

REVISÃO:

Fundamento dos testes estatísticos e sua aplicabilidade em ensaios experimentais com animais

Éder Clementino dos Santos. Docente do IFSM, - Campus de Inconfidentes-MG. Doutor em Zootecnia – ederclementino@gmail.com

Elesandre Clementino dos Santos – Médico Veterinário

Marília Fortunato Silva Mesquita – Gestor Ambiental

RESUMO

A palavra Estatística foi cunhada pelo acadêmico alemão Gottfried Achenwall (1719-1772), que foi um notável continuador dos estudos de Hermann Conrig (1606-1681). Gottfried determinou os objetivos da Estatística e suas relações com as demais ciências. A Estatística é uma parte da Matemática Aplicada que fornece métodos para coleta, organização, descrição, análise e interpretação de dados e para a utilização dos mesmos na tomada de decisões. Atualmente, os estudos estatísticos têm avançado rapidamente e, com seus processos e técnicas, têm contribuído para a organização dos negócios e recursos do mundo moderno. Um problema que é necessário aprender a resolver é o de testar uma hipótese, isto é, feita uma determinada afirmação sobre um parâmetro populacional, por exemplo, sobre uma média populacional ou uma proporção populacional, será que os resultados de uma amostra contrariam ou não tal afirmação?

As variáveis dependentes escolhidas e obtidas nos experimentos com animais de forma planejada e conduzidas adequadamente, podem ser submetidas a vários tipos de análises estatísticas, sendo a de comparações múltiplas de testes de médias uma importante ferramenta para verificar a possível existência do efeito de dois ou mais fatores testados no desempenho, rendimento de carcaça ou outro parâmetro zootécnico de qualquer animal. Assim, desde que a validade do erro não possa ser questionada, o aumento de precisão da experiência, que corresponde à diminuição do erro experimental, permite detectar diferenças dos tratamentos que de outra maneira seriam atribuídos a efeitos de flutuação aleatória. Nesse sentido, o objetivo de um teste estatístico de hipóteses é fornecer ferramentas que nos permitam aceitar ou rejeitar uma hipótese estatística através dos resultados de uma amostra. Sendo assim, este estudo objetiva apresentar de forma introdutória algumas noções em estatística experimental e suas variadas aplicações com ensaios com animais, visando permitir uma melhor compreensão da utilização de seus principais instrumentos de análise nos testes estatísticos. O importante desafio na utilização da análise em estatística é o de mostrar a grande amplitude de seu campo de aplicação e a articulação e fundamentação deste com a base conceitual e teórica no segmento de produção animal.

Palavras-chave: Teste estatístico, experimento animal, tipo de teste.

Basis of statistical tests and their applicability in experimental animal

ABSTRACT

The Statistical word was coined by the German academic Gottfried Achenwall (1719-1772), that was a notable prosecuting of Hermann Conrig's studies (1606-1681). Gottfried de-

terminated the Statistics objectives and your relationships with the other sciences. The Statistics is a part of the Applied Mathematics that supplies methods for collection, organization, description, analysis and interpretation of data and for the use of the same ones in the electric outlet of decisions. Nowadays, the statistical studies have been moving forward quickly and, with your processes and techniques, they have been contributing to the organization of the businesses and resources of the modern world. Is a problem that is necessary to learn to the one to solve of testing a hypothesis, that is, done make a certain declaration? The chosen dependent variables and obtained in the experiments with animals in a planned way and led appropriately, they can be submitted to several types of statistical analyses, being the one of multiple comparisons of tests of averages an important tool to verify the possible existence of the effect of two or more factors tested in the acting, carcass revenue or other parameter of the any animal. Like this, since the validity of the mistake cannot be questioned, the increase of precision of the experience, that corresponds to the decrease of the experimental mistake, allows to detect differences of the treatments that would be attributed to effects of aleatory flotation in another way. On a population parameter, for instance, about a population average or a population proportion, will it be that the results of a sample thwart or not such to declaration? In that sense, is the objective of a statistical test of hypotheses to supply tools that allow to accept us or to reject a statistical hypothesis through the results of a sample. Being like this, this study aims at to present in an introductory way some notions in experimental statistics and your varied applications with rehearsals with animals, seeking to allow a better understanding of your principal analysis instruments use in the statistical tests. The important challenge in the use of the analysis in statistics is it of showing the

great width of your application field and the articulation and to base of this with the conceptual and theoretical base in the segment of animal production.

Key-words: Statistical test, animal experiment, test type.

INTRODUÇÃO

Diversos testes estatísticos de ensaios com animais permitem aos pesquisadores inferir sobre os resultados experimentais. No entanto, para a aplicação correta dos testes estatísticos, é de fundamental importância que o pesquisador tenha conhecimento sobre quais os tipos de fatores, de variáveis respostas, de tratamentos e do delineamento experimental que irão constituir o seu trabalho. Segundo Silva (1999), a fundamentação teórica do delineamento de tratamento de um experimento com animal, tal como os tipos de fatores envolvidos, é fundamental para as inferências do pesquisador. Ainda, de acordo com o mesmo autor, a estrutura dos fatores experimentais, a estrutura dos fatores de unidade e a relação entre essas duas estruturas, determinada pela casualização, constituem a estrutura do experimento ou o delineamento experimental zootécnico.

Os tratamentos utilizados como grupos experimentais são submetidos à análise de variância para posterior comparação de suas médias. Os delineamentos conhecidos contemplam as mais diversas situações onde limitações de amostragem ou de infra-estrutura impedem a utilização de esquemas mais simples (Gomes, 1990).

Ensaio mais sofisticados, a utilização de mais de um fator experimental exige como consequência a investigação das interações entre os fatores. Assim, a modelagem não permite tão imediatamente a apreciação dessas interações. Isto ocorre, basicamente porque uma interação que pode ser facilmente identificada em um quadro de médias atra-

vés de testes elegidos, estará decomposta no modelo em tantos termos quantos forem os seus graus de liberdade quando esta apresentar significativa. Por sua vez, a decomposição fatorial na análise de variância, também não nos permite uma justa avaliação das referidas interações (Sampaio, 2007; Monteiro, 2002; Cardellino & Siewerdt 1992).

A possibilidade de utilizar um dos vários testes estatísticos existentes para a comparação de médias, de alguma maneira, seduz o experimentador que, entre comedido e inovador, deseja optar por um deles. Assim, a escolha de um teste deve contemplar aquele cujas conclusões advindas de seu uso sejam menos sujeitas a erros tidos como indesejáveis.

A experimentação animal é uma atividade considerada relativamente preocupante, uma vez que diversos autores descrevem que muitos artigos de periódicos em ciências veterinárias, zootecnia ou biologia poucos se verifica um envolvimento de parâmetros de significância expressiva de experimentação (Monteiro, 2002; Aquino, 1992).

Os mesmos autores destacam que são usados dados, muitas vezes coletados por rotina pelas associações de criadores, entidades zootécnicas ou de pesquisa que envolvem uma quantidade muito grande de números e/ou resultados do desenvolvimento e aplicação de tecnologias, nova ou conhecidas que não contribuem satisfatoriamente com a metodologia científica e experimental.

É bem verdade que a experimentação animal é central para a qualificação do método científico, pois questões que não possam ser respondidas por observações planejadas não podem ser qualificadas como ciência. Assim, a instalação de qualquer tipo de experimento com animal demanda o conhecimento de técnicas experimentais relativas à seleção não somente de animais a serem utilizados, mas também da aplicação de maneira adequada dos tratamentos a serem desenvol-

vidos em conformidade com o delineamento experimental apropriado (Sampaio, 2007).

O problema de comparação de médias na análise da variância de experimentos é uma questão que vem sendo estudada já algum tempo e para o qual há numerosas soluções, nenhuma, porém, inteiramente satisfatória (Chew, 1976; Pimentel Gomes, 1990 e Sampaio, 1993).

Os procedimentos para comparações múltiplas (PPCM) têm assim, merecido atenção de muitos pesquisadores.

Essa questão além de ser ampla apresenta uma maior importância de sua aplicabilidade, isto que há relatos referentes ao assunto, tais como: Scheffé (1959); Bancroft (1968); Neter & Wasserman (1974); Miller (1981) e muitas revisões com mais de 200 citações, como de O'Neill & Wetherill, (1971), Ferreira et al.; (1997), Sampaio, Monteiro (2002) e Sampaio (2007).

Embora a literatura sobre esse assunto seja vasta, não se pode afirmar que o problema esteja completamente resolvido, sendo ainda um campo aberto para novas pesquisas.

No esclarecimento da necessidade dos PPCM, considere-se uma análise de variância de um experimento inteiramente casualizado com g tratamentos. Existem infinitas possíveis comparações (contrastes) entre os g médias, das quais só $(g - 1)$ são ortogonais. Só entre comparações de duas médias, existem $g(g-1) / 2$. Por exemplo, com 06 médias, têm-se 05 comparações ortogonais, 15 comparações entre duas médias e inúmeras outras (Sampaio, 2007; Alvarez & Alvarez, 2006; Gomes, 1990).

Segundo Sampaio (2007), um par pré-escolhido entre as g médias pode ser comparado de maneira ótima pelo teste t de **student**; o intervalo de confiança para a diferença entre essas duas médias pode ser encontrado com uma confiança pré-fixada, diga-se $1-\alpha$. Para $(g-1)$ comparações ortogonais, se cada comparação tem coeficiente de

confiança ($1-\alpha$), o coeficiente de confiança conjunta será $(1-\alpha)^{g-1}$.

Para um conjunto de comparações não ortogonais, o coeficiente de confiança conjunta não é facilmente estabelecido e o mesmo não se dá para comparações sugeridas pelos dados, daí a necessidade de verificar a aplicabilidade de outros tipos de testes de PPCM aos delineamentos propostos nos ensaios com animais. Nesse sentido, há um número relativamente elevado de procedimentos para tal fim, cada um fundamentado em um particular conjunto de suposições que o tornam efetivo para propósitos específicos.

É comum nos experimentos com animais a obtenção de pelo menos dois conjuntos de dados para análise, o consumo de ração e o ganho médio de peso, que são correlacionados e, portanto, a análise de qualquer destas variáveis em separado poderá não fornecer informações confiáveis sobre as diferenças entre os tratamentos. Além disso, é prática comum o uso da análise da variância da conversão alimentar, variável esta que é o quociente do consumo de ração sobre o ganho de peso, com distribuição não normal (Bertoldo et al., 2008).

Como alternativas alguns autores recomendam o uso do ganho de peso ajustado para o consumo de ração como critério mais correto para avaliar o valor nutricional dos alimentos. Em termos de porcentagem de diferenças significativas, observou-se que quando o número de tratamento é menor ou igual a 3, a conversão alimentar é mais eficiente, e para maior número de tratamentos a análise de covariância é mais eficiente.

A IMPORTÂNCIA DA ESTATÍSTICA EM ENSAIOS COM ANIMAIS

As situações que conduzem à comparação de valores médios precisam considerar sempre dois aspectos essenciais para a escolha do teste adequado. Um diz-se

respeito à caracterização da resposta a ser medida (variável alvo) quanto a sua natureza (qualitativa ou quantitativa), à sua distribuição (normal ou não), à sua continuidade (contínua ou discreta) e à sua instabilidade (muito ou pouco instável).

Para as variáveis qualitativas e geralmente discretas, requerem a utilização de testes não paramétricos que não exigem a condição de normalidade para sua execução. Por outro lado, as variáveis de caráter qualitativo apesar de assumir valores quantitativos e com distribuição normal são utilizadas diversos tipos de testes de médias. Assim, o primeiro elemento de decisão para a escolha do teste estatístico está na percepção da instabilidade da variável a ser avaliada. A noção de instabilidade origina-se nos valores já registrados de coeficientes de variação para uma determinada variável. Independentemente da origem, ensaios zootécnicos apresentam CV de 5 a 35%, embora em alguns casos ocorram valores menores de CV (Sampaio, 2007; Aquino, 1992).

