



UNIVERSIDADE ESTADUAL DA PARAÍBA - UEPB
CAMPUS I
CENTRO DE CIÊNCIAS E TECNOLOGIA
CURSO DE BACHARELADO EM ESTATÍSTICA

Fábio Danilo da Silva Cabral

**Avaliação do erro tipo I de seis testes de
comparação múltipla de médias**

Campina Grande
2017

Fábio Danilo da Silva Cabral

**Avaliação do erro tipo I de seis testes de comparação
múltipla de médias**

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.

Área de Concentração:

Orientador: Prof. Dr. João Gil de Luna

Campina Grande

2017

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.

C117a Cabral, Fábio Danilo da Silva.

Avaliação do erro tipo I de seis testes de comparação múltipla de médias [manuscrito] / Fábio Danilo da Silva Cabral. - 2017.

41 p. : il. color.

Digitado.

Trabalho de Conclusão de Curso (Graduação em Estatística) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2017.

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

1. Análise de variância. 2. Métodos de comparações múltiplas. 3. RStudio. 4. Taxa de erro. I. Título.

21. ed. CDD 519.538

Fábio Danilo da Silva Cabral

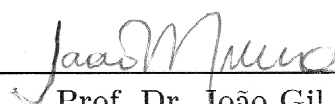
Avaliação do erro tipo I de seis testes de comparação múltipla de médias

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.

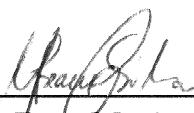
Área de Concentração:

Aprovado em: 04 de Agosto de 2017

BANCA EXAMINADORA



Prof. Dr. João Gil de Luna(Orientador)
Universidade Estadual da Paraíba



Prof.ª Dra. Maria Joseane Cruz da Silva
Universidade Federal de Campina Grande



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

*Este trabalho é dedicado às crianças adultas que,
quando pequenas, sonharam em se tornar cientistas.*

Agradecimentos

A Deus por ter me dado a oportunidade de estudar e concluir meus estudos. Aos meus pais, Manoel Bento e Maria do Socorro, que se esforçaram muito para me educar da melhor forma possível e nunca desistiram de fazer de mim uma pessoa vencedora na vida.

A minha namorada Jessiany por sempre estar me apoiando nas minhas escolhas e por estar ao meu lado nos momentos mais difíceis da minha vida acadêmica e pessoal.

Aos meus amigos e companheiros de turma, João Batista e Jonatha Felipe, com os quais passei por uma jornada inesquecível de quatro anos, de muito trabalho, muitas manhas estudando juntos, com um ajudando o outro.

Aos demais companheiros e colegas de turma, Carla Calorine, Douglas Costa, Felipe Costa, Nayara Lima e Roseane Alcântara.

Ao meu amigo e excelentíssimo presidente Rodolfo Crystian por ter-me ajudado com o Latex.

Ao professor João Gil de Luna, por ter aceitado ser o meu orientador e ter-me ajudado muito nessa tarefa difícil.

Ao professor coordenador Tiago Almeida de Oliveira e a professora Maria Joseane Cruz por terem aceitado o convite para fazer parte da banca examinadora.

A coordenação do curso, que mostrou ter uma imensa preocupação conosco, alunos de estatística, e com o curso.

A professora Giselly Oliveira por ter acreditado no meu potencial no início do curso.

A todos os professores do departamento de estatística que, direta ou indiretamente, me ajudaram a chegar onde cheguei. Ao longo desses anos, tive a oportunidade de aprender com os melhores e mais competentes professores que eu poderia ter.

*“Só duas coisas são infinitas:
o universo e a estupidez humana,
e só tenho dúvidas quanto ao universo“
(Albert Einstein)*

Resumo

Esta monografia teve como objetivo apresentar a teoria estatística dos métodos de comparações múltiplas, demonstrando a hipótese de interesse, estatística do teste, regra de decisão e conclusões para cada teste. Com o sentido de avaliar os testes t, Tukey, Bonferroni, Duncan, Scott-Knott e Student-Newman-Keuls (SNK) em relação ao erro tipo I por experimento, em cenários formados por combinações de número de tratamentos, número de repetições e variâncias, em condição de normalidade dos resíduos. Foram simulados, via RStudio, 60 cenários para avaliar a taxa do erro tipo I. Os testes de Tukey, Bonferroni, SNK e Scott-Knott controlam a taxa de erro tipo I por experimento (TPE), comparando tais resultados, pode-se dizer que o teste de Scott-Knott é conservador, embora não tanto quanto os testes de Tukey e Bonferroni. Já o teste t e teste de Duncan não controla a TPE. Há um acréscimo da TPE nos teste t e Duncan com o aumento do número de tratamentos, enquanto os testes de Tukey, SNK e Scott-Knott controlam bem essa taxa independentemente da variação dos tratamentos, repetição ou variância e o teste de Bonferroni apresentou estimativas das taxas do erro do tipo I por experimento mais baixa do que o nível de significância proposto.

Palavras-chave: Métodos de comparações múltiplas; Simulação; RStudio; Taxa do erro por experimento.

Abstract

This monograph aimed to present the statistical theory of multiple comparison methods, demonstrating the hypothesis of interest, test statistic, decision rule and conclusions for each test. With the purpose of evaluating the t-test, Tukey, Bonferroni, Duncan, Scott-Knott and Student-Newman-Keuls (SNK) in relation to the type I error by experiment, in scenarios formed by combinations of number of treatments, number of repetitions and variances, in normal condition of the residues. Sixty scenarios were simulated using RStudio to evaluate the type I error rate. The Tukey, Bonferroni, SNK and Scott-Knott tests control the type I error rate per experiment (TPE), comparing such results, it can be said that the Scott-Knott test is conservative, though not as much as the Tukey and Bonferroni. The Duncan test and t test, however, do not control TPE. There is an increase in TPE in the Duncan test with the increase in the number of treatments, while the Tukey, SNK and Scott-Knott tests control this rate regardless of the variation of the treatments, repetition or variance, and the Bonferroni test presented estimates of the rates Of type I error per experiment lower than the level of significance proposed.

Keywords: Multiple comparison methods; Simulation; RStudio; Error rate per experiment.

Lista de ilustrações

Figura 1 – Taxas de erro tipo I por experimento (TPE), em percentual, para diferentes testes de comparação múltipla de médias ao nível nominal de significância de 5%, em função do número de tratamentos e de variância igual a 5.	34
Figura 2 – Taxas de erro tipo I por experimento (TPE), em percentual, para diferentes testes de comparação múltipla de médias ao nível nominal de significância de 5%, em função do número de tratamentos e de variância igual a 10.	35
Figura 3 – Taxas de erro tipo I por experimento (TPE), em percentual, para diferentes testes de comparação múltipla de médias ao nível nominal de significância de 5%, em função do número de tratamentos e de variância igual a 15.	36

Lista de tabelas

Tabela 1 – Possibilidades de erros em decisões de testes de hipóteses.	15
Tabela 2 – Taxas de erro tipo I por experimento (TPE), em percentual, para diferentes testes de comparação múltipla de médias ao nível nominal de significância de 5%, em função do número de tratamentos e de repetições e de variância igual a 5.	34
Tabela 3 – Taxas de erro tipo I por experimento (TPE), em percentual, para diferentes testes de comparação múltipla de médias ao nível nominal de significância de 5%, em função do número de tratamentos e de repetições e de variância igual a 10.	35
Tabela 4 – Taxas de erro tipo I por experimento (TPE), em percentual, para diferentes testes de comparação múltipla de médias ao nível nominal de significância de 5%, em função do número de tratamentos e de repetições e de variância igual a 15.	36

Sumário

1	INTRODUÇÃO	12
2	FUNDAMENTAÇÃO TEÓRICA	14
2.1	Análise de Variância	14
2.1.1	Erro Tipo I e Poder do Teste	15
2.2	Métodos de Comparações Múltiplas	16
2.3	Método de comparação baseado na distribuição t-Student	17
2.3.1	Método de Bonferroni	20
2.4	Método de amplitude estudentizada	22
2.4.1	Método de Tukey	23
2.4.2	Teste de Duncan	25
2.4.3	Teste de Student-Newman-Keuls	26
2.5	Teste de Scott-Knott	28
3	MATERIAL E MÉTODOS	30
4	RESULTADO E DISCUSSÃO	32
5	CONCLUSÃO	37
	REFERÊNCIAS	38
	APÊNDICES	40
	APÊNDICE A – ROTINA	41

1 Introdução

Um dos principais objetivos da estatística é a obtenção de conclusões válidas para toda a população com base em amostras da mesma (BANZATTO; KRONKA, 2006). Existem diversas metodologias que podem ser utilizadas para analisar os dados. Para que um pesquisador realize e avalie uma pesquisa corretamente, é essencial um bom conhecimento de estatística, para compreender as potencialidades e as limitações das técnicas utilizadas. Em particular, a Estatística Experimental tem a finalidade de tirar conclusões para as características de várias populações baseadas nos dados de experimentos planejados.

A Estatística Experimental, pode ser definida como a parte da Matemática aplicada aos dados observados. Esses dados são provenientes de experimentos previamente planejados, que depois de depurados, são analisados por meio de métodos estatísticos e interpretados. Toda experimentação envolve basicamente as etapas de planejamento, instalação e condução do experimento, coleta e análise dos dados usando a análise de variância e interpretação dos resultados. O sucesso do trabalho desenvolvido está na execução criteriosa destas etapas, sendo necessário o conhecimento de alguns conceitos básicos (FINNEY, 1972).

A análise de variância é a técnica estatística que permite avaliar afirmações sobre as médias das populações. Tal metodologia visa, fundamentalmente, verificar se existe uma diferença significativa entre as médias dos tratamentos e se os fatores exercem influência em alguma variável dependente. Como com qualquer outro teste de hipóteses, existem chances de cometermos erros. Um dos possíveis erros é rejeitar a hipótese nula, quando esta é verdadeira (Erro Tipo I) ou então não rejeitar a hipótese nula, quando esta é falsa (Erro Tipo II). Os métodos de comparações múltiplas de médias, por sua vez, vêm complementar a análise de variância, pois tal metodologia detecta quais contrastes de médias diferem estatisticamente entre si.

