

**UTILIZAÇÃO DE TESTES DE COMPARAÇÃO DE
MÉDIAS NA ANÁLISE DE EXPERIMENTOS
UNIFATORIAIS COM TRATAMENTOS
QUANTITATIVOS**

Josiane Rodrigues e Sônia Maria De Stefano Piedade

UTILIZAÇÃO DE TESTES DE COMPARAÇÃO DE MÉDIAS NA ANÁLISE DE EXPERIMENTOS UNIFATORIAIS COM TRATAMENTOS QUANTITATIVOS

Josiane Rodrigues¹, Sônia Maria De Stefano Piedade²

¹ Departamento de Tecnologia Agroindustrial e Socioeconomia Rural, Centro de Ciências Agrárias – CCA, Universidade Federal de São Carlos – UFSCar, Rodovia Anhanguera, km 174, CEP 13600-970, Araras, SP, Brasil. E-mail: josirodrigues@ufscar.br

² Departamento de Ciências Exatas, Escola Superior de Agricultura “Luiz de Queiroz” – ESALQ/USP. Av. Pádua Dias, 11, Caixa Postal 9, CEP 13418-900, Piracicaba, SP, Brasil. E-mail: soniamsp@usp.br

RESUMO

O presente trabalho teve por finalidade verificar se testes de comparação de médias são comumente empregados na análise de experimentos cujos tratamentos são representados por níveis de um fator quantitativo, bem como as eventuais vantagens e/ou limitações da aplicação dos testes na análise desses experimentos quando comparados à análise de regressão. De acordo com os objetivos propostos, foi realizado um levantamento acerca de artigos cujo objetivo principal foi o de fazer uma investigação de trabalhos científicos publicados em jornais, revistas ou periódicos – nos quais se utilizou algum procedimento de comparação de médias – verificando a adequação desses testes às análises estatísticas realizadas. Com isso buscou-se verificar se testes de médias são comumente empregados na análise de experimentos com tratamentos quantitativos, bem como levantar a opinião de pesquisadores sobre o assunto. Em seguida, foi feita a análise de um experimento que comparou cinco diferentes doses de pentóxido de difósforo na produção de milho, por meio da análise de regressão e da aplicação de um teste para comparação múltipla das médias dos tratamentos, de modo a verificar as eventuais diferenças nos resultados obtidos com cada uma delas. A revisão dos artigos demonstrou que um número significativo de pesquisadores utiliza de procedimentos de comparações múltiplas em análises estatísticas de experimentos com tratamentos quantitativos, o que é considerado por alguns como um procedimento inadequado. Na análise do experimento, por sua vez, ficou claro que o uso de procedimentos de comparações múltiplas na análise de experimentos envolvendo tratamentos quantitativos pode resultar na redução de informações e da eficiência dos resultados. Assim sendo, conclui-se que o uso dos testes na análise desses experimentos não é um procedimento errôneo, mas sim menos informativo e, portanto, menos eficiente quando comparado à análise de regressão.

Palavras-chave: Análise estatística de dados, análise de variância, análise de regressão

1. INTRODUÇÃO

Em experimentos agronômicos, o objetivo de uma análise estatística dos dados é conhecer de que forma as unidades experimentais respondem aos tratamentos que são aplicados a elas, de forma a verificar se existem ou não diferenças entre as médias dos tratamentos em estudo, além de encontrar também, no caso dessa diferença ser significativa, qual(is) o(s) tratamento(s) produz(em) a resposta desejada.

