DETERMINAÇÃO DO TAMANHO ÓTIMO DE PARCELAS EM ENSAIOS AGRÍCOLAS[1]

ADROALDO GUIMARÃES ROSSETTI[2] e FREDERICO PIMENTEL GOMES[3]

RESUMO - Este estudo visa a fornecer um embasamento teórico a trabalhos de tamanho de parcelas experimentais. A estimativa do tamanho ótimo de parcela é enfocada sob dois aspectos: em função do custo mínimo do experimento, em unidades básicas, e em função do número de repetições necessárias para se obter determinada diferença de médias, fixada "a priori", a um nível α de probabilidade. Estimou-se o coeficiente b, ou índice de variabilidade do solo partindo-se de um modelo de regressão em que os dados são correlacionados e têm variâncias desiguais, minimizando-se a soma dos quadrados dos erros através do método dos quadrados mínimos generalizados. Desenvolveu-se uma metodologia para estimar os pesos ω^{ij} a partir dos próprios dados. Tal procedimento é válido tanto para dados de ensaios de uniformidade como para dados de outros ensaios. Estudou-se o primeiro caso através de um esquema de classificação hierárquica e o segundo através de um modelo de blocos ao acaso com parcelas subdivididas.

Termos para indexação: tamanho de parcela, observações correlacionadas, estatística experimental, variâncias.

DETERMINATION OF THE OPTIMUM PLOT SIZE IN AGRICULTURAL ESSAYS

ABSTRACT - This study aims to provide a theorical basis for research on experimental plot size. The optimum plot size estimate is focused on two aspects: In terms of the minimum cost of the experiment, in basic units and in terms of the number of replications necessary to obtain a predetermined difference of mean, fixed "a priori", at the α level of probability. The b coefficient or soil heterogeneity coefficient was estimated using a regression model in which the data are correlated and have unequal variances and minimizing the sum of squares of the errors through the generalized least square method. A methodology was developed to estimate the ω^{ij} weight to the appropriate data. Such a procedure is valid for uniformity trial data as well as for other trial experimental data. The first case was studied through a hierarchical classification scheme and the second through of a randomized block design, with split-plot model.

Index terms: plot size, correlationed data, experimental statistic, variances.

INTRODUÇÃO

Para minimizar a variabilidade dos resultados de um experimento, o pesquisador deve atentar para fatos importantes, tais como: forma e colocação das parcelas no campo, forma do bloco, tamanho das parcelas, número de repetições, delineamento experimental, falhas de plantas nas parcelas, efeito de bordaduras entre parcelas e forma de condução do experimento (Calzada Benza 1966).

Muitos experimentadores têm usado tamanhos e formas de parcelas inteiramente arbitrários. Nas parcelas pequenas, a forma tem pouca ou quase nenhuma influência sobre o erro experimental, enquanto que em parcelas grandes a influência é notável (Torrie et al. 1963).

A literatura sobre metodologia do tamanho de parcelas em experimentos agrícolas é extensa, sendo os estudos, na grande maioria, conduzidos através de ensaios de uniformidade. Várias tentativas visando a criar métodos estatísticos para determinar o tamanho ideal de parcelas em ensaios agrícolas têm sido desenvolvidas utilizando várias culturas e procedimentos, tais como: Método do Desvio Padrão, Método da Máxima Curvatura, Coeficiente de Variação e Método da Informação Relativa, entre outros (Smith 1938). Há consideráveis diferenças que variam com o procedimento utilizado, com o tipo de solo, com a cultura em estudo, embora a conclusão mais freqüente evidencie que

1 Aceito para publicação em 3 de maio de 1983.
Extrato da tese apresentada à Esc. Sup. de Agric. "Luiz de Queiroz", Univ. de São Paulo, para obtenção do título de Mestre em Estatística e Experimentação Agronômica.

2 Matemático, M.Sc., em Estatística e Experimentação Agronômica, EMBRAPA/Centro Nacional de Pesquisa de Seringueira e Dendê (CNPSD), Caixa Postal 319, CEP 69000 - Manaus, AM.

3 Engº - Agrº, Doutor em Agronomia e Prof. Catedrático do Depart. de Matemática e Estatística da ESALQ/USP, Caixa Postal 09, CEP 13400 - Piracicaba, SP.

o coeficiente de variação decresce à medida que aumenta o tamanho da parcela.

O método que se revela mais eficiente é a Lei da Variância (Smith 1938), que estabelece uma relação linear entre o logarítmo da variância de parcelas de um mesmo tamanho com o logarítmo do tamanho da parcela.

Um aspecto bastante curioso é que os trabalhos acerca de metodologias do tamanho de parcelas nem sempre consideram a correlação que deve existir entre as observações, fato que ocorre com grande freqüência quando suas variâncias são estimadas a partir dos mesmos dados. Por ser este fato, de grande importância nas pesquisas experimentais, o presente trabalho visa a um estudo teórico da metodologia do tamanho de parcelas partindo de um modelo onde as observações são correlacionadas e de variâncias desiguais, através do qual o tamanho ótimo da parcela será estimado levando-se em conta a relação entre o tamanho da parcela e a variabilidade do solo, com base ainda no custo mínimo na condução do experimento.

O procedimento apresenta um método de ponderação das variâncias observadas de parcelas de diferentes tamanhos que conduzem a um estimador imparcial de β, com variância mínima assintótica (Róger Arroyo & Chaves 1966). Tal procedimento é aplicável tanto a dados de ensaio de uniformidade como a dados de outros ensaios (Koch & Rigney 1951).

