

## A CONSIDERAÇÃO DA ESTRUTURA DAS UNIDADES EM INFERÊNCIAS DERIVADAS DO EXPERIMENTO<sup>1</sup>

JOÃO GILBERTO CORRÊA DA SILVA<sup>2</sup>

RESUMO - O modelo estatístico deve exprimir corretamente a estrutura do experimento. Isso é necessário para garantir que os componentes de variância que afetam efeitos referentes a fatores experimentais sejam idênticos aos componentes de variância usados para julgar a significância desses efeitos, exceto pelas próprias variâncias atribuíveis a esses efeitos. O modelo estatístico usualmente formulado ignora a estrutura das unidades que resulta de restrições à casualização. Como consequência, propriedades que decorrem da casualização não são apropriadamente levadas em conta, e inferências podem se tornar tendenciosas. Sugere-se um procedimento para identificação dos efeitos referentes à estrutura das unidades, e sua consideração no modelo estatístico e em inferências derivadas do experimento. Em particular, é proposto um algoritmo para a determinação prática dos valores esperados de quadrados médios que levam em conta apropriadamente a estrutura do experimento.

Termos para indexação: estrutura do experimento, modelo estatístico, análise da variação, valores esperados de quadrados médios, inferência do experimento.

### THE CONSIDERATION OF THE UNIT STRUCTURE IN INFERENCES FROM THE EXPERIMENT

ABSTRACT - The statistical model should correctly express the structure of the experiment. This is necessary to guarantee that the variance components which affect the experimental factor effects are identical to the variance components used to judge the significance of these effects, except for the variance attributable to these effects. The usually formulated statistical model ignores the unit structure which results from restrictions to randomization. Consequently, properties implied by randomization are not appropriately taken into account and inferences may turn out biased. A procedure for identifying the effects related to the unit structure and considering them in the statistical model and in inferences from the experiment is suggested. Particularly, an algorithm for the practical derivation of the expected values of mean squares which take into account the experiment structure is proposed.

Index terms: experiment structure, statistical model, analysis of variance, expected values of mean squares, inference from experiment.

---

<sup>1</sup> Aceito para publicação em 18 de agosto de 1998.

<sup>2</sup> Eng. Agr., Ph.D., Prof. Titular, Dep. de Matemática, Estatística e Computação, Inst. de Física e Matemática, Univ. Fed. de Pelotas, Caixa Postal 354, CEP 96010-900 Pelotas, RS. Bolsista do CNPq. E-mail: jgcs@ufpel.tche.br

## INTRODUÇÃO

As restrições à casualização estabelecem uma relação entre a estrutura dos fatores experimentais e as unidades elementares do experimento (unidades de observação). Essas restrições têm implicações para as propriedades derivadas da casualização, já que o erro experimental resulta, por consequência, estruturado.

O modelo estatístico usualmente formulado não distingue efeitos de fatores experimentais e efeitos de unidade. Como consequência, por exemplo, atribui a fatores experimentais intrínsecos efeitos de unidade que com eles estão confundidos, de modo que inferências referentes a esses fatores resultam tendenciosas.

Essa falha de especificação do modelo estatístico tem implicações para inferências derivadas do experimento. Em particular, os valores esperados de quadrados médios (*EQM*) resultam incorretos e, conseqüentemente, também podem se tornar incorretos os procedimentos de inferência baseados no método dos momentos.

Assim, por exemplo, o modelo estatístico formulado da forma usual não distingue delineamentos de experimentos com dois fatores experimentais completamente casualizado, e blocos casualizados generalizados (com mais de uma repetição por bloco) e blocos casualizados com subamostragem, ambos com um dos fatores em blocos.

Essa falha tem duas origens: a) ausência de distinção entre as duas categorias de fatores experimentais, ou seja, fatores de tratamentos, cujos níveis são atribuídos aleatoriamente às unidades elementares, sob o controle do pesquisador, e fatores intrínsecos, cujos níveis são determinados pelas próprias unidades; e b) ausência de distinção entre fatores experimentais e fatores de unidade, estes constituídos por agrupamentos das unidades elementares decorrentes de controle local, de fatores intrínsecos e de conveniência.

Fisher (1935) distingue duas estruturas independentes: uma, relacionada às questões de interesse a serem respondidas através do experimento, denominada de “estrutura de tratamentos”, e a outra, associada à classificação das unidades experimentais segundo características dessas unidades, que ele denomina “topográfica”. Ele observa que o delineamento experimental pode ser entendido como a relação dessas duas estruturas no experimento, determinada pela casualização.

Esse conceito de delineamento experimental, ignorado por muitos anos, foi retomado por Nelder (1965a, 1965b), e, a partir de então, tem sido explorado em alguns trabalhos (Houtman, 1980; Beltrão & Pinheiro, 1989).

A distinção entre fator de tratamento e fator intrínseco, entretanto, tem sido geralmente ignorada, com poucas exceções (Cox, 1958; Cox & Snell, 1981). Esses autores, entretanto, não fazem uso dessa distinção nos procedimentos de inferência estatística.

Como decorrência, é geralmente ignorada a existência de “fatores parceiros” quando estão presentes no experimento fatores intrínsecos, como, por exemplo, local, ano e bloco, constituindo classificações das unidades de interesse no experimento. Esse fato tem conseqüências para as inferências deri-

vadas do experimento, decorrentes da impossibilidade de distinção de delineamentos experimentais com diferentes estruturas, conforme ilustrado anteriormente.

