ISSN 0034-7655

REVISTA DE AGRICULTURA

Vol. 67 Dezembro Nº 3

INTERVALOS DE CONFIANÇA PARA AS COORDENADAS DO PONTO CRÍTICO DA EQUAÇÃO DE REGRESSÃO A DUAS VARIÁVEIS INDEPENDENTES

Marli de Bem Gomes[1]

INTRODUÇÃO

Na pesquisa científica agronômica é bastante comum o ajustamento de superfícies de resposta a dados de experimentos de adubação. Nestes trabalhos procura-se determinar o ponto crítico de interesse, que é, geralmente, o ponto de máximo. Mas se encontram freqüentemente pontos de sela e, às vezes, até de mínimo. Outra dificuldade séria, porém, é que os estimadores das coordenadas do ponto crítico adequado são variáveis aleatórias, de tal sorte que a sua localização deve ser dada por uma região de confiança, para cuja obtenção não há método conhecido.

Há muitos trabalhos que poderíamos citar sobre o ajustamento de superfícies de resposta, como os de CAMPOS (1967), ZAGATTO & PIMENTEL-GOMES (1967), MORAES (1969), PIMENTEL-GOMES (1969), JORGE & CONAGIN (1977), MALHEIROS & PERECIN (1983), ZIMMERMANN & CONAGIN (1986), SANCHES

[1] Professora Associada, Departamento de Matemática e Estatística da ESALQ/USP.

(1986), PIMENTEL-GOMES & CONAGIN (1987), entre outros. Porém vamos comentar com maior detalhe apenas aqueles que têm relação mais direta com este artigo.

ZAGATTO & PIMENTEL-GOMES (1967), em estudo das funções de produção sob o ponto de vista econômico, abordaram equações com uma, duas ou mais variáveis independentes. aplicando-as a ensaios fatoriais de 3^2. Indicaram uma série de dificuldades que surgem e sugeriram:

1) Usar sempre grupos de ensaios numerosos ou ensaios isolados com muitas repetições de boa precisão.

2) Não confiar em doses ótimas obtidas a partir de polinômios cujos coeficientes para os termos de 2º grau não sejam significativamente diferentes de zero.

3) Verificar se o ponto crítico obtido corresponde realmente a um máximo.

4) Calcular intervalos de confiança para as doses ótimas encontradas. Não fizeram, porém, qualquer referência a método de cálculo da variância da dose econômica no caso de regressão polinomial.

CAMPOS (1967) estudou 50 ensaios fatoriais de 3^2 individualmente e agrupados de diferentes formas, ajustando regressão polinomial. Fazendo cortes, calculou a dose ótima, as variâncias e os intervalos de confiança, dessa dose, entre outros estudos. Concluiu que não é aconselhável aplicar regressão polinomial a um único ensaio.

PIMENTEL-GOMES & CONAGIN (1987) resumiram com exemplos agronômicos a teoria de curvas e superfícies de resposta. Citam dois métodos para cálculo do intervalo de confiança para a dose econômica de adubação, no caso do polimônio de segundo grau, mas apenas para o caso de uma só variável independente.

Como podemos ver pela bibliografia citada, não havia nada sobre o assunto, deste trabalho, quando GOMES (1989) estudou com detalhes as variâncias das coordenadas das doses ótimas e a determinação de seus intervalos de confiança, sugerindo uma fórmula para o cálculo da variância das coordenadas do ponto crítico para duas variáveis independentes, no caso da equação de regressão de segundo grau.

Neste trabalho foram simulados experimentos fatoriais de 3^2, e calculados intervalos de confiança para as coordenadas de seus pontos críticos.

MATERIAL E MÉTODOS

Consideraremos um experimento fatorial de 3^2, onde tenhamos dois nutrientes em níveis indicados por X_1 e X_2, com três doses para cada um deles 0, 40, 80 kg/ha. Portanto, a produção Y é uma função.

