

Aplicação dos testes de comparação de médias em ensaios de cevada

Eduardo Caierão¹

Resumo - Os testes de comparação de médias (TCM) são ferramentas estatísticas de apoio à experimentação agrícola, mas nem sempre são empregados corretamente. O trabalho teve como objetivo avaliar o nível de adequação deste procedimento estatístico na rede experimental de cevada. Foram analisados os trabalhos incluídos nos anais das Reuniões da Comissão Brasileira de Pesquisa de Cevada, de 2001 a 2005. A aplicação dos TCM foi classificada como adequada, como parcialmente adequada e como inadequada, de acordo com as características dos tratamentos a que foram aplicados. O percentual de uso dos TCM oscilou de 40 a 60%, conforme o ano de avaliação, sendo que o teste de Tukey esteve presente em 80% das situações. Na média dos cinco anos de avaliação, os TCM foram aplicados adequadamente em 40% dos experimentos. Em 46% dos artigos, esses testes estavam parcialmente corretos e, nos 14% dos trabalhos restantes, incorretamente aplicados. O aprofundamento dos conhecimentos estatísticos, especialmente no que tange à existência e à aplicabilidade de outros TCM, como por exemplo Duncan e Scott-Knott, pode ser uma estratégia eficaz de melhorar o percentual de uso e adequabilidade dos TCM.

Palavras-Chave - *Hordeum vulgare*, estatística, Tukey, Duncan, Scott-Knott.

Mean comparison tests in barley trials

Abstract - The mean comparison tests (MCT) are statistical tools supporting agricultural research, but are not always appropriately applied. The objective of this work was to evaluate the level of adequability of this statistical procedure in barley experimentation network. Papers from the Proceedings of the Brazilian Meetings of the Barley Research Comittee, from 2001 to 2005, were analyzed. The applications of MCT were classified as adequate, partially adequate or inadequate, according to the traits of treatments involved. The usage percent of MCT range from 40 to 60%, according to the evaluation year. Tukey's test has been used in 80% of situations. In the mean of five years, the MCT were adequately applied in 40% of trials. In 46% of papers, this tests were partially adequate and, in the other 14%, inadequately applied. To broaden statistical knowledge, mainly for existence of other MCT, for example, Duncan and Scott-Knott, may be an efficient strategy to improve the use and adequability use of MCT.

Key-Words - *Hordeum vulgare*, statistics, Tukey, Duncan, Scott-Knott.

¹ Engenheiro Agrônomo, Pesquisador da Embrapa Trigo. Msc. Melhoramento Vegetal. Embrapa - CNPT, Rodovia BR 285, km 294 - C. P. 451, CEP: 99001-970, Passo Fundo/RS. e-mail: caierao@cnpt.embrapa.br.
Recebido para publicação em 17/03/2006

Introdução

A estatística aplicada à experimentação agrícola consiste numa ferramenta de grande importância ao pesquisador, já que permite a estimativa do erro experimental e a verificação da significância dos contrastes analisados (CARDELLINO e SIEWERDT, 1992). Contudo, a análise realizada e a interpretação biológica dos fenômenos avaliados devem revelar sincronia, onde as inferências propostas devem estar de acordo com a quantidade de informações obtidas.

Freqüentemente, a escolha do tipo de procedimento estatístico a ser realizado, após a análise de variância, não é o mais recomendado para o tipo de experimento conduzido, principalmente quando são aplicados os testes de comparação de médias em fatores quantitativos, como doses de produtos, concentrações, densidade, entre outros. A escolha inadequada do tipo de análise pode representar dificuldades na interpretação dos dados experimentais, pela limitação nas inferências, ou mesmo pela indução ao erro sobre o exposto (CARDELLINO e SIEWERDT, 1992).

Quarenta por cento dos trabalhos avaliados na revista *Agronomy Journal* utilizavam algum tipo de teste de comparação de médias (PETERSEN, 1977) e pelo menos 40% destes o utilizavam inapropriadamente, considerando a sua correta utilização para fatores qualitativos e não correlacionados (PIMENTEL-GOMES, 1990). Da mesma forma, Cardellino e Siewerdt (1992), em levantamento feito na Revista da Sociedade Brasileira de Zootecnia, de 1984 a 1989, concluíram que dos trabalhos que empregavam algum tipo de teste de comparação de médias, 64,2 % não o aplicavam corretamente. Situação semelhante foi encontrada por Santos et al. (1998), na revista *Pesquisa Agropecuária Brasileira*, no período de 1980 a 1994, onde 57% dos artigos também utilizaram incorretamente este procedimento estatístico. Situação diferenciada e específica foi encontrada por Caierão et al. (1999), em ensaios de aveia, onde mais de 80% dos trabalhos apresentados nos anais de pesquisa das reuniões da Comissão de Pesquisa deste cereal, de 1996 a 1998, caracterizaram-se por aplicação correta dos testes de comparação de médias. Em levantamento realizado na revista *Horticultura Brasileira*, no período de 1983 a 2000, Bezerra Neto et al. (2002) observou percentual de uso adequado dos testes de comparação de médias de 65,6%.