Para abordar os erros relacionados a análise de variância, utilizamos os métodos de comparações múltiplas de médias, onde são realizadas comparações independentes por meio de um método adequado para solucionar o problema, minimizando o erro experimental. Existem vários métodos de comparações múltiplas que podem ser aplicados em diversas situações, como: Método de Diferença Mínima Significativa-DMS, Método de Bonferroni, Teste de Scheffé, Teste de Dunnett, Teste de Duncan, Teste de Student-Newman-Keuls, Teste de Tukey, Scott-Knott, entre outros.

Desta forma, este trabalho objetiva estudar os casos em que a análise de variância não detecta diferença significativa entre as médias dos tratamentos, quando algum dos

métodos de comparações múltiplas de médias estudados detecta evidências significativas ao nível de 5% de probabilidade. Para esta finalidade, foi utilizado os seguintes métodos de comparações de médias: Método de diferença mínima significativa (DMS), método de Bonferroni, método de Tukey, o método de Duncan, método de Student-Newman-Keuls (SNK) e o método de Scott-Knott. Para melhor entender sobre este assunto faremos uma aplicação dos métodos usando simulação de dados para ilustrar a teoria envolvida. Partindo desta ideia, o objetivo principal deste trabalho é apresentar a utilização teórica dos métodos de comparações múltiplas de médias e os cenários que ocorrem a rejeição da hipótese nula quando ela é verdadeira, com o sentido de avaliar a taxa do erro do tipo I por experimento.

2 Fundamentação Teórica

2.1 Análise de Variância

Fisher et al. (1951) desenvolveu uma técnica que teve grande repercussão na pesquisa experimental. Esta técnica foi denominada de análise de variância e consiste na decomposição dos graus de liberdade e da variância total de um material heterogêneo em partes atribuídas a causas conhecidas e independentes e a uma porção residual de origem desconhecida e de natureza aleatória. Em outras palavras, a técnica da análise de variância é a que nos permite fazer partições dos graus de liberdade (denotados por GL) e das somas de quadrados (denotadas por SQ), sendo que cada uma das partes nos proporciona uma estimativa de variância (denominada de quadrado médio e denotada por QM).

A análise visa, fundamentalmente, verificar a inexistência de efeitos dos níveis de um fator sobre uma variável resposta y . Por exemplo, conforme Luna (2015) no delineamento inteiramente ao acaso DIC, o modelo proposto para descrever as observações experimentais é $y_{ij} = \mu + \tau_i + \epsilon_{ij}$, onde y_{ij} é a observação obtida da unidade experimental que recebeu o i -ésimo nível do fator em estudo (tratamento) da j -ésima replicação, μ é a média geral, τ_i é o efeito do fator e ϵ_{ij} é o erro experimental. A hipótese estatística de maior interesse no modelo em estudo é a hipótese de não existência dos efeitos dos tratamentos sobre a variável resposta, isto é, $H_0 : \tau_1 = \tau_2 = \dots = \tau_I = 0$ ($H_0 : \tau_i = 0, \forall_i$), que é equivalente a testar a hipótese de igualdade das médias da variável resposta em relação aos tratamentos, ou seja, $H_0 : \mu_1 = \mu_2 = \dots = \mu_I = \mu$ onde $\mu_i = \mu + \tau_i$.

O teste F de Snedecor na análise de variância - ANOVA, tem a finalidade de comparar estimativas de variâncias dos tratamentos. Tal metodologia, pode detectar se pelo menos duas médias de tratamentos diferem entre si. O teste apenas indica se há uma diferença significativa entre, no mínimo, duas médias de tratamentos, no caso de mais de duas médias, ele não indica aonde estão estas diferenças. Uma vez que foi detectado essa diferença é importante utilizar um método de comparação múltipla que detecte quais médias diferem entre si (CANTERI; CLAUDIA, 2001).

No entanto, para que esta técnica seja empregada é necessário que algumas pressuposições sejam satisfeitas: os efeitos dos fatores que ocorrem no modelo devem ser aditivos (aditividade), os erros ϵ_{ij} devem ser independentes e identicamente distribuídos (independência), os erros ϵ_{ij} devem ter variância comum, denotada por σ^2 (homocedasticidade ou homogeneidade de variâncias), os erros ϵ_{ij} devem seguir a distribuição normal (normalidade).

2.1.1 Erro Tipo I e Poder do Teste

Iremos usar uma estatística amostral apropriada para construir o nosso teste de hipótese, e nesse contexto, essa estatística é chamada de estatística do teste. O procedimento de decisão é definido em termos da hipótese nula H_0 , as decisões possíveis são: rejeitar ou não rejeitar H_0 . Conforme resumo apresentado na tabela abaixo, existem duas possibilidades de erro:

Tabela 1 – Possibilidades de erros em decisões de testes de hipóteses.

Possibilidade	Decisão Sobre H_0	
	Rejeita H_0	Não Rejeita H_0
H_0 é verdadeira	Erro Tipo I	Decisão Correta
H_0 é falsa	Decisão correta	Erro Tipo II

A partir da estatística do teste é possível fazer uma comparação com valores pertencentes ao que chamamos de região crítica, que é o conjunto de valores cujas probabilidades de ocorrência sob H_0 são pequenas. A definição de “probabilidade pequena” se faz através da escolha do nível de significância α do teste, que é a probabilidade do erro tipo I, isto é:

$$\alpha = P(\text{erro tipo I}) = P(\text{rejeitar } H_0 \mid H_0 \text{ verdadeira}).$$

O procedimento de decisão é definido em termos da hipótese nula, em que será possível rejeitar H_0 (se a estatística do teste pertencer à região crítica), ou não (se a estatística do teste não pertencer à região crítica), a um certo nível de significância α . Em geral, o valor de α é pequeno e as escolhas mais comuns são $\alpha = 0,05$ e $\alpha = 0,01$ (FARIAS, 2008).

No procedimento de teste de hipótese, as decisões possíveis são rejeitar ou não rejeitar H_0 . Definem-se, assim, as seguintes funções em termos das probabilidades de cada uma delas: A **função característica de operação** é definida como a probabilidade de não se rejeitar H_0 , quando H_0 é falsa. Ou seja,

$$\beta(\theta) = P(\text{não rejeitar } H_0 \mid \theta).$$

Define-se a função **poder do teste** como a probabilidade do teste rejeitar H_0 quando H_0 é falsa. Ou seja, o poder de um teste é igual a

$$Q(\theta) = 1 - \beta(\theta) = P(\text{rejeitar } H_0 \mid \theta).$$

Para valores de θ fora da região crítica, essas probabilidades correspondem à probabilidade de um acerto na escolha de H_0 .

Estas funções de probabilidades estão condicionadas ao verdadeiro e desconhecido valor do parâmetro θ . Se este valor estiver no conjunto de valores definidos pela hipótese

alternativa, então $Q(\theta)$ corresponde a uma probabilidade de acerto, ou seja, ela mede a probabilidade de se rejeitar H_0 quando H_0 é falsa. Por outro lado, se a hipótese nula é $H_0 : \theta = \theta_0$, então

$$\begin{aligned} Q(\theta_0) &= 1 - \beta(\theta_0) \\ &= 1 - P(\text{n\~ao rejeitar } H_0 \mid \theta_0) \\ &= 1 - P(\text{n\~ao rejeitar } H_0 \mid H_0 \text{ verdadeiro}) \\ &= P(\text{rejeitar } H_0 \mid H_0 \text{ verdadeiro}) \\ &= \alpha. \end{aligned}$$

O poder do teste dependerá de alguns fatores, como:

- nível de significância escolhido
- a diferença entre o valor real do parâmetro e o valor considerado verdadeiro em H_0
- variabilidade da população
- tamanho da amostra coletada

Se o valor real do parâmetro for muito próximo do considerado real em H_0 , o teste terá maior dificuldade para detectar a diferença. Daí, teremos menor poder no teste e menos gravidade do erro. Por outro lado, se o valor real do parâmetro for muito distante do considerado real em H_0 , o teste terá maior poder e maior gravidade do erro (BARBOSA, 2016).

2.2 Métodos de Comparações Múltiplas

A estatística é baseada em dados obtidos de observações. Esses dados geralmente vêm de trabalhos feitos propositamente e em condições previamente especificadas. O que dificulta o trabalho do experimentador e exige a análise estatística é a presença, em todos os dados obtidos, de efeitos de fatores não controlados. Esses efeitos são indicados pela designação geral de variação do acaso ou variação aleatória. Cabe ao experimentador verificar se as diferenças observadas num experimento têm ou não têm valor, isto é, se são ou não significativas. Uma diferença não significativa é aceita como possivelmente devida ao acaso e é deixada de lado, até que novos resultados venham a confirmá-la ou negá-la (CANTERI; CLAUDIA, 2001).

Para uma correta análise dos dados se faz necessário o conhecimento da natureza dos fatores estudados. Na estatística experimental é bastante comum a comparação das médias dos tratamentos, que tem como ideia central provar se as médias de tratamentos

são iguais, em outras palavras, assume-se como verdade a hipótese inicial de que não existe diferença entre as médias dos tratamentos (SILVA, 2011)

Segundo Cardellino e Siewerdt (1992) quando um experimento é conduzido dois são os objetivos a alcançar. O primeiro consiste em obter uma estimativa do erro experimental e o segundo em testar a significância de contrastes. É importante a estimação correta do erro experimental, pois este é utilizado em qualquer dos processos posteriores de comparações de médias. Para lograr sucesso de estimação do erro experimental é fundamental a correta escolha do delineamento experimental e do modelo matemático a ser utilizado.