METODOLOGIA

Em aplicações usuais da análise de regressão, admite-se, em geral, que os valores da variável dependente não são correlacionados, que suas variâncias são iguais e que os valores da variável independente são determinados sem erro. Neste caso, a aplicação simples e direta do método dos quadrados mínimos, para estimar os coeficientes de regressão, é eficiente, e a quantidade a ser minimizada é a soma dos quadrados dos desvios.

Um caso de ocorrência menos freqüente é aquele no qual os valores da variável dependente Y, são correlacionados e têm variâncias desiguais. Embora o método simples e direto dos quadrados mínimos ainda dê estimadores imparciais para a e b, este método é, em geral, ineficiente, pois as estimativas $\hat{a}$ e $\hat{b}$ não são tão exatas como quando obtidas pelo método agora exposto. Mais propícia é a aplicação do método dos quadrados mínimos generalizados, pois, se as variâncias e covariâncias dos valores de Y são conhecidas, podem ser determinadas estimativas eficientes dos parâmetros; e mesmo se as variâncias e covariâncias forem apenas estimadas dos dados, seu uso permite que sejam determinadas estimativas de mínima variância.

Na estimação do tamanho ótimo de parcelas, o coeficiente de regressão b, do método de Smith (1938), tem particular importância, pois representa um índice de heterogeneidade do solo; é uma medida de correlação entre unidades adjacentes, portanto de grande interesse.

ESTIMAÇÃO DE b BASEADA EM DADOS DE ENSAIO DE UNIFORMIDADE

Ao avaliar um delineamento em blocos ao acaso com parcelas subdivididas ou em reticulado quadrado, de dados de ensaios de uniformidade, visando a determinar o tamanho ótimo das parcelas, não é necessário indicar o componente de tratamentos no modelo matemático. Por exemplo, para um reticulado quadrado k x k, a área toda é dividida em tantas repetições quantas possíveis, cada uma com k blocos de k parcelas, e pode ser analisado segundo um esquema de classificação hierárquica (Robinson et al. 1948), cujo modelo é:

$$Y_{ijkl} = m + r_i + b_{ij} + p_{ijk} + s_{ijkl}, \qquad (1)$$

onde:
m é a média geral, r_i é a i-ésima "repetição"; i = 1,2, . . . , I; b_{ij} é o j-ésimo bloco dentro da "repetição" i; j = 1, 2, . . . , J; p_{ijk} é a k-ésima parcela dentro do bloco ij; k = 1,2, . . . , k; s_{ijkl} é a l-ésima subparcela dentro da parcela ijk; l = 1, 2, . . . , L.

O que chamamos aí repetição é o conjunto de cada repetição ortogonal, colocadas em grupo compacto e formando um grande bloco que encerre todos os tratamentos.

Aplicando-se o método dos quadrados mínimos à equação (1), obtêm-se os componentes de variância da Tabela 1.

Formação das parcelas

Seja V'_1 a variância de parcelas de diversos tamanhos, reduzidas a uma parcela base. Neste caso, V_1 coincide com o quadrado médio de "repetições", como aparece na Tabela 1.

Considerem-se agora os blocos como parcelas menores, dentro das "repetições". Vê-se que:

$$E\ (SQB) = I\ (J-1)\ V_2 + (I-1)\ V_1.$$

Mas há, nas I "repetições", IJ blocos, portanto IJ-1 graus de liberdade, para estimar a respectiva variância. Logo, a variância entre blocos será:

$$V'_2 = \left[I\ (J-1)\ V_2 + (I-1)\ V_1 \right] / (IJ-1). \qquad (2)$$

No que se refere à soma de quadrados de parcelas tem-se:

TABELA 1. Análise de variância.

Fonte de variação	G.L.	QM	E(QM)
Repetições	I -1	V_1	$\sigma^2 + L\sigma_P^2 + KL\sigma_B^2 + JKL\sigma_R^2$
Blocos d. repetições	I(J-1)	V_2	$\sigma^2 + L\sigma_P^2 + KL\sigma_B^2$
Parcelas d. blocos	IJ(K-1)	V_3	$\sigma^2 + L\sigma_P^2$
Subparcelas d. parcelas	IJK(L-1)	V_4	σ^2

$$E\,(SQP) = IJ\,(K\text{-}1)\,V_3 + I\,(J\text{-}1)\,V_2 + (I\text{-}1)\,V_1.$$

Como há IJK parcelas, portanto há IJK-1 graus de liberdade. Assim, a variância entre parcelas será:

$$V_3' = \left[IJ\,(K\text{-}1)\,V_3 + I\,(J\text{-}1)\,V_2 + (I\text{-}1)\,V_1\right] / (IJK\text{-}1). \quad (3)$$

Analogamente, as subparcelas dão

$$E\,(SQS) = IJK\,(L\text{-}1)\,V_4 + IJ\,(K\text{-}1)\,V_3 + I\,(J\text{-}1)\,V_2 + (I\text{-}1)\,V_1.$$

Assim, a variância entre subparcelas fica:

$$V_4' = \left[IJK\,(L\text{-}1)\,V_4 + IJ\,(K\text{-}1)\,V_3 + I\,(J\text{-}1)\,V_2 + (I\text{-}1)\,V_1\right] / (IJKL\text{-}1). \quad (4)$$

Determinação do coeficiente b

Smith (1938) definiu, através da "Lei da Variância", o coeficiente de regressão b, pela relação

$$V\,(x) = V_1 / x^b, \quad (5)$$

onde:

x é o tamanho da parcela em unidades básicas, V_1 é a variância entre parcelas de tamanho correspondente a uma unidade, e V(x) é a variância do rendimento médio por unidade de área, para parcelas que têm x unidades de tamanho.