O primeiro passo na análise estatística dos dados geralmente consiste em submetê-los à análise de variância (ANOVA), procedimento este que compara a variação devida aos tratamentos com a variação devida ao acaso, ou resíduo. Se os tratamentos do ensaio são de efeito fixo, a hipótese testada na ANOVA por meio do teste F é a de igualdade entre as médias populacionais dos tratamentos comparados:

$$H_0: \mu_1 = \mu_2 = \dots = \mu_I.$$

Caso o teste F seja significativo, então vão existir evidências para não aceitar a hipótese de igualdade entre as médias populacionais dos tratamentos. Quando isso ocorre, o passo seguinte da pesquisa é então determinar quais são as médias que diferem entre si (CHEW, 1976). De acordo com Nogueira (2007), existem dois procedimentos possíveis nessa análise continuada dos dados, de forma a comparar os tratamentos em estudo: a aplicação da análise de regressão, ou a aplicação de testes para a comparação das médias dos tratamentos. Os testes de médias, segundo a autora, devem ser aplicados quando os tratamentos envolvidos no experimento são qualitativos e, a análise de regressão, quando eles são quantitativos.

A análise de regressão consiste em ajustar uma curva aos dados obtidos em um experimento, curva essa que descreve, quando significativa, uma relação entre variável(is) explanatória(s) e variável resposta. Os testes de comparação de médias, por sua vez, também permitem comparar as médias de tratamentos. Esses testes, segundo Cardellino e Siewerdt (1992), consistem em fazer a comparação de todas as médias de tratamentos entre si, no caso de tratamentos qualitativos não estruturados, por meio de testes de comparações múltiplas (comparações duas a duas) ou, no caso de tratamentos qualitativos estruturados, por meio da aplicação de contrastes ortogonais para a comparação entre grupos de médias. Aqui, entendemos por tratamentos estruturados os casos em que no conjunto deles apareçam tratamentos formados pela adição de um ou mais fatores.

Diferentemente do que acontece na análise de regressão, na qual não há como empregá-la para analisar experimentos cujos tratamentos são de natureza qualitativa, o uso de testes de comparação de médias é também utilizado por pesquisadores para analisar experimentos cujos tratamentos são níveis de um fator quantitativo. Dal'Col Lúcio et al. (2003) relataram em sua pesquisa que testes de comparação de médias são também utilizados por pesquisadores para analisar experimentos cujos tratamentos são quantitativos. Entretanto, a aplicação de testes de médias nestes casos pode ser vista como inadequada. De acordo com Souza et al. (2012), por exemplo, os testes de médias de tratamentos são de grande interesse na pesquisa aplicada, mas quando o objetivo do trabalho é comparar tratamentos que são qualitativos.

Assim sendo, os objetivos do presente trabalho são i) realizar um estudo acerca do uso de testes de médias em pesquisas científicas, de forma especial na análise de experimentos que trabalham com níveis de um fator quantitativo, verificando se eles são empregados, de uma maneira geral, na análise desses experimentos, assim também como levantar a opinião de

pesquisadores sobre o assunto e ii) analisar dados de um experimento cujos tratamentos são níveis de um fator quantitativo por meio da análise de regressão e da aplicação de um teste de comparação múltipla das médias, verificando as eventuais vantagens e/ou limitações de cada uma dessas análises, fazendo assim uma reflexão acerca do tema.

2. MATERIAL E MÉTODOS

2.1 APLICAÇÃO DE TESTES DE MÉDIAS EM ARTIGOS CIENTÍFICOS

Dentro dos objetivos propostos foi feita uma revisão acerca de artigos cujo intuito principal foi o de fazer um levantamento de trabalhos científicos publicados em jornais, revistas ou periódicos – nos quais se utilizou algum procedimento de comparação de médias – verificando a adequação desses testes às análises estatísticas realizadas, e, conseqüentemente, identificando quais as principais adequações e inadequações na aplicação desses testes em trabalhos científicos.

Para tanto foram estudados os trabalhos de Petersen (1977), Cardellino e Siewerdt (1992), Santos et al. (1998), Bezerra Neto et al. (2002), Bertoldo et al. (2008a) e Bertoldo et al. (2008b). Cada um desses trabalhos teve como objetivo realizar um levantamento de artigos científicos que utilizaram procedimentos de comparação de médias em suas análises, bem como classificar este uso.

