

# TESTE DE COMPARAÇÃO DE MÉDIAS: DIFICULDADES E ACERTOS EM ARTIGOS CIENTÍFICOS

## TEST OF COMPARISON OF MEANS: DIFFICULTIES AND SUCCESSES IN SCIENTIFIC PAPERS

Juliano Garcia Bertoldo<sup>1</sup>; Fabiani da Rocha<sup>1</sup>; Jefferson Luís Meirelles Coimbra<sup>1\*</sup>; Danieli Zitterell<sup>1</sup>; Vanessa de Fátima Grah<sup>1</sup>;

### RESUMO

Este trabalho teve como objetivo identificar erros e acertos com relação à aplicação dos testes de comparação de médias em trabalhos científicos, e sugerir alternativas para melhorar a interpretação dos resultados pelos pesquisadores. Foi realizada uma revisão de quatrocentos e oitenta e três trabalhos científicos publicados na Revista Brasileira de Agrociência, no período entre 1995 a 2005. Os trabalhos foram classificados quanto ao uso dos testes de comparação de médias (TCM): i) correto; ii) parcialmente correto e; iii) incorreto. Experimentos com apenas um fator (71% dos artigos foram classificados como correto), não há grande dificuldade por parte dos pesquisadores na aplicação correta dos TCM. Ainda assim, em experimentos com apenas um fator, a principal dificuldade corroborada esta relacionada com trabalhos que envolvem fator de classificação quantitativo (8% dos artigos foram classificados como incorreto). Por outro lado, em experimentos com mais de um fator (67% dos artigos foram classificados como incorreto), os resultados encontrados são diferentes em relação aos trabalhos unifatorial. Demonstrando assim, dificuldade por parte dos pesquisadores em experimentos com este tipo de fator. A aplicação dos TCM para fatores de natureza quantitativa foi o abuso mais comum. Nos experimentos fatoriais constatou-se a não decomposição dos graus de liberdade da interação, quando esta é significativa. Os principais aspectos que afetam diretamente a fidedignidade na interpretação dos resultados por parte dos pesquisadores são: i) a não consideração das pressuposições dos diferentes testes estatísticos e; ii) conhecimento incipiente em relação quais os tipos de fatores.

*Palavras-chave:* fator qualitativo específico, fator quantitativo, unifatorial, fatorial, interação.

### ABSTRACT

This work had the objective of identify mistakes and successes regarding the application of tests of comparison of means in scientific papers and to suggest alternatives to improve the interpretation of the results for the researchers. It was evaluated four hundred and eighty three papers published from 1995 to 2005 in the Revista Brasileira de Agrociência, a scientific Brazilian Journal. According to the use of the tests of comparison of means (TCM), the studies were classified as: i) correct; ii) partially correct; iii) and incorrect. In

experiments dealing with only one factor, 71% of the studies were classified as correct, showing there was no great difficulty by the researchers in the correct application of TCM. Nevertheless, when the evaluated factor was quantitative, the TCM was used incorrectly in 8% of the studies. When the experiments involved more than one factor, the number of studies that used the TCM incorrectly increased to 67% demonstrating the researchers have difficulty in handling the statistical analysis. The most common abuse of application of TCM was found for experiments dealing with quantitative factors regardless of the number of factors involved. In the factorial experiments the researchers' deficiency was verified in partitioning the degrees of freedom of the interaction, when it was significant. The main aspects that affect the reliable interpretation of the results by the researchers are: i) lack of information about the assumptions of the different statistical tests and; ii) their incipient knowledge in relation to the types of classification of the factors.

*Key words:* specific qualitative factor, quantitative factor, unifatorial, factorial, interaction.

### INTRODUÇÃO

Freqüentemente, os pesquisadores utilizam inadequadamente testes estatísticos devido à não utilização das pressuposições necessárias. Em consequência, os pesquisadores podem ter dificuldades tanto na análise dos dados coletados, quanto na interpretação de resultados obtidos, podendo assim, originar conclusões errôneas.

