DETERMINAÇÃO DO TAMANHO ÓTIMO DE PARCELAS EM ENSAIOS AGRÍCOLAS

ADROALDO GUIMARÃES ROSSETTI

Orientador: Dr. Frederico Pimentel Gomes

Dissertação apresentada à Escola Superior de Agricultura "Luiz de Queiroz", da Universidade de São Paulo, para obtenção do título de Mestre em Estatística e Experimentação Agronômica.

P I R A C I C A B A
Estado de São Paulo - Brasil
Julho, 1979

À Rosália, minha esposa,

aos meus filhos Adroaldo Júnior,
Rosaldo e
Lialda Lúcia,

ao meu pai (em memória) e

a minha mãe,

D E D I C O

A G R A D E C I M E N T O S

Ao Dr. Frederico Pimentel Gomes, Coordenador do Curso de Pós Graduação em Estatística e Experimentação Agronômica da Escola Superior de Agricultura "Luiz de Queiroz" e orientador deste trabalho, pelo apoio e pela segura orientação, não só a este trabalho, mas de um modo geral, ao longo de todo o curso.

Ao Dr. Izaias Rangel Nogueira, Chefe do Departamento de Matemática e Estatística da ESALQ, pelos ensinamentos e valiosas sugestões apresentadas.

Ao Dr. Humberto de Campos, Professor Adjunto do Departamento de Matemática e Estatística da ESALQ, pelo apoio, atenção, ensinamentos e sugestões apresentadas.

Ao Corpo Docente do Departamento de Matemática e Estatística da ESALQ, pelos conhecimentos transmitidos, que concorreram eficientemente, para a elaboração deste trabalho.

À CAPES e ao CNPq, pelas bolsas concedidas para a realização do Curso.

Ao Dr. Arnaldo Felisberto Imbiriba da Rocha, Professor Titular do Instituto de Ciências Exatas da Universidade do Amazonas, pelo apoio e incentivo.

A todos que de algum modo concorreram para a realização do curso e deste trabalho.

Í N D I C E

1 - RESUMO

O estudo do tamanho ótimo de parcelas, na experimentação de campo, visa a minimizar os custos, em função do erro experimental.

As inúmeras pesquisas existentes na literatura indicam sua grande importância. A principal questão envolvida é a heterogeneidade do solo que influi, em grande escala, no estabelecimento do tamanho ideal da parcela.

Neste trabalho estima-se o coeficiente de regressão _b_ , ou índice de variabilidade do solo, proposto por FAIRFIELD SMITH (1938), 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. Tal estimativa,

$$\hat{\beta} = - \frac{\sum_i \sum_j w^{ij} Y_i (X_i - \bar{X})}{\sum_i \sum_j w^{ij} X_i (X_i - \bar{X})} ,$$

bem como as estimativas das variâncias observadas, dos diferentes tamanhos de parcelas, são ponderadas pelos pesos w^{ij}.

Desenvolve-se uma metodologia que visa a estimar eficientemente os pesos w^{ij} dos próprios dados. Tal procedimento é válido tanto para dados de ensaios de uniformidade, aqui estudados através de um esquema de classificação hierárquica, ou seja,

$$y_{ijkl} = m + r_i + b_{ij} + p_{ijk} + s_{ijkl} ,$$

como para dados de experimentos, tipo comparativo de variedades, onde haja efeitos de tratamentos, conduzido aqui através da análise de um experimento em delineamento de blocos ao acaso com parcelas subdivididas, ou seja

$$y_{ijkl} = m + r_i + t_j + \varepsilon_{1ij} + t'_k + (t\,t')_{jk} + \varepsilon_{2ijk} + e_{ijkl}$$

Demonstra-se que a estimativa $\hat{\beta}$, obtida por tal procedimento tem variância mínima assintótica.

O desvio de regressão,

$$V - \frac{U^2}{T} ,$$

é testado através do teste de Qui-quadrado com n - 2 graus de liberdade, uma vez que, como tal é distribuído.

A estimativa do tamanho ótimo de parcela é enfocado sob dois aspectos:

a - Em função do mínimo custo envolvido no experimento, em unidades básicas, através de

$$x = \frac{b\ K_1}{(1 - b)\ K_2}$$

b - E em função do número de repetições necessárias para se obter determinada diferença de médias,

$$d = \sqrt{2\ (t_1 + t_2)(C\ V)^2 / (r\ x^b)}\ ,$$

fixada "a priori", a um nível $\underline{\alpha}$ de probabilidade.

2 - INTRODUÇÃO

Na análise de variância de um experimento qualquer, é mais comum determinarem-se os efeitos devidos aos tratamentos e às repetições, do que os efeitos aparentemente inclusos no resíduo, e que são devidos a inúmeros fatores, muitos dos quais, incontroláveis pelo experimentador. Consequentemente a feitura da análise de variância deve estar condicionada a algumas hipóteses básicas, cuja observância proporciona validade aos testes de significância. Tais hipóteses são:

2.1 - Os diversos efeitos do modelo são aditivos;

2.2 - Os erros experimentais são independentes entre si;

2.3 - Os erros experimentais têm todos a mesma variância σ^2;

2.4 - Os erros experimentais têm distribuição normal.

CALZADA BENZA (1965), adverte que o pesquisador, para reduzir ao mínimo a variabilidade sobre os resultados de um experimento, deve atentar para os seguintes pontos importantes:

- Forma e colocação das parcelas em campo;
- Forma do bloco ou repetição;
- Tamanho das parcelas;
- Número de repetições;
- Delineamento experimental;
- Falhas de plantas nas parcelas;
- Efeitos de bordaduras entre parcelas;
- Forma de condução do experimento.

Muitos experimentadores usaram tamanhos e formas de parcelas inteiramente arbitrárias. Para CALZADA BENZA (1965), tratando-se de parcelas pequenas, a forma tem pouca ou quase nenhuma influência sobre o erro experimental, enquanto que em parcelas grandes, a influência é notável.

A questão da técnica de parcelas tem, durante muito tempo, atraído atenções de pesquisadores, conforme indicam numerosos trabalhos a respeito. Na literatura, as referências acerca de metodologia do tamanho de parcelas, em experimentos agrícolas, são extensas, sendo na grande maioria, estudadas através de ensaios de uniformidade. Várias tentativas visando a criar métodos estatísticos para determinar tamanho ideal de parcelas em ensaios agrícolas têm sido desenvolvidas, como in dicam os trabalhos a respeito. Diversos autores, com esse objetivo utilizaram, em diversas culturas, procedimentos, tais como: Método do Desvio Padrão , Método da Máxima Curvatura, Coeficiente de Variação, Método da Informação Relativa e outros fundamentaram suas pesquisas basicamente no empirismo, etc. A conclusão mais frequente dos pesquisadores foi que o coeficiente de variação decrescia, com o aumento do tamanho da parcela, havendo entretanto consideráveis diferenças que variam com o procedimento utilizado, com o tipo de solo, com a cultura em estudo, etc.

Até então, o método que se há revelado mais eficiente é a Lei Empírica introduzida por FAIRFIELD 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 ponto bastante curioso e que os inúmeros trabalhos acerca de metodologias do tamanho de parcelas, nem sempre consideram a correlação que deve existir entre as observações, fato que ocorre com grande frequência, quando suas variâncias são calculadas dos mesmos dados. Tendo em vista tal aspecto de grande importância nas pesquisas experimentais, e que 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 desiguais, através do qual, o tamanho ótimo da parcela será estimado levan do-se em conta a relação entre o tamanho da parcela e a variabilidade do solo, bem como, com base 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 $\underline{\beta}$, com variância mínima assintótica. Tal procedimento é aplicável tanto a dados de ensaio de uniformidade como a dados de outros ensaios, como foi sugerido por KOCH e RIGNEY (1951).

Visa-se, com isso, a fornecer um embasamento mais sólido a futuras pesquisas do gênero, visto que os trabalhos até então desenvolvidos, têm sido com culturas específicas ; nada se houve encontrado na literatura consultada acerca de um estudo teórico especificamente.

3 - REVISÃO DE LITERATURA

HARRIS (1915), tomando um relativamente grande número de ensaios de uniformidade, com diversos cultivos, como capim Naxenim (*Eleusine coracana*, Gaertn) , cultivado na Índia, beterraba, trigo, etc., em outros países, propôs o uso do coeficiente de correlação intraclasse da produção de áreas adjacentes, como um "coeficiente de heterogeneida". Tal método entretanto, segundo FAIRFIELD SMITH (1938), serve apenas para demonstrar que as fertilidades das parcelas adjacentes são correlacionadas.

DAY (1920) , JUSTENSEN (1932) , REYNOLDS *et alii* (1934) , WIEBE (1935) , BOSE (1935) , LOFSELL (1936) , KULKARNI e BOSE (1936), aplicaram o método do coeficiente de variação a dados de ensaios de uniformidade, como um indicador do tamanho e forma da parcela, para as culturas de trigo, bata -

tas, algodão, cevada, feijão branco e sorgo vassoura (*Andropogon sorghum*), respectivamente. A experiência comum destes pesquisadores foi que o coeficiente de variação decresceu proporcionalmente ao aumento do tamanho da parcela. Verificaram também que a economia do terreno utilizado decresceu com o aumento do tamanho da parcela. Por este método, o tamanho ótimo de parcela foi determinado graficamente, através da projeção do Coeficiente de Variação (CV) nas ordenadas, contra o tamanho da parcela nas abscissas, sendo considerado o tamanho ótimo, o ponto da curva onde o raio de curvatura e menor. Deste modo, com excessão apenas de WIEBE (1935), que determinou que parcelas quadradas são mais uniformes que parcelas compridas e estreitas, no sentido das linhas do experimento, os pesquisadores concluíram que parcelas compridas e estreitas são menos variáveis que parcelas curtas e largas, com a mesma área.

Sabe-se, no entanto, que os resultados obtidos com o uso deste método dependem largamente da escala de coordenadas, onde as observações são projetadas.

