

BEZERRA NETO, F.; NUNES, G.H.S.; NEGREIROS, M.Z. Avaliação de procedimentos de comparações múltiplas em trabalhos publicados na revista Horticultura Brasileira de 1983 a 2000. *Horticultura Brasileira*, Brasília, v. 20, n. 1, p. 05-09, março, 2.002.

## Avaliação de procedimentos de comparações múltiplas em trabalhos publicados na revista Horticultura Brasileira de 1.983 a 2.000.

Francisco Bezerra Neto; Glauber Henrique S. Nunes; Maria Zuleide de Negreiros.

ESAM, Depto. Fitotecnia, Km 47 BR 110, C. Postal 137, 59.625-900, Mossoró-RN; E-mail: bezerra@esam.br

### RESUMO

Vários procedimentos de comparações múltiplas têm sido usados para explicar a resposta de tratamentos testados em experimentos agrônomicos. Entre esses, está o uso de testes de comparação de médias, muitas vezes apropriados a alguns tratamentos e inteiramente inapropriados a outros. Um levantamento e uma avaliação do uso e mau uso desses testes foi realizado com os trabalhos publicados na revista Horticultura Brasileira no período de 1.983 a 2.000. Procedimentos alternativos são sugeridos para situações onde os testes de comparação de médias não são apropriados. Dos 294 trabalhos analisados, algum tipo de teste foi utilizado para a comparação de médias. Em 65,6% dos artigos consultados, os testes como o de Tukey, Duncan e a DMS de Fisher foram adequadamente aplicados com relação ao tipo de dados envolvidos. Em 22,8% dos artigos, esses testes estavam parcialmente corretos e nos 11,6% dos artigos restantes estavam incorretamente aplicados ao tipo de tratamento utilizado na pesquisa.

**Palavras-chave:** Tratamentos qualitativos, tratamentos quantitativos, contrastes, procedimentos de comparações múltiplas, técnicas de regressão.

### ABSTRACT

**Assessment of multiple-comparison procedures of published articles in the journal Horticultura Brasileira from 1,983 to 2,000.**

Several procedures of multiple-comparison have been used to explain the response of treatments applied in experimental units of agronomic experiments. Among these are the mean comparison tests, sometimes appropriate for some kinds of treatments and completely inappropriate for others. The use and misuse of these tests were surveyed and assessed in the articles published in the journal Horticultura Brasileira from 1,983 to 2,000. Alternative procedures are suggested for situations in which mean comparison tests are inappropriate. We observed that in 294 articles some kind of test was used for mean comparison. In 65,6% of these articles multiple comparison tests (as Tukey, Duncan and LSD) were appropriately used for the type of involved data. In another 22,8% tests were used in a partially correct form. For the 11,6% remaining articles multiple comparison tests were inappropriately used for the type of treatments used in the research.

**Keywords:** Qualitative treatments, quantitative treatments, contrasts, multiple comparison tests, regression techniques.

(Aceito para publicação em 07 de dezembro de 2.001)

Em experimentos agrônomicos, o objetivo da análise estatística dos dados é fornecer tantas informações quanto possíveis a respeito da maneira como as unidades experimentais respondem aos tratamentos aplicados. Para isso, é muito comum submeter os dados a uma análise de variância (ANOVA) para saber se existem ou não diferenças significativas entre as médias dos tratamentos aplicados. Feita a análise, a tentativa é explicar a resposta desses tratamentos de maneira mais detalhada. Vários procedimentos podem ser usados com esse propósito, como o ajustamento de funções de resposta, usando as técnicas de regressão (Little, 1981; Dawkins, 1983), os conjuntos de contrastes planejados entre as médias ou grupos de médias (Gill, 1973; Jones, 1984; Swallow, 1984; Klockars & Sax, 1990), os procedimentos de comparações múltiplas de médias (Chew, 1976; Carmer & Walker, 1985) ou o método aglomerativo, tal como o proposto por Scott & Knott (1974).

Nenhum desses procedimentos deve ser usado indiscriminadamente, pois, alguns deles são apropriados a certos tipos de tratamentos e inteiramente inapropriados a outros. Um dos mais frequentes e incorretamente utilizados é o teste de comparação múltipla de médias (Petersen, 1977; Aflakpui, 1995). O uso inapropriado de um teste pode conduzir o pesquisador a interpretações equivocadas dos resultados de um experimento bem como a tirar conclusões erradas.