Assim sendo, é de fundamental importância para à pesquisa a escolha do delineamento experimental adequado, bem como o emprego correto do modelo estatístico, no intuito de estimação correta do erro experimental.

A escolha do teste inadequado pode causar dificuldade na interpretação dos resultados de um experimento ou mesmo levar o pesquisador a realizar conclusões erradas (CARDELINO & SIEWERDT, 1992).

Por exemplo, os principais aspectos que afetam diretamente a fidedignidade são: i) falta de informação sobre as pressuposições dos diferentes testes estatísticos e;

(Recebido para Publicação em 11/05/2007, Aprovado em 08/02/2007)

ii) conhecimento incipiente da grande maioria dos pesquisadores quanto aos tipos de fatores e tipos de variáveis. Assim, 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 et al., 1999, BANZATTO & KRONKA, 2006), uma vez que, tendo pleno conhecimento da classificação dos diferentes tipos de fatores quanto à escala de medida e do tipo de distribuição das variáveis (normal, binomial, etc.), as inferências sobre os resultados podem ganhar acuidade e precisão.

Para uma correta análise dos dados, se faz necessário o conhecimento da natureza dos fatores estudados. De modo geral, os fatores podem ser classificados em qualitativos (específicos estruturados e não-estruturados, ordenados e amostrados) e em quantitativos. Fatores qualitativos estruturados são aqueles cujos níveis podem ser classificados em grupos, cujas comparações constituem o objetivo do trabalho e os fatores qualitativos não estruturados apresentam como objetivo comparações entre todos os níveis do fator, ou seja, todos contra todos. Os fatores qualitativos não-estruturados apresentam como objetivo comparações entre todos os níveis do fator experimental, ou seja, todos contra todos (CHEW, 1976).

Os fatores qualitativos, por exemplo, cultivares, se diferenciam dos quantitativos, por exemplo, densidade de plantas por unidade de área, doses de um fertilizantes ou hormônio, por serem expressos numa variável nominal, de escala ordinal ou ainda aleatória, enquanto que os quantitativos são expressos numa variável intervalar ou racional. Entretanto, muitas vezes um fator de natureza quantitativa pode ser classificado como qualitativo, de acordo com o objetivo do trabalho. Por exemplo, se um experimento é constituído por 3 níveis de um hormônio (0, 45 e 90 mg L<sup>-1</sup>) e se objetiva verificar as concentrações específicas, o fator é classificado como qualitativo. Por outro lado, sendo o objetivo verificar os intervalos compreendidos entre as concentrações (níveis) do hormônio, a natureza do fator em estudo é quantitativa.

Quando apresentam características qualitativas e não relacionadas são comumente empregados os testes de comparação de médias, pois, segundo AFLAKPUI (1995), os testes de comparação de médias freqüentemente são os

mais utilizados pelos pesquisadores. Este é recomendado quando o objetivo é determinar os melhores tratamentos dentro de um grupo, ou seja, comparar todos os tratamentos entre si. Ainda, o uso de comparação de médias se restringe às situações específicas, tolerado quando não se conhecem as características dos fatores (CHEW, 1976). Além dos testes de comparação de médias, dependendo do objetivo e da estrutura dos tratamentos, as comparações podem ser realizadas por contrastes, possibilitando uma melhor inferência do pesquisador sobre os resultados. De forma geral, os contrastes podem ser divididos em comparações, quando os efeitos de um grupo de tratamentos das variáveis em estudo são comparados com os de outro grupo (ALVAREZ & ALVAREZ, 2006). Se os fatores forem de natureza quantitativa, com mais de dois níveis, recomenda-se a análise de regressão. Segundo CHEW (1976), quando os fatores são quantitativos, a análise de regressão é a técnica mais apropriada para o pesquisador inferir sobre os resultados. Neste caso, o objetivo do pesquisador é observar o comportamento da variável resposta entre os intervalos avaliados, como por exemplo, doses de um fertilizante, e não os níveis específicos.