FAIRFIELD SMITH (1938), pesquisando o melhor tamanho de parcela para a cultura de trigo, através de ensaios de uniformidade, fez uma revisão crítica dos trabalhos existentes na época, sobre técnicas de parcelas, a partir do que introduziu uma lei empírica - "Lei da variância" que usualmente se aceita como uma das medidas mais úteis da variabilidade do solo. O autor mostrou, empiricamente, que o logarítmo da variância entre parcelas de um mesmo tamanho, está linearmente relacionado ao logarítmo do tamanho da parcela, ou seja:

$$V(x) = \frac{V_1}{x^b} \longrightarrow \log V(x) = \log V_1 - b \log x \qquad (3.1)$$

onde:

- x é o tamanho da parcela em unidades básicas,
- V_1 é a variância entre parcelas de tamanho correspondente a uma unidade,
- $V(x)$ é a variância do rendimento médio por unidade de área entre parcelas de área de $\underline{x}$ unidades,
- b é uma característica do solo, isto é, mede a variabilidade do solo, é uma medida de correlação entre unidades adjacentes:

$$0 \leq b \leq 1$$

Se $b = 0$ indica perfeita correlação e que o solo é bem uniforme.

Se $b = 1$ indica que não há correlação entre as unidades, sendo o solo experimental muito heterogêneo.

Em seu trabalho original, com a cultura de trigo, o autor computou dados de 38 ensaios de uniformidade, verificando que os valores de $\underline{b}$, variaram entre 0,2 e 0,8.

Em combinação com a fórmula empírica

$$V(x) = V_1 / x^b ,$$

o autor usou uma função de custo, que minimizada, dá o tama -

nho ótimo da parcela. Esta função de custo por parcela é dada por

$$C(x) = K_1 + K_2 x \quad ,$$

onde:

x é o tamanho da parcela, em unidades básicas,

K_1 é a parte do custo que está associado somente ao número de parcelas e

K_2 é a parte do custo por unidade de área.

Então o custo por unidade de informação será mínimo se se tiver,

$$x = \frac{K_1 \; b}{(1 - b) \; K_2} \quad . \qquad (3.2)$$

O autor recomenda que os efeitos da estimação do coeficiente de regressão <u>b</u> e as variâncias dos diferentes tamanhos de parcelas devem ser ponderados por seus respectivos números de graus de liberdade, ajustando uma regressão pelo método dos mínimos quadrados.

ROBINSON, RIGNEY e HARVEY (1948), utilizaram o método de FAIRFIELD SMITH (1938), em dados de um ensaio de uniformidade com a cultura de amendoim, a fim de estimar o melhor tamanho de parcela. Estes autores concluíram que 30% do custo total foi proporcional à área utilizada. Determinaram 70 / 30 para a razão K_1 / K_2 , na equação (3.2) e calcularam o tamanho ótimo da parcela, <u>x</u> , partindo dessa equação, em função das unidades básicas, no ensaio de uniformidade.

KELLER (1949), propôs o uso do método da informação relativa para estimar o tamanho e forma de parcela mais efi - cientes, para lúpulo (*Humulus lupulus*, L.), com dados de um ensaio de uniformidade. A variância entre parcelas foi compu tada para cada tipo de parcela proposto e logo dividida pelo número de unidades básicas que formavam a parcela, para obter uma variância que seria comparável com a variância da unidade básica. A variância entre as unidades básicas, foi atribuída como fornecendo 100% de informação relativa. Dividida esta variância pela variância comparável de cada tipo de parcela , obteve a porcentagem de informação relativa, correspondente a cada tipo de parcela. Concluiu então que, em lúpulo, de modo geral, a variância comparável aumentou e a informação relati- va diminuiu a medida que aumentou o tamanho da parcela.

KOCH e RIGNEY (1951), aplicaram a lei empírica de FAIRFIELD SMITH (1938), a dados de 15 experimentos de tabaco e 10 de algodão, para os quais foram determinados os valores de b , obtendo como valores médios para cada grupo de ensaios respectivamente 0,55 e 0,49 , a fim de determinar o tamanho ó timo de parcela. Os experimentos foram delineados em blocos incompletos com parcelas subdivididas, onde há efeitos de tra tamentos. Concluíram então, que o coeficiente de regressão b , do logarítmo da variância, sobre o logarítmo do tamanho da parcela, poderia ser estimado de ensaios experimentais, em que os efeitos de tratamentos se façam presentes, do mesmo modo como de dados de ensaios de uniformidade. O valor de b foi estimado pela relação

$$\hat{b} = \frac{\sum_i (x_i' - \bar{x}')\, y}{\sum_i (x_i' - \bar{x}')^2}$$

onde:

$$y = \log (V/x) \quad ; \quad x_i' = \log x_i \quad e \quad \bar{x}' = \frac{\sum_i x_i'}{n}$$

Os autores utilizaram os componentes de variância para estimar a regressão do logarítmo da variância da parcela, sobre o logarítmo do tamanho da parcela. Mencionam ainda que o ajustamento, pelo número de graus de liberdade é inviável quando se trabalha com dados experimentais, visto que as variâncias são constituídas de diferentes componentes estimados. Mesmo assim, um ajustamento imponderado, através do método dos mínimos quadrados seria razoável.

O tamanho ótimo de parcela foi estimado através da fórmula (3.2), com base nos valores de b citados anteriormente.

AMARAL (1951), estudando o melhor tamanho de parcela na experimentação com o cafeeiro, com dados de um experimento de competição de variedades, através do método de FAIRFIELD SMITH (1938), determinou que o comprimento e a largura das parcelas são fatores que agem independentemente sobre a variância. Este autor, usando parcelas do tipo: a , b , c, d , e , f , g (parcelas alongadas) e parcelas do tipo: a', b' , c' , d' , e' , f' , g' , (parcelas de mesma dimensão, mas no sentido perpendicular às primeiras) e representando graficamente, com log V(r) nas ordenadas e log V(x) nas abscissas, obteve dois quadriláteros que são praticamente paralelogramos de lados homólogos paralelos. Esse paralelismo ofereceu a indicação de que comprimento e largura das parcelas são fatores que agem independentemente sobre a variância. Determinou o valor do coeficiente de regressão b, na equação (3.1) usando o método dos mínimos quadrados, tendo como peso os respectivos números de graus de liberdade. Trabalhou com a variância reduzida e determinou que

$$\log V(r) = \log V + B \log x \qquad (3.3)$$

onde:

$B = 1 - b$, possibilitando calcular uma equação do tipo:

$$\log V(r) = \log V + b_1 \log x_1 + b_2 \log x_2 , \qquad (3.4)$$

onde:

x_1 representa o comprimento da parcela (número de parcelas unitárias no sentido do comprimento) e

x_2 é a largura da parcela (número de parcelas unitárias no sentido da largura).

Para o ajustamento, usou como uma regressão múltipla, satisfazendo o princípio exigido pelo método dos mínimos quadrados. WEBER e HORNER (1957), para estudar a relação entre o tamanho da parcela e sua variabilidade, usaram o método de FAIRFIELD SMITH (1938), em dados de um ensaio de uniformidade com a cultura de soja, observaram que o coeficiente de variação decrescia, a medida que aumentava o tamanho da parcela. Entretanto, parcelas longas foram mais variáveis do que parcelas curtas e largas. Para estes autores, esse resultado pouco comum, foi devido a orientação das parcelas, com respeito ao gradiente de fertilidade.

HATHEWAY e WILLIAMS (1958), utilizando dados de ensaios de uniformidade com soja, descreveram um método de ponderação das variâncias observadas de diferentes tamanhos de parcelas que conduzem a um estimador não viciado de $\underline{b}$, com variância mínima assintótica. Os autores conseguiram tal minimização, ponderando os diferentes termos nas somas de quadrados e produtos, que definem o coeficiente de regressão.

O procedimento descrito é aplicável a ensaios de uniformidade e a dados de outros ensaios. Neste último caso , a análise de variância é construída de modo a simular haver sido derivada diretamente de resultados de ensaios de uniformidade, segundo sugerem KOCH e RIGNEY (1951).

Vários autores consideram que o procedimento proposto por estes pesquisadores, conduz a uma estimativa mais eficiente do parâmetro $\underline{b}$, desde que obedeça as considerações teóricas sugeridas pelas observações.

BRIM e MASON (1959), pesquisando o melhor tamanho de parcela para a cultura da soja, usam a técnica introduzida por KOCH e RIGNEY (1951), em dados de experimentos de competi ção de variedades. Os ensaios foram delineados de modo a obter vários tamanhos de parcela, os quais simulavam um ensaio de uniformidade. Foram computadas parcelas de 4,88 x 0,91 m e 2,44 x 1,83 m. Dentro dos limites dos ensaios a variabilidade das parcelas compridas foi a mesma das parcelas curtas. Os pesquisadores obtiveram um coeficiente de regressão não ponderado $\hat{b}$, através da equação

$$\hat{b} = \frac{\sum_i (x'_i - \bar{x}') \, y}{\sum_i (x'_i - \bar{x}')^2}$$

onde: $x'_i = \log x_i$, $\bar{x}' = \sum_i x'_i / n$ e $y = \log (V/x)$.

A aplicação desta fórmula, no entanto, pode apresentar, algumas vezes estimadores de b maiores que 1 , o que não é razoável do ponto de vista do parâmetro b . A estimativa $\hat{b}$ é nesse caso, tendenciosa.

Determinaram que os custos K_1 e K_2 foram aproximadamente 70% e 30% do custo total, respectivamente.

