

METODOLOGIA E TÉCNICAS EXPERIMENTAIS

DECOMPOSIÇÃO ORTOGONAL DE GRAUS DE LIBERDADE DE TRATAMENTOS UTILIZANDO VARIÁVEIS AUXILIARES E O PROC GLM DO SAS ⁽¹⁾

CRISTIANO NUNES NESI ⁽²⁾; ANTONIO AUGUSTO FRANCO GARCIA ⁽³⁾

RESUMO

Os testes de hipóteses têm sido amplamente utilizados, em análise de variância, para se concluir a respeito das fontes de variação. O teste F, em análise de variância para tratamentos com mais de um grau de liberdade, proporciona informações gerais, relacionadas com o comportamento médio dos tratamentos. Informações mais específicas são obtidas por meio de desdobramentos ortogonais dos graus de liberdade, comumente empregando contrastes. Entretanto, essa técnica torna-se complexa e de difícil implementação à medida que o número de tratamentos aumenta. Diante desse fato, utilizando-se como exemplo do uso da técnica, dados provenientes de um experimento de competição entre nove híbridos de milho, classificados em três tipos (simples, duplo e triplo), no delineamento em blocos casualizados, explicitou-se o emprego das variáveis auxiliares, que facilitam o desdobramento ortogonal dos graus de liberdade de tratamentos. Cada fator no modelo, após a decomposição, recebe uma codificação específica, tendo como valor, o nível do fator quando o tratamento for desse tipo, ou valor zero em caso contrário. O uso de variáveis auxiliares é uma forma simples e prática para determinar diferenças dentro dos grupos de interesse, embora não seja possível testar diretamente diferenças entre os grupos, o que pode também ser feito com a função CONTRAST do PROC GLM do SAS, mas não de forma tão simples.

Palavras-chave: desdobramento ortogonal, contrastes, variáveis auxiliares.

ABSTRACT

ORTHOGONAL DECOMPOSITION OF TREATMENT DEGREES OF FREEDOM USING AUXILIAR VARIABLES AND THE PROC GLM/SAS

Hypothesis tests have been largely used in analysis of variance in order to study sources of variation. The F test for treatments when there are more than one degree of freedom provides only a general information, related to the comparison of means. More specific information is obtained by orthogonal decomposition of the degrees of freedom, often using contrasts. However, this technique becomes complex and difficult to implement as the number of treatments increases. In order to overcome this problem it was studied auxiliary variables to test differences among groups and facilitate orthogonal

⁽¹⁾ Parte da dissertação de mestrado do primeiro autor, apresentada à Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo. Recebido para publicação em 25 de julho de 2003 e aceito em 20 de outubro de 2004.

⁽²⁾ Centro de Pesquisa para Agricultura Familiar - CEPAF/EPAGRI. Caixa Postal 791, 89801-970 Chapecó (SC) E-mail: cristiano@epagri.rct-sc.br

⁽³⁾ Departamento de Genética - ESALQ/USP, Caixa Postal 83, 13400-970 Piracicaba (SP). Bolsista do CNPq - Brasil. E-mail: aafgarci@esalq.usp.br

decomposition of treatment degrees of freedom. It was used as example, experimental data from a maize trial consisting of nine hybrids, classified as simple, double or triple assayed in randomized complete block design. After the decomposition, each factor in the model, receives a specific codification, the value referring the level of the factor when the treatment is of that type or, alternatively, zero if not. The use of auxiliary variables showed to be a useful and simple way to test differences between groups. This can be also done with the CONTRAST function of the PROC GLM of SAS software, but with a more complex procedure.

Key words: orthogonal decomposition, contrasts, auxiliary variables.

1. INTRODUÇÃO

Quando se aplica o teste F em análise de variância para a fonte de variação tratamentos com mais de um grau de liberdade, obtêm-se informações gerais, relacionadas com o comportamento médio dos tratamentos (BANZATTO E KRONKA, 1995). Se apenas uma das comparações envolve uma diferença marcante, um teste F médio pode falhar para evidenciar a diferença existente. Por essa razão, devem-se planejar comparações objetivas, fazendo-se desdobramentos ortogonais dos graus de liberdade de tratamentos para obter informações mais específicas, relacionadas com o comportamento médio de cada um dos componentes do desdobramento.

Após essa decomposição, pode-se aplicar o teste F aos componentes do desdobramento, e para I tratamentos pode-se estabelecer $(I-1)$ comparações independentes. Uma técnica de análise, usada na prática, baseia-se na utilização de contrastes, que devem ser ortogonais entre si para que as comparações sejam independentes. Entretanto, a utilização de contrastes com essa finalidade torna-se trabalhosa à medida que o número de tratamentos aumenta.

