição, tem-se que:

$$SQResiduo = SQTotal - SQLinhas - SQColunas - SQTrat.$$

Assim,

$$\begin{aligned}
 E(SQResiduo) &= E(SQTotal - SQLinhas - SQColunas - SQTrat.) \\
 &= E(SQTotal) - E(SQLinhas) - E(SQColunas) - E(SQTrat.)
 \end{aligned}$$

$$\begin{aligned}
&= \sum_i \alpha_i^2 + R \sum_j \beta_j^2 + R \sum_k \tau_k^2 + (R^2 - 1)\sigma^2 - \left(R \sum_i \alpha_i^2 + (R - 1)\sigma^2 \right) \\
&\quad - \left(R \sum_j \beta_j^2 + (R - 1)\sigma^2 \right) - \left(R \sum_k \tau_k^2 + (R - 1)\sigma^2 \right) \\
&= (R^2 - 1)\sigma^2 - 3 \times (R - 1)\sigma^2 \\
&= R^2\sigma^2 - 3R\sigma^2 + 2\sigma^2 \\
&= (R - 2)(R - 1)\sigma^2
\end{aligned}$$

Tendo que:

$$QMResiduo = \frac{SQResiduo}{(R - 2)(R - 1)}$$

Logo:

$$E(QMResiduo) = E \left[\frac{SQResiduo}{(R - 2)(R - 1)} \right] = \frac{1}{(R - 2)(R - 1)} (R - 2)(R - 1)\sigma^2 = \sigma^2$$

Como sugere (BARBIN, 1993), resume-se os resultados obtidos em uma tabela de análise de variância.

Tabela 2: A análise de variância com os valores esperados dos quadrados médios.

<i>F.V</i>	<i>G.L</i>	<i>S.Q</i>	<i>E(Q.M)</i>
<i>Linhas</i>	$R - 1$	$\frac{1}{R} \sum_i y_{i..} - C$	$\sigma^2 + \frac{R}{R-1} \sum_i \alpha_i^2 = \sigma^2 + \frac{R}{R-1} \sum_i (\mu_i - \mu)^2$
<i>Colunas</i>	$R - 1$	$\frac{1}{R} \sum_j y_{.j} - C$	$\sigma^2 + \frac{R}{R-1} \sum_j \beta_j^2 = \sigma^2 + \frac{R}{R-1} \sum_j (\mu_j - \mu)^2$
<i>Trat.</i>	$R - 1$	$\frac{1}{R} \sum_k y_{..k} - C$	$\sigma^2 + \frac{R}{R-1} \sum_k \tau_k^2 = \sigma^2 + \frac{R}{R-1} \sum_k (\mu_k - \mu)^2$
<i>Residuo</i>	$(R - 2)(R - 1)$	$\sum_{i,j} (y_{ijk} - \hat{y}_{ijk})^2$	σ^2
<i>Total</i>	$R^2 - 1$	$\sum_{i,j} y_{ijk}^2 - C$	-

2.8 Análises estatísticas

No delineamento em quadrado latino, o efeito de linha e colunas no delineamento servem, na maioria das vezes, para estabelecer o controle local do ambiente no qual estão sendo alocados os tratamentos, não sendo necessário fazer inferências sobre estes efeitos. O que mais nos interessa no estudo em questão são o efeito dos tratamentos, os contrastes

de interesse entre as médias dos tratamentos e a comparação múltipla das médias dos tratamentos.

As hipóteses de interesse a serem testadas neste trabalho são:

$$\left\{ \begin{array}{l} H_0^{(\tau)} : \tau_k = 0, \forall k = 1, 2, \dots \\ H_1^{(\tau)} : \tau_k \neq 0, \text{ para pelo menos um } \tau_k \end{array} \right. \text{ ou } \left\{ \begin{array}{l} H_0^{(\tau)} : \mu_k = \mu_{k'}, \forall k = k' \\ H_1^{(\tau)} : \mu_k \neq \mu_{k'}, \forall k \neq k' \end{array} \right.$$