Assim, se V' representa a variância de parcelas de vários tamanhos, V(x) será calculado através da divisão de V' pelo número de unidades por parcela ("repetição", bloco, parcela e subparcela), ficando assim relacionada a uma unidade base, como se segue:

$$V(x) = V' / x.$$

Quando os valores da equação (5) são transformados para a escala logarítmica, resulta a forma linear:

$$\log V\,(x) = \log V_1 - b \log x, \quad (6)$$

que pode ainda ser expressa sob a forma

$$Y_i = a + B\,X_i + e_i, \quad (7)$$

onde os valores de Y_i podem ser correlacionados e podem ter variâncias desiguais. Admitiremos e_i com distribuição normal.

Para os n valores observados, a forma matricial será:

$$Y = X\beta + \epsilon \quad (8)$$

onde:

$$E\,(Y) = X\beta$$

e

$$V(Y) = \begin{bmatrix} V(Y_1) & Cov\,(Y_1Y_2) & \ldots\ldots & Cov\,(Y_1Y_n) \\ Cov\,(Y_1Y_2) & V\,(Y_2) & \ldots\ldots & Cov\,(Y_2Y_n) \\ \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots \\ Cov\,(Y_1Y_n) & Cov\,(Y_2Y_n) & \ldots\ldots & V\,(Y_n) \end{bmatrix} =$$

$$= \begin{bmatrix} \omega_{11} & \omega_{12} & \ldots & \omega_{1n} \\ \omega_{12} & \omega_{22} & \ldots & \omega_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ \omega_{1n} & \omega_{2n} & \ldots & \omega_{nn} \end{bmatrix} = W, \quad (9)$$

com $\omega_{ij} = c_{ij}\sigma^2$. A matriz W é simétrica e positiva definida, e σ^2 é a variância da parcela base. Assim,

$$Y \cap N\,(X\beta, W),$$

onde N indica a distribuição multinormal.

Para minimizar a soma dos quadrados dos erros na equação (8), através do método dos quadrados mínimos generalizados, seja a matriz D tal que

$$W = DD, \quad (10)$$

onde D é simétrica e não singular (Rao 1973).

Pré-multiplicando-se e pós-multiplicando-se W por D^{-1}, obtêm-se:

$$D^{-1} W D^{-1} = I. \tag{11}$$

Considere-se a variável

$$Z = D^{-1} Y. \tag{12}$$

Então, substituindo-se (8) em (12) vem:

$$Z = D^{-1} X \beta + D^{-1} \epsilon. \tag{13}$$

Seja agora

$$D^{-1} X = M \tag{14}$$

$$D^{-1} \epsilon = \zeta. \tag{15}$$

Logo (13) toma a forma

$$Z = M \beta + \zeta \tag{16}$$

Portanto,

$E(Z) = M\beta$ e $V(Z) = I$, ou seja $Z \cap N(M\beta, I)$.

A minimização da soma dos quadrados dos erros ζ, em (16), através do método simples e direto dos quadrados mínimos, conduz ao sistema de equações normais

$$M' M\hat{\beta} = M' Z$$

e ao estimador dos parâmetros

$$\hat{\beta} = (M' M)^{-1} M' Z. \tag{17}$$

Substituindo-se os valores de (12) e (14) em (17), resulta:

$$\hat{\beta} = \left[X' D^{-1} D^{-1} X\right]^{-1} X' D^{-1} D^{-1} Y \rightarrow \hat{\beta} = \left[X' (DD)^{-1} X\right]^{-1} X' (DD)^{-1} Y$$

substituindo agora (10), tem-se

$$\hat{\beta} = (X' W^{-1} X)^{-1} X' W^{-1} Y, \tag{18}$$

de onde se obtêm as estimativas:

$$\hat{a} = \bar{Y} - \hat{b} \bar{X},$$

$$\hat{b} = \frac{\sum_i \sum_j \omega^{ij} Y_i (X_i - \bar{X})}{\sum_i \sum_j \omega^{ij} X_i (X_i - \bar{X})} \tag{19}$$

onde: ω^{ij} são os elementos da matriz de informação das Y_i (Hatheway & Williams 1958); são os pesos pelos quais são ponderadas as estimativas do coeficiente de regressão e as estimativas das variâncias dos diferentes tamanhos de parcelas. $\bar{X}$ e $\bar{Y}$ são médias ponderadas, ou seja:

$$\bar{X} = \frac{\sum_i \sum_j \omega^{ij} X_i}{\sum_i \sum_j \omega^{ij}}, \quad \bar{Y} = \frac{\sum_i \sum_j \omega^{ij} Y_i}{\sum_i \sum_j \omega^{ij}}$$

A variância de $\hat{b}$ será determinada como segue:

Tomando-se a equação (18) e substituindo-se os valores de (8), obtêm-se:

$$\hat{\beta} = \beta + (X' W^{-1} X)^{-1} X' W^{-1} \epsilon \rightarrow \hat{\beta} - \beta = (X' W^{-1} X)^{-1} X' W^{-1} \epsilon.$$

Mas $V(\hat{\beta}) = E\left[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'\right]$, portanto

$$V(\hat{\beta}) = E\left\{\left[(X' W^{-1} X)^{-1} X' W^{-1} \epsilon\right]\left[\epsilon' W^{-1} X (X' W^{-1} X)^{-1}\right]\right\}$$

$$V(\hat{\beta}) = (X' W^{-1} X)^{-1}, \tag{20}$$

de onde se obtêm as variâncias dos parâmetros. Assim:

$$V(\hat{b}) = 1/\sum_i \sum_j \omega^{ij} X_i (X_i - \bar{X}) = 1/T. \tag{21}$$

Os pesos ω^{ij} serão estimados a partir dos dados. Mesmo assim, a estimativa $\hat{b}$ será de mínima variância assintótica, como se demonstra a seguir.

Para que $V(\hat{\beta})$ seja mínima, deve existir um outro estimador imparcial $\tilde{\beta}$, como função linear de Y (Rao 1973), tal que:

$$V(\hat{\beta}) < V(\tilde{\beta}).$$

Seja o estimador

$$\tilde{\beta} = \hat{\beta} + AY,$$

onde A é fixo. Assim, pela substituição de (18), tem-se

$$\tilde{\beta} - \beta = (X' W^{-1} X)^{-1} X' W^{-1} \epsilon + A X \beta + A\epsilon, \tag{22}$$

$$E(\tilde{\beta}) = \beta + \phi + A X \beta + \phi \rightarrow E(\tilde{\beta}) = \beta + A X \beta.$$

Portanto, $\tilde{\beta}$ será imparcial se para todo $\beta \in R^k$, $AX = \phi$.

Em tais condições, a variância de $\tilde{\beta}$ será como se segue:

$$V(\tilde{\beta}) = E\left[(\tilde{\beta} - \beta)(\tilde{\beta} - \beta)'\right],$$

que por substituição de (22), fica

$$V(\tilde{\beta}) = V(\hat{\beta}) + AWA' \rightarrow V(\hat{\beta}) = V(\tilde{\beta}) - AA'W.$$

portanto,

$$V(\hat{\beta}) < V(\bar{\beta}),$$

e a igualdade ocorrerá se e somente se

$$A'A = \phi + A = \phi$$

isto é, quando $\bar{\beta} = \hat{\beta}$. Logo $\hat{\beta}$ tem variância mínima assintótica.

Quando, como é freqüente o caso, existem mais de duas estimativas de variâncias, das quais se calcula a regressão, pode-se também testar a significância do desvio de regressão, para o que se deve determinar a soma de quadrados dos desvios de regressão (SQDR). Assim, minimizando-se a soma dos quadrados dos erros em (16), através do método dos mínimos quadrados (Anderson & Bancroft 1952), obtêm-se:

$$SQDR = Z'Z - \hat{\beta}'M'Z. \qquad (23)$$

Substituindo-se (12) e (14), vem

$$SQDR = Y'W^{-1}Y - \hat{\beta}'X'W^{-1}Y,$$

ou, ainda,

$$SQDR = (Y'W^{-1}Y - C) - (\hat{\beta}X'W^{-1}Y - C), \qquad (24)$$

onde

$$V = Y'W^{-1}Y - C = \text{SQ Total},$$

portanto,

$$V = \sum_i \sum_j \omega^{ij} Y_j (Y_j - \bar{Y}),$$

com (n-1) graus de liberdade, onde n é o número de variâncias estimadas.

Do mesmo modo, a soma de quadrados atribuída à regressão, devido às X_i, será obtida com base na estimativa $\hat{b}$, por

$$SQReg = \frac{\left[\sum_i \sum_j \omega^{ij} Y_i (X_i - \bar{X})\right]^2}{\sum_i \sum_j \omega^{ij} X_i (X_i - \bar{X})} = \frac{U^2}{T},$$

portanto, (24) é escrita agora como

$$SQDR = V - U^2/T,$$

que tem distribuição de qui-quadrado com n-2 graus de liberdade (Rao 1973).

Estimação dos pesos ω^{ij}

Os pesos apropriados, pelos quais são ponderadas as variâncias observadas de parcelas de diferentes tamanhos, que conduzem a um estimador imparcial de β, com variância mínima assintótica, são os elementos da matriz de informação dos valores de Y_i.

Como Y_i representa as variâncias das parcelas bases ("repetições", blocos, parcelas e subparcelas), cujas variâncias V_1', V_2', V_3' e V_4' são funções lineares dos quadrados médios independentes, V_1, V_2, V_3 e V_4, as estimativas de suas variâncias são respectivamente:

$$V(V_1) = 2V_1^2/(I-1), V(V_2) = 2V_2^2/I(J-1), V(V_3) = 2V_3^2/IJ(K-1) \text{ e } V(V_4) = 2V_4^2/IJK(L-1). \qquad (25)$$

Se S^2 é a variância de uma amostra aleatória de tamanho n, de uma população normal de variância σ^2, então

$$E(S^2) = \sigma^2, \quad V(S^2) = 2\sigma^4/(n-1),$$

onde (n-1) representa o número de graus de liberdade.

