

Uso ou abuso em testes de comparações de média: conhecimento científico ou empírico?

Use or abuse in mean comparison tests: scientific or empiric knowledge?

Juliano Garcia Bertoldo¹ Jefferson Luís Meirelles Coimbra^{1*} Altamir Frederico Guidolin¹
Aquidauana Miqueloto¹ Diego Toaldo¹

-NOTA-

RESUMO

O objetivo deste trabalho foi identificar quais os principais erros e acertos na aplicação de testes de comparação de médias em trabalhos científicos, demonstrando alternativas viáveis no sentido de aumentar a imparcialidade dos resultados obtidos pelos pesquisadores. Um dos maiores desafios do pesquisador é a interpretação dos resultados de forma fidedigna. Apesar da preocupação com a análise dos dados, muitas vezes pode ser observado certo descaso com a interpretação dos resultados, em que a aplicação incorreta de testes estatísticos pode levar à divulgação de informações pouco confiáveis. Foram revisados 226 trabalhos científicos publicados na revista *Ciência Rural* de 2002 a 2006, somente na área de Fitotecnia, sendo utilizados 148 trabalhos para discussão. A maioria dos trabalhos que estudaram mais de um fator foram classificados como incorretos (72%) devido ao abuso dos testes de comparações de médias. Por outro lado, 4 e 24% foram classificados como parcialmente corretos e corretos, respectivamente.

Palavras-chave: testes de comparação de médias, fatores qualitativos e quantitativos, contrastes.

ABSTRACT

The objective of this research was to verify which are the main mistakes and the successes in the application of mean comparison tests in scientific studies, demonstrating viable alternatives in the sense of increasing the impartiality of the results obtained by researchers. One of the researcher's largest challenges is the interpretation of the obtained results in a trustworthy way. In spite of the concern of most researchers with the data analysis, many times certain disregard is observed in the interpretation of the results. Thus, the incorrect application of statistical tests leads the researchers to publish information

not completely reliable. One hundred and forty-eight papers dealing with one or more than one factor were evaluated. All of them are related to the crop production major area, published from 2002 to 2006, in the '*Ciência Rural*', a Brazilian scientific journal. Most of the studies (72%) were classified as incorrect due to the abuse of means comparison tests. In addition, only 4% and 24% were classified respectively as partially correct and correct.

Key words: means comparison tests, qualitative and quantitative factors, contrasts.

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. 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 (CHEW, 1976), sendo que, os testes de comparação de médias frequentemente são os mais utilizados pelos pesquisadores (AFLAKPUI, 1995). PETERSEN (1977), em seus estudos, constatou que 40% dos autores pesquisados usaram, em suas análises, algum tipo de teste de comparação de médias e, destes, 40% utilizavam esses testes, de forma inteiramente inapropriada ao tipo dos dados envolvidos.

¹Departamento de Fitotecnia, Universidade do Estado de Santa Catarina (UDESC). Av. Luiz de Camões, 2090, Conta Dinheiro, 88520-000, Lages, SC, Brasil. E-mail: coimbrajefferson@cav.udesc.br. *Autor para correspondência.

Constantemente, o primeiro passo é submeter os dados a uma análise de variância para determinar se existe ou não diferença significativa entre as médias dos tratamentos (PETERSEN, 1977). Geralmente, o passo seguinte é a comparação de médias por meio de algum teste específico.

O objetivo deste trabalho foi verificar quais os principais erros e acertos na aplicação de testes de comparação de médias em trabalhos científicos, demonstrando alternativas viáveis no sentido de aprimorar a interpretação dos resultados pelos pesquisadores. Para tanto, foram revisados 226 trabalhos científicos publicados neste periódico entre os anos de 2002 e 2006, somente na área de Fitotecnia, sendo utilizados 148 trabalhos para discussão. Os trabalhos que envolviam análise multivariada não foram revisados, uma vez que este não era o objetivo do presente trabalho.

Os artigos foram classificados quanto ao uso de testes de comparação de médias em: i) correto; ii) parcialmente correto e iii) incorreto. O critério utilizado foi o de classificar como uso correto (comparação de médias aos tratamentos de natureza qualitativa e não relacionados e regressão para fatores quantitativos) quando o teste de comparação de médias foi utilizado em função dos objetivos do trabalho e da estrutura de fatores. O critério empregado para o uso incorreto foi usado para os fatores quantitativos em que foi aplicado algum teste de comparação de médias e quando não foi verificado o efeito da interação em experimentos fatoriais (com mais de um fator). Por fim, a denominação de parcialmente correto foi aplicada aos testes de comparações de médias (todos contra todos) quando o correto seria testar contrastes previamente planejados (fatores qualitativos estruturados, em que o adequado seria a aplicação de contrastes ortogonais). Os trabalhos foram divididos, para exemplificação, em três grupos: i) Grupo I, para os trabalhos corretos; ii) Grupo II, para os parcialmente corretos e iii) Grupo III, para os incorretos.