A casualização permite duas propriedades fundamentais ao delineamento experimental: o controle, em um sentido estatístico, do erro experimental, e a determinação de estimativas válidas das incertezas das inferências (Wilk & Kempthorne, 1956). Entretanto, essas propriedades requerem que o processo de atribuição aleatória dos tratamentos às unidades experimentais e, particularmente, as restrições envolvidas nesse processo, sejam adequadamente consideradas no modelo estatístico e nos conseqüentes procedimentos de inferência.

Fisher (1935) salienta a importância da consideração correta da estrutura do experimento, quando enfatiza que ela deve determinar completamente os procedimentos estatísticos para inferência, em particular para a determinação de estimativas válidas do erro experimental.

Essa importância também é enfatizada por Addelman (1970), ao salientar que o primeiro passo no esboço da análise de um experimento, que deve acompanhar o plano do experimento, é a dedução de um modelo estatístico apropriado para a situação experimental. O modelo estatístico deve conter termos para representar todas as fontes de variação presentes no experimento. Completa Addelman (1970), afirmando que é irreal supor que um modelo estatístico seja apropriado para mais do que um delineamento experimental.

Muito freqüentemente, a discussão de um experimento particular é baseada em um modelo estatístico, sem uma descrição clara da estrutura do experimento que o modelo é suposto representar. Em particular, descrições referentes à condução de experimentos são, muitas vezes, vagas, relativamente à forma da atribuição dos tratamentos às unidades experimentais. Essa vaguidade é freqüentemente estendida à análise pelo uso da palavra “célula” para denotar um grupo de unidades tratadas de modo semelhante: uma “célula” pode ser um conjunto de unidades experimentais às quais é atribuído um mesmo tratamento, ou um conjunto de unidades de observação em uma mesma unidade experimental que recebe um tratamento específico. Dessa forma, a designação “classificação dupla com E observações por célula” não descreve adequadamente uma situação experimental.

Este artigo tem por finalidade discutir as falhas que podem decorrer da forma usual de especificação do modelo estatístico, e sugerir um procedimento para a formulação do modelo estatístico que represente corretamente a estrutura do experimento, levando em conta os efeitos de unidade.

### MODELO ESTATÍSTICO FORMULADO USUALMENTE

Para ilustração, considere-se um experimento com dois fatores de interesse, ou seja, dois fatores experimentais,  $A$  e  $B$  com um mesmo número  $E$  de observações para cada uma das combinações dos  $A$  níveis do fator  $A$  e dos  $B$  níveis do fator  $B$ .

Dados provenientes de um experimento com tal estrutura fatorial são referidos em alguns textos como constituindo uma classificação dupla balanceada (ou com número igual de observações nas células) (Kempthorne, 1952; Li,

1964; Snedecor & Cochran, 1967), ou classificação cruzada de dois fatores (ou classificação dupla cruzada), com igual número de observações nas células (Johnson & Leone, 1964; Brownlee, 1965; Gill, 1978).

A equação do modelo estatístico para experimentos com essa especificação é apresentada em muitos textos na forma:

$$Y_{abe} = m + a_a + b_b + ab_{ab} + e_{abe}, \quad a=1,2,\dots,A, \quad b=1,2,\dots,B, \quad e=1,2,\dots,E,$$

onde  $m$  é a média geral esperada,  $a_a$ ,  $b_b$  e  $ab_{ab}$  são os efeitos principais e a interação dos fatores  $A$  e  $B$  e  $e_{abe}$  é o termo do erro.

Genericamente, pressupõe-se que os  $A$  níveis do fator  $A$  são uma seleção aleatória dos  $A$  níveis populacionais deste fator, e que os  $B$  níveis constituem uma amostra aleatória dos  $B$  níveis populacionais do fator  $B$ . Estas pressuposições genéricas abrangem as situações de modelos fixo, aleatório e misto; por exemplo,  $A=A$ ,  $A<A$  e  $A<<A$  ( $A$  consideravelmente maior que  $A$ ) correspondem, respectivamente, às situações em que o fator  $A$  é fixo, aleatório com número de níveis populacionais finito e aleatório com número de níveis populacionais infinito, respectivamente.

Adotem-se, ademais, as seguintes definições e notações:

$$\sum_{a=1}^A a_a = \sum_{b=1}^B b_b = \sum_{a=1}^A ab_{ab} = \sum_{b=1}^B ab_{ab} = 0,$$

$$\sigma_A^2 = \frac{1}{A-1} \sum_{a=1}^A a_a^2, \quad \sigma_B^2 = \frac{1}{B-1} \sum_{b=1}^B b_b^2, \quad \sigma_{AB}^2 = \frac{1}{(A-1)(B-1)} \sum_{a=1}^A \sum_{b=1}^B ab_{ab}^2$$

$$\text{e } \sigma_E^2 = E(e_{abe}^2),$$

e a pressuposição de que as covariâncias de dois distintos efeitos correspondentes a um mesmo termo do modelo estatístico e dos efeitos de dois distintos termos são todas nulas.

Sob essas condições, as expressões dos  $EQM$  são as apresentadas na Tabela 1. Essas expressões podem ser facilmente obtidas pelo emprego de

**TABELA 1.  $EQM$  para experimento com estrutura de dados correspondente à classificação dupla balanceada com dois fatores experimentais  $A$  e  $B$ , ignorando efeitos de unidade.**

Fonte de variação / efeito		$EQM$
$A:$	$a_a$	$\sigma_E^2 + (1 - \frac{B}{B})E\sigma_{AB}^2 + BE\sigma_A^2$
$B:$	$b_b$	$\sigma_E^2 + (1 - \frac{A}{A})E\sigma_{AB}^2 + AE\sigma_B^2$
$AB:$	$ab_{ab}$	$\sigma_E^2 + E\sigma_{AB}^2$
$E:$	$e_{abe}$	$\sigma_E^2$

métodos práticos (Bennett & Franklin, 1954; Schultz Junior, 1955; Cornfield & Tukey, 1956; Hicks, 1973).