2.3 Método de comparação baseado na distribuição t-Student

Os testes de comparação múltiplas as médias mais usadas se baseiam na distribuição *t-Student*, o interesse está em comparar $\binom{I}{2} = \frac{I(I-1)}{2}$ pares de média para I efeitos de tratamentos, ou seja, contrastar as hipóteses

$$H_0 : \mu_i = \mu_{i'} \quad vs \quad H_1 : \mu_i \neq \mu_{i'}, \quad \forall i \neq i', \quad i < i' = 2, 3, \dots, I.$$

Sabendo que y_{ij} segue uma distribuição normal e que \bar{y}_i é uma combinação linear de y_{ij} e que toda combinação linear de variáveis aleatórias normais, também segue uma distribuição normal, assim \bar{y}_i é normal. Agora, pode-se encontrar as características da distribuição. Isto é,

$$\bar{y}_i = \frac{1}{r_i} \sum_{j=1}^{r_1} y_{ij} = \frac{1}{r_i} (r_i \mu + r_i \tau_i + \sum_{j=1}^{r_1} \epsilon_{ij}) = \mu + \tau_i + \frac{1}{r_i} \sum_{j=1}^{r_i} \epsilon_{ij} = \mu_i + \frac{1}{r_i} \sum_{j=1}^{r_1} \epsilon_{ij},$$

em que \bar{y}_i unidades experimentais submetidas ao i -ésimo tratamento em que r_i corresponde ao número de repetições.

$$E(\bar{y}_i) = \mu_i,$$

uma vez que $E(\epsilon_{ij}) = 0$.

$$\begin{aligned} Var(\bar{y}_i) &= E[\bar{y}_i - (\bar{y}_i)]^2 = E\left[\mu_i + \frac{1}{r_i} \sum_{j=1}^{r_1} \epsilon_{ij} - \mu_i\right]^2 = E\left[\frac{1}{r_i} \sum_{j=1}^{r_i} \epsilon_{ij}\right]^2 \\ &= \frac{1}{r_i^2} E[\epsilon_{i1} + \epsilon_{i2} + \dots + \epsilon_{ir_i}]^2 = \frac{1}{r_i^2} E[\epsilon_{i1}^2 + \epsilon_{i2}^2 + \dots + \epsilon_{ir_i}^2 + dp] = \frac{1}{r_i^2} r_i \sigma_i^2 \\ &= \frac{\sigma^2}{r_i} \end{aligned} \tag{2.1}$$

sendo que, dp são os duplos produtos.

Portanto, conclui-se que a média amostral relativa ao i -ésimo tratamento, segue uma distribuição normal com média μ_i e variância $\frac{\sigma^2}{r}$ independente e identicamente distribuída e será denotada por $\bar{y}_i \sim N(\mu_i; \frac{\sigma^2}{r_i})$, $i = 1, 2, \dots, I$.

Por outro lado, tem-se que

$$\begin{aligned}
 \bar{y}_i - \bar{y}_{i'} &= \mu_i + \frac{1}{r_i} \sum_{j=1}^{r_i} \epsilon_{ij} - \mu_{i'} - \frac{1}{r_{i'}} \sum_{j=1}^{r_{i'}} \epsilon_{i'j} \\
 &= \mu_i - \mu_{i'} + \frac{1}{r_i} \sum_{j=1}^{r_i} \epsilon_{ij} - \frac{1}{r_{i'}} \sum_{j=1}^{r_{i'}} \epsilon_{i'j} \\
 E(\bar{y}_i - \bar{y}_{i'}) &= \mu_i - \mu_{i'} \tag{2.2}
 \end{aligned}$$

Logo, temos que a variância é

$$\begin{aligned}
 Var(\bar{y}_i - \bar{y}_{i'}) &= E[\bar{y}_i - \bar{y}_{i'} - E(\bar{y}_i - \bar{y}_{i'})]^2 \\
 &= E[\bar{y}_i - \bar{y}_{i'} - (\mu_i - \mu_{i'})]^2 \\
 &= E\left[(\mu_i - \mu_{i'}) + \frac{1}{r_i} \sum_{j=1}^{r_i} \epsilon_{ij} - \frac{1}{r_{i'}} \sum_{j=1}^{r_{i'}} \epsilon_{i'j} - (\mu_i - \mu_{i'})\right]^2 \\
 &= \left[\frac{1}{r_i} (\epsilon_{i1} + \epsilon_{i2} + \dots + \epsilon_{ir_i} + dp) + \frac{1}{r_{i'}} (\epsilon_{i'1} + \epsilon_{i'2} + \dots + \epsilon_{i'r_{i'}} - dp) \right]^2 \\
 &= \left[\frac{1}{r_i^2} (\epsilon_{i1}^2 + \epsilon_{i2}^2 + \dots + \epsilon_{ir_i}^2 + dp) + \frac{1}{r_{i'}^2} (\epsilon_{i'1}^2 + \epsilon_{i'2}^2 + \dots + \epsilon_{i'r_{i'}}^2 - dp) \right] \\
 &= \frac{1}{r_i^2} r_i \sigma_i^2 + \frac{1}{r_{i'}^2} r_{i'} \sigma_i^2 = \left(\frac{1}{r_i} + \frac{1}{r_{i'}} \right) \sigma^2.
 \end{aligned}$$

Portanto,

$$Var(\bar{y}_i - \bar{y}_{i'}) = \left(\frac{1}{r_i} + \frac{1}{r_{i'}} \right) \sigma^2.$$

Além disso, $\bar{y}_i - \bar{y}_{i'}$ é uma combinação linear de variáveis aleatórias normais e conforme foi dito que, combinação linear de variáveis aleatórias normais é também normal, então conclui-se que

$$\bar{y}_i - \bar{y}_{i'} \sim \left(\mu_i - \mu_{i'}; \left(\frac{1}{r_i} + \frac{1}{r_{i'}} \right) \sigma^2 \right)$$

e segue que

$$Z = \frac{\bar{y}_i - \bar{y}_{i'} - (\mu_i - \mu_{i'})}{\sqrt{\left(\frac{1}{r_i} + \frac{1}{r_{i'}} \right) \sigma^2}} \sim N(0; 1)$$

Como σ^2 é desconhecido, temos de acordo com (MOOD, 1950)

$$T = \frac{\bar{y}_i - \bar{y}_{i'} - (\mu_i - \mu_{i'})}{\sqrt{\left(\frac{1}{r_i} + \frac{1}{r_{i'}} \right) s^2}} \sim t_{(\nu)} \tag{2.3}$$

sendo ν o número de graus de liberdade associado a $s^2 = \text{QMRes}$

Portanto, sob $H_0 : \mu_i = \mu_{i'}$ (ou $H_0 : \mu_i - \mu_{i'} = 0$) a estatística do teste é dada por

$$T = \frac{\bar{y}_i - \bar{y}_{i'}}{\sqrt{\left(\frac{1}{r_i} + \frac{1}{r_{i'}}\right) s^2}} \sim t_{(\nu)}. \quad (2.4)$$

Daí, para um dado nível de significância α , rejeita-se a hipótese H_0 , se

$$T = \frac{|\bar{y}_i - \bar{y}_{i'}|}{\sqrt{\left(\frac{1}{r_i} + \frac{1}{r_{i'}}\right) s^2}} > t_{(\nu; 1-\frac{\alpha}{2})}. \quad (2.5)$$

Método da Diferença Mínima Significativa - DMS

Sugerido por Fisher em 1935, este procedimento consiste em testar hipóteses de pares de médias baseado na distribuição t -Student, este método é recomendável quando a hipótese das igualdades de todas as I médias dos tratamentos, $H_0 : \mu_1 = \mu_2 = \dots = \mu_I$, for rejeitada pelo teste F da análise de variância.

A rejeição de $H_0 : \mu_i = \mu_{i'}$ implica na não rejeição da hipótese alternativa, $H_1 : \mu_i \neq \mu_{i'}$, para pelo menos um par de médias $(\mu_i, \mu_{i'})$, isto é, que existe pelo menos um par de médias que diferem estatisticamente entre si, e considerando-se a expressão 2.5 tem-se que a estatística de teste é dada por

$$T = \frac{\bar{y}_i - \bar{y}_{i'}}{\sqrt{\left(\frac{1}{r_i} + \frac{1}{r_{i'}}\right) s^2}}. \quad (2.6)$$

Essa estatística segue os pressupostos do modelo $y_{ij} = \mu + \tau_i + \epsilon_{ij}$ e segue uma distribuição de t -Student com ν graus de liberdade.

