O PROBLEMA DO TAMANHO DAS PARCELAS EM EXPERIMENTOS COM PLANTAS ARBÓREAS[1]

F. PIMENTEL-GOMES[2]

RESUMO - Este trabalho desenvolve uma teoria sobre a determinação do tamanho ótimo de parcelas baseada no coeficiente de correlação intraclasse (ρ) entre unidades (subparcelas) dentro das parcelas do experimento. Essa abordagem tem a vantagem de utilizar uma estatística bem conhecida, que varia dentro de limites previamente conhecidos, que são: $[-1/(k-1) < \rho < 1]$, onde k é o número de subparcelas dentro da parcela. Considera-se ótimo o número de subparcelas (k), que dá a variância mínima da média de um tratamento, quando se fixa a área total relativa a esse tratamento. Discute-se detalhadamente a influência de bordadura completa ou de meia bordadura, e ainda da ausência de bordadura. A teoria se aplica especialmente ao caso de experimentos com árvores, e conduz a conclusões de grande generalidade.

Termos para indexação: coeficiente de correlação intraclasse, bordaduras, tamanho ótimo de parcelas.

THE PROBLEM OF PLOT SIZE IN EXPERIMENTS WITH TREES

ABSTRACT - This paper develops a theory on the determination of optimum plot size, based on the intraclass correlation coefficient (ρ). This method has the advantage of using a well known statistic, which belongs always to the closed interval $[-1/(k-1) < \rho < 1]$, where k is the number of units (subplots) within each plot. The plot taken as optimum is that with k units, which leads to minimum variance of treatment means, when the total area per treatment is constant. Guard rows are taken in consideration in the solution. The theory is specially appropriate for research on plot size in experiments with trees, with or without guard rows.

Index terms: intraclass correlation coefficient, guard rows, optimum plot size.

INTRODUÇÃO

O tamanho ótimo das parcelas em experimentos de campo com vegetais é problema relativamente simples de resolver e de interesse apenas moderado quando se trata de plantas de pequeno porte, como os cereais, o feijoeiro e o algodoeiro herbáceo, e muito principalmente quando não há bordadura. Mas a situação muda de figura quando se trata de plantas grandes, como as fruteiras perenes, principalmente quando há necessidade de bordadura.

A bibliografia sobre a determinação do tamanho ótimo de parcelas é relativamente abundante, especialmente no que se refere a plantas anuais. Oliveira & Biava (1982) publicaram, aliás, uma bibliografia muito completa sobre tudo que saiu sobre o assunto de 1890 para cá. Entre os trabalhos que mais nos interessam podem-se citar os de Smith (1938), Koch & Rigney (1951), Fraga Junior & Conagin (1956), Oliveira (1976), Rossetti & Pimentel-Gomes (1983), Cordeiro & Miranda (1983).

Um método tradicional de estudo do tamanho ótimo de parcelas são os ensaios em branco. Na verdade, porém, o melhor é utilizar os experimentos correntes (Koch & Rigney 1951), de tal sorte que tenhamos, em cada um, estimativas de quadrados médios relativas a pelo menos dois tamanhos distintos de parcelas. Bons resultados se obtêm especialmente nos experimentos em parcelas subdivididas ou nos ensaios sem essa particularidade em que se considerem, de propósito, subparcelas dentro de cada parcela, sem diferença de tratamento entre elas. Por exemplo, num experimento inteiramente casualizado ou em blocos ao acaso, com parcelas de três mangueiras úteis, basta colher separadamente as frutas de cada mangueira. Obtém-se assim um quadrado médio relativo a parcelas (grupos de três mangueiras), e outro, referente à variação entre mangueiras dentro da parcela.

É tradicional também (Smith 1938) utilizar no cálculo do tamanho ótimo das parcelas o custo da experimentação, avaliado sob dois pontos de vista: a. Um coeficiente K_1, que dá o custo por parcela;

[1] Aceito para publicação em 3 de outubro de 1984.

[2] Eng.-Agr., Dr., Prof. Catedrático, ESALQ-USP (aposentado), Consultor do IICA/EMBRAPA, CEP 70333 Brasília, DF.

b) Um coeficiente K_2, que dá o custo por área. O custo total seria, pois:

$$C(x) = N(K_1 + K_2 x),$$

onde N é o número de parcelas e x é a área de cada uma.