A variação do CV que é descrita por: $CV \frac{S}{\bar{X}} \cdot 100$ está mais fortemente ligada à variação do erro experimental \underline{s} , além da possibilidade de conter eventuais erros experimentais, caracteriza definitiva e descritivamente cada resposta. Variáveis muito instáveis sempre o serão. Esta instabilidade refletir-se-á no intervalo de confiança das médias do ensaio e conseqüentemente na comparação entre elas (Sampaio, 2007).

Por outro lado, o CV permite a comparação de resultados de diferentes experimentos envolvendo uma mesma variável-resposta, permitindo ao pesquisador quantificar a precisão de suas pesquisas. Conforme Monteiro (2002) e Gomes (1990) que alertam para o fato de que o pesquisador deve estar atento aos valores razoáveis dessa medida em sua área de pesquisa, pois os valores variam de acordo com a espécie animal estudada e com a variável-resposta utilizada.

Outro aspecto é que após a utilização de um teste na comparação de médias, existem dois tipos de erro reconhecidos pela teoria estatística, que é o *erro tipo I* - atribuir uma significância quando ela realmente não existir e o *erro tipo II* - atribuir uma equivalência quando realmente houver uma diferença significativa (Sampaio, 2007; Ferreira, 2000; Gomes, 1990).

Dessa forma, torna-se impossível o controle absoluto e simultâneo de ambos em um só ensaio, posto que seja antagônico. Ao se restringir o aparecimento de um, permite-se ocorrência do outro.

De acordo com Bertoldo et al., (2008), Sampaio (2007) e Gomes (1990), o pesquisador deve eleger a opção que mais lhe convier frente aos objetivos propostos. Em uma comparação onde o resultado favorável seria uma equivalência (repostas sorológicas de vacinas) seria inadmissível a utilização de um teste estatístico que beneficiasse a ocorrência do erro tipo II.

Por outro lado, a competição de tratamentos experimentais, onde se espera melhor eficiência de alguns, o erro indesejável seria o tipo I que acusaria artificialmente diferenças onde elas não existem.

A estratégia de controle dos erros tipo I e II devem sempre ser consideradas, mas a avaliação dos níveis probabilísticos dos mesmos dependeria do conhecimento prévio dos parâmetros populacionais. Assim, percebe-se que, devido à estrutura do intervalo de confiança da média que inclui como respostas típicas de um grupo experimental apenas aquelas contidas no intervalo central do nível de significância, as respostas muito instáveis estarão mais sujeitas ao erro tipo II e, portanto, deverão ser contempladas com testes estatísticos que reduzam esse risco (Sampaio, 2007; Silva, 1999).

Todos os estudos até então desenvolvidos, preocuparam-se em controlar aquele tipo de erro em particular, criando testes mais

rigorosos no controle do erro tipo I e consequentemente mais permissivo no do erro II.

Todas as críticas contra a utilização de comparações múltiplas apontam mais as falhas interpretativas de seus usuários e as estratégias de execução do que qualquer evidência que as desabone (Campos, 1984; Little, 1978).

A lógica dos testes estatísticos dá ao pesquisador condições de fazer inferências aos seus dados amostrais, evitando assim, conclusão restrita às amostras com que trabalhou, proporcionando ao pesquisador melhor entendimento aos resultados que obteve para toda a população. Com base nos resultados da amostra, o teste estabelece se, na população, a média de **A** é diferente da média de **B**, em determinado nível de significância. Em estatística, a palavra significância implica a idéia de que é muito provável que um resultado similar ao que foi obtido na amostra, teria sido obtido se toda a população tivesse sido estudada. Ou seja, que se possa realmente ser representativa de toda a população amostral (Sampaio, 1993; Gomes, 1990).

Mas muito provável não significa certo. Então a conclusão do teste estatístico está associada a algum tipo de erro. Se o pesquisador concluir que as médias são diferentes, pode estar cometendo um erro. Isto não significa que o pesquisador errou em seu procedimento, significa apenas que – por causa da flutuação amostral – o pesquisador pode ter observado uma amostra que apresenta uma diferença entre médias que não existia na população. A probabilidade dessa ocorrência é o nível de significância do teste.

Segundo Aquino (1992), de forma geral a primeira etapa no estudo dos dados de um experimento é realizar-se a análise da variância, para que, a partir desta, se possa testar a hipótese de igualdade das médias populacionais dos tratamentos testados pela estatística F. O mesmo autor ressalta que é muito importante antes de se realizar a análise

se diz respeito ao conhecimento da natureza dos tratamentos estudados, isto é, se estes são de efeito aleatório ou fixo. No primeiro caso, visa a estimação dos componentes de variância e, no segundo, a estimação e comparação dos efeitos, por meio de testes de comparação de médias. No entanto, tais procedimentos vêm sendo utilizados indiscriminadamente, conforme o relato de diversos autores.

De acordo com Sampaio (2007), os testes estatísticos põem à prova hipóteses a respeito da população. O pesquisador pode fazer duas ou mais hipóteses: uma é a de que a média de uma variável avaliada (quaisquer animal, não apenas os da amostra) por um método de pesagem **A** é igual à média da mesma variável avaliada por outro método de pesagem **B**. Essa hipótese da nulidade (se as médias são iguais a diferença entre elas é zero). Indica-se por H_0 (lê-se agá-zero) e se escreve:

H_0 : as médias são iguais

A outra hipótese é a de que a média da variável estudada pelo método de pesagem **A** (quaisquer animal, não apenas os da amostra) é diferente da média da mesma variável estudada pelo método de pesagem **B**. Essa hipótese é denominada hipótese alternativa. Indica-se por H_1 (lê-se agá-um) e se escreve:

H_1 : as médias são diferentes

Para decidir por uma das hipóteses, o pesquisador submete seus dados a um teste estatístico. Se o pesquisador concluir que as médias são diferentes, pode estar cometendo um erro. Só que o pesquisador não sabe, quando tomou essa decisão, se está ou não cometendo erro. No entanto, pode estabelecer a probabilidade de cometer esse tipo de erro. Essa probabilidade é o que os estatísticos denominam nível de significância do teste e indicam pela letra grega α (lê-se alfa). Então, nível de significância é a probabilidade de rejeitar H_0 , quando H_0 é verdadeira.

Finalmente, a escolha do valor de α é arbitrária. Quando se escolhe o nível de significância de 5%, é usual afirmar que o re-

sultado é *significante* e indicar isso com um asterisco. Quando se escolhe o nível de significância de 1%, é usual afirmar que o resultado é *altamente siginificante* e indicar isso com dois asteriscos.

A teoria estatística sempre deixou claro que frente a vários tratamentos (t) apenas os $(t-1)$ graus de liberdade poderiam ser decompostos e testados enquanto na realidade existiam $t(t-1)/2$ comparações possíveis. Os pesquisadores sempre se interessam mais por estas $t(t-1)/2$ comparações entre médias, já que os $t-1$ contrastes permitidos nem sempre retratavam a comparação de dois tratamentos e sim de grupos deles.

Segundo Fisher (1935), Gomes (1990), recomendaram discretamente que fora daquela decomposição, as comparações múltiplas fossem utilizadas com cautela, principalmente quando o teste F não fosse significativo.

Na realidade, os ensaios experimentais com animais têm apresentado um aumento do número de tratamento e, portanto, da amplitude da menor para a maior média que, aumentava a probabilidade de se detectarem diferenças significativas entre os $t(t-1)/2$ contrastes possíveis. Assim, o erro tipo I estaria sendo beneficiado à medida que a distância entre duas médias fosse aumentando.

Pode-se considerar de uma maneira simplista, que o objetivo de uma análise estatística dos dados de um experimento zootécnico é conhecer o efeito de seus tratamentos. Para este fim, os dados são submetidos à análise de variância (Anava), possibilitando, então, concluir sobre a existência ou não de diferenças entre as respostas dos tratamentos. Existindo evidências de efeitos diferenciados para tais, faz-se necessário o detalhamento do estudo para a identificação destas diferenças, através das medis obtidas nos experimentos. Existem dois procedimentos possíveis para o estudo das médias dos tratamentos: comparações múltiplas e análise de regressão. A distinção

para a aplicação destes recursos é dependente da natureza do fator em estudo, qualitativo ou quantitativo (Bertoldo et al., 2008). Da mesma forma, Cardellino & Sieverdt (1992) apontam que se o fator em estudo no experimento é uma variável qualitativa, o procedimento apropriado é o das comparações entre as médias dos tratamentos através de testes de comparações múltiplas. Sendo uma variável quantitativa, utiliza-se a análise de regressão para o estudo dos efeitos dos tratamentos na variável resposta. Os mais freqüentes erros detectados referem-se aos testes de comparação múltipla de médias, os quais podem tender a interpretações e conclusões incertas dos resultados de um experimento.

O pesquisador em geral gostaria de aplicar um teste para comparar médias, duas a duas e para isso, é necessário utilizar um método que forneça a *diferença mínima significativa* entre duas médias e, essa diferença seria o instrumento de medida. Dessa forma, toda vez que o valor absoluto da diferença entre duas médias é igual ou maior do que a diferença mínima significativa, as médias são consideradas estatisticamente diferentes, ao nível de significância estabelecida (Sampaio, 2007; Gomes, 1990).

A fundamentação teórica do delineamento de tratamento de um experimento, tal como os tipos de fatores envolvidos, é fundamental para as inferências do pesquisador (Silva, 1999). Apesar da preocupação dos pesquisadores com a análise dos dados, muitas vezes pode-se observar certo descaso com o emprego de testes de comparação de médias. Nesse sentido, o autor enfatiza que para explicar a resposta dos dados obtidos de maneira mais detalhada, vários procedimentos podem ser utilizados, como, por exemplo, os procedimentos de comparação múltipla de médias, sendo que, os testes de comparação de médias freqüentemente são os mais utilizados pelos pesquisadores.

Porém, apesar de uma diversidade de

testes estatísticos à disposição dos pesquisadores, a maioria os utiliza de modo incorreto. Em seus estudos, Petersen (1977) verificou que 40% dos autores pesquisados utilizaram em suas análises, algum tipo de teste de comparação de médias e que destes, 40% os utilizaram de forma inteiramente inapropriada quanto ao tipo dos dados envolvidos. De acordo com o mesmo autor, os testes de comparação de médias (t Student, Duncan, Tukey, Bonferroni, etc.) são apropriados para situações em que os tratamentos são níveis de fator qualitativo e não relacionados.

Os testes de comparação múltipla procuram atender aos interesses da experimentação moderna que utiliza um número maior de grupos de investigados. Devido às limitações amostral e de infraestrutura, este número varia entre 3 e 6 na grande maioria dos ensaios, exceto nas competições de variedades em agricultura e em geral em esquemas fatoriais (Sampaio, 2007).

TESTE F

O teste F foi obtido por Snedecor (1980) e tem por finalidade comparar estimativa de variâncias. O teste F mostra se existe diferença entre os tratamentos, mas não localiza a diferença, exceto no caso t=2 tratamentos. Portanto, a análise de um experimento exige a aplicação de um método de comparação múltipla para determinar a diferença de um tratamento com os demais. No entanto, Silva (1999) recomenda discretamente, que fora daquela decomposição, as comparações múltiplas fossem utilizadas com cautela, principalmente quando o teste F não fosse significativo.

De acordo com Gomes (1990), a estatística de F, denominada de razão de variâncias, pode ser definida como sendo o quociente de duas estimativas de variância, s^2_1 e s^2_2 , supostas independentes e calculadas com n_1 e n_2 graus de liberdade, respectivamente.

$$\text{Então: } F = \frac{s_1^2}{s_2^2}$$

Consideraremos sempre $s_1^2 > s_2^2$, de forma que $F > 1$, o que caracteriza o teste unilateral.

As hipóteses que testamos quando aplicamos o teste F, são as seguintes:

a) $H_0: \sigma_1^2 = \sigma_2^2$
ou seja, a hipótese da nulidade estabelece que as duas populações possuem variâncias iguais, o que equivale a admitir que as duas amostras foram retiradas de uma mesma população;

b) $H_1: \sigma_1^2 > \sigma_2^2$ (teste unilateral)
ou seja, a hipótese alternativa estabelece que a população 1 possui variância superior à da população 2 ou, equivalentemente, que as amostras são provenientes de populações diferentes.