HATHEWAY (1961), estudando a relação entre tamanho de parcela e variabilidade do solo, a fim de determinar o tamanho ótimo da parcela, usando a fórmula (3.2), através de dados de um ensaio de uniformidade com batatas, associou o índice b , de heterogeneidade do solo, à fórmula de COCHRAN e COX (1957), para número de repetições e estabeleceu uma fórmula para o cálculo do tamanho conveniente da parcela, independente do custo. A fórmula proposta pelo autor permite estabelecer o tamanho ótimo da parcela para uma determinada diferença de médias que o experimentador deseja detectar ou conhece "a priori". Esta fórmula é dada por

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

RAMPTON e PETERSEN (1962), estudando a eficiência relativa de tamanhos de parcelas e número de repetições, como indicador de produção de capim pé-de-galinha (*Dactylis glomerata*, L.), conduziu um ensaio de uniformidade, a cujos dados aplicou o procedimento de FAIRFIELD SMITH (1938), concluindo que parcelas maiores ou maior número de repetições de uma parcela de certo tamanho, permitiria a comprovação de menor diferença, do que menor número de repetições ou parcelas menores. Concluíram portanto, que a forma da parcela tem pouco efeito na variância, que foi mais precisamente estimada por redução ao mínimo, o tamanho da parcela, quando foi aumentado o número de repetições. Em tais condições, o tamanho ótimo da parcela decresceu proporcionalmente ao aumento do custo da terra.

MARANI (1963), baseado no procedimento original de FAIRFIELD SMITH (1938), conduziu um ensaio de uniformidade com a cultura de milho, a fim de estimar o tamanho ótimo de parcela. Em seu trabalho o autor afirma que K_1 e K_2 são constantes para diferentes tamanhos de parcelas, mas K_1, está associado ao número de unidades básicas por parcela e K_2 está associado à unidade de área básica. As porcentagens dos dois tipos de custos em relação ao custo total, são porconseguinte proporcionais a K_1 e K_2x , respectivamente. Assim a razão determinada pelos autores, na equação (3.2) foi realmente $K_1 / K_2 x$ e não K_1 / K_2 . Esta razão depende portanto do tamanho $\underline{x}$, da parcela.

TORRIE, SCHMIDT e TENPAS (1963), estudando o tamanho ótimo, forma de parcela e número de repetições, para medir a produção de alfafa (*Medicago sativa,* L.), aos dados de dois ensaios de uniformidade, através do procedimento de FAIRFIELD SMITH (1938), determinaram as estimativas do número de repetições para detectar diferenças de 25% da média, ao nível de 5% de probabilidade com 80% de segurança.

Com tal procedimento os autores concluíram que os dados indicaram que o número de repetições exigido para conseguir uma grande precisão decresceu com aumento do tamanho da parcela, mas a área total exigida, por tratamento, aumentou. Concluíram ainda, que as parcelas mais compridas, tiveram redução mais efetiva na variação que as parcelas largas.

Para detectar uma diferença entre tratamentos de um dado tamanho, poucas repetições, mas uma grande área por tratamento, foi exigida, a medida que crescia o tamanho da parcela. O mais eficiente tamanho de parcela, usando o método do coeficiente de variação foi o que teve como unidade básica 1,37 x 3,05 m , quando o custo proporcional ao número de parcelas por tratamento (K_1), e o custo proporcional à área total

(K_2), foram 70% e 30% , respectivamente. O tamanho ótimo de parcela, para $б = 0{,}35$, foi estimado ser de 1,26 unidades básicas.

ARROYO e CHÁVEZ (1966), pesquisando o melhor tamanho de parcela, através de um ensaio de uniformidade com batata, a cujos dados aplicaram a fórmula de FAIRFIELD SMITH (1938), e os processos introduzidos por KOCH e RIGNEY (1951), e HATHEWAY e WILLIAMS (1958), determinaram um tamanho ótimo de parcela, em unidades básicas, de 3,53 m^2, em função de um coeficiente б de heterogeneidade igual a 0,38769 , portanto entre 0,3 e 0,7. Deste modo, com base no que salienta FEDERER (1955), de que se $0{,}3 < б < 0{,}7$, pode-se tomar o dobro ou a metade da área ótima encontrada, sem afetar os resultados, os autores recomendam que se tomem, para este cultivo, parcelas entre 1,77 m^2 e 7,06 m^2 .

HERNANDEZ e VERGARA (1968), pesquisando o melhor tamanho de parcela para a cultura de arroz, com dados de um experimento de variedades, através da lei empírica de FAIRFIELD SMITH (1938) e da técnica de KOCH e RIGNEY (1951), concluíram que, comparando esta, com pesquisas anteriores, é conveniente utilizar parcelas de 4 m de largura, o número de 6 plantas, para colher as fileiras centrais ou parcelas de duas fileiras de 4 m de largura para a colheita total, em caso de não existir efeito de bordaduras. Salientam que suas conclusões apresentam fortes razões para sustentar que as áreas uteis que vinham sendo usadas em ensaios de adubação, de 25 m^2, poderiam ser reduzidas a 4 x 3 m, em função do estimador do parâmetro b , que indica alta associação entre as unidades do solo, ao menos para as condições experimentais em que se realizou o ensaio de uniformidade. Com um valor de $\hat{b}$ = 0,30854, o tamanho ideal de parcela, em função dos custos K_1 e K_2 foi de 1,97 m^2 . Como $0,3 < \hat{b} < 0,7$, os autores, baseados em FEDERER (1955), recomendam que o tamanho ótimo de parcela pode ser tomado entre 0,99 m^2 e 3,9 m^2 , sem perda de eficiência.

SILVA (1972), usando dados de um ensaio de uniformidade com a cultura de soja, para estudar o melhor tamanho e forma de parcela, através da lei empírica de FAIRFIELD SMITH (1938), do mesmo modo que AMARAL (1951), projetou, com log V(r) nas ordenadas e log x nas abscissas, os tamanhos de parcela, encontrando um paralelismo, que não refletiu a mesma nitidez, encontrada por este autor. Usando a variância reduzida, incluiu o parâmetro b_3 , responsável pela interação x_1 x_2 (comprimento x largura), ficando agora a equaçao como

$$\log V(r) = \log V + b_1 \log x_1 + b_2 \log x_2 + b_2 \log x_1 + b_3 \log x_1 \cdot \log x_2$$

que além de x_1 (comprimento) , x_2 (largura) , estudou a interação comprimento x largura, indicada por $x_1 . x_2$, que representou a independência entre o comprimento e a largura da parcela.

A análise de variância revelou que a interação comprimento x largura não foi estatisticamente significativa, donde concluiu que β_3 não influiu na regressão. Consequentemente, o comprimento e largura das parcelas são fatores que agem independentemente sobre a variância, portanto, a mesma conclusão de AMARAL (1951). O custo por parcela foi estudado em função do seu comprimento e largura, através da equação

$$T_{x_1 x_2} = K_1 + K_2 \, x_1 \cdot x_2 + K_3 \, x_2 / x_1$$

donde

$$x_1 = \sqrt{\frac{K_3 \, (\beta_2 + \beta_1)}{K_2 \, (\beta_2 - \beta_1)}} \quad ; \quad \text{para} \quad \beta_1 < \beta_2$$

$$x_2 = \frac{K_1 \, (\beta_2 - \beta_1)}{2 \, K_3 \, (1 - \beta_2)} \sqrt{\frac{K_3 \, (\beta_2 + \beta_1)}{K_2 \, (\beta_2 - \beta_1)}} \; ;$$

para $\quad 0 < \beta_1 < \beta_2 < 1$

$$x_1 \cdot x_2 = \frac{K_1 \, (\beta_2 + \beta_1)}{2 \, K_2 \, (1 - \beta_2)} \quad .$$

CAMPOS (1972), usando três ensaios de uniformidade com a cultura do girassol (*Helianthus annuus*, L.), a cujos dados usou diversos procedimentos para estimar o melhor tamanho de parcela, chegando às conclusões que se seguem.

O método da máxima curvatura, forneceu um tamanho ótimo de parcela, em média, para os três ensaios, de 5,5 unidades experimentais ou 5,5 m^2 .

O método da informação relativa, apenas através da utilização das variâncias comparáveis das parcelas, deu um tamanho ótimo de 2,7 unidades experimentais ou 2,7 m^2 de área , em média para os três ensaios.

O método de FAIRFIELD SMITH (1938), cujo valor de b foi em media, para os três ensaios, de 0,5919 , deu tamanho ótimo, também em média para os três ensaios, de 5 unidades experimentais ou 5,0 m^2 de área.

Sabe-se, no entanto que o método da máxima curvatura não leva em conta o custo e varia com a escala de coordenadas, do mesmo modo que o método da informação relativa.

IGUE, SOUZA e NAGAI (1972), utilizando a fórmula de HATHEWAY (1961), para a determinação do tamanho mais conveniente da parcela, para ensaios agrícolas, usaram dados de um ensaio com a cultura de arroz, em condições de sequeiro e de irrigação, através de um delineamento de blocos ao acaso, com 10 tratamentos e 2 , 3 , 4 , 5 ou 6 repetições. Foram usados três experimentos, sendo um em condição de sequeiro e dois de irrigação, conduzidos em quadrado latino, instalados no Centro Experimental de Campinas, SP.

Para o ensaio em condições de sequeiro, o índice de heterogeneidade do solo foi de 0,28966 e nos em condições de irrigação, estes foram de 0,38854 e 0,43878, cuja média foi de 0,41366, conduzindo à conclusão de que o solo usado tem pouca heterogeneidade.

Concluíram os autores que para valores fixos de r e d, o tamanho mais conveniente de parcela teve aumento proporcional ao aumento do coeficiente de variação; para um CV fixo o tamanho ótimo de parcela aumentou na mesma proporção que aumentou o número de repetições; para CV e r fixos, o tamanho ótimo da parcela diminuiu a medida que aumentaram as diferenças porcentuais d, entre dois tratamentos a serem comparados estatisticamente.

IGUE e MASCARENHAS (1974), utilizaram os dados de três ensaios de uniformidade, da cultura de soja, da variedade Pelicano, instalados nas Estações Experimentais de Pindorama, Lins, Marília e Ribeirão Preto, SP.

Os autores empregaram o método de FAIRFIELD SMITH (1938), para estimar o tamanho ótimo da parcela e, como não avaliaram as constantes K_1 e K_2, tomaram, por serem mais frequentes na literatura, os seguintes pares de valores: 50 - 50, 60 - 40, 70 - 30 e 80 - 20. Os valores de b encontrados para os três ensaios foram de 0,70235 ; 1,02959, e 0,52908, cuja média, para os três ensaios foi de 0,75367.