A situação experimental que origina esse modelo estatístico e os consequentes processos de inferência estatística é descrita pela estrutura dos dados, que podem ser dispostos em uma tabela de dupla-entrada com um número igual de observações nas células. Ela ignora o delineamento experimental e, conseqüentemente, origina um modelo estatístico incompletamente especificado.

Esse modelo estatístico e os consequentes *EQM* da Tabela 1 são comuns para experimentos com qualquer dos três seguintes delineamentos, que correspondem a três distintos processos de atribuição aleatória dos níveis dos fatores *A* e *B* às unidades elementares:

### 1) Completamente casualizado com dois fatores

As *AB* combinações dos níveis dos fatores *A* e *B* são atribuídas de modo completamente aleatório às *ABE* unidades elementares do experimento (parcelas), de modo que cada combinação resulta em *E* dessas unidades.

### 2) Blocos casualizados generalizados

As *ABE* unidades elementares são agrupadas em *B* blocos de *AE* dessas unidades (parcelas); a cada um desses blocos corresponde um dos níveis do fator *B*; cada um dos *A* níveis do fator *A* é atribuído aleatoriamente a *E* das *AE* unidades de cada bloco, com atribuição separada e independente para cada bloco.

### 3) Blocos casualizados com subamostragem

*AB* unidades são agrupadas em *B* blocos de *A* unidades (parcelas); a cada bloco corresponde um nível do fator *B*; os *A* níveis do fator *A* são atribuídos aleatoriamente às *A* unidades de cada um dos blocos, com casualização separada e independente para cada bloco; são registradas observações em *E* subdivisões (ou amostras) de cada unidade, que constituem as unidades elementares do experimento.

Na situação do delineamento completamente casualizado, como as combinações dos níveis dos fatores *A* e *B* são atribuídas às unidades experimentais de modo completamente aleatório, é natural que o erro para inferências referentes aos efeitos principais e à interação dos fatores *A* e *B* provenha da variação entre as unidades experimentais individuais. Entretanto, no tocante ao delineamento blocos casualizados generalizados com o fator *B* em blocos, a restrição à casualização impõe atribuição aleatória dos níveis do fator *A* dentro de cada bloco, separada e independente em relação a cada bloco. Nessas circunstâncias, é de esperar que o erro para inferências referentes ao fator *A* provenha da variação interna a blocos e o erro para inferências referentes ao fator *B*, da variação externa aos blocos.

Os *EQM* da Tabela 1 revelam que esses requisitos para o delineamento completamente casualizado são satisfeitos pela adoção do modelo especificado, mas não são satisfeitos para o delineamento blocos casualizados generalizados no que se refere ao fator *B*.

Assim, aquela especificação do modelo estatístico é apropriada para o delineamento completamente casualizado, mas não o é para o delineamento blocos casualizados generalizados. Por semelhante argumento, pode-se observar que aquele modelo também é inadequado para o delineamento blocos casualizados com subamostragem.

Para ilustração, suponha-se um experimento com o objetivo de pesquisar os efeitos do tempo (fator *A*) e da temperatura de armazenamento (fator *B*) sobre a qualidade da semente de uma cultivar de trigo. *AE* porções de sementes da cultivar, de uma mesma procedência, são armazenadas em cada um de *B* ambientes, cada ambiente com uma diferente temperatura. Em cada um de *A* tempos de armazenamento, *E* porções de sementes são retiradas aleatoriamente do conjunto das *AE* porções de cada um dos *B* ambientes para avaliação.

Essa é, formalmente, a situação de delineamento blocos casualizados generalizados, com o fator *B* em blocos. O erro experimental apropriado para inferências referentes à temperatura é a variação entre ambientes. Entretanto, como cada temperatura é atribuída a apenas um ambiente, a variação atribuível à temperatura fica completamente confundida com a variação entre ambientes não devida à temperatura, ou seja, com uma fonte de variação do erro experimental. Nessas circunstâncias, o experimento não permitirá a derivação de inferências válidas com referência à temperatura.

Observe-se que, em geral, se cada um dos níveis de um fator experimental é atribuído a um e apenas um grupo de um agrupamento de unidades, os efeitos desse fator resultam completamente confundidos com os efeitos atribuíveis ao agrupamento de unidades. Semelhantemente, se cada combinação dos níveis de dois ou mais fatores experimentais é atribuída a um e somente um diferente grupo de um agrupamento de unidades, os efeitos principais e interações desses fatores experimentais ficam completamente confundidos com os efeitos do agrupamento. No delineamento completamente casualizado, não há confundimento; entretanto, nos delineamentos em blocos com um fator experimental em blocos, o efeito desse fator fica confundido com o efeito de bloco.

Os efeitos de agrupamentos de unidades contribuem com componentes de variância para os *EQM* que podem influir nas relações entre os *EQM*. Nessas circunstâncias, a desconsideração desses efeitos pode conduzir a expressões incorretas de estimadores de componentes de variância, de razões *F* apropriadas para testes de significância e de variâncias de contrastes de médias.

