

AVALIAÇÃO DO EMPREGO DOS TESTES DE COMPARAÇÃO DE MÉDIAS NA REVISTA PESQUISA AGROPECUÁRIA BRASILEIRA (PAB) DE 1980 A 1994¹

JOSÉ WELLINGTON DOS SANTOS, JOSÉ DE ALENCAR NUNES MOREIRA²
e NAPOLEÃO ESBERARD DE MACÉDO BELTRÃO³

RESUMO - O acerto no uso dos testes de comparações de médias (TCM) reveste-se de grande importância na análise de dados dos experimentos. Desta forma, procedeu-se ao levantamento dos trabalhos publicados na revista Pesquisa Agropecuária Brasileira (PAB) de 1980 a 1994, em que se utilizou algum tipo de TCM, com o objetivo de averiguar a adequação desses testes às análises estatísticas realizadas. O estudo envolveu 628 trabalhos e os TCM adotados foram classificados nas categorias inadequada, adequada e parcialmente adequada. Constatou-se que 57,01%, 31,53% e 11,46% dos trabalhos levantados, aplicaram os TCM segundo as três categorias adotadas, respectivamente.

Termos para indexação: análise de variância, planejamento experimental.

EVALUATION OF THE AVERAGE COMPARISON TESTS IN THE "PESQUISA AGROPECUÁRIA BRASILEIRA" JOURNAL FROM 1980 TO 1994

ABSTRACT - The choice of average comparison tests (ACT) is very important in the data analysis of experimental designs. Due to this fact, it was carried out a survey of the papers published in the "Pesquisa Agropecuária Brasileira - PAB" journal from 1980 to 1994 in which some type of ACT was used. The objective of this paper was to point up the growing problems of the indiscriminate use of the ACT applied to the statistical analysis in the papers published in this journal. The study was carried out using 628 papers, and the ACT were classified in the following categories: appropriate, not appropriate and partially appropriate. As a consequence of this study, it was found that 57.01%, 31.53% and 11.46% of the surveyed papers applied the appropriate, not appropriate and partially appropriate ACT, respectively.

Index terms: analysis of variance, experimental planning.

INTRODUÇÃO

As variáveis dependentes, obtidas nos experimentos planejados e conduzidos adequadamente, podem ser submetidas a vários tipos de análises estatísticas, sendo a mais importante a análise de variância, a qual, além do emprego em pesquisas

agrícolas e biológicas, é também utilizada em outras áreas de pesquisa, como nas ciências humanas e exatas.

Em geral, a primeira etapa no estudo dos dados de um experimento é realizar-se a análise da variância, para que, a partir desta, se possa testar a hipótese de igualdade das médias populacionais dos tratamentos testados pela estatística F. Um aspecto essencial antes de se realizar a análise diz respeito ao conhecimento da natureza dos tratamentos estudados, isto é, se estes são de efeito aleatório ou fixo. No primeiro caso, o objetivo da pesquisa é a estimação dos componentes de variância e, no segundo, a estimação e comparação dos efeitos, por meio de testes de

¹ Aceito para publicação em 9 de julho de 1997.

² Eng. Agr., M.Sc., Embrapa-Centro Nacional de Pesquisa de Algodão (CNP), Caixa Postal 174, CEP 58107-720 Campina Grande, PB. E-mail: jwsantos@cnpa.embrapa.br

³ Eng. Agr., Dr., Embrapa-CNP.

comparação de médias (TCM). Acontece que tais procedimentos vêm sendo usados indiscriminadamente, conforme relatado por diversos autores. Petersen (1977) constatou que 40% dos autores pesquisados usaram, em suas análises, algum tipo de TCM, e destes, 40% o utilizaram de forma inteiramente inapropriada para o tipo de dados envolvidos, tendo em vista que os TCM são apropriados para situações em que os tratamentos são de caráter qualitativo e não relacionados. Segundo Little (1978), os TCM são aplicáveis em experimentos cujos tratamentos consistem em um conjunto de materiais não relacionados, tais como cultivares, raças, métodos, tipo de produtos, configuração de consórcio, métodos de plantio e tipo de fertilizantes. Não ocorre o mesmo, no entanto, quando os tratamentos são de natureza quantitativa, com mais de dois níveis e, neste caso, recomenda-se o estudo de regressão (Pimentel-Gomes, 1984).

