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

Russilano Costa Silva

Testes de Comparações de Médias

Campina Grande
Setembro-2010

Russilano Costa Silva

Testes De Comparações 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.

Orientador:
Prof. Dr. João Gil de Luna

Campina Grande
Setembro-2010

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL – UEPB

S586t Silva, Russilano Costa.

Testes de comparações de médias [manuscrito] / Russilano Costa Silva. – 2010.

27 f.: il.

Digitado.

Trabalho de conclusão de Curso (Graduação em Estatística) – Universidade Estadual da Paraíba, Centro de Ciências e Tecnologias, 2010.

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

1. Estatística. 2. Análise de Variância. 3. Comparações. I. Título.

21. ed. CDD 519.5

Russilano Costa Silva

Aprovado em: 10 / 10 /2010

Banca Examinadora:

Prof. Dr. João Gil De Luna
Departamento de Matemática e Estatistica
- CCT/EUPB
Orientador

Prof. Ms. Juarez Fernabdes de Oliveira -
CCT/UEPB
Examinador

Prof. Dr. Gustavo Henrique Esteves
Departamento de Matemática e Estatistica
- CCT/EUPB
Examinador

Dedicatória

Dedico esta monografia a todos que sempre me incentivaram a concluir meu curso. Em especial meus pais e mestres. Me sinto muito feliz e é com muito carinho, que faço esta dedicação.

Agradecimentos

À Deus.

Ao Professor Orientador, Prof. Dr João Gil de Luna pela orientação e dedicação a este trabalho e pelo apoio e colaboração durante a orientação da pesquisa.

Aos meus familiares: Evaldo (pai), Fatima (mãe), Raissa e Russien (irmãos)e a Juliana (minha noiva), que sempre me apoiaram e incentivaram e por estarem sempre presentes em todas as fases da minha vida.

Ao departamento de Matemática e Estatística do CCT/EUPB, pela estrutura e dedicação ao curso de Bacharelado em Estatística.

Aos funcionários do departamento de Matemática e Estatística, Agnaldo e Mercias por todo serviço prestado durante o curso.

E a todos os amigos que conquistei durante os cinco anos de curso, que me incentivaram e vibraram comigo a cada conquista durante o curso.

Resumo

Esta Monografia, cujo conteúdo foi elaborado para servir como apresentação da estrutura conceitual e metodológica de algumas técnicas estatísticas de comparações de médias, tem como objetivo comparar entre si as médias dos tratamentos ou de grupos de tratamentos, com a realização da análise de variância, conhecida como (ANOVA). Se a mesma confirmar a existência de diferenças significativas entre os tratamentos, é conveniente analisar que médias ou tratamentos são estatisticamente diferentes e em quanto oscila o valor dessas diferenças, para tal aplica-se os testes de comparações de médias.

Abstract

This Monograph, whose content was elaborated to serve as presentation of the conceptual and metodológica structure of some statistical techniques of comparisons of averages, has as objective to compare between itself the averages of the treatments or groups of treatments, with the accomplishment of the analysis of variance, known as (ANOVA). If the same one to confirm the existence of significant differences between the treatments, is convenient to analyze that average or treatments are statistical different and in how much oscillates the value of these differences, for such applies the tests of comparisons of averages.

Sumário

Lista de Tabelas

1 Introdução

O presente trabalho visa proporcionar uma fundamentação teórica da estatística, aos pesquisadores das diversas áreas do conhecimento, tais como, biologia, agronomia, ciências sociais e engenharia entre outras. Apresenta-se um conjunto de testes de hipóteses, particularmente, de comparações de médias, que podem levar o pesquisador a tomar decisões importantes na sua pesquisa.

As comparações de médias, aqui apresentadas, refem-se a estudos experimentais, cujo objetivo principal é verificar se existem diferenças significativas entre os tratamentos utilizados na pesquisa experimental. Utiliza-se o teste F de Fischer-Snedecor no processo da análise de variância (ANOVA), para testar a hipótese $H_0 : \mu_1 = ... = \mu_I = \mu$, isto é, a hipótese de nulidade dos efeitos das médias dos I tratamentos, contra a hipótese alternativa $H_1 : \mu_i \neq \mu_j$, para algum par $i, j(i, j = 1, 2, \cdots, I; i \neq j)$. Sempre que o resultado do teste F for significativo, existem fortes indícios de que pelo menos um dos contrastes dos I tratamentos seja significativo ao nivel de significancia α adotado. Os dados obtidos nos experimentos podem ser submetidos a vários tipos de análise estatística, sendo que uma das mais importantes é a análise de variância (ANOVA), (SANTOS; GHEYI, 2003).

Executada a análise de variância, e constatado que houve efeito significativo nos tratamentos, aplica-se os testes de comparações de médias, a fim de verificar quais médias diferem entre sí e em quanto oscila essa diferença. Apresenta-se então, a fundamentação teórica dos testes de comparações de médias. Contraste da diferença mínima significativa, que foi sugerido por Fischer em 1935; Método de Bonferroni, onde utiliza-se a desigualdade de Bonferroni; Método de Tukey, método que compara as médias, duas a duas; Método de Duncan, realizado de forma sequêncial e depende do número de médias que se comparam; e por último o método de Scheffé, que se propõe a realizar qualquer contraste entre médias de tratamentos.