Para $K_2 = 40\%$, a parcela ótima foi de 3,54 unidades básicas, que corresponde a uma área de 1,24 m^2, para Pindorama; de 1,68 unidades básicas ou 0,59 m^2 para Ribeirão Preto e para as médias dos ensaios de Pindorama o tamanho ótimo foi de 9,69 unidades básicas ou 5,23 m^2 de área. Em média, o tamanho ótimo da parcela foi de 4,59 unidades básicas ou 1,60 m^2 de área, podendo serem usadas parcelas de 0,40 m^2 a 6,40 m^2, sem perda de eficiência.

4 - METODOLOGIA

Em aplicações usuais da análise de regressão, admite-se frequentemente 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 simples e direta aplicação do método dos mínimos quadrados, para estimar o coeficiente de regressão é eficiente e a quantidade a ser minimizada é a soma dos quadrados dos desvios. Assim, se X e Y são respectivamente a variável independente e dependente, no modelo estatístico

$$Y_i = a + b X_i + e_i ,$$

o sistema de equações normais é então

$$X' X \hat{\beta} = X' Y ,$$

que conduz ao conhecido estimador

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

cujas estimativas dos parâmetros são respectivamente

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

e

$$\hat{b} = \frac{\sum_i Y_i (X_i - \bar{X})}{\sum_i (X_i - \bar{X})^2} .$$

Um caso de ocorrência menos frequente é no qual os valores de Y são correlacionados e têm variâncias desiguais. Embora o método simples e direto dos mínimos quadrados ainda dê, neste caso, estimadores imparciais para a e b , este método é, em geral ineficiente, pois as estimativas $\hat{a}$ e $\hat{b}$, não serão tão exatas quando obtidas dos dados. Mais propícia é a aplicação do método dos mínimos quadrados 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 são apenas estimadas dos dados, seu uso permite que sejam determinadas estimativas com aproximadamente mínima variância.

Na estimação do tamanho ótimo de parcelas, o coeficiente de regressão b tem particular importância, pois nesse contexto ele representa um índice de heterogeneidade do solo, uma medida de correlação entre unidades adjacentes, constituindo-se deste modo, de grande interesse para esse propósito.

4.1 - ESTIMAÇÃO DE $\underline{b}$ BASEADA EM DADOS DE ENSAIOS DE UNIFORMIDADE

Segundo ilustram ROBINSON, RIGNEY e HARVEY (1948), ao avaliar um delineamento em blocos ao acaso com parcelas subdivididas ou com reticulados quadrados, de dados de ensaios de uniformidade, não é necessário, na verdade, atribuir uma variedade ou tratamento a cada parcela. Em lugar disso, para um reticulado k x k, a área toda é dividida em tantas repetições quantas possíveis, cada uma com $\underline{k}$ blocos de $\underline{k}$ parcelas, que pode ser analisado segundo um esquema de classificação hierárquica, que se adapta ao modelo

$$y_{ijkl} = m + r_i + b_{ij} + p_{ijk} + s_{ijkl} \qquad (4.1.1)$$

onde:

m é o efeito da média geral,

r_i é a i-ésima repetição; $i = 1, 2, \ldots, I$

b_{ij} é o j-ésimo bloco dentro da repetição $\underline{i}$; $j = 1, 2, \ldots, J$,

p_{ijk} é a k-ésima parcela dentro do bloco $\underline{j}$; $k = 1, 2, \ldots, K$,

s_{ijkl} é a l-ésima subparcela dentro da parcela k; $l = 1, 2, \ldots, L$.

Pela aplicação direta do método dos mínimos quadrados, à equação (4.1.1) obtém-se as somas de quadrados e os outros componentes do quadro (4.1).

QUADRO 4.1 - Análise de variância

Causa de Variação	G.L.	Q.M.	E (Q.M.)
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	$\sigma^2 + L\,\sigma_p^2 + KL\,\sigma^2_B$
Parcelas d. blocos	$I\,J(K - 1)$	V_3	$\sigma^2 + L\,\sigma_p^2$
Subparcelas d. parcelas	$I\,J\,K\,(L - 1)$	V_4	σ^2

4.1.1 - Formação das parcelas básicas

Forme-se, com base no experimento, as parcelas básicas como se segue.

Considere-se, inicialmente, cada repetição como uma parcela grande.

Chamando-se V_i' a variância de parcelas de diversos tamanhos, reduzidas a uma parcela base ; tem-se, neste caso, que V_i' coincide com o quadrado médio de repetições, como aparece na análise de variância do quadro (4.1). Assim

$$V_i' = V_1 \qquad (4.1.1.1)$$

Considerando-se agora os blocos como o próximo menor tamanho de parcela, vê-se que a variância entre blocos contém, além da variância devido a blocos dentro de repetições, aquela que é eliminada pela extratificação de grupos de blocos dentro de repetições, na análise de variância. Assim,

$$S\,Q\,B = \frac{\sum_{i,j} B_{ij}^2}{K\,L} - C \quad ,$$

que escrevendo em função de V_1 e V_2 , vem

$$E\,(S\,Q\,B) = I\,(J - 1)\,V_2 + (I - 1)\,V_1 \;.$$

Mas existem, nas I repetições, I J blocos, portanto I J - 1 graus de liberdade, para estimar sua respectiva variância. Assim, a variância entre blocos será:

$$V_2' = \frac{I\,(J - 1)\,V_2 + (I - 1)\,V_1}{I\,J - 1} \qquad (4.1.1.2)$$

Analogamente, considerando-se as parcelas como o próximo menor tamanho de parcela, tem-se,

$$S\,Q\,P = \frac{\sum_{ijk} P_{ijk}^2}{L} - C \quad ,$$

ou, expressa em função de V_1 , V_2 e V_3

$$E\,(S\,Q\,P) = I\,J\,(K - 1)\,V_3 + I\,(J - 1)\,V_2 + (I - 1)\,V_1$$

Mas como existem I J K parcelas, portanto I J K - 1 graus de liberdade, a variância entre parcelas sobre toda a área será

$$V_3' = \frac{I\,J\,(K - 1)\,V_3 + I\,(J - 1)\,V_2 + (I - 1)\,V_1}{I\,J\,K - 1} \qquad (4.1.1.3)$$

Do mesmo modo, tomando-se as subparcelas, tem-se

$$S\,Q\,S = \sum_{ijkl} Y_{ijkl}^2 - C \quad ,$$

que escrita em função de V_1 , V_2 , V_3 e V_4 fica

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

Neste caso, existem para as I J K L subparcelas I J K L - 1 graus de liberdade para estimar sua variância, logo

$$V'_4 = \frac{IJK\ (L-1)\ V_4 + IJ\ (K-1)\ V_3 + I\ (J-1)\ V_2 + (I-1)\ V_1}{IJKL-1} \qquad (4.1.1.4)$$

4.1.2 - Determinação do coeficiente de regressão b

FAIRFIELD SMITH (1938), definiu implicitamente, através de sua "lei de variância" o coeficiente de regressão b pela relação

$$V(x) = \frac{V_1}{x^b} \quad , \qquad (4.1.2.1)$$

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), colocando assim sobre uma unidade base. Deste modo tem-se que

$$V\ (x) = \frac{V'}{x}$$

e b é o coeficiente de regressão, uma característica do solo, uma medida de correlação entre unidades adjacentes.

Quando os valores, na equação (4.1.2.1) são transformados para a escala logarítmica, resulta, na relação cuja forma linear será

$$\log V\ (x) = \log V_1 - b \log x \qquad (4.1.2.2)$$

Seja

$\log V\ (x) = Y_i$, $\log V_1 = a$ e $\log x = X_i$,

então

$$Y_i = a - b\, X_i + e_i$$

onde

$$- b = B ,$$

logo

$$Y_i = a + B\, X_i + e_i \qquad (4.1.2.3)$$

onde os valores de Y_i são correlacionados e têm variâncias desiguais.

Assim, para os n valores observados, a forma matricial será:

$$Y = X \beta + \varepsilon \qquad (4.1.2.4)$$

donde

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

e

$$V(Y) = \begin{bmatrix} V(Y_1) & Cov(Y_1Y_2) & \dots\ Cov(Y_1Y_n) \\ Cov(Y_1Y_2) & V(Y_2) & \dots\ Cov(Y_2Y_n) \\ \dots\ \dots\dots & \dots\ \dots\dots & \dots\ \dots\ \dots\dots \\ Cov(Y_1Y_n) & Cov(Y_2Y_n) & \dots\ V(Y_n) \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1n} \\ w_{12} & w_{22} & \dots & w_{2n} \\ \dots & \dots & \dots & \dots \\ w_{1n} & w_{2n} & \dots & w_{nn} \end{bmatrix}$$

então

$$V(Y) = W , \qquad (4.1.2.5)$$

portanto W é a matriz de informação dos Y_i , onde cada $w_{ij} = c\,\sigma^2$. A matriz W é simétrica e positiva definida, c é uma constante e σ^2 é a variância da parcela base. Assim,

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

Para minimizar a soma dos quadrados dos erros na equação (4.1.2.4), através do método dos mínimos quadrados generalizados, faça-se, como RAO (1973), as transformações:

Seja a matriz D tal que

$$W = D\,D\ , \qquad (4.1.2.6)$$

onde D é simétrica e não-singular.

