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

Márcia Marciele dos Santos Silva

Análise estatística de um experimento em blocos ao acaso abordando os aspectos teóricos e práticos

Campina Grande
Junho de 2015

Márcia Marciele dos Santos Silva

Análise estatística de um experimento em blocos ao acaso abordando os aspectos teóricos e práticos

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 as exigências legais para obtenção do título de bacharel em Estatística.

Orientadora:
Nyedja Fialho Morais Barbosa

Campina Grande
Junho de 2015

S237a Silva, Márcia Marciele dos Santos.
Análise estatística de um experimento em blocos ao acaso abordando os aspectos teóricos e práticos. [manuscrito] / Marcia Marciele dos Santos Silva. - 2015.
54 p.

Digitado.
Trabalho de Conclusão de Curso (Graduação em Estatística) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2015.
"Orientação: Profa. Ma. Nyedja Fialho Morais Barbosa, Departamento de Estatística".

1. Estatística experimental. 2. Delineamento em blocos casualizados. 3. Testes de comparações múltiplas. I. Título.

21. ed. CDD 519.5

Márcia Marciele dos Santos Silva

Análise Estatística de um experimento em blocos ao acaso abordando os aspectos teóricos e práticos

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 as exigências legais para obtenção do título de bacharel em Estatística.

Aprovado em: 17/06/2015

Banca Examinadora:

Nyedja Fialho Morais Barbosa

Profª Ma. Nyedja Fialho Morais Barbosa
UEPB - DE/CCT
Orientadora

Ana Patrícia Bastos Peixoto

Profª Drª Ana Patrícia Bastos Peixoto
UEPB - DE/CCT
Examinadora

Tiago Almeida de Oliveira

Profº. Drº Tiago Almeida de Oliveira
UEPB - DE/CCT
Examinador

Dedicatória

A DEUS que me deu força para enfrentar as dificuldades e que me proporcionou para que conseguisse chegar até aqui.

Aos meus pais, Antonio José e Maria Lusinete,

Aos meus irmãos, José Severino e Marcio Maciel,

Ao meu esposo Rubeilson, por todo amor, cuidado, apoio, confiança e paciência.

Agradecimentos

Primeiramente, a Deus, por ter me concedido a vida, por me dar forças para seguir em frente e ter permitido em minha vida mais uma vitória.

A toda a minha família, em especial aos meus pais Antonio José da Silva e Maria Lusinete dos Santos Silva, que são os responsáveis pelo que sou hoje. Aos meus irmãos, José Severino e Márcio Maciel pelo carinho, apoio e dedicação.

Ao meu esposo Rubeilson dos Santos, pela paciência, apoio, incentivo e por todo amor.

À minha orientadora Nyedja Fialho Morais Barbosa, pela paciência, amizade e dedicação.

À UEPB por proporcionar condições para meu aprendizado durante esses quatro anos e meio. Especialmente, aos professores, Ana Patrícia, Tiago, Silvio, Gustavo, Juarez, João Gil e aos demais que tanto contribuíram para minha formação.

À minha amiga e companheira de curso, Andreza Jardelino por toda amizade, confiança, incentivo, por todos os momentos vividos juntos que jamais esquecerei.

Gostaria de agradecer também a todos os meus amigos Dark, Josevânia, Geilza, Poliana, Cleber, Simone, Francilene que sempre me deram forças para chegar até aqui.

A todos que acreditaram em mim e fazem parte da minha vida, o meu muito obrigada, pois, sem vocês nunca teria conseguido realizar mais este sonho.

"Diga o que você pensa com esperança.
Pense no que você faz com fé.
Faça o que você deve fazer com amor!"

Ana Carolina.

Resumo

A estatística experimental é um conjunto de procedimentos que se preocupa em obter conclusões a partir de dados experimentais e é utilizada em várias áreas do conhecimento. Por meio do planejamento de experimentos é possível testar hipóteses de interesse a partir de procedimentos estatísticos como análise de variância e testes de comparações múltiplas com a finalidade de oferecer ao pesquisador subsídios suficientes para a tomada de decisões. Este trabalho tem como objetivo apresentar os aspectos téoricos referentes a utilização da estatística experimental, sobretudo na análise estatística de um experimento em blocos ao acaso. Portanto, foram apresentados alguns conceitos utlilizados na experimentação, definindo o modelo matemático associado às observações experimentais e utilizando a técnica da análise de variância para comparar os efeitos dos tratamentos e dos blocos com o intuito de aplicar os testes de comparacões múltiplas.

Palavras-chave: Estatística experimental; Delineamento em blocos casualizados; Testes de comparações múltiplas

Abstract

The experimental statistic is a set of procedures that cares to draw conclusions from experimental data and is used in various fields of knowledge. Through the design of experiments it is possible to test hypotheses of interest from statistical procedures such as analysis of variance and multiple comparison tests in order to offer the researcher sufficient subsidies for decision making. This work aims to present the theoretical aspects about the use of experimental statistics, particularly in the statistical analysis of an experiment in randomized blocks. So we were shown some utlilizados concepts on trial by setting the mathematical model associated with the experimental observations and using the technique of analysis of variance to compare the effects of treatments and blocks in order to apply the multiple comparison tests.

Keywords: Experimental statistics; randomized block design; Multiple comparison tests

Sumário

Lista de Tabelas

1 Introdução

Os fundamentos da mordena estatística teórica e experimental surgiram a partir do início do século 20, principalmente na Inglaterra. Grande parte dos desenvolvimentos realizados é devido a Ronald Fisher, que trabalhou na Rothamsted Experimental Sation localizado ao norte de Londres (RESENDE, 2007).

Seguindo R.A. Fisher, podemos definir a estatística como a parte da matemática aplicada aos dados observados. Quando tais dados, em muitos casos, colhidos através de trabalhos feitos propositalmente e em condições previamente especificadas: temos então dados experimentais, obtidos de experimentos (GOMES, 1990). O objetivo da estatística experimental é constituído pelo planejamento, execução e análise dados.

Em uma pesquisa científica o procedimento geral é formular hipóteses de interesse e verificá-las diretamente ou por suas consequências. As hipóteses são verificadas com a utilização de métodos de análise estatística que dependem da maneira sob a qual as observações foram obtidas. Portanto, planejamento de experimentos e análise dos resultados estão intimamente ligados e devem ser utilizadas em sequência nas pesquisas científicas das diversas áreas do conhecimento.

O experimentador tem um papel muito importante e o que dificulta o seu trabalho é que na análise estatística exige a presença, em todos os dados obtidos, de efeitos de fatores não controlados que podem ser controláveis ou não, como por exemplo, pequenas diferenças de fertilidade do solo, na semeadura, na genética, etc. Esses efeitos são conhecidos como variação ao acaso ou variação aleatória, dessa maneira ocorrendo um efeito da variação aleatória pode mudar totalmente os resultados do experimento.

O intuito principal deste trabalho é apresentar os aspectos teóricos referentes a utilização da estatística experimental, sobretudo na análise estatística de um experimento em blocos ao acaso. Para isto será utilizado um exemplo de variedades de mandioca proposto por Gomes (1990), oriundo de um delineamento em blocos ao acaso para esclarecer teoria e interpretações adequadas sobre o problema proposto.

2 Fundamentação Teórica

O conteúdo desta seção relata os principais aspectos referentes a utilização da estatística experimental, sobretudo em experimentos realizados com delineamento em blocos ao acaso, utilizando-se de artigos, teses e dissertações para extrair os aspectos mais relevantes que permitam alcançar o objetivo proposto sobre o assunto abordado.

2.1 Marco Histórico

Boa parte da experimentação que existe hoje foi introduzida por Sir Ronald A. Fisher (1890-1962), um estatístico que trabalhou na Estação Experimental de Agricultura de Rothamstead, na Inglaterra. E em 1925 com o propósito de eliminar o efeito da heterogeneidade presentes nas unidades experimentais sobre comparação dos tratamentos, propôs que fizessem grupos de parcelas homogêneas (blocos), onde cada um recebia uma repetição de todos os tratamentos (SOUZA et al., 2002.). Este delineamento ficou conhecido como blocos completos casualizados, porque todos os tratamentos apareciam em cada bloco, ou seja, uma situação de completa ortogonalidade (OLINDA, 2012). É a origem agrícola da experimentação que explica o uso de vários termos técnicos (SOUZA et al., 2002).

A estatística experimental se aplica a diferentes áreas de estudo tais como: agronomia, medicina, engenharia, psicologia, além de outras áreas. Na agronomia, por exemplo é utilizada na elucidação de princípios biológicos e na solução de problemas agrícolas (SOUZA et al., 2002).

2.2 Planejamento Experimental

O planejamento é a etapa inicial, ou seja, é o processo de arquitetar e conduzir um experimento, incluindo a sua implantação, de forma que seja possível recolher dados que possam ser analisados, usando as metodologias estatísticas apropriadas, e que conduzam a conclusões válidas e objetivas. Como o objetivo da pesquisa é verificar, cientificamente, as suas consequências, um planejamento de experimentos é fundamental para organizar a forma como esses experimentos serão considerados, pois após o seu planejamento um método estatístico deverá ser utilizado para testar as hipóteses propostas.

Para a realização do planejamento é necessário que o pesquisador leve em conta uma série de fatores para aumentar a precisão do experimento conforme Banzatto e Kronka como: estabelecer objetivos bem definidos, ter conhecimento sobre o local de instalação do experimento, escolha dos tratamentos, tamanho e forma das parcelas, o delineamento experimental, o número de repetições, o croqui, as características a serem mensuradas e a condução de experimento.

2.2.1 Conceitos básicos da estatística experimental

Os conceitos básicos necessários para entendimento e concepção de um experimento são: (BARBOSA, 2013).

Experimento é execução do que foi planejado anteriormente com o objetivo de analisar o seu desenvolvimento sob condições controladas, que nos permite aceitar ou rejeitar as hipóteses levantadas sobre o fenômeno observado. Um experimento é composto por um conjunto de unidades experimentais, onde são aplicados os tratamentos, de maneira aleatória a fim de obtermos a estimativa do erro experimental.

Tratamento é a condição imposta à parcela cujo objetivo deseja-se medir ou comparar. Como por exemplo, podemos utilizar diferentes variedades, espaçamentos, densidades, além de outros. Os tratamentos de acordo com o interesse do experimento se dividem em dois tipos: quantitativos e qualitativos. O tratamento quantitativo trabalha com o mesmo produto, só que em diferentes quantidades, com o intuito de verificar o comportamento da característica estudada a respeito de dosagens diferentes impostas às parcelas. O tratamento qualitativo trabalha com diferentes tipos de tratamentos, sejam eles vermiculita, herbicidas, etc. Em um mesmo experimento pode-se utilizar dois ou mais tratamentos quantitativos, como exemplo: doses de fósforo, temperatura, doses de herbicida, entre outros.

Unidade experimental ou parcela é a unidade em que é feita a aplicação do tratamento testado, de modo a fornecer as informações necessárias para a análise estatística. Por exemplo, para um determinado experimento uma parcela pode ser: uma área de terreno com plantas, um lotes de sementes, uma caixa de madeira, etc.

Delineamento experimental é a forma como os tratamentos são conduzidos às unidades experimentais. Por exemplo: Delineamento em Blocos Casualizados, Delineamento Inteiramente Casualizado, Delineamento em Quadrado Latino entre outros.

Fatores de tratamento são aqueles em que o pesquisador tem o objetivo de verificar a sua influência sobre as variáveis dependentes. Os experimentos apresentam mais de um efeito de tratamento, onde se dividem em dois tipos: fatores de efeitos fixos e fatores de efeitos aleatórios. Nos de efeitos fixos o pesquisador antes de iniciar a aplicação do experimento estabelece ou escolhe o tratamento que irá comparar. Nos de efeitos aleatórios os tratamentos a ser comparados são sorteados, ou seja, os níveis são sorteados de uma população de níveis aleatoriamente.

Fatores de restrição admitem que os resultados fiquem independentes de outras fontes de variação, por exemplo, blocos.

Variáveis respostas nos experimentos os níveis e fatores são determinados e aplicados e os seus efeitos observados nas variáveis respostas.

Fonte de variação é tudo o que faz variar uma resposta.

Erro experimental é o efeito de fatores que atuam de forma aleatória e que não são passíveis de controle pelo experimentador.

2.2.2 Princípios básicos da experimentação

Os princípios básicos da experimentação são: repetição, casualização e controle local (BARBOSA, 2013).

O princípio da repetição basea-se na aplicação do mesmo tratamento a várias unidades experimentais, ou seja, consiste na reprodução do experimento básico. Este princípio tem por finalidade proporcionar a obtenção da estimativa do erro experimental. Logo, as repetições tem dois objetivos: propiciar obtenção de uma estimativa do resíduo experimental, melhora a precisão do experimento e fazer com que o teste de hipóteses seja possível.

O princípio da casualização basea-se em distribuir aleatoriamente os tratamentos às unidades experimentais. Este princípio tem por finalidade proporcionar a cada unidade experimental a mesma chance de serem designados, desta forma evita-se que algum dos tratamentos seja favorecido ou desfavorecido por fatores que o pesquisador não possa controlar. Logo as variações que contribuem para o erro experimental são transformadas em variáveis aleatórias. Para um experimento no ponto de vista estatístico, a casualização permite a validade da estimativa do erro experimental e garante o uso de testes de significância, pois nas unidades experimentais os erros atuam de maneiras independentes.