Na prática, porém, tem sido considerado difícil avaliar devidamente esses custos, o que prejudica o uso desse método.

No presente artigo, o problema é abordado através do uso da correlação intraclasse e com cuidadosa consideração das bordaduras. O método é relativamente simples, não exige ensaio em branco e leva a conclusões de grande generalidade, passíveis de recomendações de uso muito amplo.

A consideração da correlação intraclasse no estudo do tamanho de parcelas não é totalmente nova. Já Pimentel-Gomes et al. (1964) a aplicaram na resolução de um problema de amostragem de colmos de cana-de-açúcar para determinações tecnológicas. E a idéia de aplicar a correlação intraclasse à variação entre subparcelas já aparece claramente em Fisher (1938) e em Cochran & Cox (1950). Note-se que Oliveira (1976) usou um conceito similar, o de "coeficiente auto-regressivo de correlação", denominação que talvez fosse substituída com vantagem por "coeficiente de correlação serial", ou por "coeficiente de autocorrelação" (Rodrigues 1970). Mas este modelo matemático é, certamente, demasiado sofisticado e irrealista para estudo de dados experimentais do tipo em consideração.

Neste artigo, discutiremos especificamente o problema do tamanho de parcelas de ensaios com árvores, muito embora toda a teoria se aplique também a plantas pequenas, desde que se considerem, em vez de árvores, linhas ou subparcelas unitárias de maneira geral (Bueno & Pimentel-Gomes 1983).

No caso específico das árvores, Pearce (1953) declara que, para economizar área, devem-se usar parcelas pequenas, e para economizar trabalho especializado ("skilled labour"), devem ser preferidas parcelas grandes. No entanto, essas recomendações só são válidas em casos muito especiais, como se verá neste artigo.

O MODELO MATEMÁTICO

Vamos admitir que temos um ensaio com parcelas de k árvores úteis, sendo os dados obtidos planta por planta. Em outras palavras, cada parcela tem k subparcelas. Por exemplo, no caso de c cultivares de macieira em b blocos casualizados, com k plantas por parcela, a análise da variância seria a seguinte. Aí o símbolo E indica esperança matemática.

Causa de variação	G.L.	Q.M.	E (Q.M.)
Blocos	b-1		
Cultivares	c-1		
Resíduo (a)	(b-1) (c-1)	V_1	$\sigma_1^2 + k\,\sigma_2^2$
Resíduo (b)	bc (k-1)	V_2	σ_1^2

Essa análise corresponde ao modelo matemático

$$Y_{ijk} = m + c_i + b_j + e_{ij} + e_{ijk}, \qquad \text{(I)}$$

onde m é a média, c_i ($i = 1, 2, \ldots, c$) é o efeito da cultivar, b_j ($j = 1, 2, \ldots, b$) é o efeito do bloco, os componentes e_{ij}, e_{ijk} são aleatórios, com distribuição normal e variâncias σ_1^2 e σ_2^2, respectivamente, e $E(e_{ij}\,e_{ijk}) = 0$.

Por esse modelo devemos ter sempre

$$E(V_1) = \sigma_1^2 + k\,\sigma_2^2 > \sigma_1^2 = E(V_2).$$

Um modelo mais geral é:

$$Y_{ijk} = m + c_i + b_j + e_{ijk}, \qquad \text{(II)}$$

com $E(e_{ijk}^2) = \sigma^2$, $E(e_{ijk}\,e_{i'j'k'}) = 0$ para $(i, j) \neq (i', j')$, $E(e_{ijk}\,e_{ijk'}) = \rho\,\sigma^2$, onde ρ é o coeficiente de correlação intraclasse.

A análise da variância dá, então, os resultados seguintes:

Causa de variação	G.L.	Q.M.	E(Q.M.)
Blocos	b-1		
Cultivares	c-1		
Resíduo (a)	(b-1) (c-1)	V_1	$\sigma^2\,[1 + (k-1)\rho]$
Resíduo (b)	bc (k-1)	V_2	$\sigma^2\,(1 - \rho)$

Com $\rho > 0$ estes resultados equivalem aos anteriores. Mas com $\rho < 0$, temos a possibilidade nova de obter $E(V_1) < E(V_2)$, o que é importante.