Pré-multiplicando-se e pós-multiplicando-se W simultaneamente, por D^{-1}, obtém-se

$$D^{-1}\,W\,D^{-1} = I \qquad (4.1.2.7)$$

Considere-se uma nova variável Z , tal que

$$Z = D^{-1}\,Y \qquad (4.1.2.8)$$

então, substituindo-se (4.1.2.4) em (4.1.2.8) ,

$$Z = D^{-1}\,X\beta + D^{-1}\,\varepsilon \qquad (4.1.2.9)$$

Sejam agora,

$$D^{-1}X = M \qquad (4.1.2.10)$$

$$D^{-1}\varepsilon = \xi \qquad (4.1.2.11)$$

Assim (4.1.2.9) toma a forma

$$Z = M\,\beta + \xi \quad , \qquad (4.1.2.12)$$

logo

$$E\,(Z) = M\,\beta \qquad e \qquad V\,(Z) = I \quad ,$$

ou seja

$$Z \cap N\,(M\beta\,,\,I)\ .$$

A minimização da soma dos quadrados dos erros $\underline{\xi}$, em (4.1. 2.12), através do método simples e direto dos mínimos quadrados, conduz ao sistema de equações normais

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

e ao estimador dos parâmetros

$$\hat{\beta} = (M'\,M)^{-1}\,M'\,Z \qquad (4.1.2.13)$$

Substituindo-se em (4.1.2.13) as expressões (4.1.2.8) e (4.1.2.10), resulta,

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

Substituindo-se agora (4.1.2.6), tem-se

$$\hat{\beta} = (X'\,W^{-1}\,X)^{-1}\,X'\,W^{-1}\,Y\ , \qquad (4.1.2.14)$$

donde se obtém as estimativas

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

$$\hat{b} = -\frac{\sum_i \sum_j w^{ij}\,Y_i\,(X_i - \bar{X})}{\sum_i \sum_j w^{ij}\,X_i\,(X_i - \bar{X})} \qquad (4.1.2.15)$$

onde:

w^{ij} são as entradas da matriz de informação inversa das Y_i ; são os pesos pelos quais são ponderadas a estimavita do coeficiente de regressão e as estimativas das variâncias dos diferentes tamanhos de parcelas. $\bar{X}$ e $\bar{Y}$ são médias ponderadas, ou seja

$$\bar{X} = \frac{\sum_i \sum_j w^{ij} X_i}{\sum_i \sum_j w^{ij}} \qquad e \qquad \bar{Y} = \frac{\sum_i \sum_j w^{ij} Y_i}{\sum_i \sum_j w^{ij}}$$

A variância de $\underline{\beta}$ será determinada como se segue.

Substituindo-se (4.1.2.4) em (4.1.2.14)

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

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

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

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

$$V(\hat{\beta}) = (X' W^{-1} X)^{-1} X' W^{-1} E(\varepsilon \varepsilon') W^{-1} X (X' W^{-1} X)^{-1}$$

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

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

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

$$V(\hat{\beta}) = (X' W^{-1} X)^{-1}, \qquad (4.1.2.16)$$

donde se obtém as variâncias dos parâmetros. Assim,

$$V(b) = \frac{1}{\sum_i \sum_j w^{ij} X_i (X_i - \bar{X})}$$

ou

$$V(b) = 1 / T \qquad (4.1.2.17)$$

Os pesos w^{ij} terão de ser estimados dos dados. Mesmo assim a estimativa $\underline{\hat{\beta}}$ será de mínima variância assintótica, como se demonstra a seguir.

Deste modo, segundo RAO (1973), para que $V(\hat{\beta})$ seja mínima, deve existir um outro estimador imparcial $\tilde{\beta}$, como função linear de $\underline{Y}$, tal que

$$V(\hat{\beta}) \leq V(\tilde{\beta})$$

Considere-se então o estimador

$$\tilde{\beta} = \hat{\beta} + A\,Y \quad ,$$

onde A é fixo. Assim, pela substituição de (4.1.2.14),

$$\tilde{\beta} = (X'\,W^{-1}\,X)^{-1}\,X'\,W^{-1}\,Y + A\,Y \quad ,$$

substituindo-se a expressão (4.1.2.4), têm-se

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

$$\tilde{\beta} - \beta = (X'\,W^{-1}\,X)^{-1}\,X'\,W^{-1}\,\varepsilon + A\,X\,\beta + A\,\varepsilon \qquad (4.1.2.18)$$

$$E(\tilde{\beta}) = \beta + \phi + A\,X\,\beta + \phi \longrightarrow E(\tilde{\beta}) = \beta + A\,X\,\beta \quad .$$

Portanto $\tilde{\beta}$ será imparcial se para todo

$$\beta \in R^k \quad , \quad A\,X = \phi \quad .$$

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

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

através de (4.1.2.18), tem-se que

$$V(\tilde{\beta}) = E\,\{[(X'W^{-1}X)^{-1}\,X'W^{-1}\varepsilon + A\varepsilon\,][\,\varepsilon'W^{-1}\,X(X'W^{-1}X)^{-1} + \varepsilon'\,A'\,]\}$$

$$V(\tilde{\beta}) = V(\hat{\beta}) + A\,W\,A' \longrightarrow V(\hat{\beta}) = V(\tilde{\beta}) - A\,A'\,W$$

portanto $\quad V(\hat{\beta}) \leq V(\tilde{\beta})$,

e a igualdade ocorrerá se e somente se

$$A' A = \phi \longrightarrow A = \phi$$

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

Quando, como é frequentemente 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 deve-se determinar a soma de quadrados dos desvios de regressão (SQDR). Assim, minimizando-se a soma dos quadrados dos erros em (4.1.2.12), através do método dos mínimos quadrados, como ANDERSON (1952), obtém-se que

$$R = SQDR = Z' Z - \hat{\beta}' M' Z \qquad (4.1.2.19)$$

Substituindo-se (4.1.2.8) e (4.1.2.10) ,

$$SQDR = Y' W^{-1} Y - \hat{\beta}' X' W^{-1} Y , \qquad (4.1.2.20)$$

ou ainda

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

onde

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

portanto

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

com (n - 1) graus de liberdade, onde $\underline{n}$ 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 de $\underline{\hat{b}}$, por

$$S\,Q\,Reg. = \frac{[\sum_i \sum_j w^{ij} \; Y_i \; (X_i - \bar{X})]^2}{\sum_i \sum_j w^{ij} \; X_i \; (X_i - \bar{X})}$$

ou seja

$$S\,Q\,Reg. = \frac{U^2}{T} \quad ,$$

portanto, (4.1.2.20) é escrita agora como

$$S\,Q\,D\,R = V - \frac{U^2}{T} \tag{4.1.2.21}$$

que tem distribuição de Qui-quadrado com (n - 2) graus de liberdade, conforme se demonstra a seguir. Desta forma, o desvio de regressão, como tal pode ser testado.

Como $Z \cap N\ (M\ \beta\ ,\ I)$, pela substituição de (4.1.2.13) em (4.1.2.19), resulta

$$R = Z'\ Z - Z'\ M\ (M'\ M)^{-1}\ M'\ Z \quad ,$$

$$R = Z'\ [\ I - M\ (M'\ M)^{-1}\ M'\ ]\ Z \quad ,$$

considere-se $I - M\ (M'\ M)^{-1}\ M' = A$, então

$$R = Z'\ A\ Z \tag{4.1.2.22}$$

Como A é uma matria idempotente , substituindo-se (4.1.2.9) em (4.1.2.22) ,

$$R = [\ \beta'\ M' + \varepsilon'\ D^{-1}]\ A\ [\ M\ \beta + D^{-1}\ \varepsilon\ ] \longrightarrow$$

$$R = \beta'\ M'\ A\ M\beta + \beta'\ M'\ A\ D^{-1}\varepsilon + \varepsilon'\ D^{-1}\ A\ M\ \beta + \varepsilon'\ D^{-1}\ A\ D^{-1}\ \varepsilon$$

como

$$\beta'\ M'\ A\ M\ \beta = \beta'\ M'\ A\ D^{-1}\varepsilon = \varepsilon'\ D^{-1}\ A\ M\ \beta = \phi \quad ,$$

então

$$R = \varepsilon' \, D^{-1} \, A \, D^{-1} \, \varepsilon \qquad (4.1.2.23)$$

De (4.1.2.9), vem que

$$D^{-1}\varepsilon = Z - M\,\beta \longrightarrow \varepsilon' \, D^{-1} = Z' - \beta' \, M'$$

portanto

$$R = (Z' - \beta' \, M') \, A \, (Z - M\,\beta)$$

Mas

$$Z \cap N\,(M\beta\,,\,I) \longrightarrow (Z - M\,\beta) \cap N\,(0\,,\,I)\,,$$

ou seja R é da forma $X' \, A \, X$ onde $X \cap N\,(0\,,\,I)$, assim

$$R = (Z' - \beta' \, M') \, A \, (Z - M\,\beta) = Y' \, W^{-1} \, Y - \hat{\beta}' \, X' \, W^{-1} \, Y = V - \frac{U^2}{T}$$

então

$$\left(V - \frac{U^2}{T}\right) \cap \chi^2_{[\rho(A)]}\,,$$

onde $\rho\,(A)$ é a característica de A, ou seja

$$\rho\,(A) = \rho\,[I - M\,(M'\,M)^{-1}\,M']$$

Mas A é idempotente, então $\rho\,(A) = tr\,(A)$,

logo

$$\rho\,(A) = tr\,(I) - tr\,[\,M\,(M'\,M)^{-1}\,M'\,]$$

$$\rho\,(A) = n - tr\,[\,M\,(M'\,M)^{-1}\,M'\,]$$

como $M\,(M'M)^{-1}\,M'$ é idempotente,

$$tr\,[\,M\,(M'\,M)^{-1}\,M'\,] = \rho\,[M\,(M'\,M)^{-1}\,M'\,] = \rho\,(M)\,,$$

logo

$$\rho\,(A) = n - \rho\,(A) \longrightarrow \rho\,(A) = n - \rho\,(D^{-1}\,X)$$

Mas D^{-1} é não singular, portanto

$$\rho\ (A) = n - \rho\ (X)$$

$$\rho\ (A) = n - 2$$

Portanto

$$(V - \frac{U^2}{T}) \cap \chi^2_{(n-2)} \qquad (4.1.2.24)$$

através do que testa-se o desvio de regressão.