O princípio do controle local consiste em aplicar o controle quando as parcelas não estão sob condições homogêneas devido à influência de um ou mais fatores. Em outras palavras, o controle é usado quando as unidades experimentais, antes de receber os tratamentos, apresentam diferenças entre si. Desta forma, é necessário fazer o agrupamento das unidades experimentais em blocos de unidades de maneira que dentro de cada bloco haja homogeneidade.

2.2.3 Contrastes de médias dos tratamentos

O estudo de contrastes é muito importante na Estatística Experimental, principalmente quando o experimento em análise é composto por mais do que dois tratamentos. Com o uso de contrastes é possível ao pesquisador estabelecer comparações, entre tratamentos ou grupos de tratamentos, que sejam de interesse (CARNEIRO et al., 2010). Desta maneira, visando dar fundamentos para estabelecer grupos de contrastes, obter a estimativa para cada contraste estabelecido, bem como estimar a variabilidade associada a cada um destes contrastes. Os contrastes são usados para fazer comparações entre médias de tratamento, seja por combinação ou por pares de médias, obedecendo a relação direta entre o número de graus de liberdade e o número de contrastes ortogonais (CARNEIRO et al., 2010).

Contrastes de médias consiste em uma comparação entre médias de tratamentos que é chamada de contraste quando puder ser expressa por uma função linear destas médias populacionais de tratamentos, como mostra a equação 2.1:

$$\psi_h = C_1\mu_1 + C_2\mu_2 +, ..., +C_I\mu_I \tag{2.1}$$

onde ψ_h será um contraste entre médias se μ_i satisfazer a seguinte condição: $\sum_{i=1}^{I} C_I\mu_I = 0$.

2.2.4 Estimadores e Estimativas dos contrastes de médias dos tratamentos

i) **Estimador do contraste das médias**

Na prática, geralmente não se conhece os valores das médias populacionais μ_i , mas suas estimativas m_i. Daí, em Estatística Experimental, não se trabalhar com o contraste ψ_h mas com o seu estimador $\hat{\psi}_h$, que também é uma função linear de médias obtidas por meio de experimentos ou amostras. Assim tem-se que o estimador para o contraste de médias é dado pela equação 2.2:

$$\hat{\psi}_h = C_1m_1 + C_2m_2+, ..., +C_Im_I \tag{2.2}$$

ii) **Estimadores das variâncias e covariância entre dois contrastes**.

Consideremos os contrastes de médias: (OLINDA, 2012).

$\psi_1 = a_1\mu_1 + a_2\mu_2 + ... + a_I\mu_I$ e $\psi_2 = b_1\mu_1 + b_2\mu_2 + ... + b_I\mu_I$

e seus estimadores dado por:

$\hat{\psi}_1 = a_1m_1 + a_2m_2 + ... + a_Im_I$ e $\hat{\psi}_2 = b_1m_1 + b_2m_2 + ... + b_Im_I$.

As variâncias e covariância dos estimadores dos contrastes, são dadas por:

$$\text{Var}(\hat{\psi}_1) = a_1^2\frac{\sigma_1^2}{r_1} + a_2^2\frac{\sigma_2^2}{r_2} + ... + a_I^2\frac{\sigma_I^2}{r_I} = \sum a_i^2\frac{\sigma_i^2}{r_i}$$

$$\text{Var}(\hat{\psi}_2) = b_1^2\frac{\sigma_1^2}{r_1} + b_2^2\frac{\sigma_2^2}{r_2} + ... + b_I^2\frac{\sigma_I^2}{r_I} = \sum b_i^2\frac{\sigma_i^2}{r_i}$$

$$\text{Cov}(\hat{\psi}_1, \hat{\psi}_2) = a_1b_1\frac{\sigma_1^2}{r_1} + a_2b_2\frac{\sigma_2^2}{r_2} + ... + a_Ib_I\frac{\sigma_I^2}{r_I} = \sum a_ib_i\frac{\sigma_i^2}{r_i}$$

Logo, admitindo que $r_1 = r_2 = ... = r_I = J$ e $\sigma_1^2 = \sigma_2^2 = ... = \sigma_I$, assim teremos:

$$\text{Var}(\hat{\psi}_1) = (a_1^2 + a_2^2 + ... + a_I^2)\frac{\sigma^2}{J} = \frac{\sigma^2}{J}\sum a_i^2$$

$$\text{Var}(\hat{\psi}_2) = (b_1^2 + b_2^2 + ... + b_I^2)\frac{\sigma^2}{J} = \frac{\sigma^2}{J}\sum b_i^2$$

$$\text{Cov}(\hat{\psi}_1, \hat{\psi}_2) = (a_1b_1 + a_2b_2 + ... + a_Ib_I)\frac{\sigma^2}{J} = \frac{\sigma^2}{J}\sum a_ib_i$$

E os estimadores das variâncias e covariância dos estimadores dos contrastes são dadas por:

$$\widehat{Var}(\hat{\psi}_1) = (a_1^2 + a_2^2 + ... + a_I^2)\frac{s^2}{J} = \frac{s^2}{J}\sum a_i^2$$

$$\widehat{Var}(\hat{\psi}_2) = (b_1^2 + b_2^2 + ... + b_I^2)\frac{s^2}{J} = \frac{s^2}{J}\sum b_i^2$$

$$\widehat{Cov}(\hat{\psi}_1, \hat{\psi}_2) = (a_1b_1 + a_2b_2 + ... + a_Ib_I)\frac{s^2}{J} = \frac{s^2}{J}\sum a_ib_i$$

onde, para σ^2 o estimador apropriado é s^2 , a variância do erro experimental.

Uma peculiaridade, da variância do estimador de um contraste entre duas médias é dada por :

$$\text{Var}(\hat{\psi}_h) = (\frac{1}{r_i} + \frac{1}{r_i'})\sigma^2 \text{ e seu estimador é } \widehat{Var}(\hat{\psi}_h) = (\frac{1}{r_i} + \frac{1}{r_i'})s^2.$$

Portanto, se r_i=r'_i= J, como em geral ocorre nos experimentos, então o estimador da variância do estimador de um contraste entre duas médias será dada por:

$$\widehat{Var}(\hat{\psi}_h) = \frac{2s_2}{J}$$

2.2.5 Contrastes Ortogonais

Em algumas situações desejamos testar um grupo de contrastes relacionados com o experimento em estudo. Alguns tipos de testes indicados para este objetivo, necessitam que os contrastes, que compõem o grupo a ser testado, sejam ortogonais entre si. A ortogonalidade entre os contrastes indica independência linear na comparação estabelecida por um contraste com a comparação estabelecida pelos outros contrastes (CARNEIRO et al., 2010).

Os contrastes ortogonais na análise da variância são importantíssimos. Portanto, a ortogonalidade indica que a variação de um contraste é inteiramente independente da variação de outro qualquer que lhe seja ortogonal (GOMES, 1990).

Dois contrastes ψ_1 e ψ_2 são ortogonais se $\sum r_ia_ib_i = 0$ ou $\sum a_ib_i = 0$ (se o número de repetições for o mesmo). Portanto, esta é a condição de ortogonalidade entre dois contrastes para um experimento com número diferente de repetições para os tratamentos. E para um experimento com o mesmo número de repetições, as pressuposições são as mesmas (médias independentes e homogeneidade de variâncias) (BARBOSA, 2013).

Para um experimento com I tratamentos, podem ser formados vários grupos de contrastes ortogonais, no entanto cada grupo deverá conter no máximo (I-1) contrastes ortogonais, o que corresponde ao número de graus de liberdade para tratamentos. Dentro de um grupo de contrastes ortogonais, todos os contrastes tomados dois a dois, serão também ortogonais.

2.2.5.1 Métodos para obtenção de grupos de contrastes mutuamente ortogonais

i) **Obtenção por meio de sistema de equações lineares:**

Neste método, deve-se estabelecer, a princípio, um contraste que seja de interesse e, a partir deste é que os demais são obtidos. Por meio da imposição da condição de ortogonalidade e da condição para ser um contraste, obtém-se equações lineares, cujas incógnitas são os coeficientes das médias que compõem o contraste. Como o número de incógnitas é superior ao número de equações existentes, será sempre necessário atribuir valores a algumas incógnitas. É desejável que os valores a serem atribuídos, permitam que os coeficientes sejam números inteiros (CARNEIRO et al., 2010).

ii) **Obtenção por meio de regras práticas:**

De acordo com esta metodologia, é possível estabelecer facilmente um grupo de contrastes ortogonais. Desta forma, a metodologia pode ser resumida nos seguintes passos: (CARNEIRO et al., 2010).

1º Divide-se o conjunto das médias de todos os tratamentos do experimento em dois grupos. O primeiro contraste é obtido pela comparação das médias de um grupo contra as médias do outro grupo. Para isso atribui-se sinais positivos para membros de um grupo e negativos para membros do outro grupo.

2º Dentro de cada grupo formado no passo anterior, que possui mais que uma média, aplica-se o passo 1, subdividindo-os em subgrupos. Repete-se este passo até que se formem subgrupos com apenas uma média. Ao final, deveremos ter formado (I-1) comparações.

Para se obter os coeficientes que multiplicam cada média que compõem os contrastes estabelecidos, deve-se, para cada contraste: (BARBOSA, 2013)

1. Verificar o número de parcelas experimentais envolvidas no 1º grupo, digamos g_1, e o número de parcelas experimentais envolvidas no 2º grupo, digamos g_2. Calcula-se o mínimo múltiplo comum (m.m.c.) entre g_1 e g_2.

2. Dividir o m.m.c. por g_1. O resultado será o coeficiente de cada média do 1º grupo.

3. Dividir o m.m.c. por g_2. O resultado será o coeficiente de cada média do 2º grupo.

4. Multiplicar os coeficientes obtidos pelo número de repetições da respectiva média. Se possível, simplificar os coeficientes obtidos por uma constante. No caso em que o número de repetições é igual para todos os tratamentos, este passo pode ser eliminado.

2.3 Testes estatísticos e provas de significância utilizados em estatística experimental

As variáveis respostas, obtidas nos experimentos, podem ser submetidas a vários tipos de testes estatísticos ou vários procedimentos existentes para comparações múltiplas, a seguir serão apresentados os testes mais utilizados.

2.3.1 Teste F

O teste básico para a análise da variância é o teste Z de R.A. Fisher, hoje geralmente substituído pelos seus equivalentes F de George Waaddel Snedecor, v^2 de A. Hald; ou U, de F.G. Brieger, todos eles tendo em vista comparar variâncias ou os respectivos desvios padrões (GOMES, 1990). Neste teste a suposição utilizada é que s_1^2 é sempre maior do que s_2^2.

Sendo assim s_1^2 e s_2^2 as variâncias amostrais obtidas respectivamente de n_1 e n_2 observações extraídas de populações normais independentes, então:

$$F = \frac{s_1^2}{s_2^2} \sim F_{[n_1-1;n_2-1;\alpha]}$$

onde, $(n_1 - 1)$ e $(n_2 - 1)$ são os graus de liberdade da distribuição F de Snedecor.

As hipóteses contrastadas ao aplicarmos o teste F são:

$$\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 > \sigma_2^2 \end{cases}$$

De acordo com a estatística calculada e o valor tabelado tomaremos a seguinte regra de decisão:

- Se $F_{calculado} > F_{tabelado}$, rejeitamos H_0 ao nível de α de significância.
- Se $F_{calculado} < F_{tabelado}$, não rejeitamos H_0 ao nível de α de significância.

Para este caso também pode usar o valor do P-valor para comparar se rejeita ou não a hipótese confrontada e o mesmo é comparado com o α fixado.

- Se P-valor $\leq \alpha$, rejeitamos H_0 ao nível de α de significância.
- Se P-valor $> \alpha$, não rejeitamos H_0 ao nível de α de significância.

2.3.2 Teste t-Student

Outro teste clássico que pode ser usado para comparar médias de populações normais é o teste t. Como requisitos para a sua aplição são os seguintes (GOMES, 1990):

1. As comparações feitas pelo teste t devem ser escolhidas antes de serem examinados os dados;

2. Podem-se fazer no máximo tantas comparações quantos são os graus de liberdade para tratamentos, e os contrastes devem ser ortogonais.

3. Que os contrastes entre as médias dos tratamentos no experimento sejam ortogonais entre si e estabelecidos no planejamento.

Neste teste as hipóteses de interesse são:

$$\begin{cases} H_0 : \psi_h = 0 \\ H_1 : \psi_h \neq 0 \end{cases}$$

A estatística do teste é calculada pela seguinte fórmula:

$$t = \frac{|\hat{\psi}_h - 0|}{\sqrt{Var\hat{(}\psi_h)}} \sim t_{[\nu;\alpha]}$$

De acordo com a estatística calculada e o valor tabelado tomaremos a seguinte regra de decisão:

- Se $t_{calculado} > t_{tabelado}$, rejeitamos H_0 ao nível de α de significância.
- Se $t_{calculado} < t_{tabelado}$, não rejeitamos H_0 ao nível de α de significância.