Os tipos de experimentos para os quais os procedimentos de comparações múltiplas de médias são apropriados, são aqueles cujo objetivo é determinar os melhores tratamentos dentro de um con-

junto qualitativo de tratamentos. Vários procedimentos são possíveis. Um deles é quando os contrastes (que envolvem grupos de médias) são definidos após prévia inspeção dos dados, sendo sua significância verificada usando o teste de scheffé (Banzatto & Kronka, 1995). Quando os contrastes são definidos a posteriori e são ortogonais o procedimento adequado para verificar as suas significâncias seria o teste F ou equivalente ao teste t (Gill, 1973).

Porém, quando os contrastes não são ortogonais e que o interesse é comparar todos os pares de médias entre si, neste caso deve-se escolher um procedimento de comparação múltipla tal como o de Tukey (Carmer & Walker, 1985; Percin & Malheiros, 1989). Se o número de tratamentos é grande e há o interesse numa separação real de grupos de médias, sem a ambiguidade de resul-

**Tabela 1.** Classificação dos trabalhos publicados na revista Horticultura Brasileira no período de 1.983 a 2.000, quanto ao uso de testes de comparação de médias. Mossoró, ESAM, 2000.

Testes	Uso dos testes de comparação de médias							
	Adequado		Parc. adequado		Inadequado		Total	
Usados	Nº	%	Nº	%	Nº	%	Nº	%
Tukey	103	61,3	41	24,4	24	14,3	168	57,1
Duncan	62	64,6	25	26,0	9	9,4	96	32,6
Skott-Knott	10	100,0	0	0,0	0	0,0	10	3,4
DMS	7	77,8	1	11,1	1	11,1	9	3,1
SNK	2	100,0	0	0,0	0	0,0	2	0,7
Dunnett	2	100,0	0	0,0	0	0,0	2	0,7
Não informado	7	100,0	0	0,0	0	0,0	7	2,4
Total	193		67		34		294	
Porcentagem		65,6		22,8		11,6		100,0

tados, o procedimento apropriado seria o teste de Scott Knott (Ferreira *et al.*, 1999; Silva *et al.*, 1999; Ramalho *et al.*, 2000). Se o interesse é comparar a média de qualquer tratamento com a média do tratamento testemunha (controle), o procedimento adequado seria usando o teste de Dunnett (Lentner & Bishop, 1986; Christensen, 1996).

Cardelino & Siewerdt (1992), revisando 260 trabalhos publicados na Revista da Sociedade Brasileira de Zootecnia (SBZ) de 1984 a 1989, constataram que 24,6% dos testes de comparação de médias estavam corretos, 11,2% parcialmente corretos e 64,2% incorretos. Santos *et al.* (1998), em levantamento realizado em 628 trabalhos publicados na revista Pesquisa Agropecuária Brasileira (PAB) de 1980 a 1994, observaram que os testes de comparação de médias foram adequados em 57,0%, parcialmente adequados em 11,5% e inadequados em 35,5%.

Este trabalho teve como objetivo fazer uma avaliação dos procedimentos de comparações múltiplas de médias empregados na análise dos trabalhos publicados na revista Horticultura Brasileira, de 1.983 a 2.000, além de apresentar alternativas adequadas aos procedimentos inapropriadamente usados.

## MATERIAL E MÉTODOS

Uma revisão dos procedimentos de comparações múltiplas de médias foi realizada nos trabalhos publicados na revista Horticultura Brasileira (HB) de

1983 a 2000, correspondendo aos volumes de 01 a 18. Esses procedimentos quanto à aplicação de testes foram classificados em “adequado”, “parcialmente adequado” ou “inadequado”. Foi considerado “adequado” quando o teste de comparação de médias consistia no procedimento mais apropriado aos tratamentos de natureza qualitativa e não estruturados. Foi considerado “parcialmente adequado” quando o teste de comparação de médias consistia no procedimento apropriado a tratamentos estruturados (tratamentos de um conjunto formado pela adição de um ou mais fatores, muito comuns nas áreas de fitotecnia e de sementes), a experimentos fatoriais (onde se compara todos os tratamentos entre si, dois a dois), e a situações onde, após ajustar uma equação de regressão aos dados, ainda é utilizado um teste de comparação múltipla. Foi considerado “inadequado” quando o teste de médias foi aplicado a tratamentos de natureza quantitativa ou, ainda, em experimentos fatoriais, em médias marginais dos fatores, sem levar em conta possíveis interações entre os efeitos principais.