Os valores críticos de F são tabelados em função dos números de graus de liberdade, n_1 e n_2 , das estimativas de variâncias, s_1^2 e s_2^2 , a diferentes níveis de probabilidade. O valor assim calculado de F deve ser comparado àquele tabelado segundo os graus de liberdade da fonte testada e do resíduo, respectivamente nas colunas e linhas daquela tabela.

Quando a fonte sendo testada não for uma interação e se referir apenas um grau de liberdade, o teste F é adequadamente aplicado e equivale ao teste t de Student ($t^2=F$).

Suponhamos, por exemplo, a seguinte análise da variância de um experimento com 4 tratamentos e 6 repetições (Quadro 1).

Quadro 1 – Análise de variância

CV	GL	S.Q.	Q.M.	F	D.P
Tratamentos	3	32,64	10,88	7,56	3,30
Resíduo	20	28,80	1,44		1,20
Total	23	61,44			

Admitindo a hipótese de nulidade, isto é, supondo-se que os tratamentos sejam todos equivalentes, o quadrado médio (Q.M.) para os tratamentos é uma estimativa da va-

riância σ^2 , da mesma forma que o quadrado médio referente ao resíduo. Sendo estimativas diferentes do mesmo parâmetro, elas não deveriam diferir a não ser por acaso. Para compará-las é que usamos o teste F.

As tabelas de F dão o valor de 3,10 para o nível de 5% de probabilidade e 4,94 para o nível de 1% com 3 graus de liberdade para n_1 e 20 graus para n_2 . Quer isto dizer que há uma probabilidade de 95% de obter, por simples acaso, um valor de F igual ou inferior de 3,10 e há probabilidade de 5% de obter os valores de F superiores a 3,10. Analogamente, é de 1% a probabilidade que o valor de F exceda 4,94 e é de 99% a probabilidade de que o F não exceda 4,94. Nesse caso, o valor obtido excede o limite de 1% e se diz significativo ao nível de 1%. Isto quer dizer que há uma probabilidade inferior de 1% de que o valor de F observado tenha ocorrido por acaso.

No caso se não houver efeito de tratamentos, os dois quadrados médios (Q.M.Trat. e Q.M.Res) estimam a mesma variância e, portanto, qualquer diferença em ordem de grandeza que ocorrer entre eles deverá ser puramente casual.

Todas as considerações feitas para tratamentos, valem para qualquer outra causa de variação devida a fatores controlados, como blocos. Assim, o teste F para blocos é aplicado, calculando:

$$F \text{ Blocos} = \text{Q.M.Blocos} / \text{Q.M.Res.}$$

O valor de F da tabela para blocos é encontrado em função do número de graus de liberdade de blocos e do número de graus de liberdade do resíduo.

TESTES DE COMPARAÇÕES MÚLTIPLAS

Os testes de comparações múltiplas, ou testes de comparações de médias, servem como um complemento do teste de F, para detectar diferenças entre os tratamentos e que

requer alguns procedimentos:

a) Contrastes de médias

Se tivermos uma função linear:

$$Y = f(x) = a_1x_1 + a_2x_2 + \dots + a_nx_n$$

E verificamos que $\sum_{i=1}^n a_i = a_1 + a_2 + \dots + a_n = 0$,

dizemos que Y constitui um contraste na variáveis x. Então, se: $Y = x_1 + x_2 - x_3 - x_4$, Y é um contraste, visto que: $a_1=1, a_2=1, a_3=-1$ e $a_4=-1$ e, portanto: $a_1 + a_2 + a_3 + a_4 = 1 + 1 + (-1) + (-1) = 0$

Se, em lugar das variáveis x, tivermos médias, obteremos um contraste de médias. Assim, se num experimento temos 05 tratamentos, cujas médias verdadeiras são: m_1, m_2, m_3, m_4 e m_5 , as relações:

$$Y_1 = m_1 + m_2 + m_3 + m_4 - m_5$$

$$Y_2 = m_1 + m_2 - m_3 - m_4$$

$$Y_3 = m_1 - m_2$$

$$Y_4 = m_1 - m_3, \text{ constituem contrastes de médias.}$$

É claro que com um grupo de médias podemos formar um número muito grande de contrastes. Numa análise estatística devemos formular aqueles que sejam de maior interesse para o pesquisador.

De modo geral, não conhecemos as médias verdadeiras, de forma que o verdadeiro valor do contraste também nos é desconhecido. Conhecemos as estimativas das médias, podemos calcular as estimativas dos contrastes. Assim, a fórmula será:

$$\hat{Y} = c_1m_1 + c_2m_2 + \dots + c_nm_n$$

$$\sum_{i=1}^n c_i = 0,$$

obtemos a estimativa:

$$\hat{Y} = c_1m_1 + c_2m_2 + \dots + c_nm_n$$

b) Covariância de dois contrastes

Consideremos as duas estimativas de contrastes:

$$\hat{Y}_1 = a_1m_1 + a_2m_2 + \dots + a_nm_n$$

$$\hat{Y}_2 = b_1m_1 + b_2m_2 + \dots + b_nm_n$$

nas quais as médias estimadas foram calculadas com r_1, r_2, \dots, r_n repetições, respectivamente.

A estimativa de covariância entre essas duas estimativas de contrastes é definida por:

$$\text{Cov}(\hat{Y}_1\hat{Y}_2) = a_1b_1V(m_1) + a_2b_2V(m_2) + \dots + a_nb_nV(m_n)$$

$$\text{Lembramos que } V(m_i) = \frac{s_i^2}{r_i} \text{ (i=1,2,\dots,n)}$$

temos:

$$\text{Cov}(\hat{Y}_1\hat{Y}_2) = a_1b_1\frac{s_1^2}{r_1} + a_2b_2\frac{s_2^2}{r_2} + \dots + a_nb_n\frac{s_n^2}{r_n}$$

Frequentemente, temos $s_1^2 = s_2^2 = \dots = s_n^2 = s^2$,

logo:

$$\text{Cov}(\hat{Y}_1\hat{Y}_2) = \left[\frac{a_1b_1}{r_1} + \frac{a_2b_2}{r_2} + \dots + \frac{a_nb_n}{r_n} \right] s^2$$

Nas análises de variância de delineamento balanceados, todas as médias possuem o mesmo número de repetições, r, e portanto:

$$\text{Cov}(\hat{Y}_1\hat{Y}_2) = (a_1b_1 + a_2b_2 + \dots + a_nb_nV(m_n)) \frac{S^2}{r}$$

c) Contrastes ortogonais

Na prática, a ortogonalidade entre dois contrastes indica uma independência entre suas comparações, ou seja, a variação de um contraste é inteiramente independente da variação do outro.

A condição necessária e suficiente para que dois contrastes sejam ortogonais entre si é que a covariância entre eles seja nula, isto é:

$$\frac{a_1b_1}{r_1} + \frac{a_2b_2}{r_2} + \dots + \frac{a_nb_n}{r_n} = 0$$

Se tivermos $s_1^2 = s_2^2 = \dots = s_n^2 = s^2$, a condição de ortogonalidade fica:

$$\frac{a_1b_1}{r_1} + \frac{a_2b_2}{r_2} + \dots + \frac{a_nb_n}{r_n} = 0 \text{ ou } \sum_{i=1}^n \frac{a_ib_i}{r_i} = 0$$

Se, além das variâncias, os números de repetições também forem iguais, a condição de ortogonalidade fica:

$$a_1b_1 + a_2b_2 + \dots + a_nb_n + 0 \text{ ou } \sum_{i=1}^n a_i b_i = 0$$

Três ou mais contrastes serão ortogonais entre si se eles forem ortogonais dois a dois. Num experimento com **I** tratamentos, podemos formular vários grupos de contrastes ortogonais entre si, porém cada grupo terá apenas (I-1) contrastes.

Se existirem cinco tratamentos de natureza qualitativa, apenas 4 gl poderão ser decompostos ortogonalmente sem confundimentos, de modo que as somas de quadrados individuais somadas correspondam a $SQ_{\text{TRATAMENTO}}$ com 4 gl. Os contrastes C_i para serem ortogonais precisam atender às seguintes exigências matemáticas:

$$\sum_{i=1}^t C_i = 0 \text{ e } \sum_{i \neq j} C_i C_j = 0$$

Então, para os cinco tratamentos de A a E do tratamento hipotético supramencionado podemos criar os seguintes (5-1=4) contrastes ortogonais, cada um exigindo 1 gl para seu estudo (Quadro 2).

Quadro 2 – Contrastes ortogonais

Tratamentos	Total de 4				
	Tratamentos	Contrastes			
	T_i	C_1	C_2	C_3	C_4
A	152	-1	0	-1	-1
B	164	1	0	-1	-1
C	100	0	1	1	-1
D	96	0	-1	1	-1
E	132	0	0	0	4
Total	644	0	0	0	0

Outras tabelas como esta poderia ser formulada, com base no interesse específico em determinadas comparações. Tanto a soma por coluna (C_i) como a soma de qualquer produto ($C_i C_j$) é nula, portanto os contrastes são ortogonais.

O contraste C_1 indica que o pesquisador deseja comparar os tratamentos A e B cuja soma de quadrados, como de qualquer, contraste será a por:

$$SQ_{C_1} = \frac{[\sum_{i=1}^t T_i C_{ij}]^2}{r \sum_{i=1}^t C_{ij}^2}$$

onde:

t é o número de tratamentos

T_i é o total das r repetições do tratamento i

C_{ij} é o coeficiente do tratamento i no contraste j

Logo, a comparação entre A e B seria julgada pela SQ_{C_1} :

$$SQ_{C_1} = \frac{[152(-1) + 164(1) + 100(0) + 96(0) + 132(0)]^2}{4[(-1)^2 + 1^2 + 0^2 + 0^2]} = \frac{(-12)^2}{8} = 18$$

O contraste C_2 também compara efetivamente os tratamentos C e D, e sua contribuição tem termos de SQ é:

$$SQ_{C_2} = \frac{[152(0) + 164(0) + 100(1) + 96(-1) + 132(0)]^2}{4[0^2 + 0^2 + 1^2 + (-1)^2 + 0^2]} = \frac{(4)^2}{8} = 2$$

Já o contraste C_3 compara a média de A + B com a de C+D, e a conclusão da significância ou não deste contraste, não esclarece o que está ocorrendo entre A e C por exemplo, ou A e D, B e C ou B e D, como o pesquisador gostaria de saber. Este fato desabona a interpretação de um contraste que envolva mais de dois tratamentos, como é o caso. De qualquer maneira, a contribuição desse contraste seria:

$$SQ_{C_3} = \frac{[152(-1) + 164(-1) + 100(1) + 96(1) + 132(0)]^2}{4[(-1)^2 + (-1)^2 + 1^2 + 1^2]} = \frac{(120)^2}{8} = 900$$

Finalmente o contraste C_4 , com 1 gl como todos anteriores, estaria comparando a média do tratamento E com a média dos demais:

$$SQ_{C_4} = \frac{[152(-1) + 164(-1) + 100(-1) + 96(-1) + 132(4)]^2}{4[(-1)^2 + (-1)^2 + (-1)^2 + (-1)^2 + 4^2]} = \frac{(-16)^2}{80} = 3,2$$

Esta decomposição da $SQ_{\text{TRATAMENTOS}}$ de 4 gl, foi ortogonal porque a $SQ_{\text{TRATAMENTOS}}$

equivale à soma das SQ de cada um dos quatro contrastes investigados:

$$SQ_{\text{TRATAMENTOS}} = \frac{152^2 + 164^2 + 100^2 + 96^2 + 132^2}{4} = 923,2$$

$$SQ_{C_1} + SQ_{C_2} + SQ_{C_3} + SQ_{C_4} = 18 + 2 + 900 + 3,2 = 923,2$$

Portanto, se existirem mais de dois tratamentos, o teste de F fornecerá uma informação efetiva quando o contraste utilizado envolver apenas dois grupos experimentais.