A escolha de qual teste de comparação de médias será utilizado depende do grau de precisão desejado e do nível de discriminação de tratamentos oferecido pelo procedimento estatístico. Conforme Lucio et al. (2003), o teste de Duncan diferencia mais os tratamentos quando comparado com o teste de Tukey.

Para experimentos onde os tratamentos são de natureza quantitativa, com pelo menos três níveis, o procedimento mais informativo é o ajustamento de equações de regressão (PIMENTEL-GOMES, 1984). A análise de regressão permite a estimativa da variável resposta no seu

comportamento máximo e mínimo em função da variável dependente, dentro da amplitude dos níveis considerados, informação esta que não encontra embasamento numa simples comparação de médias dos níveis do fator quantitativo (PETERSEN, 1977).

Devido a importância da cultura da cevada para o sul do Brasil, os trabalhos científicos devem ter a maior precisão possível. Isto implica em aproveitamento das informações obtidas, o que pode ser alcançado pela aplicação mais adequada dos testes de comparação de médias e outros procedimentos estatísticos. Assim, o trabalho teve como objetivo avaliar a adequabilidade da aplicação dos testes de comparação de médias em ensaios de cevada.

Materiais e métodos

Os dados foram retirados dos trabalhos publicados nos anais das Reuniões da Comissão Brasileira de Pesquisa de Cevada, de 2001 a 2005. Os anais incluem trabalhos desenvolvidos por instituições de pesquisa, em todas as regiões produtoras de cevada, desde o Rio Grande do Sul até São Paulo, conferindo representatividade às análises realizadas. Oitenta por cento das publicações analisadas faziam referência à comparação de cultivares e os demais, comparação de outros tipos de tratamentos, como fungicidas, inseticidas, doses de nitrogênio, redutor de crescimento, densidade e época de semeadura.

Os trabalhos foram submetidos a uma primeira classificação, de acordo com a utilização ou não de algum tipo de teste de comparação de médias (TCM). Todas as publicações que foram analisadas via este procedimento estatístico foram, posteriormente, agrupadas conforme a sua aplicação, sendo utilizadas as categorias: adequada, parcialmente adequada e inadequada. A classificação adequada foi empregada quando os fatores considerados no trabalho eram de natureza qualitativa não estruturada, ou seja, para fatores onde os níveis eram de interesse específico no experimento com ausência de relação entre os mesmos, como, por exemplo, diferentes cultivares e diferentes tipos de herbicidas. Foram julgados inadequados os procedimentos referentes a utilização de TCM aos níveis de fatores quantitativos com mais de dois níveis, como, por exemplo, doses de nitrogênio e, parcialmente adequados, a experimentos fatoriais (mais de um fator), que não consideraram a decomposição dos níveis do fator principal para a comparação ou aos ensaios compostos de fatores qualitativos estruturados (fatores onde os níveis estão relacionados e são de interesse específico para o experimento), onde o procedimento mais adequado seria a comparação através de contrastes. Aqueles trabalhos que apresentaram TCM, foram então classificados de acordo com o teste utilizado, nos diferentes anos de avaliação. Também foi determinado o tipo de procedimento estatístico utilizado nas demais publicações que não utilizaram TCM.

Alguns trabalhos foram ilustrados para a melhor discussão dos resultados. As inferências realizadas a partir do teste empregado em cada uma das situações e àquelas possíveis pela aplicação do procedimento estatístico adequado foram discutidos.

Resultados e discussão

Foram analisados 258 trabalhos completos publicados nos anais das Reuniões da Comissão Brasileira de Pesquisa de Cevada, de 2001 a 2005. O percentual de utilização dos TCM oscilou de 40 a 60% conforme o ano de avaliação (Figura 1). Contudo, estes valores ficaram abaixo dos encontrados por Caierão et al. (1999), em ensaios de aveia, onde o percentual de uso dos TCM chegou a 73%.