$$\left\{ \begin{array}{l} H_0^{(\alpha)} : \alpha_i = 0, \forall i = 1, 2, \dots \\ H_1^{(\alpha)} : \alpha_i \neq 0, \text{ para pelo menos um } \alpha_i \end{array} \right. \text{ ou } \left\{ \begin{array}{l} H_0^{(\alpha)} : \mu_i = \mu_{i'}, \forall i = i' \\ H_1^{(\alpha)} : \mu_i \neq \mu_{i'}, \forall i \neq i' \end{array} \right.$$

$$\left\{ \begin{array}{l} H_0^{(\beta)} : \beta_j = 0, \forall j = 1, 2, \dots \\ H_1^{(\beta)} : \beta_j \neq 0, \text{ para pelo menos um } \beta_j \end{array} \right. \text{ ou } \left\{ \begin{array}{l} H_0^{(\beta)} : \mu_j = \mu_{j'}, \forall j = j' \\ H_1^{(\beta)} : \mu_j \neq \mu_{j'}, \forall j \neq j' \end{array} \right.$$

$$\left\{ \begin{array}{l} H_0^{(\Psi)} : \Psi_h = \sum_k c_{h_k} \mu_k = 0, \\ H_1^{(\Psi)} : \Psi_h = \sum_k c_{h_k} \mu_k \neq 0 \end{array} \right.$$

Considerando-se os resultados encontrados no desenvolvimento teórico deste trabalho, pode-se verificar que:

1. De acordo com o estudo realizado $E(QMResiduo) = \sigma^2$, isto é, o quadrado médio do resíduo é um estimador não viciado para σ^2 , independentemente de que $H_0^{(\tau)}$ seja verdadeiro.
2. Observou-se também que $E(QMTratamento) = \sigma^2 + \frac{R}{R-1} \sum_k (\mu_k - \mu)^2$. Caso $H_0^{(\tau)} : \mu_1 = \mu_2 = \dots = \mu_k = 0$ seja verdadeira, então $\frac{R}{R-1} \sum_k (\mu_k - \mu)^2 = 0$ e com isso o QMTratamento será um estimador não viciado para σ^2 . No entanto, se $H_0^{(\tau)}$ for falsa, então $E(QMTratamento) > \sigma^2$. Assim pode-se comparar QMTratamento e QMResiduo para testar se $H_0^{(\tau)}$ é verdadeira, tendo em vista que quanto maior for o QMTratamento comparado ao QMResiduo mais evidências se tem que $H_0^{(\tau)}$ será rejeitada.
3. Foi mostrado também que $E(QMLinhas) = \sigma^2 + \frac{R}{R-1} \sum_i (\mu_i - \mu)^2$. Caso $H_0^{(\alpha)} : \mu_1 = \mu_2 = \dots = \mu_i = 0$ seja verdadeira, então $\frac{R}{R-1} \sum_k (\mu_k - \mu)^2 = 0$ e com isso o QMLinha será um estimador não viciado para σ^2 . No entanto, se $H_0^{(\alpha)}$ for falsa,

então $E(QMLinha) > \sigma^2$. Assim pode-se comparar QMLinha e QMResiduo para testar se $H_0^{(\alpha)}$ é verdadeira, tendo em vista que quanto maior for o QMLinha comparado ao QMResiduo mais evidências se tem que $H_0^{(\alpha)}$ será rejeitada.

4. De modo análogo ao item 2 e 3 se testa $H_0^{(\beta)}$
5. Da mesma forma como $E[QM(\Psi)] = \sigma^2 + \frac{R}{K_h} \sum_k c_{h_k} \mu_k$ caso $H_0^{(\Psi)} : \sum_k c_{h_k} \mu_k = 0$ seja verdadeira, então $QM(\Psi)$ será um estimador não viciado para σ^2 . Porém, se $H_0^{(\Psi)}$ for falsa, então $E[QM(\Psi)] > \sigma^2$. Assim pode-se comparar $QM(\Psi)$ e QMResiduo para testar se $H_0^{(\Psi)}$ é verdadeira, tendo em vista que quanto maior for o $QM(\Psi)$ comparado ao QMResiduo mais evidências se tem que $H_0^{(\Psi)}$ será rejeitada.
6. Segundo Angela e Daniel (1999), a razão entre duas variáveis aleatórias qui-quadrado independentes, cada um dividido por seus graus de liberdade, tem uma distribuição F de snedecor. Tomando-se como base os resultados encontrados, pode-se concluir que:

$$F_\tau = \frac{\frac{SQTrat.}{\sigma^2(R-1)}}{\frac{SQResiduo}{\sigma^2(R-2)(R-1)}} = \frac{QMTrat.}{QMResiduo} \sim F_{[(R-1); (R-2)(R-1)]}$$

$$F_\alpha = \frac{\frac{SQLinhas}{\sigma^2(R-1)}}{\frac{SQResiduo}{\sigma^2(R-2)(R-1)}} = \frac{QMLinhas}{QMResiduo} \sim F_{[(R-1); (R-2)(R-1)]}$$

$$F_\beta = \frac{\frac{SQColunas}{\sigma^2(R-1)}}{\frac{SQResiduo}{\sigma^2(R-2)(R-1)}} = \frac{QMColunas}{QMResiduo} \sim F_{[(R-1); (R-2)(R-1)]}$$

$$F_\Psi = \frac{\frac{SQ[\Psi_h]}{\sigma^2}}{\frac{SQResiduo}{\sigma^2(R-2)(R-1)}} = \frac{QM(\Psi_h)}{QMResiduo} \sim F_{[1; (R-2)(R-1)]}$$

7. Os resultados apresentados no item 6 serão as estatísticas do teste para testar as Hipóteses $H_0^{(\tau)}, H_0^{(\alpha)}, H_0^{(\beta)}, H_0^{(\Psi)}$, para as quais nossa regra de decisão será:

Se $F_\tau = \frac{QMTrat.}{QMResiduo} < F_{[(R-1); (R-2)(R-1); \omega]}$, não rejeita-se $H_0^{(\tau)}$ ao nível de significância ω , isto é, as médias dos tratamentos não diferem entre si, caso contrário, rejeita-se $H_0^{(\tau)}$, isto é, pelo menos um par de médias dos tratamentos diferem entre si.

Se $F_\alpha = \frac{QMLinhas}{QMResiduo} < F_{[(R-1); (R-2)(R-1); \omega]}$, não rejeita-se $H_0^{(\alpha)}$ ao nível de significância ω , isto é, as médias das linhas não diferem entre si, caso contrário rejeita-se $H_0^{(\alpha)}$, isto é, pelo menos um par de médias das linhas diferem entre si.

Se $F_\beta = \frac{QMColunas}{QMResiduo} < F_{[(R-1); (R-2)(R-1); \omega]}$, não rejeita-se $H_0^{(\beta)}$ ao nível de significância ω , isto é, as médias das colunas não diferem entre si, caso contrário rejeita-se $H_0^{(\beta)}$, isto é, pelo menos um par de médias das colunas diferem entre si.

Se $F_\Psi = \frac{QM(\Psi_h)}{QMResiduo} < F_{[(1); (R-2)(R-1); \omega]}$, não rejeita-se $H_0^{(\Psi)}$ ao nível de significância ω , isso quer dizer que a média dos contrastes não diferem entre si. Caso contrário rejeita-se $H_0^{(\Psi)}$ isso é a médias dos contrastes diferem entre si.

2.9 Estruturação da tabela da análise de variância - ANOVA

Com base nos resultados demonstrados nas seções anteriores e com o objetivo de fazer inferência sobre os efeitos das linhas, das colunas e dos tratamentos sobre a variável resposta ou sobre a igualdade das médias das linhas, colunas e tratamentos, organiza-se a seguinte tabela:

Tabela 3: A análise de variância para testar as hipóteses da não existência de efeitos das linhas, das colunas e dos tratamentos ou de igualdade das médias das linhas, das colunas e dos tratamentos.

F.V	G.L	S.Q	Q.M	F
Linha	$R - 1$	$SQLinhas$	$QMLinhas = \frac{SQLinhas}{R-1}$	$F_\alpha = \frac{QMLinhas}{QMResiduo}$
Colunas	$R - 1$	$SQColunas$	$QMColunas = \frac{SQColunas}{R-1}$	$F_\beta = \frac{QMColunas}{QMResiduo}$
Trat.	$R - 1$	$SQTrat.$	$QMTrat. = \frac{SQTrat.}{R-1}$	$F_\tau = \frac{QMTrat.}{QMResiduo}$
Residuo	$(R - 2)(R - 1)$	$SQResiduo$	$QMResiduo = \frac{SQResiduo}{(R-2)(R-1)}$	-
Total	$R^2 - 1$	$SQTtotal$	-	-

2.10 Comparação de médias duas a duas dos tratamentos, das linhas e das colunas

O desenvolvimento das teorias de comparações múltiplas de médias não faz parte dos objetivos deste trabalho. Sendo assim, apresenta-se apenas um procedimento prático para comparação de médias duas a duas, o teste de Tukey.