NOGUEIRA (2004) apresenta uma revisão bastante completa sobre contrastes ortogonais. Diversos autores como KEMPTHORNE (1952), SNEDECOR e COCHRAN (1981), BANZATTO E KRONKA (1995) e GOMES (2000) abordam o assunto detalhadamente. Basicamente, considere-se que $\mu_1, \mu_2, \dots, \mu_I$ são as verdadeiras médias dos I tratamentos de um experimento, todos eles com o mesmo número (n) de repetições. Qualquer contraste linear entre essas médias pode ser representado pela função:

$$A = \lambda_1 \mu_1 + \lambda_2 \mu_2 + \dots + \lambda_I \mu_I \quad \text{com} \quad \lambda_1 + \lambda_2 + \dots + \lambda_I = 0$$

A variância dessa função linear é $(\sum v_i^2) \sigma^2$. Suponha-se outro contraste linear B dado por:

$$B = v_1 \mu_1 + v_2 \mu_2 + \dots + v_I \mu_I \quad \text{com} \quad v_1 + v_2 + \dots + v_I = 0$$

Então, a variância de B é $(\sum v_i^2) \sigma^2$ e a covariância de A e B é $(\sum \lambda_i v_i) \sigma^2$. A covariância entre A e B será zero se $\sum \lambda_i v_i = 0$, condição necessária e

suficiente para que dois contrastes sejam ortogonais entre si. Isto significa que as comparações A e B serão independentes e, assim, usando A para estimar um parâmetro e B para estimar outro parâmetro, os erros na estimação dos respectivos parâmetros não serão relacionados. Se a distribuição não for normal, as quantidades A e B não serão correlacionadas se a condição for satisfeita, mas não necessariamente independentes. Conforme foi dito, para I tratamentos podem ser estabelecidas $(I-1)$ comparações independentes. Porém, esse conjunto de comparações não é único. Assim, devem ser consideradas as comparações de interesse do pesquisador, pois geralmente as comparações que se tinham em mente no planejamento do experimento são as mais informativas.

No PROC GLM do SAS, um pacote estatístico muito usado, existem as funções CONTRAST e ESTIMATE, com as quais se podem aplicar testes de hipóteses sobre combinações lineares dos parâmetros e estimar contrastes, permitindo ao usuário testar hipóteses de seu interesse. A função CONTRAST proporciona um mecanismo simples para obter teste de hipóteses de interesse do pesquisador, tais como $H_0: B'\theta = \phi$, em que B' é um vetor que indica a combinação linear estimável dos parâmetros contidos no vetor θ (SAS INSTITUTE, 1989). Para uma hipótese não testável ($B'\theta$ não estimável), o PROC GLM informa o problema ao usuário; quando a hipótese é testável (a função paramétrica é estimável), sua soma de quadrados é calculada e o valor da estatística F fornecido.

A função ESTIMATE é utilizada para estimar funções lineares dos parâmetros, multiplicando-se o vetor B pelo vetor de estimativas de mínimos quadrados dos parâmetros (θ), ou seja, calculando $B\theta$. A estimativa da função linear é verificada por meio de B' , isto é, verificando se B' é linearmente dependente ou independente das linhas/colunas da matriz do delineamento (SAS INSTITUTE, 1989). A seguir, é fornecido o valor da estatística t para o usuário testar se as estimativas obtidas são diferentes de zero. Assim, tanto para o uso da função CONTRAST quanto para a função ESTIMATE, é necessário que o usuário explicita os contrastes que deseja testar e/ou estimar.

SCOTT E MILLIKEN (1993) propuseram uma técnica, com variáveis auxiliares, para decomposição dos graus de liberdade de tratamentos na realização de análises de variância usando o PROC GLM do SAS no delineamento em blocos aumentados, sem a necessidade de explicitar os contrastes, uma vez que isso pode ser muito trabalhoso. Nessa situação, são dadas as somas de quadrados dos tipos I e III, mas somente as somas de quadrados do tipo III foram consideradas apropriadas pelos autores. PERECIN et al. (2000) apresentaram exemplos do uso de variáveis auxiliares para expressar desdobramentos de graus de liberdade e contrastes utilizando o sistema estatístico SAS, mostrando que as variáveis auxiliares podem ser utilizadas para modelos de experimentos fatoriais e de parcelas subdivididas, quando a interação for significativa. Além disso, com essa técnica é possível ajustar modelos de regressão quando uma variável tem níveis quantitativos e a interação entre os fatores é significativa.

Contudo, nenhum desses trabalhos aborda uma situação comum na experimentação agrícola, que são os experimentos realizados para comparar variedades comerciais. Em muitos desses experimentos, são encontrados grupos de tratamentos com mesmas características, sendo de interesse do pesquisador verificar se há diferenças entre e dentro desses grupos. Usualmente é feito com base em contrastes ortogonais que podem demandar grande tempo para serem apresentados.