Asssim, como V_1, V_2, V_3 e V_4 são independentes e V_i (i = 1, 2, 3, 4) é a variância da i-ésima amostra aleatória, de população normal, então

$$V(V_i) = 2V_i^2/(n-1).$$

Visto que V_1', V_2', V_3' e V_4' são funções lineares de V_1, V_2, V_3 e V_4, não apenas a variância de V_1', mas também a sua covariância com outro V_i' são proporcionais a V_1^2, ou seja:

$$V(V_1') = 2V_1^2/(I-1).$$

Desta forma, determinam-se os elementos da matriz de informação de V_i'. Ou seja:

$$V(V_1) = 2V_1^2/(I-1), \quad V(V_2) = 2V_2^2/I(J-1),$$

$$V_2' = \left[I(J-1)V_2 + (I-1)V_1\right]/(IJ-1), \quad V(V_2') = E\left[V_2' - E(V_2')\right]^2$$

$$E(V_2') = E\left[I(J-1)V_2/(IJ-1) + (I-1)V_1/(IJ-1)\right]$$

$$V(V_2') = E\left\{\frac{\left[I(J-1)\right]^2}{(IJ-1)^2} E\left[V_2 - E(V_2)\right]^2 + \frac{(I-1)^2}{(IJ-1)^2} E\left[V_1 - E(V_1)\right]^2\right\}$$

Por interdependência de V_1 e V_2,

$$V(V_2') = \frac{\left[I(J-1)\right]^2}{(IJ-1)^2} V(V_2) + \frac{(I-1)^2}{(IJ-1)^2} V(V_1),$$

que, pela substituição de (25), tem-se

$$V(V'_2) = \frac{[I(J-1)]^2}{(IJ-1)^2} \cdot \frac{2}{I(J-1)} V_2^2 + \frac{(I-1)^2}{(IJ-1)^2} \cdot \frac{2}{(I-1)} V_1^2,$$

$$V(V'_1) = \frac{2\left[I(J-1)\,V_2^2 + (I-1)\,V_1^2\right]}{(IJ-1)^2}.$$

Usando o mesmo procedimento, determina-se

$$V(V'_3) = \frac{2\left[IJ(K-1)\,V_3^2 + I(J-1)\,V_2^2 + (I-1)\,V_1^2\right]}{(IJK-1)^2},$$

$$V(V'_4) = \frac{2\left[IJK(L-1)\,V_4^2 + IJ(K-1)\,V_3^2 + I(J-1)\,V_2^2 + (I-1)\,V_1^2\right]}{(IJKL-1)^2}$$

Do mesmo modo, como

$$\mathrm{Cov}\,(V'_i, V'_j) = E\left\{\left[V'_i - E\,(V'_i)\right]\left[V'_j - E\,(V'_j)\right]\right\},$$

suas expressões são respectivamente:

$\mathrm{Cov}\,(V'_1, V'_2) = 2\,(I-1)\,V_1^2/(I-1)(IJ-1)$, $\mathrm{Cov}\,(V'_1, V'_3) = 2\,(I-1)\,V_1^2/(I-1)\,(IJK-1)$,

$\mathrm{Cov}\,(V'_1, V'_4) = 2\,(I-1)V_1^2/(I-1)\,(IJKL-1)$, $\mathrm{Cov}\,(V'_2, V'_3) = \left[2\,I(J-1)V_2^2 + 2\,(I-1)\,V_1^2\right]/(IJ-1)\,(IJK-1)$,

$\mathrm{Cov}\,(V'_2, V'_4) = \left[2\,I(J-1)\,V_2^2 + 2\,(I-1)\,V_1^2\right]/(IJ-1)\,(IJKL-1)$,

$\mathrm{Cov}\,(V'_3, V'_4) = \left[2\,IJ(K-1)\,V_3^2 + 2\,I(J-1)\,V_2^2 + 2\,(I-1)\,V_1^2\right]/(IJK-1)\,(IJKL-1)$.

Logo, a matriz de informação das V'_i é a que se segue:

$$\begin{bmatrix}
\dfrac{D}{(I-1)^2} & \dfrac{D}{(I-1)\,(IJ-1)} & \dfrac{D}{(I-1)\,(IJK-1)} & \dfrac{D}{(I-1)\,(IJKL-1)} \\[2ex]
\dfrac{D}{(I-1)\,(IJ-1)} & \dfrac{C+D}{(IJ-1)^2} & \dfrac{C+D}{(IJ-1)\,(IJK-1)} & \dfrac{C+D}{(IJ-1)\,(IJKL-1)} \\[2ex]
\dfrac{D}{(I-1)\,(IJK-1)} & \dfrac{C+D}{(IJ-1)\,(IJK-1)} & \dfrac{B+C+D}{(IJK-1)^2} & \dfrac{B+C+D}{(IJK-1)\,(IJKL-1)} \\[2ex]
\dfrac{D}{(I-1)\,(IJKL-1)} & \dfrac{C+D}{(IJ-1)\,(IJKL-1)} & \dfrac{B+C+D}{(IJK-1)\,(IJKL-1)} & \dfrac{A+B+C+D}{(IJKL-1)^2}
\end{bmatrix}$$

onde

$A = 2\,IJK(L-1)\,V_4^2$, $B = 2\,IJ(K-1)\,V_3^2$, $C = 2\,I(J-1)\,V_2^2$, $D = (I-1)\,V_1^2$.