A estatística F , numa análise de variância para tratamentos com mais de um grau de liberdade, proporciona informações não-conclusivas, relacionadas diretamente com o comportamento médio dos tratamentos. Por essa razão, devem ser planejadas comparações objetivas, como, por exemplo, decompondo os graus de liberdade de tratamentos para se obter informações mais específicas e, principalmente, mais coerentes com o plano experimental.

Considerando-se (Tabela 1) os diferentes tipos de fatores encontrados nos trabalhos revisados, bem como sua classificação quanto ao uso em correto, parcialmente correto e incorreto, pode ser notado que

o percentual de trabalhos classificados como adequados para a categoria unifatorial foi de 68% (30 artigos), enquanto que para a categoria fatorial foi de 24% (25 artigos). Já o percentual de trabalhos classificados como inadequados para as categorias unifatorial e fatorial foi de 9% (3 artigos) e 72% (76 artigos), respectivamente.

A maior parte dos erros encontrados foi relacionada a trabalhos que utilizaram mais de um fator, nos quais os autores não consideraram a interação entre os fatores, realizando testes de comparação de médias de forma separada. O uso de teste de comparação de médias nessas circunstâncias é um abuso comum dessa técnica (CHEW, 1976). Assim, em experimentos fatoriais, o primeiro passo é analisar o efeito da interação entre os vários fatores (PETERSEN, 1977). Sendo assim, quando o teste F for significativo para o efeito da interação, os graus de liberdade devem ser rearranjados de modo a comparar os níveis de um fator dentro dos níveis do outro (CARDELLINO & SIEWERDT, 1992). Nesse caso, fixando-se um fator e variando-se dentro dos níveis do outro fator, realiza-se o que é conhecido por teste de efeitos simples (WINER, 1971).

Com relação à utilização de testes de comparação de médias, nos experimentos unifatoriais revisados, 17, 8 e 1 trabalhos, foram classificados quanto ao uso de forma correta, parcialmente correta e incorreta, respectivamente (Tabela 2). De forma que, nos experimentos fatoriais, 58 trabalhos revisados estavam incorretos com relação aos testes de comparações de médias, enquanto 16 estavam corretos e 4 parcialmente corretos (Tabela 2). As comparações entre tratamentos podem ser afetadas pelas condições que ocorrem, e interpretações claras dos efeitos de um tratamento precisam ser levadas em conta para o efeito de outros tratamentos (KUEHL, 1994). Isso significa que a interação entre fatores pode estar presente em delineamentos e, sendo a interação significativa, é necessário se fazer um estudo da interação através de sua decomposição em fatores aninhados (NOGUEIRA & CORRENTE, 2000).

As causas principais da má aplicação dos testes de comparação de médias observadas foram: i) desconhecimento do procedimento de aplicação, bem como dos conceitos de condições de uso de cada teste; ii) inabilidade quanto à interpretação dos resultados, levando muitos pesquisadores a afirmações errôneas; iii) descaso quanto à aplicabilidade dos testes, em que alguns pesquisadores utilizaram testes que mais discriminassem (maior número de diferença significativa entre as médias testadas) os resultados esperados pelos pesquisadores negligentes. Com relação aos

Tabela 1 - Número de artigos revisados quanto ao número de fatores (um fator ou mais de um fator), quanto ao tipo de fator (qualitativo não-estruturado (Qlne), qualitativo estruturado (Qle), quantitativo (Qt)) e quanto ao uso (correto, parcialmente correto e incorreto), editados na revista *Ciência Rural*, no período 2002 a 2006.