O objetivo deste trabalho é propor o uso de uma técnica, empregando variáveis auxiliares e o PROC GLM do SAS, para facilitar o desdobramento ortogonal dos graus de liberdade de tratamentos, sem necessitar a construção de grande número de contrastes.

2. MATERIAL E MÉTODOS

Material

Os dados da Tabela 1, usados para ilustrar a técnica proposta, referem-se à produtividade (t.ha⁻¹) de um experimento fictício de avaliação de nove híbridos de milho, em que foi utilizado o delineamento em blocos completos casualizados.

Os tratamentos utilizados agrupam-se em três categorias: híbridos simples (HS), híbridos duplos (HD) e híbridos triplos (HT). O interesse dessa pesquisa está em verificar se há diferença de produtividade entre os três grupos de híbridos e entre os híbridos dentro de cada grupo.

É importante destacar que, em situações reais, o número de tratamentos avaliados é muito maior que o aqui apresentado, chegando a dezenas, ou até mesmo centenas, em muitos casos.

Métodos

Adotou-se, segundo IEMMA (1987), o modelo linear de Gauss-Markov Normal (efeitos fixos), dado por

$$y = X\theta + e \quad (1)$$

em que ${}_n y_1$: vetor de observações; ${}_n X_p$: matriz de "uns" e "zeros" que identifica o delineamento experimental; ${}_p \theta_1$: vetor de parâmetros desconhecidos; ${}_n e_1$: vetor de variáveis aleatórias não observáveis, tal que $e \sim N(\phi, I\sigma^2)$.

Tabela 1. Produtividade de híbridos de milho (t.ha⁻¹), no delineamento em blocos casualizados. Dados fictícios

Híbridos	Blocos			
	1	2	3	4
(1) HS1	3,4	3,0	3,1	3,2
(2) HS2	2,7	2,9	2,8	3,0
(3) HS3	2,6	2,5	2,1	2,4
(4) HD1	2,9	3,5	3,2	3,3
(5) HD2	2,6	2,8	3,0	3,1
(6) HD3	2,3	2,6	2,2	2,3
(7) HT1	3,1	3,3	3,4	3,0
(8) HT2	2,0	1,9	2,2	2,2
(9) HT3	3,2	3,4	3,3	3,1

Utilizando-se o PROC GLM de forma convencional, fez-se uma análise de variância (GOMES, 2000), calculando-se somas de quadrados para blocos, tratamentos e resíduo. Adotou-se como caracterização para o modelo linear:

$$y_{ij} = \mu + \beta_j + \tau_i + e_{ij} \quad (2)$$

em que y_{ij} : valor observado para o i -ésimo tratamento no j -ésimo bloco; μ : constante inerente a todas as observações; β_j : efeito do j -ésimo bloco; τ_i : efeito do i -ésimo tratamento (híbrido); e_{ij} : erro aleatório relativo à observação, y_{ij} tal que $e_{ij} \sim N(0, \sigma^2)$.

Essa análise de variância testa a hipótese de que os híbridos não diferem entre si, ou seja:

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6 = \mu_7 = \mu_8 = \mu_9, \text{ em que}$$

μ_1 : média de produção do HS1; μ_2 : média de produção do HS2; μ_3 : média de produção do HS3; μ_4 : média de produção do HD1; μ_5 : média de produção do HD2; μ_6 : média de produção do HD3; μ_7 : média de produção do HT1; μ_8 : média de produção do HT2; μ_9 : média de produção do HT3 (Tabela 1).

Como há interesse em testar se os híbridos simples, duplos e triplos diferem entre si dentro de cada grupo, propõe-se o uso de variáveis auxiliares da forma descrita a seguir. Os tratamentos foram divididos em três grupos e definidas três variáveis auxiliares (C, D e F): i) variável C: assume o nome dos híbridos simples quando o tratamento corresponde a esses tipos de híbridos e zero, caso contrário; ii) variável D: assume o nome dos híbridos duplos, quando o tratamento corresponde a esses tipos de híbridos e zero, caso contrário; iii) variável F: assume o nome dos híbridos triplos quando o tratamento corresponde a esses tipos de híbridos e zero caso contrário.

Assim, a caracterização do modelo linear foi reescrita em função das variáveis auxiliares:

$$y_{ij} = \mu + \beta_j + c_i + d_i + f_i + e_{ij} \quad (3)$$

em que y_{ij} : valor observado para o i -ésimo híbrido no j -ésimo bloco; μ : constante inerente a todas as observações; β_j : efeito do j -ésimo bloco; c_i : efeito dos híbridos simples; d_i : efeito dos híbridos duplos; f_i : efeito dos híbridos triplos; e_{ij} : erro aleatório relativo à observação y_{ij} .