Cardellino & Siewerdt (1992) concluíram, a partir de revisão nos trabalhos publicados na Revista da Sociedade Brasileira de Zootecnia, de 1984 a 1989, que 260 autores utilizaram alguns tipos de TCM e que entre esses 64,2% aplicaram tais procedimentos de maneira inadequada, 24,6% adequada e 11,2% parcialmente adequada e, ainda, apresentaram alternativas para o uso correto de tais testes.

O presente trabalho relata casos de inadequação de análises estatísticas, no caso dos TCM, e tem por objetivo identificar as deficiências no uso de referidos procedimentos na revista Pesquisa Agropecuária Brasileira - PAB, além de apresentar alternativas adequadas para a solução desses problemas.

MATERIALE MÉTODOS

Os dados fazem parte de um levantamento realizado nos trabalhos publicados na revista Pesquisa Agropecuária Brasileira - PAB, em que se utilizou algum tipo de TCM, no período de 1980 a 1994. Os procedimentos adotados quanto à aplicação dos TCM foram classificados nas categorias adequada, inadequada e parcialmente adequada. Considerou-se adequado o procedimento no qual a aplicação dos TCM era feita para tratamentos de natureza qualitativa e não relacionados; inadequado, quando aplicado a tratamentos de natureza quantitativa; e parcialmente ade-

quado para a aplicação dos TCM a tratamentos relacionados (casos em que no conjunto dos tratamentos apareçam tratamentos formados pela adição de um ou mais fatores, situação muito freqüente em experimentos nas áreas de nutrição animal e práticas culturais), combinação fatorial e, ainda, nos casos em que os tratamentos foram de natureza quantitativa e quando foram aplicados, simultaneamente, os TCM e a análise de regressão.

Após as informações obtidas no levantamento, procedeu-se ao cálculo dos percentuais referentes a cada uma das três categorias adotadas para os seguintes testes de comparação de médias: Tukey, Duncan, lsd, t (aplicado a contrastes) e Student - Newman-Keuls (SNK) (Winer, 1971; Neter & Wasserman, 1974; Snedecor & Cochran, 1974; Steel & Torrie, 1980).

RESULTADOS E DISCUSSÃO

Na Tabela 1 são encontrados os números totais e respectivos percentuais de classificação, segundo o critério adotado, dos trabalhos publicados na PAB no período de 1980 a 1994, com algum tipo de comparação de médias. Verifica-se, nesta tabela, que dos 628 artigos analisados, 11 não nos permitiram identificar o tipo de teste nem o nível de significância adotados na comparação das médias; no entanto, nos demais, 57,01%, 31,53% e 11,46%, respectivamente, aplicaram-se os TCM de forma adequada, inadequada e parcialmente adequada.

Considerando-se os testes individualmente, observa-se, na Tabela 1, que nos trabalhos em que foi empregado o teste de Tukey para comparação de médias dos tratamentos, 33,4% utilizaram o procedimento de forma inadequada. No teste t, e com a mesma finalidade, este percentual elevou-se para 85,7% nos casos em que os autores compararam as médias duas a duas de K tratamentos, sendo seu uso mais freqüente na experimentação zootécnica. Para os que usaram o teste SNK, a percentagem de acerto foi de 100%; todavia, esse procedimento foi o de menor preferência entre os trabalhos analisados.

Uma das prováveis causas do emprego inadequado dos testes de comparação de médias pode estar associada ao desconhecimento, por parte do pesquisador, da adequabilidade dos testes aos tipos de dados estudados; com isto, pode-se tirar conclusões que não refletem as respostas corretas aos materiais biológicos investigados.