Nesse estudo se buscou saber a proporção em que procedimentos de comparações múltiplas são empregados para se analisar experimentos unifatoriais com tratamentos quantitativos, além da opinião dos autores diante desse tema. Além disso, buscou-se criar uma unificação das categorias utilizadas para classificação do uso dos testes de médias, os quais, de modo geral, podem ser divididas em: adequado, parcialmente adequado e inadequado.

Não é intenção do estudo, no entanto, criticar os trabalhos que de alguma forma fizeram uso indiscriminado de algum procedimento de comparação de médias, mas sim procurar melhorar as análises estatísticas de experimentos científicos, de forma particular aqueles cujos tratamentos são representados por níveis de um fator quantitativo, o que contribui para a confiabilidade acerca dos resultados e das inferências que são feitas dos tratamentos investigados.

2.2. TESTES DE MÉDIAS X ANÁLISE DE REGRESSÃO

Os dados da Tabela 1 se referem a um experimento de produção de milho feita pelos engenheiros agrônomos Glauco Pinto Viegas e Erik Smith (PIMENTEL, 1990), segundo o delineamento em blocos ao acaso. Os tratamentos constaram de adubação com 0, 25, 50, 75 e 100 kg de pentóxido de difósforo (P_2O_5) por hectare.

A fim de comparar as técnicas da análise de regressão e de testes de comparação de médias na análise de experimentos unifatoriais com tratamentos quantitativos, os dados da Tabela 1 foram submetidos inicialmente à ANOVA e, posteriormente, a aplicação da análise de regressão e ao teste de Tukey, ao nível de significância de 5%. Todas as análises estatísticas foram feitas no *software* R (R CORE TEAM, 2019).

Tabela 1. Produção de milho (kg/parcela) de acordo com o nível de adubação de P_2O_5 (kg/ha) utilizado.

Bloco	Tratamento				
	0	25	50	75	100
I	3,38	7,15	10,07	9,55	9,14
II	5,77	9,78	9,73	8,95	10,17
III	4,90	9,99	7,92	10,24	9,75
IV	4,54	10,10	9,48	8,66	9,50

3. RESULTADOS E DISCUSSÃO

3.1 APLICAÇÃO DE TESTES DE MÉDIAS EM ARTIGOS CIENTÍFICOS

Cada um dos autores dos artigos científicos estudados utilizou categorias para classificar os procedimentos de comparação de médias dos artigos analisados, as quais, de uma maneira geral, estão apresentadas a seguir.

Foi considerado “adequado” ou “correto” o procedimento no qual a aplicação do teste de médias foi feita para tratamentos de natureza qualitativa e não estruturados, ou seja, não relacionados. Foi considerado “parcialmente adequado” “ou parcialmente correto” os casos em que algum teste de comparação múltipla (todos contra todos) foi utilizado na análise de experimentos com tratamentos qualitativos estruturados, em que o mais adequado seria o teste de contrastes previamente planejados. Também se enquadram dentro dessa categoria os experimentos fatoriais, em que se comparou todos os tratamentos entre si, dois a dois, e, por fim, os casos em que, mesmo após se ajustar uma equação de regressão a dados de natureza quantitativa, ainda foi aplicado um teste de comparação de médias. Por fim, foi considerado “inadequado” ou “incorreto” quando o teste de comparação de médias foi aplicado para comparar níveis de um fator quantitativo, ou em experimentos fatoriais, nas médias marginais dos fatores principais, sem levar em conta possíveis interações entre eles.

Com base nas categorias estabelecidas, os autores puderam então quantificar a proporção de artigos em que o procedimento de comparação de médias utilizado foi adequado, parcialmente adequado ou inadequado, respectivamente (Tabela 2). Com exceção do levantamento realizado por Bezerra Neto et al. (2002), fica evidente a alta proporção de artigos científicos em que os testes de comparação de médias são aplicados de maneira inadequada. Fica claro ainda que, mesmo com o decorrer dos anos, os abusos na utilização dos testes continuaram muito frequentes.