Uma fonte de variação com mais de 1 gl pode ser testada pela razão de variâncias F, entretanto o que estará sendo julgado é a variação média desta fonte. O resultado do teste F avaliará, portanto a magnitude dessa variação média mas não deixará entrever se existe algum contraste que poderia ser significativo (Quadro 3). Em um exemplo hipotético onde a análise de variância fosse, seria:

Quadro 3 – Análise de variância

FV	gl	SQ	QM	F
Tratamento	4	932,2	233,05	3,00
Erro	15	1165,25	77,68	
Total	19	2097,45		

Note que os 4 gl de tratamentos não foram decompostos ainda, fazendo com que o quadrado médio de tratamentos seja avaliado como $932,2 / 4 = 233,05$.

A razão de variâncias de $233,05 / 77,68 = 3,00$ com 4 gl e 15 gl respectivamente é inferior à sua correspondente na Tabela de F onde o valor é 3,06. Logo poderíamos concluir que a variação média observada entre os tratamentos não foi significativa. Entretanto, não poderíamos dizer que não ocorreu diferenças significativas entre os cinco tratamentos. Para chegarmos a esta conclusão seria preciso fazer um teste específico que permitisse aquelas comparações, ou então, menos eficientemente, fazer uma decomposição ortogonal da $SQ_{\text{TRATAMENTO}}$ (Quadro 4).

Quadro 4 – Decomposição ortogonal

FV	gl	SQ	QM	F
Contraste 1	1	18,00	18,00	n.s
Contraste 2	1	2,00	2,00	n.s
Contraste 3	1	900,00	900,00	11,58*
Contraste 4	1	3,20	3,20	n.s
Erro	15	1165,25	77,68	
Total	19	2097,45		

* Significativo ($p < 0,05$)

Podemos então concluir que pelo menos entre os grupos A e B, como também entre C e D não houve diferença significativa (contrastos C_1 e C_2), nada se podendo dizer para as demais comparações entre dois grupos.

Como o teste F para o contraste 3 foi significativo (valor tabelado de $F_{1e15} = 4,54$), percebemos que quando utilizamos o teste F para mais de um grau de liberdade, algumas comparações significativas poderão passar despercebidas.

Um outro teste específico, como t de Student, através do cálculo da diferença mínima significativa seria de ação mais definitiva:

$$Dms = t_{15 \text{ gl}} \sqrt{2s_{2e}/r} = 2,131 \sqrt{\frac{2 \times 77,68}{4}} = 13,2$$

Podemos verificar que, embora o teste F não tenha sido significativo quando avaliou a variação média dos tratamentos (4gl), ele teria acusado a significância de contrastes não executados, mas correspondendo as seguintes comparações: BxC, BxD e AxD, que foram percebidas pelo teste t de Student, equivalente ao F para fontes simples de 1 gl (Quadro 5).

Quadro 5 - Médias segundo o tratamento e comparações

Tratamento	Médias	Comparação
B	41	a
A	38	ab
E	33	abc
C	25	bc
D	24	c

Nos modelos matemáticos lineares com mais de uma variável independente, cada um desses elementos no modelo assume apenas 1 gl, razão pela qual caberia a utilização do teste F. Cada variável independente no modelo seria um contraste, com 1 gl. Esses contrastes poderiam ser de origem qualitativa como o exemplo anterior, mas também podem ser de origem quantitativa. Isto ocorre quando os tratamentos estudados são níveis equidistantes e variáveis de um mesmo fator. Nesse caso, embora fosse possível sugerir qualquer grupo de contrastes ortogonais, o mais adequado seria investigar o tipo de função observada através dos coeficientes dos polinômios ortogonais. Assim sendo, o primeiro contraste seria aquele que investigaria se o efeito observado é linear. O segundo verificaria se o efeito é quadrático e o terceiro, se o efeito é cúbico e assim por diante, dependendo do número de níveis estudados.

Teste de Student

O teste **t** é um teste clássico e com uso mais regular. É recomendado para testar contrastes envolvendo 2 ou mais médias, devendo as comparações serem planejadas antes do acesso aos dados, em número não superior aos graus de liberdade de tratamentos e os contrastes devem ser ortogonais.

Seja o contraste:

$$\hat{Y} = c_1 \tilde{Y}_1 + c_2 \tilde{Y}_2 + \dots + c_t \tilde{Y}_t$$

Onde \tilde{Y}_1 foi determinada com r_1 repetições, \tilde{Y}_2 com r_2 repetições e \tilde{Y}_t com r_t repetições.

Uma estimativa da variância dos contrastes é:

$$V(\hat{Y}) = \left[\begin{array}{c} c_1^2 \\ \sum_i s^2 \\ \quad r_i \end{array} \right], s^2 \text{ QME}$$

A diferença mínima significativa (DMS) é dada por:

$$DMS = T = t_{\alpha/2} (v) \sqrt{V(\hat{Y})}$$

ou $DMS = t \sqrt{(2 \times QMR) / r}$

Onde **t** é obtido na tabela **t** num nível α de probabilidade com **v** graus de liberdade do erro experimental ($v = n - t$, DIC).

Se $|\hat{Y}| > t$ conclui-se que o contraste é significativo. Se as médias forem determinadas com o mesmo número de repetições, substitua r_1 por r e v por t ($r-1$) para DIC.

Considere dados de um experimento hipotético para estudar a implantação de hormônios. As médias experimentais foram: $m_1 = 2,55$, $m_2 = 3,09$, $m_3 = 2,68$ e $m_4 = 2,96$.

Temos, nesse caso:

- controle versus tratamentos experimentais, o contraste é:

$$\hat{Y}_1 = -3m_1 + m_2 + m_3 + m_4$$

e

$$V(\hat{Y}_1) = (1/24) \times 12 \times 0,291 = 0,1455$$

A diferença mínima significativa é:

$$DMS_1 = T_1 = 1,99 \sqrt{0,1455} = 0,76$$

O contraste é significativo, pois $\hat{Y}_1 = 1,08$ é maior que 0,76.

- Estilbestriol vs Estradiol

O contraste é:

$$\hat{Y}_2 = 2\hat{Y}_2 - \hat{Y}_3 - \hat{Y}_4$$

e

$$V(\hat{Y}_2) = (1/24) \times 6 \times 0,291 = 0,07275$$

A diferença mínima significativa é:

$$DMS_2 = T_2 = 1,99 \sqrt{0,07275} = 0,53$$

O contraste $\hat{Y}_2 = 0,54$ é significativo.

- 10 mg vs 20 mg Estradiol

O contraste é:

$$\hat{Y}_3 = \hat{Y}_3 - \hat{Y}_4$$

e:

$$V(\hat{Y}_3) = (1/24) \times 2 \times 0,291 = 0,02425$$

A diferença mínima significativa é:

$$DMS_3 = T_3 = 1,99 \sqrt{0,02425} = 0,309$$

O contraste $\hat{Y}_3 = -0,28$ é não significativo.

O intervalo de confiança de $(1-\alpha)100\%$ para Y_1 é dado por:

$$P\left[\hat{Y}_1 - \overset{T}{t_{\alpha/2}(v)} \times s(\hat{Y}_1) \leq Y_1 \leq \hat{Y}_1 + \overset{T}{t_{\alpha/2}(v)} \times s(\hat{Y}_1)\right] = 1-\alpha$$

O IC de 95% para Y_3 é;

$$P(-0,28 - 1,986 \times 0,1557 \leq \hat{Y}_3 \leq 0,28 + 1,986 \times 0,1557) = 95\%$$

$$P(-0,589 \leq Y_3 \leq 0,029) = 95\%$$

Portanto, com 95% de probabilidade, a precisão de \hat{Y}_3 é:

$$-0,28 \pm 0,309$$

Obs.: Feita a escolha dos contrastes a testar antes de examinar os dados, nada há a objetar contra o uso do teste *t*. Mas, no entanto, talvez não seria muito coerente por exemplo, comparar por esse processo a maior média com a menor delas, ou as duas maiores com a menor. Isto pode ser feito, porém, por outros métodos de testes de médias, que assim reduziria a tendência de forçar o erro Tipo I experimental.

No caso de utilizar o dms, para obter a diferença mínima significativa estabelecida pelo teste *t*, basta calcular pela fórmula:

$$d.m.s = t \sqrt{\frac{2.QMR}{r}}$$

onde *t* é um valor dado em tabela, QMR é o quadrado médio do resíduo da análise de variância e *r* é o número de repetições de cada tratamento. Toda vez que o valor absoluto da diferença entre duas médias é igual ou maior do que o valor da d.m.s., as médias são estatisticamente diferentes. É fácil comparar médias usando a d.m.s. pelo teste de *t*.

Teste de Tukey

Tukey (1953) considerou muito trabalhoso o procedimento proposto por Newman (1939) mas compactuava com a preocupação no controle do erro tipo I. A opção proposta de apenas um valor de diferença mínima significativa, a despeito da existência de várias médias, caracterizou o teste como extremamente rigoroso, que embora controlasse muito bem o erro tipo I, permitia o aparecimento do erro tipo II. O valor único proposto por Tukey coincide com o valor máximo do SNK, ou seja, equivalente à comparação entre a maior e a menor média.

Um teste assim muito rigoroso aplicado em um ensaio com muitos tratamentos (*q* aumenta com o número de tratamentos) e envolvendo uma variável muito instável ($CV > 25\%$) favorecerá sobremaneira o aparecimento do erro tipo II. No entanto, esse teste de média (Tukey) pode ser utilizado para comparar todo e qualquer contraste entre médias 2 a 2. As comparações 2 a 2 são indicadas para as situações em que os tratamentos são qualitativos e não existe uma estrutura entre eles que permita a formação de contrastes mutuamente ortogonais. O teste é exato e de uso muito simples quando o número de repetições é o mesmo para todos os tratamentos.

Se as médias forem determinadas com o mesmo número de repetições, a diferença mínima significativa (DMS) é dada por:

$$DMS = \Delta = q_{\alpha} \sqrt{\frac{QMR}{r}}$$

onde:

$q_{\alpha}(t,v)$ = valor da amplitude total estudentizada, num nível α de probabilidade, obtido da tabela de Tukey, em função do número de tratamentos (*t*) e do número de graus de liberdade do erro (*v*).

s= desvio padrão residual

r= número de repetições

Se as médias forem determinadas com diferentes números de repetições, a diferença mínima significativa é dada por:

$$DMS' = q_{\alpha}(t, v) \sqrt{\frac{1}{2} V(\hat{Y})}$$

$$V(\hat{Y}) = \frac{1}{r_i} + \frac{1}{r_j} s^2, i \neq j, s^2 = QME$$

Se $|\hat{Y}| > DMS$ ou $|\hat{Y}| > DMS'$, conclui-se que o contraste é significativo.

Considere o experimento para estudar a mobilidade de proteína no soro de animais de uma determinada espécie. Os tratamentos eram constituídos de 5 raças e foram usados 10 animais por raça. Os valores médios foram:

$$m_1 = 2,54; m_2 = 2,89; m_3 = 2,48; m_4 = 2,80 \text{ e } m_5 = 2,77 - QME = 0,0395$$

A diferença mínima significativa é dada por:

$$q_{0,05}(5;45) \times \frac{s}{\sqrt{r}} = 4,20 \times \sqrt{\frac{0,0395}{10}} = 0,25$$

As médias ordenadas e os contrastes entre 2 médias são:

$$m_2 = 2,89 \quad m_4 = 2,80 \quad m_5 = 2,77 \quad m_1 = 2,54 \quad m_3 = 2,48$$

$$\begin{aligned} \hat{Y}_1 &= \hat{Y}_2 - \hat{Y}_4 = 0,09 & \hat{Y}_6 &= \hat{Y}_4 - \hat{Y}_1 = 0,26 \\ \hat{Y}_2 &= \hat{Y}_2 - \hat{Y}_5 = 0,12 & \hat{Y}_7 &= \hat{Y}_4 - \hat{Y}_3 = 0,32 \\ \hat{Y}_3 &= \hat{Y}_2 - \hat{Y}_1 = 0,35 & \hat{Y}_8 &= \hat{Y}_5 - \hat{Y}_1 = 0,23 \\ \hat{Y}_4 &= \hat{Y}_2 - \hat{Y}_3 = 0,41 & \hat{Y}_9 &= \hat{Y}_5 - \hat{Y}_3 = 0,29 \\ \hat{Y}_5 &= \hat{Y}_2 - \hat{Y}_5 = 0,03 & \hat{Y}_{10} &= \hat{Y}_1 - \hat{Y}_3 = 0,06 \end{aligned}$$

Os resultados estão apresentados no seguinte quadro 6:

Quadro 6 - Valores médios da mobilidade de proteína em 5 raças de certa espécie animal

Raças	Médias ¹
A	2,54 bc
B	2,89 a
C	2,48 c
D	2,80 a
E	2,77 ab

¹ médias seguidas de mesma letra são estatisticamente iguais pelo teste de Tukey (P<0,05)

O teste de Tukey pode ser adaptado também ao julgamento de contrastes mais complicados, mas não é conveniente optar para utilizá-lo devido o fato de forçar para o erro tipo I, assim, optar-se-á pelo teste de Scheffé que parece ser em geral, o mais conveniente.