Para ilustrar a fundamentação teórica mencionada, usou-se um experimento de competição de variedades de cana-de-açucar, no qual, foi feita uma análise de variância e aplicado o teste F, após a verificação de significância de pelo menos um dos tratamentos, foram aplicados os seguintes testes de comparações de médias: Teste t-Student, teste de Tukey, teste de Duncan e por último o teste de Scheffé. A metodologia empregada, baseia-se em revisão de textos especializados e livros que contemplam tanto a teoria dos testes quanto a sua aplicação. Seu objetivo é apresentar os presupostos e a fundamentação teórica envolvida nos principais métodos de comparações de médias utilizadas na Estatística Experimental, bem como utilizar exemplos reais para ilustrar tais teorias.

2 Fundamentação Teórica

2.1 Comparações múltiplas de médias

Ao se estudar o comportamento de uma variável resposta em relação aos níveis do fator (Y), objetiva-se saber, se existe diferenças entre as médias dos níveis desse fator (Y), através da análise de variância.

A análise proposta é feita através da técnica estatística de comparações múltiplas, que permitem testar hipóteses do tipo $H_0 : C = 0$ Vs $H_1 : C \neq o$, onde $C = \sum_{i=1}^{K} C_i\mu_i$, com a restrição de que $\sum_{i=1}^{K} C_i = 0$, onde, K é o número de níveis do fator e C_i denominado de contraste.

Neste trabalho, apresenta-se diversos métodos de comparações múltiplas de médias, entre as quais, destacam-se o método LSD (Diferença Mínima Significante), Bonferroni, Tukey, Duncan e Scheffé.

2.2 Comparações baseadas na distribuição *t-Student*

Os testes estatísticos para comparações de médias mais utilizados baseiam-se na distribuição *t-Student*. Cujo interesse seja de comparar os efeitos das médias de I tratamentos e se quer contrastar quaisquer hipóteses da forma:

$$\begin{cases} H_0 : \mu_i = \mu_{i'} \\ \text{vs} \\ H_1 : \mu_i \neq \mu_{i'}, \quad \forall\, i \neq i',\ i, i' = 1, 2, \cdots, I. \end{cases}$$

A técnica mais antiga e popular para efetuar estas comparações mútiplas é o procedimento LSD(Least Significant Difference), que apresenta-se a seguir.

2.3 Contraste da diferença mínima significativa

Este procedimento foi sugerido por Fisher em 1935 e será o primeiro método de comparações de médias a ser utilizado neste trabalho. Deve ser aplicado quando a hipótese de igualdade de todas as médias dos tratamentos tiver sido rejeitada. A base teórica se sustenta nos seguintes argumentos:

Sendo

$$\bar{y}_i - \bar{y}_{i'} \sim N\left[\mu_i - \mu_{i'}\ ;\ \left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right)\sigma^2\right] \tag{2.1}$$

segue que,

$$t = \frac{(\bar{y}_i - \bar{y}_{i'}) - (\mu_i - \mu_{i'})}{\sqrt{(\frac{1}{n_i} + \frac{1}{n_{i'}})\hat{S}_R^2}} \sim t(\nu), \tag{2.2}$$

onde, $\hat{S}_R^2$ é o quadrado médio dos resíduos, obtido a partir da análise da variância, n_i e $n_{i'}$ é o número de observações correspondentes à cada média, e ν é o número de graus de liberdade associados a $\hat{S}_R^2$. Assim sendo, intervalos com $(1-\alpha)100\%$ de confiança para $\mu_i - \mu_{i'}$ serão obtidos com base em

$$Pr\left[-t_{\frac{\alpha}{2}} < \frac{(\bar{y}_i - \bar{y}_{i'}) - (\mu_i - \mu_{i'})}{\sqrt{(\frac{1}{n_i} + \frac{1}{n_{i'}})\hat{S}_R^2}} < t_{\frac{\alpha}{2}}\right] = 1 - \alpha \tag{2.3}$$

que resulta em,

$$\begin{aligned} IC_{[(1-\alpha)100\%]}(\mu_i - \mu_{i'}) \quad &: \quad \left[(\bar{y}_i - \bar{y}_{i'}) - t_{\frac{\alpha}{2}}\sqrt{\left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right)\hat{S}_R^2};\ (\bar{y}_i - \bar{y}_{i'}) + t_{\frac{\alpha}{2}}\sqrt{\left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right)\hat{S}_R^2}\right] \\ &: \quad [(\bar{y}_i - \bar{y}_{i'}) - LSD\ ;\ (\bar{y}_i - \bar{y}_{i'}) + LSD], \end{aligned}$$

onde,

$$LSD = t_{\frac{\alpha}{2}}\sqrt{\left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right)\hat{S}_R^2}.$$

Para contrastar as hipóteses $H_0 : \mu_i = \mu_{i'} \quad vs \quad H_1 : \mu_i \neq \mu_{i'}$, com $i \neq i'$, argumenta-se que sob H_0 e sob as hipóteses associadas (modelo proposto no experimento), utiliza-se a seguinete estatística.

$$t = \frac{\bar{y}_i - \bar{y}_{i'}}{\sqrt{\left(\frac{1}{n_i} + \frac{1}{n'_i}\right)\hat{S}_R^2}} \sim t_{(\nu)} \tag{2.4}$$

Em palavras, a estatística t segue uma distribuição de probabilidade *t-Student* com ν graus de liberdade. Portanto, se conclui que as médias μ_i e $\mu_{i'}$ são estatisticamente diferentes se

$$|\bar{y}_i - \bar{y}_{i'}| > LSD, \tag{2.5}$$

onde, a quantidade LSD, denominada diferença mínima significativa, é dada por

$$LSD = t_{(\frac{\alpha}{2};\,\nu)}\sqrt{\left(\frac{1}{n_i}+\frac{1}{n_{i'}}\right)\hat{S}_R^2}. \tag{2.6}$$