Suponhamos que nosso interesse agora é comparar os pares de médias dos I tratamentos. Nossas hipóteses de interesse vão ser,

$$H_0 : \mu_i = \mu_{i'} \quad vs \quad H_1 : \mu_i \neq \mu_{i'}, \quad \forall i \neq i'$$

a regra de decisão será: rejeitar H_0 , ao nível de significância α se,

$$|\bar{y}_i - \bar{y}_{i'}| > DMS$$

em que a quantidade DMS é dada por,

$$DMS = t_{(\nu; \frac{\alpha}{2})} \sqrt{\left(\frac{1}{r_i} + \frac{1}{r_{i'}}\right) s^2}, \quad (2.7)$$

em que

ν o número de graus de liberdade associado ao resíduo
 r_i e $r_{i'}$ o número de repetições dos tratamentos envolvidos no contraste.
 s^2 o quadrado médio do resíduo

Para um experimento em particular. No caso em que $r_1 = r_2 = \dots = r_I = r$ a equação 2.7 reduz-se a

$$DMS = t_{(\nu; \frac{\alpha}{2})} \sqrt{\frac{2s^2}{r}} \quad (2.8)$$

Segundo Leal e Porra (1998) é possível usar intervalos de confiança para as diferenças das médias dos tratamentos, baseados nos seguintes argumentos

$$Pr \left[\left| \frac{\bar{y}_i - \bar{y}_{i'} - (\mu_i - \mu_{i'})}{\sqrt{(\frac{1}{r_1} + \frac{1}{r_{i'}})s^2}} \right| \leq t_{(\nu; \frac{\alpha}{2})} \right] = 1 - \alpha,$$

$$Pr \left[t_{(\nu; \frac{\alpha}{2})} \leq \frac{\bar{y}_i - \bar{y}_{i'} - (\mu_i - \mu_{i'})}{\sqrt{(\frac{1}{r_1} + \frac{1}{r_{i'}})s^2}} \leq t_{(\nu; \frac{\alpha}{2})} \right] = 1 - \alpha,$$

$$Pr \left[(\bar{y}_i - \bar{y}_{i'}) - t_{(\nu; \frac{\alpha}{2})} \sqrt{(\frac{1}{r_1} + \frac{1}{r_{i'}})s^2} \leq (\mu_i - \mu_{i'}) \leq (\bar{y}_i - \bar{y}_{i'}) + t_{(\nu; \frac{\alpha}{2})} \sqrt{(\frac{1}{r_1} + \frac{1}{r_{i'}})s^2} \right] = 1 - \alpha.$$

Como

$$DMS = t_{(\nu; \frac{\alpha}{2})} \sqrt{(\frac{1}{r_1} + \frac{1}{r_{i'}})s^2},$$

Então,

$$Pr[(\bar{y}_i - \bar{y}_{i'}) - DMS \leq (\mu_i - \mu_{i'}) \leq (\bar{y}_i - \bar{y}_{i'}) + DMS] = 1 - \alpha.$$

Logo, temos

$$IC(\mu_i - \mu_{i'})_{100(1-\alpha)\%} = [(\bar{y}_i - \bar{y}_{i'}) - DMS; (\bar{y}_i - \bar{y}_{i'}) + DMS], \quad (2.9)$$

Neste caso se o intervalo de confiança contiver a origem é um indicador da não rejeição da hipótese nula ($H_0 : \mu_i = \mu_{i'}$), ao nível α de significância.

2.3.1 Método de Bonferroni

O método de Bonferroni é usado para neutralizar os problemas das comparações múltiplas. Este teste estatístico parte do princípio de que se deseja testar h hipóteses dependentes ou independentes em uma amostra. Para o método de Bonferroni, se fixa um

nível α de significância que é distribuído entre as possíveis comparações. As comparações formuladas baseiam-se na desigualdade de Bonferroni

$$Pr\left(\bigcup_{m=1}^M A_m\right) \leq \sum_{m=1}^M Pr(A_m). \quad (2.10)$$

Sendo A um evento qualquer associado a uma σ -álgebra.

Uma característica do teste é manter a taxa do erro "FamilyWise¹", que seria testar as hipóteses individualmente ao nível de significância $\frac{\alpha}{n}$, o objetivo é que o nível de significância para o conjuntos de hipóteses seja no máximo α . Sendo estatisticamente significativas, quer dizer, que é improvável que um determinado resultado tenha ocorrido devido ao acaso, assumindo que H_0 é realmente verdadeira (SILVA, 2011).

Consideremos que desejamos realizar estimações de intervalos para as $M = \binom{I}{2} = \frac{I(I-1)}{2}$ comparações possíveis e sempre cada uma ao nível de significância $\alpha^* = \frac{\alpha}{M}$. Isto dar origem a M intervalos de confiança que contém uma das possíveis diferenças $\Psi_i = \mu_i - \mu_{i'}$ com probabilidade de $(1-\alpha^*)$.

Seja C_m o m -ésimo intervalo que deve ser

$$Pr[\Psi_m = \mu_{1m} - \mu_{2m} \in C_m] = 1 - \alpha^* \quad m = 1, 2, \dots, M. \quad (2.11)$$

sendo que μ_{1m} é μ_{2m} , a primeira e a segunda média de comparação correspondente, onde assume-se que $1 \leq 1m < 2m < I$.

Aplicando-se a desigualdade de Bonferroni temos que

$$Pr\left(\bigcap_{m=1}^M C_m\right) = 1 - Pr\left(\bigcup_{m=1}^M \bar{C}_m\right) \geq 1 - \sum_{m=1}^M Pr(\bar{C}_m) = 1 - \sum_{m=1}^M \alpha^* \quad (2.12)$$

sendo \bar{C}_m o complementar do intervalo C_m .

Segundo Leal e Porra (1998) para garantir que o nível de significância α seja o mesmo para as M comparações em pares, o nível de confiança será pelo menos $(1-\alpha)$ para o conjunto de comparações, por pares, o nível de confiança para o conjunto de intervalos, bastar tomar

$$\alpha^* = \frac{\alpha}{M}$$

Sendo assim a probabilidade de que todos os intervalos C_m contenha a diferença entre as médias Ψ_m , será de pelo menos $(1-\alpha)$ e os intervalos serão de forma

$$\begin{aligned} \bar{y}_{1m} - \bar{y}_{2m} \pm t_{(\nu; \frac{\alpha}{2M})} \sqrt{s^2 \left(\frac{1}{r_m} + \frac{1}{r_{m'}} \right)} \\ \hat{\psi}_m \pm t_{(\nu; \frac{\alpha}{2M})} \sqrt{s^2 \left(\frac{1}{r_m} + \frac{1}{r_{m'}} \right)} \end{aligned} \quad (2.13)$$

sendo

¹ É a probabilidade de descobrir que cometeu uma ou mais vezes o erro tipo I entre todas as hipóteses.

$\widehat{\psi}_m$ a diferença entre as médias $\bar{y}_m - \bar{y}_{m'}$

r_m e $r_{m'}$ os tamanhos amostrais que correspondem as estimativas de médias $\bar{y}_m - \bar{y}_{m'}$
 s^2 o quadrado médio do erro

A hipótese de interesse é dada por

$$H_0 : \mu_i = \mu_{i'} \quad vs \quad H_1 : \mu_i \neq \mu_{i'}$$

Estatística do teste

No caso do experimento ser balanceado a DMS será a mesma para todas as comparações

$$DMS = t_{(\nu; \frac{\alpha}{2M})} \sqrt{\frac{2s^2}{r}} \quad (2.14)$$

mas quando o experimento é não balanceado a DMS é

$$DMS = t_{(\nu; \frac{\alpha}{2M})} \sqrt{\left(\frac{1}{r_m} + \frac{1}{r_{m'}}\right) s^2} \quad (2.15)$$

sendo que

r_m é o número de repetições de cada efeito do tratamento

$t_{(\nu; \frac{\alpha}{2M})}$ é o valor crítico da distribuição t-Student

ν é o número de graus de liberdade associado ao resíduo

Regra de decisão

Se $|\widehat{\psi}_m| \geq DMS \rightarrow$ Rejeita-se H_0

Se $|\widehat{\psi}_m| < DMS \rightarrow$ Não rejeita-se H_0

Conclusão

Se $|\widehat{\psi}_m| = |\bar{y}_m - \bar{y}_{m'}| \geq DMS$, rejeita-se a hipótese nula, $H_0 : \mu_m = \mu_{m'}$, ao nível α de significância, e concluímos, ao nível de significância α , que o contraste $\Psi_m = \mu_m - \mu_{m'}$ não é nulo (ou que as médias dos tratamentos m e m' diferem entre si).

2.4 Método de amplitude estudentizada

Segundo Leal e Porra (1998) os contrastes se baseiam na distribuição da amplitude estudentizada sendo sua definição dada em relação ao número de grupos que serão comparados e dos graus de liberdade do estimador da variância do erro.

Os testes baseados na distribuição da amplitude estudentizada permitem superar dificuldades que surgem ao aumentar o número de grupos a ser comparados e que não são

capazes de controlar a rejeição da hipótese nula (H_0), dado que é verdadeiramente falso. Os métodos de modo geral são conservadores, a saber, a probabilidade real de rejeitar a hipótese inicial, quando é verdadeira, é menor que o nível de significância α fixado.

Para definir a amplitude estudentizada, suponha-se que se dispõe de I observações $y_1, y_2, y_3, \dots, y_I$ independentes de uma distribuição normal com média μ e variância σ^2 , $N \sim (\mu; \sigma^2)$. Suponhamos, também, que dispõe-se do estimador da variância s^2 com ν graus de liberdade que são independentes das observações y_i 's (MOOD, 1950). Sendo R a amplitude do conjunto de observações

$$R = \text{Max}(y_i) - \text{Min}(y_i). \quad (2.16)$$

Sob esta condição se define a estatística da amplitude total estudentizada como,

$$\frac{R}{S}. \quad (2.17)$$

2.4.1 Método de Tukey

O teste de Tukey é baseado na amplitude total estudentizada. Tukey e o método usado em análise de variância para criar intervalos de confiança para todas as diferenças entre as médias dos níveis dos fatores controlando a taxa de erro global para um nível de significância α especificado.

Este procedimento foi primeiro usado para modelos univariados e equilibrados que consiste em construir intervalos de confiança (IC) para todas as possíveis combinações de pares de I medias, ao nível de confiança de $1-\alpha$.

Para construção desses intervalos foi considerado os I desvios da forma

$$(\bar{y}_1 - \mu_1), (\bar{y}_2 - \mu_2), \dots, (\bar{y}_I - \mu_I)$$

Sendo cada um desses desvios $(\bar{y}_i - \mu_i)$ variáveis aleatórias independentes e normalmente distribuídas com média 0 e variância $\frac{\sigma^2}{n}$, $N \sim (0, \frac{\sigma^2}{n})$, com o estimador da variância $\frac{\sigma^2}{n}$ sendo $\frac{s^2}{n}$ independente para cada desvio. Então da definição da amplitude estudentizada, tem-se que

$$\frac{\text{Max}(\bar{y}_i - \mu_i) - \text{Min}(\bar{y}_i - \mu_i)}{\sqrt{\frac{s^2}{n}}} \quad (2.18)$$

segue uma distribuição de probabilidade conhecida por distribuição da amplitude total estudentizada de Tukey, para I médias a serem comparadas e ν graus de liberdade dos resíduos, no qual é denotada por

$$\frac{\text{Max}(\bar{y}_i - \mu_i) - \text{Min}(\bar{y}_i - \mu_i)}{\sqrt{\frac{s^2}{n}}} \sim q_{[I;\nu]} \quad (2.19)$$

sendo,

$Max(\bar{y}_i - \mu_i)$ o maior desvio entre as médias dos tratamentos;

$Min(\bar{y}_i - \mu_i)$ o menor desvio entre as médias dos tratamentos;

ν os graus de liberdade associado ao resíduo.