É interessante que, muito raramente, pode acontecer que, embora o teste F não tenha sido significativo na análise de variância, obtenha-se um ou mais contrastes significativos pelo teste de Tukey. Ou pode ocorrer o contrário, em que o teste F seja significativo, sem que nenhuma diferença ente médias seja significativa pelo teste de Tukey.

Fatos semelhantes ocorrem com o teste de Duncan, que não concorda inteiramente com o teste F. Tais discordâncias se devem à aceitação de hipóteses diferentes nas deduções teóricas, e são, aliás, de pouca importância prática.

Teste de Student-Newman-Keus (SNK)

Newman (1939) apresentou um teste que contornava os inconvenientes do teste t para ensaios com mais de dois tratamentos. Ajustava o valor de t dependendo da distância entre as médias então ordenadas.

É próprio para comparação entre duas médias. É menos rigoroso que Tukey e pode evidenciar diferenças não discriminadas por ele. A técnica da aplicação do teste exige que as médias estejam ordenadas e são calculadas tantas DMS's quantos são o número de

médias ordenadas abrangidas pelo contraste em estudo.

A condição de balanceamento ($r_1 = r_2 = \dots = r_g = r$) é exigida, em princípio, e a estatística q é considerada com parâmetros g' e f , onde g' é o número de médias ordenadas abrangidas pelo contraste em estudo. Assim sendo, o valor de g' varia durante a aplicação do teste.

Os passos para o teste de Newman-Keuls são:

- a) ordena-se as médias (ordem crescente ou decrescente);
- b) compara-se a maior com a menor média ($g' = g$). Em uma relação decrescente de t médias, duas delas (A e F) apresentarão a diferença mínima significativa se:

$$\frac{A - B}{\sqrt{QMR/r}} \geq q(i, g', f)$$

Onde QMR é o valor do quadrado médio do resíduo, pela análise de variância, r é o número de repetições, comum a todos os tratamentos e q_i é o valor tabelado proposto (t ajustado) obtido em função da distância entre as médias ($i = p + 2$, p sendo o número de médias existentes entre as duas médias comparadas, na relação decrescente) e os graus de liberdade do resíduo, QMR.

Os valores de q_i diminuem com o aumento dos gl mas aumentam com a distância entre as médias, corrigindo os excessos de erro tipo I.

A diferença mínima significativa entre duas médias com distância i entre elas e com igual número de repetições é:

$$dms (SNK) = q_i \sqrt{\frac{QMR}{r}}$$

Quando as médias comparadas A e B apresentarem diferente número de repetições, a diferença mínima significativa seria:

$$dms (SNK) = q_i \sqrt{\frac{QMR}{2} \left(\frac{1}{r_A} + \frac{1}{r_B} \right)}$$

Se o valor da d.m.s for maior do que o módulo da estimativa do contraste, as médias são ligadas por um traço, indicando que não há diferença significativa entre elas; caso contrário, determina-se o seguinte:

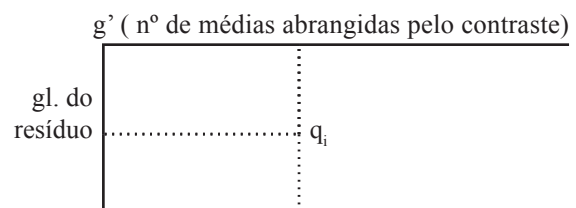
- c) reduz-se de uma unidade o valor de g' . Calcula-se o novo valor da d.m.s e, para todos os pares de médias que não estejam ligados por um mesmo traço e que envolvem g' médias, repete-se o mesmo procedimento dos itens (a e b).
- d) repete-se o procedimento anterior c, até que $g' = 1$.

Este teste tem como inconveniente, além de ser um teste trabalhoso, o fato das médias ordenadas não serem independentes e o valor de $q(i, g', f)$ em consequência, não ser exato.

Quando num conjunto de médias ordenadas a comparar, a maior não difere significativamente da menor pelo teste SNK, não se admite diferença significativa, pelo teste, entre médias intermediárias.

O teste de Newman-Keuls é um procedimento sequencial, válido para a totalidade dos contrastes de 2 médias. Esse teste exige, em princípio, balanceamento dos dados e além do mais, é um teste aproximado. Assim, torna-se um teste menos rigoroso nas comparações de médias.

Para determinar a d.m.s. é necessário conhecer o valor de q_i na tabela ao nível de α de probabilidade em função do número de médias abrangidas pelo contraste (g') e os graus de liberdade do resíduo (f).



Em um experimento onde os dados referem-se aos pesos totais, em Kg, das parcelas no final de um experimento com frangos de corte, onde foram comparados quatro

rações. O delineamento foi DIC com seis repetições e cada parcela foi um lote de 12 frangos (Quadro 7).

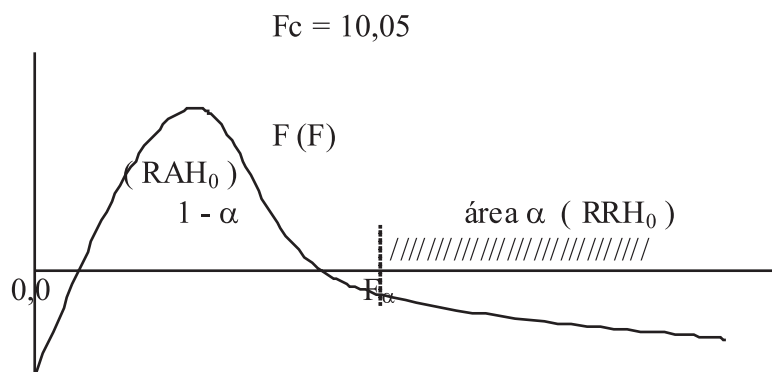
Quadro 7 – Delineamento inteiramente ao acaso

Rações	Repetições						Total (Y_i)	Média (\bar{Y}_i)
	1	2	3	4	5	6		
A	16,1	15,4	15,2	14,8	17,9	16,6	96,0	16,0
B	17,4	18,5	17,6	18,1	18,6	17,8	108,0	18,0
C	16,3	18,1	18,4	16,9	17,2	18,1	105,0	17,5
D	20,0	18,0	18,4	22,0	18,5	18,9	115,8	19,3
Total							424,8	17,7

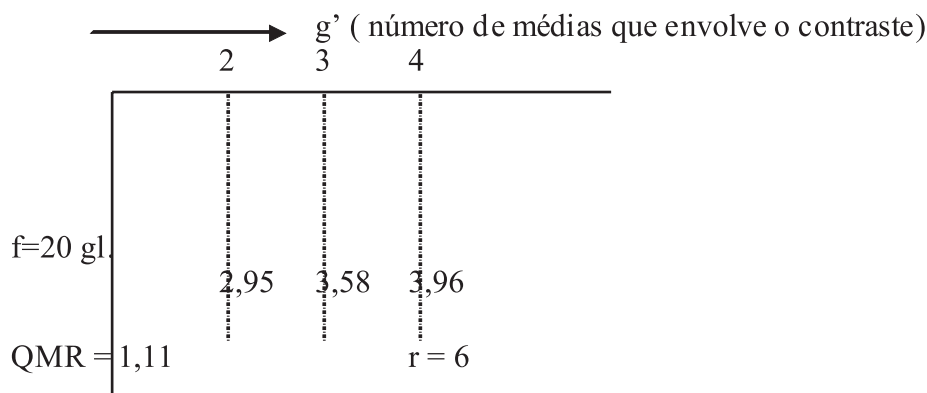
Quadro de ANOVA

FV	gl	SQ	QM	F
Tratamento	3	33,48	11,16	10,05
Resíduo	20	22,10	1,11	
Total	23	55,58		

Na tabela de Tukey ($\alpha = 5\%$)



RAH₀ = região de aceitação de H₀
RRH₀ = região de rejeição de H₀

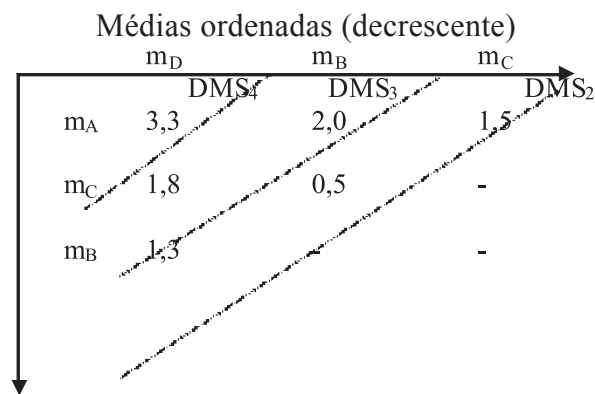


$DMS_4 = 3,96\sqrt{1,11/6} = 1,70$ este valor vai ser usado para contrastes que envolve 4 médias;

$DMS_3 = 3,58\sqrt{1,11/6} = 1,54$ este valor vai ser usado para contrastes que envolve 3 médias;

$DMS_2 = 2,95\sqrt{1,11/6} = 1,27$ este valor vai ser usado para contraste que envolve 2 médias.

Obs.: qualquer diferença das médias, envolvendo as 4 médias que seja \leq ao valor de 1,70 é estatisticamente nula ou seja, zero.



Quadro de médias do peso médio em Kg dos lotes de frangos em cada tratamento ao final do ensaio (Quadro 8).

Quadro 8 – Relação de médias de um delineamento em DIC

Tratamento	Médias ¹
A	16,0 c
B	18,0 b
C	17,5 b
D	19,3 a

¹ médias seguidas de mesma letra são estatisticamente iguais pelo teste de Tukey (P<0,05)

A ração “D” deveria ser recomendada, porque é a que tem um melhor conceito.

Teste de Scheffé

Não é considerada uma boa técnica deixar os resultados experimentais sugerir as hipóteses que deverão ser testadas. Contudo, em algumas situações práticas é difícil evitar

este inconveniente. O teste desenvolvido por Scheffé permite que os contrastes sejam formados após a inspeção de resultados.

A flexibilidade proposta por Scheffé (1954) para comparar qualquer contraste entre médias e permitindo diferentes números de observações por tratamento definiu um teste um pouco mais rigoroso que aquele de Tukey, merecendo portanto os mesmos com relação ao perigoso aumento do erro tipo II.

É um teste mais rigoroso que o teste t, porém é mais flexível que ele, tendo em vista a não exigência de ortogonalidade e nem que os contrastes sejam estabelecidos antes de se examinar os dados. Este só deve se aplicado quando o teste F da análise de variância for significativo. Se o valor de F obtido for não significativo, nenhum contraste poderá se significativo e, assim sendo, a aplicação do teste de Scheffé não se justifica. Por outro lado, quando F for significativo, pelo menos um dos contrastes entre médias de tratamentos será significativo. Pode acontecer que o contraste em questão é muito complicado ou sem interesse prático.

O teste de Scheffé é de uso bem mais geral que os de Tukey, SNK, e outros e permite julgar qualquer contraste. É o teste mais indicado para contrastes com mais de duas médias, não sendo, no entanto, indicado para contrastes de duas médias.