Sendo, n_i e $n_{i'}$ é o número de observações correspondente a cada média observada no experimento, $t_{(\frac{\alpha}{2};\,\nu)}$ é o valor crítico da distribuição t-Student com ν graus de liberdade, cuja a área a sua direita corresponde a $\frac{\alpha}{2}$. Caso o experimento seja balanceado, isto é, se cada média dos tratamentos for calculada a partir do mesmo número de unidades experimentais, n, então tem-se que

$$LSD = t_{(\frac{\alpha}{2};\,\nu)}\sqrt{\frac{2\hat{S}_R^2}{n}}. \tag{2.7}$$

O procedimento LSD é simples de utilizar, pode ser aplicado tanto em modelos equilibrados como não equilibrados. Além disso, proporciona também intervalos de confiança para diferenças de médias, estes intervalos são da forma:

$$((\bar{y}_i - \bar{y}_{i'} - LSD)\ ;\ (\bar{y}_i - \bar{y}_{i'} + LSD)) \tag{2.8}$$

2.4 Método de Bonferroni

Neste procedimento se fixa um nível de significância α que se divide entre cada uma das comparações consideradas e se utiliza a desigualdade de Bonferroni (Luna, J.G., 2007)

$$Pr\left(\bigcup_{m=1}^{M} A_m\right) \leq \sum_{m=1}^{M} Pr(A_m). \tag{2.9}$$

$$Pr\left(A_1 \cup A_2 \cup \cdots \cup A_M\right) \leq P(A_1) + P(A_2) + \cdots + P(A_M) \tag{2.10}$$

Realiza-se estimações por intervalos para as $M = \binom{I}{2}$ possíveis comparações, cada uma ao nível de significância $\alpha^* = \alpha/M$; isto dá origem a M intervalos de confiança contendo cada uma das possíveis diferenças $\mu_i - \mu_{i'}$ com probabilidade $1-\alpha^*$. Chamando C_m o m-ésimo intervalo, tem-se que:

$$Pr[\mu_{1m} - \mu_{2m} \in C_m] = 1 - \alpha^*, \quad m = 1, 2, \cdots, M \tag{2.11}$$

Sendo μ_{1m} e μ_{2m} a primeira e a segunda média da correspondente comparação, onde supomos que $1 \leq 1m < 2m \leq I$.

Aplicando a desigualdade de Bonferroni, (expressão 2.9), tem-se:

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

denota-se por $\bar{C}_m$ é o complemento do intervalo C_m.

Tendo em conta esse resultado e objetivando garantir um nível de significância α para um conjunto de M comparações por pares, ou um nível de confiança $1-\alpha$ para o conjunto de intervalos, basta tomar:

$$\alpha^* = \frac{\alpha}{M} \tag{2.13}$$

Com isso, evidentemente, a probabilidade de que todos os intervalos, C_m, contenham as correspondentes diferenças de médias, será ao menos $1-\alpha$. Assim os intervalos ficarão na forma:

$$\bar{y}_{1m} - \bar{y}_{2m} \pm t_{\frac{\alpha}{2M}} \sqrt{\hat{S}_R^2 \left(\frac{1}{n_{1m}} + \frac{1}{n_{2m}}\right)}. \tag{2.14}$$

Onde $\bar{y}_{1m}, \bar{y}_{2m}$ e n_{1m}, n_{2m}, são as médias e os tamanhos amostrais correspondentes a comparação m-ésima.

Denotando por $\theta_m = \mu_{1m} - \mu_{2m}, m = 1, 2, \cdots, M$, uma das M comparações lineares por pares de médias, para as quais interessa contrastar $H_0 : \theta_m = 0$ vs $H_1 : \theta_m \neq 0$.

Então, rejeita-se H_0, se

$$|\theta_m| > B_m,$$

e aceita-se H_0, em caso contrário. Onde

$$B_m = t_{\frac{\alpha}{2M}} \sqrt{\hat{S}_R^2 \left(\frac{1}{n_{1m}} + \frac{1}{n_{2m}}\right)} \tag{2.15}$$

No caso do modelo equilibrado os valores de B_m coincidem, tais valores se denotam por BSD e têm a seguinte expressão

$$BSD = t_{\frac{\alpha}{2M}} \sqrt{\hat{S}_R^2 \left(\frac{1}{n} + \frac{1}{n}\right)}, \tag{2.16}$$

Sendo n o número de observações de cada grupo e $t_{\frac{\alpha}{2M}}$ é o valor crítico da distribuição t (valor observado em tabela), com o mesmo número de graus de liberdade associado à variância residual, que deixa uma probabilidade $\alpha/2M$ a sua direita.

2.5 Método de Tukey (Honestly-significant-difference).

Este procedimento pode ser utilizado para testar todo e qualquer contraste envolvendo duas médias de tratamentos e considera em primeiro lugar, o caso do modelo equilibrado. Neste modelo deve-se construir intervalos conjuntos com $1-\alpha$ de confiança para todas as possíveis comparações entre duas médias associados aos I níveis, isto é, o número possível de comparações entre duas médias é $\binom{I}{2}$.

O coeficiente de confiança conjunto $1-\alpha$ indica que de cada 100 amostras em $(1-\alpha)\times 100$ delas, cada um dos intervalos contém a sua correspondente diferença de médias. Portanto, o coeficiente de confiança de cada um dos intervalos será ao menos de $1-\alpha$.