Portanto, para cada tipo de fator, há um teste estatístico específico e apropriado, o que reforça a importância do pesquisador conhecer a classificação dos fatores que envolvem o seu trabalho. Assim, quando os fatores forem quantitativos o ajustamento de funções de resposta, através de técnicas de regressão (JONES, 1984), é o apropriado; o adequado para fatores qualitativos são os conjuntos de contrastes planejados entre as médias ou grupos de médias (GILL, 1973), e os procedimentos de comparação múltiplas de médias para fator qualitativo específico não estruturado, como por exemplo, cultivares (CHEW, 1976).

Desta forma, este trabalho teve como objetivo identificar erros e acertos com relação à aplicação de testes de comparação de médias em trabalhos científicos, e sugerir alternativas para uma melhor interpretação dos resultados pelos pesquisadores.

## MATERIAL E MÉTODOS

Foi realizada uma revisão de quatrocentos e oitenta e três (483) trabalhos científicos publicados na Revista

Brasileira de Agrociências, no período entre 1995 a 2005, sendo utilizados trezentos e quarenta (340) trabalhos. Os artigos publicados e que não foram revisados, foram aqueles que não envolveram testes de comparação de médias, por não se relacionarem com o objetivo deste trabalho; ou ainda, nenhuma análise estatística foi utilizada. Os demais trabalhos foram classificados quanto ao uso dos testes de comparação de médias: *i)* correto; *ii)* parcialmente correto e; *iii)* incorreto. A categoria correto (teste de comparação de médias para fatores qualitativos não-estruturados e não relacionados, análise de regressão para fatores quantitativos e contrastes ou teste de Dunnett, para fatores qualitativos estruturados) refere-se aos trabalhos cuja realização de teste de comparação de médias, foi em função dos objetivos do trabalho e da estrutura de fatores. A categoria parcialmente correto caracteriza a aplicação de testes de comparação de médias em experimentos com fatores qualitativos estruturados e análise de regressão simultaneamente a testes de comparação de médias para fator quantitativo. A

categoria incorreto foi usada para trabalhos com níveis de um fator quantitativo no qual se aplicou um teste de comparação de médias, análise de regressão para fatores qualitativos e ao não ser verificado o efeito da interação em experimentos fatoriais (com mais de um fator), nesse caso, quando o efeito da interação não foi estudado, o trabalho foi considerado incorreto, e se o efeito foi considerado, foram avaliados os efeitos dos fatores separadamente, da mesma forma que nos trabalhos com experimentos unifatorial. Se pelo menos um dos fatores estudados foram avaliados de forma errônea, o trabalho foi considerado como incorreto.

Para melhor compreensão, foram separados exemplos mais contrastantes, sendo divididos em três classes: classe I corretos; classe II parcialmente corretos e; classe III incorretos. Além dos exemplos, para facilitar o entendimento dos pesquisadores, a Tabela 1, adaptada de GILL (1978), demonstra alternativas para a análise dos dados, com relação aos fatores em estudo.

Tabela 1 - Alternativas para a escolha de testes estatísticos em relação com os fatores de estudo.

Tipo de Fatores	Alternativas para inferência
1. Qualitativos	Sim.....2 Não.....5
2. Estruturado	Sim.....3 Não.....4
3. Tratamento com testemunha	Sim.....Dunnett Não.....Contrastes
4. Não estruturado	Balanceados .....Tukey Desbalanceados.....Bonferroni, Scheffé

## RESULTADOS E DISCUSSÃO

Nos artigos revisados, a inadequada consideração dos tipos de fatores em estudo, ficou evidente, principalmente no uso incorreto de testes de comparações de médias em experimentos fatoriais. Em experimentos unifatoriais, não houve grande dificuldade na utilização de testes de comparações de médias, conforme demonstrado na Tabela 2, onde a aplicação de testes de comparação de médias

para a categoria unifatorial foi correta em 71% (109 trabalhos publicados) dos trabalhos revisados, em 21% (32 trabalhos publicados) foi parcialmente correta e em 8% (13 trabalhos publicados) incorreta. A principal dificuldade verificada, de acordo com a Tabela 2, está relacionada a trabalhos que envolvem fatores quantitativos, sendo que em 13 trabalhos, o uso de teste de comparação de médias foi incorreto.