A aplicação dessa técnica no PROC GLM do SAS viabiliza-se com a utilização de algumas linhas de programação de fácil implementação, apresentadas no Anexo A. As funções estimáveis dos tipos I, II, III e IV referentes a esse modelo foram calculadas pelo

PROC GLM para facilitar o entendimento das hipóteses que foram testadas.

No caso das variáveis auxiliares, três hipóteses foram testadas: i) Hipótese de que as médias dos híbridos simples não diferem entre si (variável auxiliar C), $H_0 : \mu_1 = \mu_2 = \mu_3$; ii) Hipótese de que as médias dos híbridos duplos não diferem entre si (variável auxiliar D), $H_0 : \mu_4 = \mu_5 = \mu_6$; iii) Hipótese de que as médias dos híbridos triplos não diferem entre si (variável auxiliar F), $H_0 : \mu_7 = \mu_8 = \mu_9$.

Com finalidade de comparação dessa forma de análise com a usualmente empregada, foram decompostos os oito graus de liberdade para tratamentos em contrastes ortogonais, da forma proposta originalmente pela função CONTRAST do PROC GLM (SAS INSTITUTE, 1989). Seguindo os objetivos do experimento foram construídos os seguintes contrastes ortogonais:

i) Híbridos simples versus Híbridos duplos e triplos:

$$Y_1 : 2\mu_1 + 2\mu_2 + 2\mu_3 - \mu_4 - \mu_5 - \mu_6 - \mu_7 - \mu_8 - \mu_9 = 0$$

ii) Híbridos simples 1 versus Híbridos simples 2 e 3:

$$Y_2 : 2\mu_1 - \mu_2 - \mu_3 = 0$$

iii) Híbridos simples 2 versus Híbridos simples 3:

$$Y_3 : \mu_2 - \mu_3 = 0$$

iv) Híbridos duplos versus Híbridos triplos:

$$Y_4 : \mu_4 + \mu_5 + \mu_6 - \mu_7 - \mu_8 - \mu_9 = 0$$

v) Híbrido duplo 4 versus Híbridos duplos 5 e 6:

$$Y_5 : 2\mu_4 - \mu_5 - \mu_6 = 0$$

vi) Híbrido duplo 5 versus Híbrido duplo 6:

$$Y_6 : \mu_5 - \mu_6 = 0$$

vii) Híbrido triplo 7 versus Híbridos triplos 8 e 9:

$$Y_7 : 2\mu_7 - \mu_8 - \mu_9 = 0$$

viii) Híbrido triplo 8 versus Híbrido triplo 9:

$$Y_8 : \mu_8 - \mu_9 = 0$$

O programa SAS usado é apresentado no Anexo B. Os contrastes ii e iii, v e vi, vii e viii, foram testados conjuntamente, formando hipóteses com dois graus de liberdade, porque o interesse nesse experimento é verificar se há diferenças dentro de cada grupo.

3. RESULTADOS E DISCUSSÃO

A Tabela 2 apresenta o resultado da análise de variância.

Tabela 2. Análise de variância para os dados de produtividade dos híbridos de milho (Tabela 1).

V.C.M.	G.L.	S.Q.	Q.M.	F	Pr > F
Blocos	$(b-1) = 3$	$R(\beta \mu, \tau) = 0,0733$	0,0244	0,69	0,5654
Híbridos	$(t-1) = 8$	$R(\tau \mu, \beta) = 6,3023$	0,7878	22,33	0,0001
Resíduo	$(b-1) \cdot (t-1) = 24$	$\Delta = 0,8467$	0,0353	-	-
Total	$(n-1) = 35$	$R(\mu, \beta, \tau) = 7,2222$	-	-	-

μ : efeito da constante; β : efeito de blocos; τ : efeito dos tratamentos; b : número de blocos; t : número de tratamentos; n : número de observações; $\Delta = R(\mu, \beta, \tau) - R(\beta | \mu, \tau) - R(\tau | \mu, \beta)$. $R(\mu, \beta, \tau)$: redução na soma de quadrados devido ao ajuste no modelo com μ , β , e τ ; $R(\beta | \mu, \tau)$: redução na soma de quadrados devido ao ajuste no modelo de β após μ e τ ; $R(\tau | \mu, \beta)$: redução na soma de quadrados devido ao ajuste no modelo de τ após μ e β .

Como o teste F para híbridos foi significativo ($Pr < 0,05$), rejeita-se H_0 e conclui-se que os híbridos possuem efeitos diferentes sobre a produtividade. Essa conclusão, por ser muito geral, nada informa com relação à comparação entre os híbridos de cada tipo (simples, duplos e triplos). Para obter essas informações, pode-se realizar um teste de comparação de médias, ou ainda decompor ortogonalmente os graus de liberdade para tratamentos. Na escolha dos componentes do desdobramento, deve-se considerar as comparações que interessam ao pesquisador. Supôs-se que a segunda alternativa é a de interesse no presente caso.