Para construir estes intervalos, considera-se os desvios:

$$(\bar{y}_1-\mu_1),(\bar{y}_2-\mu_2),\cdots,(\bar{y}_I-\mu_I). \tag{2.17}$$

Estes desvios são variáveis aleátorias independentes com média 0 e variância σ^2/n. Além disso, $\frac{\hat{S}_R^2}{n}$ é um estimador de $\frac{\sigma^2}{n}$ que é independente de tais desvios. Então da definição de intervalo estudentizado, tem-se que:

$$\frac{max(\bar{y}_i-\mu_i)-min(\bar{y}_i-\mu_i)}{\sqrt{\frac{\hat{S}_R^2}{n}}}\sim q_{[I;\nu]} \tag{2.18}$$

Onde:

- ν, é o número de graus de liberdade associado a $\hat{S}_R^2$,
- $max(\bar{y}_i-\mu_i)$, é o desvio maior,
- $min(\bar{y}_i-\mu_i)$, é o desvio menor.
- $q_{[I;\nu]}$ é a amplitude total estudentizada, cujo valor é encontrado em tabelas.

Então verifica-se que:

$$Pr\left\{\frac{max(\bar{y}_i-\mu_i)-min(\bar{y}_i-\mu_i)}{\sqrt{\frac{\hat{S}_R^2}{n}}}\leq q_{[I;\,\nu;\,\alpha]}\right\}=1-\alpha. \tag{2.19}$$

Considerando-se a desigualdade a seguir, gerada para todos os pares de tratamentos i e i'.

$$|(\bar{y}_i-\mu_i)-(\bar{y}_{i'}-\mu_{i'})|\leq max(\bar{y}_i-\mu_i)-min(\bar{y}_{i'}-\mu_{i'}) \tag{2.20}$$

Visto que na desigualdade (2.20) se verifica para todos os pares μ_i e $\mu_{i'}$ segue de (2.19) que:

$$Pr\left\{\left|\frac{max(\bar{y}_i-\mu_i)-min(\bar{y}_{i'}-\mu_{i'})}{\sqrt{\frac{\hat{S}_R^2}{n}}}\right|\leq q_{[I;\,\nu;\,\alpha]}\right\}=1-\alpha. \tag{2.21}$$

inclue todos os pares $I(I-1)/2$ de comparações entre as médias dos I níveis do fator.

Reordenando a desigualdade, tem-se os limites de confiança, ao nível de confiança conjunto $1-\alpha$ para todas as $\binom{I}{2}$ comparações por pares das I médias, dadas por:

$$(\bar{y}_i-\bar{y}_{i'})-HSD\leq\mu_i-\mu_{i'}\leq(\bar{y}_i-\bar{y}_{i'})+HSD, \tag{2.22}$$

onde,

$$HSD=q_{[I;\,\nu;\,\alpha]}\sqrt{\frac{\hat{S}_R^2}{n}}. \tag{2.23}$$

Se o intervalo (2.22) não contém o *zero* conclue-se que as médias μ_i e $\mu_{i'}$ diferem significativamente entre sí.

O coeficiente de confiança $(1-\alpha)$ indica que de cada 100 amostras em $(1-\alpha)\times100$ delas se determinará corretamente qual das $\binom{I}{2}$ comparações por pares de médias são significativas. Portanto, o método de Tukey resolve o contraste:

$H_0:\mu_i=\mu_i'$ vs $H_1:\mu_i\neq\mu_i'$,

Da seguinte maneira:

Se $[\bar{y}_i-\bar{y}_{i'}]\leq HSD\Rightarrow$ Aceita-se H_0

Se $[\bar{y}_i-\bar{y}_{i'}]>HSD\Rightarrow$ Regeita-se H_0

No caso do modelo não ser equilibrado aplica-se o ajuste de Tukey-Kramer, que consiste em trocar n pela média harmônica, n_h do tamanho dos grupos ou seja;

$$n_h=\frac{2}{\sum_{i=1}^{2}\frac{1}{n_i}} \tag{2.24}$$

2.6 Método Duncan

O teste de Duncan utiliza, como HSD de Tukey, a distribuição da amplitude estudentizada. A aplicação desse teste é sequencial, no sentido de que não se utiliza o mesmo valor crítico para todas as diferenças de médias, como é feito no teste de Tukey, mas um valor crítico que dependa do número de médias ordenadas compreendidas entre as médias que se comparam, sendo estas colocadas previamente em ordem crescente.

Será considerado em primeiro lugar, o caso do modelo equilibrado e depois será generalizado para o caso não equilibrado.

A regra de decisão:

Rejeitar $H_0:\mu_i=\mu_{i'}$ (que é equivalemte a rejeitar $H_0:\Psi_h=0$, onde $\Psi_h=\mu_i-\mu_{i'}$), se o valor absoluto da diferença entre a maior e a menor das p médias envolvidas no contraste, for maior que R_p. Isto é,

$$|\bar{y}_i - \bar{y}_{i'}| > R_p, \tag{2.25}$$

em que,

- A estatística de teste é dada por:

$$R_p = q_{[\alpha_p;p;\nu]}\sqrt{\frac{\hat{S}_R^2}{n}}, \quad p = 2, 3, \cdots, I; \tag{2.26}$$