Uma distribuição de frequência absoluta e percentual foi realizada para as três categorias dos testes de comparação de médias. Entre os testes levantados estão a DMS de Fisher (teste ‘t’ de amplitude fixa), o de Tukey, o de Duncan, o de SNK (Student-Newman-Keuls), o de Scott-Knott e o de Dunnett. Nos trabalhos em que os autores não informaram o tipo de comparação múltipla empregado, o teste foi classificado como “Não informado”.

tipla empregado, o teste foi classificado como “Não informado”.

## RESULTADOS E DISCUSSÃO

A distribuição de frequência absoluta e percentual dos testes de comparação de médias aplicados aos trabalhos publicados na revista Horticultura Brasileira no período de 1.983 a 2.000 estão apresentados na Tabela 1. Nesses dezoito anos, 294 trabalhos foram analisados, dos quais 193 (correspondendo a 65,6%) foram classificados como “adequados”, com relação à aplicação dos testes de comparação de médias, 67 (correspondendo a 22,8%) foram considerados como “parcialmente adequados” e 34 (correspondendo a 11,6%) foram classificados como “inadequados”. Entre os testes utilizados, os mais frequentes foram o de Tukey (57,1%) e o de Duncan (32,6%). Estes resultados estão em concordância de certa maneira com os obtidos por Cardelino & Siewerdt (1992) que obtiveram percentuais em torno de 63% e 25%, para os teste de Tukey e de Duncan, respectivamente. Por outro lado, os percentuais encontrados estão mais próximos dos obtidos por Santos *et al.* (1998), que encontraram resultados próximos a 59% e 32%, para os respectivos testes.

Observando-se os procedimentos de forma individual, dentro da categoria “adequado”, verificou-se que 61,3% dos testes de Tukey, 64,6% dos testes de Duncan e 77,8% da DMS de Fisher fo-

**Tabela 2.** Porcentagem de germinação de sementes de *Maytenus ilicifolia* após tratamentos de secagem e armazenamento.

Tratamentos	Médias* (%)
T <sub>1</sub> : Testemunha	100 a
T <sub>2</sub> : Secagem (22°C, 15 % UR)	82 b
T <sub>3</sub> : Armazenamento (5°C)	93 a b
T <sub>4</sub> : Armazenamento (-20°C)	85 a b
T <sub>5</sub> : Secagem + Armazenamento (5°C)	82 b
T <sub>6</sub> : Secagem + Armazenamento (-20°C)	75 b

\* Médias seguidas pela mesma letra não diferem entre si, pelo teste de Tukey a 5% de probabilidade.

**Tabela 3.** Efeito do nível de adubação orgânica sobre o rendimento de óleo essencial de *Lipia Alba*.

Níveis de adubação orgânica (kg/m <sup>2</sup> )	Rendimento de óleo essencial (ml/g x 100%)*
0	0,3395 a
1	0,3091 a
2	0,2940 ab
4	0,2696 bc
8	0,2642 c

\* Médias seguidas pela mesma letra não diferem entre si, pelo teste de Tukey, a 5% de probabilidade.

**Tabela 4.** Efeito do teor de cinza vegetal sobre peso médio da cabeça de alface, cv. Brasil 303.

Níveis de cinza vegetal (t/ha)	Peso médio (g)*
0	170,2 b
10	265,8 a
15	262,9 a
20	233,3 a
30	220,4 a

\* Médias seguidas pela mesma letra não diferem entre si, pelo teste de Duncan, a 5% de probabilidade.

ram utilizados de forma apropriada. Para os outros tipos de teste estas porcentagens foram 100%. Ainda dentro dessa categoria, em sete trabalhos de competição de cultivares, usou-se procedimentos de comparações múltiplas apropriados para identificar as melhores cultivares, porém observou-se o descuido dos pesquisadores em não mencionarem o tipo de procedimento usado nos ensaios (Tabela 1, “Não informado”), indispensável na elaboração da tomada de decisão.