Esses abusos, de acordo com Bezerra Neto et al. (2002), estão ligados ao desconhecimento do pesquisador com relação a outros procedimentos que não a aplicação de testes de comparação de médias que poderiam ser utilizados na análise dos dados. Ainda, segundo os autores, pode ser citado o desconhecimento dos pesquisadores no que diz respeito às condições de uso desses testes aos tipos de dados estudados e, por fim, a inabilidade na interpretação dos resultados obtidos com a aplicação dos testes.

Tabela 2. Proporção de artigos analisados de forma adequada, parcialmente adequada e inadequada, quanto ao uso de testes de comparação de médias.

Autor	Procedimento adequado	Procedimento parcialmente Adequado	Procedimento inadequado
Petersen (1976)	30,00%	30,00%	40,0%
Cardellino e Siewerdt (1992)	24,60%	11,20%	64,20%
Santos et al. (1998)	31,53%	11,46%	57,01%
Bezerra Neto et al. (2002)	65,60%	22,80%	11,60%
Bertoldo et al. (2008a)	27,27%	11,19%	61,54%
Bertoldo et al. (2008b)	31,73%	11,54%	56,73%

No que se refere aos procedimentos classificados como inadequados ou incorretos, verificou-se que a maioria se deve a aplicação dos testes de médias a experimentos com tratamentos quantitativos (BERTOLDO et al., 2008b). Nos experimentos fatoriais em que também se notou inadequação na aplicação dos testes de médias, ela está relacionada a uma deficiência na decomposição da interação entre os fatores principais.

Verificou-se, portanto, que o uso de procedimentos de comparação de médias na análise de experimentos com tratamentos quantitativos é considerado pelos autores como procedimento inadequado. De acordo com Petersen (1977), para experimentos nos quais os tratamentos são níveis de um fator quantitativo, um procedimento mais apropriado na análise dos dados é ajustar uma curva aos mesmos pelas técnicas de regressão, pois, de acordo com o autor, esse é um procedimento mais informativo nesses casos.

Segundo Bertoldo et al. (2008a), a escolha inadequada do teste a ser utilizado na análise dos dados pode resultar em afirmações incompletas ou errôneas. De acordo com Kozak (2009), isso não significa que um procedimento complexo deva ser aplicado na análise dos dados, mas sim um procedimento correto, que forneça resultados adequados, evitando desse modo falsas conclusões e interpretações. Assim sendo, Cardellino e Siewerdt (1992) sugeriram, de acordo com o tipo de fator que o trabalho abrange, alguns procedimentos para inferência do pesquisador, de modo a minimizar os equívocos na análise e interpretação de experimentos (Figura 1).

3.2. TESTES DE MÉDIAS X ANÁLISE DE REGRESSÃO

A ANOVA dos dados referentes à produção de milho, sob diferentes níveis de adubação, está apresentada na Tabela 3. Nota-se que o teste F foi significativo, ao nível de significância de 5%, indicando diferenças entre as doses de adubo utilizadas.