Assim, o intervalo de confiança para $\underline{b}$ será

$$\hat{b} \pm t_{\alpha} \sqrt{T^{-1}} \longrightarrow \hat{b} \pm t_{\alpha} T^{-\frac{1}{2}} , \qquad (4.1.2.25)$$

onde t_{α} é o valor tabelado de $\underline{t}$ de "Student", ao nível $\underline{\alpha}$ de probabilidade, para a precisão desejada.

4.1.3 - Estimação dos Pesos w^{ij}

Os pesos apropriados, pelos quais são ponderadas as variâncias observadas de parcelas de diferentes tamanhos que conduzem, como foi visto no item anterior, a um estimador imparcial de $\underline{\beta}$, com variância mínima assintótica, são os elementos da matriz de informação inversa, dos valores de Y_1.

Desde que 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 de V_1, V_2, V_3 e V_4, as quais são independentes, portanto, suas variâncias são respectivamente

$$V(V_1) = \frac{2\ V_1^2}{I - 1} \quad , \qquad V(V_2) = \frac{2\ V_2^2}{I\ (J - 1)} \quad ,$$

$$V(V_3) = \frac{2\ V_3^2}{I\ J\ (K - 1)} \quad , \qquad V(V_4) = \frac{2\ V_4^2}{I\ J\ K\ (L - 1)} \quad ,$$

(4.1.3.1)

pois, se S^2 é a variância de uma amostra aleatória de tamanho $\underline{n}$, de uma população normal de variância σ^2, então

$$E(S^2) = \sigma^2 \qquad\qquad V(S^2) = \frac{2}{n - 1}\ \sigma^4 \quad ,$$

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

Assim, como V_1, V_2, V_3 e V_4 são independentes e V_i $(i = 1, 2, 3, 4)$ é a variância da amostra aleatória, cuja população é normal, então

$$V(V_i) = \frac{2\ V_i^2}{n - 1}$$

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

$$V(V_1') = \frac{2\ V_1^2}{I - 1}$$

Desta forma determinam-se os elementos da matriz de informação de V_i', como se segue. Como,

$$V(V_1) = \frac{2\,V_1^2}{I - 1} \quad , \qquad V(V_2) = \frac{2\,V_2^2}{I(J-1)} \quad ,$$

$$V_2' = \frac{I(J-1)\,V_2 + (I-1)\,V_1}{IJ - 1} \quad \text{e} \quad V(V_2') = E\,[\,V_2' - E(V_2')\,]^2$$

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

$$\longrightarrow \quad E(V_2') = \frac{I(J-1)}{IJ-1} E(V_2) + \frac{I-1}{IJ-1} E(V_1)$$

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

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

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

Por independência de V_1 e V_2

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

que pela substituição de (4.1.3.1)

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

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

Pelo mesmo procedimento determina-se que

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

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

Do mesmo modo,

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

portanto suas expressões são respectivamente:

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

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

$$Cov\,(V'_1\,,\,V'_4) = \frac{2\,(I-1)\,V_1^2}{(I-1)(IJKL-1)}$$

$$Cov\,(V'_2\,,\,V'_3) = \frac{2\,I\,(J-1)\,V_2^2 + 2\,(I-1)\,V_1^2}{(IJ-1)(IJK-1)}$$

$$\mathrm{Cov}\,(V_2', V_4') = \frac{2\,I\,(J-1)\,V_2^2 + 2\,(I-1)\,V_1^2}{(I\,J-1)(I\,J\,K\,L-1)}$$

$$\mathrm{Cov}\,(V_3', V_4') = \frac{2\,I\,J\,(K-1)\,V_3^2 + 2\,I\,(J-1)\,V_2^2 + 2\,(I-1)\,V_1^2}{(I\,J\,K-1)(I\,J\,K\,L-1)}$$

Logo a matriz de informação das V_i' é a que se segue

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

onde

$$A = 2\,I\,J\,K\,(L-1)\,V_4^2 \quad , \qquad B = 2\,I\,J\,(K-1)\,V_3^2 \quad ,$$

$$C = 2\,I\,(J-1)\,V_2^2 \quad , \qquad D = 2\,(I-1)\,V_1^2 \quad .$$

Portanto a matriz inversa das V_i' será:

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

Isto posto, determinam-se os pesos w^{ij}, multiplicando-se cada linha e cada coluna da matriz de informação inversa das V_i', pelas correspondentes V_i', cujo resultado é uma aplicação imediata do método da "programação dos erros", isto é:

Seja

$$\text{Cov}\,(\log V_i' \,,\, \log V_j') =$$

$$= E\,\{[\,\log V_i' - E\,(\log V_i)\,]\,[\,\log V_j' - E\,(\log V_j')\,]\,\}$$

$$\text{Cov}\,(\log V_i' \,,\, \log V_j') \simeq E\{[\,\frac{1}{V_i'}\,d\,(V_i')\,]\,[\,\frac{1}{V_j'}\,d\,(V_j')\,]\}$$

$$\text{Cov}\,(\log V_i' \,,\, \log V_j') \simeq \frac{1}{V_i' \,.\, V_j'}\,E\{[\,d\,(V_i')\,]\,[d\,(V_j')\,]\}$$

$$\text{Cov}\,(\log V_i' \,,\, \log V_j') \simeq \frac{\text{Cov}\,(V_i' \,,\, V_j')}{V_i' \,.\, V_j'}$$

então

$$\text{Cov}\,(V_i' \,,\, V_j') \simeq V_i' \,.\, V_j' \;\text{Cov}\,(\log V_i' \,,\, \log V_j') \quad .$$

Assim, designando-se por

$$w\,t\,(Y_i \,,\, Y_j) \qquad e \qquad w\,t\,(V_i' \,,\, V_j')$$

os elementos na posição $(i \,,\, j)$ das matrizes de informação inversas de Y_i e V_i', respectivamente, tem-se que

$$w\,t\,(Y_i \,,\, Y_j) = w_{ij} = V_i' \,.\, V_j' \; w\,(V_i' \,,\, V_j')$$

Desta forma, a matriz W^{-1} será a de (4.1.3.2).

$$\begin{bmatrix} [(I-1)V_1']^2 \left(\frac{1}{C} + \frac{1}{D}\right) & -\frac{(I-1)(IJ-1)}{C} V_1'V_2' & 0 & 0 \\ -\frac{(I-1)(IJ-1)}{C} V_1'V_2' & [(IJ-1)V_2']^2 \left(\frac{1}{C} + \frac{1}{B}\right) & -\frac{(IJ-1)(IJK-1)}{B} V_2'V_3' & 0 \\ 0 & -\frac{(IJ-1)(IJK-1)}{B} V_3'V_4' & [(IJK-1)V_3']^2 \left(\frac{1}{A} + \frac{1}{B}\right) & -\frac{(IJK-1)(IJKL-1)}{A} V_3'V_4' \\ 0 & 0 & -\frac{(IJK-1)(IJKL-1)}{A} V_3'V_4' & \frac{[(IJKL-1)V_4']^2}{A} \end{bmatrix}$$

(4.1.3.2)

Sendo portanto, os pesos w facilmente estimados da própria matriz de informação inversa das V_i' . Assim, w^{11} por exemplo, é determinado por

$$w^{11} = [(I - 1) V_1']^2 \left(\frac{1}{C} + \frac{1}{D}\right) ,$$

seguindo-se do mesmo modo, determinam-se os outros.

Observa-se que a base de todos os cálculos expostos é a matriz de informação inversa das V_i' , das parcelas de diferentes tamanhos. Porque essas variâncias são expressas como combinações lineares dos quadrados médios originais, os quais são independentes, e não em termos dos componentes de variância, os quais são correlacionados uns com os outros. A matriz de informação resultante, e sua inversa, tomam, como se vê, uma forma relativamente simples.

Se,porém, se empregarem logarítmos de base 10 , os pesos ficam multiplicados pelo fator

$$(\log_e 10)^{-2} ,$$

que indica o ajuste necessário, em caso de se haver empregado logarítmos de base 10 , pois

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

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

Assim sendo, quando se usam logarítmos de base 10, embora o valor de $\underline{\hat{b}}$ continue como em (4.1.2.15) mas sua variância é agora

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

Consequentemente os limites do intervalo de confiança para $\underline{b}$ (4.1.2.25) serão também alterados, ficando como se segue

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

Do mesmo modo que (4.1.2.24) fica agora

$$5{,}302\ (V - \frac{U^2}{T})$$

Demonstra-se a seguir que é preciso o procedimento aqui usado para estimar os pesos w^{ij}. Isto é, que a soma dos pesos é igual à metade do número de graus de liberdade para as somas de quadrados, das quais foram estimadas as variâncias. Isto é demonstrado por serem verificados, como é o caso, os seguintes aspectos.

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 da variância, são independentes, e têm variância assintótica, ou seja:

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

$$V_i' = 2 \cdot \frac{1}{n} \quad ,$$

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

Por outro lado, a "informação" é dada por

$$\mathrm{Inf} = \frac{1}{V_i'}$$

No caso

$$\mathrm{Inf} = \sum_i \sum_j w^{ij}$$

logo

$$\sum_i \sum_j = \frac{1}{\frac{2}{n}} \longrightarrow \sum_i \sum_j w^{ij} = \frac{1}{2}\, n \quad ,$$

portanto, para dados de ensaios de uniformidade

$$\sum_i \sum_j w^{ij} = \frac{1}{2}\,(I\ J\ K\ L - 1) \qquad (4.1.3.1)$$

A verificação desta igualdade constitue uma evidência do cálculo correto dos pesos, no caso específico que se está estudando de ensaio de uniformidade.

Será visto adiante, como se apresenta esta igualdade, em casos de dados de outros ensaios.

4.2 - Estimação de $\underline{b}$ Baseada em Dados de Outros Ensaios Experimentais

Quando os componentes de variância são estimados de dados de outros ensaios experimentais, as estimativas são calculadas do mesmo modo que no caso de dados de ensaios de uniformidade. Mas, desde que um número de contrastes é dado para a estimação dos efeitos de tratamentos, as diferentes variâncias de parcelas e blocos são estimadas com um número menor de graus de liberdade, portanto, com menor precisão que as estimadas de ensaios de uniformidade. Afastadas essas complicações, para as quais deve-se fazer concessões na determinação dos pesos, para os vários componentes, a determinação de um estimador linear imparcial, com variância mínima assintótica para $\underline{\beta}$, segue o mesmo procedimento apresentado no item (4.1).