A matriz do delineamento obtida com emprego da técnica de variáveis auxiliares é apresentada no Anexo C. Diferentemente de SCOTT E MILIKEN (1993),

que usaram apenas as somas de quadrados do tipo III, por se tratar de delineamentos em blocos completos aumentados, no presente caso foram utilizadas as somas de quadrados do tipo II, III e IV, que para esse conjunto de dados e caracterização do modelo linear adotada, têm as respectivas funções estimáveis equivalentes (Tabela 3), resultando em somas de quadrados iguais. A função estimável do tipo I não é equivalente às demais, resultando em uma soma de quadrados diferente. A hipótese do tipo I é de interesse do pesquisador somente para o fator F, uma vez que não permite comparações dos níveis dos fatores C e D. Dessa forma, caso sejam empregadas variáveis auxiliares, devem ser utilizadas somente as somas de quadrados dos tipos II, III e IV, que testam hipóteses de interesse para todos os fatores.

Tabela 3. Funções estimáveis para as variáveis auxiliares C, D e F

		Tipo I			Tipo II = Tipo III = Tipo IV		
		Fator C	Fator D	Fator F	Fator C	Fator D	Fator F
μ		0	0	0	0	0	0
Blocos	1	0	0	0	0	0	0
	2	0	0	0	0	0	0
	3	0	0	0	0	0	0
	4	0	0	0	0	0	0
C	0	L6	0	0	0	0	0
	1	L7	0	0	L7	0	0
	2	L8	0	0	L8	0	0
	3	-L6-L7-L8	0	0	-L7-L8	0	0
D	0	-0,5L6	L10	0	0	0	0
	4	0,167L6	L11	0	0	L11	0
	5	0,167L6	L12	0	0	L12	0
	6	0,167L6	-L10-L11-L12	0	0	-L11-L12	0
F	0	-0,5L6	-L10	0	0	0	0
	7	0,167L6	0,333L10	L15	0	0	L15
	8	0,167L6	0,333L10	L16	0	0	L16
	9	0,167L6	0,333L10	-L15-L16	0	0	-L15-L16

Diversos autores abordam a questão dos quatro tipos de somas de quadrados (OVERALL e SPIEGEL (1969); SPEED et al. (1978); SEARLE (1987); SAS INSTITUTE (1989); MONDARDO (1994); IEMMA (1995); CAMARINHA FILHO (1995)). Segundo MONDARDO (1994) e MONDARDO e IEMMA (1998), os quatro tipos de somas de quadrados testam o seguinte: a soma de quadrados do tipo I (SQI) testa hipóteses sobre médias ponderadas ajustadas para os efeitos precedentes no modelo; a SQII testa hipóteses sobre médias ponderadas ajustadas para todos os demais efeitos, exceto para interações; a SQIII testa hipótese sobre as médias não ponderadas de linhas e/ou colunas; a SQIV testa hipóteses construídas pelo PROC GLM do SAS, como função das médias não ponderadas e dependem da posição e do número de caselas vazias.

De acordo com o SAS INSTITUTE (1989), o número de letras diferentes dentro de cada tipo de função estimável e fator considerado, define o número de graus de liberdade associado à hipótese. Por exemplo, para o fator C, tem-se três graus de liberdade (letras L6, L7 e L8). As hipóteses testadas podem ser obtidas atribuindo-se valores arbitrários a esses coeficientes. Nota-se que para os fatores C e D, em funções estimáveis do tipo I, não é possível construir hipóteses específicas para os fatores em questão, razão pela qual as funções estimáveis de interesse são dos tipos II, III e IV. Para o fator C, a primeira linha da hipótese foi escrita atribuindo-se arbitrariamente valor 1 para o coeficiente L7 e zero para o coeficiente L8. Em seguida, atribui-se valor 1 para o coeficiente L8 e zero para o coeficiente L7, obtendo-se a seguinte hipótese, com dois graus de liberdade:

$$H_0 : \begin{cases} \mu_{1.} - \mu_{3.} = 0 \\ \mu_{2.} - \mu_{3.} = 0 \end{cases}$$

Portanto, a soma de quadrados para essa hipótese está associada a um teste de que as médias de produtividade dos híbridos simples não diferem entre si, que é de interesse nesse experimento. Essa hipótese é equivalente à obtida com a decomposição ortogonal com dois graus de liberdade:

Híbrido simples 1 versus Híbrido simples 2, Híbrido simples 3 (1 g.l.), e

Híbrido simples 2 versus Híbrido simples 3 (1 g.l.)