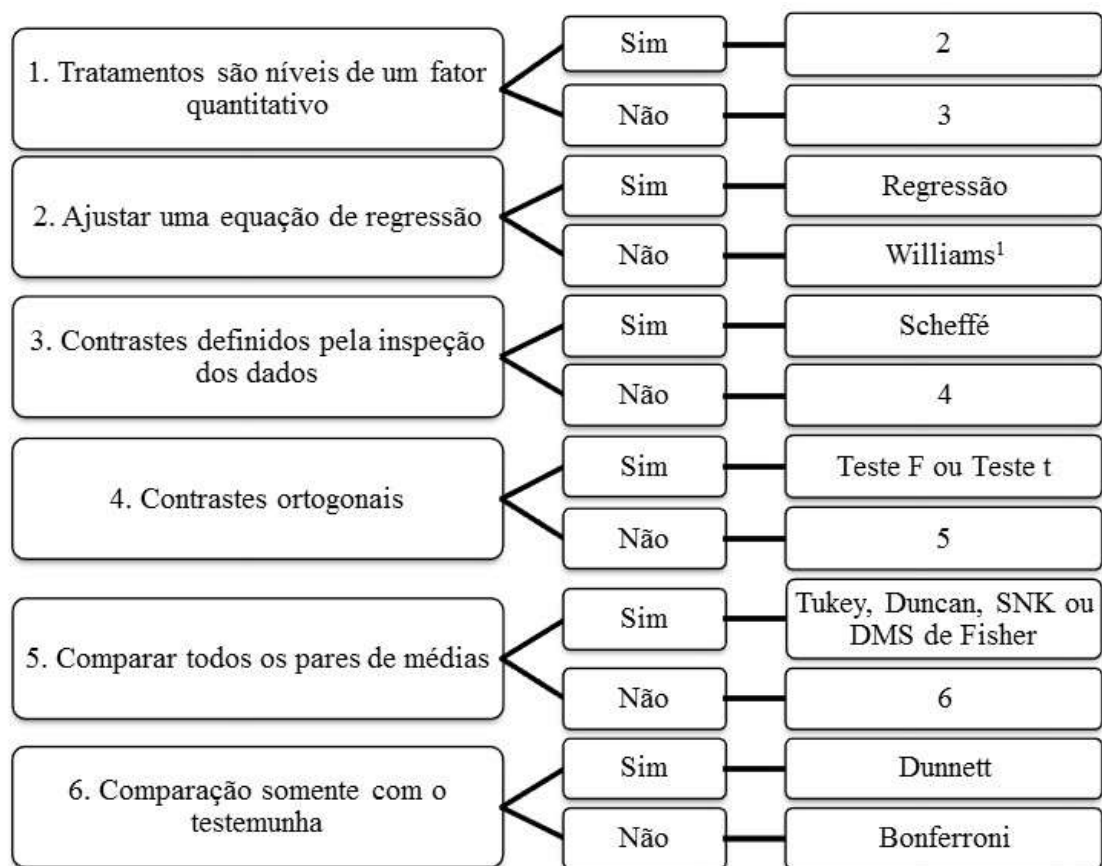


Figura 1. Chave para escolha do teste de comparação de médias (adaptado de Cardellino e Siewerdt (1992)).

Tabela 3. ANOVA dos dados referentes à produção de milho (kg/parcela) de acordo com o nível de adubação de P_2O_5 (kg/ha) utilizado.

Causas de variação	G.L.	S.Q.	Q.M.	F	Valor de p
Blocos	3	2,74	-	-	-
Tratamentos	4	72,22	18,06	19,85	<0,05
Resíduo	12	10,91	0,91	-	-
Total	19	85,87	-	-	-

A fim de investigar eventuais diferenças entre os tratamentos comparados, inicialmente os dados foram submetidos à análise de regressão. Para tanto, considerou-se a ANOVA com os desdobramentos para os componentes lineares e quadrático (Tabela 4). Uma regressão quadrática foi considerada uma vez que, no gráfico de dispersão dos dados, eles mostraram evidências de uma tendência não linear. Nota-se que o componente quadrático foi significativo, ao nível de 5%, o que justifica o ajuste de uma curva de segundo grau aos dados do experimento, curva esta representada na Figura 2.

Tabela 4. ANOVA dos dados referentes à produção de milho (kg/parcela) com desdobramento de graus de liberdade de tratamentos nos componentes lineares e quadráticos.