O emprego adequado dos TCM nos trabalhos de cevada não foi inferior a 32%, chegando a atingir 44% no ano de 2001 (Tabela 1). Na média dos 5 anos de

avaliação, em 40% das publicações, o procedimento estatístico foi usado de maneira adequada (Tabela 1). Esses valores são baixos frente a taxa de uso adequado de 65,6% encontrada por Bezerra Neto et al. (2002) nas publicações da revista Horticultura Brasileira, no período de 1983 a 2000. Os tipos de experimentos mais frequentes nessa categoria foram os de comparação de linhagens. A utilização inadequada dos TCM oscilou de 4 a 22%, com média de 14% ao final dos 5 anos de avaliação. Em todas as situações consideradas ele foi empregado para comparar níveis de fatores quantitativos, como, por exemplo, doses de nitrogênio. Percentual expressivo dos trabalhos foi enquadrado como parcialmente adequado (Tabela 1). Na maioria destas situações, fatores qualitativos estruturados foram considerados como desestruturados, e comparados individualmente. Para a obtenção de melhores inferências, o uso de contrastes ortogonais seria a ferramenta estatística mais apropriada.

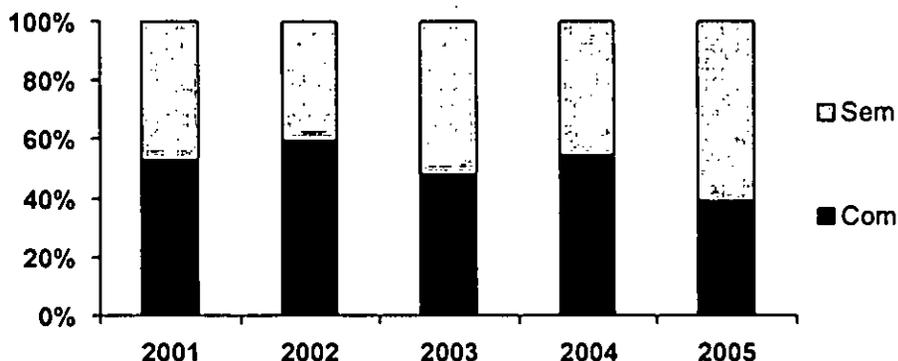


Figura 1 - Percentagens de trabalhos publicados com e sem a utilização de testes de comparação de médias, de 2001 a 2005, em ensaios de cevada.

Tabela 1 - Classificação da aplicação dos testes de comparação de médias, de 2001 a 2005, em ensaios de cevada.

Classificação	2001		2002		2003		2004		2005		Total	
	N	%	N	%	N	%	N	%	N	%	N	%
Adequado	12	44	8	32	11	46	11	38	9	39	51	40
Parcialmente adequado	10	37	16	64	9	38	15	51	9	39	59	46
Inadequado	5	19	1	4	4	17	3	10	5	22	18	14
Total	27	100	25	100	24	100	29	100	23	100	128	100

Nos trabalhos onde foram aplicados TCM, o teste de Tukey foi o mais comum, sendo observado em 103 trabalhos e representando mais de 80% do total. O teste de Duncan foi aplicado em, aproximadamente, 15% dos trabalhos, seguido pelo teste de Scott-Knott (2,3%), pelo teste t (2,3%), e por aqueles em que não foi possível sua identificação (aproximadamente 1%) (Tabela 2). A elevada aplicação do teste de Tukey pode ser um indicativo da rigorosidade desejada pelos pesquisadores no momento da escolha do procedimento (BANZATTO e KRONKA, 1995) ou falta de familiari-

dade com outros tipos de testes, por vezes, mais adequados. Em levantamento realizado na revista Ciência Rural, no período de 1971 a 2000, Lucio et al. (2003) detectaram que os testes de Tukey e Duncan foram aplicados quando o número de tratamentos a serem testados foi inferior a 20 e, quando este aumentava, havia preferência pelo uso dos testes Qui-quadrado (χ^2) e de F. A opção de não aplicar um TCM quando o número de tratamentos é superior a 20 não é fundamentada, já que nestes casos há a opção do teste de Scott-Knott (1974). A baixa utilização do teste "t" pode

Tabela 2 - Classificação dos trabalhos analisados quanto ao tipo de teste de comparação de médias utilizado, em ensaios de cevada.