As hipóteses para os fatores D e F foram construídas atribuindo-se valor 1 para o coeficiente L11 e zero para o coeficiente L12. Em seguida, atribui-se valor 1 para o coeficiente L12 e zero para o coeficiente L11 procedendo-se de forma análoga. Assim, tem-se para o fator a hipótese com 2 g.l.:

$$H_0 : \begin{cases} \mu_{4.} - \mu_{6.} = 0 \\ \mu_{5.} - \mu_{6.} = 0 \end{cases}$$

O teste associado a essa hipótese verifica se as médias de produtividade dos híbridos duplos não diferem entre si, sendo também de interesse. Os dois graus de liberdade para D (Tabela 4) podem ser decompostos analogamente, com dois graus de liberdade, em:

Híbrido duplo 1 versus Híbrido duplo 2, Híbrido duplo 3 (1 g.l.), e

Híbrido duplo 2 versus Híbrido duplo 3 (1 g.l.)

Para o fator F:

$$H_0 : \begin{cases} \mu_{7.} - \mu_{9.} = 0 \\ \mu_{8.} - \mu_{9.} = 0 \end{cases}$$

O teste está associado à hipótese de que as médias de produtividade dos híbridos triplos não diferem entre si. Decompondo-se os graus de liberdade de em contrastes com dois graus de liberdade, tem-se:

Híbrido triplo 1 versus Híbrido triplo 2, Híbrido triplo 3 (1 g.l.)

Híbrido triplo 2 versus Híbrido triplo 3 (1 g.l.)

Nota-se claramente que o emprego das variáveis auxiliares permitiu que as hipóteses de interesse fossem testadas, sem necessitar construir os contrastes. A análise de variância, apresentada na Tabela 5, têm os oito graus de liberdade no modelo com efeito τ_i decompostos em dois graus de liberdade para híbridos simples (C), 2 graus de liberdade para híbridos duplos (D) e 2 graus de liberdade para híbridos triplos (F). Neste caso, restam ainda dois graus de liberdade, que correspondem às comparações entre os grupos HS, HD e HT.

Os contrastes que representam estas comparações podem ser testados facilmente por meio da função CONTRAST do PROC GLM, considerando conjuntamente os contrastes YHS vs HD, HT e YHD vs HT. É importante destacar que essa hipótese tem 2 graus de liberdade, correspondendo ao número de grupos menos 1. Assim, mesmo que o número de tratamentos seja elevado, é possível construí-la facilmente no SAS e completar o quadro da análise da variância.

Tabela 4. Análise de variância para o modelo .

V.C.M.	G.L.	S.Q.	Q.M.	F	Pr > F
Blocos	$(b-1) = 3$	$R(\beta \mu, \tau) = 0,0733$	0,0244	0,69	0,5654
C	$(c'-1) = 2$	$R(c \mu, \beta, d, f) = 1,2117$	0,6058	17,17	0,0001
D	$(d'-1) = 2$	$R(d \mu, \beta, c, f) = 1,5517$	0,7758	21,99	0,0001
F	$(f'-1) = 2$	$R(f \mu, \beta, c, d) = 3,5317$	1,7658	50,06	0,0001
Grupos*	$(NG-1) = 2$	$R(g \mu, \beta, \tau) = 0,0072$	0,0036	0,10	0,9031
Resíduo	$(b-1) \cdot (t-1) = 24$	$\Delta = 0,8467$	0,0353	-	-
Total*	$(n-1) = 35$	$R(\mu, \beta, \tau) = 7,2222$	-	-	-

μ : efeito da constante. β : efeito de blocos. g : efeito de grupos. τ : efeito dos tratamentos; c : efeito dos HS. d : efeito dos HD. f : efeito dos HT. b : número de blocos. c' : número de níveis que assume a variável auxiliar C, exceto c_0 ; d' : número de níveis que assume a variável auxiliar D, exceto d_0 ; f' : número de níveis que assume a variável auxiliar F, exceto f_0 . $\Delta = R(\mu, \beta, \tau) - R(\beta | \mu, \tau) - R(\tau | \mu, \beta)$; *Grupos: testa simultaneamente os contrastes Y_{HS} vs Y_{HD} , Y_{HT} e Y_{HD} vs Y_{HT} ; NG = número de grupos.

Como o teste F para os fatores C, D e F foi significativo (Tabela 4), conclui-se que há diferença entre os efeitos de pelo menos dois híbridos de cada um dos grupos. Ressalta-se que essas são as hipóteses de interesse nesse experimento e foram facilmente construídas.

No caso dos contrastes ortogonais construídos por meio da função CONTRAST do PROC GLM, foi possível repetir os mesmos resultados obtidos com as variáveis auxiliares. Contudo, conforme observa-se no programa SAS apresentado no Anexo B, são necessárias várias linhas de programação, representando os contrastes ortogonais. Esse número pode causar dificuldades na implementação dessa abordagem. Na Tabela 5 são apresentadas as análises de variância com desdobramentos dos graus de liberdade para híbridos, fornecidas pela função CONTRAST do PROC GLM. Obviamente, têm-se as mesmas conclusões que aquelas obtidas com as variáveis auxiliares.