Causas de variação	G.L.	S.Q.	Q.M.	F	Valor de p
Componente linear	1	40,64	40,64	44,66	<0,05
Componente quadrático	1	21,28	21,28	23,38	<0,05
Desvios de regressão	2	10,3	5,15	5,66	<0,05
(Doses)	(4)	(72,22)	(18,06)	(19,85)	(<0,05)
Resíduo	12	10,91	0,91	-	-

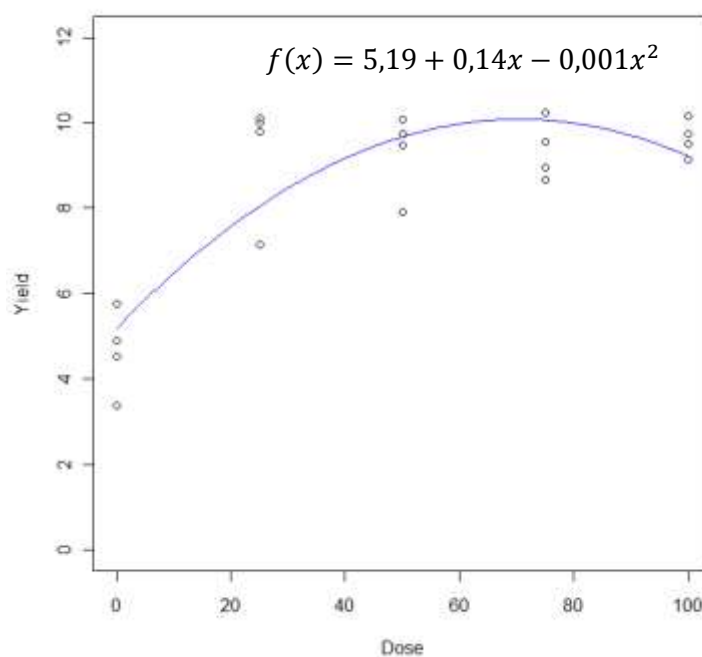


Figura 2. Curva ajustada aos dados de produção de milho (kg/parcela) de acordo com o nível de adubação de P_2O_5 (kg/ha) utilizado.

Investigando a função ajustada, é possível notar que em $x = 70$ ocorre um ponto de máximo. Assim sendo, temos que a dose de P_2O_5 que resulta na maior produção de milho é a de aproximadamente 70 kg/ha, que sequer é uma dose envolvida diretamente no experimento. Além disso, por meio da função ajustada é possível estudar também qual a produção esperada para uma determinada dose de adubo, bem como qual dose deve ser utilizada para que se obtenha uma determinada produção de milho.

Os mesmos dados foram também submetidos ao teste de Tukey, para comparação das médias de tratamentos duas a duas. Os resultados obtidos, ao nível de significância de 5%, estão apresentados na Tabela 5. Verifica-se que apenas a dose nula diferiu significativamente das demais quanto aos seus resultados, sendo que todas as demais foram consideradas equivalentes entre si.

Tabela 5. Teste de Tukey aplicado às médias de produção de milho (kg/parcela) de acordo com o nível de adubação de P_2O_5 (kg/ha) utilizado.

Dose	Média	Grupo
0	4,648	b
25	9,255	a
50	9,300	a
75	9,350	a
100	9,640	a

Médias seguidas pela mesma letra não diferem entre si, ao nível de significância de 5%.

A aplicação do teste de Tukey permitiu-nos, portanto, dentro da gama dos tratamentos aplicados diretamente no experimento, a separá-los em grupos dentro dos quais as respostas obtidas não foram significativamente diferentes. Assim sendo, concluiu-se de que apenas a dose nula, ou seja, o grupo controle, diferiu significativamente dos demais grupos. Logo, seria possível ao experimentador, diante desses fatos e diante do custo do produto, optar pelo uso da dose de 25 kg de P_2O_5 por hectare, visto que para a decisão do nível ótimo, deve ser considerada também a avaliação econômica do produto. Assim, verifica-se uma conclusão diferente daquela obtida com a aplicação da análise de regressão.

4. CONCLUSÕES

De acordo com a pesquisa realizada, fica evidente que o uso de testes de comparação de médias é frequente no caso em que os tratamentos envolvidos no experimento são representados por níveis de um fator quantitativo.