O uso mais comum dos procedimentos de comparações múltiplas de médias é efetuar comparações de cada uma das médias com cada uma das outras, com o objetivo de detectar possíveis

grupos entre um conjunto de tratamentos não estruturados. Um exemplo que ilustra esta categoria foi extraído de um trabalho em que se comparou a produção de nove cultivares de cebola (*Allium cepa* L.), através do teste de Tukey, identificando-se quatro grupos. Em outro exemplo, avaliando a produtividade de vinte linhagens de abóbora (*Curcubita moschata Duchesne*), através do teste Scott-Knott, os autores identificaram dois grupos.

Foram incluídos na categoria “parcialmente adequado”, 24,4% dos trabalhos onde se empregou o teste de Tukey, 26% dos trabalhos onde se empregou o teste de Duncan e 11,1% dos trabalhos onde se empregou a DMS de Fisher. A

aplicação deste tipo de procedimento é comum em tratamentos estruturados (tratamentos formados pela adição de um ou mais fatores) ou em experimentos fatoriais onde se comparam todos os tratamentos entre si, dois a dois ou em situações após ajustar uma equação de regressão, ainda se utiliza um teste de comparação de médias.

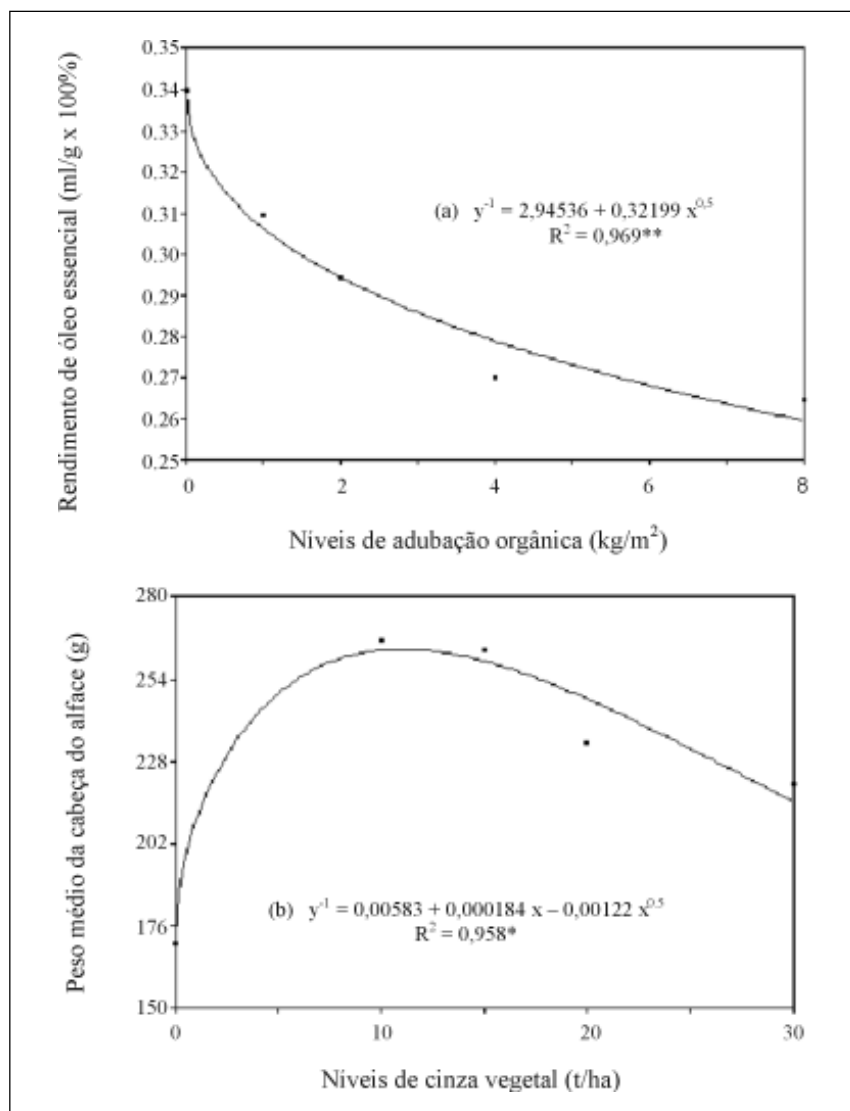
O exemplo da Tabela 2 ilustra a situação, onde os pesquisadores avaliaram a porcentagem de germinação de sementes de *Maytenus ilicifolia* após tratamentos de secagem e armazenamento. Embora parcialmente correta a aplicação do teste de Tukey, os autores poderiam ter formulado contrastes importantes como:  $Y_1 = (m_3 + m_4) - 2m_1$  (armazenamento x testemunha);  $Y_2 = m_3 + m_6 - m_3 - m_4$  (secagem e armazenamento x armazenamento);  $Y_3 = m_5 + m_6 - 2m_2$  (secagem e armazenamento x secagem). Outra situação está ilustrada em um experimento fatorial 2 x 2 x 3, onde os autores estudaram a influência de duas fontes de potássio, duas doses de cloreto de sódio e três doses de potássio sobre o teor de nitrogênio da matéria seca da parte aérea de tomateiro. Pode-se constatar que, o emprego do teste de Duncan é parcialmente apropriado, embora o procedimento correto, no caso de nenhuma interação significativa, fosse comparar as médias marginais dos fatores principais (caso de natureza qualitativa) ou ajustar uma equação de regressão a eles (caso de natureza quantitativa). No caso de interação significativa, deve-se comparar médias dos níveis de um fator (de natureza qualitativa) dentro dos níveis de outro fator. Por exemplo, comparar as médias dos efeitos das fontes de potássio dentro de cada dose de cloreto de sódio. O ajustamento de uma equação de regressão aos teores de nitrogênio em função das doses de potássio dentro de cada dose de cloreto de sódio também seria um procedimento correto, apesar de apenas três níveis para a estimação dos parâmetros da equação. Quando este número de níveis é menor do que quatro pode-se ter problema no processo de estimação dos parâmetros.