Realizar a análise empregando contrastes requer a construção de todos os contrastes de interesse e, por meio de suas funções paramétricas, testá-los utilizando a função CONTRAST. Este procedimento é simples para experimentos pequenos, com poucos tratamentos, mas torna-se trabalhoso para um grande número de tratamentos. Embora seja possível testar as hipóteses de interesse, muitas vezes essa tarefa é trabalhosa e pode ser facilmente substituída com emprego das variáveis auxiliares. Por exemplo, caso o experimento tenha 100 tratamentos, representando dois grupos com 50 tratamentos cada um, o teste das diferenças dentro de cada grupo necessitaria que fossem construídos 49 contrastes ortogonais, implicando no final 98 linhas adicionais no programa. Com uso das variáveis auxiliares, seriam necessárias apenas duas linhas de programação. É importante salientar que tais variáveis são facilmente definidas em cada situação específica, bastando para tanto adaptar o modelo apresentado em (3) e o programa no Anexo A à situação específica.

Tabela 5. Análise de variância com desdobramento dos graus de liberdade de híbridos, fornecida pela função CONTRAST do PROC GLM

Contraste	GL	SQ Contrastos	QM	F	Pr > F
Blocos	3	0,0733	0,0244	0,69	0,5654
Híbridos	8	6,3023	0,7878	22,33	0,0001
(Dentro de HS)	(2)	(1,2117)	(0,6058)	(17,16)	(0,0001)
(Dentro de HD)	(2)	(1,5516)	(0,7758)	(21,98)	(0,0001)
(Dentro de HT)	(2)	(3,5316)	(1,7658)	(50,02)	(0,0001)
Resíduo	24	0,8467	0,0353	-	-
Total	35	7,2220	-	-	-

HS1: híbrido simples 1; HS2: híbrido simples 2; HS3: híbrido simples 3; HD1: híbrido duplo 1; HD2: híbrido duplo 2; HD3: híbrido duplo 3; HT1: híbrido triplo 1; HT2: híbrido triplo 2; HT3: híbrido triplo 3.

Com base no exposto, nota-se que o sistema estatístico SAS nem sempre fornece automaticamente o teste para hipóteses de interesse imediato, pois em muitas situações há interesse em desdobrar graus de liberdade de tratamentos. Para os dados da Tabela 1, deseja-se verificar se os híbridos diferem entre si, dentro de cada grupo. Nessa situação, o PROC GLM do SAS oferece como opção a construção de contrastes e o teste por meio da função CONTRAST. Entretanto, quando o número de elementos dentro de cada grupo é elevado, torna-se trabalhoso construir os contrastes entre eles. Diante disso, propõe-se o uso de variáveis auxiliares como uma forma simples para minimizar esses problemas, principalmente para determinar diferenças dentro dos grupos. Embora por meio de variáveis auxiliares não seja possível determinar diferenças entre os grupos, sendo necessário utilizar a função CONTRAST, isso é facilmente implementado em situações práticas, uma vez que a maior dificuldade é construir os contrastes dentro de cada grupo, e não entre os grupos.

4. CONCLUSÕES

1) O uso da técnica de variáveis auxiliares facilita o desdobramento ortogonal dos graus de liberdade de tratamentos para a comparação dos tratamentos dentro de grupos pré-estabelecidos;

2) Essa técnica tem resultados equivalentes aos obtidos utilizando-se a função CONTRAST do PROC GLM do SAS sendo, contudo, mais facilmente implementada.