Em seguida verifica-se que

$$P \left[\frac{Max(\bar{y}_i - \mu_i) - Min(\bar{y}_i - \mu_i)}{\sqrt{\frac{s^2}{n}}} \leq q_{[I;\nu]} \right] = 1 - \alpha \quad (2.20)$$

Considerando-se todas as desigualdades para os pares de médias dos tratamentos, tem-se

$$|(\bar{y}_i - \mu_i) - (\bar{y}_{i'} - \mu_{i'})| \leq Max(\bar{y}_i - \mu_i) - Min(\bar{y}_{i'} - \mu_{i'}) \quad (2.21)$$

Logo, tem-se que

$$P \left[\left| \frac{(\bar{y}_i - \mu_i) - (\bar{y}_{i'} - \mu_{i'})}{\sqrt{\frac{s^2}{n}}} \right| \leq q_{[I;\nu;\alpha]} \right] = 1 - \alpha \quad (2.22)$$

Incluindo todos os $\frac{I(I-1)}{2}$ pares de comparações entre as médias dos tratamentos, temos

$$\left| \frac{(\bar{y}_i - \mu_i) - (\bar{y}_{i'} - \mu_{i'})}{\sqrt{\frac{s^2}{n}}} \right| \leq q_{[I;\nu;\alpha]}$$

$$q_{[I;\nu;\alpha]} \leq \frac{(\bar{y}_i - \mu_i) - (\bar{y}_{i'} - \mu_{i'})}{\sqrt{\frac{s^2}{n}}} \leq q_{[I;\nu;\alpha]}$$

$$q_{[I;\nu;\alpha]} \sqrt{\frac{s^2}{n}} \leq (\bar{y}_i - \mu_i) - (\bar{y}_{i'} - \mu_{i'}) \leq q_{[I;\nu;\alpha]} \sqrt{\frac{s^2}{n}}$$

sendo,

s^2 é o quadrado médio do resíduo (QMRes)

ν é o número de graus de liberdade do erro quadrado médio

Portanto,

$$(\bar{y}_i - \bar{y}_{i'}) - DMS \leq \mu_i - \mu_{i'} \leq DMS + (\bar{y}_i - \bar{y}_{i'}) \quad (2.23)$$

onde

$$q_{[I;\nu;\alpha]} \sqrt{\frac{s^2}{n}} = DMS$$

Se o intervalo não contém a origem, então conclui-se que, é fortemente provável que as médias defiram estatisticamente entre si.

O procedimento acima é equivalente a contrastar as hipóteses

$$H_0 : \mu_i = \mu_{i'} \quad vs \quad H_1 : \mu_i \neq \mu_{i'}, \quad i \neq i' = 1, 2, \dots, I.$$

No caso do experimento não ser equilibrado, se aplica o ajuste de Tukey-Kramer, que consiste em substituir n pela média harmônica, n_h , ou seja,

$$n_h = \frac{n}{\sum_{i=1}^n \frac{1}{n_i}} = \frac{2}{\frac{1}{n_i} + \frac{1}{n_{i'}}} \quad (2.24)$$

Logo, temos que

$$\begin{aligned} DMS &= q_{[I;\nu;\alpha]} \sqrt{\frac{s^2}{n_h}} \\ &= q_{[I;\nu;\alpha]} \sqrt{\frac{s^2}{\frac{2}{\frac{1}{n_i} + \frac{1}{n_{i'}}}}} \\ DMS &= q_{[I;\nu;\alpha]} \sqrt{\frac{1}{2} \left(\frac{1}{n_i} + \frac{1}{n_{i'}} \right) s^2} \end{aligned} \quad (2.25)$$

Regra de decisão

Se $|\bar{y}_i - \bar{y}_{i'}| \leq DMS \rightarrow$ Não rejeita H_0

Se $|\bar{y}_i - \bar{y}_{i'}| > DMS \rightarrow$ Rejeita H_0

Conclusão

Se $|\bar{y}_i - \bar{y}_{i'}| > DMS$, rejeita-se a hipótese $H_0 : \mu_i = \mu_{i'}, i \neq i'$, e conclui-se, ao nível α de significância que, é muito provável que as médias defiram estatisticamente entre si.

2.4.2 Teste de Duncan

Segundo Banzatto e Kronka (2006) os resultados obtidos pelo teste de Duncan discriminam com mais facilidade entre os efeitos de tratamentos e também é menos rigoroso do que o teste de Tukey, em outras palavras ele mostra mais significância onde Tukey não mostra.

O teste de Duncan também é baseado na amplitude estudentizada, onde essa amplitude é utilizada de uma forma sequencial onde não existe um único valor para realizar as comparações. Tal como o teste de Tukey, o de Duncan exige, para ser exato, que todos os tratamentos tenham o mesmo número de repetições (GOMES, 1990)

Os contrastes testados envolvem apenas um par de média, sabendo que a amplitude do contraste pode ter um número maior de médias, em um conjunto de médias elas são ordenadas de forma decrescente e após verifica-se a maior e a menor média são ou não significativas, sendo não significativa o par de médias, todas as médias contidas neste intervalo também são não significativas. Da mesma forma sendo significativa as médias contidas no intervalo são significativas para a DMS calculada e, sendo assim, calcula-se uma nova DMS e continua a verificar se as $(I - 1)$ médias contidas no intervalo são significativas. O próxima comparação será feita com a menor média com a segunda maior, caso ainda exista significância calcula-se uma nova DMS e continua novamente o processo de comparação até realizar todas as comparações possíveis ou um contraste ser não significativo.

Hipóteses de Interesse

$$H_0 : \mu_i = \mu_{i'} \quad vs \quad H_1 : \mu_i \neq \mu_{i'}, \quad i \neq i' = 1, 2, \dots, I.$$

Estatística do teste

Partindo do principio que o número de repetições pode ser diferente entre os fatores e tratamentos, temos

$$DMS = q_{[k;\nu;\alpha]} \sqrt{\frac{1}{2} \left(\frac{1}{r_i} + \frac{1}{r_{i'}} \right) s^2}$$

Generalizando para um experimento completamente balanceado a equação reduzi-se a

$$DMS = q_{[k;\nu;\alpha]} \sqrt{\frac{s^2}{r}}$$

onde $q_{[k;\nu;\alpha]}$ é o valor da tabelado para o número k de médias ordenadas abrangidas no contraste de interesse, número de graus de liberdade dos resíduos ν , e nível de significância α , $s^2 = QMRes$ e $\left(\frac{1}{r_i} + \frac{1}{r_{i'}} \right) s^2$ é a estimativa da variância da estimativa do contraste entre duas médias $\hat{\psi}_h = \bar{y}_i - \bar{y}_{i'}$.

De acordo com a estatística calculada $\hat{\psi}_h$ e o valor da DMS tomaremos a seguinte decisão:

Se $|\hat{\psi}_h| = |\bar{y}_i - \bar{y}_{i'}| \geq DMS$, rejeitamos $H_0 : \mu_i = \mu_{i'}, \forall i \neq i'$ ao nível α de significância.

Se $|\hat{\psi}_h| = |\bar{y}_i - \bar{y}_{i'}| < DMS$, não rejeitamos $H_0 : \mu_i = \mu_{i'}, \forall i \neq i'$ ao nível α de significância.

2.4.3 Teste de Student-Newman-Keuls

Esse método desenvolvido por Newman e ampliado por Keuls é uma forma de solucionar problemas relacionados às comparações quando existe mais de dois efeitos de

tratamentos envolvidos. O teste de Student-Newman-Keuls é utilizado para comparar todo e qualquer contraste entre duas médias e caracteriza-se por utilizar a metodologia do teste de Duncan e a amplitude total estudentizada do teste de Tukey (GARCIA; LEITE, 2006).

Para a aplicação ele usa a mesma metodologia do teste de Duncan, em um conjunto de I médias, elas são colocadas em ordem decrescente e então verifica-se a diferença entre a maior e a menor média que é não significativa, todas as diferenças entre as médias do intervalo também serão não significantes.

Para cada contraste a ser testado apenas um par de médias é usado para ser comparado com a estatística do teste, porém a amplitude do contraste pode possuir um número maior de médias sobre as quais deseja-se concluir.

Hipóteses de Interesse

$$H_0 : \mu_i = \mu_{i'} \quad vs \quad H_1 : \mu_i \neq \mu_{i'}, \quad i \neq i' = 1, 2, \dots, I.$$

Estatística do teste

Partindo do princípio que o número de repetições pode ser diferente entre os fatores e tratamentos, temos

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

Generalizando para um experimento completamente balanceado a equação reduz-se a

$$DMS = q_{[I;\nu;\alpha]} \sqrt{\frac{s^2}{r}}$$

onde, DMS é a diferença mínima significativa calculada para um contraste de amplitude I , ou seja, um contraste que abrange I médias, $q_{[I;\nu;\alpha]}$ é a amplitude total estudentizada, cujo valor é tabelado em função do nível de significância α do teste, do número de médias I abrangidas pelo contraste e do número de graus de liberdade do resíduo ν .

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

Se $|\psi_h| = |\mu_i - \mu_{i'}| \geq DMS$, rejeitamos H_0 ao nível α de significância.

Se $|\psi_h| = |\mu_i - \mu_{i'}| < DMS$, não rejeitamos H_0 ao nível α de significância.