Para este caso também pode usar o valor do P-valor para comparar se rejeita ou não a hipótese confrontada e o mesmo é comparado com o α fixado.

- Se P-valor $\leq \alpha$, rejeitamos H_0 ao nível de α de significância.
- Se P-valor $> \alpha$, não rejeitamos H_0 ao nível de α de significância.

2.3.3 Teste de Scheffé

O teste de Scheffé só deve ser aplicado quando o teste F tiver dado resultado significativo. Se o valor de F obtido não for significativo, nenhum contraste poderá ser significativo, e a aplicação do teste de Scheffé não se justifica. Quando, porém, o valor de F obtido é significativo, pelo menos um dos contrastes entre tratamentos será significativo

(GOMES, 1990).

Este teste permite ser aplicado para testar todo e qualquer contraste entre médias, mesmo quando sugerido pelos dados. É também considerado um teste mais conservador que o teste t, porém não exige que os contrastes a serem testados sejam ortogonais e nem que estes contrastes sejam estabelecidos antes de se examinar os dados. Sendo assim, Scheffé testa grupos de médias no contraste entre si.

Neste teste as hipóteses de interesse são:

$$\begin{cases} H_0 : \psi_h = 0 \\ H_1 : \psi_h \neq 0 \end{cases}$$

A estatística do teste é calculada pela seguinte fórmula:

$$S_{\psi_h} = \sqrt{(I-1)\widehat{Var}(\hat{\psi}_h F_{[(I-1);(n-1);\alpha]}},$$

em que, I é o número de tratamentos, $\widehat{Var}(\hat{\psi}_h)$ é a estimativa da variância do estimador do contraste h, e $F_{[(I-1);(n-1);\alpha]}$ é o valor da tabela da distribuição F, α ao nível de significância, correspondente ao número de graus de liberdade dos resíduos.

De acordo com a estatística calculada e o valor tabelado tomaremos a seguinte regra de decisão:

- Se $|\hat{\psi}_h| > S_{\psi_h}$, rejeitamos H_0 ao nível de α de significância.
- Se $|\hat{\psi}_h| < S_{\psi_h}$, não rejeitamos H_0 ao nível de α de significância.

Neste caso, podemos concluir da seguinte forma: de acordo com o teste Scheffé aplicado ao nível de ao nível de α de significância, e podemos concluir que há diferença estatisticamente entre os grupos de médias.

2.3.4 Teste de Tukey

O teste de Tukey, baseado na amplitude total estudentizada, *(studentized range)* pode ser utilizado para comparar todo e qualquer contraste entre duas médias de tratamentos. O teste é exato e de uso simples quando o número de repetições é o mesmo para todos os tratamentos, o que admitiremos de início (GOMES, 1990).

Para esse teste as hipóteses de interesse são:

$$\begin{cases} H_0 : \mu_i = \mu_i' \\ H_1 : \mu_i \neq \mu_i', \forall i \neq \ i' \end{cases}$$

A estatística do teste é calculada pela seguinte fórmula:

$$\Delta = q_{[I;\nu;\alpha]}\sqrt{\frac{1}{2}\left(\frac{1}{r_i}+\frac{1}{r_i'}\right)s^2},$$

em que, $q_{[I;\nu;\alpha]}$ é o valor da tabela da amplitude total estudentizada de Tukey correspondente ao número de tratamentos I, o ν é o número de graus de liberdade dos resíduos ,o α é o nível de significância , s^2 = QMRes e $\left(\frac{1}{r_i}+\frac{1}{r_i'}\right)s^2$ é a estimativa da variância do estimador do contraste entre duas médias $\psi = \mu_i - \mu_i'$.

De acordo com a estatística calculada e o valor tabelado tomaremos a seguinte regra de decisão:

- Se $\hat{\psi_h} = |m_i - m_i'| > \Delta$, rejeitamos H_0 ao nível de α de significância.
- Se $\hat{\psi_h} = |m_i - m_i'| < \Delta$, não rejeitamos H_0 ao nível de α de significância.

Neste caso, podemos concluir da seguinte forma: de acordo com o teste Tukey aplicado ao nível de ao nível de α de significância, e podemos concluir que os seguintes pares de médias $m_i - m_i'$ diferem estatisticamente entre si.

2.3.5 Intervalos de confiança

O conceito de intervalo de confiança *(confidence interval)* que aí fica foi introduzido por J. Neyman, e veio substituir os limites ou intervalos fiduciais de R.A. Fisher. Os dois conceitos conduzem aos mesmos resultados na maioria dos casos, mas os fundamentos lógicos são diferentes (GOMES, 1990).

Consideremos uma estimativa $\hat{y}$ de um contraste y e seja s($\hat{y}$) seu erro padrão com *n'* graus de liberdade. Se o valor de t para este número de graus de liberdade e ao nível de 5% de probabilidade t_0 , então há uma probabilidade de 95% de que tenhamos

$$\hat{y} - t_0 s(\hat{y}) < y < \hat{y} + t_0 s(\hat{y})$$

ou seja, de 95% dos casos o intervalo de confiança (variável) de extremos $\hat{y} - t_0 s(\hat{y})$ e $\hat{y} + t_0 s(\hat{y})$ conterá o verdadeiro valor do contraste. Quer dizer que se repetirmos muitas vezes o experimento, em 95% dos casos o intervalo de extremos $\hat{y} - t_0 s(\hat{y})$ e $\hat{y} + t_0 s(\hat{y})$ conterá o valor verdadeiro Y do contraste. Em outras palavras, dizendo que é de 95% a probabilidade fiducial de que o verdadeiro valor do contraste esteja dentro do intervalo de confiança determinado num certo experimento (GOMES, 1990).

No caso de termos, não um contraste, mas apenas uma média, $y = m_i$, pode ainda ser aplicada e obtemos intervalos de confiança para a média em questão pelos os seguintes extremos:

$$m_i - t_0 s(\hat{m_i}) \quad \text{e} \quad m_i + t_0 s(\hat{m_i})$$

No caso em que aplica o teste de Tukey, intervalo de confiança baseados nesses testes são obtidos pelos os seguintes extremos:

$$\hat{y} - q\frac{s}{\sqrt{r}} \quad \text{e} \quad \hat{y} + q\frac{s}{\sqrt{r}}$$

em que, $\hat{y}$ é a estimativa do contrates, $q_{[I;\nu;\alpha]}$ é o valor da tabela da amplitude total estudentizada de Tukey correspondente ao número de tratamentos I, o ν é o número de graus de liberdade dos resíduos, o α é o nível de significância, o s é desvio padrão e o r é o número de repetições.

No caso em que aplica-se do teste de Scheffé, os extremos do intervalo de confiança são:

$$\hat{y} - \sqrt{(n-1)\hat{V}(\hat{y})F_{[(I-1);(n-1);\alpha]}} \quad \text{e} \quad \hat{y} + \sqrt{(n-1)\hat{V}(\hat{y})F_{[(I-1);(n-1);\alpha]}},$$

em que, o $\hat{y}$ é o contraste em questão, n é o número de tratamentos e $F_{[(I-1);(n-1);\alpha]}$ é o valor da tabela ao nível de significância, correspondente aos números de graus de liberdade para tratamentos e para o resíduo. E $\hat{V}(\hat{y})$ é dado por: $\hat{V}(\hat{y}) = s^2\left(\frac{1}{r_i} + \frac{1}{r'_i}\right)$.

2.3.6 Número ideal de repetições

Um dos problemas mais interessantes da experimentação e a determinação prévia do número necessário de repetições (GOMES, 1990).

Na experimentação agrícola a experiência indica que dificilmente se conseguem resultados razoáveis com ensaios que tenham menos de 20 parcelas. Neste caso este número deve ser tomado como mínino. Outra forma útil e a de que devemos ter, em geral, pelo menos 10 graus de liberdade para o resíduo. Mas, estas duas limitações, embora muito úteis na prática, porém, podem ser deixadas de lado em alguns casos. Como nos experimentos de grande rigor (alguns ensaios físicos ou químicos, por exemplo) ou então quando temos um grupo numeroso de experimentos, que serão estudados em conjunto, tendo em vista unicamente resultados gerais. Neste caso, cada experimento tem individualmente pouco valor e se necessário, podemos reduzir um pouco o número de repetições, de maneira que com os recursos disponíveis, poder aumentar o número de ensaios (GOMES, 1990).

Uma solução rigorosa interessante e recente para o problema pode ser obtida pelo uso do teste de Tukey, utilizando a seguinte fórmula (GOMES, 1990):

$$r = \frac{q^2 s_2^2 F}{d^2}$$

em que, r é o número de repetições, o q é a amplitude total estudentizada para o experimento a ser feito, e o F é o valor tabelado, ao nível a escolhido de probabilidade, com numero de graus de liberdade n_1(do novo experimento) e n_2, o s_2 é uma estimativa prévia do desvio padrão, obtido de ensaios anteriores em condições análogas, o d é a diferença mínima fixar que deverá ser estatisticamente comprovada pelo ensaio.

Outra forma para se obter o número ideal de repetições, quando conhecemos o coeficiente de variação é dado por (C.V.) e a diferença mínima d, em porcentagem, a ser comprovada, é dado por: (GOMES, 1990).

$$r = \frac{q^2 (C.V.)^2 F}{d^2}$$

Outra solução prática e rápida para a determinação do número necessário de repetições também pode ser obtida pela fórmula de Tukey, com pequena modificação com efeito (GOMES, 1990), é dada pela fórmula:

$$\Delta = q \frac{s}{\sqrt{r}},$$

se dividida pela estimativa da média ($\hat{m}$) e multiplicada por 100, nos dá:

$$\frac{\Delta \times 100}{\hat{m}} = q \frac{\frac{s}{\hat{m}} \times 100}{\sqrt{r}} \ ,$$

isto é,

$$d = q \frac{C.V.}{\sqrt{r}} \ ,$$

em que, d é a diferença mínima significativa Δ em porcentagem da media $\hat{m}$. Com esta fórmula podemos calcular a diferença mínima significativa (em porcentagem) pelo teste de Tukey para um número r repetições e tendo em vista um dado coeficiente de variação.

2.4 Ensaios em Blocos ao acaso

O delineamento em blocos casualizados ou blocos ao acaso constituem o tipo de delineamento mais relevante e nele leva em consideracão, além dos pricínpios da repetição e da casualização, também o pricínpio do controle local que é representado pelos blocos, onde cada um dos quais inclui todos os tratamentos. Este delineamento deverá ser utilizado sempre que tivermos dúvida sobre a homogeneidade ambiental, ou se temos certeza de sua heterogeneidade, utilizamos este delineamento que nestas condições é muito eficiente. Cada bloco constitui uma repetição e nele os tratamentos são distribuídos inteiramente ao acaso e devemos resaltar que dentro de cada bloco, deve ser o mais homogêneo possível (CAMPOS, 1984).

Vantagens de um delineamento em blocos ao acaso:

Controla as diferenças que ocorrem nas condições experimentais, de um bloco para outro; permite, dentro de certos limites, utilizar qualquer número de tratamentos e de blocos; conduz a uma estimativa mais exata para a variância residual, uma vez que a variação ambiental entre blocos é isolada; a análise de variância é relativamente simples, sendo apenas um pouco mais demorada que a do delineamento inteiramante casualizado, visto que existe mais uma causa da variação que deve ser isolada (BANZATTO e KRONKA, 2006).

Desvantagens de um delineamento em blocos ao acaso:

Pela utilização do princípio do controle local, há uma redução no núnero de graus de liberdade do resíduo; a exigência de homogeneidades das arcelas dentro de cada bloco limita o número de tratamentos, que não pode ser muito limitado (BANZATTO e KRONKA, 2006).

2.4.1 Modelo matemático e suposições associadas ao modelo

Para o experimento realizado através de um delineamento de blocos casualizados, em que $i = 1, 2, ..., I$ e $j = 1, 2, ..., r_i$, o modelo estatístico linear apropriado para descrever as observações é dado por:

$$y_{ij} = \mu + \tau_i \ + \beta_j + \epsilon_{ij}$$

em que y_{ij} é a observação o obtida da j-ésima unidade experimental que recebeu o tratamento i; μ é a média geral; τ_i é o efeito do i-ésimo tratamento sobre a variável resposta, considerado fixo e aleatório; β_j é o efeito do j-ésimo bloco,também considerado aleatório; ϵ_{ij} é o erro aleatório associado á observação y_{ij}.

2.4.1.1 Suposições associadas ao modelo estatístico-matemático

Nas suposições associadas aos componentes do modelo temos que os erros aleatórios são variáveis aleatórias que se distribuem normalmente e identicamente, com média zero e mesma variância, ou seja, $\epsilon_{ij} \sim N(0, \sigma^2)$ (BARBOSA, 2012).