Tabela 2 - Número de artigos revisados em experimentos fatoriais, quanto ao tipo de fator (qualitativo não estruturado, qualitativo estruturado, quantitativo) e quanto ao uso (correto, parcialmente correto e incorreto), editados na Revista Brasileira de Agrociência, no período 1995 a 2005.

Classificação	UNIFATORIAL			Total	%
	Ql <sup>1</sup>	Qle <sup>2</sup>	Qt <sup>3</sup>		
Correto	59	4	46	109	71
Parcialmente correto	0	24	8	32	21
Incorreto	0	0	13	13	8
Total	59	28	67	154	
%	38	18	44		

Ql<sup>1</sup> – qualitativo não estruturado; Qle<sup>2</sup> – qualitativo estruturado; Qt<sup>3</sup> – quantitativo

Esses resultados estão de acordo com CARDELINO & SIEWERDT (1992) e SANTOS et. al (1998), que revisaram trezentos e noventa (390) trabalhos científicos publicados na Revista da Sociedade Brasileira de Zootecnia e trabalhos publicados na Pesquisa Agropecuária Brasileira (PAB) no período de 1980 a 1994, com algum tipo de comparação de médias, respectivamente, sendo que o uso inadequado mais

freqüente dos testes de comparação de médias ocorreu nos casos em que os tratamentos foram de natureza quantitativa.

Por outro lado, em experimentos fatoriais, os resultados foram diferentes, como observado na Tabela 3, onde 31% (58) dos trabalhos revisados foram classificados como corretos, 2% (4) parcialmente corretos e 67% (124) como incorretos.

Tabela 3 - Número de artigos revisados quanto ao número de fatores (fatorial), quanto ao tipo de fator (qualitativo não estruturado, qualitativo estruturado, quantitativo) e quanto ao uso (correto, parcialmente correto e incorreto), editados na Revista Brasileira de Agrociência no período 1995 a 2005.

Fatorial	Correto	Parcialmente correto	Incorreto	Total	%
Ql x Ql <sup>1</sup>	22	0	33	55	29,60
Ql x Qt	13	0	30	43	23,00
Ql x Qle <sup>2</sup>	1	0	14	15	8,10
Qt x Qt <sup>3</sup>	6	3	9	18	9,70
Qt x Qle	0	1	1	2	1,10
3 Ql	6	0	11	17	9,10
2 Ql x Qt	9	0	11	20	10,80
Ql x 2 Qt	1	0	10	11	6,00
3 Ql x Qt	0	0	2	2	1,10
4 Ql	0	0	1	1	0,50
2 Ql x Qle x Qt	0	0	1	1	0,50
3 Qt	0	0	1	1	0,50
Total	58	4	124	186	
%	31	2	67		

Ql<sup>1</sup> – qualitativo não estruturado; Qle<sup>2</sup> – qualitativo estruturado; Qt<sup>3</sup> – quantitativo

Os resultados demonstram a dificuldade por parte dos pesquisadores na aplicação de testes de comparação de médias em experimentos com fatores quantitativos, em experimentos fatoriais. O uso de um esquema fatorial é importante quando se objetiva investigar as possíveis interações entre os fatores, e ainda, segundo WECHSLER (1998), o arranjo fatorial além de permitir investigar as interações entre os fatores, permite estimar a variância do erro experimental com maior precisão, aumentando a potência dos testes estatísticos. Porém, em trabalhos com esquema fatorial ( $n \times n$ ;  $n \times n \times n$ ), se o efeito da interação entre os fatores for significativo, a análise dos fatores deve-se dar em conjunto, uma vez que os níveis de um fator ocorre em combinação de níveis de um outro fator (PETERSEN, 1976). Sendo assim, em experimentos