Devemos ter sempre

$$\sigma^2\,[1 + (k-1)\rho] > 0,$$

isto é:

$$\rho > -1/(k-1) \qquad (k > 1). \qquad \text{(III)}$$

Essa desigualdade nos mostra que, quando aumenta o valor de k, o valor de ρ, se inicialmente negativo, estará cada vez mais próximo de zero, isto é, para $\rho < 0$ temos

$$\lim_{k \to \infty} \rho = 0$$

Em qualquer caso, temos sempre

$$-1/(k-1) \leq \rho \leq 1 \qquad (k > 1).$$

Da análise de variância exposta acima, deduz-se facilmente um estimador de ρ:

$$\hat{\rho} = \frac{V_1 - V_2}{V_1 + (k-1)V_2} \qquad (k > 1).$$

Daí se conclui que:

$$V_1 > V_2 \implies \hat{\rho} > 0,$$

$$V_1 < V_2 \implies \hat{\rho} < 0.$$

Por outro lado, sendo V_1 e V_2 ambos positivos, temos necessariamente

$$-1/(k-1) < \hat{\rho} < 1.$$

Note-se que, para $1 > \rho > 0$, temos:

$$\rho/(1-\rho) = (\sigma_2/\sigma_1)^2.$$

O coeficiente de variação

A variância relativa a cada parcela tem a expressão

$$V_1 = \sigma^2 [1 + (k-1)\rho],$$

de onde se conclui que, sendo m a média geral das subparcelas, o coeficiente de variação, relativo ao resíduo (a) (de parcelas), será dado pela expressão:

$$CV = (100/km) \sqrt{\sigma^2 [1 + (k-1)\rho]}$$

$$= (100\,\sigma/m) \sqrt{\frac{(1-\rho) + k\rho}{k^2}}$$

$$= (100\,\sigma/m) \sqrt{\frac{1-\rho}{k^2} + \frac{\rho}{k}},$$

com

$$1 + (k-1)\rho > 0.$$

Com $\rho > 0$, o máximo de CV se dá para $k = 1$, evidentemente, pois ambos os termos do radicando

$$\frac{1-\rho}{k^2} + \frac{\rho}{k}$$

são funções decrescentes de k. (Com $\rho = 1$ o primeiro termo é nulo, mas isto não altera o resultado).

Com $\rho < 0$, o numerador do radicando

$$(1-\rho) + k\rho$$

decresce quando k cresce, mas seu valor é restringido pela condição necessária de que tenhamos

$$(1-\rho) + k\rho > 0.$$

Ao mesmo tempo, cresce o denominador k^2, de sorte que o radicando é monotonicamente decrescente, para $k > 1$.

A conclusão final é, pois, de que o coeficiente de variação é função decrescente do número k de árvores úteis por parcela, o que favoreceria o uso de parcelas grandes.

A variância da média de um tratamento

O mais importante, porém, é reduzir a variância da média de cada tratamento, sem aumentar o número de árvores do experimento, isto é, para uma área fixa ou um número fixo de árvores, tornar mínima a variância da média de cada tratamento. Considerado o modelo (II), que é geral, a variância da média de r repetições de um tratamento é dada pela expressão

$$V(\hat{m}) = (\sigma^2/kr)\,[1 + (k-1)\rho].$$

Não havendo bordadura e sendo N o número total de árvores por tratamento, temos $N = kr$, logo:

$$V(\hat{m}) = (\sigma^2/N)\,[1 + (k-1)\rho].$$

Com $\rho > 0$, essa variância é mínima para $k = 1$, isto é, para parcelas de uma única árvore. Já com $\rho < 0$, essa variância é função decrescente de k. Por exemplo, com $N = 10$, $\rho = -0{,}050$, temos:

$$f(k) = V(\hat{m}) = (\sigma^2/10)\,[1-(k-1)\,0{,}050],$$

$$f(1) = (\sigma^2/10), \qquad f(2) = (\sigma^2/10)\,0{,}95,$$

$$f(5) = (\sigma^2/10)\,0{,}80, \qquad f(10) = (\sigma^2/10)\,0{,}55.$$

O mínimo se daria, pois, para $k = N$ ou $r = 1$, mas com duas restrições:

a. Devemos ter $1 + (k-1)\rho > 0$;
b. Deve haver um numero razoável de repetições $r > 2$, que permita obter um número razoável de graus de liberdade para o resíduo (em geral, pelo menos 10 g.l.).