Se verificarmos que $|\hat{Y}| \geq d.m.s._s$, dizemos que o contraste é significativo ao nível de α de probabilidade, indicando que os grupos de médias confrontados no contraste diferem entre si a esse nível de probabilidade.

A diferença mínima significativa para qualquer contraste (que pode ser a comparação entre apenas duas médias) é:

$$d.m.s._{Scheffé} = \sqrt{(t - 1).F.Var(contraste)}$$

onde t é o número de tratamentos, F é o valor tabelado de F com (t-1) e x graus de liberdade

de, sendo x os graus de liberdade do resíduo (relativo a (QMR)).

Como a variância de um contraste é definida por:

$$\text{Var}_{(\text{contraste})} = \text{QMR} \sum_{i=1}^t C_i^2 / r_i$$

onde C_i é o coeficiente do tratamento i com r_i repetições. Para a comparação de duas médias oriundas de um igual número de repetições, tem-se:

$$\text{d.m.s.}_{\text{Scheffé}} = \sqrt{(t-1) \cdot F \cdot 2\text{QMR} / r}$$

ou
$$\text{d.m.s.}_{\text{Scheffé}} = \sqrt{(t-1) V(Y) F}$$

$$V(Y) = \text{QMR} \cdot \left[\frac{c_1^2}{r_1} + \frac{c_2^2}{r_2} + \dots + \frac{c_n}{r_n} \right]$$

Suponhamos, por exemplo, a seguinte análise da variância de um experimento com 4 tratamentos e 6 repetições (Quadro 9).

Quadro 9 – Análise de variância de uma delimitação em DIC

CV	GL	S.Q.	Q.M.	F
Tratamentos	3	32,64	10,88	7,56
Resíduo	20	28,80	1,44	
Total	23	61,44		

Onde : $m_1 = 26,0$; $m_2 = 24,8$;
 $m_3 = 22,8$ $m_4 = 24,0$

Se considerar o contraste:

$$Y = 3m_1 - m_2 - m_3 - m_4 (3 -1 -1 -1)$$

$\hat{Y} = 3(26,0) - 24,8 - 22,8 - 24,0 \therefore = 6,4$, ao nível de 5% de probabilidade, sendo o valor de F com 20 gl. do erro = 3,10;

Assim,

$$\text{Dms}_{\text{SCHEFFÉ}} = 1,20 \sqrt{3 \left[\frac{9}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} \right]} 3,10 = 5,17$$

A estimativa obtida (6,4) excede o valor de d.m.s._{Scheffé} calculado, logo o contraste será significativo ao nível de 5% de probabilidade. Já para o nível de 1% não é significativo, pois o valor do contraste de média supramencionado (6,4) é menor que o d.m.s._{Scheffé} ao nível de 1% (6,53).

Teste de Duncan

Em 1955, Duncan sugeriu a utilização de um teste baseando-se na mesma argumentação do teste de Student-Newman-Keuls. A percepção e o controle do erro tipo I era sem dúvida flagrante no SNK, mas a comparação de médias mais afastadas criava uma oportunidade para o aparecimento do erro tipo II (já que o maior valor da diferença mínima significativa se equiparava com aquela do teste de Tukey). Sua aplicação é bem mais trabalhosa do que o teste de Tukey, mas se chega a resultados mais detalhados e se discrimina com mais facilidade entre os tratamentos, isto é, o teste de Duncan indica resultados significativos em casos em que o teste de Tukey não permite obter significação estatística. O teste exige que as médias sejam colocadas em ordem decrescente e que todas elas possuam o mesmo número de repetições, para ser exato.

Para uso do teste necessitamos de tabelas especiais para o nível de 5% de probabilidade, e outra para o de 1%. Entretanto, geralmente o teste é aplicado ao nível de 5% de probabilidade e a significância do teste é indicada ligando-se por uma barra as médias que não diferem entre si, embora atualmente se encontrem trabalhos que indicam sua significância utilizando o sistema de letras diferentes.

Cada contraste testado envolve apenas 2 (duas) médias, embora a amplitude do

contraste possa abranger um número maior de médias.

Utilizando o exemplo do teste anterior, começamos por calcular uma amplitude total mínima significativa pelo teste de Duncan para o contraste entre a maior e a menor das médias. No caso de um mesmo número de repetições, utiliza-se a d.m.s._{DUNCAN}:

$$d.m.s._{DUNCAN} = Z \sqrt{\frac{QMR}{r}}$$

onde z é tirado das tabelas para o número de médias ordenadas abrangidas pelo contraste em estudo e o número de graus de liberdade do resíduo, e r é o número de repetições. Os valores de z não sobem tão rapidamente quanto aqueles do teste SNK, controlando assim o aparecimento do erro tipo II na comparação de médias mais afastadas. No caso de um experimento que tenha comparado duas médias (A e B) e que contiverem diferente número de repetições, a d.m.s. será:

$$d.m.s._{DUNCAN} = Z \sqrt{\frac{QMR}{2} \cdot \left[\frac{1}{r_A} + \frac{1}{r_B} \right]}$$

Suponhamos, por exemplo, a seguinte análise da variância de um experimento com 4 tratamentos e 6 repetições (Quadro 10).

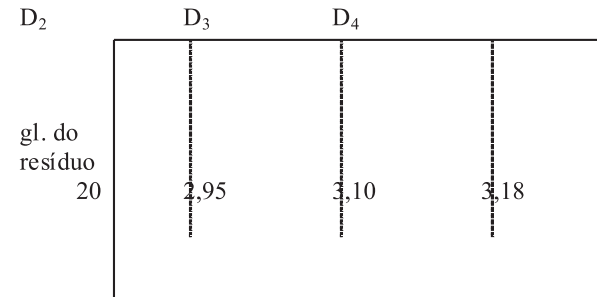
Quadro 10 – Análise de variância de uma delimitação em DIC

CV	GL	S.Q.	Q.M.	F
Tratamentos	3	32,64	10,88	7,56
Resíduo	20	28,80	1,44	
Total	23	61,44		

Onde: $m_1 = 26,0$; $m_2 = 24,8$; $m_3 = 22,8$ $m_4 = 24,0$

Na tabela de Z com 5%, os valores correspondentes as três médias calculadas, são:

Número de médias abrangidas pelo contraste



Logo, ao nível de 5% de probabilidade, temos:

$$D_4 = 3,18 \sqrt{(1,44 / 6)} = 1,56$$

A comparação entre a maior e a menor das 4 médias abrange todas as médias do ensaio. Para comparar a segunda colocada, logo abaixo da maior, coma menor de todas, teremos um contraste que abrange 3 médias.

$$D_4 = 3,10 \sqrt{(1,44 / 6)} = 1,52$$

Este valor servirá para testar também o contraste entre a maior média e a penúltima por ordem de grandeza, isto é, entre $m_1 - m_4$.

Finalmente, para as duas médias consecutivas temos:

$$D_4 = 2,95 \sqrt{(1,44 / 6)} = 1,45$$

e este resultado corresponde ao que nos daria o teste t.

Sempre que houver uma barra unindo duas médias, elas não diferem significativamente. Assim, m_1 não é significativamente diferente de m_2 , pois:

$m_1 - m_2 = 1,2$ e não excede 1,45. Mas m_1 é significativamente superior a m_4 , pois $m_1 - m_4 = 2,0$, que excede $D_3 = 1,53$. Assim, quando a barra e/ou a mesma barra aparece com as duas médias, a diferença entre elas não é significativa; quando não há nenhuma letra comum às duas médias, sabe-se que a diferença alcançou significação estatística.

Tanto o método de barras como o das letras hoje se usa comumente também para resumir o resultado da comparação de médias de tratamentos por outros critérios, tais como o teste de Tukey. Dessa forma, o resultado fica:

$m_1 = 26,0$	$m_1 = 26,0$ a
$m_2 = 24,8$	$m_2 = 24,8$ a c
$m_3 = 24,0$	$m_3 = 24,0$ b c
$m_4 = 22,8$	$m_4 = 22,8$ b

Quando o número de médias é grande (superior a 10) a aplicação do teste de Duncan se torna muito trabalhosa.

Teste de Dunnett

Este teste é utilizado quando as únicas comparações que interessam ao experimentador são aquelas feitas entre um determinado tratamento padrão, geralmente a testemunha, e cada um dos demais tratamentos, não havendo interesse na comparação dos demais tratamentos entre si. O objetivo maior é apontar os tratamentos que são melhores que um padrão ou controle. Assim, as comparações múltiplas onde apenas um tratamento serve de referência para os demais no sentido de comparar todos com apenas um.

A estatística para o teste é definida por:

$$t_D = \frac{m_i - m_c}{\left[\frac{1}{r_i} + \frac{1}{r_c} \right] \cdot \text{QMR}}$$

onde,

m_i = a média de tratamentos que queremos comparar com o controle ($i = 1, 2, \dots, m, m = t - 1$);

m_c = média do tratamento controle;

r_i = número de repetições de cada tratamento;

r_c = número de repetições do controle;

QMR = variância residual

Quando todos os tratamentos têm igual número de observações, a d.m.s é:

$$d.m.s._{\text{Dunnett}} = D \cdot \frac{2 \cdot \text{QMR}}{r}$$

que é semelhante ao teste **t** de Student, exceto pelo valor de **D**, aqui ajustado para um maior número de tratamentos.

A utilização deste teste, que impede outras comparações não contidas em suas condições iniciais é pouco frequente dado a esta limitação.

Assim, um experimento com *i* tratamentos (um dos quais a testemunha ou padrão, P) permite a aplicação do teste a *i* - 1 comparações. Para aplicação da técnica do teste Dunnett é necessário utilizar alguns procedimentos importantes:

- a) os valores tabelados $t_{D, \alpha, m, v}$ são válidos para comparar as médias de *m* tratamentos com a média do controle, quando as variâncias das médias são homogêneas,

$\sigma^2 m_i = \sigma^2 m_c$ ($1, 2, \dots, m$) e as repetições são balanceadas, $r_i = r_c$ ($1, 2, \dots, m$).

- b) calcular a estimativa de cada contraste:

$$\hat{Y}_1 = m_1 - m_p$$

$$\hat{Y}_2 = m_2 - m_p$$

$$\hat{Y}_{(i-1)} = m_{(i-1)} - m_p$$

- c) calcular a estimativa de variância da estimativa de cada controle

$$V(Y) = \left[\frac{1}{r_i} + \frac{1}{r_p} \right] s^2 \quad (i = 1, 2, \dots, I \text{ e } i \neq p)$$

- d) calcular o erro padrão do contraste

$$s(Y) = \sqrt{V(Y)}$$

- e) calcular o valor do teste - *d'* - dado por:

$$d' = t_d \cdot s(\hat{Y})$$

onde t_d é o valor dado na tabela para uso no teste de Dunnett (5% e 1%), em função do número de graus de liberdade de tratamento ($i - 1$) e do número de graus de liberdade do resíduo.

f) comparar cada estimativa de contraste, em valor absoluto, com o valor d' .

Todo $|Y| \geq d'$ será significativo, indicando que a média da testemunha (ou padrão) difere significativamente da média do tratamento com ela comparado.

Todo $|Y| < d'$ será não significativo e as médias desse contraste não diferem entre si.

g) indicar a significância do teste no valor da estimativa do contraste.

Num experimento com aditivos funcionais na ração para frangos de corte, foram utilizados 4 tratamentos (3 com aditivo e 1 testemunha) e 6 repetições, num delineamento inteiramente ao acaso (Quadro 11). O esquema da ANAVA foi:

Quadro 11 – Análise de variância de uma delineamento em DIC

CV	GL	S.Q.	Q.M.	F
Tratamentos	3	32,64	10,88	7,56
Resíduo	20	28,80	1,44	
Total	23	61,44		

Onde: $m_1 = 28,8$; $m_2 = 34,8$; $m_3 = 36,0$ $m_4 = 34,0$, que corresponde o peso da parcela com 20 aves.

t_d 5 gl. Tratamento $t_d = 2,73$
20 gl. Resíduo

$$V(Y) = \left[\frac{1}{4} + \frac{1}{4} \right] \cdot 1,44 = 0,72$$

$$s(Y) = \sqrt{V(Y)} = \sqrt{0,72} = 0,84 \text{ Kg / parcela}$$

$$d' = 2,73 \times 0,84 \therefore 2,29 \text{ Kg / parcela}$$

os contrastes são:

$$Y_1 = m_2 - m_1 = 6,0 \text{ Kg / parcela}$$

$$Y_2 = m_3 - m_1 = 7,2 \text{ Kg / parcela}$$

$$Y_3 = m_4 - m_1 = 5,2 \text{ Kg / parcela}$$

Verificamos, portanto, que todos os tratamentos com aditivos diferem da testemunha e foram mais eficientes que ela.