fatoriais, o primeiro passo é analisar os efeitos da interação entre os níveis dos fatores. Se for ignorado o efeito da interação significativa, ou seja, a dependência do efeito de um fator nos níveis do outro, a premissa de que os níveis de um fator apresentam comportamento diferenciado na presença do outro fator pode ser comprometida. As comparações entre tratamentos podem ser afetadas substancialmente pelas condições em que ocorrem, e, interpretações claras dos efeitos de um tratamento precisam ser levadas em conta para o efeito de outros tratamentos. Isso significa que, uma vez significativa a interação entre os fatores, é necessário fazer um estudo dessa interação através de sua decomposição em fatores aninhados (KUEHL,1994).

De forma sucinta, geralmente o primeiro passo é submeter os dados a uma análise de variância, através do teste *F* global. O teste de hipóteses ou de significância permite decidir se aceita-se ou rejeita-se a hipótese de nulidade ( $H_0$ ), ou se a amostra observada difere estatisticamente dos valores esperados (BANZATTO & KRONKA, 2006), ou ainda se a o efeito da interação entre os fatores (esquema fatorial) é significativo. Se os graus de liberdade se equivalerem a um (1), o teste *F* se torna decisivo, pois de acordo com CHEW (1976), quando um experimento inclui apenas dois tratamentos, nenhum teste adicional faz-se necessário, uma vez que existindo apenas duas médias, e sendo o teste *F* significativo, o único contraste existente também é significativo. Por outro lado, se o fator em estudo for de natureza quantitativa, o teste *F* não é necessário, sendo aplicada análise de regressão diretamente. Assim, nos casos em que os tratamentos forem de níveis de um fator quantitativo, o apropriado é decompor os graus de liberdade em componentes polinomiais ortogonais (CHEW, 1976). Uma vez não sendo nenhum desses casos, os contrastes devem ser comparados por testes de comparação de médias. Um número de procedimentos estatísticos pode ser utilizado com esse propósito. Alguns deles são: planejar contrastes entre as médias, ou grupos de médias e; utilizar teste de comparação de médias (PETERSEN, 1976). Se a interação entre os fatores for significativa, os graus de liberdade devem ser rearranjados de modo a comparar os níveis de um fator dentro dos níveis do outro (CARDELINO & SIEWERDT, 1992). Nesse caso, fixando um fator e variando os níveis do outro fator, assim chamado de efeito simples. De acordo com HINKELMANN & KEMPTHORNE (1994), o procedimento de

variar um fator, por vez, via de regra se aplica quando o objetivo é estabelecer uma lei fundamental, o que conduziria ao conhecimento detalhado do efeito de um fator, quando os outros são mantidos constantes. Ainda, outro procedimento experimental seria considerar todas as possíveis combinações de níveis simultaneamente. Isso permitiria obter informação acerca dos fatores principais e, especialmente, da interação entre os vários fatores (NOGUEIRA, 1997).

#### Exemplos

##### Classe I – Trabalhos corretos

O trabalho publicado no volume 4, n.3, 183-186, 1998, foi classificado como correto quanto ao emprego de testes de comparação de médias (Tabela 4). O objetivo do trabalho foi avaliar a qualidade fisiológica das sementes de *Delphinium* que estavam a venda no mercado, em Pelotas/RS, no ano de 1995. Como apresentou um único fator, qualitativo específico não estruturado (quatro lotes de três diferentes empresas de importação de sementes de *Delphinium*) foi aplicado teste de Duncan para a comparação das médias de maneira correta, cujo objetivo era comparar todos contra todos. Apesar da correta utilização de teste de comparação de média nesse caso, a escolha do teste de Duncan não é adequada, uma vez que, com três ou mais médias sendo comparadas, a teoria do teste de Duncan é inadequada, pois o nível de significância global não é mantido (GILL, 1978). Como alternativa, os autores poderiam utilizar teste de Tukey, Scheffé ou ainda Bonferroni.