Portanto, a matriz inversa das V'_i será:

$$\begin{bmatrix} (I-1)^2 (\frac{1}{C} + \frac{1}{D}) & -\frac{(I-1)(IJ-1)}{C} & 0 & 0 \\ -\frac{(I-1)(IJ-1)}{C} & (IJ-1)^2 \left(\frac{1}{C} + \frac{1}{D}\right) & -\frac{(IJ-1)(IJK-1)}{B} & 0 \\ 0 & -\frac{(IJ-1)(IJK-1)}{B} & (IJK-1)^2 \left(\frac{1}{C} + \frac{1}{D}\right) & -\frac{(IJK-1)(IJKL-1)}{A} \\ 0 & 0 & -\frac{(IJK-1)(IJKL-1)}{A} & \frac{(IJKL-1)^2}{A} \end{bmatrix}$$

Isto posto, determinam-se os pesos ω^{ij}, multiplicando-se cada linha e cada coluna da matriz de informação das V'_j pelas correspondentes V'_i.

Assim, designando-se por

$$\omega t\,(Y_i, Y_j), \omega t\,(V'_i, V'_j)$$

os elementos na posição (i, j) das matrizes de informação de Y_i e V'_i, respectivamente, tem-se:

$$\omega t\,(Y_i, Y_j) = \omega^{ij} = V'_i \,.\, V'_j\, \omega\,(V'_i, V'_j).$$

Desta forma, a matriz W^{-1} será:

$$\begin{bmatrix} \left[(I-1)\, V'_1\right]^2 (\frac{1}{C} + \frac{1}{D}) & -\frac{(I-1)(IJ-1)}{C} V'_1 V'_2 & 0 & 0 \\ -\frac{(I-1)(IJ-1)}{C} V'_1 V'_2 & \left[(IJ-1)\, V'_2\right]^2 (\frac{1}{C} + \frac{1}{B}) & -\frac{(IJ-1)(IJK-1)}{B} V'_2 V'_3 & 0 \\ 0 & -\frac{(IJ-1)(IJK-1)}{B} V'_3 V'_4 & \left[(IJK-1)\, V'_3\right]^2 (\frac{1}{A} + \frac{1}{A}) & -\frac{(IJK-1)(IJKL-1)}{A} V'_3 V'_4 \\ 0 & 0 & -\,-\frac{(IJK-1)(IJKL-1)}{A} V'_3 V'_4 & \frac{\left[(IJKL-1)\, V'_4\right]^2}{A} \end{bmatrix} \quad (26)$$

São, portanto, os pesos ω^{ij} facilmente estimados da própria matriz de informação V'_i. Assim, ω^{11}, por exemplo, é estimado como a seguir.

$$\omega^{11} = \left[(I-1)\, V'_i\right]^2 (\frac{1}{C} + \frac{1}{D}).$$

Observa-se que a base de todos os cálculos expostos é a matriz inversa, da matriz de variância e covariância (i. é. a matriz de informação) das V_i', das parcelas de diferentes tamanhos. Essas variâncias são expressas como combinações lineares dos quadrados médios originais, independentes, e não em termos dos componentes de variância, os quais são correlacionados entre si. A matriz de informação resultante toma, como se vê, uma forma relativamente simples.

Se, porém, empregarem-se logarítmos decimais, os pesos ficam multiplicados pelo fator

$(\log_e 10)^{-2}$, que indica o ajuste necessário para tal, pois

$$\log_e V_i' \log_e V_j' = \frac{\log_{10} V_i' \cdot \log_{10} V_j'}{\log_{10} e \cdot \log_{10} e}$$

$$\log_e V_i' \cdot \log_e V_j' = \log_{10} V_i' \cdot \log_{10} V_j' \cdot (\log_{10} e)^{-2}.$$

Em tal caso, embora o valor de $\hat{b}$ continue como em (19), sua variância passa a ser

$$V(\hat{b}) = 1/5{,}302\ T \text{ ou } V(\hat{b}) = 0{,}1886\ T^{-1}.$$

Conseqüentemente, os limites do intervalo de confiança para b serão agora expressos sob a forma:

$$\hat{b} \pm t_\alpha \sqrt{0{,}1886\ T^{-1}} \quad \hat{b} \pm 0{,}4343\ t_\alpha T^{-1/2}.$$

O procedimento apresentado para estimar os pesos ω^{ij} é preciso, como se demonstra a seguir.

Se as variâncias não são afetadas pelo tamanho da parcela, então todas as somas de quadrados disponíveis são estimadas da mesma variância básica. Mas as estimativas do logarítmo das variâncias calculadas das diferentes fontes da análise são independentes e têm variância assintótica, ou seja:

Se $\hat{\lambda}_i = \log V_i'$, cuja $V(V_i) = V_i'$

é assintótica,

$$V_i' = 2/n,$$

onde n é o número de graus de liberdade usado para estimar V_i'.

Por outro lado, a "informação" é dada por

$$\text{Inf} = 1/V_i'.$$

No caso $\text{Inf} = \sum_i \sum_j \omega^{ij} + \sum_i \sum_j \omega^{ij} = \frac{1}{2/n}$,

logo

$$\sum_i \sum_j \omega^{ij} = n/2$$

portanto, para dados de ensaio de uniformidade

$$\sum_i \sum_j \omega^{ij} = (IJKL-1)/2. \qquad (27)$$

ESTIMAÇÃO DE b BASEADA EM DADOS DE OUTROS ENSAIOS

O procedimento para se estimar $\hat{b}$, neste caso, é o mesmo do caso anterior. Mas se existirem tratamentos, as diferentes variâncias de parcelas são estimadas com um número menor de graus de liberdade, portanto, com menor precisão que em ensaios de uniformidade. Afastada essa hipótese, o procedimento é exatamente o mesmo.