### MODELO ESTATÍSTICO COMPLETO

A relação entre os níveis e combinações dos níveis dos fatores experimentais constitui a estrutura dos fatores experimentais. A atribuição desses níveis e combinações de níveis às unidades de observação pode ser sujeita a restrições. A cada uma dessas restrições corresponde um agrupamento ou estratificação dessas unidades. Por conveniência conceitual, considera-se, também, o agrupamento das unidades de observação para constituir o mate-

rial experimental, globalmente. Os grupos de cada um desses agrupamentos são os níveis de um fator de unidade, que constituem um estrato do material experimental. A relação entre os níveis e as combinações dos níveis dos fatores de unidade constitui a estrutura dos fatores de unidade.

A estrutura dos fatores experimentais, a estrutura dos fatores de unidade e a relação entre essas duas estruturas, determinada pela casualização, constitui a estrutura do experimento ou o delineamento experimental.

Os níveis de um fator de unidade são as unidades experimentais para os níveis do fator experimental, ou as combinações dos níveis dos fatores experimentais, com os quais guardam relação de correspondência na estrutura do experimento. A variação entre as unidades experimentais para os níveis de um fator experimental, ou para as combinações dos níveis de dois ou mais fatores experimentais, constitui o erro experimental para esses níveis, ou combinações de níveis. Assim, a estrutura dos fatores de unidade estabelece uma estruturação do erro experimental.

De modo geral, postula-se que o modelo estatístico completo para um experimento representa a resposta observada na unidade de observação  $u$  com a condição experimental (combinação dos níveis dos fatores experimentais)  $c$  pela equação:

$$Y(c,u) = c_c + u_u,$$

onde  $c_c$  é o efeito da condição experimental  $c$  e  $u_u$  é o efeito da unidade  $u$ .

Cada um dos dois termos  $c_c$  e  $u_u$  compreende um conjunto de efeitos relacionados com fatores experimentais e com fatores de unidade, respectivamente.

Essa forma geral do modelo estatístico pressupõe a aditividade dos efeitos atribuíveis aos fatores experimentais e aos fatores de unidade. Essa pressuposição é estendida para as correspondentes composições desses dois efeitos.

A especificação do modelo estatístico completo, com as decomposições dos efeitos atribuíveis aos fatores experimentais e aos fatores de unidade, é derivada a partir das correspondentes estruturas desses dois conjuntos de fatores.

Procedimentos práticos para a derivação da estrutura do experimento caracterizada na forma usual e para a especificação da equação do correspondente modelo estatístico para experimentos com delineamento completo balanceado, são apresentados por Lee (1975), Silva (1977) e Silva & Zonta (1991).

Esses procedimentos práticos podem ser utilizados para as determinações separadas da estrutura dos fatores experimentais e da estrutura dos fatores de unidade, e para as especificações dos correspondentes modelos estatísticos.

Para ilustração, considerem-se novamente os três distintos delineamentos correspondentes à classificação dupla balanceada, com dois fatores experimentais  $A$  e  $B$ . O componente referente à condição experimental, comum para os três delineamentos, tem a seguinte composição:

$$c_c = m + a_a + b_b + ab_{ab}.$$

Para determinar o componente que exprime o efeito de unidade, observe-se que as estruturas de unidade correspondentes aos três delineamentos são as

seguintes: completamente casualizado: um fator de unidade - parcela; blocos casualizados generalizados: dois fatores de unidade - bloco, confundido com o fator  $B$ , e parcela; blocos casualizados com subamostragem: três fatores de unidade-bloco, confundido com o fator  $B$ , parcela, confundido com as combinações dos níveis dos fatores  $A$  e  $B$ , e observação.

Observe-se que nesse exemplo, como em muitas situações de controle local, pode não haver um fator experimental associado ao fator de unidade "bloco"; ou seja, pode não haver qualquer interesse relacionado com a variação entre blocos, que nesse caso constitui apenas variação atribuível ao erro experimental. Entretanto, para maior generalidade do procedimento indicado a seguir, admitir-se-á que há um fator experimental associado a cada fator de unidade. Da mesma forma, é sugerido considerar todos os efeitos principais e interações atribuíveis aos fatores experimentais. Após a derivação do modelo estatístico e, mais convenientemente, após o conhecimento dos *EQM*, o pesquisador poderá eliminar do modelo os termos que representem fatores experimentais inexistentes e efeitos de unidade que ele julgue desnecessários, e agregar com os correspondentes erros experimentais interações consideradas irrelevantes. Esse procedimento é conveniente para evitar a desconsideração de efeitos cujas implicações somente sejam percebidas através da apreciação do modelo e das composições dos *EQM*.

Então, os efeitos de unidade, para os três delineamentos em consideração, são, respectivamente, os seguintes:

$$u_u = e_{(abe)},$$

$$u_u = u_b + e_{ae(b)},$$

$$u_u = u'_b + u_{a(b)} + e_{e(ab)},$$

onde o efeito de unidade elementar é denotado por  $e$ , e os demais efeitos de unidade, por  $u$ , com apóstrofos para distinguir distintos efeitos de unidade. A estrutura dos fatores de unidade é especificada pelos índices: índices fora de parênteses correspondem a fatores aninhados; índices entre parênteses, a fatores ninhos.