Segundo Porras (1998), o teste de Tukey é baseado na amplitude estudentizada e pode ser utilizado para comparar todo e qualquer contraste entre duas médias de tratamentos.

As hipóteses de interesse deste teste são:

$$\begin{cases} H_0 : \mu_k = \mu_{k'} \\ H_1 : \mu_k \neq \mu_{k'}, \forall k \neq k' \end{cases}$$

A estatística do teste é dada por:

$$\Delta = q_{[R; (R-2)(R-1); \omega]} \sqrt{\frac{QMResiduo}{R}},$$

em que $q_{[R; (R-2)(R-1); \omega]}$ é o valor de tabela da amplitude total estudentizada de Tukey correspondente ao número de tratamentos R , número de graus de liberdade do resíduo $(R-2)(R-1)$, e o nível de significância ω .

De acordo com a estatística calculada, toma-se a seguinte conclusão:

Se $|m_i - m_{i'}| < \Delta$, não rejeita-se H_0 ao nível de significância ω , caso contrário rejeita-se H_0 .

3 Aplicação em um exemplo real

Vamos agora apresentar em um exemplo prático a teoria desenvolvida neste trabalho, o exemplo apresentado foi retirado do livro de Gomes (1985)

3.1 Um exemplo de aplicação

Num experimento de competição de variedades de cana de açúcar foram utilizadas 5 variedades: V1=Co290, V2=Co412, V3=Co419, V4=POJ2827, V5=CO36-13, dispostas em um quadrado latino 5x5. As produções de cada variedade, em *kg/parcelas*, são dadas na tabela 5.

Tabela 4: Produção *kg/parcelas* de cinco variedades de cana-de-açúcar.

Linhas (L_i)	Colunas (C_j)					Total $y_{i..}$	Media $\bar{y}_{i..}$
	C_1	C_2	C_3	C_4	C_5		
L_1	V4(432)	V1(518)	V2(458)	V3(583)	V5(331)	2322	464,4
L_2	V3(724)	V5(478)	V1(524)	V2(550)	V4(400)	2676	535,2
L_3	V5(489)	V2(384)	V3(556)	V4(297)	V1(420)	2146	429,2
L_4	V2(494)	V4(500)	V5(313)	V1(486)	V3(501)	2294	458,8
L_5	V1(515)	V3(660)	V4(438)	V5(394)	V2(318)	2325	465,0
Total $y_{.j}$	2654	2540	2289	2310	1970	11763	-
Media $\bar{y}_{..}$	530,8	508,0	457,8	462,0	394,0	-	470,52

Os totais e as estimativa das médias dos tratamentos, isto é; $y_{.k}$ e $\bar{y}_{..k}$, para $k = 1, 2, 3, 4, 5$. foram os seguintes:

$$\begin{aligned}
y_{..1} &= 518 + 524 + 420 + 486 + 515 = 2463 \text{ e } \bar{y}_{..1} = 492,6; \\
y_{..2} &= 458 + 550 + 384 + 494 + 318 = 2204 \text{ e } \bar{y}_{..2} = 440,8; \\
y_{..3} &= 583 + 724 + 556 + 501 + 660 = 3024 \text{ e } \bar{y}_{..3} = 604,8; \\
y_{..4} &= 432 + 400 + 297 + 500 + 438 = 2067 \text{ e } \bar{y}_{..4} = 413,4; \\
y_{..5} &= 331 + 478 + 489 + 313 + 394 = 2005 \text{ e } \bar{y}_{..5} = 401,0.
\end{aligned}$$