Logo, temos que:

Como ϵ_{ij} são aleatórios, μ, τ_i e β_j são fixos, verifica-se que:

i) $E(\epsilon_{ij}) = 0$ e $E(\epsilon_{ij}^2) = \sigma^2$

ii) $E(\mu) = \mu$ e $E(\mu^2) = \mu^2$

iii) $E(\tau_i) = \tau_i$ e $E(\tau_i^2) = \tau_i^2$

iv) $E(\beta_j) = \beta_j$ e $E(\beta_j^2) = \beta_j^2$

v) $E(\mu\beta_j) = \mu\beta_j$ e $E(\mu\tau_i) = \mu\tau_i$

vi) $E(\mu\epsilon_{ij}) = 0$ e $E(\tau_i\epsilon_{ij}) = 0$

vii) $E(\tau_i\beta_j) = \tau_i\beta_j$ e $E(\epsilon_{ij}\epsilon_{rs}) = 0$, para todo i $\neq$ r ou j $\neq$ s

Além disso, sendo μ a média geral, τ_i e β_j de efeitos fixos, verifica- se que:

$$\sum_{i=1}^{I} \tau_i = 0 \text{ e } \sum_{j=1}^{J} \beta_j = 0.$$

2.4.2 Hipóteses confrontadas

Nas hipóteses confrontadas neste tipo de delineamento é possível verificar duas situações: Os efeitos dos tratamentos sobre a variável resposta e o efeito dos blocos. Veremos a seguir cada caso.

1. Para verificar os efeitos dos tratamentos:

$$\begin{cases} H_0^{(\tau)} : \tau_i = 0, \forall_i \\ H_1^{(\tau)} : \tau_i \neq 0, \text{ para pelo menos um } i \neq i' \end{cases}$$

Se H_0 for verdadeira, conclui-se que os tratamentos não produzem nenhum efeito sobre a variável resposta. Caso contrário, podemos concluir que pelo menos um tratamento produz efeito sobre a variável resposta, que é equivalente a testar :

$$\begin{cases} H_0^{(\tau)} : \mu_1 = \mu_2 = ... = \mu, \\ H_1^{(\tau)} : \mu_i \neq \mu_i', \text{ para pelo menos um } i \neq i' \end{cases}$$

Se H_0 for verdadeira, conclui-se que as médias da variável resposta relativas ao tratamento são todas iguais a μ. Caso contrário, podemos concluir que existe pelo menos um par de médias que diferem estatisticamente entre si.

2. Para verificar os efeitos dos blocos:

$$\begin{cases} H_0^{(\beta)} : \beta_i = 0, \forall_i \\ H_i^{(\beta)} : \beta_i \neq \beta_i', \text{ para pelo menos um } i \neq i' \end{cases}$$

Se H_0 for verdadeira, conclui-se que os blocos da variável resposta relativas ao tratamento são todas iguais. Caso contrário, podemos concluir que existe pelo menos um par de blocos que diferem estatisticamente entre si.

2.4.3 Estimativas dos parâmetros do modelo

Os parâmetros populacionais e os erros experimentais geralmente são desconhecidos, por isso é necessário estimá-los, portanto, o modelo em termos de estimadores poderá ser escrito como:

$$y_{ij} = m + t_i + b_j + e_{ij}$$

em que m é o estimador de μ, t_i é o estimador de τ_i, b_j é o estimador de β_j e e_{ij} é o esimador de ϵ_{ij}.

Assim, os resíduos poderão ser obtidos por:

$$e_{ij} = y_{ij} - m - t_i - b_j$$

Desse modo, para encontrar os estimadores que são funções de observações experimentais, utilizou-se o método dos mínimos quadrados, onde consiste de encontrar os valores de m, t_i e b_j que minimizam a soma dos quadrados dos resíduos. Neste sentido, tem-se:

$$e_{ij} = y_{ij} - m - t_i - b_j \Rightarrow \; e_{ij}^2 = (y_{ij} - m - t_i - b_j)^2,$$

e segue que:

$$Z = \sum_{ij} e_{ij}^2 = \sum_{ij} (y_{ij} - m - t_i - b_j)^2.$$

Para obter os valores de m, t_i, e b_j que minimizam a soma dos quadrados dos resíduos, precisamos derivar a função Z em relação a cada parâmetro, igualarmos a zero e em seguida explicitamos cada estimador. Assim tem-se:

$$\begin{cases} \frac{\partial Z}{\partial m} = \frac{\partial \sum_{ij}(y_{ij}-m-t_i-b_j)^2}{\partial m} = -2\sum_{ij}(y_{ij} - m - t_i - b_j)^2 (I) \\ \frac{\partial Z}{\partial t_i} = \frac{\partial \sum_{j}(y_{ij}-m-t_i-b_j)^2}{\partial t_i} = -2\sum_{j}(y_{ij} - m - t_i - b_j)^2 (II) \\ \frac{\partial Z}{\partial b_j} = \frac{\partial \sum_{i}(y_{ij}-m-t_i-b_j)^2}{\partial b_j} = -2\sum_{i}(y_{ij} - m - t_i - b_j)^2 (III) \end{cases}$$

Agora, igualamos a zero as equações (I), (II) e (III):

De (I), segue:

$$-2\sum_{ij}(y_{ij} - m - t_i - b_j)^2 = 0 \Rightarrow$$
$$-2\sum_{ij} y_{ij} + 2\sum_{i=1}^{I} r_i m + 2\sum_{i=1}^{I} r_i t_i + 2\sum_{j=1}^{J} r_i b_j = 0 \Rightarrow$$
$$2\sum_{ij} y_{ij} = 2\sum_{i=1}^{I} r_i m + 2\sum_{i=1}^{I} r_i t_i + 2\sum_{j=1}^{J} r_i b_j = 0 \Rightarrow$$
$$\sum_{ij} y_{ij} = r_i m + \sum_{i=1}^{I} r_i t_i + \sum_{j=1}^{J} r_i b_j$$

De (II), tem-se que:

$$-2\sum_{j}(y_{ij} - m - t_i - b_j)^2 = 0 \Rightarrow$$
$$-2\sum_{j} y_{ij} + 2r_i m + 2r_i t_i + 2r_i b_j = 0 \Rightarrow$$
$$2\sum_{j} y_{ij} = 2r_i m + 2r_i t_i + 2r_i b_j = 0 \Rightarrow$$
$$\sum_{j} y_{ij} = r_i m + r_i t_i + \sum_{j=1}^{J} r_i b_j$$

De (III), verifica-se que:

$$-2\sum_{i}(y_{ij} - m - t_i - b_j)^2 = 0 \Rightarrow$$
$$-2\sum_{i} y_{ij} + 2r_i m + 2r_i t_i + 2r_i b_j = 0 \Rightarrow$$
$$2\sum_{i} y_{ij} = 2r_i m + 2r_i t_i + 2r_i b_j = 0 \Rightarrow$$

$$\sum_{i} y_{ij} = r_i m + \sum_{i=1}^{I} r_i t_i + r_i b_j$$

As equações (I), (II) e (III) formam o conjunto de equações normais que é dado por:

$$\begin{cases} \sum_{ij} y_{ij} = r_i m + \sum_{i=1}^{I} r_i t_i + \sum_{j=1}^{J} r_i b_j \\ \sum_{j} y_{ij} = r_i m + \ r_i t_i + \sum_{j=1}^{J} r_i b_j \\ \sum_{i} y_{ij} = r_i m + \sum_{i=1}^{I} r_i t_i + r_i b_j \end{cases}$$

Agora, por convenção, chamaremos $\sum_{ij} y_{ij}$ de $y_{..}$, $\sum_{j} y_{ij}$ de $y_{i.}$ e $\sum_{i} y_{ij}$ de $y_{.j}$. Assim tem-se:

$$\begin{cases} y_{..} = r_i m + \sum_{i=1}^{I} r_i t_i + \sum_{j=1}^{J} r_i b_j \\ y_{i.} = r_i m + r_i \ t_i + \sum_{j=1}^{J} r_i b_j \\ y_{.j} = r_i m + \sum_{i=1}^{I} r_i t_i + r_i b_j \end{cases}$$

O sistema de equações normais é indeterminado, uma vez que apresenta três equações e (I + J + 1) incógnitas e (I, J $>$ 1). Para resolvê-lo, precisamos supor que $\sum_{i=1}^{I} r_i t_i = 0$ e $\sum_{j=1}^{J} r_i b_j = 0$. Desta forma podemos encontrar os estimadores dos parâmetros $\mu, \tau_i \ e \beta_j$.

Ou seja,

$$m = \bar{y}_{..}, t_i = \bar{y}_{i.} - \bar{y}_{..} \ e \ \ b_j = \bar{y}_{.j} - \bar{y}_{..}.(IV)$$

Segue que os estimadores de y_{ij} e ϵ_{ij} são dados por:

$$\hat{y}_{ij} = \hat{\mu} + \hat{\tau}_i \ + \hat{\beta}_j \ \text{ e } \ \hat{\epsilon}_{ij} = e_{ij} = y_{ij} - \hat{y}_{ij}.$$

Agora substituindo os resultados em encontrados em (IV), obtém- se:

$$\hat{y}_{ij} = \bar{y}_{i.} + \bar{y}_{.j} - \bar{y}_{..} \ \text{ e } \ \epsilon_{ij} = y_{ij} - \bar{y}_{.i} - \bar{y}_{.j} + \bar{y}_{..}$$

Portanto, os resíduos e_{ij} são elementos importantes a serem utilizados para comprovar as suposições impostas aos termos do modelo: aditividade, normalidade, homocedasticidade e independência.

A Tabela 1 apresenta um resumo dos resultados encontrados dos parâmetros e das estimativas do modelo.

Tabela 1: Parâmetros e estimadores do modelo

Parâmetro do modelo	Estimador
μ	$m = \bar{y}_{..}$
τ_i	$t_i = \bar{y}_{i.} - \bar{y}_{..}$
β_j	$b_j = \bar{y}_{.j} - \bar{y}_{..}$
$E(y_{ij}) = \mu + \tau_i + \beta_j$	$\hat{y}_{ij} = \bar{y}_{i.} + \bar{y}_{.j} - \bar{y}_{..}$
ϵ_{ij}	$e_{ij} = y_{ij} - \bar{y}_{.i} - \bar{y}_{.j} + \bar{y}_{..}$
σ^2	$S^2 = QMRes$

Observação: O QM Res. é um estimador não viciado para σ^2 , que é a variância do erro experimental ϵ_{ij}.

2.4.4 Decomposição da variabilidade total

Para comparar globalmente os efeitos dos tratamentos e/ou dos blocos, utiliza-se a técnica da análise da variância que beseia-se na decomposição da variabilidade total dos dados em partes componentes. O modelo estatístico-matemático, como já sabemos, é dado por $y_{ij} = m + t_i + b_j + e_{ij}$, onde, $m = \bar{y}_{..}$, $t_i = \bar{y}_{i.} - \bar{y}_{..}$e $b_j = \bar{y}_{.j} - \bar{y}_{..}$.

Logo, reescrevendo o modelo teremos:

$$y_{ij} = \underbrace{\bar{y}_{..}}_{m} + \underbrace{(\bar{y}_{i.} - \bar{y}_{..})}_{t_i} + \underbrace{(\bar{y}_{.j} - \bar{y}_{..})}_{b_j} + \underbrace{(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})}_{e_{ij}}$$

ou ainda pode ser escrito como:

$$(y_{ij} - \bar{y}_{..}) = (\bar{y}_{i.} + \bar{y}_{..}) + (\bar{y}_{.j} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})$$

Elevando-se ao quadrado ambos os lados da equação, obteremos:

$$(y_{ij} - \bar{y}_{..})^2 = (\bar{y}_{i.} + \bar{y}_{..})^2 + (\bar{y}_{.j} - \bar{y}_{..})^2 + (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2$$

E somando-se para todas as observações, teremos:

$$\sum_I \sum_J (y_{ij} - \bar{y}_{..})^2 = J \sum_I (\bar{y}_{i.} + \bar{y}_{..})^2 + I \sum_J (\bar{y}_{.j} - \bar{y}_{..})^2 + \sum_I \sum_J (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2 \Rightarrow$$
$$\underbrace{\sum_I \sum_J (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2}_{SQResiduos} = \underbrace{\sum_I \sum_J (y_{ij} - \bar{y}_{..})^2}_{SQTotal} - \underbrace{J \sum_I (\bar{y}_{i.} + \bar{y}_{..})^2}_{SQTratamentos} - \underbrace{I \sum_J (\bar{y}_{.j} - \bar{y}_{..})^2}_{SQBlocos}.$$

Assim, desenvolvendo algebricamente, a expressão anterior resulta em:

$$SQResiduos = \underbrace{\left(\sum_{i,j} y_{ij}^2 - C\right)}_{SQTotal} - \underbrace{\left(\frac{1}{J}\sum_{I=1}^{I} y_{i.}^2 - C\right)}_{SQTratamentos} - \underbrace{\left(\frac{1}{I}\sum_{j=1}^{J} y_{.j}^2 - C\right)}_{SQBlocos}.$$

onde,

$$C = \tfrac{(y_{..})^2}{IJ} \text{ ou } C = \tfrac{(y_{..})^2}{n}.$$

Dessa maneira, observa-se que a variabilidade total é decomposta por três partes componentes: a variabilidade devida aos efeitos dos blocos, a variabilidade devida aos efeitos dos tratamentos e a variabilidade do acaso, que pode ser escrita como:

$$SQTotal = SQTratamentos + SQBlocos + SQResiduos.$$

Ou seja, a variabilidade total pode acontecer devido aos tratamentos utilizados na experimentação, mas também leva em conta a aleatoriedade do processo (BARBOSA, 2013).