TABELA 1. Classificação dos trabalhos considerados neste documento e publicados na PAB, em que se utilizou algum tipo de teste de comparação de médias (TCM), no período 1980 a 1994.

Teste usado	Aplicação dos TCM							
	Adequada		Inadequada		Parc. adequada		Total	
	Nº	%	Nº	%	Nº	%	Nº	%
Tukey	183	49,7	124	33,4	64	17,3	371	59,08
Duncan	154	76,3	40	19,8	8	3,9	202	32,17
lsd	10	66,7	5	33,3	0	0,0	15	2,39
t	3	14,3	18	85,7	0	0,0	21	3,34
SNK	8	100,0	0	0,0	0	0,0	8	1,27
Não identificado	0	0,0	11	100,0	0	0,0	11	1,75
Total	358		198		72		628	
Porcentagem	57,01		31,53		11,46		100,0	

O uso inadequado mais freqüente dos TCM ocorreu nos casos em que os tratamentos foram de natureza quantitativa.

No levantamento realizado constatou-se que em 11 dos trabalhos, mesmo tendo-se realizado algum tipo de comparação de médias, não foi feita nenhuma referência acerca do tipo de teste realizado.

Com base nas informações obtidas neste trabalho, concluiu-se que parte dos pesquisadores não tem tido o cuidado necessário com o uso das técnicas estatísticas, indispensáveis na elaboração das tomadas de decisão. Uma situação comum foi a indefinição do delineamento experimental usado na pesquisa e dos TCM dos tratamentos. Assim, evidencia-se falha na definição completa do método estatístico a ser usado nos trabalhos, em particular dos procedimentos adequados para comparação das médias dos tratamentos.

Face a tais constatações, sugere-se aplicar os testes de comparação de médias em experimentos cujo objetivo é determinar os melhores tratamentos, dentro de um conjunto qualitativo. Nesse tipo de experimento o uso mais comum desses procedimentos é efetuar comparações sobre as médias, com o objetivo de detectar possíveis grupos entre um conjunto de tratamentos não relacionados.

Ressalta-se que os TCM são aplicáveis para fatores quantitativos, com dois níveis e, também, não são válidos para tratamentos resultantes de combinação fatorial, sejam fatores qualitativos ou quantitativos. Sugere-se, ainda, que nos casos em que os fatores são de natureza quantitativa, com mais de dois níveis, o emprego da análise de regressão é a alternativa mais apropriada e que, em termos gerais, consiste em ajustar uma função entre a resposta (Y) e os fatores em estudo (X).

Quanto aos tratamentos que apresentam certa relação, formular, antes da realização do experimento, as comparações de interesse e, posteriormente, procedidas por meio de contrastes ortogonais ou de interesse particular.

A seguir são detalhados exemplos para melhor entendimento da aplicação dos TCM.

Tratamentos de natureza quantitativa

A Tabela 2 apresenta dados de um trabalho em que o pesquisador estudou o efeito de níveis de calcário sobre o peso seco de nódulos, na cultura da soja.

Nesse exemplo a aplicação do TCM foi inadequada, pois os tratamentos (níveis de calcário) são quantitativos e o método apropriado e mais informativo deveria ter sido a análise de regressão.

Outro exemplo de uso inadequado da aplicação dos TCM foi extraído de trabalho de investigação do efeito de níveis de nitrogênio aplicado no solo, na cultura do feijão (Tabela 3). Nesse caso, o correto seria proceder a uma análise de regressão e, a partir de uma equação, identificar a dose de nitrogênio que proporcionasse o maior rendimento.

Tratamentos formados pela combinação dos níveis de dois ou mais fatores

O exemplo da Tabela 4 foi extraído de um trabalho em que se estudou os efeitos de dois fatores sobre os valores constituintes do sangue em bubalinos: os fatores são dois tratamentos (T_1 = controle e T_2 = experimental) e três níveis de estresse térmico (P_1 = pré-estresse, P_2 = estresse térmico e P_3 = pós-estresse).