Dentro da categoria “inadequada”, registrou-se que, nos trabalhos onde foi empregado o teste de Tukey, 14,3%

utilizaram-no de forma inapropriada, seguido pela DMS de Fisher com 11,1% e pelo teste de Duncan com 9,4% (Tabela 1). Esses testes foram responsáveis pelos 11,6% do uso inadequado dos testes de comparação de médias nos trabalhos publicados na revista *Horticultura Brasileira*.

Nas Tabelas 3 e 4 estão ilustrados exemplos de má aplicação dos testes de comparações múltiplas a tratamentos quantitativos (como níveis de adubação orgânica e níveis de cinza vegetal). O procedimento estatístico mais correto para essas situações, seria examinar os dados das variáveis observadas através de uma relação funcional com os níveis dos tratamentos aplicados. As equações resultantes proporcionam uma descrição das variações ocorridas nas variáveis observadas em função dos níveis dos tratamentos aplicados, permitindo estudar qualquer nível intermediário dos tratamentos, mesmo que eles não estejam diretamente incluídos no estudo (Figura 1). Por exemplo, na Figura 1 (b), pode-se identificar o nível de cinza vegetal que proporciona o maior peso médio (cerca de 263 g) de cabeça de alface, que é de aproximadamente 11 t/ha.

As causas prováveis do mau uso desses testes podem estar associadas ao desconhecimento de procedimentos alternativos aos testes de comparações múltiplas de médias, como a técnica de análise de regressão, bem como a falta de conhecimento das condições de uso adequado desses testes aos tipos de dados estudados. Além disso, pode também ser devido à falta de habilidade dos pesquisadores na interpretação dos resultados, podendo levá-los a fazer inferências errôneas acerca dos tratamentos investigados. Cardellino & Siewerdt (1992) e Pearce (1993) advertem que deficiências como estas podem levar o pesquisador à simplificação do uso de um teste de comparação múltipla, mesmo quando ele não se constitui na melhor opção para a análise dos dados experimentais. Por outro lado, Matos (1993) evidencia que a formação acadêmica da maioria dos pesquisadores tem sido baseada em receitas estatísticas com muita ênfase na parte matemática e pouca ou nenhuma consideração na adequação dos métodos ou na interpretação dos resultados obtidos.



**Figura 1.** (a) Rendimento de óleo essencial em função de diferentes níveis de adubação orgânica e (b) Peso médio da cabeça de alface em função de diferentes níveis de cinza vegetal.