- $q_{[\alpha_p;\, p;\, \nu]}$ é o valor crítico da amplitude estudentizada baseado na comparação da média maior e a média menor de p médias.
- $\widehat{S}_R^2$ é a quadrado médio dos resíduos com ν graus de liberdade;
- α_p é o nível de significância conjunto relativo às p médias consecutivas; quer dizer, é a probabilidade de rejeitar erroneamente ao menos uma das $p-1$ comparações independentes associadas as médias consideradas. Tal nível de significância está relacionado com o nível de significância α, de uma comparação individual, através da equação

$$\alpha_p = 1 - (1-\alpha)^{p-1} \tag{2.27}$$

Para a aplicação do teste de amplitude múltiplos de Duncan, uma vez que as médias estejam em ordem crescente, calcula-se as diferenças entre as médias, começando pelo valor menor em relação ao mais alto das $p = I$ médias dos tratamentos, comparando esta diferença com o valor R_I na equação (2.27) com um nível de significância α_I. Se estas duas médias não são significativamente diferentes, então se termina o contraste e se conclui que nenhuma das médias são significativamente diferentes entre si ao nível de significância α_I. Isto é, equivalente a não rejeitar $H_0 : \mu_1 = \mu_2 = \cdots = \mu_I$. Se as duas médias das extremidades são significativamente diferentes, o contraste continua.

No passo seguinte calcula-se o valor da diferença entre o valor menor e o segundo maior e esta diferença se compara com R_{I-1}. Se este contraste não for estatisticamente significativo, o teste para nessa comparação e apenas as duas médias extremas se consideram significativamente diferentes. Se este contraste é estatisticamente significativo, o teste continua até encontrar o primeiro par de médias que não seja significativamente distintas. A seguir, calcula-se a diferença entre a segunda média menor e a maior e se compara com R_{I-1}. Este processo continua até que se tenha considerado as diferenças entre todos as $I(I-1)/2$ pares possíveis.

Para modelos não equilibrados, a estatística do teste é dada por:

$$R_p = q_{[\alpha_p;\, P;\, \nu]}\sqrt{\frac{\hat{S}_R^2}{n_h}}, \quad p = 2, 3, \cdots, I, \tag{2.28}$$

Onde n_h é a média harmônica dada pela expressão (2.24).

A probabilidade de se rejeitar erroneamente ao menos uma hipótese nula ou seja, a probabilidade de detectar incorretamente como significativa a diferença entre duas médias de um grupo de tamanho p, é o nível de significância conjunto α_p relacionado com o nível de significância α por meio da expressão (2.27).

Por exemplo, se $\alpha = 0,05$, então

- Ao se comparar pares de médias adjacentes, o nível de significância conjunto é

$$\alpha_2 = 1 - (1 - 0,05)^{2-1} = 0,05,$$

  que logicamente coincide com o nível de significância α.

- Ao se comparar pares de médias separadas por uma média, o nível de significância conjunto é

$$\alpha_3 = 1 - (1 - 0,05)^{3-1} = 0,10.$$

- Ao se comparar pares de médias separadas por duas méidias, o nível de significância conjunto é

$$\alpha_4 = 1 - (1 - 0,05)^{4-1} = 0,142.$$

Observe que ao aumentar p, aumenta o nível de significância conjunto.

2.7 O teste de Scheffé

Scheffé (1953) propôs um método para realizar qualquer contraste entre duas ou mais médias de tratamentos. Este procedimento não requer que o modelo seja equilibrado.

Considere uma família de contrastes da forma:

$$\Psi = \sum_i \ell_i \mu_i$$

O objetivo deste procedimento é decidir, para cada um destes contrastes, entre as hipóteses:

$$H_0 : \Psi = 0 \quad vs \quad H_1 : \Psi \neq 0 \tag{2.29}$$

O método de Scheffé está baseado na construção de intervalos de confiança para todos os possíveis contrastes da forma (2.29). Esses intervalos têm um nível de confiança simultâneo $1 - \alpha$, isto é, a probabilidade de que todos os intervalos sejam iguais simultaneamente é igual a $1 - \alpha$. Schéffe demonstrou que esses intervalos de confiança têm a seguinte expressão:

$$\hat{\Psi} \pm S(\hat{\Psi})\sqrt{(I-1)F_{[\alpha;(I-1),(\nu)]}}, \tag{2.30}$$

onde $\hat{\Psi} = \sum_i \ell_i m_i$ é um estimador de Ψ, m_i é a média observada no tratamento i, ℓ_i são constantes reais, tais que, $\sum_i \ell_i = 0$ e $S(\hat{\Psi}) = \sqrt{\hat{S}_R^2 \sum_i \frac{\ell_i^2}{n_i}}$, sendo $\hat{S}_R^2$ a variação residual com ν graus de liberdade. Como regra de decisão, temos que, rejeita-se a hipótese H_0 sobre um contraste Ψ, se o intervalo de confiança para Ψ,

$$\left[\hat{\Psi} - S(\hat{\Psi})\sqrt{(I-1)F_{[\alpha;\,(I-1);\,(\nu)]}} \quad ; \quad \hat{\Psi} + S(\hat{\Psi})\sqrt{(I-1)F_{[\alpha;\,(I-1);\,(\nu)]}}\right],$$

não inclui o zero. Isto é, se:

$$|\hat{\Psi}| \geq S(\hat{\Psi})\sqrt{(I-1)F_{[\alpha;(I-1);\,(\nu)]}}.$$

Na prática procede-se do seguinte modo:

- Calcula-se :
$$S = S(\hat{\Psi})\sqrt{(I-1)F_{[\alpha;(I-1);(\nu)]}},$$
onde, I é o número de tratamentos envolvidos no experimento, e F é obtido da tabela da distribuição F, considerando o número de graus de liberdade de tratamento e do resíduo com nível de significância α.