Conclusão

Se $|\psi_h| = |\mu_i - \mu_{i'}| \geq DMS$, rejeita-se a hipótese nula, ou seja, o contraste entre as médias é estatisticamente diferente de zero ao nível α de probabilidade (ou, existe diferença significativa entre as médias dos dois tratamentos).

2.5 Teste de Scott-Knott

Segundo Scott e Knott (1974), uma vez ordenada as médias, procede-se do seguinte modo, fazendo inicialmente o número de tratamentos $g = n$:

1. Determinar a partição entre dois grupos que maximiza a soma de quadrados entre grupos. Essa soma de quadrados será definida por β_0 , e será estimada da seguinte forma: Sejam T_1 e T_2 os totais dos grupos com k_1 e k_2 tratamentos em casa um.

$$\beta_0 = \frac{T_1^2}{k_1} + \frac{T_2^2}{k_2} - \frac{(T_1 + T_2)^2}{k_1 + k_2}$$

sendo

$$T_1 = \sum_{i=1}^{k_1} \bar{Y}_{(i)} \quad e \quad T_2 = \sum_{i=k_1+1}^g \bar{Y}_{(i)}$$

em que $\bar{Y}_{(i)}$ é a média do tratamento da posição i . Os dois grupos deverão ser identificados por meio da inspeção das somas de quadrados das $g - 1$ partições possíveis, sendo g o número de tratamentos envolvidos no grupo de médias considerados.

2. Determinar o valor da estatística λ da seguinte forma:

$$\lambda = \frac{\pi}{2(\pi - 2)} \times \frac{\beta_0}{\hat{\sigma}_0^2}$$

em que $\hat{\sigma}_0^2$ é o estimador de máxima verossimilhança de σ_Y^2 .

Seja $s_Y^2 = \frac{QME}{r}$ o estimador não viesado de σ_Y^2 e ν os graus de liberdade associados a este estimador.

$$\hat{\sigma}_0^2 = \frac{1}{g + \nu} \left[\sum_{i=1}^g (\bar{Y}_{(i)} - \bar{Y})^2 + \nu s_Y^2 \right]$$

3. Se $\lambda \geq \chi_{(\alpha; \frac{g}{\pi-2})}$, rejeita-se a hipótese de que os dois grupos são idênticos em favor da hipótese alternativa de que os dois grupos diferem.
4. No caso de rejeitar a essa hipótese, os dois subgrupos formados serão, independentemente, submetidos aos passos 1 a 3, fazendo, respectivamente $g = k_1$ e $g = k_2$. O processo em cada subgrupo se encerra ao não rejeitar H_0 no passo 3 ou se cada subgrupo contiver apenas uma média.

Segundo SILVA e BEARZOTI (1999) a grande vantagem de usar este método é a ausência de ambiguidade presente nos procedimentos de comparações múltiplas, já que este teste e um método aglomerativo.

Para um experimento com $n = 5$ tratamentos, com as médias ordenadas, A, B, C, D e E, existem $(2^{(n-1)} - 1) = (2^{(5-1)} - 1) = 15$ partições possíveis destas 5 médias

em dois grupos distintos. Fisher (1958) demonstrou, que é necessário considerar apenas $(n-1)$ partições obtidas pela ordenação das médias e sua divisão entre duas partições sucessivas, por essa simplificação é necessário considerar apenas $(5-1)=4$ partições das médias ordenadas, em dois grupos distintos:

Primeira partição: A versus B C D E

Segunda partição: A B versus C D E

Terceira partição: A B C versus D E

Quarta partição: A B C D versus E

Na aplicação do teste, quando muitas médias são consideradas, dificilmente a operação termina com apenas uma partição. Após encontrar a melhor separação entre dois grupos, repete-se o processo em cada subgrupo. A partir prossegue-se com a subdivisão até que os grupos resultantes sejam considerados estatisticamente iguais pelo teste χ^2 (SILVA, 2007).

3 Material e métodos

Por meio de um programa implementado no software R Core Team (2016) (APÊNDICE A), criado especialmente para esta monografia, foram simuladas 60000 experimentos. Foi utilizada esta simulação para avaliar as taxas de erro tipo I por experimento (TPE) dos testes t ou DMS, Tukey, t de Bonferroni ou Bonferroni, Duncan, Student-Newman-Keuls e Scott-Knott. Foi adotado o modelo matemático de um delineamento inteiramente casualizado: $y_{ij} = \mu + \tau_i + \epsilon_{ij}$, com $i = 1, 2, \dots, I$ e $j = 1, 2, \dots, J$, no qual, y_{ij} é o valor simulado na j-ésima repetição do i-ésimo tratamento, μ é a média geral (constante) arbitrada como 100, τ_i é o efeito do tratamento i , estipulado de tal forma que $\sum_{i=1}^I \tau_i = 0$ é o resíduo ou erro aleatório na j-ésima repetição do i-ésimo tratamento, simulado com distribuição normal com média zero e variância (σ^2).

Para simular os dados experimentais, realizar as análises de variância e os testes de comparações múltiplas de médias (TCMM) e calcular as taxas de erro tipo I por experimento, foram utilizados os pacotes Ripley et al. (2013), Silva (2015) e Mendiburu (2012) em conjunto com o aplicativo Microsoft Excel 2010 (PINTO, 2011).

Para cada cenário foram simulados 1000 experimentos. Tomou-se o cuidado de, para qualquer particular cenário, estimar as taxas de erro tipo I dos TCMM com os mesmos resultados simulados (com base na mesma semente aleatória), para assegurar que diferenças nessas estimativas não se devam ao erro aleatório do processo de simulação via o software RStudio e sim as diferenças entre os TCMM. O nível nominal de significância adotado em todos os cenários foi de 5%.

Foram simulados 60000 experimentos (60 cenários x 1000 experimentos por cenário) sem efeito de tratamento, ou seja, em situação de nulidade completa ($\tau_1 = \tau_2 = \dots = \tau_I = 0$). Os 60 cenários foram formados pela combinação entre o número de tratamentos ($i = 3, 6, 9, 12$ e 15), o número de repetições ($j = 4, 8, 12$ e 16) e as variâncias (σ^2) iguais a 5, 10 e 15.

Foram estimadas as taxas de erro tipo I por experimento (TPE) dos testes t ou DMS, Tukey, Bonferroni, Duncan, Student-Newman-Keuls (SNK) e Scott-Knott em cada cenário.

De acordo com o seguinte cenário demonstrativo ($i = 6$; $r = 4$ e $\sigma^2 = 10$). Nesse cenário foram simulados 1000 experimentos com seis tratamentos, quatro repetições e variância igual a dez, e, em cada experimento quinze estimativas de contrastes de médias, duas a duas, podem ser comparadas (μ_1 vs μ_2 , μ_1 vs μ_3 , μ_1 vs μ_4 , μ_1 vs μ_5 , μ_1 vs μ_6 , μ_2 vs μ_3 , μ_2 vs μ_4 , μ_2 vs μ_5 , μ_2 vs μ_6 , μ_3 vs μ_4 , μ_3 vs μ_5 , μ_3 vs μ_6 , μ_4 vs μ_5 , μ_4 vs μ_6 , μ_5 vs μ_6) com o valor da diferença mínima significativa de cada um dos cinco TCMM considerados. A

TPE foi estimada por meio da razão entre o número de experimentos com no mínimo uma inferência errada entre as 15 testadas e o número total de experimentos (1000), multiplicada por 100.

Para verificar se a estimativa da taxa do erro do tipo I por experimento diferiu do nível nominal de significância estabelecido ($\alpha = 5\%$), utilizou-se o limite inferior (3,225%) e o limite superior (6,775%), com base no intervalo de confiança (IC) exato de 99% para uma proporção $\hat{p} = 0,05$ expresso por,

$$IC = \hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$$

no qual $n = 1000$ experimentos e $z_{\frac{\alpha}{2}} = 2,5758$ obtido da normal padrão com média zero e desvio padrão um. Assim, taxas de erro dentro deste intervalo não foram assumidas como diferentes do valor nominal (GIRARDI; FILHO; STORCK, 2009).

4 Resultado e discussão

Nos 60 cenários formados pela combinação entre o número de tratamentos ($i = 3, 6, 9, 12$ e 15), o número de repetições ($j = 4, 8, 12$ e 16) e as variâncias (σ^2) iguais a $5, 10$ e 15 , as taxas de erro tipo I por experimento (TPE), de maneira geral, não apresentaram alterações de elevada magnitude com variabilidade do número de repetições e da variância (Tabelas 2 a 4). Menor variabilidade entre a TPE é observada com o acréscimo do número de tratamentos. Com elevado número de tratamentos o efeito das repetições sobre a estimativa da TPE pode ser considerado desprezível, pois mesmo com poucas repetições os graus de liberdade do erro experimental são altos.

As taxas de erro tipo I para o teste t por experimento (TPE) foram sempre superiores ao nível nominal de significância estabelecido ($\alpha = 5\%$) evidenciando que esse procedimento não controla a TPE, conforme destacado em Oliveira, Ferreira e Ramalho (2000). Houve aumento da TPE com o acréscimo do número de tratamentos chegando a aproximadamente 82% nos cenários com 15 tratamentos. Com 15 tratamentos sempre houve ao menos um contraste, entre pares de médias, significativo em cada experimento, independentemente do número de repetições e da precisão experimental. Em 100% dos cenários a TPE foi superior ao nível nominal de significância estabelecido ($\alpha = 5\%$) (Tabelas 2 a 4), de acordo com intervalo de confiança de 99% [$3,225\%$; $6,775\%$].

O teste de Tukey não apresentou nenhuma estimativa de TPE fora do intervalo de confiança de 99% [$3,225\%$; $6,775\%$], independentemente da variação do número de tratamentos, repetições e variância, indicando que não difere do nível nominal de significância estabelecido ($\alpha = 5\%$) (Tabelas 2 a 4). Então, observa-se que esse procedimento controla a TPE, conforme destacado em Oliveira, Ferreira e Ramalho (2000). Esses resultados concordam parcialmente, com Oliveira, Ferreira e Ramalho (2000) que afirmam que o teste de Tukey controla adequadamente a TPE, preservando o nível nominal de significância (α), no entanto, esta taxa de erro tipo I é inferior ao nível nominal estabelecido.