Portanto, na prática a variabilidade do acaso é calculada de forma análoga por diferença como visto anteriormente. Desse modo, tem-se que:

$$SQResiduos = SQTotal - SQTratamentos - SQBlocos.$$

Os quadrados médios das partes dos componentes são obtidos pela a razão entre as somas dos quadrados e seus respectivos graus de liberdade. Desse modo, tem-se que:

$$QMTrat. = \tfrac{SQTrat.}{I-1}, QMBlocos = \tfrac{SQBlocos}{J-1} \text{ e } QMRes. = \tfrac{SQRes.}{(I-1)(J-1)}.$$

2.4.5 Análise da variância

A partir da decomposição da variabilidade total as somas de quadrados total, dos tratamentos, dos blocos e dos resíduos, segue as demonstrações das distribuições de probabilidade dos estimadores e das somas de quadrado.

1. **Distribuições de probabilidade de μ e de C**

Sabe-se que m é o estimador da média geral μ, onde:

$$m = \bar{y}_{..} = \frac{1}{n}\sum_{i,j} y_{ij}$$

Desse modo, m é uma combinação linear dos y_{ij}, os quais segue uma distribuição normal, então m segue uma distribuição normal. Encontrando as características da distribuição, temos:

$$m = \frac{1}{n}\sum_{ij} y_{ij} = \frac{1}{n}\sum_{ij} y_{ij}(\mu + \tau_i + \beta_j + \epsilon_{ij})$$
$$m = \frac{1}{n}\left(n\mu + \sum_{i} r_i\tau_i + \sum_{j} r_i\beta_j + \sum_{ij}\epsilon_{ij}\right)$$
$$m = \mu + \frac{1}{n}\sum_{i,j}\epsilon_{ij}.$$

Logo, teremos:

$$E(m) = \mu$$

e

$$Var(m) = E[m - E(m)]^2$$
$$Var(m) = E\left[\mu + \frac{1}{n}\sum_{ij}\epsilon_{ij} - \mu\right]^2$$
$$Var(m) = \frac{1}{n^2}E\left[\sum_{ij}\epsilon_{ij}\right]^2$$
$$Var(m) = \frac{1}{n^2}E(\varepsilon_{11} + ... + \varepsilon_{Ir_I})^2$$
$$Var(m) = \frac{\sigma^2}{n}.$$

Sendo assim, m segue uma distribuição normal com média μ e variância $\frac{\sigma^2}{n}$, então pode ser denotada por:

$$m = \bar{y}_{..} \sim N\left(\mu; \frac{\sigma^2}{n}\right) \text{ ou } m = \bar{y}_{..} \sim N\left(\mu; \frac{\sigma^2}{IJ}\right).$$

Segue que:

$$\frac{\bar{y}_{..} - \mu}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1) \Leftrightarrow \frac{\sqrt{n}(\bar{y}_{..}\mu)}{\sigma} \sim N(0,1).$$

E sob a hipótese $H_0 : \mu = 0$, as estatísticas $\frac{\sqrt{n}(\bar{y}_{..})}{\sigma} \sim N(0,1)$ e $\frac{n(\bar{y}^2_{..})}{\sigma^2} = \frac{\frac{(\bar{y}^2_{..})}{n}}{\sigma^2} \sim \chi^2_{(1)}$.

Portanto, sob $H_0 : \mu = 0$, a estatística $\frac{C}{\sigma^2}$, segue uma distribuição de qui-quadrado central com um grau de liberdade.

2. Distribuições de probabilidade de τ_i e da SQTratamentos.

Sabe-se que t_i é o estimador τ_i, onde:

$$t_i = \bar{y}_{i.} - \bar{y}_{..} = \frac{1}{ri}\sum_j y_{ij} - \frac{1}{n}\sum_{ij} y_{ij}$$

Logo, t_i é uma combinação linear dos y_{ij}, os quais segue uma distribuição normal e, portanto, t_i segue uma distribuição normal.

As característica da distribuição t_i serão obtidas da seguinte forma:

$$t_i = \frac{1}{r_i}\sum_j y_{ij} - \frac{1}{n}\sum_{ij} y_{ij}$$
$$t_i = \frac{1}{r_i}\sum_{j=1}^{r_i}(\mu + \tau_i + \beta_j + \epsilon_{ij}) - \frac{1}{n}\sum_{ij}(\mu + \tau_i + \beta_j + \epsilon_{ij})$$
$$t_i = \frac{1}{r_i}\left(r_i\mu + r_i\tau_i + \sum_j r_i\beta_j + \sum_j \epsilon_{ij}\right) - \frac{1}{n}\left(n\mu + \sum_i r_i\tau_i + \sum_j r_i\beta_j + \sum_{ij}\epsilon_{ij}\right)$$
$$t_i = \tau_i + \frac{1}{r_i}\sum_j \epsilon_{ij} - \frac{1}{n}\sum_{ij}\epsilon_{ij}.$$

Portanto, desta forma obtém- se que:

$$E(t_i) = \tau_i$$

e

$$Var(t_i) = E[t_i - E(t_i)]^2$$
$$Var(t_i) = E\left[\frac{1}{r_i}\sum_j \epsilon_{ij} - \frac{1}{n}\sum_{ij}\epsilon_{ij}\right]^2$$
$$Var(t_i) = E\left[\frac{1}{r_i^2}\left(\sum_j \epsilon_{ij}\right)^2 + \frac{1}{n^2}\left(\sum_{ij}\epsilon_{ij}\right)^2 - \frac{2}{nr_i}\sum_j \epsilon_{ij}\left(\sum_{ij}\epsilon_{ij}\right)\right]$$
$$Var(t_i) = \frac{(n-r_i)\sigma^2}{nr_i}.$$

Sendo assim, obtém- se:

$$t_i = \bar{y}_{i.} - \bar{y}_{..} \sim N\left(\tau_i; \frac{(n-r_i)\sigma^2}{nr_i}\right) \text{ ou } t_i = \bar{y}_{i.} - \bar{y}_{..} \sim N\left(\tau_i; \frac{(I-1)\sigma^2}{IJ}\right).$$

Onde,

$$\frac{(\bar{y}_{i.} - \bar{y}_{..} - \tau_i)}{\sqrt{\frac{(n-r_i)\sigma^2}{nr_i}}} \sim N(0,1).$$

Sob a hipótese $H_0 : \tau_i = 0$, temos:

$$\frac{(\bar{y}_{i.}-\bar{y}_{..})}{\frac{(n-r_i)\sigma^2}{nr_i}} \sim N(0,1) \Rightarrow \frac{r_i(\bar{y}_{i.}-\bar{y}_{..})^2}{\frac{(n-r_i)\sigma^2}{nr_i}} \sim \chi^2_{(1)}.$$

ou

$$\frac{r_i(\bar{y}_{i.}-\bar{y}_{..})^2}{\sigma^2} \sim \frac{(n-r_i)}{n}\chi^2_{(1)}.$$

Como $r_1(\bar{y}_{1.} - \bar{y}_{..}),\ r_2(\bar{y}_{2.} - \bar{y}_{..}), ...,\ r_I(\bar{y}_{I.} - \bar{y}_{..})$, são variáveis independente e estão sujeita a restrição $\sum\limits_{i=0}^{I} r_i t_i = \sum\limits_{i=0}^{I} r_i(\bar{y}_{i.} - \bar{y}_{..}) = 0$, segue:

$$\frac{r_i \sum\limits_{i=1}^{I}(\bar{y}_{i.}-\bar{y}^2_{..})}{\sigma^2} = \frac{SQTrat.}{\sigma^2} \sim \frac{(In-n)}{n}\chi^2_{(1)} = (I-1)\chi^2_{(1)} = \chi^2_{(I-1)}.$$

Portanto, sob a hipótese, $H_0 : \tau_i = 0$, a estatística $\frac{SQTrat.}{\sigma^2}$, segue uma distribuição de qui-quadrado central com $(I-1)$ graus de liberdade, isto é:

$$\frac{SQTrat.}{\sigma^2} \sim \chi_{(I-1).}$$

3. Distribuições de probabilidade de $\boldsymbol{\beta_j}$ e da SQBlocos.

Sabe-se que b_j é um estimador para β_j, em que,

$b_j = \bar{y}_{.j} - \bar{y}_{..}$ é uma combinação linear de normais, logo b_j é normal.

$$b_j = \frac{1}{r_i}\sum_i y_{ij} - \frac{1}{n}\sum_{i,j} y_{ij}$$

$$b_j = \frac{1}{r_i}\left(r_i\mu + \sum_i r_i\tau_i + r_i\beta_j + \sum_i \epsilon_{ij}\right) - \frac{1}{n}\left(n\mu + \sum_i r_i\tau_i + \sum_j r_i\beta_j + \sum_{i,j}\epsilon_{ij}\right)$$

$$b_j = \beta_j + \frac{1}{r_i}\sum_i e_{ij} - \frac{1}{n}\sum_{ij} e_{ij}.$$

Portanto, desta forma obtém- se que:

$$E(b_j) = \beta_j$$

e

$$Var(b_j) = E[b_j - E(b_j)]^2$$

$$Var(b_j) = E\left[\frac{1}{r_i}\sum_i e_{ij} - \frac{1}{n}\sum_{ij} e_{ij}\right]^2$$

$$Var(b_j) = E\left[\frac{1}{r_i^2}\left(\sum_j \epsilon_{ij}\right)^2 + \frac{1}{n^2}\left(\sum_{ij}\epsilon_{ij}\right)^2 - \frac{2}{nr_i}\sum_j \epsilon_{ij}\left(\sum_{ij}\epsilon_{ij}\right)\right]$$

$$Var(b_j) = \frac{(n-r_i)\sigma^2}{nr_i}.$$

Sendo assim, obtém- se:

$$b_j = \bar{y}_{.j} - \bar{y}_{..} \sim N\left(\beta_j; \frac{(n-r_i)\sigma^2}{nr_i}\right) \text{ ou } b_j = \bar{y}_{.j} - \bar{y}_{..} \sim N\left(\beta_j; \frac{(J-1)\sigma^2}{IJ}\right).$$

Onde,

$$\frac{(\bar{y}_{.j} - \bar{y}_{..} - \beta_j)}{\sqrt{\frac{(n-r_i)\sigma^2}{nr_i}}} \sim N(0,1).$$

Sob a hipótese $H_0 : \beta_j = 0$, temos:

$$\frac{(\bar{y}_{.j} - \bar{y}_{..})}{\frac{(n-r_i)\sigma^2}{nr_i}} \sim N(0,1) \Rightarrow \frac{r_i(\bar{y}_{.j} - \bar{y}_{..})^2}{\frac{(n-r_i)\sigma^2}{nr_i}} \sim \chi^2_{(1)}.$$

ou

$$\frac{r_i(\bar{y}_{.j} - \bar{y}_{..})^2}{\sigma^2} \sim \frac{(n-r_i)}{n}\chi^2_{(1)}.$$

Como $r_1(\bar{y}_{.1} - \bar{y}_{..}),\ r_2(\bar{y}_{.2} - \bar{y}_{..}), ...,\ r_I(\bar{y}_{.I} - \bar{y}_{..})$, são variáveis independentes e estão sujeita a restrição $\sum\limits_{j=0}^{J} r_i b_j = \sum\limits_{j=0}^{J} r_i(\bar{y}_{.j} - \bar{y}_{..}) = 0$, segue:

$$\frac{r_i \sum\limits_{i=1}^{I} (\bar{y}_{.j} - \bar{y}^2_{..})}{\sigma^2} = \frac{SQBlocos}{\sigma^2} \sim \frac{(Jn-n)}{n}\chi^2_{(1)} = (J-1)\chi^2_{(1)} = \chi^2_{(J-1)}.$$

Portanto, sob a hipótese, $H_0 : \beta_j = 0$, a estatística $\frac{SQBlocos.}{\sigma^2}$, segue uma distribuição de qui-quadrado central com $(J-1)$ graus de liberdade, isto é:

$$\frac{SQBlocos}{\sigma^2} \sim \chi^2_{(J-1).}$$

4. Distribuições de probabilidade de ϵ_{ij} e da SQResiduos.

Sabe-se que e_{ij} é uma estimador para ϵ_{ij}, em que,

$$e_{ij} = y_{ij} - \hat{y}_{ij} \text{ ou } e_{ij} = y_{ij} - \bar{y}_{.i} - \bar{y}_{.j} + \bar{y}_{..}$$

é uma combinação linear dos y_{ij} que têm distribuição normal, logo e_{ij} segue uma distribuição normal.