Teste de Bonferroni

O teste proposto pôr Bonferroni é indicado para aquelas situações práticas em que não conseguimos formar um conjunto de contrastes mutuamente ortogonais para as comparações de interesse. Em geral, o teste de Bonferroni é mais poderoso para estas situações do que os processos alternativos adaptados de comparações após exame dos dados (Scheffé, por exemplo).

O teste **t** de Bonferroni não exige um F significativo para tratamentos na análise de variância. É definido pôr:

$$T_B = \sqrt{\frac{Y_i}{V(Y_i)}}$$

onde,

$$Y_i = c_1 m_1 + c_2 m_2 + \dots + c_t m_t, \sum c_i = 0$$

i = número de contrastes

$$V(Y_i) = \sum_{i=1}^t \frac{c_i^2}{r_i} \cdot s^2, \quad s^2 = \text{QMR}$$

O valor crítico encontrado na tabela de Bonferroni para alternativas bilaterais é $\pm t_{B,\alpha/2, m, v}$ onde **m** é o número de contrastes selecionados e **v** o número de graus de liberdade do erro. Os valores críticos para alternativas unilaterais são

$$+ t_{B,\alpha/2, m, v} \text{ e } - t_{B,\alpha/2, m, v}$$

Se **m** for muito grande em relação ao número de tratamentos do experimento

($m > 16$), ganha-se em sensibilidade nas comparações usando a técnica de Scheffé proposta para contrastes selecionados, após observação dos resultados.

O intervalo de confiança de $1 - \alpha$ para Y_i é dado pôr:

$$DMS = B = t_{B, \alpha/2, m, v} \sqrt{V(Y_i)}$$

Onde,

$$DMS = B = t_{B, \alpha/2, m, v} \sqrt{V(Y_i)}$$

Para o exemplo de um experimento proposto para estudar os efeitos de implantação de hormônios nos ganhos de peso médios de perus na faixa de 15 a 20 semanas de idade. Os tratamentos em estudo foram: (1) controle, (2) 24 mg de dietilestilbestrol, (3) 10 mg de estradiol e (4) 20 mg de estradiol. O experimento foi conduzido com 24 repetições pôr tratamento e o totais de tratamentos = 61,23; 74,16; 64,37 e 71,01 respectivamente (Quadro 12).

Quadro 12 – Análise de variância de uma delimitação em DIC

CV	GL	S.Q.	Q.M.	F
Tratamentos	3	4,4015	1,4671	5,041
Residuo	92	26,772	0,2910	
Total	95	31,1735		

Nesse exemplo de implantação de hormônios, considere o conjunto de 4 contrastes não ortogonais: (1) controle vs estilbestrol; (2) controle vs estradiol; (3) 10 mg vs 20 mg estradiol e (4) estilbestrol vs estradiol,

Considerando o terceiro contraste;

$$Y_3 = m_3 - m_4 = 2,68 - 2,96 = -0,28$$

Uma estimativa da variância do contraste é:

$$V(Y_3) = \left(\frac{1}{24} + \frac{1}{24} \right) \cdot 0,291 = 0,291/12 = \mathbf{0,02425}$$

$$t_B = \frac{-0,28}{\sqrt{0,02425}} = -1,798$$

O valor crítico pela tabela de Bonferroni é $\pm t_{B; 0,025; 4; 92} = \pm 2,55$. Como o t_B calculado caiu neste intervalo, concluímos que o contraste é não significativo, não havendo portanto diferença entre as duas doses de estradiol.

Teste de Scott e Knott

O procedimento de Scott e Knott (1974) citado por Ferreira (1998), utiliza a razão de verossimilhança para testar a significância de que os n tratamentos podem ser divididos em dois grupos que maximizem a soma de quadrados entre grupos.

Seja um exemplo de 3 tratamentos, A, B e C. O processo consiste em determinar uma partição, em dois grupos, que maximize a soma de quadrados. Nesse caso são possíveis $2^n - 1$ grupos, isto é A vs B e C, B vs A e C e C vs A e B. Com um número pequeno de tratamentos como o exemplo, é fácil obter todos os grupos. Contudo, quando o número (n) de tratamentos é grande, o número de grupos cresce exponencialmente, dificultando a aplicação do teste. Para atenuar esse problema, basta ordenar as médias dos tratamentos. Nessa situação, o número de partições possíveis passa a ser obtido pôr $n - 1$. Uma vez ordenada as médias, procede-se do seguinte modo, fazendo inicialmente $g = n$:

- a) determinar a partição entre dois grupos que maximiza a soma de quadrados entre grupos. Essa soma de quadrados será definida pôr B_0 , e será estimada da seguinte forma: seja T_1 e T_2 os totais dos dois grupos com k_1 e k_2 tratamentos em cada um.

$$B_0 = \frac{T_1^2}{K_1} + \frac{T_2^2}{k_2} - \frac{(T_1 + T_2)^2}{k_1 + k_2}$$

$$T_1 = \sum_{i=1}^{k_1} Y_{(i)} \text{ e } T_2 = \sum_{i=k_1+1}^g Y_{(i)}$$

em que $Y_{(i)}$ é a média do tratamento da posição ordenada i .

Os dois grupos deverão ser identificados pôr meio da inspeção das somas de quadrados da $g - 1$ partições possíveis, sendo g o número de tratamentos envolvidos no grupo de médias considerado.

b) determinar o valor da estatística λ da seguinte forma:

$$\lambda = \frac{\pi}{2(\pi - 2)} \times \frac{B_0}{\sigma_0^2}$$

em que σ_0^2 é o estimador de máxima verossimilhança de σ_y^2 . Seja $s_y^2 = \text{QME} / r$ o estimador não viesado de σ_y^2 e o v os graus de liberdade associados a este estimador.

$$\sigma_0^2 = \frac{1}{g + v} \left[\sum_{i=1}^g (\bar{Y}_{(i)} - \bar{Y})^2 + v s_y^2 \right]$$

c) se $\lambda \geq \chi^2_{(\alpha; g/\pi-2)}$ rejeita-se a hipótese de que os dois grupos são idênticos em favor da hipótese alternativa de que os dois grupos diferem.

d) no caso de rejeitar esta hipótese, os

$$\sigma_0^2 = \frac{1}{g + v} \left[\sum_{i=1}^g (\bar{Y}_{(i)} - \bar{Y})^2 + v s_y^2 \right]$$

$$\sigma_0^2 = \frac{1}{4 + 20} (21,16 - 23,90)^2 + \dots + (28,14 - 23,90)^2 + 20 \times \left[\frac{10,19}{6} \right]$$

$$\sigma_0^2 = \frac{1}{24} [26,98 + 33,9667] = 2,5394$$

dois subgrupos formados serão independentemente submetidos aos passos (a) e (c), fazendo respectivamente $g = k_1$ e $g = k_2$. O processo em cada subgrupo se encerra ao se aceitar H_0 no passo (c) ou se cada subgrupo contiver apenas uma média.

Para ilustrar a aplicação deste algoritmo, será considerado um exemplo de um experimento em que o QME = 10,19 com 20 graus de liberdade, cujas médias dos 4 tratamentos foram estimadas a partir de 6 repetições, dadas pôr:

$$\begin{aligned} Y_{(1)} &= 21,16 \\ Y_{(2)} &= 22,72 \\ Y_{(3)} &= 23,58 \\ Y_{(4)} &= 28,14 \end{aligned}$$

A somas de quadrados da partição 1 vs 2, 3 e 4 é obtida pôr:

$$B_0 = \frac{T_1^2}{K_1} + \frac{T_2^2}{k_2} - \frac{(T_1 + T_2)^2}{k_1 + k_2}$$

$$B_0 = \frac{21,16^2}{1} + \frac{(22,72 + 23,58 + 28,14)^2}{3} - \frac{(21,16 + 22,72 + 23,58 + 28,14)^2}{4}$$

De modo análogo, obtém-se as somas de quadrados das partições d 1 e 2 vs 3; e 4 ($B_0 = 15,37$) e 1, 2 e 3 vs 4 ($B_0 = 23,97$). Assim, essa última partição foi a que maximizou a soma de quadrados entre grupos.

A estimativa σ_0^2 é:

Assim,

$$\lambda = \frac{\pi}{2(\pi-2)} \times \frac{B_0}{\sigma_0^2}$$

$$\lambda = \frac{\pi}{2(\pi-2)} \times \frac{23,9701}{2,5394} = 12,9879$$

O valor de $\chi^2_{(\alpha; g/\pi-2)}$ é 8,65. Como $\lambda > 8,65$ rejeita-se H_0 , ou seja, dois grupos são

$$\sigma_0^2 = \frac{1}{4+20} (21,16 - 22,4867)^2 + \dots + (23,58 - 22,4867)^2 + 20 \times \left(\frac{10,19}{6} \right)$$

$$\sigma_0^2 = \frac{1}{24} [3,0099 + \dots 33,9667] = 1,6077$$

Assim:

$$\lambda = \frac{\pi}{2(\pi-2)} \times \frac{2,6401}{1,6077} = 2,2596$$

O valor de $\chi^2_{(\alpha; g/\pi-2)}$ é 7,133, logo como $\lambda < 7,133$ aceita-se H_0 , ou seja, os tratamentos 1,2 e 3 pertencem a um mesmo grupo. O resultado final do teste é portanto:

Trat.	Médias	Grupos
$Y_{(1)}$	21,16	a
$Y_{(2)}$	22,72	a
$Y_{(3)}$	23,58	a
$Y_{(4)}$	28,14	b

Pode-se verificar que no teste de Scott e Knott não ocorre ambiguidade das comparações de médias, o contrário que apresenta

formados ao nível de 5%. O grupo 1 com os tratamentos 1, 2 e 3 e o grupo 2 com o tratamento 4. Assim o procedimento é repetido apenas para os subgrupos que contém mais de 1 tratamento, no caso apenas o grupo 1.

As partições 1 vs 2 e 3 e e 2 vs 3 possuem as seguintes somas de quadrados entre grupos: $B_0 = 2,6401$ e $B_0 = 1,7931$. Assim, a partição 1 vs 2 e 3 é a que maximiza B_0 e as estimativas σ_0^2 e λ devem ser obtidas pôr:

na maioria dos outros métodos.

Considerando os testes aqui apresentados, é possível observar o efeito específico da utilização de cada um deles. Em uma situação de um ensaio com o delineamento inteiramente casualidade, sempre com 4 repetições pôr tratamento, observa-se que ocorre um aumento da d.m.s dos tipos de testes de comparação de média quando aumenta concomitantemente o número de tratamentos.

Assim, é possível verificar no quadro 13:

Valores da diferença mínima significativa segundo o número de tratamentos estudados e o teste de significância aplicado

Quadro 13 – Diferença mínima significativa em função número de tratamentos

Teste	Número de tratamentos					
	2	3	4	5	10	20
t Student	1,73	1,60	1,54	1,51	1,44	1,41
Duncan	1,73	1,67	1,66	1,66	1,68	1,74
Dunnett	1,73	1,87	1,90	1,93	2,02	2,11
SNK	1,73	1,84	2,10	2,19	2,42	1,62
Tukey	1,73	1,98	2,10	2,19	2,42	2,62
Scheffé	1,73	2,06	2,29	2,47	3,13	4,13
gl. resíduo	06	09	12	15	30	60

Pode-se notar que enquanto os valores das d.m.s. para os demais testes aumentam com a elevação do número de tratamentos, aqueles do teste t diminuem. Isto ocorre porque o teste t está em função tão somente do gl do resíduo e não consideram o número de tratamentos como os demais.