De modo geral, pode-se observar a verificação pouca cuidadosa da adequabilidade dos testes de comparações múltiplas aos tipos de tratamentos considerados na análise. Além disso, nota-se que o uso inadequado dos testes é comum, principalmente, nos casos nos quais estão envolvidos tratamentos de natureza quantitativa. O uso indiscriminado desses testes pode resultar em perda de informação e eficiência reduzida quando procedimentos mais adequados estão disponíveis. Por fim, ressalta-se que o pesquisador deve consultar, sempre que possível, um estatístico quando existir dúvidas na escolha do teste a ser aplicado, de modo que os dados possam ser explorados de manei-

ra correta e conseqüentemente, os resultados advindos da pesquisa possam auxiliar nas tomadas de decisões.

## LITERATURA CITADA

- AFLAKPUI, G.K.S. Some uses/abuses of statistics in crop experimentation. *Tropical Science*, v. 35, n. 2, p. 347-353, 1995.
- CARDELLINO, R.A.; SIEWERDT, F. Utilização correta e incorreta dos testes de comparação de médias. *Revista da Sociedade Brasileira de Zootecnia*, Viçosa, v. 21, n. 6, p. 985-995, 1992.
- CARMER, S.G.; WALKER, W.M. Pairwise multiple comparisons of treatment means in agronomic research. *Journal of Agronomic Education*, v. 14, n. 1, p. 19-26, 1985.
- CHEW, V. Comparing treatment means: a compendium. *Hortscience*, v. 11, n. 4, p. 348-357, 1976.

- CHRISTENSEN, R. *Analysis of variance, design and regression*. London: Chapman & Hall, 1996. 587 p.
- DAWKINS, H.C. Multiple comparisons misused: Why so frequently in response curve studies? *Biometrics*, v. 39, n. 3, p. 789-790, 1983.
- FERREIRA, D.F.; MUNIZ, J.A.; AQUINO, L.H. Comparações múltiplas em experimentos com grande número de tratamentos – utilização do teste de Scott Knott. *Ciência e Agrotecnologia*, Lavras, v. 23, n. 3, p. 745-752, 1999.
- GILL, J.L. Current status of multiple comparisons of means in designed experiments. *Journal of Dairy Science*, v. 56, n. 8, 1973.
- JONES, D. Use, misuse, and role of multiple comparison procedures in ecological and agricultural entomology. *Environmental Entomology*, v. 13, n. 3, p. 635-649, 1984.
- KLOCKARS, A.J.; SAX, G. *Multiple comparisons*. 3 ed. California: SAGE, 1990. 87p.
- LENTNER, M.; BISHOP, T. *Experimental design and analysis*. Blacksburg, VA: Valley Book Company, 1986. 565 p.
- LITTLE, T.M. Interpretation and presentation of results. *Hortscience*, v. 16, n. 5, p. 637-640, 1981.
- MATOS, L.L. O pesquisador, o estatístico e a sociedade. In: REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE ZOOTECNIA, 30, 1993, Rio de Janeiro, RJ. *Anais ...* Rio de Janeiro: SBZ, 1993. p. 121.
- PEARCE, S.C. Data analysis in agricultural experimentation. III. Multiple comparisons. *Experimental Agriculture*, v. 29, n. 1, p. 1-8, 1993.
- PERECIN, D.; MALHEIROS, E.B. *Procedimentos para comparações múltiplas*. Lavras: UFLA, 1989. 67p.
- PETERSEN, G.R. Use and misuse of multiple comparison procedures. *Agronomy Journal*, v. 69, n. 2, p. 205-208, 1977.
- RAMALHO, M.A.P.; FERREIRA, D.F.; OLIVEIRA, A.C. *Experimentação em genética e melhoramento de plantas*. Lavras, UFLA, 2000. 326 p.
- 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.
- SCOTT, A.J.; KNOTT, M.A. A cluster analysis method for grouping means in the analysis of variance. *Biometrics*, v. 30, n. 2, p. 507-512, 1974.
- 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.
- SWALLOW, W.H. Those overworked and oft-misused mean separation procedures – Duncan's, LSD, etc. *Plant Disease*, v. 68, n. 10, p. 919-921, 1984.