Sendo assim desenvolvendo a expressão temos:

$$
\begin{gathered}
e_{ij} = y_{ij} - \bar{y}_{.i} - \bar{y}_{.j} + \bar{y}_{..} \\
e_{ij} = (\mu + \tau_i + \beta_j + \epsilon_{ij}) - \tfrac{1}{r_i}\left(r_i\mu + r_i\tau_i + \sum_J r_i\beta_j + \sum_j e_{ij}\right) \\
-\tfrac{1}{r_i}\left(r_i\mu + \sum_i r_i\tau_i + r_i\beta j + \sum_i e_{ij}\right) + \tfrac{1}{n}\left(n\mu + \sum_i r_i\tau_i + \sum_j r_i\beta_j + \sum_{ij}\epsilon_{ij}\right) \\
e_{ij} = \epsilon_{ij} - \tfrac{1}{r_i}\sum_i \epsilon_{ij} - \tfrac{1}{r_i}\sum_j \epsilon_{ij} + \tfrac{1}{n}\sum_{ij}\epsilon_{ij}.
\end{gathered}
$$

Daí, encontramos:

$$
E(e_{ij}) = 0
$$

e

$$
\begin{gathered}
Var(e_{ij}) = E[e_{ij} - E(e_{jj})]^2 \\
Var(e_{ij}) = E\left[\epsilon_{ij} - \tfrac{1}{r_i}\sum_i \epsilon_{ij} - \tfrac{1}{r_i}\sum_j \epsilon_{ij} + \tfrac{1}{n}\sum_{ij}\epsilon_{ij}\right]^2 \\
Var(e_{ij}) = \tfrac{(r_i-1)(r_i-1)}{n}\sigma^2.
\end{gathered}
$$

Assim sendo, tem-se:

$$
\begin{gathered}
e_{ij} = y_{ij} - \bar{y}_{.i} - \bar{y}_{.j} + \bar{y}_{..} \sim N\left(0; \tfrac{(r_i-1)(r_i-1)}{n}\sigma^2\right) \\
e_{ij} = y_{ij} - \bar{y}_{.i} - \bar{y}_{.j} + \bar{y}_{..} \sim N\left(0; \tfrac{(I-1)(J-1)}{IJ}\sigma^2\right).
\end{gathered}
$$

Como,

$$
e_{ij} = y_{ij} - \bar{y}_{.i} - \bar{y}_{.j} + \bar{y}_{..} \sim N\left(0; \tfrac{(I-1)(J-1)}{n}\sigma^2\right) \Rightarrow \frac{y_{ij} - \bar{y}_{.i} - \bar{y}_{.j} + \bar{y}_{..}}{\sqrt{\frac{(I-1)(J-1)}{IJ}\sigma^2}} \sim N(0,1).
$$

Segue que:

$$
\frac{(y_{ij} - \bar{y}_{.i} - \bar{y}_{.j} + \bar{y}_{..})^2}{\frac{(I-1)(J-1)}{IJ}\sigma^2} \sim \chi^2_{(1)} \Rightarrow \frac{(y_{ij} - \bar{y}_{.i} - \bar{y}_{.j} + \bar{y}_{..})^2}{\sigma^2} \sim \frac{(I-1)(J-1)}{IJ}\chi^2_{(1)}.
$$

Somando-se para todos os resíduos, ou dada a independência entre os resíduos, segue que:

$$
\frac{\sum_{ij}(y_{ij} - \bar{y}_{.i} - \bar{y}_{.j} + \bar{y}_{..})^2}{\sigma^2} = \frac{SQRes.}{\sigma^2} \sim (I-1)(J-1)\chi^2_{(1)} = \chi^2_{[(I-1)(J-1)]}.
$$

5. Distribuições de probabilidade da SQTotal.

Neste caso, é necessário conhecer a distribuição de probabilidade da SQTotal sob a hipótese de que não existe variabilidade entre os efeitos dos tratamentos e nem dos blocos. Isto é, sob a hipótese de que as observações podem ser escritas pelo o modelo $y_{ij} = \mu + \epsilon_{ij}$, com i=1,...,I e j=1,...,J.

Sejam as dierenças entre cada observação e a média geral, temos:

$$(y_{ij} - y_{..}) = \mu + \epsilon_{ij} - \tfrac{1}{n}\left(n\mu + \sum_{ij}\epsilon_{ij}\right)$$
$$(y_{ij} - y_{..}) = \epsilon_{ij} - \tfrac{1}{n}\sum_{ij}\epsilon_{ij}.$$

Logo, obtém- se que:

$$E(y_{ij} - y_{..}) = 0$$

e

$$Var(y_{ij} - y_{..}) = E[y_{ij} - y_{..} - E(y_{ij} - y_{..})]^2$$
$$Var(y_{ij} - y_{..}) = E\left[\epsilon_{ij} - \tfrac{1}{n}\sum_{ij}\epsilon_{ij}\right]^2$$
$$Var(y_{ij} - y_{..}) = E\left[\epsilon_{ij}^2 + \tfrac{1}{n^2}\left(\sum_{ij}\epsilon_{ij}\right)^2 - \tfrac{2\epsilon_{ij}}{n}\sum_{ij}\epsilon_{ij}\right]$$
$$Var(y_{ij} - y_{..}) = \tfrac{(n-1)}{n}\sigma^2.$$

Portanto,

$$(y_{ij} - y_{..}) \sim N\left(\tau_i + \beta_j, \tfrac{(n-1)}{n}\sigma^2\right) \text{ ou } (y_{ij} - \bar{y}_{..}) \sim N\left(\tau_i + \beta_j, \tfrac{(IJ-1)}{IJ}\sigma^2\right).$$

Sob a hipótese de que não existe variabilidade entre os efeitos dos tratamentos. Isto é, $H_0^{\tau} : \tau_i = 0$ e $H_0^{\beta} : \beta_j = 0$.

$$\frac{(y_{ij} - y_{..})}{\sqrt{\frac{(IJ-1)}{IJ}\sigma^2}} \sim N(0,1) \;\; e \;\; \frac{(\, y_{ij} - \; y_{..})^2}{\frac{(IJ-1)}{IJ}\sigma^2} \sim \chi^2_{(1)}.$$

E ainda,

$$\frac{(y_{ij} - y_{..})^2}{\sigma^2} \sim \frac{(IJ-1)}{IJ}\chi^2_{(1)}.$$

Somando- se para todas as diferenças, teremos:

$$\frac{\sum\limits_{i,j}(\bar{y}_{i.}-\bar{y}_{..}^2)}{\sigma^2} = \frac{SQTotal}{\sigma^2} \sim (IJ-1)\chi^2_{(1)} = \chi^2_{(IJ-1)}.$$

Portanto, sob $H_0^{\tau}: \tau_i = 0$ e $H_0^{\beta}: \beta_j = 0$, a estatística $\frac{SQTotal}{\sigma^2}$, segue uma distribuição de qui-quadrado central com $(IJ-1)$ graus de liberdade, isto é:

$$\frac{SQTotal}{\sigma^2} \sim \chi^2_{(IJ-1).}$$

A Tabela 2 apresenta um resumo dos resultados encontrados das distribuições de probabilidade das estimativas dos parâmetros e das somas de quadrados.

Tabela 2: Distribuições de probabilidade dos parâmetros e das somas de quadrados

Distribuição do estimador	Sob a(s) hipóteses	Distribuição da SQ sob H_0
$m = \bar{y}_{..} \sim N\left(\mu; \frac{\sigma^2}{IJ}\right)$	$H_0^{(\mu)}: \mu = 0$	$\frac{C}{\sigma^2} \sim \chi^2_{(1)}$
$t_i = \bar{y}_{i.} - \bar{y}_{..} \sim N\left(\tau_i; \frac{(I-1)\sigma^2}{IJ}\right)$	$H_0^{(\tau)}: \tau_i = 0$	$\frac{SQTrat.}{\sigma^2} \sim \chi^2_{(I-1)}$
$b_j = \bar{y}_{.j} - \bar{y}_{..} \sim N\left(\beta_j; \frac{(J-1)\sigma^2}{IJ}\right)$	$H_0^{(\beta)}: \beta_j = 0$	$\frac{SQBlocos}{\sigma^2} \sim \chi^2_{(J-1)}$
$e_{ij} = y_{ij} - \bar{y_{ij}} \sim N\left(0; \frac{(I-1)(J-1)\sigma^2}{IJ}\right)$	*indep. de* $H_0^{(\tau)}$ e $H_0^{(\beta)}$	$\frac{SQRes.}{\sigma^2} \sim \chi^2_{[(I-1)(J-1)]}$
$y_{ij} - \bar{y}_{..} \sim N\left(\tau_i + \beta_j, \frac{(IJ-1)}{IJ}\sigma^2\right)$	$H_0^{(\tau)}$ e $H_0^{(\beta)}$	$\frac{SQTotal}{\sigma^2} \sim \chi^2_{(IJ-1)}.$

Os procedimentos práticos para se contrastar as H_0 e H_1 .

São estabelecer as hipóteses a serem contrastada, onde o maior interesse é testar as hipóteses de que os efeitos dos tratamentos e dos blocos sobre a variável resposta não existem, em outras palavras, testar a hipótese de que as médias da variável resposta relativa aos tratamentos são todas iguais e testar a hipótese de não existência dos efeitos dos blocos, calcular as somas dos quadrados e organizar a Tabela da Anova a fim de encontrar o valor da estatística F, conforme a Tabela 3, a seguir em que pode-se ver de forma organizada, as fontes de variação, os graus de liberdade, as somas dos quadrados, os quadrados médios e a estatística $F_{calculado}$, logo todos estes elementos são necessários para a realização da análise da variância (ANOVA).

Tabela 3: Análise de variância: ANOVA

F. Variação	G. de Liberdade	S. dos Quadrados	Q. Médio	$F_{calculado}$
$Blocos\ (H_0^{(\beta)})$	$(J-1)$	$SQBlocos$	$QMBlocos$	$\frac{QMBlocos}{QMRes.}$
$Tratamentos\ (H_0^{(\tau)})$	$(I-1)$	$SQTratamentos$	$QMTrat.$	$\frac{QMTrat.}{QMRes.}$
$Residuos$	$(I-1)(J-1)$	$SQResiduos$	$QMRes. = S^2$	$-$
$Total$	$IJ-1$	$SQTotal$	$-$	$-$

Agora é possível comparar o valor estatística $F^{\beta}_{(cal.)}$ e $F^{\tau}_{(cal.)}$ com $F^{\beta}_{(tab.)}$ e $F^{\tau}_{(tab.)}$, onde são obtido da tabela da distribuição de probabilidade F.

Assim, é possível fazer uma escolha sob H_0 com base na comparação feita, de modo que:

Se $F_{(cal.)} > F_{(\beta)} = \frac{QMBlocos}{QMRes.}$, rejeitamos $H_0^{(\beta)}$ ao nível de α de significância.

Se $F_{(cal.)} > F_{(\tau)} = \frac{QMTrat.}{QMRes.}$, rejeitamos $H_0^{(\tau)}$ ao nível de α de significância.

Logo, rejeitamos as hipóteses $H_0^{(\beta)}$ e $H_0^{(\tau)}$, ao nível de significância α e concluimos que os efeitos dos blocos e dos tratamentos sobre a variável resposta não são nulos, ou seja, os efeitos dos blocos e dos tratamentos sobre a variável resposta são estatisticamente significativos.

2.4.5.1 Componentes da variância

Para estimar os componentes da variância tem-se que a variância de qualquer observação experimental, y_{ij}, é denominada de variância total e representada por σ^2_T, sendo assim é dada por:

$$\sigma^2_T = Var(y_{ij}) = \sigma^2_\tau + \sigma^2_\beta + \sigma^2$$

O método da análise da variância é um procedimento para estimar σ^2_τ, σ^2_β e σ^2, onde utilizar as linhas da tabela da análise da variância. O procedimento incide em igualar os valores esperados dos quadrados médios a seus valores observados na análise da variância e resolver a a fórmula em relação aos componente da variância. Sendo assim para ese caso, obtém-se:

$$Var_1 = QMTrat. = \sigma^2 + J\sigma^2_\tau$$

$$Var_2 = QMBlocos = \sigma^2 + I\sigma^2_\beta$$

$$Var_3 = QMRes. = \sigma^2$$

Daí, encontramos que os estimadores dos componentes da variância são:

$$\widehat{\sigma}^2 = Var_3 = QMRes.$$

$$\widehat{\sigma}_\tau^2 = \frac{Var_1 - Var_3}{J} = \frac{QMTrat. - QMRes.}{J}$$

$$\widehat{\sigma}_\beta^2 = \frac{Var_2 - Var_3}{I} = \frac{QMBlocos. - QMRes.}{I}$$

Logo, o estimador da variabilidade total é obtida pela seguinte fórmula:

$$\widehat{\sigma}_T^2 = \widehat{\sigma}_\tau^2 + \widehat{\sigma}_\beta^2 + \widehat{\sigma}^2$$

De acordo com o estimador da variabilidade total que sugere, se são positivas todas as estimativas dos componentes da variância, logo o estimador da proporção da variância total explicada pelos tratamentos e blocos, daí tem-se que:

$$R^2 = \frac{100(\widehat{\sigma}_\tau^2 + \widehat{\sigma}_\beta^2)}{\widehat{\sigma}_T^2}$$

onde, para cada componente de variância individual temos:

$$\begin{cases} R_\tau^2 = \frac{100\widehat{\sigma}_\tau^2}{\widehat{\sigma}_T^2} \\ R_\beta^2 = \frac{100\widehat{\sigma}_\beta^2}{\widehat{\sigma}_T^2} \\ R_\epsilon^2 = \frac{100\widehat{\sigma}^2}{\widehat{\sigma}_T^2} \end{cases}$$

em que: R_τ^2 é o estimador da proporção da variação total explicada pelos tratamentos, R_β^2 é o estimador da proporção da variação total explicada pelos blocos e R_ϵ^2 é o estimador da proporção da variação total explicada pelo acaso.