Tome-se um experimento em blocos ao acaso, com parcelas subdivididas, no qual são calculadas quatro estimativas de variância (Koch & Rigney 1951). Considere-se, nesse modelo, que não exista interação bloco x tratamento. Assim,

$$Y_{ijkl} = m + b_i + t_j + \epsilon_{1ij} + t_k' + (tt')_{jk} + \epsilon_{2ijk} + e_{ijkl} \quad (28)$$

onde

m é a média geral, b_i é o efeito do bloco i (i = 1, 2, ..., I), t_j é o efeito do tratamento j (j = 1, 2, ..., J), ϵ_{1ij} é o efeito da interação (b x t) ou erro 1 e representa bloco dentro de repetição, t_k' é o efeito do tratamento k (k = 1, 2, ..., k), ϵ_{2ijk} é o efeito da interação (b x t x t') ou erro 2 e representa parcelas dentro de tratamentos, dentro de blocos, e_{ijkl} é o efeito de subparcelas dentro de parcelas l (l = 1, 2, ..., L).

Minimizando-se a soma dos quadrados dos erros, em (28), obtêm-se os componentes de variância da Tabela 2.

Analogamente ao caso anterior, a variância estimada de parcelas do tamanho de uma repetição é V_1. Esta é estimada com o número de graus de liberdade para blocos (I-1), portanto sua vaiância V_1' será:

$$V_1' = 2\,V_1^2/(I-1).$$

Na estimação do quadrado médio de blocos (variância para blocos), aqui representada pela interação blocos x tratamentos, deve-se atentar para o fato de que, dos

$$I(J-1) = (I-1)(J-1) + (J-1)$$

contrastes de blocos dentro de "repetições", apenas (I-1) (J-1) são úteis para estimar a variância, sendo que os outros (J-1) contêm efeitos de tratamentos. Logo, a variância estimada entre blocos (2) é agora

$$V_2' = 2\left[I^2 (J-1) V_2^2 + (I-1) V_1^2\right] / (IJ-1)^2.$$

TABELA 2. Análise de variância.

Causa de variação	G.L.	Q.M.	E (Q.M.)
Blocos	I-1	V_1	$\sigma^2 + L\sigma_P^2 + KL\sigma_B^2 + JKL\sigma_R^2$
Tratamentos (1)	J-1		$\sigma^2 + L\sigma_P^2 + KL\sigma_B^2 + \frac{IKL}{J-1}\sum_j t_j^2$
Resíduo (1)	(I-1) (J-1)	V_2	$\sigma^2 + L\sigma_P^2 + KL\sigma_B^2$
Tratamentos (2)T'	K-1		$\sigma^2 + L\sigma_P^2 + \frac{IJL}{K-1}\sum_k t_k'^2$
Interação (T x T')	(J-1) (K-1)		$\sigma^2 + L\sigma_P^2 + \frac{IJL}{(J-1)(K-1)}\sum_{j,k} (t\,t')^2_{jk}$
Resíduo (2)	J(I-1) (K-1)	V_3	$\sigma^2 + L\sigma_P^2$
Subparcelas	IJ(K-1)		
Erro amostral	IJK(L-1)	V_4	σ^2
Total	IJKL-1		

Analogamente, a variância V_1' de V_1 é ajustada pelo fator 1/ (I-1).

A análise continua agora, como no caso de ensaios de uniformidade, sendo a matriz de informação das V_i' a que se segue,

$$\begin{bmatrix} (I-1)^2\left(\frac{1}{C}+\frac{1}{D}\right) & -\frac{(I-1)(IJ-1)}{C} & 0 & 0 \\ -\frac{(I-1)(IJ-1)}{C} & (JK-1)^2\left(\frac{1}{B}+\frac{1}{C}\right) & -\frac{(IJ-1)(IJK-1)}{B} & 0 \\ 0 & -\frac{(IJ-1)(IJK-1)}{B} & (IJK-1)^2\left(\frac{1}{A}+\frac{1}{B}\right) & -\frac{(IJK-1)(IJKL-1)}{A} \\ 0 & 0 & -\frac{(IJK-1)(IJKL-1)}{A} & \frac{(IJKL-1)^2}{A} \end{bmatrix}$$

onde

$A = 2\,IJK(L-1)\,V_4^2$, $B = \left[2\,I^2J(K-1)\,V_3^2\right]/(I-1)$, $C = \left[2\,I^2(J-1)\,V_2^2\right]/(I-1)$, $D = 2(I-1)V_1^2$.

A estimação dos pesos é feita como no caso anterior, sendo a matriz W^{-1} a de (26). Por sua vez, para este experimento, a soma dos pesos será

$$\sum_i \sum_j \omega^{ij} = (1/2)\, JK\, (IL-1).$$

Idênticos resultados podem ser obtidos com o uso de reticulados quadrados e outros tipos de delineamentos experimentais.