Logo, as equações dos modelos estatísticos completos para os três delineamentos em consideração são, respectivamente, as seguintes:

$$Y_{abe} = m + a_a + b_b + ab_{ab} + e_{(abe)},$$

$$Y_{abe} = m + a_a + [b_b + u_b] + ab_{ab} + e_{ae(b)},$$

$$Y_{abe} = m + a_a + [b_b + u'_b] + [ab_{ab} + u_{a(b)}] + e_{e(ab)}.$$

De modo geral, efeitos com o mesmo conjunto de letras nos índices estão completamente confundidos. Dessa forma, não podem ser estimados separadamente, e constituem, de fato, uma mesma fonte de variação. Para manter a relação biunívoca entre os componentes da equação do modelo estatístico, excluído  $m$ , e as fontes de variação independentes, os efeitos confundidos são colocados dentro de colchetes, de modo que os termos dentro de um mesmo par de colchetes correspondem a uma única fonte de variação.

Para ilustração mais ampla, suponha-se um experimento agrícola de campo com um fator experimental cultivar, com delineamento blocos casualizados,



conduzido em diversos ambientes (locais ou anos). Trata-se de um experimento com três fatores  $A$ ,  $B$  e  $D$ , onde  $A$  é o fator cultivar, suposto fixo, e  $B$  e  $D$  correspondem a ambiente e bloco, respectivamente, ambos considerados aleatórios.

A equação do modelo estatístico usual postulado para esse experimento, com  $A$  níveis do fator  $A$  em cada um de  $D$  blocos ( $D$ ), conduzido em  $B$  ambientes ( $B$ ), é:

$$Y_{abd} = m + a_a + b_b + d_{d(b)} + ab_{ab} + ad_{ad(b)}, \quad a=1,2,\dots,A, \quad b=1,2,\dots,B, \quad d=1,2,\dots,D,$$

onde  $m$  é a média geral esperada,  $a_a$ ,  $b_b$  e  $d_{d(b)}$  são os efeitos principais de cultivar e de ambiente, e o efeito de bloco dentro de ambiente, respectivamente, e  $ab_{ab}$  e  $ad_{ad(b)}$  são as interações cultivar x ambiente e cultivar x bloco dentro de ambiente.

Sob as pressuposições usuais, os valores esperados dos quadrados médios são os apresentados na Tabela 2.

Neste experimento, as unidades elementares são as unidades experimentais para o fator  $A$ , essas unidades estão fisicamente aninhadas nos blocos, e estes, nos ambientes. Assim, identificam-se três fatores de unidade: parcela, bloco e ambiente, cujos correspondentes efeitos são denotados, respectivamente, por  $e_{a(bd)}$ ,  $u_{d(b)}$  e  $u'_b$ . Então, o efeito de unidade tem a seguinte composição:

$$u_u = u'_b + u_{d(b)} + e_{a(bd)}.$$

A equação do modelo estatístico completo para a observação na parcela com a cultivar  $a$  no bloco  $d$  do ambiente  $b$  é, portanto,

$$Y_{abd} = m + a_a + [b_b + u'_b] + [d_{d(b)} + u_{d(b)}] + ab_{ab} + [ad_{ad(b)} + e_{a(bd)}].$$

**TABELA 2. EQM para experimento com  $A$  níveis de um fator  $A$ , em  $D$  blocos casualizados (bloco fator  $D$ ), conduzido em  $B$  ambientes (ambiente fator  $B$ ), ignorando a distinção entre fatores experimentais e fatores de unidade.**

Fonte de variação / efeito		EQM
$A:$	$a_a$	$\sigma_E^2 + \sigma_{AD}^2 + D\sigma_{AB}^2 + BD\sigma_A^2$
$B:$	$b_b$	$\sigma_E^2 + A\sigma_D^2 + AD\sigma_B^2$
$D(B)$	$d_{d(b)}$	$\sigma_E^2 + A\sigma_D^2$
$AB$	$ab_{ab}$	$\sigma_E^2 + \sigma_{AD}^2 + D\sigma_{AB}^2$
$AD(B)$	$ad_{ad(b)}$	$\sigma_E^2 + \sigma_{AD}^2$

### CONSIDERAÇÃO DOS EFEITOS DE UNIDADE NOS EQM

O procedimento descrito nesta Seção permite a derivação dos EQM corretos através da inclusão dos componentes de variância correspondentes aos efeitos de fatores de unidade nos EQM derivados pelo método usual. Ele se baseia na propriedade de que a variação atribuível a um fator de unidade se manifesta em todas as fontes de variação que recebem contribuição da variação entre os níveis desse fator. Não se manifesta nas fontes de variação internas ao correspondente agrupamento de unidades elementares.

Para ilustração, considere-se o experimento agrícola de campo conduzido em diversos ambientes. Denotem-se os componentes de variância correspondentes aos três efeitos de unidade - parcela dentro de bloco,  $e_{a(bd)}$ , bloco dentro de ambiente,  $u_{d(b)}$ , e ambiente  $u'_b$ , respectivamente por  $\sigma_E^2$ ,  $\sigma_U^2$  e  $\sigma_{U'}^2$ .

O primeiro componente de variância está presente em todas as fontes de variação, o segundo, apenas nas fontes de variação não internas a bloco, e o terceiro, apenas na fonte de variação referente a ambiente.

Essas circunstâncias são expressas pela composição dos índices dos correspondentes efeitos. O efeito de unidade "bloco", ou seja,  $u_{d(b)}$ , contém apenas os índices d e b; logo, o correspondente componente de variância  $\sigma_U^2$  deve compor apenas as fontes de variação referentes a efeitos cujos índices incluam somente b e d, ou seja,  $b_b$  e  $d_{d(b)}$ . O efeito de unidade "ambiente",  $u'_b$ , contém apenas o índice b; logo, o componente de variância  $\sigma_{U'}^2$  deve aparecer apenas na fonte de variação  $b_b$ .