Observação: Quando há uma ou mais estimativas de componente de variância negativas, seus resultados fornecidos por R_ϵ^2 não é tão simples de se interpretar. Mas, pode acontecer do método da análise de variância produzir uma estimativa negativa de um componente da variância , portanto quando isso ocorre podemos tomar algumas alternativas tais como: estima e usá-la como evidência de que o verdadeiro valor do componente da variância é zero e reestimar o componete negativo por um método onde sempre resulte em estimativas não-negativas.

2.4.6 Experimento com parcelas perdidas

Para a realização da análise deste experimento com parcela perdida é necessário estimar o valor que substituirá o que deveria ser obtido nessa parcela. Porém, esta estimativa de maneira alguma representa o valor que seria obtido, pois ninguém pode saber qual, mas é apenas um artifício de cálculo que conduz ao mesmo resultado a que se chegaria, por processos muito mais complexos, considerando apenas os dados realmente obtidos (GOMES, 1990).

Neste experimento em blocos ao acaso, o valor da variável y representará a parcela perdida que pode ser escrita da seguinte fórmula:

$$y = \frac{rB+nT-G}{(r-1)(n-1)}.$$

em que r é o número de blocos, B é o total das parcelas restantes no bloco em que representa a parcela perdida, n é o número de tratamentos, T é o total do tratamento de parcela perdida nos blocos e G é o total das parcelas disponíveis.

Após obtido este valor, ele é substituído no lugar do dado perdido e a análise é feita como anteriormente, mas com uma única diferença de que se perde um grau de liberdade para o resíduo. Em outras palavras, refaz uma nova tabela incluindo o valor estimado e a análise segue o mesmo processo, ou de modo análogo ao que já foi feito nesta seção.

2.4.7 Experimento em blocos ao acaso com tratanento comum

Na experimentação com cana-de-açúcar, não muito raro ocorre de termos um número excessivo de tratamentos a serem ensaiados. Se estruturarmos o delineamemento tradicional de blocos, em todos os tratamentos, isto acarretará um tamanho excessivo de cada bloco, comprometendo a sua homogeneidade (CAMPOS, 1984).

Dentre as alternaivas para se contornar o problema, destacamos o seguinte procedimento: (CAMPOS, 1984).

1) Os tratamentos são subdivididos em grupos;

2) Cada grupo de tratmentos constituirá um experimento em blocos ao acaso;

3) São tomados alguns tratamentos, que integrarão todos os grupo. São os chamados *tratamentos comuns*; os demais tratamentos são denominados *regulares*;

4) Procede-se à análise de variância da maneira usual, independentemente para cada

experimento;

5) Tendo como *elo de ligação* os tratamentos comuns, é feita uma análise conjuntada de todos os experimentos.

Esa é em suma a *Análise de Experimento em Blocos ao Acaso com Tratamentos Comuns*.

2.4.8 Experimento em blocos com tratanento repetidos

Nos experimentos em blocos casualizados em que o número de tratamentos é pequeno, para obtermos o número mínino de parcelas recomendado (20) e o número mínino de graus de liberdade para o resíduo (10), devemos utilizar muitos blocos. Entretanto, se a homogeneidade das parcelas de cada bloco nos permitir, podemos utilizar duas ou mais repetições dos tratamentos dentro de cada bloco, que com um número maior de parcelas. Este procedimento permitir obter um número de graus de liberdade para o resíduo, com o mesmo número de parcelas. A designação dos tratamentos às parcelas é feita de forma casual dentro de cada bloco, isto é, procedemos como se tivéssemos 6 tratamentos e efetuássemos suas casualizações dentro de cada bloco (BANZATTO e KRONKA, 2006).

3 Aplicação

Nesta seção encontram-se as principais metodologias que serviram de base para este trabalho, no que se refere a análise estatística de um experimento em blocos ao acaso.

3.1 Métodos

Os dados utilizados para a realização deste trabalho foram provenientes de um ensaio de competição de variedades de mandioca, em blocos ao acaso, realizado pelo Instituto de Pesquisas Agronômicas do Leste (atual Centro Nacional de Pesquisa de Mandioca e Fruticultura, da EMBRAPA), em Cruz das Almas, BAHIA. As produções foram as seguintes: Aipim Bravo, Milagrosa, Sutinga, Salangó Preta, Mamão e Escondida em t/ha. O intuito na elaboração e obtenção desses dados foi o de verificar quais as variedades que diferem estatisticamente, umas das outras, e qual a variedade que produz mais, ou seja, as variedades consideradas ideais para esse tipo de experimento e para tanto utilizou-se 24 unidades básicas (GOMES, 1990).

Tabela 4: Dados da produção de variedades de mandioca em (t/ha) obtidos de experimento em blocos ao acaso com quatro repetições (GOMES, 1990).

Variedades	1º Bloco	2º Bloco	3º Bloco	4º Bloco	totais ($y_{i.}$)	Médias ($\bar{y}_{i.}$)
$AipimBravo$	$14,5$	$15,8$	$24,0$	$17,0$	$71,3$	$17,83$
$Milagrosa$	$5,7$	$5,9$	$10,5$	$6,6$	$28,7$	$7,18$
$Sutinga$	$5,3$	$7,7$	$10,2$	$9,6$	$32,8$	$8,2$
$SalangoPreta$	$4,6$	$7,1$	$10,4$	$10,8$	$32,9$	$8,23$
$Mamao$	$14,8$	$12,6$	$18,8$	$16,0$	$62,2$	$15,55$
$Escondida$	$8,2$	$8,2$	$12,7$	$17,5$	$46,6$	$11,65$
$Totais$	$53,1$	$57,3$	$86,6$	$77,5$	$274,5$	$68,64$

Para este estudo existem várias possibilidades de combinações de contrastes, mas foram escolhidos apenas 4 contrastes. No 1º contraste o objetivo é testar se a média da variedade Aipim Bravo é diferente das médias das variedades Milagrosa, Sutinga, Salangó

Preta, Mamão e Escondida. No 2º contraste pode-se testar se existe diferença entre as médias das variedades Aipim Bravo e Milagrosa versos as médias das variedades Sutinga e Salango Preta. No 3º contraste será testado se existe diferença entre as médias das variedades Aipim Bravo, Milagrosa e Sutinga versos as médias das variedades Salango Preta, Mamão e Escondida. No 4º contraste é possível testar se existe diferença entre as médias das variedades Aipim Bravo e Salango Preta versos as médias das variedades Mamão e Escondida.

Os contrastes citados a cima podem ser escritos matematicamente como:

$\psi_1 = 5\mu_1 - (\mu_2 + \mu_3 + \mu_4 + \mu_5 + \mu_6)$;

$\psi_2 = \mu_1 + \mu_2 - (\mu_3 + \mu_4)$;

$\psi_3 = \mu_1 + \mu_2 + \mu_3 - (\mu_4 + \mu_5 + \mu_6)$ e

$\psi_4 = \mu_1 + \mu_4 - (\mu_5 + \mu_6)$

onde os mesmos serão utilizados para os testes t-student e Scheffé.

Sendo assim: μ_1 é a variedade Aipim Bravo, μ_2 é a variedade Milagrosa, μ_3 é a variedade Sutinga, μ_4 é a variedade Salangó Preta, μ_5 é a variedade Mamão, μ_6 é a variedade Escondida.

Como o principal objetivo é apresentar os aspectos teóricos referentes a utilização da estatística experimental, principalmente na análise estatística de um experimento em blocos ao acaso, foi desenvolvida a teoria estatística para as análises dos dados e aplicação dos testes de comparações múltiplas que foram realizados com o uso do *software* R versão 3.1.1 (R DEVELOPMENT CORE TEAM, 2012).

4 Resultados

4.1 Análise da variância

Na Tabela 5 é apresentado os resultados da análise da variância (ANOVA), verificando-se que os efeitos dos blocos e dos tratamentos sobre a variável resposta são estatisticamente significativos.

Tabela 5: Análise de variância: ANOVA

F. Variação	G. de Liberdade	S. do Quadrado	Q. Médio	$F_{calculado}$	P-valor
Blocos ($H_0^{(\beta)}$)	3	$128,5245$	$42,8415$	$10,679$	$0,0005^*$
Tratamentos ($H_0^{(\tau)}$)	5	$386,9137$	$77,3827$	$19,288$	$4,725e^{-06*}$
Residuos	15	$60,1780$	$4,0119$	–	–
Total	23	$574,6162$	–	–	–

* significativo ao nível de 5%

Conclusões: De acordo com a **Tabela 5** acima, podemos observar que o valor do $F_{cal.} = 10,679 > F^{(\beta)}_{[3,15,5\%]} = 3,29$ e P-valor= 0,0005 $< \alpha$, rejeita-se a hipótese ($H_0^{(\beta)}$) ao nível de 5% de significância e pode-se dizer que há evidência estatística significativa de que pelo menos um par de médias dos tratamentos diferem entre si. De modo análogo, como $F_{cal.} = 19,288 > F^{(\tau)}_{[5,15,5\%]} = 2,90$ e o P-valor= 4,725$e^{-06} < \alpha$, rejeita-se a hipótese ($H_0^{(\tau)}$) ao nível de 5% de significância e pode-se dizer que há evidência estatística significativa de que pelo menos um par de médias dos blocos diferem entre si. Portanto, como a análise de variância foi significativa, é possível recorrer aos testes de comparações múltiplas.

Na Tabela 6 é apresentado os resultados do teste de Tukey para os pares médias dos tratamentos da variedades de mandioca e verificando-se qual os pares de médias dos tratamentos que diferem estatisticamente entre si.

$$\begin{cases} H_0 : \mu_i = \mu_i' \\ H_1 : \mu_i \neq \mu_i', \forall i \neq \ i' \end{cases}$$

Tabela 6: Comparação de médias das variedades

Variedades	Médias
$AipimBravo$	$17,83$ a
$Mamao$	$15,55$ ab
$Escondida$	$11,65$ bc
$SalangoPreta$	$8,23$ c
$Sutinga$	$8,2$ c
$Milagrosa$	$7,18$ c

Observação: Médias seguidas da mesma letra não diferem estatisticamente pelo teste de Tukey, ao nível de 5%de significância.

Conclusão: De acordo com o teste de tukey aplicado ao nível de 5% de significância, rejeita-se a hipótese H_0 e pode-se dizer que há evidência estatística significativa de que existe diferença entre os pares de médias dos tratamentos da média da variedade Aipim Bravo com as médias das variedades Milagrosa, Sutinga, Salangó Preta e Escondida. E a média da variedade Mamão com as variedades Milagrosa, Sutinga e Salangó Preta.

Primeiramente, calculando as médias dos grupos dos contrastes e as variâncias de cada contraste para o teste t-Student e Scheffé, segue-se que:

$\hat{\psi}_1 = 5m_1 - (m_2 + m_3 + m_4 + m_5 + m_6)$
$\hat{\psi}_1 = 5 \times 17,83 - (7,18 + 8,20 + 8,23 + 15,55 + 11,65) = 38,34;$

$\hat{\psi}_2 = m_1 + m_2 - (m_3 + m_4)$
$\hat{\psi}_2 = 17,83 + 7,18 - (8,20 + 8,23) = 8,58;$

$\hat{\psi}_3 = m_1 + m_2 + m_3 - (m_4 + m_5 + m_6)$
$\hat{\psi}_3 = 17,83 + 7,18 + 8,20 - (8,23 + 15,55 + 11,65) = 6,02;$

$\hat{\psi}_4 = m_1 + m_4 - (m_5 + m_6)$
$\hat{\psi}_4 = 17,83 + 8,23 - (15,55 + 11.65) = -1,14;$

E suas respectivas variâncias são dadas por: $\text{Var}(\hat{\psi}_h) = \frac{s^2}{J} \sum a_h^2$

$\widehat{Var}(\hat{\psi}_1) = \frac{4{,}0119}{4}(5^2 + (-1)^2 + (-1)^2 + (-1)^2 + (-1)^2 + (-1)^2) = 30,09;$

$\widehat{Var}(\hat{\psi}_2) = \frac{4{,}0119}{4}(1^2 + 1^2 + (-1)^2 + (-1)^2) = 4,01;$

$\widehat{Var}(\hat{\psi}_3) = \frac{4{,}0119}{4}(1^2 + 1^2 + 1^2 + (-1)^2 + (-1)^2 + (-1)^2) = 6,02;$

$\widehat{Var}(\hat{\psi}_4) = \frac{4{,}0119}{4}(1^2 + 1^2 + (-1)^2 + (-1)^2) = 4,01.$

Agora, aplicando o teste t-Student para os grupos de médias dos tratamentos, segue-se que:

$$\begin{cases} H_0 : \psi_h = 0 \\ H_1 : \psi_h \neq 0 \end{cases}$$

Para ψ_1 temos que: $t = \frac{|\hat{\psi}_1 - 0|}{\sqrt{Va\hat{r}(\psi_1)}} = \frac{38,34}{\sqrt{30,09}} = 6,989^*$

Para ψ_2 temos que: $t = \frac{|\hat{\psi}_2 - 0|}{\sqrt{Va\hat{r}(\psi_2)}} = \frac{8,58}{\sqrt{4,01}} = 4,285^*$

Para ψ_3 temos que: $t = \frac{|\hat{\psi}_3 - 0|}{\sqrt{Va\hat{r}(\psi_3)}} = \frac{2,22}{\sqrt{6,02}} = 0,905^{NS}$

Para ψ_4 temos que: $t = \frac{|\hat{\psi}_4 - 0|}{\sqrt{Va\hat{r}(\psi_4)}} = \frac{1,17}{\sqrt{4,01}} = 0,569^{NS}$

De acordo com o teste de t-Student aplicado ao nível de 5% de significância, como $t_{cal.} > t_{[24;5\%]} = 2,064$, rejeita-se a hipótese H_0 para ψ_1 e pode-se dizer que há evidência estatística significativa de que existe diferença entre os grupos da média do primeiro tratamento com a média dos cincos últimos tratamentos do contraste e de forma análoga para ψ_2, rejeitamos a hipótese de nulidade e pode-se dizer que a média dos dois primeiros tratamentos parece ser diferente da média dos dois últimos tratamentos do contraste. Já, para ψ_3 e ψ_4 como $t_{cal.} < t_{[24;5\%]} = 2,064$, não rejeita-se a hipótese de nulidade e pode-se dizer que as médias são consideradas estatiscamente iguais. Ou seja, é possível supor que há diferença entre os grupos da média do primeiro tratamento representado pela variedade Aipim Bravo com a média dos cincos últimos tratamentos do contraste representados pelas variedades Milagrosa, Sutinga, Salangó Preta, Mamão e Escondida. E também que há diferença entre os grupos da média dos dois primeiros tratamentos representados pelas variedades Aipim Bravo e Milagrosa com a média dos dois últimos tratamentos do contraste representados pelas variedades Sutinga e Salangó Preta.