ESTIMAÇÃO DO TAMANHO ÓTIMO DA PARCELA

Tamanho ótimo da parcela em função do custo

A variabilidade do solo e a relação entre o custo fixo e o que varia com o número de unidades (custos da condução do experimento) influenciam o tamanho da parcela (Smith 1938).

Seja o custo de r repetições dado pela função linear

$$C(x) = r K_1 + r K_2 x, \qquad (29)$$

onde,

r é o número de repetições, K_1 é a parte do custo que é proporcional ao número de parcelas por tratamento, K_2 é a parte do custo proporcional à área total (t) por tratamento.

Usando-se o multiplicador de Lagrange (λ), do Cálculo Diferencial, sejam

$$\log V(x) = L,\ F(x) = C(x) + \lambda \left[\log V(x) - L\right],$$

onde λ é o multiplicador de Lagrange, C (x) é dado em (29) e log V (x) em (6).

Assim,

$$F(x) = r K_1 + r K_2\, x + \lambda \left[\log V_1 - \log r - b \log x - L\right].$$

Minimizando-se F (x) em relação a x e r, e resolvendo-se ambas as equações em λ, tem-se

$$x = \frac{bK_1}{(1-b)\, K_2}, \qquad (30)$$

que estima o tamanho ótima da parcela, em função dos custos determinados pela razão K_1/K_2.

Tamanho ótimo da parcela em função do número de repetições

Associando-se o coeficiente b à equação que estima o número de repetições (Cochran & Cox 1957),

$$r = \left[2\,(CV)^2\,(t_1 + t_2)^2\right] / d^2, \qquad (31)$$

obtêm-se o tamanho da parcela, independentemente do custo, para uma determinada diferença de médias, que se deseje conhecer "a priori" (Hatheway 1961).

Considerando-se (5) e substituindo-se V_1 pelo seu valor correspondente, vem

$$V(x) = V'(x)/x \text{ ou } V(x) = s^2(x)/x.$$

Mas o coeficiente de variação para parcelas de tamanho x é calculado por

$$CV(x) = s(x)/m(x),$$

onde m (x) é a média aritmética de parcelas de tamanho x.

Note-se que, como a variância, a média também é reduzida à unidade básica:

$$m_1 = m(x)/x,$$

portanto,

$$CV(x) = V(x)/x\, m_1$$

que, substituindo em (31), resulta

$$x^b = \frac{2\,(t_1 + t_2)\,(CV_1)^2}{r d^2}, \qquad (32)$$

que dá o tamanho ótimo da parcela em função de r, onde: r é o número de repetições exigido para se detectar uma diferença de d unidades; d é a diferença entre dois tratamentos (medida em percentagem da média): é a diferença que se deseja comprovar, CV_1 é o coeficiente de variação de parcelas de tamanho unitário, t_1 e t_2 são os valores do teste t correspondentes às probabilidades desejadas.

O cálculo "a priori", das diferenças entre médias, ao nível α de probabilidade, é feito com diferentes valores de CV, para vários tamanhos de parcelas.

O cálculo das diferenças é feito através de

$$d = \sqrt{\frac{2\,(t_1 + t_2)\,(CV_1)^2}{(r\,.\,x^b)}},$$

que resulta imediatamente de (32).

CONCLUSÕES

1. Para se estimar o tamanho ideal de parcelas, o elemento principal é o coeficiente de heterogeneidade do solo (b).

2. O tamanho ideal de parcelas pode ser estimado em função dos custos na condução do experimento.

3. O número de repetições para estabelecer o tamanho ideal de parcelas, para determinada diferença de médias, fixada "a priori", não pode ser generalizado, pois varia com o solo e com a cultura.

4. O tamanho ideal de parcelas pode ser determinado tanto através de ensaios de uniformidade como através de outros ensaios.

REFERÊNCIAS

ANDERSON, R.L. & BANCROFT, T.A. Statistical theory in research. New York, Mc Graw-Hill, 1952.

CALZADA BENZA, J. Métodos estadísticos para la investigación. Lima, s.ed., 1966.

COCHRAN, W.G. & COX, G.M. Experimental design. New York, John Wiley, 1957. 611p.

HATHEWAY, W.H. Convenient plot size. Agron. J., 53: 279-80, 1961.

HATHEWAY, W.H. & WILLIAMS, E.J. Efficient estimation of the relationship between plot size and the variability of crop yields. Biometrics, 44:207-22, 1958.

KOCH, E.J. & RIGNEY, J.A. A method of estimating optimum plot size from experimental data. Agron. J., 43:17-21, 1951.

RAO, C.R. Linear estatistical inference and its applications. 2 ed. New York, John Wiley, 1973.

ROBINSON, H.F.; RIGNEY, J.A. & HARVEY, P.H. Investigations in plot technique with peanuts. N.C. Agric. Exp. Stn. Res. Tech. Bull., 86:1-19, 1948.

RÓGER ARROYO, V.J. & CHAVES, A.M. Estimación eficiente de parámetros em la determinación del tamaño óptimo de parcela. Estac. Exp. Agric. de La Molina Boletin, (15):1-32, 1966.

SMITH, H.F. An empirical law describing heterogeneity in the yields of agricultural crops. J. Agric., 28: 1-23, 1938.

TORRIE, J.H.; SCHMIDT, D.R. & TENPAS, G.H. Estimates of optimum plot size and shape and replicate number for forage yield of alfalfa-bromegrass mixtures. Agron. J., 55(3):258-60, 1963.