Ano	Tipo de teste de comparação de médias utilizado				
	Tukey	Duncan	Scott-Knott	Teste t	Não identificado
2001	17	6	1	0	0
2002	17	9	1	2	0
2003	21	3	0	0	0
2004	24	0	0	1	0
2005	24	1	1	0	1
Total	103 (79,9%)	19 (14,7%)	3 (2,3%)	3 (2,3%)	1 (0,8%)

ser um indicativo da não aplicação do método dos contrastes para comparação de tratamentos, como também, pode revelar insuficiência no planejamento prévio dos ensaios. A rigorosidade do teste de Tukey, freqüentemente, dá origem a interpretações confusas, principalmente quando o procedimento não discrimina satisfatoriamente os tratamentos. Nestas situações, o uso da estatística (no caso TCM) dificulta e não auxilia na interpretação dos resultados. Sugere-se, portanto, o uso de um TCM que tenha por característica uma maior discriminação dos tratamentos, como o teste de Duncan ou Skott-Knott (LUCIO et al., 2003).

Freqüentemente melhoristas e fitotecnistas têm optado por ferramentas estatísticas mais simples no auxílio às interpretações, como por exemplo, o desvio padrão da média, o percentual em relação a testemunha de melhor desempenho, entre outros. Nos trabalhos analisa-

dos em que não foram aplicados TCM, não houve aplicação de qualquer ferramenta estatística em 77 publicações e, em 31 delas, o procedimento utilizado foi o percentual em relação a melhor testemunha (Tabela 3). Em 80% destas situações o contexto do trabalho era a comparação de linhagens e genótipos, o que dá um indicativo de que melhoristas de cevada, em muitas situações, estão optando por interpretações subjetivas de campo em apoio aos dados de produtividade de grãos. As regressões, quando utilizadas, foram aplicadas corretamente. O percentual reduzido deste procedimento estatístico não pode ser atribuído a uma limitação na utilização dos fatores quantitativos nos ensaios de cevada, mas sim à aplicabilidade errônea dos TCM, que substituíram o ajustamento de equações para fatores quantitativos em 14% (média dos anos de avaliação) nos trabalhos analisados (Tabela 1).

Tabela 3 - Classificação dos trabalhos que não apresentaram teste de comparação de médias (TCM), quanto ao tipo de procedimento estatístico utilizado em ensaios de cevada.

Ano	Procedimento estatístico utilizado quando não se aplicou TCM				
	Nenhum	% Melhor Testemunha	χ^2	Regressão	Outros
2001	20	6	1	7	1
2002	14	7	0	2	1
2003	16	7	1	3	0
2004	12	5	2	1	1
2005	15	6	1	3	0
Total	77	31	5	16	3

Os exemplos 1 e 2, incluídos nas tabelas 4 e 5, respectivamente, ilustram algumas situações encontradas nos anais da Comissão de Pesquisa de Cevada, no período analisado.

No exemplo 1, o objetivo do ensaio foi comparar o efeito das doses de nitrogênio em cobertura sobre o rendimento de grãos do cultivar de cevada BRS 195. Pelo exposto, o autor comparou as doses com o teste de Tukey, o que limita as inferências possíveis sobre os dados. Em fatores quantitativos, os níveis não são de interesse específico (PIMENTEL-GOMES, 1990), mas escolhidos de modo a permitir a estimativa apropriada da variável resposta em função da variável explanatória. Neste caso, o autor não pode inferir sobre qualquer ou-

tra dose que não seja aquelas testadas, ao passo que o ajustamento de uma curva de regressão poderia proporcionar informações importantes sobre a resposta do cultivar. Embora, no trabalho, o uso de TCM sobre fatores quantitativos tenha sido classificada como inadequada, pelos princípios da estatística, a nível prático, ela não é inadequada, mas sim incompleta. O que se considera, é que os dados poderiam ser melhor explorados com a aplicação da ferramenta apropriada.

O exemplo 2 demonstra uma das situações onde os TCM foram classificados como parcialmente adequados. O objetivo do ensaio foi comparar o grau de infecção de um isolado, na primeira e segunda folha da cevada, entre tratamentos com gengibre (24, 48 e 72h) e uma

APLICAÇÃO DOS TESTES DE COMPARAÇÃO DE MÉDIAS EM ENSAIOS DE CEVADA

Tabela 4 - Exemplo 1: Rendimento de grãos da cultivar de cevada BRS 195 pela aplicação de doses de N em cobertura, sobre restavas de soja e milho.