Parcelas com meia bordadura

Consideraremos inicialmente o caso em que as árvores úteis estão dispostas em uma única linha. Nestas condições, o número total de árvores por parcela (K) que tenha k árvores úteis é dado pela expressão:

$$K = 2\,(k + 1),$$

e o número total de árvores por tratamento (N) será:

$$N = r\,K = 2r\,(k + 1).$$

Então, a variância da média de um tratamento, no caso do modelo (II), será:

$$V(\hat{m}) = (\sigma^2/rk)\;[1 + (k\text{-}1)\,\rho]$$

$$= \frac{2\,(k + 1)\,\sigma^2}{N\,k}\;[1 + (k\text{-}1)\,\rho]$$

$$= (2\,\sigma^2/N)\;[(1 + 1/k) + (k\text{-}1/k)\rho\,].$$

Com $\rho > 0$, esta função tem mínimo para

$$k = \sqrt{\frac{1 - \rho}{\rho}} \qquad (\rho > 0).$$

Para $\rho > 0{,}50$, a solução é, pois, $k = 1$; para $\rho > 0{,}20$, temos $k < 2$; para $\rho > 0{,}10$, conclui-se que temos $k < 3$.

Estudemos, agora, $V(\hat{m})$ no caso de $\rho < 0$. Temos:

$$\frac{dV}{dk} = (2\,\sigma^2/N)\;\;[(1/k^2)\,(\rho - 1) + \rho].$$

Para $\rho < 0$, esta derivada é negativa, qualquer que seja k. Nestas condições, para reduzir o valor de $V(\hat{m})$ devemos aumentar k o mais possível, mas com a condição de ter um número razoável de graus de liberdade para o resíduo.

No caso de meia bordadura com duas linhas de árvores úteis temos:

$$K = (3/2)\,(k + 2)\,,$$

$$V(\hat{m}) = (3\,\sigma^2/2N)\;[\,(1 + 2/k) + (k + 1 - 2/k)\,\rho\,],$$

onde $k > 2$ é um número par. Está claro que com $k = 2$ temos somente uma árvore em cada uma das duas linhas úteis, o que equivale a ter uma única linha com duas árvores. Conclui-se, pois, que devemos ter, na verdade, $k > 2$.

Com $\rho > 0$, a variância $V(\hat{m})$ é mínima para

$$k = \sqrt{\frac{2\,(1 - \rho)}{\rho}} \qquad (\rho > 0). \qquad \text{(V)}$$

Mas a condição de termos $k > 2$ importa em ter $\rho < 1/3$. Em tais condições, a expressão (V) sugere os valores mais próximos (pares) de k a serem tentados. Por exemplo, para $\rho = 0{,}10$ temos:

$$\sqrt{\frac{2\,(1 - 0{,}10)}{0{,}10}} = 4{,}24.$$

Para $k = 4$ temos:

$$V(\hat{m}) = (\sigma^2/N)\;2{,}92,$$

e para $k = 6$:

$$V(\hat{m}) = (\sigma^2/N)\;3{,}00.$$

Então, a solução seria $k = 4$.

Passemos, agora, ao caso de $\rho < 0$, ainda com duas filas de árvores úteis. Temos:

$$\frac{dV}{dk} = (3\,\sigma^2/2N)\;[\,-(2/k^2) + (1 + 2/k^2)\,\rho\,].$$

Para $\rho < 0$, esta derivada é negativa, qualquer que seja k (número natural par). Nestas condições, para reduzir o valor de $V(\hat{m})$ devemos aumentar k o mais possível, mas com a condição de ter um número razoável de graus de liberdade para o resíduo.

Parcelas com bordadura completa

No caso do modelo (II), estando as k árvores úteis dispostas em uma única linha, temos:

$$K = 3(k + 2)\,,$$

$$V(\hat{m}) = (\sigma^2/rk)\,[1 + (k\text{-}1)\,\rho\,]$$

$$= (3\sigma^2/N)\;\;[(1 + 2/k) + (k + 1 - 2/k)\,\rho\,].$$

Com $\rho > 0$, essa variância é mínima para

$$k = \sqrt{\frac{2(1 - \rho)}{\rho}} \qquad (\rho > 0). \qquad \text{(VI)}$$

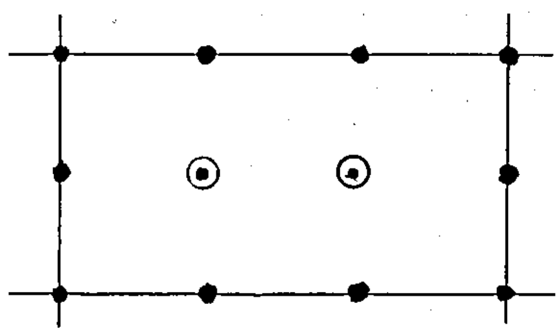

FIG. 1. Parcela de árvores com meia bordadura e duas plantas úteis.