Entre os 60 cenários, em 47 ($78,33\%$) a estimativa da TPE do teste de Bonferroni não diferiu do nível nominal de 5% , pois a mesma esteve entre $3,225\%$ e $6,775\%$. Nos demais 13 cenários ($21,67\%$) a TPE foi inferior ao nível nominal (Tabelas 2 a 4). De certa forma, este é um resultado interessante, pois implica em afirmar que o teste de Bonferroni é “melhor” do que se propõe, uma vez que fornece taxas mais baixas de erro tipo I, pois, analisando as propriedades analíticas de um teste, espera-se que ele atenda a taxa de erro nominal fixada.

Quanto ao teste de Duncan, há falta de controle da taxas de erro tipo I, igualmente o teste t. Ocorreu aumento da TPE com o aumento do número de tratamentos,

independentemente do número de repetições e da variância. Em 100% dos cenários a TPE foi superior ao nível nominal de significância estabelecido ($\alpha = 5\%$) (Tabelas 2 a 4), de acordo com intervalo de confiança de 99% [3,225%; 6,775%].

O teste de Student-Newman-Keuls (SNK) controlou a TPE, com exceção de dois cenários que envolvem 3 e 9 tratamentos, cujas estimativas excederam o limite superior de confiança de 99% (6,775%) (Tabelas 2 e 3).

Com as constatações, pode-se dizer que o teste de Scott-Knott controlou bem a taxa de erro tipo I por experimento. No entanto, para um pequeno número de tratamentos, os níveis empíricos superaram significativamente os nominais em algumas situações.

Ao se comparar tais resultados com os obtidos por SILVA e BEARZOTI (1999), pode-se dizer que o teste de Scott-Knott é conservador, embora não tanto quanto os testes de Tukey, Bonferroni e SNK. Porém, também não atinge níveis elevados como os procedimentos de Duncan e Teste t, que atingem taxas de mais de 50% e 30% respectivamente para 9 tratamentos (Figuras 1, 2 e 3).

De maneira geral, o teste t e o teste de Duncan não controlam a taxa do erro tipo I, conforme o aumento do número de tratamentos a taxa vai se aproximando de 100%, enquanto os testes de Tukey, Bonferroni, SNK e Scott-knott controlam a taxa de erro tipo I por experimento adequadamente.

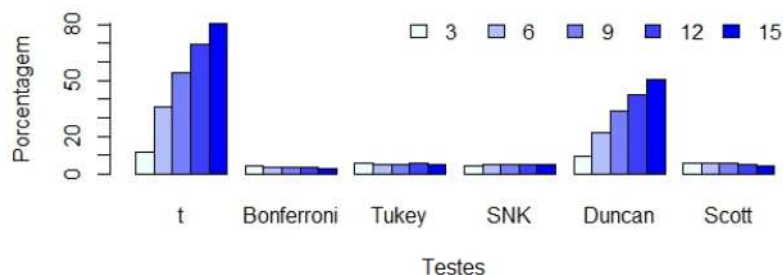


Figura 1 – Taxas de erro tipo I por experimento (TPE), em percentual, para diferentes testes de comparação múltipla de médias ao nível nominal de significância de 5%, em função do número de tratamentos e de variância igual a 5.

Tabela 2 – Taxas de erro tipo I por experimento (TPE), em percentual, para diferentes testes de comparação múltipla de médias ao nível nominal de significância de 5%, em função do número de tratamentos e de repetições e de variância igual a 5.

Tratamento	Repetição	Teste t	Bonferroni	SNK	Tukey	Duncan	Scott-Knott
		TPE	TPE	TPE	TPE	TPE	TPE
3	4	13,3**	5,0	7,7**	4,6	11,9**	6,7
3	8	11,7**	5,0	6,1	5,1	9,5**	6,3
3	12	11,3**	3,8	4,9	4,4	8,8**	4,4
3	16	11,3**	3,7	5,2	4,2	9,4**	6,2
6	4	34,5**	3,8	5,6	5,5	23,3**	5,6
6	8	33,8**	3,0*	4,3	4,1	22,1**	6,2
6	12	36,1**	4,6	6,3	6,0	21,9**	6,0
6	16	38,8**	4,1	5,4	5,1	22,6**	5,4
9	4	52,6**	4,0	6,0	6,0	35,6**	4,5
9	8	54,7**	3,7	5,3	5,1	33,2**	5,9
9	12	55,2**	4,6	5,7	5,6	33,5**	6,4
9	16	55,4**	3,4	5,4	5,3	33,0**	6,8**
12	4	68,0**	3,5	6,0	6,0	44,4**	4,4
12	8	67,6**	2,7*	4,7	4,7	39,1**	4,8
12	12	70,9**	4,2	6,1	6,0	42,8**	6,6
12	16	71,5**	4,2	5,8	5,7	42,2**	6,4
15	4	78,3**	3,8	5,4	5,4	51,8**	4,9
15	8	81,2**	3,4	5,5	5,5	50,2**	4,7
15	12	81,6**	3,5	5,2	5,1	47,2**	5,2
15	16	80,3**	3,3	4,8	4,8	53,6**	4,5

* Taxas de erro tipo I (por experimento) que ficaram abaixo do limite inferior do intervalo de confiança com 99% de probabilidade (3,225%) para a proporção empírica desta taxa.

** Taxas de erro tipo I por experimento que ficaram acima do limite superior do intervalo de confiança com 99% de probabilidade (6,775%) para a proporção empírica desta taxa.

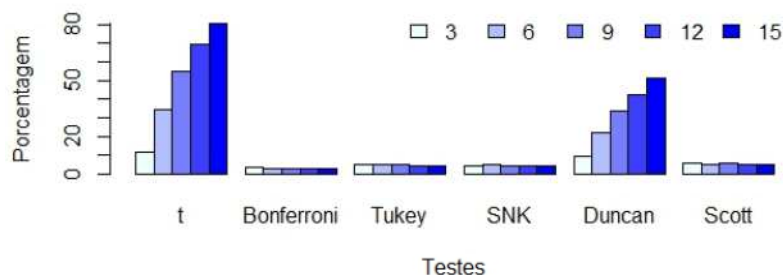


Figura 2 – Taxas de erro tipo I por experimento (TPE), em percentual, para diferentes testes de comparação múltipla de médias ao nível nominal de significância de 5%, em função do número de tratamentos e de variância igual a 10.

Tabela 3 – Taxas de erro tipo I por experimento (TPE), em percentual, para diferentes testes de comparação múltipla de médias ao nível nominal de significância de 5%, em função do número de tratamentos e de repetições e de variância igual a 10.

Tratamento	Repetição	Teste t	Bonferroni	SNK	Tukey	Duncan	Scott-Knott
		TPE	TPE	TPE	TPE	TPE	TPE
3	4	11,7**	4,8	6,4	5,9	10,1**	5,7
3	8	11,1**	3,7	4,5	4,0	8,9**	4,9
3	12	11,9**	4,3	5,6	4,9	10,2**	5,7
3	16	11,0**	3,8	5,3	4,4	9,1**	6,9**
6	4	32,4**	4,0	6,0	5,8	21,7**	5,8
6	8	35,2**	3,2*	4,8	4,4	21,7**	5,3
6	12	32,6**	3,1*	3,9	5,1	20,6**	5,6
6	16	37,0**	3,1*	5,2	4,7	23,8**	5,5
9	4	53,3**	3,4	5,4	5,2	34,5**	5,6
9	8	54,0**	2,9*	4,4	4,4	33,9**	6,3
9	12	57,9**	4,7	7,1**	6,0	35,6**	5,7
9	16	54,7**	2,4*	3,9	4,0	32,0**	6,6
12	4	68,7**	3,5	5,6	5,5	44,6**	4,3
12	8	69,0**	2,7*	4,7	4,7	40,7**	5,6
12	12	69,1**	3,5	5,5	5,5	42,0**	5,4
12	16	70,1**	3,4	4,8	4,8	41,8**	6,4
15	4	78,4**	3,4	5,2	5,2	48,9**	4,0
15	8	81,1**	3,4	4,7	4,7	53,1**	5,5
15	12	81,2**	3,5	4,9	4,8	52,8**	5,7
15	16	80,6**	3,7	4,8	4,8	49,8**	7,1**

* Taxas de erro tipo I (por experimento) que ficaram abaixo do limite inferior do intervalo de confiança com 99% de probabilidade (3,225%) para a proporção empírica desta taxa.

** Taxas de erro tipo I por experimento que ficaram acima do limite superior do intervalo de confiança com 99% de probabilidade (6,775%) para a proporção empírica desta taxa.

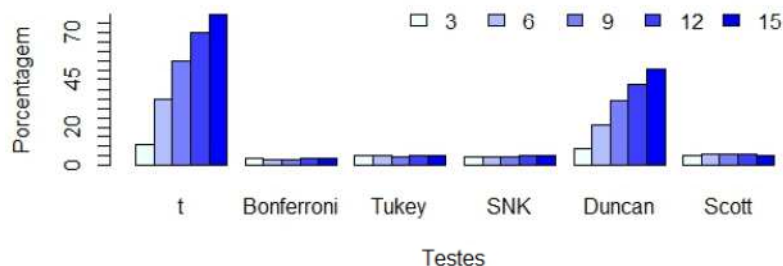


Figura 3 – Taxas de erro tipo I por experimento (TPE), em percentual, para diferentes testes de comparação múltipla de médias ao nível nominal de significância de 5%, em função do número de tratamentos e de variância igual a 15.

Tabela 4 – Taxas de erro tipo I por experimento (TPE), em percentual, para diferentes testes de comparação múltipla de médias ao nível nominal de significância de 5%, em função do número de tratamentos e de repetições e de variância igual a 15.