Dose de N	Soja		Milho	
0	2.910	c	3.353	c
20	3.682	bc	4.260	b
40	3.954	ab	4.327	
60	4.241	ab	4.754	ab
80	4.641	a	5.200	a
100	4.685	a	5.308	a

Médias seguidas da mesma letra na coluna não diferem significativamente entre si, pelo teste de Tukey (5%).

FONTE: XXV Reunião Anual de Pesquisa de Cevada. Guarapuava, 2005.

Tabela 5 - Exemplo 2: Percentual de proteção em folhas de plantas de cevada da cultivar Embrapa 128, contra o respectivo isolado, utilizando o extrato de gengibre como indutor.

Tratamentos	1º Folha (%)		2º Folha (%)	
1 - Gengibre 24 h 1º FT	82,6	b	86,9	b
2 - Gengibre 24 h 2º FT	89,5	b	84,2	b
3 - Gengibre 48 h 1º FT	83,3	b	100,0	b
4 - Gengibre 8 h 2º FT	100,0	b	92,8	b
5 - Gengibre 72 h 1º FT	92,8	b	95,3	b
6 - Gengibre 72 h 2º FT	99,0	b	98,9	b
7 - Infectada	0,0	a	0,0	a

FONTE: XXV Reunião Anual de Pesquisa de Cevada. Guarapuava, 2005. Médias seguidas da mesma letra na coluna não diferem significativamente entre si pelo teste de Tukey (5%)

testemunha. Se a comparação tivesse sido realizada através de contrastes, o autor poderia contrastar o tempo de ação do gengibre ($C_1 = m_1 + m_2 - m_3 - m_4$, $C_2 = m_1 + m_2 - m_5 - m_6$, $C_3 = m_3 + m_4 - m_5 - m_6$), o efeito do bioprotetor sobre a posição das folhas ($C_1 = m_1 + m_3 + m_5 - m_2 - m_4 - m_6$) ou o efeito do gengibre ($C_1 = m_1 + m_2 + m_3 + m_4 + m_5 + m_6 - m_7$). Nestas situações, as inferências seriam mais completas e informativas.

Conclusões

O percentual de uso e de adequabilidade dos TCM para comparação de tratamentos em ensaios de cevada é baixo. O aprofundamento dos conhecimentos estatísticos, especialmente no que tange à existência e à aplicabilidade de outros TCM, como por exemplo Duncan e Scott-Knott, pode ser uma estratégia eficaz de melhorar estes índices.

Referências

BANZATTO, D.A.; KRONKA, S. do N. Experimentação Agrícola. 3. ed. Jaboticabal: FUNEP, 1995. 247 p.

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. 5-9, 2002.

CAIERÃO, E.; CARVALHO, F.I.F.; FLOSS, E.L. Eficiência da Aplicação dos Testes de Comparação de Médias em Ensaio de Avcia. Pesquisa Agropecuária Gaúcha, Porto Alegre, v. 5, n. 2, p. 317-324, 1999.

CARDELLINO, R.; SIEWERDT, F. Utilização Correta e Incorreta dos Testes de Comparação de Médias. Revista da Sociedade Brasileira Zootecnia, Viçosa, v. 21, n. 6, p. 985 - 995, 1992.

LUCIO, A.D.; LOPES, S.J.; STORCK, L.; CARPES, R.H.; LIEBERKNECHT, D.; NICOLA, M.C. Características Experimentais das Publicações da Ciência Rural de 1971 a 2000. Ciência Rural, Santa Maria, v. 33, n.1, p. 161-164, 2003.

PETERSEN, R.G. Use and Misuse of Multiple Comparison Procedures. Agronomy Journal, Madison, v. 69, p. 205-208, 1977.

PIMENTEL-GOMES, F. Curso de Estatística Experimental. 13. ed. Piracicaba: USP/ESALQ, 1990. 468 p.

PIMENTEL-GOMES, F. A Estatística Moderna na Pesquisa Agropecuária. 2. ed. Piracicaba: POTAFÓS, 1984. 162 p.

SANTOS, J.W.; MOREIRA, J. de A.N.; BELTRÃO, N.E. de M. Avaliação do Emprego dos Testes de Comparação de Médias na Revista Pesquisa Agropecuária Brasileira (PAB-1980 a 1994). Pesquisa Agropecuária Brasileira, Brasília, v. 33, n.3, p. 225-230, 1998.

SCOTT, A.J.; KNOTT, M.A. A Cluster Analysis Methods for Grouping Means in the Analysis of Variance. Biometrics, Raleigh, v. 30, n.3, p. 507-512, 1974.