Para $\rho > 2/3$, a solução é $k = 1$. Para $0 < \rho < 2/3$, a expressão (VI) sugere os valores mais próximos, números naturais, a serem tentados.

Com $\rho < 0$, temos:

$$\frac{dV}{dk} = (3\,\sigma^2/N)\,[-(2/k^2) + (1 + 2/k^2)\,\rho].$$

Com $\rho < 0$, esta derivada é negativa, qualquer que seja o número natural k. Nestas condições, para reduzir o valor de $V(\hat{m})$ devemos aumentar k o mais possível, mas de tal sorte que tenhamos um número razoável de graus de liberdade para o resíduo.

No caso de linha dupla de plantas úteis, temos:

$$K = 2\,(k + 4),$$

$$V(\hat{m}) = (2\,\sigma^2/N)\,[(1 + 4/k) + (k + 3 - 4/k)\,\rho],$$

onde $k > 2$ é um número natural par. Está claro que com $k = 2$ temos somente uma árvore em cada uma das duas filas úteis, o que equivale a ter uma única linha com duas árvores. Conclui-se, pois, que devemos ter, na verdade, $k > 2$.

Com $\rho > 0$, a variância $V(\hat{m})$ é mínima para

$$k = 2\sqrt{\frac{1-\rho}{\rho}} \qquad (\rho > 0).$$

Mas a condição de termos $k > 2$ importa em ter $\rho < 1/2$.

Com $\rho < 0$, a derivada

$$\frac{dV}{dk} = (2\,\sigma^2/N)\,[-(4/k^2) + (1 + 4/k^2)\,\rho]$$

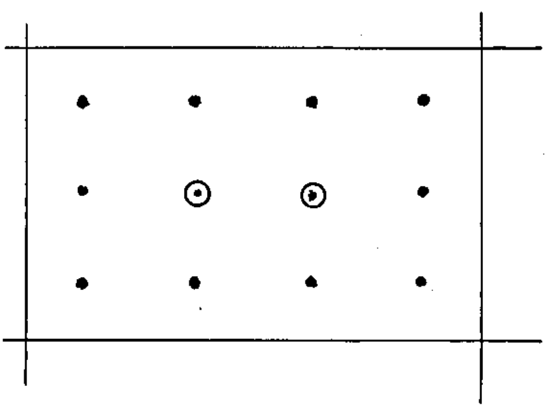

FIG. 2. Parcela de árvores com bordadura completa e duas plantas úteis.

é negativa, qualquer que seja o número natural par k. Logo, para reduzir o valor de $V(\hat{m})$ devemos aumentar k o mais possível, mas com a condição de ter um número razoável de graus de liberdade para o resíduo.

CONCLUSÕES

São as seguintes, relativamente ao número recomendado k de árvores úteis por parcela.

A) Parcelas sem bordadura

A1) $\rho > 0$ Recomenda-se $k = 1$.

A2) $\rho < 0$ k deve ser o maior possível, compatível com um número razoável de graus de liberdade para o resíduo.

B) Parcelas com meia bordadura

B1) Linha simples de plantas úteis

a) $\rho > 0{,}50$ Recomenda-se $k = 1$.

b) $0 < \rho < 0{,}50$ Solução dada pela expressão

$$k = \sqrt{\frac{1-\rho}{\rho}}$$

ou por um dos valores de k número natural mais próximos da valor da raiz.

c) $\rho < 0$ k deve ter o maior valor possível compatível com um número razoável de graus de liberdade para o resíduo.

B2) Linha dupla de plantas úteis

a) $\rho > 0{,}50$ Recomenda-se $k = 2$.

b) $0 < \rho < 0{,}50$ Solução dada pela expressão

$$k = \sqrt{\frac{2(1-\rho)}{\rho}},$$

com k número natural par, ou por um dos valores de k números naturais pares mais próximos do valor da raiz.

c) $\rho < 0$ k deve ser o maior número natural par possível, compatível com um número razoável de graus de liberdade para o resíduo.