Assim, pode-se derivar os EQM que consideram os efeitos de unidade através da inclusão dos correspondentes componentes de variância nos EQM derivados pelos métodos usuais, comparando a composição dos índices dos efeitos de unidade com as composições dos índices dos efeitos que constituem fontes de variação ignorando efeitos de unidade.

Essa propriedade pode ser generalizada para a seguinte regra: Para a consideração de cada fator de unidade presente no delineamento do experimento nos EQM, o correspondente componente de variância deve ser adicionado aos EQM obtidos pelo processo prático usual, através dos seguintes passos: Passo 1. Para cada fator de unidade  $U$ , adiciona-se o componente de variância  $\sigma_U^2$  para cada fonte de variação cuja composição de letras do correspondente efeito seja um subconjunto da composição de letras do efeito de unidade  $u$ , ignorando parênteses. Passo 2. Se há  $E$  observações por nível do fator de unidade  $U$ , multiplica-se  $\sigma_U^2$  por  $E$ , de modo que o componente de variância resultante é  $E\sigma_U^2$ . Esses dois passos completam o algoritmo para a situação mais comum de fator de unidade aleatório com número infinito de níveis na população. Para fator de unidade fixo ou aleatório com número finito de níveis na população, deve ser acrescentado o seguinte: Passo 3. Para cada fonte de variação que aninha o fator de unidade  $U$  inclui-se no coeficiente de  $\sigma_U^2$  o fator de correção para população finita  $1 - \frac{U}{U}$ , onde  $U$  e  $U$  são, respectivamente, os números de níveis de  $U$  na amostra e na população.

Os EQM completos para o experimento agrícola em consideração são apresentados na Tabela 3. Esses EQM revelam que a fonte de variação  $AD(B)$ , ou seja, o erro combinado dos B ambientes, é o erro apropriado para inferências referentes à interação  $AxB$  e esta interação é o erro para inferências referentes ao efeito principal do fator  $A$ . Esses resultados concordam com as indicações de textos. Entretanto, não há um erro apropriado para inferências referentes a ambiente, exceto se o componente da variância entre ambientes atribuível ao erro experimental,  $\sigma_{U'}^2$ , for nulo; neste caso,  $D(B)$ , ou seja, bloco dentro de ambiente é o erro apropriado. Os textos divergem nas indicações referentes a estas inferências, que podem ser importantes em algumas situações. Kempthorne (1952) e Cochran & Cox (1957) não fazem qualquer indicação para tais inferências. Outros autores indicam como erro apropriado para essas inferências: bloco dentro de local (Gomez & Gomez, 1984), ou a interação tratamento x ambiente (Pimentel-Gomes, 1990), ou uma combinação de fontes de variação (Federer, 1955), ou o erro combinado.

Os EQM para os três delineamentos que correspondem à classificação dupla balanceada são apresentados na Tabela 4. Pode-se verificar que o delineamento blocos casualizados generalizados não propicia inferências válidas referentes ao fator experimental  $B$ , alocado a blocos, exceto se o componente do erro experimental  $\sigma_U^2$ , confundido com a variação atribuível ao fator  $B$ , for nulo (neste caso, o erro apropriado para tais inferências é a interação  $AxB$  ou o erro  $E$ , respectivamente se o fator  $A$  for aleatório ou fixo). O delineamento blocos casualizados com subamostragem não permite inferências válidas para o fator  $A$  quando o fator  $B$  é fixo, exceto se o componente de variância relativo ao efeito de parcela  $\sigma_U^2$  for nulo (neste caso, o erro apropriado é  $e_{e(ab)}$ ). Este delineamento também não permite inferências válidas relativas ao fator  $B$  quando o fator  $A$  é aleatório, exceto

**TABELA 3. EQM para experimento com A níveis de um fator A, em D blocos casualizados (fator D), conduzido em B ambientes (fator B), distinguindo efeitos de fatores experimentais e de fatores de unidade.**

Fonte de variação / efeito		EQM
A:	$a_a$	$\sigma_E^2 + \sigma_{AD}^2 + D\sigma_{AB}^2 + BD\sigma_A^2$
B:	$b_b$	$\sigma_E^2 + A[\sigma_U^2 + \sigma_D^2] + AD[\sigma_B^2 + \sigma_{U'}^2]$
$D(B)$ :	$d_{d(b)}$	$\sigma_E^2 + A[\sigma_U^2 + \sigma_D^2]$
$AB$	$ab_{ab}$	$\sigma_E^2 + \sigma_{AD}^2 + D\sigma_{AB}^2$
$AD(B)$ :	$ad_{ad(b)}$	$\sigma_E^2 + \sigma_{AD}^2$

se  $\sigma_{U'}^2=0$  (neste caso o erro apropriado é  $ab_{ab}$ ), e quando o fator  $A$  é fixo, exceto se ambos  $\sigma_U^2=0$  e  $\sigma_{U'}^2=0$  (quando o erro apropriado é  $e_{(abe)}$ ). Também não permite inferências referentes à interação  $AxB$ , exceto se o componente de variância relativo ao efeito de parcela  $\sigma_U^2$  for nulo.