REFERÊNCIAS

- BANZATTO, D.A.; KRONKA, S. N. **Experimentação agrícola**. 3.ed. Jaboticabal: FUNEP, 1995. 245p.
- CAMARINHA FILHO, J.A. **Testes de hipóteses em modelos lineares com dados desbalanceados e caselas vazias**. 1995. 142f. Dissertação (Mestrado) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba.
- IEMMA, A.F. **Modelos lineares: uma introdução para profissionais da pesquisa agropecuária**. 2.ed. Londrina: Imprensa Oficial do Estado do Paraná, 1987. 263p.
- GOMES, F. P. **Curso de estatística experimental**. 14.ed. Piracicaba: Frederico Pimentel Gomes, 2000. 477p.
- IEMMA, A.F. Que hipóteses estatísticas testamos através do SAS em presença de caselas vazias? **Scientia Agricola**, Piracicaba, v.52, n.2, p. 210-220, 1995.
- KEMPTHORNE, O. **The design and analysis of experiments**. New York: John Wiley e Sons, 1952. 631p.
- MONDARDO, M. **Estimabilidade de funções paramétricas com dados desbalanceados através do PROC GLM do SAS: Aplicações à pesquisa agropecuária**. 1994. 166f. Dissertação (Mestrado) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba.
- MONDARDO, M.; IEMMA, A. F. Considerações sobre as funções estimáveis fornecidas pelo PROC SAS/GLM para dados desbalanceados. **Scientia Agricola**, Piracicaba, v.55, n.2, p.172-182, 1998.
- NOGUEIRA, M.C.S. Orthogonal contrasts: definitions and concepts. **Scientia Agricola**, Piracicaba, v.61, n.1, p. 118-124, 2004.
- OVERALL, J.E.; SPIEGEL, D.K. Concerning least squares analysis of experimental data. **Psychological Bulletin**, Washington, v.72, p.311-322, 1969.
- PERECIN, D.; MALHEIROS, E. B.; PEREIRA, G.T. Variáveis auxiliares para expressar desdobramentos de graus de liberdade e contrastes com o programa SAS. In: REUNIÃO ANUAL DA REGIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA, 45. , 2000. São Carlos. **Anais...** São Carlos: 45ªRBRAS, 2000. p.137-140.
- SAS INSTITUTE. **SAS/STAT user's guide**. Version 6. 4. ed. Cary, 1989. v.2, 846p.
- SCOTT, R.A.; MILLIKEN, G.A. A SAS program for analysing augmented randomized complete block design. **Crop Science**, Madison, v.33, p.865-867, 1993.
- SEARLE, S. R. **Linear models for unbalanced data**. New York: John Wiley, 1987. 536p.
- SNEDECOR, G.W.; COCHRAN, W.G. **Statistical methods**. 70.ed. Ames: The Iowa State University, 1981. 507p.
- SPEED, F.M.; HOCKING, R.R.; HACKNEY, O.P. Methods of analysis of linear models with unbalanced data. - Alexandria, v. 73, p.105-112, 1978.

ANEXO A

Programa SAS para análise do experimento apresentado na Tabela 1 utilizando variáveis auxiliares.

```
data HIBRI;
input hib bloco prod;
c=hib; if hib>3 then c="0";
d=hib; if 4>hib or hib>6 then d="0";
f=hib; if hib<7 then f="0";
cards;
 1 1 3.4
 1 2 3.0
 1 3 3.1
 1 4 3.2
 2 1 2.7
 2 2 2.9
 2 3 2.8
 2 4 3.0
 3 1 2.6
 3 2 2.5
 3 3 2.1
 3 4 2.4
 4 1 2.9
 4 2 3.5
 4 3 3.2
 4 4 3.3
 5 1 2.6
 5 2 2.8
 5 3 3.0
 5 4 3.1
 6 1 2.3
 6 2 2.6
 6 3 2.2
 6 4 2.3
 7 1 3.1
 7 2 3.3
 7 3 3.4
 7 4 3.0
 8 1 2.0
 8 2 1.9
 8 3 2.2
 8 4 2.2
 9 1 3.2
 9 2 3.4
 9 3 3.3
 9 4 3.1
;
proc print;
proc glm;
class bloco c d f;
model prod=bloco c d f/ss1 ss2 ss3 ss4;
run;
```

ANEXO B

Programa SAS para análise do experimento apresentado na Tabela 1 utilizando contrastes.

```

data HIBRI;
input hib bloco prod;
cards;
  1 1 3.4
  1 2 3.0
  1 3 3.1
  1 4 3.2
  2 1 2.7
  2 2 2.9
  2 3 2.8
  2 4 3.0
  3 1 2.6
  3 2 2.5
  3 3 2.1
  3 4 2.4
  4 1 2.9
  4 2 3.5
  4 3 3.2
  4 4 3.3
  5 1 2.6
  5 2 2.8
  5 3 3.0
  5 4 3.1
  6 1 2.3
  6 2 2.6
  6 3 2.2
  6 4 2.3
  7 1 3.1
  7 2 3.3
  7 3 3.4
  7 4 3.0
  8 1 2.0
  8 2 1.9
  8 3 2.2
  8 4 2.2
  9 1 3.2
  9 2 3.4
  9 3 3.3
  9 4 3.1
;
proc glm;
class bloco hib;
model prod=bloco hib/ss3;
contrast "Y(dentro HS)"      HIB 2 -1 -1 0 0 0 0 0 0,
HIB 0 1 -1 0 0 0 0 0 0;
contrast "Y(dentro HD)"      HIB 0 0 0 2 -1 -1 0 0 0,
HIB 0 0 0 0 1 -1 0 0 0;
contrast "Y(dentro HT)"      HIB 0 0 0 0 0 0 2 -1 -1,
HIB 0 0 0 0 0 0 0 1 -1;
contrast "Y(HSvsHD, HT e HDvsHT)" HIB 2 2 2 -1 -1 -1 -1 -1 -1,
HIB 0 0 0 1 1 1 -1 -1 -1;
run;

```