C) Parcelas com bordadura completa

C1) Linha simples de plantas úteis

a) $\rho > 2/3$ Recomenda-se $k = 1$.

b) $0 < \rho < 2/3$ Solução dada pela expressão

$$k = \sqrt{\frac{2(1-\rho)}{\rho}},$$

com k número natural, ou por um dos valores de k números naturais mais próximos do valor da raiz.

c) $\rho \leq 0$ k deve ser o maior número natural possível, compatível com um número razoável de graus de liberdade para o resíduo.

C2) Linha dupla de plantas úteis

a) $\rho > 0{,}50$ Recomenda-se k = 2.

b) $0 < \rho < 0{,}50$ Solução dada pela expressão

$$k = 2\sqrt{\frac{1-\rho}{\rho}}$$

ou por um dos valores de k números naturais pares mais próximos do valor da raiz.

c) $\rho \leq 0$ k deve ser o maior número natural par possível, compatível com um número razoável de graus de liberdade para o resíduo.

Em resumo - As parcelas grandes só são aconselháveis no caso de valores negativos ou, se positivos, muito próximos de zero, para o coeficiente de correlação intraclasse (ρ). Mesmo com ρ = 0,050, por exemplo, a parcela recomendável, no caso de bordadura completa, teria k = 6 árvores úteis, em linha simples, e este número cairia para k = 1, no caso de parcelas sem bordadura, ou para k = 4, se se usar meia bordadura. Com $\rho > 0{,}20$, as recomendações seriam:

A) Para parcelas sem boradura: k = 1.

B1) Para parcelas com meia bordadura e linha simples de plantas úteis: k = 1 ou 2, conforme o valor de ρ.

B2) Para parcelas com meia bordadura e linha dupla de plantas úteis: k = 2 (na verdade seria uma linha simples).

C1) Para parcelas com bordadura completa e linha simples de plantas úteis: k = 1, 2 ou 3, conforme o valor de ρ.

C2) Para parcelas com bordadura completa e linha dupla de plantas úteis: k = 2 ou 4, conforme o valor de ρ.

Por outro lado, a experiência tem indicado, como norma, valores positivos para ρ, em geral acima de 0,20. E a estimação de ρ deve basear-se, de preferência, em dados de, pelo menos, dois anos de colheita, no caso de plantas perenes.

REFERÊNCIAS

BUENO, A. & PIMENTEL-GOMES, F. Estimativa do tamanho de parcela em experimentos de mandioca. R. bras. mandioca, 2(2):39-44, 1983.

COCHRAN, W.G. & COX, G.M. Experimental designs. New York, John Wiley, 1950.

CORDEIRO, C.M.T. & MIRANDA, J.E.C. Tamanho de parcela e eficiência experimental em batata doce usando a potência do teste F. Pesq. agropec. bras., Brasília, 18(7):707-13, jul. 1983.

FISHER, R.A. Statistical methods for research workers. Londres, Oliver and Boyd, 1938.

FRAGA JUNIOR, C.G. & CONAGIN, A. Delineamentos e análises de experimentos com cafeeiros. Bragantia, 15:177-91, 1956.

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

OLIVEIRA, E.B. Estudo comparativo de alguns métodos de estimação do tamanho adequado de parcelas experimentais. Brasília, EMBRAPA-DMQ, 1976. Tese Mestrado.

OLIVEIRA, E.B. & BIAVA, M.L. Bibliografia sobre tamanho e forma de parcelas experimentais. Brasília, EMBRAPA-DID, 1982.

PEARCE, S.C. Field experimentation with fruit trees and other perennial plants. East Malling, Commonwealth Agric. Bureaux, 1953.

PIMENTEL-GOMES, F.; VALSECHI, O.; ABREU, C.P. & OLIVEIRA, E.R. A amostragem da cana-de-açúcar para determinações tecnológicas. An. Esc. Sup. Agric. Luiz de Queiroz, 20:89-114, 1964.

RODRIGUES, M.S. Dicionário brasileiro de estatística. 2.ed. Rio de Janeiro, IBGE, 1970.

ROSSETTI, A.G. & PIMENTEL-GOMES, F. Determinação do tamanho ótimo de parcelas em ensaios agrícolas. Pesq. agropec. bras., Brasília, 18(5):477-87, maio 1983.

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