### DISCUSSÃO

A distinção dos significados dos fatores experimentais e dos fatores de unidade confundidos em um experimento é fundamental para a compreensão da implicação da consideração destes últimos para as inferências referentes aos fatores experimentais. Essa distinção pode ser óbvia quando o fator experimental é um fator de tratamento. Esse é o caso do fator experimental “temperatura” e do correspondente fator de unidade “ambiente” do experimento para pesquisa dos efeitos do tempo e da temperatura de armazenamento sobre a qualidade da semente de trigo. O fator de unidade “ambiente”, compreende todas as características permanentes do ambiente (referentes ao local de armazenamento da semente), ou seja: umidade, ventilação, microorganismos, etc., excluída a temperatura.

A distinção pode não ser tão nítida quando o fator experimental é um fator intrínseco. Por exemplo, a distinção entre o fator experimental “ambiente” e o correspondente fator de unidade no experimento agrícola de campo não é óbvia. Para lograr a distinção, será necessário definir o significado do fator experimental “ambiente”. Esse fator, em geral, compreende as características permanentes do ambiente, correspondentes aos aspectos gerais do clima e do tipo de solo inerentes ao ambiente; as características do ambiente que flutuam em uma maneira sistemática, como o comprimento do dia; as caracte-

**TABELA 4. EQM para as três diferentes estruturas de experimento que correspondem à classificação dupla balanceada com dois fatores experimentais  $A$  e  $B$ , distinguindo os efeitos de fatores experimentais e de fatores de unidade.**

Fonte de variação / efeito	Componentes de variância <sup>1</sup>		
	Fatores experimentais		Fatores de unidade
	(1)	(2)	(3)
$A:$ $a_a$	$\sigma_E^2 + (1 - \frac{B}{B})E\sigma_{AB}^2 + BE\sigma_A^2$		$E\sigma_U^2$
$B:$ $a_b$	$\sigma_E^2 + (1 - \frac{A}{A})E\sigma_{AB}^2 + AE\sigma_B^2$	$AE\sigma_U^2$	$E\sigma_U^2 + \sigma_{U'}^2$
$AB:$ $ab_{ab}$	$\sigma_E^2 + E\sigma_{AB}^2$		$E\sigma_U^2$
$E:$ $e_{(abe)}$	$\sigma_E^2$		

<sup>1</sup> Delineamento completamente casualizado: EQM: (1); delineamento blocos casualizados generalizados com fator  $B$  em blocos: EQM: (1)+(2); delineamento blocos casualizados com subamostragem, com fator  $B$  em blocos: EQM: (1)+(3).

terísticas do ambiente que flutuam por condições intrínsecas aos ambientes em consideração, tais como as distribuições de chuva e temperatura e incidências de pragas e doenças; e as características do ambiente que podem ser determinadas pelo homem, tais como datas de plantio e de colheita, densidade de semeadura, e outras práticas de cultivo, quando inerentes e específicas para os ambientes em consideração. Por exclusão, o correspondente fator de unidade compreende as características ambientais não propriamente atribuíveis aos ambientes em consideração, que, idealmente, deveriam se manter constantes, mas que variam, circunstancialmente, ou acidentalmente, como pode ocorrer em um mesmo ambiente, tais como: qualidade da semente, ocorrências de clima e solo, incidências de pragas e doenças, e práticas de cultivo.

Em alguns experimentos em blocos casualizados, o fator experimental que corresponde a blocos é alguma característica específica de interesse, como, por exemplo, localização topográfica na área de um experimento agrícola de campo e posição no tempo de um experimento conduzido em etapas. Nessas situações, a caracterização da distinção entre o fator experimental e o fator de unidade “bloco” também poderá não ser óbvia.

Em muitos experimentos com delineamento experimental em blocos, não há um fator experimental associado a blocos, ou seja, blocos não correspondem a níveis de uma característica de interesse particular. Mesmo assim, o pesquisador pode ter interesse em verificar se o controle local logrado pelo agrupamento das unidades experimentais foi efetivo para a redução da estimativa do erro experimental para inferências de interesse. Nessas circunstâncias, ele poderá testar a hipótese de nulidade da soma dos componentes de variância  $\sigma_B^2$  e  $\sigma_U^2$ , ou seja:  $H_0: \sigma_B^2 + \sigma_U^2 = 0$ .

A inspeção dos *EQM* com os componentes de variância referentes aos efeitos de unidade, ainda na fase de planejamento do experimento, pode permitir a verificação da adequabilidade do plano do experimento para as inferências que o experimento tem como objetivo derivar. Ela também pode sugerir possíveis alterações apropriadas, se o plano elaborado não contemplar a derivação de inferências válidas (não tendenciosas) importantes. Assim, por exemplo, a inspeção dos *EQM* da Tabela 4 corrobora o fato de que o experimento para pesquisa dos efeitos do tempo e da temperatura de armazenamento sobre a qualidade fisiológica da semente, com o delineamento descrito em ilustração anterior (blocos casualizados generalizados) não permitiria a derivação de inferências referentes à temperatura. Poder-se-ia verificar, através da inspeção dos *EQM*, que tais inferências se tornariam viáveis com a duplicação dos ambientes e a atribuição aleatória de cada uma das *B* temperaturas a dois ambientes.

Semelhante solução, entretanto, não viabilizaria inferências referentes a ambiente (local ou ano) no experimento agrícola de campo, pela impossibilidade da repetição de ambiente.

## CONCLUSÕES

1. O modelo estatístico usualmente formulado para experimentos não leva adequadamente em conta a estrutura do experimento.