O procedimento será ilustrado pela análise de um experimento em blocos ao acaso com parcelas sub-subdivididas, como sugerem KOCH e RIGNEY (1951), no qual calculam quatro estimativas básicas de variância. Nesse modelo considera-se que não existe a interação bloco x tratamento. Assim o experimento segue o modelo estatístico

$$y_{ijkl} = m + r_i + t_j + \varepsilon_{1ij} + t'_k + (t\,t')_{jk} + \varepsilon_{2ijk} + e_{ijkl} \quad (4.2.1)$$

onde:

m é o efeito da média geral,

r_i é o efeito do bloco $\underline{i}$ ($i = 1, 2, \ldots, I$),

t_j é o efeito do tratamento j (j = 1 , 2 ,..., J) ,

ε_{1ij} é o efeito da interação (r x t) ou erro 1 , e representa blocos dentro de repetições,

t'_k é o efeito do tratamento (k = 1 , 2 ,..., K),

ε_{2ijk} é o efeito da interação (r x t x t') ou erro 2 e representa parcelas dentro de tratamentos, dentro de repetições

e_{ijkl} é o efeito sub-subparcelas dentro de subparcelas (l = 1 , 2 ,..., L)

Através da aplicação do método dos mínimos quadrados, para minimizar a soma dos quadrados dos erros, na equação (4.2.1), obtém-se as somas de quadrados e outros componentes para a análise de variância do Quadro (4.2).

Deste modo, caso haja interesse, pode-se testar o efeito de subparcelas com o erro amostral, que provém das subsubparcelas.

Do mesmo modo que em ensaios de uniformidade, a variância estimada de parcelas do tamanho de uma repetição completa e V_1 . Esta variância é estimada com o número total de graus de liberdade para os blocos (I - 1), portanto sua variância V'_1 será:

$$V'_1 = \frac{2\ V_1^2}{I - 1} \quad ,$$

exatamente como foi visto no parágrafo (4.1).

QUADRO 4.2 - 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	$\sigma^2 + L\,\sigma_p^2 + K\,L\sigma_B^2$
Tratamentos (2) T'	K - 1		$\sigma^2 + L\,\sigma_p^2 + \frac{IJL}{K-1} \sum_k t'^2_k$
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		

Na estimação do quadrado médio de blocos (variância para blocos), aqui representada pela interação repetição x tratamento, 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 $(J - 1)$ contém efeitos de tratamentos. Assim, a variância estimada entre blocos (4.1.1.2), é agora

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

Analogamente, a variância V_1' de V_1 , é ajustada pelo fator

$$\frac{I}{I - 1}\ .$$

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

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

onde

$$A = 2\ I\ J\ K\ (L-1)\ V_4^2 \quad ,$$

$$B = \frac{2\ I^2\ J\ (K-1)\ V_3^2}{(I-1)} \quad ,$$

$$C = \frac{2\ I^2\ (J-1)\ V_2^2}{(I-1)} \quad ,$$

$$D = 2\ (I-1)\ V_1^2$$

A estimação dos pesos, neste caso, é feita do mesmo modo que para dados de ensaios de uniformidade, sendo a matriz W^{-1}, a de (4.1.3.2). Por sua vez, para dados de experimentos em parcelas subdivididas, a soma dos pesos será

$$\sum_i \sum_j w^{ij} = 1/2\ J\ K\ (I\ L-1) \quad ,$$

e constitue, como no caso anterior, uma evidência do cálculo correto dos pesos.

Idênticos resultados podem ser determinados para reticulados quadrados e outros tipos de delineamentos experimentais.

4.3 - ESTIMAÇÃO DO TAMANHO ÓTIMO DA PARCELA

4.3.1 - Tamanho ótimo da parcela em função do custo

O tamanho ótimo da parcela depende da variabilidade do solo e da relação entre o custo fixo e o que varia com o número de unidades (custo dos diversos fatores envolvidos na condução do experimento). Partindo-se, como FAIRFIELD SMITH (1938), da linearidade, suponha-se que o custo de r repetições seja dado pela regressão linear

$$C(x) = r\, K_1 + r\, K_2\, x \quad , \qquad (4.3.1.1)$$

onde:

r é o número de repetições,

K_1 é a parte do custo que e proporcional ao número de parcelas por tratamento,

K_2 é a parte do custo que é proporcional à área total por tratamento, que minimizado dá o tamanho ótimo $\underline{x}$, da parcela, expresso em unidades básicas, conforme se demonstra a seguir.

Usando multiplicadores não determinados, no cálculo diferencial, sejam

$$\log V(x) = L \quad \text{e} \quad F(x) = C(x) + \lambda\,[\log V(x) - L]$$

onde:

λ é o multiplicador indeterminado,

$C(x)$ é dado em (4.3.1.1) e

$\log V(x)$ é dado em (4.1.2.2).

Assim

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

Minimizando-se $F(x)$ em relação a $\underline{x}$ e $\underline{r}$, vem

$$\left.\begin{aligned} \frac{\partial F}{\partial x} &= r\,K_2 - \lambda\,b\,\frac{1}{x} \\ \frac{\partial F}{\partial r} &= K_1 + K_2\,x - \lambda\,\frac{1}{r} \end{aligned}\right. \longrightarrow \begin{cases} r\,K_2\,x - \lambda\,b = 0 \\ r\,K_1 + r\,K_2\,x - \lambda = 0 \end{cases}$$

Resolvendo-se ambas as equações em λ, vem

$$r\,K_2\,x = r\,K_1\,b + r\,K_2\,b\,x$$

$$r\,K_2\,x - r\,b\,K_2\,x = r\,b\,K_1 \longrightarrow x\,K_2\,(1 - b) = b\,K_1$$

$$x = \frac{b\,K_1}{(1 - b)\,K_2}\,, \qquad (4.3.1.2)$$

onde K_1 e K_2 estão definidos anteriormente e $\underline{b}$ é o coeficiente de heterogeneidade do solo. Deste modo se obtém, por (4.3.1.2) o tamanho ótimo da parcela em função dos custos, determinados pela razão K_1 / K_2.

4.3.2 - Tamanho ótimo da parcela independente do custo

Associando-se o coeficiente $\underline{b}$ de FAIRFIELD SMITH (1938), à fórmula de COCHRAN e COX (1957), para determinar o numero de repetições, pode-se estabelecer uma fórmula, segundo foi sugerido por HATHWAY (1961), para o cálculo do tamanho conveniente da parcela, independente do custo. Tal aplicação é útil para estabelecer o tamanho da parcela, para uma determinada diferença entre médias, que se deseje conhecer "a priori" e poderá ser comprovada entre médias para um determinado tamanho de parcela.

Assim, se

$$V(x) = \frac{V_1}{x^b} \qquad (4.3.2.1)$$

é a variância do rendimento por unidade de área, então, em parcelas de $\underline{x}$ unidades

$$V(x) = \frac{V'(x)}{x} \longrightarrow V(x) = \frac{s^2(x)}{x}$$

Mas, o coeficiente de correlação para parcelas de tamanho $\underline{x}$ é calculado segundo

$$CV(x) = \frac{s(x)}{m(x)}$$

onde:

$m(x)$ é a média aritmética calculada do rendimento de parcelas de tamanho $\underline{x}$.

Note-se que, como a variância, a média da parcela é também reduzida à unidade básica, ou seja

$$m_1 = \frac{m(x)}{x}$$

Portanto

$$CV(x) = \frac{s(x)}{m_1 \; x} \longrightarrow CV^2(x) = \frac{s^2(x)}{m_1^2 \; x^2} \longrightarrow$$

$$\longrightarrow \quad CV^2(x) = \frac{V(x)}{m_1^2} \qquad (4.3.2.2)$$

Substituindo-se este valor na fórmula de COCHRAN e COX (1957), ou seja, na fórmula

$$r = \frac{2 \; (CV)^2 \; (t_1 + t_2)^2}{d^2}$$

tem-se

$$r = \frac{2 \; (t_1 + t_2)^2 \; \dfrac{V(x)}{m_1^2}}{d^2} \longrightarrow r = \frac{2 \; (t_1 + t_2)^2 \; V(x)}{m_1^2 \; d^2}$$

Substituindo-se $V(x)$, da expressão (4.1.2.1) resulta

$$r = \frac{2 \; (t_1 + t_2) \; V_1}{x^b \; m_1^2 \; d^2}$$

onde $\dfrac{V_1}{m_1^2} = (CV_1)^2$, logo

$$x^b = \frac{2\ (t_1 + t_2)(CV_1)^2}{r\ d^2}, \qquad (4.3.2.3)$$

onde:

r é o número exigido de repetições para detectar uma verdadeira diferença de d unidades;

d é a verdadeira diferença entre dois tratamentos (medida em porcentagem da média — é a diferença que se deseja comprovar);

CV_1 é o coeficiente de variação de parcelas de uma unidade de tamanho;

t_1 é o valor significativo do teste t, ao nível α de probabilidade;

t_2 é o valor de tabela, do teste t correspondendo a 2 (1 - p), onde p é a probabilidade "a priori" de se obter diferenças significativas entre médias.

A fim de se conseguir o conhecimento "a priori" das diferenças entre as médias, dadas em porcentagem da média, que o teste de significância irá detectar, ao nível α de probabilidade, sendo por exemplo, de 80% a probabilidade de se encontrar resultados significativos, calculam-se essas diferenças para vários tamanhos de parcelas, com diferentes valores de CV.