-----Unifatorial-----								
Classificação	Qlne	Qle	Qt	Total	%			
Correto	17	0	13	30	68			
Parcialmente correto	0	8	2	10	23			
Incorreto	1	0	2	3	9			
Total	18	8	17	43				
%	42	19	39		100			
-----Fatorial-----								
Classificação	¹ Ql x Ql	Ql x Qt	² Qt x Qt	Ql x Ql x Ql	Ql x Ql x Qt	Ql x Qt x Qt	Total	%
Correto	9	7	4	3	1	1	25	24
Parcialmente correto	2	2	0	0	0	0	4	4
Incorreto	24	18	17	4	10	3	76	72
Total	35	27	21	7	11	4	105	
%	33	26	20	7	10	4		100

¹/ Ql = qualitativo. ²/ Qt = quantitativo.

abusos registrados, os mais habituais foram: i) experimento com fator qualitativo estruturado (o correto seria utilizar contrastes, previamente planejados), tendo sido utilizado teste de comparação de médias; ii) fatores quantitativos (o correto seria ajustar equação (ões) de regressão), tendo sido utilizado teste de comparação de médias, por exemplo, Duncan; iii) experimentos fatoriais com interação significativa entre os fatores (é relevante decompor tanto os graus de liberdade da interação e dos efeitos envolvidos), tendo sido aplicados testes de comparação de médias tanto para discriminar o efeito principal quanto da interação.

O pesquisador pode optar por alguns procedimentos de acordo com o tipo de fator envolvido. Por exemplo, é possível particionar os graus de liberdade em experimentos fatoriais ou, se os tratamentos consistirem em níveis de fatores quantitativos, através da decomposição em fatores aninhados. O ajuste de uma equação de regressão permite estudar qualquer nível intermediário do fator em questão, mesmo que não diretamente incluído no estudo (CARDELLINO & SIEWERDT, 1992). Quando os contrastes (que envolvem grupos de médias) são definidos após prévia inspeção dos dados, sua

Tabela 2 - Classificação dos trabalhos unifatoriais (um fator somente) e fatoriais (mais de um fator) editados na revista *Ciência Rural* (2002 a 2006), quanto ao uso de testes de comparação de médias.

Teste	-----Unifatorial-----		
	correto	Par. correto	incorreto
Duncan	7	6	0
Tukey	4	1	0
Scott & Knott	5	1	0
Dms	1	0	0
Dunett	0	0	1
Total	17	8	1
Teste	-----Fatorial-----		
	correto	par. correto	incorreto
Duncan	4	2	26
Tukey	10	1	24
Scott & Knott	1	1	4
Dms	1	0	3
Dunnett	0	0	1
Total	16	4	58

significância pode ser testada usando-se o teste de Scheffé (BANZATTO & KRONKA, 1995). Se o pesquisador deseja comparar todos os pares de médias entre si, deve ser escolhido um teste de comparação de médias, como, por exemplo, Tukey (PERECIN & MALHEIROS, 1989). Quando o interesse é comparar a média de qualquer tratamento com a testemunha (controle), o procedimento correto é o emprego do teste de Dunnett (LENTNER & BISHOP, 1986).

Portanto, existe dificuldade na escolha de qual o procedimento correto em relação ao tipo de fator experimental estudado. O uso incorreto dos testes de comparação de médias é mais comum em situações em que os tratamentos são de natureza quantitativa. Em experimentos fatoriais, existe deficiência na decomposição da interação. Falhas na escolha do teste correto têm implicação para inferências derivadas do experimento.

REFERÊNCIAS

- AFLAKPUI, G.K.S. Some uses/abuses of statistics in crop experimentation. **Tropical Science**, v.35, n.2, p.347-353, 1995.
- BANZATTO, D.A.; KRONKA, S.N. **Experimentação agrícola**. 3.ed. Jaboticabal: FUNEP, 1995. 247p.
- CARDellino, R.A.; SIEWERDT, F. Utilização adequada e inadequada dos testes de comparação de médias. **Revista da Sociedade Brasileira de Zootecnia**, v.21, n.6, p.985-995, 1992.
- CHEW, V. Comparing treatment means: a compendium. **Hortscience**, v.11, n.4, p.348-357, 1976.
- GILL, J.L. **Design and analysis of experiments**. Ames: Iowa State University, 1978. V.1, 410p.
- KUEHL, R.O. **Statistical principles of research design and analysis**. Belmont, California: Duxburys, 1994. 686p.
- LENTNER, M.; BISHOP, T. **Experimental design and analysis**. Blacksburg, VA: Valley Book, 1986. 565p.
- NOGUEIRA, M.C.S.; CORRENTE, J.E. Decomposição da interação tripla significativa utilizando o comando *CONTRAST* do PROC GLM do SAS aplicado a modelo de classificação tripla para dados balanceados. **Bragantia**, v.59, n.1, p.109-115, 2000.
- 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.
- SILVA, J.G.C. da. A consideração da estrutura das unidades em inferências derivadas do experimento. **Pesquisa Agropecuária Brasileira**, v.34, n.3, p.911-925, 1999.
- WINER, B.J. **Statistical principles in experimental design**. 2.ed. New York: McGraw-Hill, 1971. 907p.