TABELA 2. Efeito de níveis de calcário sobre o peso seco de nódulos, na cultura da soja.

Níveis de calcário (t/ha)	Peso seco de nódulos (mg/planta) ¹
0	77ab
2	94a
4	105a
6	56 b

¹Médias seguidas pela mesma letra não diferem entre si, pelo teste de Tukey, a 5% de probabilidade.

TABELA 3. Efeito de níveis de nitrogênio na cultura do feijão.

Dose de N (kg/ha)	Rendimento (kg/ha) ¹
0	487b
20	544b
40	640b
60	569b
80	853a
100	787a

¹Médias seguidas pela mesma letra não diferem entre si, pelo teste de Tukey a 5% de probabilidade.

Pode-se observar a aplicação parcialmente adequada dos TCM, embora o correto seria comparar as médias dos fatores principais isolados, no caso da interação tratamento (T) x tratamento (P) não significativa, ou comparar as médias dos níveis de tratamento (P) dentro de cada nível do tratamento (T) e vice-versa, no caso da interação significativa tratamento (T) x tratamento (P).

Ainda exemplificando a aplicação parcialmente adequada dos TCM nos experimentos fatoriais apresentam-se, na Tabela 5, os resultados de uma pesquisa na área animal, quando foram estudados, por

TABELA 4. Efeito do estresse térmico sobre constituintes do sangue, em bubalinos.

Tratamentos	Uréia (mg%) ¹
$T_1 P_1$	31,33a
$T_1 P_2$	31,68a
$T_1 P_3$	31,11a
$T_2 P_1$	31,06a
$T_2 P_2$	29,07b
$T_2 P_3$	31,53a

¹Médias seguidas pela mesma letra não diferem entre si, pelo teste de Tukey, a 5% de probabilidade.

TABELA 5. Efeito de raças, épocas e vermifugação sobre o número de eritrócitos em cordeiros.

Interações	Classificação	Eritrócitos/mm ³ (x10 ⁶) ¹
Raça x época	$R_1 E_1$	10,39c
	$R_1 E_2$	12,06a
	$R_2 E_1$	11,57abc
	$R_2 E_2$	11,31ab
	$R_3 E_1$	11,92bc
	$R_3 E_2$	11,22bc
Raça x vermifugação	$R_1 V_1$	10,52b
	$R_1 V_2$	11,93a
	$R_2 V_1$	10,79b
	$R_2 V_2$	12,15a
	$R_3 V_1$	10,36b
	$R_3 V_2$	12,79a
Época x vermifugação	$E_1 V_1$	10,12c
	$E_1 V_2$	12,42a
	$E_2 V_1$	11,01b
	$E_2 V_2$	12,14a

¹Valores para a mesma interação seguidos de letras distintas, diferem estatisticamente ($P < 0,05$) pelo teste Isd.

meio de um fatorial 3 x 2 x 2, os efeitos de três raças, duas épocas e dois níveis de vermifugação sobre o número de eritrócitos em cordeiros. Nesse caso, a alternativa mais correta seria comparar as médias em cada interação, de acordo com a orientação do exemplo apresentado na Tabela 4.

Tratamentos relacionados

Os exemplos apresentados a seguir são referentes à aplicação parcialmente adequada dos TCM, para tratamentos relacionados. Foi pesquisado o efeito de rações balanceadas em camundongos e os resultados estão apresentados na Tabela 6.

Poderia ter sido formulado importantes contrastes como: $C_1 = 3 m_A - m_B - m_C - m_D$ (farinha de trigo x (farinha de trigo + suplemento)); $C_2 = 2 m_D - m_B - m_C$ (leite em pó x suplemento) e $C_3 = m_B - m_C$ (endosperma x germe).