Tabela 4 - Médias das percentagens de germinação impressas na embalagem das sementes e avaliadas em laboratório, percentagem de germinação após pré-esfriamento, índice de velocidade de emergência (IVE) e estande final de sementes de *Delphinium* de quatro lotes diferentes<sup>(1)</sup>.

Lotes	Germinação impr. na emb. (%)	Germinação em lab. (%)	Germinação após pré-esfriamento (%)	IVE	Estande Final (%)
A	65 Ac	14 Bb	62 Ab	0,31 bc	20 aB

B	94 Aa	55 aC	80 aB	0,38 ab	27 aD
C <sub>1</sub>	80 Ab	48 aC	67 Bb	0,58 a	28 aD
C <sub>2</sub>	80 Ab	0 cC	24 Bc	0,13 c	7 bC

<sup>(1)</sup>Médias seguidas da mesma letra minúscula a coluna e da mesma letra maiúscula na linha, não diferem significativamente entre si pelo Teste de Duncan a nível de 5% de probabilidade.

Um exemplo da correta aplicação de testes de comparação de médias, foi verificada no trabalho publicado no volume 6, n.2, 107-111, 2000, que tinha como objetivo avaliar a resposta das plantas quanto às características fenométricas do crisântemo em diferentes populações e duas épocas de plantio, conduzido em haste única e cultivado em estufa plástica. Os fatores nesse experimento foram duas épocas de plantio e oito populações de plantas (2x8). Os autores não verificaram significância na interação entre os fatores. Como a interação não foi significativa, analisaram os fatores de forma separada, onde para as épocas de plantio aplicaram teste de Duncan, uma vez que é um fator qualitativo, e para as populações de plantas, como é um fator quantitativo ajustaram curvas de regressão para as variáveis.

#### Classe II – Trabalhos parcialmente corretos

Dependendo do objetivo do experimento, alguns testes estatísticos podem ser mais eficazes do que outros. Nesse exemplo (Tabela 5), o esquema era unifatorial: três tratamentos pré-geminativos (*calor seco*: exposição ao ar seco à temperatura de 50+- 2°C, durante seis dias;

*escarificação*: escarificação mecânica em um escarificador elétrico provido de lixa número 180; *calor úmido*: imersão em água a 60°C por cinco minutos e; *testemunha*). Os autores utilizaram teste de Tukey para a comparação entre os tratamentos, considerando assim, o fator como qualitativo não estruturado. Porém, uma vez que o trabalho apresenta uma testemunha, o fator deve ser classificado como qualitativo estruturado. Assim, o adequado seria a utilização de teste de Dunnett para a inferência sobre as médias, pois segundo CARDELINO & SIEWERDT (1992), o teste de Dunnett é apropriado quando se deseja comparações apenas entre os tratamentos e uma testemunha. De acordo com os mesmos autores, o teste de Dunnett pode ser aplicado de modo unilateral ou bilateral. Quando a *priori* sabe-se que o tratamento testemunha terá o melhor ou o pior resultado dentre todos os tratamentos, o teste pode ser aplicado de forma unilateral, aumentando a probabilidade de detecção de diferenças significativas entre o tratamento testemunha e os demais.

Tabela 5 - Percentagens de germinação (PG), dureza, plântulas anormais e sementes mortas de *Desmodium incanum* submetidas a tratamentos de superação de dormência<sup>(1)</sup>. Estação Experimental Agrônômica, Universidade Federal do Rio Grande do Sul, Eldorado do Sul, RS.

Tratamentos	PG	Dureza	Plântulas anormais	Sementes mortas
Escarificado	80 a	0 b	19 a	1 a

Calor seco	24 b	75 a	1 c	0 a
Calor úmido	21 b	75 a	4 b	0 a
Testemunha	18 b	82 a	0 c	0 a
CV (%)	15,081	11,495	32,183	230,940

<sup>(1)</sup>Médias na mesma coluna, seguidas por letras distintas diferem entre si pelo teste de Tukey, ao nível de 1% de significância.