3.1.1 Cálculo das somas de quadrado e análise de variância

A partir dos dados da subseção anterior, obtemos os seguintes resultados:

$$C = \frac{y^2}{R^2} = \frac{2325^2}{5^2} = 5534726,76$$

$$\begin{aligned}
SQ_{Total} &= \sum_{i,j} y_{ijk}^2 - C \\
&= (432^2 + 518^2 + \dots + 318^2) - 5534726,76 \\
&= 257724,24
\end{aligned}$$

$$\begin{aligned}
SQ_{Linhas} &= \frac{1}{R} \sum_i y_{i..}^2 - C \\
&= \frac{1}{5} [2322^2 + 2676^2 + \dots + 2325^2] - 5534726,76 \\
&= 30480,64
\end{aligned}$$

$$\begin{aligned}
SQ_{Colunas} &= \frac{1}{R} \sum_j y_{.j.}^2 - C \\
&= \frac{1}{5} [2654^2 + 2540^2 + \dots + 1970^2] - 5534726,76 \\
&= 55640,64
\end{aligned}$$

$$\begin{aligned}
SQ_{Trat.} &= \frac{1}{R} \sum_k y_{.k}^2 - C \\
&= \frac{1}{5} [2463^2 + 2204^2 + \dots + 2005^2] - 5534726,76
\end{aligned}$$

$$= 137488,72$$

$$\begin{aligned} SQResiduo &= SQTotal - SQLinhas - SQColunas - SQTrat. \\ &= 257724,24 - 30480,64 - 55640,64 - 137488,72 \\ &= 34114,72 \end{aligned}$$

Com estes resultados, podemos organizar a tabela de análise de variância.

Tabela 5: Análise de variância para testar os efeitos das linhas, colunas e variedades sobre a produção (ton/ha).

F.Variação	G.L	S.Q	Q.M	F
Linhas($H_0^{(\alpha)} : \alpha_1 = \alpha_2 = \dots = \alpha_5 = 0$)	4	30480,64	7620,16	2,68
Colunas($H_0^{(\beta)} : \beta_1 = \beta_2 = \dots = \beta_5 = 0$)	4	55640,64	13910,16	4,89*
Trat.($H_0^{(\tau)} : \tau_1 = \tau_2 = \dots = \tau_5 = 0$)	4	137488,72	34372,06	12,09**
Residuo	12	34114,72	2842,89	-
Total	24	257724,24	-	-

Pode-se observar que os efeitos das linhas sob a variável resposta não foram estatisticamente significativo pelo teste F, isto é, $H_0^{(\alpha)} : \alpha_1 = \alpha_2 = \dots = \alpha_5 = 0$ não rejeita-se ao nível de 5% de significância. Por outro lado, $H_0^{(\beta)} : \beta_1 = \beta_2 = \dots = \beta_5 = 0$ foi rejeitada ao nível de significância de 5%, isso quer dizer que pelo menos um par de médias relativo as colunas diferem estatisticamente entre si. O mesmo acontece com a hipótese $H_0^{(\tau)} : \tau_1 = \tau_2 = \dots = \tau_5 = 0$ que foi rejeitada ao nível de significância de 1%, assim, pode-se afirmar que pelo menos um par de produção de variedade diferem estatisticamente entre si.

Como foi dito, pode-se desdobrar os graus de liberdade dos tratamentos em contrastes de interesse. Suponha que seja do interesse do pesquisador verificar as seguintes hipóteses sobre as médias das variedades:

Se a média combinada das variedades V1=Co290, V2=Co412 e V3=Co419 diferem da média combinada das variedades V4=POJ2827 e V5=CO36-13.

Se a média combinada das variedades V1=Co290 e V2=Co412 diferem da média da variedade V3=Co419.

Se a média da variedade V1=Co290 diferem da média da variedade V2=Co412.

Se a média da variedade V4=POJ2827 diferem da média da variedade V5=CO36-13.