Em outro exemplo apresentado na Tabela 7 estudou-se o efeito do desfolhamento artificial sobre o

diâmetro da capitula do girassol e aplicou-se TCM sem considerar a natureza dos tratamentos. Observa-se que importantes contrastes poderiam ter sido formulados, tais como: $C_1 = 3 m_A - m_D - m_E - m_F$ (terço inferior x (terço inferior + complemento)).

Os TCM foram discutidos nos vários textos de metodologia estatística (Scheffé, 1959; Winer, 1971; Snedecor & Cochran, 1974; Dittrich, 1976; Steel & Torrie, 1980) porém a maioria se restringiu aos aspectos algébricos e computacionais, não enfatizando a adequação dos procedimentos estatísticos aos tipos de dados em estudo.

Uma orientação da aplicação adequada dos TCM pode ser encontrada nos artigos de Chew (1976); Petersen (1977); Carmer & Walker (1982); Percini & Malheiros (1989) e Cardellino & Siewerdt (1992).

CONCLUSÕES

1. Existe deficiência de verificação cuidadosa da adequabilidade dos testes de comparação de médias aos tipos de tratamentos estudados.

2. O uso inadequado dos testes de comparação de médias é mais comum nos casos em que os tratamentos são de natureza quantitativa.

3. Dos artigos publicados na revista PAB, de 1980 a 1994, em que se utilizou algum tipo de testes de comparações de médias, 57,01% usaram, de forma adequada, os testes de comparação de médias 31,53%, de forma inadequada, e 11,46%, parcialmente adequada.

TABELA 6. Efeito de ração balanceada em camundongo.

Tratamentos	Ganho de peso (g) ¹
Farinha de trigo (T _A)	34,2c
Farinha de trigo + endosperma de trigo (T _B)	50,0b
Farinha de trigo + germe de trigo (T _C)	85,2a
Farinha de trigo + leite em pó (T _D)	77,8a

¹Médias seguidas pela mesma letra não diferem entre si pelo teste de Tukey a 5% de probabilidade.

TABELA 7. Efeito do desfolhamento artificial sobre o diâmetro da capitula do girassol.

Desfolhamento	Diâmetro da capitula (cm) ¹
T _A = Terço inferior (T _i)	188,1a
T _B = Terço médio (T _m)	16,6b
T _C = Terço superior (T _s)	15,9c
T _D = Terço inferior + T _m	16,7b
T _E = Terço inferior + T _s	15,5bc
T _F = Terço inferior + T _m + T _s	5, 0e
T _G = Terço superior + T _m	7,9d
T _H = Sem remoção (S _s)	18,0a

¹Médias seguidas pela mesma letra não diferem entre si pelo teste de Duncan a 5% de probabilidade.

REFERÊNCIAS

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.

CARMER, S.G.; WALKER, W.M. Baby bear's dilemma: A statistical tale. *Agronomy Journal*, v.74, p.122-124, 1982.

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

DITTRICH, R.C. Estudo de alguns métodos de comparações de médias. Brasília: UNB, 1976. 108p. Dissertação de Mestrado.

- LITTLE, T.M. If Galileo published in HortScience. HortScience, v.13, n.5, p.504-506, 1978.
- NETER, J.; WASSERMAN, W. Applied linear statistical models. Womewood: Richard D. Irwin Inc., 1974. 842p.
- PERECINI, 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.
- PIMENTEL-GOMES, F. A estatística moderna na pesquisa agropecuária. 2.ed. Piracicaba: POTAFOS, 1984. 162p.
- SCHEFFÉ, H. The analysis of variance. London: John Wiley & Sons, 1959. 477p.
- SNEDECOR, G.W.; COCHRAN, W.G. Statistical methods. 6.ed. Ames: The Iowa State Univ. Press, 1974. 593p.
- STEEL, R.G.D.; TORRIE, J.H. Principles and procedures of statistics. 2.ed. London: McGraw-Hill Book Co., 1980. 633p.
- WINER, B.J. Statistical principles in experimental design. 2.ed. New York: McGraw-Hill Book Co., 1971. 907p.