2. Os métodos práticos usuais para derivar valores esperados de quadrados médios, baseados em modelo estatístico que ignora os efeitos de unidade, podem conduzir a expressões incompletas que induzam inferências incorretas referentes à estimação de componentes de variância e a testes de significância de efeitos principais e interações de fatores, e de contrastes de médias.

3. A distinção das duas categorias de fatores experimentais, ou seja, fatores de tratamento e fatores intrínsecos, é importante para a validade das inferências derivadas do experimento.

4. A formulação do modelo estatístico completo deve considerar os efeitos de unidade, através da inclusão de um termo para representar cada agrupamento de unidades resultante das restrições à casualização.

5. Os correspondentes valores esperados de quadrado médios corretos podem ser obtidos através dos procedimentos práticos usuais e da ulterior inclusão dos componentes de variância correspondentes aos efeitos de unidade.

6. O modelo estatístico completo, reconhecendo os fatores de unidade, e os correspondentes valores esperados de quadrados médios devem ser derivados antes da condução do experimento, de modo que possíveis inadequações do delineamento experimental possam ser corrigidas em tempo.

#### REFERÊNCIAS

- ADDELMAN, S. Variability of treatments and experimental units in the design and analysis of experiments. **Journal of the American Statistical Association**, v.65, n.331, p.1095-1108, 1970.
- BELTRÃO, K.I.; PINHEIRO, J.C.R.C. **Planejamento de experimentos com um enfoque de análise de dados estruturados**. Lavras: ESAL, 1989. 44p. Trabalho apresentado no Curso Análise de Dados Estruturados, realizado durante o 3º Simpósio de Estatística Aplicada à Experimentação Agronômica e a 34ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria, em Lavras, 1989.
- BENNETT, C.A.; FRANKLIN, N.A. **Statistical analysis in chemistry and the chemical industry**. New York: John Wiley, 1954. 724p.
- BROWNLEE, K.A. **Statistical theory and methodology in science and engineering**. 2.ed. New York: John Wiley, 1965. 590p.
- COCHRAN, W.G.; COX, G.M. **Experimental designs**. 2.ed. New York: John Wiley, 1957. 617p.
- CORNFIELD, J.; TUKEY, J.W. Average values of mean squares in factorials. **The Annals of Mathematical Statistics**, v.27, n.4, p.907-949, 1956.
- COX, D.R. **Planning of experiments**. New York: John Wiley & Sons, 1958. 308p.
- COX, D.R.; SNELL, E.J. **Applied statistics, principles and examples**. London: Chapman and Hall, 1981. 189p.
- FEDERER, W.T. **Experimental design: theory and application**. New York: The Macmillan Company, 1955. 544p.

- FISHER, R.A. **The design of experiments**. Edinburgh: Oliver and Boyd, 1935. 248p.
- GILL, J.L. **Design and analysis of experiments in the animal and medical sciences**. Ames: The Iowa State Univ. Press, 1978. v.1, 410p.
- GOMEZ, K.A.; GOMEZ, A.A. **Statistical procedures for agricultural research**. 2.ed. New York: John Wiley, 1984. 680p.
- HICKS, C.R. **Fundamental concepts in the design of experiments**. 2.ed. New York: Holt, Rinehart and Winston, 1973. 349p.
- HOUTMAN, A.M. **The analysis of designed experiments**. Princeton: Princeton Univ., 1980. Ph.D. Thesis.
- JOHNSON, N.L.; LEONE, F.C. **Statistics and experimental design: in engineering and the physical sciences**. New York: John Wiley, 1964. v.2, 399p.
- KEMPTHORNE, O. **The design and analysis of experiments**. New York: John Wiley, 1952. 631p.
- LEE, W. **Experimental design and analysis**. San Francisco: W.H. Freeman, 1975. 353p.
- LI, C.C. **Introduction to experimental statistics**. New York: McGraw-Hill, 1964. 460p.
- NELDER, J.A. The analysis of randomized experiments with orthogonal block structure. I. Block structure and the null analysis of variance. **Proceedings of the Royal Society, Series A**, London, v.273, p.147-162, 1965a.
- NELDER, J.A. The analysis of randomized experiments with orthogonal block structure. II. Treatment structure and the general analysis of variance. **Proceedings of the Royal Society, Series A**, London, v.273, p.163-178, 1965b.
- PIMENTEL-GOMES, F. **Curso de Estatística Experimental**. 13.ed. São Paulo: Nobel, 1990. 467p.
- SCHULTZ JUNIOR, E.F. Rules of thumb for determining expectations of mean squares in analysis of variance. **Biometrics**, v.11, n.2, p.123-135, 1955.
- SILVA, J.G.C. da. **A estrutura do delineamento e seu uso na especificação do esquema da análise de variação**. Embrapa-DMQ, Brasília, 1977. 20p.
- SILVA, J.G.C. da; ZONTA, E.P. **Conceitos e algoritmos úteis em estatística experimental**. [S.l.: s.n.], 1991. 211p. Trabalho apresentado em Mini-curso realizado durante o IV Simpósio de Estatística Aplicada à Experimentação Agronômica, em Goiânia, GO, 1991.
- SNEDECOR, G.W.; COCHRAN, W.G. **Statistical methods**. 6.ed. Ames: The Iowa State Univ. Press, 1967. 593p.
- WILK, M.B.; KEMPTHORNE, O. **Analysis of variance: Preliminary tests, pooling, and linear models - Derived linear models and their use in the analysis of randomized experiments**. Ohio: Wright Air Development Center/Wright-Patterson Air Force Base, 1956. v.2, 126p. (Technical Report, 55-244).