Fica claro ainda que, ao se utilizar procedimentos de comparação de médias na análise desses experimentos, a quantidade de informações extraída acerca dos mesmos é menor, tratando-se de procedimento menos informativo e, portanto, menos eficiente quando comparado à análise de regressão. Enquanto a análise de regressão, por meio da análise da função ajustada, permite estimar os níveis do fator nos quais a resposta é um máximo ou um mínimo mesmo que eles não estejam sendo testados diretamente no experimento, a aplicação de um teste de comparação de médias permite, apenas dentro da gama dos níveis aplicados diretamente no experimento, a escolher o(s) melhor(es) tratamento(s), dentro das expectativas da pesquisa. Não se trata, porém, de procedimento errôneo, desde que as interpretações e conclusões feitas sejam coerentes com o método utilizado na análise.

O que se pode dizer, portanto, no que diz respeito à análise de experimentos unifatoriais com tratamentos quantitativos, é o de que cada caso é um caso. Se o interesse do pesquisador está em estudar apenas os níveis do fator trabalhados diretamente no ensaio, então o teste de comparação de médias pode ser utilizado. No entanto, se o interesse do pesquisador está em

fazer um estudo mais detalhado do experimento, tal como colocado por Dal'Col Lúcio et al. (2003), em que se busca obter informações sobre tratamentos intermediários e pontos de máxima eficiência técnica ou econômica, por exemplo, então a análise de regressão é a alternativa mais adequada.

5. REFERÊNCIAS

BERTOLDO, J. G.; COIMBRA, J. L. M.; GUIDOLIN, A. F.; MANTOVANI, A.; VALE, N. M. do. 2008a. Problemas relacionados com o uso de testes de comparação de médias em artigos científicos. **Revista Biotemas**, v.21, n.2, p.145-153.

BERTOLDO, J. G.; COIMBRA, J. L. M.; GUIDOLIN, A. F.; MIQUELOTO, A.; TOALDO, D. 2008b. Uso ou abuso em testes de comparações de média: conhecimento científico ou empírico? **Ciência Rural**, Santa Maria, v.38, n.4, p.1145-1148.

BEZERRA NETO, F.; NUNES, G. H. S.; NEGREIROS, M. Z. de. 2002. 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.

CARDELLINO, R. A.; SIEWERDT, F. 1992. Utilização correta e incorreta dos testes de comparação de médias. **Revista da Sociedade Brasileira de Zootecnia**, Viçosa, v.21, n.6, p.985-995.

CHEW, V. 1976. Comparing treatment means: a compendium. **Hortscience**, Alexandria, v.11, n.4, p.348-357.

DAL'COL LÚCIO, A.; LOPES, S. J.; STORCK, L.; CARPES, R. H.; LIEBERKNECHT, D.; NICOLA, M. C. 2003. Características experimentais das publicações da *Ciência Rural* de 1971 a 2000. **Ciência Rural**, Santa Maria, v.33, n.1, p.161-164.

KOZAC, M. 2009. Analyzing one-way experiments: a piece of cake or a pain in the neck? **Scientia Agricola**, Piracicaba, v.66, n.4, p.556-562.

NOGUEIRA, M. C. S. 2007. **Experimentação Agrônômica I – conceitos, planejamentos e análise estatística**, Piracicaba: Esalq, 479p.

PETERSEN, R. G. 1977. Use and misuse of multiple comparison procedures. **Agronomy Journal**, Madison, v.69, n.2, p.205-208.

PIMENTEL GOMES, F. **Curso de Estatística Experimental**. 13.ed., Piracicaba: Livraria Nobel, 1990. 468p.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>

SANTOS, J. W. dos; MOREIRA, J. de A. N.; BELTRÃO, N. E. de M. 1998. 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.

SOUZA, C. A.; LIRA JUNIOR, M. A.; FERREIRA, R. L. C. 2012. Avaliação de testes estatísticos de comparações múltiplas de médias. **Revista Ceres**, Viçosa, v.59, n.3, p.350-354.