- Calcula-se a estimativa do contraste $\hat{\Psi} = \sum_i \ell_i \bar{y}_i$;

- Caso a estimativa do contraste $\hat{\Psi}_h$ seja maior ou igual que o valor de S, então dizemos que o contraste é significativo ao nível de significância α adotado.

3 Aplicação

Para ilustrar as metodologias de comparações múltiplas das médias dos tratamentos consideradas neste trabalho, levou-se em conta um experimento de competição de variedades de cana-de-açúcar, instalado num delineamento em quadrados latinos com cinco variedades (V1=Co290, V2=Co421, V3=Co419, V4=POJ2878, V5=CP36-13). As produções, em kg por parcela, observadas no experimento, são dadas na **Tabela** 1.

Tabela 1: Produções de cana-de-açúcar em kg por parcela observadas no experimento

Linhas	Colunas(C_j)					Totais
(L_i)	C_1	C_2	C_3	C_4	C_5	($y_{i..}$)
L_1	432(V_4)	518(V_1)	458(V_2)	583(V_3)	331(V_5)	2322
L_2	724(V_3)	478(V_5)	524(V_1)	550(V_2)	400(V_4)	2676
L_3	489(V_5)	384(V_2)	556(V_3)	297(V_4)	420(V_1)	2146
L_4	494(V_2)	500(V_4)	313(V_5)	486(V_1)	501(V_3)	2294
L_5	515(V_1)	660(V_3)	438(V_4)	394(V_5)	318(V_2)	2325
Totais($y_{.j.}$)	2654	2540	2289	2310	1970	11763

3.0.1 Cálculo das somas dos quadrados

Primeiro deve-se montar a tabela da análise de variância (ANOVA), inicialmene calcula-se:

$$C = \frac{(y_{...})^2}{I \times J} = \frac{(11763)^2}{5 \times 5} = 5.534.726,76;$$

$$\begin{aligned} SQTotal &= \sum_{i,j} y_{ij(k)}^2 - C \\ &= 5.792.451,00 - 5.534.727,76 = 257.724,24; \end{aligned}$$

$$\begin{aligned} SQLinhas &= \frac{1}{J}\sum_{i=1}^{5} y_{i..}^2 - C \\ &= \frac{1}{5}\left[(2322)^2 + \cdots + (2325)^2\right] - 5.534.726,76 = 30.480,64; \end{aligned}$$

$$\begin{aligned} SQColunas &= \frac{1}{I}\sum_{j=1}^{5} y_{.j.}^2 - C \\ &= \frac{1}{5}\left[(2654)^2 + \cdots + (1970)^2\right] - 5.534.726,76 = 55.640,64; \end{aligned}$$

$$\begin{aligned} y_{..(1)} &= 515 + 518 + 524 + 486 + 420 = 2463 \text{ (Total de } V_1) \\ y_{..(2)} &= 494 + 384 + 458 + 550 + 318 = 2204 \text{ (Total de } V_2) \\ y_{..(3)} &= 724 + 660 + 556 + 583 + 501 = 3024 \text{ (Total de } V_3) \\ y_{..(4)} &= 432 + 500 + 438 + 297 + 400 = 2067 \text{ (Total de } V_4) \\ y_{..(5)} &= 489 + 478 + 313 + 394 + 331 = 2005 \text{ (Total de } V_5) \end{aligned}$$

$$\begin{aligned} SQVariedades &= \frac{1}{J}\sum_{k=1}^{5} y_{..k}^2 - C \\ &= \frac{1}{5}\left[(2463)^2 + \cdots + (2005)^2\right] - 5.534.726,76 = 137.488,24; \end{aligned}$$

$$\begin{aligned} SQRes. &= SQTotal - SQLinhas - SQColunas - SQVariedades \\ &= 257.724,24 - 30.480,64 - 55.640,64 - 137.488,24 = 34.114,72. \end{aligned}$$

Tabela 2: Análise de variância para o caso geral.

F.Variação	GL	SQ	QM	F
Linhas	$(I-1)$	SQL	QML	$\frac{QML}{QMRes}$
Colunas	$(J-1)$	SQC	QMC	$\frac{QMC}{QMRes}$
Variedades	$(K-1)$	SQV	QMV	$\frac{QMV}{QMRes}$
Resíduo	$(I-2)(J-1)$	$SQRes$	$QMRes$	-
Total	$IJ-1$	$SQTotal$	-	-

Fazendo as devidas substituições tem-se:

Conforme pode ser observado na (tabela 3) da ANOVA, a hipótese de igualdade das médias de produção das variedades foi rejeitada, ao nível de 1% de significância, pois $F = 12,09 > F_{[4\,;\,12;\,0,01]} = 5,41$. Isto significa, que existem fortes evidências de que pelo menos um par de médias das variedades diferem estatisticamente entre si.

Tabela 3: Análise de variância.