Para $t = 2$, todos os testes têm o mesmo valor de d.m.s. As diferenças entre esses valores, passam a se manifestar mais fortemente a partir de $t = 4$ onde representam quase 30% da dms de maior valor. Isto significa que o uso do teste t em ensaios de até 4 tratamentos não está tão penalizado pela incorporação do erro tipo I, como é o caso de experimentos com 10 tratamentos (onde aquelas diferenças podem alcançar 5%).

Também pode observar que os testes mais sensíveis de maiores chances de erro tipo I são os mais próximo do teste t de Student e vão se tornando mais rigorosos, maiores chances de erro tipo II à medida que se aproxima de Scheffé, o mais rigoroso.

O teste de Dunnett, pôr sua especialidade deixa de ser considerado como alternativa junto aos demais, mas pode-se observar que ele se situa no centro da tabela e poderia representar o controle balanceado de ambos os erros (I e II) se não fosse sua limitação condicional.

Variáveis muito instáveis ($CV \geq 30\%$) estarão mais sujeitas ao erro tipo II. Usar teste t de Student se até 4 tratamentos ou de Duncan se houver mais de 4 tratamentos.

Variáveis pouco instáveis ($CV \geq 15\%$) estarão mais sujeitas ao erro tipo I. Utilizar testes mais rigorosos como SNK, Tukey ou Scheffé ($t \geq 5$). Para $t > 5$, SNK ou Tukey, pois o teste de Scheffé para mais de 5 tratamentos aumentaria a ocorrência do erro tipo II. Se este tipo de erro precisar ser controlado, o teste SNK será melhor opção.

Variáveis mediamente instáveis ($15\% < CV < 30\%$) medidas em até 4 tratamentos abonam a utilização do teste t de Student que

pôr seu expediente torna-se mais prático a despeito do ligeiro aumento na ocorrência do erro tipo I. Para 5 ou mais tratamentos, utilizar Duncan ou SNK dependendo do controle que se queira dar a ambos erros (I e II) ou somente ao erro tipo I, respectivamente (Bertoldo et al., 2008; Sampaio, 2007).

A escolha do teste estatístico feita sob essas considerações apontará resultados menos conflitantes com o julgamento do experimentador, proveniente de sua percepção prática e de sua lógica dedutiva.

Sempre que possível, observe a estrutura dos tratamentos e formule as comparações de interesse pôr meio de um grupo de contrastes mutuamente ortogonais. Os contrastes devem ser organizados ainda na fase de planejamento, sem acesso aos resultados experimentais. Se isto for feito, a hipótese $H_0 : t_1 = 0$, resposta à tratamentos, pode ser ignorada e a ANAVA será processada para fornecer uma estimativa válida e confiável do QME.

Caso haja problemas em se conseguir um conjunto de contrastes mutuamente ortogonais, use os testes de Bonferroni ou Scheffé.

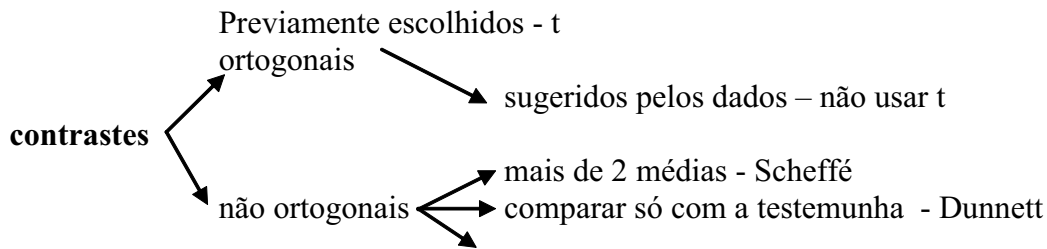
Se o objetivo principal é comparar cada média de tratamento com o controle e não há interesse em comparar os tratamentos entre si, use o teste de Dunnett.

Se os tratamentos forem quantitativos não faça comparações 2 a 2 (Tukey, SNK e outros). Para estes casos, use a técnica de polinômios ortogonais (análise de regressão).

Se os tratamentos forem qualitativos e não existe uma estrutura entre eles que permita partir para grupos de contrastes, use as comparações 2 a 2 como último recurso.

Se a opção foi comparação 2 a 2, use os testes Tukey e SNK que são mais poderosos. Evite testes fracos.

Se os contrastes de interesse forem determinados após a realização do experimento (após obtenção dos dados), o teste t de Student não deve ser utilizado.



A escolha do método adequado para comparar média exige, ainda, que se leve em consideração tanto o nível de significância como o poder do teste. Assim, o nível de significância de um teste é a probabilidade de rejeitar a hipótese de que as médias são iguais, quando esta hipótese é, na realidade, verdadeira. O poder do teste é a probabilidade de rejeitar a hipótese de que as médias são iguais quando esta hipótese é, na realidade, falsa. Dessa forma, é prudente que o pesquisador ao realizar um ensaio experimental possa utilizar um teste com baixo nível de significância e poder elevado. Como nível de significância e poder do teste crescem juntos, é necessário que o pesquisador fique atento no verdadeiro método de comparação de médias, uma vez que diminuir o nível de significância implica diminuir o poder do teste.

CONSIDERAÇÕES FINAIS

A comparação de médias só pode ser feita após a análise de variância. Isto porque todos os procedimentos para obter a d.m.s. exigem o cálculo do quadrado médio do resíduo. Mas a análise de variância dá o valor de F, que permite decidir se as médias são ou não iguais, a determinado nível de significância.

Variáveis muito instáveis ($CV \geq 30\%$) estão mais sujeitas ao erro tipo II, podendo ser usado o teste *t* de Student se até 04 tratamentos ou de Duncan se houver mais de 04 tratamentos. Já para variáveis pouco instáveis ($CV \leq 15\%$) estão sujeitas ao erro tipo I, onde deve-se utilizar os testes mais rigorosos como SNK, Tukey ou Scheffé. E para vari-

áveis medianamente instáveis ($15\% < CV < 30\%$) medidas em até 04 tratamentos abonam a utilização do uso do teste *t* de Student que por seu expediente torna-se mais prático a despeito do ligeiro aumento na ocorrência do erro do tipo I. No entanto, para maior número de tratamentos deve utilizar o Duncan ou SNK de acordo com a forma de controle dos erros Tipo I e II.

Na medida em que os contrastes que envolvem grupos de médias são definidos após prévia inspeção dos dados analisados, a sua significância pode ser testada com o uso do teste de Scheffé. No entanto, se o pesquisador deseja comparar todos os pares de médias entre si, deve ser escolhido um teste de comparação de médias mais rígido, como, por exemplo, Tukey. Quando o interesse é comparar somente a média de qualquer tratamento com a testemunha, o procedimento correto é o emprego do teste de Dunnett.

De fato, todos os procedimentos que o pesquisador de ensaios zootécnicos for utilizar para a comparação de média apresentarão algumas vantagens e desvantagens, embora ainda não existe um teste definitivamente melhor que todos os outros. Convém estar atento aos procedimentos para comparar médias no sentido de adotar um método formal para comparar médias. Isto evita que as conclusões fiquem totalmente dependentes da opinião meramente do pesquisador. Mesmo assim, existe uma grande margem de opção tanto na escolha do teste, como no estabelecimento do nível de significância que permite decidir aceitar ou rejeitar uma determinada hipótese. Assim, o pesquisador não está livre de incorrer em um dos erros (tipo I ou II).

REFERÊNCIAS BIBLIOGRÁFICAS

- AQUINO, L.H.de. Técnica experimental com animais I. ESAL:Lavras, 1992, 167p.
- ALVAREZ, V. H.; ALVAREZ, G. A. M. Comparação de médias ou teste de hipóteses? Contrastes! **Sociedade Brasileira de Ciências do Solo**, 3 (1): 24-33. 2006.
- BANCROFT, T.A. Topics in intermediate statistical methods. Iowa State University Press, v.1, 1968, 129p.
- BERTOLDO, J. B; COIMBRA, J. L. M; GUIDOLIN, A. F; MANTOVANI, A; VALE, N.M. Problemas relacionados com o uso de testes de comparação de médias em artigos científicos. **Revista Biotemas**. v.21, n 2. p. 145-153. 2008.
- CHEW, V. Comparing treatment means: a compedium. **HortScience**, v.11, n.4, p.348-357, 1976.
- BANZATTO, D.A.; KRONKA, S N. **Experimentação agrícola**, FUNEP:Jaboticabal, 3.ed., p.23-53, 1
- BUSSAB, W.O.; MORETTIN, P.A. **Estatística básica**. 4 ed. Atual Editora Ltda:São Paulo, 1987.
- CARDELLINO, R.A.; SIEWERDT, F. Utilização correta e incorreta dos testes de comparação de médias. **Revista da Sociedade Brasileira de Zootecnia**, v.21, p.985-995. 1992.
- CAMPOS, H. de. **Estatística aplicada à experimentação com cana-de-açúcar**. 4 ed., FEALQ:Piracicaba, 1984, 292p.
- COCHRAN, W.G.; COX, G.W. **Experimental designs**. New York:Wiley, 1968.
- DUNCAN, B.B. Multiple range and multiple F-tests. **Biometrics**, v.11, p.1-42, 1955.
- DUNNETT, C.W. A multiple comparison procedure for comparing several treatments with a control. **J. Am. Stat. Assoc.**, v.50, p.1096-1121, 1955.
- FERREIRA, D.F. **Sistema de Análise de Variância de Dados Balanceados (SISVAR)**. Lavras:UFLA-DEX, 2000, 25p.
- FISHER, R..A. **The design of experiments**. Edingburgh:Oliver & Body Ltd., 1935.
- GILL, J.L **Design and analysis of experiments in the animal and medical sciences**. Ames:The Iowa State Univ. Press, v.3, 1978.
- LITTLE, T.M. Is galileo published in hortscience. **HortScience**, v.13, n.5, p.504-506, 1978.
- MILLER Jr., R.C. **Simultaneous statistical inference**. New York, McGraw-Hill, 1966, (2 ed: Spring-Verlag, 1981).
- MILLER Jr., R.C. **Development in multiple comparisons 1966-1976**. **JASA**, 72:779-788, 1977.
- MONTEIRO, L.S. A experimentação e a experiência. **Revista portuguesa de ciências veterinárias**. Suplemento:117. 1-17. 2002.
- NEWMAN, D. The distribution of range in samples from a normal population expressed in terms of na independent estimate of standard deviation. **Biometrika**, v.31, p.20-30, 1939.
- O'NEILL, R.; WETHERILL, G.B. The present state of multiple comparison methods. **J. Royal Stat. Soc.**, B, 33:218-250, 1971.

PETERSEN, G.R. Use and misuse of multiple comparison procedures. **Agronomy Journal**, v.69, p.205-208. 1977.

PIMENTEL GOMES, F. **Curso de estatística experimental**. 13 ed. Piracicaba, Livraria Nobel, 1990, 468p.

SAMPAIO, I.B.M. Esclarecimento aos pesquisadores que utilizam testes de comparação de médias. **Revista da Sociedade Brasileira de Zootecnia**., v.22, n.5, encarte, 1993.

SAMPAIO, I.B.M. **Estatística aplicada à experimentação animal**. Belo Horizonte:Fundação de Ensino e Pesquisa em Medicina Veterinária e Zootecnia, ed. 3. 2007, 265p.

SCHEFFÉ, H. A method for judging all contrasts in the analysis of variance. **Biometrika**, v.40, p.87-104, 1954.

SCOTT .A.J.; KNOTT, M. A cluster analysis method for grouping means in the analysis of variance. **Biometrics**, 30:507-512, 1974.

SHAFFER, J.P. Multiple comparisons emphasizing selected contrasts:Na extension and generalization of Dunnett's procedure. **Biometrics**, 33:293-303, 1977.

SNEDECOR, G.W.; COCHRAN, W.G. **Statistical methods**. Ames:Iowa State Univ. Press, 1980.

SILVA, J. G. C. da. A consideração da estrutura das unidades em inferências derivadas do experimento. **Pesquisa Agropecuária Brasileira**, **34** (3): 911-925. 1999.