### Classe III – Trabalhos incorretos

O primeiro exemplo incorreto foi extraído do trabalho que avaliou efeito do ácido indolbutírico no enraizamento de estacas semilenhosas de kiwi cultivar *Hayward* (Tabela 6). Nesse experimento os autores tinham um único fator: cinco doses de AIB (0, 2.000, 4.000, 6.000 e 8.000 ppm). De acordo com o objetivo do trabalho, o fator em estudo é quantitativo, uma vez que o interesse dos pesquisadores foi verificar os intervalos entre os níveis de AIB, ou seja, verificar quais as concentrações ideais de AIB no enraizamento das estacas. Para suas conclusões utilizaram teste de comparação de médias (Teste de Duncan ao nível de 5% de significância) e apesar de as concentrações 4.000, 6.000 e 8.000 não se diferenciarem, concluíram que a concentração 6.000 ppm proporcionou os melhores resultados de estacas

enraizadas. Como se objetiva a análise dos intervalos de doses, a aplicação de teste de comparação está incorreta. Portanto, o ajuste de uma equação de regressão para as diferentes concentrações de AIB é o adequado, podendo indicar que uma concentração entre o intervalo 4.000 ppm e 6.000 ppm foi a mais eficaz. Em comparações de tratamentos com esquema fatorial ou níveis de fatores quantitativos, os graus de liberdade ou a soma dos quadrados devem ser particionados para corresponder aos efeitos principais ou interações, sendo a regressão a técnica apropriada nesse caso; se a regressão for significativa, nenhuma comparação múltipla é necessária e todos os tratamentos, incluindo os intermediários não utilizados no experimento, são significativamente diferentes em seus efeitos (CHEW, 1976).

Tabela 6 - Percentagem de enraizamento de estacas semilenhosas de kiwi da cultivar *Hayward*<sup>(1)</sup>, coletadas em novembro.

Concentrações de AIB (ppm)	Enraizamento (%)
Zero	25,06 c
2.000	50,07 b
4.000	64,78 a
6.000	75,59 a
8.000	69,49 a

<sup>(1)</sup>Médias seguidas por letras iguais, não diferem significativamente entre si pelo Teste de Duncan ao nível de 5 % de probabilidade.

O segundo exemplo incorreto foi extraído do trabalho que visou avaliar a influência de diferentes concentrações de ácido indolbutírico e número de folhas no enraizamento do pessegueiro cv. Charme (Tabela 7). O delineamento experimental foi o inteiramente casualizado com 4 repetições e 8 estacas por parcela, em arranjo fatorial 4 x 3 (número de folhas x concentrações de AIB). No trabalho não foi observado nenhum efeito significativo da interação entre os fatores: número de folhas e concentrações de AIB. Sendo assim, os autores optaram pela comparação múltipla de

médias, onde utilizaram teste de Scott-Knott. Os fatores em estudo são de natureza qualitativa não estruturada (número de folhas) e quantitativa (concentrações de AIB). A priori, os autores acertaram em verificar se havia significância da interação. Porém, como essa não foi significativa, deveriam realizar a análise separada dos fatores, ou seja, aplicar teste de comparação de médias para o número de folhas e ajustar uma equação de regressão para as concentrações de AIB, uma vez que, o efeito de um fator não está relacionado com o do outro.

Tabela 7 - Porcentagem de estacas enraizadas de pessegueiro cv. Charme em função de concentrações de AIB e número de folhas <sup>(1)</sup>. Canoinhas, SC, 2004.

Número de Folhas	Concentração de AIB (mg L <sup>-1</sup> )		
	0	1000	2000
0	0,0 A a	3,1 A b	3,1 A b
2	0,0 C a	34,4 A a	15,6 B a
4	3,1 C a	37,5 A a	21,9 B a
8	6,2 C a	40,6 A a	18,7 B a
C.V (%)	21,8		

<sup>(1)</sup>Médias seguidas pela mesma letra maiúscula, na linha, e minúscula, na coluna, não diferem entre si pelo teste de Scott-Knott (p = 0,05).