F.Variação	GL	SQ	QM	F
Linhas	4	30.480,64	7620,16	2,68
Colunas	4	55.640,64	13.910,16	$4,89^{*}$
Variedades	4	137.488,24	34.372,06	$12,09^{**}$
Resíduos	12	34.114,72	2.842,89=s^2	-
Total	24	257.724,24	-	-

3.1 Testes de comparações múltiplas:

3.1.1 Teste *t-Student*

Suponha que o contraste $\Psi_1 = -2\mu_1 + \mu_2 + \mu_3$ foi estabelecido no planejamento, a fim de testar os efeitos das médias dos tratamentos ao nível de 5% de significância. Então, a estimativa do contraste e respectiva estimativa da variância do seu estimador é:

$$\hat{\psi}_1 = -2\hat{m}_1 + \hat{m}_2 + \hat{m}_3 = 60,40 \; e \; \widehat{Var}(\hat{\psi}_1) = \sum \frac{c_i^2}{r_i} s^2 = \frac{6}{20} \times 2842,89 = 852,87$$

Assim, para contrastar as hipóteses:

$$\begin{cases} H_0 : \Psi_1 = 0 \\ vs \\ H_1 : \Psi_1 \neq 0 \end{cases}$$

Calculamos:

$$t = \frac{\hat{\psi}_1 - 0}{\sqrt{\widehat{Var}(\hat{\psi}_1)}}$$

$$t = \frac{60,40 - 0}{\sqrt{852,87}} = 2,068.$$

Como $t = 2,068 < t_{[12;\,5\%]} = 2,179$, então há fortes evidências de que os efeitos das médias dos tratamentos são iguais, logo, aceita-se H_0 e conclui-se, ao nível de 5% de segnificância, que o contraste é estatisticamente nulo.

3.1.2 Teste de Tukey

Para contrastar as hipóteses:

$$\begin{cases} H_0 : \mu_i = \mu_{i'} \\ vs \\ H_1 : \mu_i \neq \mu_{i'} \end{cases}$$

Calcula-se:

$$\Delta = q_{[K;\, \nu;\, \alpha]}\sqrt{\frac{QMres}{K}} = 4,51\sqrt{\frac{2.842,89}{5}} = 107,5.$$

Os valores absolutos das diferenças entre os pares de médias, são calculados conforme se segue:

$$\begin{aligned} &|m_1 - m_2| = |492,6 - 440,8| = 51,8; \\ &|m_1 - m_3| = |492,6 - 604,8| = 112,2; \\ &\vdots \\ &|m_4 - m_5| = |413,4 - 401,0| = 12,4. \end{aligned}$$

A tabela a seguir apresenta os confrontos entre todos os pares das médias.

$$\Delta = 107,5.$$

Tabela 4: Confronto entre todos os pares de médias.

	m_1 492,6	m_2 440,8	m_3 604,8	m_4 413,4
$m_2 = 440,8$	51,8	-	-	-
$m_3 = 604,8$	112,2	164,0	-	-
$m_4 = 413,4$	79,0	27,4	191,4	-
$m_5 - 401,0$	91,6	39,8	203,0	12,4

A regra de decisão, consiste em:

Aceita-se H_0 se, e somente se, $|\hat{m}_i - \hat{m}_{i'}| \leq 107,5.$

Rejeita-se H_0 se, e somente se, $|\hat{m}_i - \hat{m}_{i'}| > 107,5.$

Conclusões: Analisando a (tabela 4), Hà evidêencias de que a média de produção da variedade V_3, difere estatisticamente das médias de produção das variedades V_1, V_2, V_4 e V_5, ao nível de 5% de significância pelo teste de Tukey. Os demais confrontos de médias não foram estatisticamente significativos.

3.1.3 Teste de Duncan

Colocando-se as médias das variedades observadas no experimento, em ordem decrescente, temos que:

$$\begin{aligned} m_3 &= 604,8 \\ m_1 &= 492,6 \\ m_2 &= 440,8 \\ m_4 &= 413,4 \\ m_5 &= 401,0 \end{aligned}$$

$$\begin{aligned} R_5 = Z\sqrt{\tfrac{QMRes.}{r}} &\rightarrow R_5 = 3,370\sqrt{\tfrac{2842,89}{5}} = 80,4; \\ R_4 = Z\sqrt{\tfrac{QMRes.}{r}} &\rightarrow R_4 = 3,313\sqrt{\tfrac{2842,89}{5}} = 79,0; \\ R_3 = Z\sqrt{\tfrac{QMRes.}{r}} &\rightarrow R_3 = 3,225\sqrt{\tfrac{2842,89}{5}} = 76,9; \\ R_2 = Z\sqrt{\tfrac{QMRes.}{r}} &\rightarrow R_2 = 3,082\sqrt{\tfrac{2842,89}{5}} = 73,5. \end{aligned}$$

$$\begin{array}{lll} \hat{\psi}_1 = m_3 - m_5 = 203,8 & \text{comparar com} & R5 = 80,4; \\ \hat{\psi}_2 = m_3 - m_4 = 191,4 & \text{comparar com} & R4 = 79,0; \\ \hat{\psi}_3 = m_1 - m_5 = 91,6 & \text{comparar com} & R4 = 79,0; \\ \hat{\psi}_4 = m_3 - m_2 = 164,0 & \text{comparar com} & R3 = 76,9; \\ \hat{\psi}_5 = m_1 - m_4 = 79,2 & \text{comparar com} & R3 = 76,9; \\ \hat{\psi}_6 = m_2 - m_5 = 39,8 & \text{comparar com} & R3 = 76,9; \\ \hat{\psi}_7 = m_3 - m_1 = 112,2 & \text{comparar com} & R2 = 73,5; \\ \hat{\psi}_8 = m_1 - m_2 = 51,8 & \text{comparar com} & R2 = 73,5; \\ \hat{\psi}_9 = m_2 - m_4 = 27,4 & \text{comparar com} & R2 = 73,5; \\ \hat{\psi}_{10} = m_4 - m_5 = 12,4 & \text{comparar com} & R2 = 73,5. \end{array}$$

Portanto, observe que os contrastes $\Psi_1 = \mu_3 - \mu_5$, $\Psi_2 = \mu_3 - \mu_4$, $\Psi_3 = \mu_1 - \mu_5$, $\Psi_4 = \mu_3 - \mu_2$, $\Psi_5 = \mu_1 - \mu_4$ e $\Psi_7 = \mu_3 - \mu_1$ foram estatisticamente significativos ao nível de 5% de significância, ou seja, o valor da diferença entre a maior e menor média, para cada contraste, foi maior que o R equivalente para cada contraste. enquanto que os demais contrates não apresentaram significância estatística a esse nível.