Aplicando o teste Scheffé para os grupos de médias dos tratamentos, temos que:

$$\begin{cases} H_0 : \psi_h = 0 \\ H_1 : \psi_h \neq 0 \end{cases}$$

Onde:

$$S_{\psi_h} = \sqrt{(I-1)Va\hat{r}(\hat{\psi}_h)F_{[(I-1);(n-1);\alpha]}} \text{ e } F_{[5;23;5\%]} = 2,64.$$

$S(\hat{\psi}_1) = \sqrt{(6-1) \times 30,09 \times 2,64} = 19,93^*$ e $|\hat{\psi}_1| = 38,34$

$S(\hat{\psi}_2) = \sqrt{(6-1) \times 4,01 \times 2,64} = 7,26^*$ e $|\hat{\psi}_2| = 8,58$

$S(\hat{\psi}_3) = \sqrt{(6-1) \times 6,02 \times 2,64} = 8,91^{NS}$ e $|\hat{\psi}_3| = 2,22$

$S(\hat{\psi}_4) = \sqrt{(6-1) \times 4,01 \times 2,64} = 7,26^{NS}$ e $|\hat{\psi}_4| = 1,14$

De acordo com o teste de Scheffé aplicado ao nível de 5% de significância, como $|\hat{\psi}_1|$ e $|\hat{\psi}_2| > S(\hat{\psi}_1)$ e $S(\hat{\psi}_1)$, rejeita-se a hipótese de nulidade H_0 e pode-se dizer que há evidência estatística significativa de que existe diferença entre os grupos de médias do primeiro e no segundo contraste. Ou seja, podemos admitir que é possível haver diferença entre o grupo da média do primeiro tratamento representado pela variedade Aipim Bravo com a média dos cincos últimos tratamentos do contraste representado pelas variedades Milagrosa, Sutinga, Salangó Preta, Mamão e Escondida. E que há diferença entre os grupos da média dos dois primeiros tratamentos representados pelas variedades Aipim Bravo e Milagrosa com a média dos dois últimos tratamentos do contraste representados pelas variedades Sutinga e Salangó Preta. Já $S(\hat{\psi}_3)$ e $|\hat{\psi}_4|$ não foram significativo, então pode-se dizer que as médias são consideradas estatiscamente iguais.

5 Conclusão

O experimento realizado obteve o resultado desejado para descrever as variedades em estudo e sendo assim, a utilização do experimento em blocos ao acaso foi adequada para representar as variedades consideradas ideais e que diferem estatisticamente, umas das outras.

Na análise realizada viu-se que há indícios da existência da variabilidade dos efeitos dos tratamentos e blocos sobre a variável resposta ao nível fixado de 5%, ou seja, rejeita-se as hipóteses $H_0^{(\tau)}$ e $H_0^{(\beta)}$ e pode-se dizer que há evidências significativa de que pelo menos um par de médias dos tratamentos e dos blocos diferem entre si, já que o resultado da análise de variância foi significativo, portanto foi possível proceder ao teste de comparação múltipla para os pares das médias dos tratamentos tukey, no qual apresentou o seguinte resultado que a média da variedade Aipim Bravo é diferente das médias das variedades Milagrosa, Sutinga, Salangó Preta e Escondida. No entanto a média da variedade Mamão é diferente das variedades Milagrosa, Sutinga e Salangó Preta.

Em relação aos testes de comparações múltiplas aplicados nos grupos de médias dos tratamentos dos contrastes obteve-se que no teste de t-Student que há diferença entre os grupos da média da variedade Aipim Bravo com a média das variedades Milagrosa, Sutinga, Salangó Preta, Mamão e Escondida no contraste. Entretanto, há diferença entre os grupos da média das variedades Aipim Bravo e Milagrosa com a média das variedades Sutinga e Salangó Preta. No teste de Scheffé foi obtido o mesmo resultado encontrado no teste t-Student, vindo a reforçar as diferenças encontradas no teste posterior.

6 Referências Bibliográficas

BANZATTO, D. A.; KRONKA S. N. **Experimentação Agrícola**. 4ª ed. Jaboticabal: FUNEP, São Paulo, 2006. p.73-74

BARBOSA, N. F. M. Planejamento de Experimentos I. Universidade Estadual da Paraíba (UEPB), Campina Grande, PB, 2013. Notas de Aula

CAMPOS, H. **Estatística Aplicada à Experimentação com Cana-de-Acúçar**. Fundação de Estudos Agrários Luíz de Queiroz. Piracicaba, São Paulo, 1984. 65p.

CARNEIRO, A. P. S. et al. Estatística Experimenal. Universidade Federal de Viçosa, Viçosa, MG:[s.n.], 2010.

GOMES, F. P. Curso de Estatística Experimental. 13ª ed. Piracicaba, São Paulo, Brasil:[s.n.], 1990.

OLINDA, R. A. Planejamento de Experimentos I. Universidade Estadual da Paraíba (UEPB), Campina Grande, PB, 2012. Notas de Aula

RESENDE, M. D. V. Matemática e Estatística na Análise de Experimentos e no Melhoramento Geético. Embrapa Florestas, Colombo, PR, 2007. 19p.

SOUZA, A.M. et al. Introdução a Projetos de Experimentos: Caderno Didático. Universidade Federal Santa Maria, Santa Maria, RS:[s.n.], 2002.

APÊNDICE A – Primeiro apêndice

A.1 Teste de Duncan

Duncan introduziu um novo teste para comparar pares de médias. Porém, este teste torna-se mais criterioso, ou seja sua aplicação é bem mais trabalhosa do que a do teste de Tukey, mas se chega a resultados mais detalhados e se discrimina com mais facilidade entre os tratamentos, isto é , o teste de Duncan indica resultados significativos em casos em que o teste de Tukey não permite obter significância estatística. Tal como o teste de Tukey, o de Duncan exige, para ser exato, que todos os tratamentos tenham o mesmo numero de repetições (GOMES, 1990). Esse teste leva em consideração a amplitude da quantidade das médias a serem testadas.

Neste teste as hipóteses de interesse são:

$$\begin{cases} H_0 : \mu_i = \mu_i' \\ H_1 : \mu_i \neq \mu_i', \forall i \neq \ i' \end{cases}$$

A estatística do teste é calculada pela seguinte fórmula:

$$D_k = \ Z_{[k;\nu;\alpha]}\sqrt{\tfrac{1}{2}\left(\tfrac{1}{r_i} + \tfrac{1}{r_i'}\right)s^2},$$

em que, $Z_{[k;\nu;\alpha]}$ é o valor da tabela para o número k de médias ordenadas abrangidas na contraste de interesse, o ν é o número de graus de liberdade dos resíduos, o α é o nível de significância, s^2 = QMRes e $\left(\frac{1}{r_i} + \frac{1}{r_i'}\right) s^2$ é a estimativa da variância do estimador do contraste entre duas médias.

De acordo com a estatística calculada e o valor tabelado tomaremos a seguinte regra de decisão:

- Se $\hat{\psi}_h = |m_i - m_i'| > D_k$, rejeitamos H_0 ao nível de α de significância.
- Se $\hat{\psi}_h = |m_i - m_i'| < D_k$, não rejeitamos H_0 ao nível de α de significância.

Neste caso, podemos concluir da seguinte fórmula: de acordo com o teste Duncan aplicado ao nível de ao nível de α de significância, e podemos concluir que os seguintes pares de médias $m_i - m_i'$ diferem estatisticamente entre si.

APÊNDICE B – Segundo apêndice

B.1 Aplicação do teste de Duncan

Aplicando o teste de Duncan para os pares médias dos tratamentos, segue-se que:

$$\begin{cases} H_0 : \mu_i = \mu_i' \\ H_1 : \mu_i \neq \mu_i', \forall_i \neq \; i' \end{cases}$$

Onde:

$$D_k = Z_{[k;\nu;\alpha]}\sqrt{\tfrac{1}{2}\left(\tfrac{1}{r_i} + \tfrac{1}{r_i'}\right)s^2}.$$

$\hat{\psi}_1 = |m_1 - m_2| = 10,65^*; \quad D_2 = 3,01\sqrt{\frac{4{,}0119}{4}} = 3,01$

$\hat{\psi}_2 = |m_1 - m_3| = 9,63^*; \quad D_3 = 3,16\sqrt{\frac{4{,}0119}{4}} = 3,16$

$\hat{\psi}_3 = |m_1 - m_4| = 9,60^*; \quad D_4 = 3,25\sqrt{\frac{4{,}0119}{4}} = 3,25$

$\hat{\psi}_4 = |m_1 - m_5| = 2,28^{NS}; \quad D_5 = 3,31\sqrt{\frac{4{,}0119}{4}} = 3,31$

$\hat{\psi}_5 = |m_1 - m_6| = 6,18^*; \quad D_6 = 3,36\sqrt{\frac{4{,}0119}{4}} = 3,36$

$\hat{\psi}_6 = |m_2 - m_3| = 1,02^{NS}; \quad D_2 = 3,01\sqrt{\frac{4{,}0119}{4}} = 3,01$

$\hat{\psi}_7 = |m_2 - m_4| = 1,05^{NS}; \quad D_3 = 3,16\sqrt{\frac{4{,}0119}{4}} = 3,16$

$\hat{\psi}_8 = |m_2 - m_5| = 8,37^*; \quad D_4 = 3,25\sqrt{\frac{4{,}0119}{4}} = 3,25$

$\hat{\psi}_9 = |m_2 - m_6| = 4,47^*; \quad D_5 = 3,31\sqrt{\frac{4{,}0119}{4}} = 3,31$

$\hat{\psi}_{10} = |m_3 - m_4| = 0,03^{NS}; \quad D_2 = 3,01\sqrt{\frac{4{,}0119}{4}} = 3,01$

$\hat{\psi}_{11} = |m_3 - m_5| = 7,35^*; \quad D_3 = 3,16\sqrt{\frac{4{,}0119}{4}} = 3,16$

$\hat{\psi_{12}} = |m_3 - m_6| = 3,45^*; \quad D_4 = 3,25\sqrt{\frac{4,0119}{4}} = 3,25$

$\hat{\psi_{13}} = |m_4 - m_5| = 7,32^*; \quad D_2 = 3,01\sqrt{\frac{4,0119}{4}} = 3,01$

$\hat{\psi_{14}} = |m_4 - m_6| = 3,42^*; \quad D_3 = 3,16\sqrt{\frac{4,0119}{4}} = 3,16$

$\hat{\psi_{15}} = |m_5 - m_6| = 3,90^*; \quad D_2 = 3,01\sqrt{\frac{4,0119}{4}} = 3,01$

De acordo com o teste de Duncan aplicado ao nível de 5% de significância, rejeita-se a hipótese H_0 e pode-se dizer que há evidência estatística significativa de que existe diferença entre os pares de médias dos tratamentos de m_1 com (m_2, m_3, m_4 e m_6), e de m_5 com (m_2, m_3 e m_4) e de m_6 com (m_2, m_3, m_4 e m_5) ambos respectivamente. Ou seja, a média da variedade Aipim Bravo é diferente das médias das variedades Milagrosa, Sutinga, Salangó Preta e Escondida. E a média da variedade Mamão é diferente das variedades Milagrosa, Sutinga e Salangó Preta. E a média da variedade Escondida é diferente das variedades Milagrosa, Sutinga, Salangó Preta e Mamão.