Com isso, pode-se formular as seguintes hipóteses:

$$H_0^{(\Psi_1)} : \frac{\mu_1 + \mu_2 + \mu_3}{3} = \frac{\mu_4 + \mu_5}{2} \quad \text{vs} \quad H_1^{(\Psi_1)} : \frac{\mu_1 + \mu_2 + \mu_3}{3} \neq \frac{\mu_4 + \mu_5}{2}$$

$$H_0^{(\Psi_2)} : \frac{\mu_1 + \mu_2}{2} = \mu_3 \quad \text{vs} \quad H_1^{(\Psi_2)} : \frac{\mu_1 + \mu_2}{2} \neq \mu_3$$

$$H_0^{(\Psi_3)} : \mu_1 = \mu_2 \quad \text{vs} \quad H_1^{(\Psi_3)} : \mu_1 \neq \mu_2$$

$$H_0^{(\Psi_4)} : \mu_4 = \mu_5 \quad \text{vs} \quad H_1^{(\Psi_4)} : \mu_4 \neq \mu_5$$

Calcula-se agora as estimativas dos contrastes de interesse.

$$\hat{\psi}_1 = \frac{m_1 + m_2 + m_3}{3} - \frac{m_4 + m_5}{2} = \frac{492,6 + 440,8 + 604,8}{3} - \frac{413,4 + 401,0}{2} = 105,53 \text{kg/parcela}$$

$$\hat{\psi}_2 = \frac{m_1 + m_2 + m_3}{3} - \frac{m_4 + m_5}{2} = \frac{492,6 + 440,8}{2} - 604,8 = -138,1 \text{kg/parcela}$$

$$\hat{\psi}_3 = m_1 - m_2 = 492,6 - 440,8 = 51,8 \text{kg/parcela}$$

$$\hat{\psi}_4 = m_4 - m_5 = 413,4 - 401,0 = 12,4 \text{kg/parcela}$$

Os resultados apresentados acima podem ser interpretados da seguinte forma: no contraste ($\hat{\psi}_1$) as variedades V1=Co290, V2=Co412 e V3=Co419 produzem em média 105,53kg/parcela a mais que as variedades V4=POJ2827 e V5=CO36-13; no contraste ($\hat{\psi}_2$) as variedades V1=Co290 e V2=Co412 produzem em média 138,1kg/parcela a menos que a variedade V3=Co419. De modo análogo, os demais contrastes serão interpretados.

Desta forma calcula-se as somas dos quadrados dos contrastes que será:

$$\begin{aligned} SQ(\Psi_1) &= \frac{(\sum_k C_{h_k} y_{..k})^2}{RK_h} = \frac{[2 \times (2463) + 2 \times (2204) + 2 \times (3024) - 3 \times (2067) - 3 \times (2005)]^2}{30 \times 5} = 66823,7067; \\ SQ(\Psi_2) &= \frac{(\sum_k C_{h_k} y_{..k})^2}{RK_h} = \frac{[2463 + 2204 - 2 \times (3024) - 0 \times (2067) - 0 \times (2005)]^2}{6 \times 5} = 63572,0333; \\ SQ(\Psi_3) &= \frac{(\sum_k C_{h_k} y_{..k})^2}{RK_h} = \frac{[2463 - 2204]^2}{2 \times 5} = 6708,1000; \\ SQ(\Psi_4) &= \frac{(\sum_k C_{h_k} y_{..k})^2}{RK_h} = \frac{[(2067) - (2005)]^2}{30 \times 5} = 384,400. \end{aligned}$$

Montando a tabela de análise de variância, tem-se que:

Tabela 6: Reestruturação da análise de variância para testar contrastes de interesse relativos às médias de produção das variedades de cana-de-açúcar estudadas.

F.Variação	G.L	S.Q	Q.M	F
$H_0^{(\Psi_1)} : \frac{\mu_1+\mu_2+\mu_3}{3} = \frac{\mu_4+\mu_5}{2}$	1	66823,7067	66823,7067	23,51**
$H_0^{(\Psi_2)} : \frac{\mu_1+\mu_2}{2} = \mu_3$	1	63572,0333	63572,0333	22,36**
$H_0^{(\Psi_3)} : \mu_1 = \mu_2$	1	6708,1000	6708,1000	2,36
$H_0^{(\Psi_4)} : \mu_4 = \mu_5$	1	384,400	384,400	0,14
Trat.	(4)	(137488,72)	(-)	(-)
Linhas($H_0^{(\alpha)} : \alpha_1 = \dots = \alpha_5 = 0$)	4	30480,64	7620,16	2,68
Colunas($H_0^{(\beta)} : \beta_1 = \dots = \beta_5 = 0$)	4	55640,64	13910,16	4,89*
Residuo	12	34114,72	2842,89	-
Total	24	257724,24	-	-

Segundo a tabela da análise de variância acima, podemos retirar as seguintes conclusões:

A hipótese $H_0^{(\Psi_1)}$ e $H_0^{(\Psi_2)}$ foram rejeitadas ao nível de 1%, isto significa que a média de produção dos tratamentos $V1$, $V2$ e $V3$ combinados, diferem significativamente da média da produção dos tratamentos $V4$ e $V5$ combinados, e que a média da produção dos tratamentos $V1$ e $V2$ diferem significativamente da média da produção do tratamento $V3$;

A hipótese $H_0^{(\Psi_3)}$ e $H_0^{(\Psi_4)}$ não foram rejeitadas ao nível de 5% , isto significa que a média da produção do tratamento $V1$ não difere estatisticamente da média da produção do tratamento $V2$ e que, a média da produção do tratamento $V4$ não difere estatisticamente da média da produção do tratamento $V5$.

3.1.2 Comparação de médias duas a duas pelo teste de tukey

Ao aplicar-se o teste de Tukey, coloca-se a seguinte hipótese,

$$\begin{cases} H_0 : \mu_k = \mu_{k'} \\ H_1 : \mu_k \neq \mu_{k'}, \forall k \neq k' \end{cases}$$

A estatística do teste é dada por:

$$\Delta = q_{[R; (R-2)(R-1); \omega]} \sqrt{\frac{QMResiduo}{R}}$$

Assim

$$\Delta = q_{[5; 12; 0,05]} \sqrt{\frac{QMResiduo}{R}} = 4,51 \sqrt{\frac{2842,89}{5}} = 107,54kg/parcela$$

Isso quer dizer que as médias na qual a sua diferença for maior que $107,54kg/parcela$ diferem estatisticamente ao nível de significância de 5%. A seguir, encontram-se apresentado as estimativas das médias de produção das variedades, todas com um erro-padrão $s(\hat{m}) = 23,84kg/parcelas$.

$$\bar{y}_{..3} = 604,8 \quad a$$

$$\bar{y}_{..1} = 492,6 \quad b$$

$$\bar{y}_{..2} = 440,8 \quad b$$

$$\bar{y}_{..4} = 413,4 \quad b$$

$$\bar{y}_{..5} = 401,0 \quad b$$

No confronto das médias de produção de variedades, observou-se que apenas a variedade V3 difere das demais. Os demais contrastes entre duas médias não foram estatisticamente significativos ao nível de significancia de 5% pelo teste Tukey.

3.1.3 Comprovação da idoneidade do modelo

Antes de se fazer conclusão sobre os resultados da pesquisa, é importante fazer uma análise dos resíduos quanto aos pressupostos básicos da experimentação (aditividade, normalidade, independência e homogeneidade).

```

> ### Teste de não-aditividade. tire conclusões.
> cana_de_açucar.lm <- lm(y ~ Rows + Columns + trat)
> anova(cana_de_açucar.lm , test="F")
Analysis of Variance Table

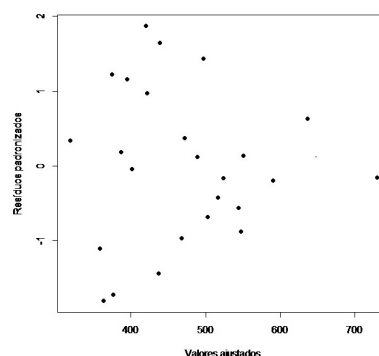
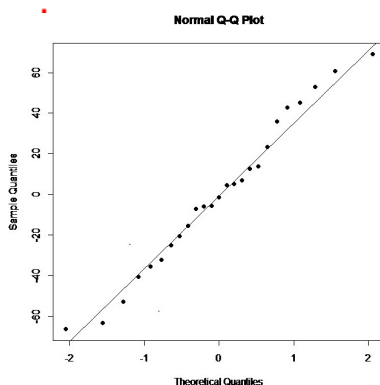
Response: y
      Df Sum Sq Mean Sq F value    Pr(>F)
Rows    4  30481     7620  2.6804 0.0831343 .
Columns  4  55641    13910  4.8930 0.0142293 *
trat    4 137488    34372 12.0905 0.0003585 ***
Residuals 12  34115     2843
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
> LP2 <- (predict(cana_de_açucar.lm))^2
> cana_de_açucar.lm2 <-update(cana_de_açucar.lm, .~. +LP2)
> anova(cana de açúcar.lm ,cana_de_açucar.lm2, test="F")
# a diferença entre os modelos
Analysis of Variance Table

Model 1: y ~ Rows + Columns + trat
Model 2: y ~ Rows + Columns + trat + LP2
      Res.Df  RSS Df Sum of Sq      F Pr(>F)
1         12 34115
2         11 33897  1    217.76 0.0707 0.7953
> # Portanto aceita-se rej Ho: (gama)=0, ou seja,
existem evidências de aditividade do modelo.
> ### 2) Homogeneidade de variâncias
>
> ### Teste do F-máximo
> F_cal = max(Var)/min(Var); F_cal
[1] 4.479106
> # Ho: as variâncias são homogêneas
> # Ha: as variâncias não são homogêneas
> # ver o F na tabela do F-máximo_tab(K,v):
k estimativas de variâncias com v graus de liberdade cada uma (repetição ou blocos)
.:
.: #F_tab(5,4) = 25.2
.: # Assim, F_cal<F_tab ==> aceita-se Ho, logo as variâncias são homogêneas!.
> ### Teste de normalidade Shapiro Wilk
> # H0: é normal
> # Ha: não é normal
> shapiro.test(cana_de_açucar.lm$res)

      Shapiro-Wilk normality test

data:  cana_de_açucar.lm$res
W = 0.977, p-value = 0.8202