3.1.4 Teste de Scheffé

Conforme foi mostrado na fundamentação teórica, o teste de Scheffé só deve ser aplicado quando o teste F for significativo, no entanto, no exemplo em questão, aplicou-se o teste à apenas a um contraste para ilustrar a realização do teste.

Seja o contraste de interesse:

$$\Psi_1 = -2\mu_1 + \mu_2 + \mu_3.$$

A estimativa desse contraste é dada por:

$$\hat{\psi}_1 = -2m_1 + m_2 + m_3 = 60,4.$$

Admitindo-se a estatística do teste com sendo S, para simplificar nossos calculos, calcula-se da seguinte forma:

$$S = \sqrt{(I-1)\hat{V}(\hat{\Psi})F_{[(I-1);\,\nu;\,\alpha]}},$$

onde,

$$\hat{V}(\hat{\Psi}) = \sum \frac{\ell_i^2}{r_i} S^2.$$

Então,

$$\hat{V}(\hat{\Psi}) = \left[\frac{(-2)^2}{5} + \frac{(+1)^2}{5} + \frac{(+1)^2}{5}]\right] 2842,89 = 3411,4680$$

e segue que

$$S = \sqrt{(5-1) \times 3411,4680 \times F_{[4;\,12;\,0,05]}} = \sqrt{4 \times 3411,4680 \times 3,26} = 210,9.$$

Portanto, como $\hat{\psi}_1 = 60,4 < S = 210,9$, há fortes evidências de que a hipótese $H_0 : \Psi_1 = -2\mu_1 + \mu_2 + \mu_3 = 0$ não é rejeitada e concluimos ao nível de 5% de significância que o contraste $\Psi_1 = 0$ (ou que o contraste Ψ_1 não é estatisticamente significativo).

4 Conclusão

Ao executar a análise de variância foi constatado ao nível de 1% de significância que a hipótese $H_0 : \mu_1 = ... = \mu_i$ foi rejeitada, ou seja, há fortes evidências de que pelo menos um par de médias das variedades diferem entre si, ($F_{calc.} = 12,09 > F_{tab} = 5,41$).

Foram aplicados os testes de comparações de médias, com o intuito de verificar se há signicância ou não entre as médias dos tratamentos. O teste *t-Student* ao nível de 5% de significância, revelou que há evidências de que a hipótese H_0 deve ser aceita, pois o $F_{calc.} = 2,08 < F_{tab.} = 2,179$, e o contraste estabelecido no experimento é nulo, ou seja, há evidências de que os tratamentos não diferem, quanto as suas médias. Os testes mais frequêntes utilizadas em comparações de médias baseim-se na distribuição *t-Student*.

O teste de *Tukey* a 5% de significância evidênciou que a variedade V_3 difere estatisticamente das variedades $V_1, V_2, V_4 e V_5$, este teste complementa o teste F e deve ser aplicado quando a análise de vaiância acusar significância estatístca e quando os tratamentos forem igualmente repetidos.

Pode-se verificar ao aplicar o teste *Duncan* quais os contrastes de médias que apresentaram significancia ao nível de 5%. De fato, observamos 6 contrastes, que foram: Ψ_1, Ψ_2, Ψ_3, Ψ_4, Ψ_5 e Ψ_7. A sua aplicação é mais trabalhosa que o teste de *Tukey*, mas se chega a resultados mais detalhados e se discrimina com mais facilidade entre os tratamnetos, isto é, o teste de *Duncan* indica resultados significativos em casos que o teste de *Tukey* não permite obter significação estatística. Tal como o *Tukey*,o *Duncan* exige para ser exato, que os tratamentos tenham o mesmo número de repetições.

O teste de *Scheffé* foi o último teste a ser aplicado, e verificou-se a 5% de significância que o contraste $\hat{\psi}_1$ não apresentou significância estatística, já que: $\hat{\psi}_1 = 60,4 < S = 210,9$. Os contrastes deste teste sempre envolvem mais de duas médias, sendo que, seus contrates podem ser estabelecidos a priori ou não do experimento.

Pode-se observar a importância de se utilizar de ferramentas estatísticas para compararmos se há diferenças entre médias e, além disso, quais médias diferem entre si. Tais ferramentas podem ser utilizadas em várias áreas do conhecimento, como por exemplo, na agronomia onde foi bastante difundido através dos estudos de *Ronald A. Fisher*, além de ser muito utilizada para verificar a viabilidade e a variabilidade da produção de certos tipos de rações, grão, variedade de plantas etc.

Referências

GOMES, Frederico Pimentel. Curso de Estatística Experimental, 13 Ed. Piracicaba SP, ESALQ, 1990. 467 p.

SANTOS, José Wellington dos; Gueyi, Hans Raj (Eds.) Estatística Experimental aplicada. Campina Grande: Editora Gráfica Marcone Ltda, 2003.213p

LUNA, J.G. Planejamento de experimento (notas de aulas) 2007.