Tratamento	Repetição	Teste t	Bonferroni	SNK	Tukey	Duncan	Scott-Knott
		TPE	TPE	TPE	TPE	TPE	TPE
3	4	10,7**	4,2	5,5	4,7	8,9**	5,1
3	8	12,0**	4,1	5,7	4,9	9,9**	6,3
3	12	10,9**	3,5	4,5	4,0	9,1**	4,5
3	16	10,0**	3,5	4,5	4,1	7,9**	5,4
6	4	31,8**	2,8*	4,7	4,1	21,2**	4,9
6	8	36,8**	3,1*	4,5	4,3	21,0**	6,1
6	12	34,1**	3,4	5,1	4,7	21,6**	5,9
6	16	36,1**	3,5	5,1	4,7	22,4**	6,5
9	4	51,2**	2,3*	4,5	4,4	33,8**	4,1
9	8	54,8**	3,4	4,8	4,6	32,9**	5,8
9	12	57,8**	3,0*	4,3	4,3	33,5**	6,5
9	16	56,6**	4,1	5,5	5,3	34,9**	6,3
12	4	67,1**	4,8	5,5	5,4	43,7**	4,4
12	8	69,7**	3,1*	4,5	4,4	42,3**	6,9**
12	12	71,4**	4,0	5,7	5,7	41,9**	5,3
12	16	70,0**	3,8	5,2	5,2	41,2**	6,3
15	4	76,3**	4,3	5,8	5,8	50,7**	4,8
15	8	80,0**	3,5	4,9	4,9	52,2**	5,9
15	12	80,4**	4,0	5,5	5,4	49,3**	5,5
15	16	81,7**	3,7	4,9	4,9	52,1**	5,4

* Taxas de erro tipo I (por experimento) que ficaram abaixo do limite inferior do intervalo de confiança com 99% de probabilidade (3,225%) para a proporção empírica desta taxa.

** Taxas de erro tipo I por experimento que ficaram acima do limite superior do intervalo de confiança com 99% de probabilidade (6,775%) para a proporção empírica desta taxa.

5 Conclusão

Nos cenários formados pela combinação entre o número de tratamentos (3, 6, 9, 12 e 15), o número de repetições (4, 8, 12 e 16) e as variâncias iguais a 5, 10 e 15, observou-se que o teste t não controla a taxa do erro tipo I por experimento, pois as TPE sempre foram superiores ao nível nominal de significância estabelecido, e conforme o aumento do número de tratamentos a sua taxa vai ser aproximado de 100%. Para o teste de Scott-Knott a taxa de erro por experimento, quase sempre esteve de acordo com os níveis nominais estabelecidos, conforme a variação do número de tratamentos, repetição e variância. Para o teste de Tukey as taxas de erro não apresentaram nenhuma estimativa fora do intervalo de confiança estimado independentemente da variação do número de tratamentos, repetições e variância o que indica que o teste controla bem esta taxa. O teste de Bonferroni em 21,67% a TPE foi inferior ao nível nominal (5%), isto implica que o teste é melhor do que se propõe, porque por sua vez ele fornece taxas mais baixas para o erro tipo I. Já para o teste de Duncan, há falta de controle da taxa de erro do tipo I, conforme o aumento do número de tratamentos a taxa do erro tende a aproximadamente 100%. Por fim, o teste SNK controla a taxa de erro tipo I por experimento, independente da variação dos tratamentos, repetição ou variância as estimativas ficaram dentro do intervalo de confiança conforme o nível nominal estabelecido.

Referências

- BANZATTO, D. A.; KRONKA, S. do N. *Experimentação Agrícola*. 4 edição. ed. Jaboticabal: [s.n.], 2006. Citado 2 vezes nas páginas 12 e 25.
- BARBOSA, N. F. M. *Notas de Aulas*. Campina Grande- PB: Universidade Estadual da Paraíba, 2016. Citado na página 16.
- CANTERI, J. S. V. F. M. G.; CLAUDIA. Sasm - agri: Sistema para análise e separação de médias em experimentos agrícola pelos métodos scott-knott, tukey e duncan. *Revista Brasileira de Agrocomputação*, v. 1, n. 2, p. 18–24, 2001. Citado 2 vezes nas páginas 14 e 16.
- CARDELLINO, R. A.; SIEWERDT, F. Utilização correta e incorreta dos testes de comparações médias. *Revista Sociedade Brasileira de Zootecnia*, v. 21, n. 06, p. 985–995, 1992. Citado na página 17.
- FARIAS, A. M. L. de. *Infêrencia Estatística*. Fluminense-RJ: Universidade Federal Fluminense, 2008. Citado na página 15.
- FINNEY, D. J. *An introduction to statistical science in agriculture*. [S.l.]: Wiley; 4th edition, 1972. Citado na página 12.
- FISHER, S. R. A. et al. *The design of experiments*. Oliver and Boyd Edinburgh, 1951. Citado na página 14.
- FISHER, W. D. On grouping for maximum homogeneity. *Journal of the American statistical Association*, Taylor & Francis Group, v. 53, n. 284, p. 789–798, 1958. Citado na página 29.
- GARCIA, S. L. R.; LEITE, H. G. *Curso de Estatística Experimental*. Viçosa, Minas Gerais: [s.n.], 2006. Citado na página 27.
- GIRARDI, L. H.; FILHO, A. C.; STORCK, L. Erro tipo i poder de cinco testes de comparação múltipla de médias. *Rev. Bras. Biom*, v. 27, n. 1, p. 23–36, 2009. Citado na página 31.
- GOMES, F. P. *Curso de estatística experimental*. 13. ed. [S.l.]: Escola Superior de Agricultura "Luiz de Queiroz", 1990. Citado na página 25.
- LEAL, J. G.; PORRA, A. M. L. *Disenõ Estadístico de Experimentos Análisis de la Varianza*. [s.n.], 1998. Disponível em: <<http://www.siapi.es/geu>>. Citado 3 vezes nas páginas 20, 21 e 22.
- LUNA, J. G. de. *Planejamento e Análises Estatísticas de Experimentos I*. 1. ed. [S.l.]: Universidade Estadual da Paraíba, 2015. Citado na página 14.
- MENDIBURU, F. de. Statistical procedures for agricultural research. *Cran R*, p. 157, jun 2012. Citado na página 30.

- MOOD, A. M. Introduction to the theory of statistics. McGraw-hill, 1950. Citado 2 vezes nas páginas 18 e 23.
- OLIVEIRA, A. de; FERREIRA, D. F.; RAMALHO, M. P. Experimentação em genética e melhoramento de plantas. *Editora Univ. Fed. de Lavras*, p. 326, 2000. Citado na página 32.
- PINTO, M. P. Microsoft excel 2010. *Edições Centro Atlântico*, 2011. Citado na página 30.
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2016. Disponível em: <<https://www.R-project.org/>>. Citado na página 30.
- RIPLEY, B. et al. Package ‘mass’. *Cran R*, Citeseer, 2013. Citado na página 30.
- SCOTT, A. J.; KNOTT, M. A cluster analysis method for grouping means in the analysis of variance. *International Biometric Society*, v. 30, p. 507 – 512, sep 1974. 3. Citado na página 28.
- SILVA, C. M. R. da. *Uso do teste de Scott-Knott e da análise de agrupamentos, na obtenção de grupos de locais para experimentos com cana-de-açúcar*. Tese (Doutorado) — Escola Superior de Agricultura “Luiz de Queiroz”, 2007. Citado na página 29.
- SILVA, C. P. da. *Métodos de comparações múltiplas*. Dissertação (Monografia) — Universidade Estadual da Paraíba, Campina Grande-PB, dec 2011. Citado 2 vezes nas páginas 17 e 21.
- SILVA, D. F. F. E. C. D.; BEARZOTI, E. Avaliação do poder e taxas de erro tipo I do teste de Scott-Knott por meio do método de Monte Carlo. *Ciênc. agrotec.*, p. 10, sep 1999. Citado 2 vezes nas páginas 28 e 33.
- SILVA, L. J. da. Duncan test, tukey test and scott-knott test. *Cran R*, p. 5, feb 2015. Citado na página 30.

Apêndices

APÊNDICE A – Rotina

Três tratamento, quatro repetições e variância igual a cinco.

```

require(MASS)
require(stats)
require(agricolae)
require(laercio)

meuve<- matrix(nrow = 1000,ncol = 6)
colnames(meuve)<- c("Teste t","Bonferroni","SNK","Tukey","Duncan","Scott")
for (i in 1:1000) {
mu<- c(rep(100,3))
Sigma<- diag(x=5,nrow = 3,ncol = 3)
a<- mvrnorm(n = 4, mu, Sigma)
y<-c(a[,1],a[,2],a[,3])
x<-c(1,1,1,1,2,2,2,2,3,3,3,3)
x1<- as.factor(x)
model<- aov(y~x1)

# Teste t
t <- LSD.test(model,"x1", main="simulação")
meuve[i,1] = length(unique(t\$groups[,3]))

# Teste de Bonferroni
Bonferroni <- LSD.test(model,"x1", p.adj="bonferroni", main="simulação")
meuve[i,2] = length(unique(Bonferroni\$groups[,3]))

# Teste de Student-Newman-Keuls
SNK <- SNK.test(model,"x1",main="simulação")
meuve[i,3] = length(unique(SNK\$groups[,3]))

# Teste de Tukey
Dms <- HSD.test(model,"x1", group=TRUE,main="simulação")
meuve[i,4] = length(unique(Dms\$groups[,3]))

# Teste de Duncan

```

```
Duncan <- duncan.test(model,"x1",main="simulação")
meuve[i,5] = length(unique(Duncan\\$groups[,3]))
```

```
# Teste de Scott-Knott
```

```
Scott <- LScottKnott(model, "x1")
meuve[i,6] = length(unique(Scott\\$' ')) }
```