```



4 Conclusão

No delineamento em quadrado latino, o modelo estatístico associado aos dados do experimento é um modelo aditivo que envolve a média geral(μ), o efeito de linhas(α_i), colunas(β_j) e tratamento(τ_k), e o erro experimental do experimento(ϵ_{ijk}) no qual $\epsilon_{ijk} \sim N(0, \sigma^2)$. Diante dessa afirmação e do que foi visto no desenvolvimento da teoria deste trabalho, podemos afirmar que:

1. Do ponto de vista prático, o delineamento em quadrado latino é pouco utilizado por ser um delineamento restritivo, pois o número de linhas, colunas e tratamentos devem ser os mesmos. E do ponto de vista teórico, as contribuições têm sido poucas, dificultando a expansão das técnicas estatísticas, envolvendo outras distribuições de probabilidade para os erros neste delineamento.
2. Para encontrar os estimadores dos parâmetros do modelo, foi utilizado o método dos mínimos quadrados, que consiste em minimizar a soma dos quadrados dos erros do modelo, é um método simples de ser utilizado e possui excelentes propriedades.
3. As estatísticas do teste são F_τ , F_α , F_β , F_Ψ no qual todos seguem distribuição F de snedecor.
4. Do ponto de vista teórico, nenhuma dificuldade foi encontrada quando se considerou que os erros experimentais seguem distribuição normal.

Referências

ALEGRI, M. Quadrados latinos e aplicações. p. 80 pg, 2006.

ANGELA, D.; DANIEL, V. *Design and analysis of experiments*. [S.l.]: New York: Spring-Verlag, 1999.

BARBIN, D. *Componentes de Variância - Teoria e Aplicações*. 2. ed. Piracicaba - SP: FEALQ - Fundação de Estudos Agrários Luiz de Queiroz, 1993.

COX, D. R.; REID, N. *The theory of the design of experiments*. [S.l.]: CRC Press, 2002.

GOMES, F. P. *Curso de Estatística Experimental*. 4. ed. Piracicaba - SP: Nobel, 1985.

LUNA, J. G. de. Apostila de planejamento de experimentos. 2010.

MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. *Introduction to the theory of Statistics*. 3. ed. Singapore: Mcgraw-Hill International Editions, 1974.

ANA MARIA LARA PORRAS, J. G. L. e. *Análisis de la varianza*. Granada - ES: Grupo editorial univessitario, 1998.

ROHATGI, V. K. *An Introduction to Probability Theory and Mathematical Statistics*. New York: John Wiley, 1976.

ROUSSAS, G. G. *A Course in Mathematical Statistics*. 2. ed. New York: Academic Press, 1997.

SANTOS, B. H. G. dos. Aspectos teóricos e práticos de um experimento em blocos completos casualizados com repetições dentro dos blocos. 2010.