## CONCLUSÕES

Existe dificuldade na escolha de quais os procedimentos corretos em relação aos tipos de fatores em estudo. Os principais aspectos que afetam diretamente a fidedignidade na interpretação dos resultados por parte dos pesquisadores são: i) a não consideração das pressuposições dos

diferentes testes estatísticos e; ii) conhecimento incipiente em relação aos tipos de fatores. Incoerências na escolha dos testes estatísticos acarretam em conclusões incompletas e/ou inapropriadas (SILVA, 1999).

## REFERÊNCIAS

- AFLAKPUI, G.K.S. Some uses/abuses of statistics in crop experimentation. **Tropical Science**, New Jersey, v. 35, n. 2, p. 347-353, 1995.
- ALVAREZ, V. H.; ALVAREZ, G.A.M. Comparação de médias ou teste de hipóteses? Contrastes! **Sociedade Brasileira de Ciências do Solo**. Viçosa, v. 3, n. 1, p. 24-33, 2006.
- BANZATTO, D.A.; KRONKA, S.N. **Experimentação Agrícola**. Jaboticabal: FUNEP, 2006. 237p.
- CARDELLINO, R.A.; SIEWERDT, F. Utilização adequada e inadequada dos testes de comparação de médias. **Revista da Sociedade Brasileira de Zootecnia**, Viçosa, v. 21, n. 6, p. 985-995, 1992.
- CHEW, V. Comparing treatment means: a compendium. **Hortscience**, Alexandria, v. 11, n. 4, p. 348-35, 1976.
- GILL, J.L. Current status of multiple comparisons of means in designed experiments. **Journal of Dairy Science**, Savoy, v. 56, n. 8, p. 973-978, 1973.
- GILL, J.L. **Design and analysis of experiments**. Ames: Iowa State University Press, 1978. 410p.
- HINKELMANN, K.; KEMPTHORNE, O. **Design and analysis of experiments: introduction to experimental design**. New York, JohnWiley & Sons, 1994. 495p.
- JONES, D. Use, misuse, and role of multiple comparison procedures in ecological and agricultural entomology. **Environmental Entomology**, Lanham, v. 13, n. 3, p. 635-649, 1984.
- KUEHL, R.O. **Statistical principles of research design and analysis**. Belmont, California: Duxbury Press, 1994. 686p.
- NOGUEIRA, M.C.S. **Estatística experimental aplicada à experimentação agrônômica**. Piracicaba: DME/ESALQ, 1997. 250 p.
- PETERSEN, G.R. Use and misuse of multiple comparison procedures. **Agronomy Journal**, Madison, v. 69, n. 2, p. 205--208, 1976.
- SANTOS, J.W.; MOREIRA, J.A.N.; BELTRÃO, N.E.M. Avaliação do emprego dos testes de comparação de médias na revista Pesquisa Agropecuária Brasileira (PAB) de 1980 a 1994. **Pesquisa Agropecuária Brasileira**, Brasília, v. 33, n. 3, p. 225-230, 1998.
- SILVA, E.C.; FERREIRA, D.F.; BEARZOTI, E. Avaliação do poder e taxas de erro tipo I do teste de Scott-Knott por meio do método de Monte Carlo. **Ciência e Agrotecnologia**, Lavras, v. 23, n. 3, p.687-696, 1999.
- SILVA, J.G.C. A consideração das estrutura das unidades em inferências derivadas do experimento. **Pesquisa Agropecuária Brasileira**, Brasília, v. 34, n. 6, p.911-925, 1999.
- WECHSLER, F.S. Fatoriais fixos desbalanceados: uma análise mal compreendida. **Pesquisa Agropecuária Brasileira**, Brasília, v. 33, n. 3, p. 231-262, 1998.