Na determinação das diferenças, utiliza-se a fórmula

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

que resulta imediatamente de (4.3.2.3)

5 - CONCLUSÕES

O número de repetições para estabelecer o tamanho ideal de parcelas, para uma determinada diferença entre médias, fixada "a priori" e o tamanho ótimo da parcela, para determinada lavoura não podem ser generalizados, pois variam com o solo, com a lavoura, devido a que cada uma delas reage diferentemente no solo, com a época, com os custos dos diferentes processos experimentais e com o procedimento para tal utilizado. Pode-se no entanto, estender para outros locais que comprovadamente não apresentem grandes diferenças quanto à natureza do solo, desde que a lavoura seja a mesma.

Para fins de estimação do tamanho ótimo de parcela, o elemento principal é o coeficiente de heterogeneidade do solo, estimado a partir da definição de FAIRFIELD SMITH (1938),

$$V(x) = \frac{V_1}{x^b} \quad ,$$

ou

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

equação que é linear em $\log x$, e que pode ser estimada pelo método dos mínimos quadrados generalizados, que conduz ao estimador de $\underline{\beta}$ com variância mínima assintótica, cuja estimativa do coeficiente $\underline{b}$, de regressão

$$\hat{b} = - \frac{\sum\limits_i \sum\limits_j w^{ij} \; Y_i \; (X_i - \bar{X})}{\sum\limits_i \sum\limits_j w^{ij} \; X_i \; (X_i - \bar{X})} \quad ,$$

é ponderado eficientemente, cujos pesos apropriados são os elementos da matriz de informação dos diferentes tamanhos de parcela $(V_i^!)$.

O procedimento utilizado conduz a uma estimativa eficiente do parâmetro $\underline{\beta}$, desde que obedeça às condições teóricas sugeridas pelas observações. Concretamente o procedimento considera a correlação que deve existir entre as variâncias dos diversos tamanhos de parcelas.

Convém ressaltar que os pesos w^{ij} são estimados de dados experimentais, são, portanto, inexatos, daí a imperiosa necessidade de serem obedecidas as condições teóricas sugeridas pelas observações, para que se obtenha estimativa imparcial e de mínima variância.

Para se estimar o coeficiente b̲ , de heterogeneidade do solo, há necessidade de grande cuidado, pois este pode conduzir a vícios e influenciar na estimativa dos componentes de blocos fazendo muito grande ou muito pequenas, consequentemente produzir desvios na linearidade da regressão.

Assim, a estimação de $\underline{\beta}$ por meio de um coeficiente de regressão não ponderado, através da fórmula

$$\hat{b} = \frac{\sum_i Y_i \ (X_i - \bar{X})}{\sum_i (X_i - \bar{X})^2} \qquad ,$$

como o fizeram KOCH e RIGNEY (1951), resulta em muitos casos, em estimativas maiores que1, o que não é esperado, do ponto de vista teórico, portanto frequentemente incorreto. A ponderação dos termos da expressão utilizada no cálculo de $\underline{\hat{b}}$ e das estimativas de variâncias observadas de diferentes tamanhos de parcela, tanto em dados de ensaios de uniformidade, como de outros ensaios experimentais, pelo respectivo número de graus de liberdade, como sugere FAIRFIELD SMITH (1938), não é precisa , uma vez que as estimativas de variâncias são constituídas de componentes comuns, portanto altamente correlacionadas.

Deste modo, o procedimento apresentado para estimar os pesos w^{ij} é seguramente eficiente, porque são estimados de variâncias que são combinações lineares dos quadrados médios originais, que são independentes e não em termos dos componen -

tes de variância que são correlacionados.

O procedimento apresentado, de ponderação do coeficiente de regressão $\underline{b}$ é aplicável tanto a dados de ensaio de uniformidade quanto a dados de outros ensaios em que estejam presentes, efeitos de tratamentos.

A significância do desvio de regressão, quando existem mais de as estimativas de variâncias, como é frequentemente o caso, pode ser testada através de

$$V - \frac{U^2}{T} \quad ,$$

que tem distribuição de Qui-quadrado com n - 2 graus de liberdade.

Deste modo, o tamanho ótimo da parcela $\underline{x}$, em unidades básicas pode ser estimado em função dos custos envolvidos no experimento por

$$x = \frac{\hat{b}\ K_1}{(1 - b)\ K_2} \quad ,$$

que torna mínimo o custo, e pela relação

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

em função do número de repetições necessárias para detectar uma determinada diferença de médias, com K% de segurança, a um nível $\underline{\alpha}$ de probabilidade, independentemente do custo.

6 - SUMMARY

The study of the optimum plot size in field experiments aims at minimizing costs, in terms of experimental error.

The great number of researches existing in the literature indicate its great importance. The principal issue involved is the soil heterogeneity which exerts an influence in the determination of the optimum plot size.

In this study, the regression coefficient b or soil variability index proposed by FAIRFIELD SMITH (1938) was estimated using a regression model in which the data are correlated and have unequal variances, minimizing the sum of squares of the errors through the generalized least square method. Such estimate,

$$\hat{b} = - \frac{\sum_i \sum_j w^{ij} \; Y_i \; (X_i - \bar{X})}{\sum_i \sum_j w^{ij} \; X_i \; (X_i - \bar{X})} \quad ,$$

as well as the variance estimates observed, of the different plot sizes, are weighted by the w^{ij} weights.

The methodology proposed by HATHEWAY and WILLIAMS (1958) is used, with the aim of effectively estimating the w^{ij} weight of the data. Such procedure is used for uniformity trial data, studied, in this case, through a hierarchic classification model, that is

$$y_{ijkl} = m + r_i + b_{ij} \; p_{ijk} + s_{ijkl} \quad ,$$

as well as for experimental data, of the comparative variety type, where there are treatment effects in this case conducted through the analysis of an experiment in randomized block design, with split-split plots, that is

$$y_{ijkl} = m + r_i + t_j + \varepsilon_{1ij} + t'_k + (t\,t')_{jk} + \varepsilon_{2ijk} + e_{ijkl}$$

It is proved that the estimate of <u>b</u> , obtained through this procedure has an asymptotic minimum variance.

The regression deviation,

$$V - \frac{U^2}{T} \quad ,$$

is studied through the chi-square test , with n - 2 degrees of freedom since, it is distributed as such.

The estimate of the optimum plot size is focused under two aspects:

a - in terms of the minimum cost involved in the experiment, in basic units, through

$$x = \frac{b\ K_1}{(1 - b)\ K_2} \quad ,$$

b - in terms of the number of replications necessary to obtain, a specific difference of average

$$\underline{d} = \sqrt{\frac{2\ (t_1 + t_2)(CV)^2}{(r\ x^b)}} \quad ,$$

fixed "a priori" at an $\underline{\alpha}$ level of probability.

7 - LITERATURA CITADA

AMARAL, E., 1951. Tamanho e Forma de Parcelas em Experimentação com o Cafeeiro. 6º Seminário de Estatística. Campinas, SP.

ANDERSON, R. L. e T. A. BANCROFT, 1952. Statistical Theory in Research. Mc Graw-Hill, Nova York.

ARROYO, J. R. e A. M. CHAVES, 1966. Estimación Eficiente de Parámetros en la Determinación del Tamaño Optimo de Parcela. Boletim nº 15. Estación Experimental de La Molina, Peru.

BRIM, C. A. e D. D. MASON, 1959. Estimates o Optimum Plot Size for Soybean Yield Trials. Agronomy Journal, 51: 331-334.

CAMPOS, G. M., 1972. Determinação do Tamanho e Forma das Parcelas para Uso em Experimentos com Girassol. (Dissertação de Mestrado - ESALQ/USP).

CHRISTIDES, B. G., 1939. Variability of Plots of Various Shapes as Affected by Plot Orientation. The Emp. J. Exp. Agr., 7: 330-342.

COCHRAN, W. G. e G. M. COX, 1957. Experimental Design. John Wiley, Nova York.

DRAPER, N. R. e H. SMITH, 1966. Applied Regression Analysis. John Wiley. Nova York.

FAIRFIELD SMITH, H., 1938. An Empirical Law Describing Heterogeneity in the Yields of Agricultural Crops. Jour. Agr., 28: 1-23.

FEDERER, W. T., 1955. Experimental Design. The Macmillan. Nova York.

HATHEWAY, W. H., 1961. Convenient Plot Size. Agronomy Journal, 53: 279-280.

HATHEWAY, W. H. e E. J. WILLIAMS, 1958. Efficient Estimation of the Relationship Between Plot Size and the Variability of Crop Yields. Biometrics, 44: 207-222.

HARRIS, J. A., 1915. On a Criterion of Substratum Homogeneity (or Heterogeneity) in Field Experiments. Amer. Naturalist, 49: 430-454.

HERNANDEZ, J. L. e J. R. ARROYO, 1968. Investigaciones sobre Tamaño Optimo de Parcela en Arroz. Boletim nº 20. Estación Experimental de La Molina, Peru.

IGUE, T. ; D. M. SOUZA e V. NAGAI, 1972. Tamanho de Parcela mais Conveniente para Experimentação de Campo com Arroz. Ciência e Cultura, 24: 1150-1153.

IGUE, T. e H. A. A. MASCARENHAS, 1974. Tamanho das parcelas para Experimentos de Campo com Soja. Boletim Técnico nº 9: 1-28, Instituto Agronômico de Campinas, SP.

KOCH, E. J. e J. A. RIGNEY, 1951. A Method of Estimating Optimum Plot Size from Experimental Data. Agr. Jour., 43: 17-21.

MARANI, A., 1963. Estimation of Optimum Plot Size Using Smith's Procedure. Agr. Jour., 43: 503-506.

PIMENTEL GOMES, F., 1978. Curso de Estatística Experimental. 7ª Edição. Piracicaba, SP.

RAO, C. R., 1973. Linear Statistical Inference and its Applications. 2ª edição. John Wiley. Nova York.

ROBINSON, H. F. ; J. A. RIGNEY e P. H. HARVEY, 1948. Investigations in Plot Technique with Peanuts. Tech. Bul., 86: 1-19.

SEARLE, S. R., 1970. Linear Models. John Wiley, Nova York.

SILVA, E. C., 1972. Estudo do Tamanho e Forma de Parcelas para Experimentos de Soja. (Dissertação de Mestrado - ESALQ - USP).