$$\hat{Y} = f\ (X_1,\ X_2)$$

e poderemos escrever que:

$$\hat{Y} = \hat{a}_{33} + 2\hat{a}_{13}X_1 + 2\hat{a}_{23}X_2 + \hat{a}_{11}X_1^2 + \hat{a}_{22}X_2^2 + 2\hat{a}_{12}X_1X_2,$$

usando o polinômio do segundo grau. Há, neste caso, porém, correlação entre os coeficientes estimados. Por isso, vamos substituir as variáveis por polinômios ortogonais ajustados ao caso:

$$\hat{Y} = \hat{a}_{33} + 2\hat{a}_{13}x_1 + 2\hat{a}_{23}x_2 + \hat{a}_{11}(x_1^2-k) + \hat{a}_{22}(x_2^2-k) + 2\hat{a}_{12}x_1x_2,$$

onde:

$$x_1 = \frac{X_1 - \overline{X}_1}{q_1}, \qquad x_2 = \frac{X_2 - \overline{X}_2}{q_2}$$

sendo $\overline{X}_1$ a dose média usada para a variável X_1 e $\overline{X}_2$ a dose média para a variável X_2. No caso presente, $q_1=q_2=q=40$ kg/ha. Nestas condições, os estimadores dos 6 parâmetros são independentes, isto é, têm covariâncias nulas.

Consideradas as doses $\underline{x}$ com níveis -1, 0, 1, obtém-se facilmente k = 2/3. Portanto a superfície de resposta ajustada tem a equação:

$$(1)\quad \hat{Y} = \hat{a}_{33}+2\hat{a}_{13}x_1+2\hat{a}_{23}x_2+\hat{a}_{11}(x_1^2-2/3)+\hat{a}_{22}(x_2^2-2/3)+2\hat{a}_{12}x_1x_2.$$

As produções de um fatorial completo de 3^2 (TABELA I), ponto de partida para a simulação de dados feita neste trabalho, foram obtidas a partir das médias de 50 ensaios de milho com N, P e K, instalados pelo Dr. Hermano Vaz de Arruda em Ribeirão Preto-SP. As doses utilizadas para os três nutrientes foram zero, 40 e 80 kg/ha. Como só consideramos os nutrientes N e P, neste trabalho, utilizamos para cada um deles a média de produção dos três níveis potássicos que lhe correspondem.

TABELA I. Produções médias de milho com os nutrientes N e P, com as doses -1, 0 e 1.

Tratamentos		Médias de Produção (kg/ha)
-1	-1	3845
-1	0	4164
-1	1	4161
0	-1	4803
0	0	4904
0	1	5146
1	-1	5218
1	0	5402
1	1	5550

Esses valores serviram de base para gerar séries de 9 dados, usando a expressão $Y_i + \sigma\, e_i$, onde $e_i \Omega\, N(0;\, 1)$. Fizemos uma série de 200 repetições para cada um dos seguintes valores de σ: 10, 50, 100, 200, 300, 400, 500 e 600, o que é equivalente a mudar o número de repetições de 3600 ($\sigma = 10$) para 1 ($\sigma = 600$). Com esses valores fizemos os cálculos que mencionamos a seguir:

1) Cálculo das coordenadas do ponto crítico para cada experimento gerado.

2) Verificação da natureza dos pontos críticos obtidos em todos os experimentos gerados.

3) Cálculo das variâncias e intervalos de confiança para x_1^* e x_2^*.

Uma vez obtidas as coordenadas dos pontos críticos $\hat{x}_1^*$ e x_2^*, 200 valores de cada, para cada valor de σ, calculamos as médias $\bar{x}_1^*$ e $\bar{x}_2^*$, e as variâncias estimadas $\sigma\ (\hat{x}_1^*)$ e $\hat{\sigma}^2(\hat{x}_2^*)$ calculadas pela fórmula comum de estimação de variância. Calculamos também $\hat{V}(\hat{x}_1^*)$ e $\hat{V}(\hat{x}_2^*)$, dados por duas fórmulas deduzidas a partir da diferenciação em relação aos parâmetros das funções seguintes (GOMES, 1989):

$$\hat{x}_1^* = \frac{-\hat{a}_{13} \cdot \hat{a}_{22} + \hat{a}_{12} \cdot \hat{a}_{23}}{\hat{a}_{11} \cdot \hat{a}_{22} - \hat{a}_{12}},$$

$$\hat{x}_1^* = \frac{\hat{a}_{13}\ \ \hat{a}_{12} - \hat{a}_{11} \cdot \hat{a}_{23}}{\hat{a}_{11} \cdot \hat{a}_{22} - \hat{a}_{12}}.$$

Isto foi feito para os 200 dados obtidos para cada coordenada e para cada valor de σ. Obtivemos também $V(\hat{x}_1^*)$ e $V(\hat{x}_2^*)$, que são as médias dos valores calculados para $V(\hat{x}_1^*)$ e $V(x_2^*)$. Foram, ainda, determinados intervalos de confiança para x_1^* e x_2^*, considerando as duas variâncias estimadas, para cada caso.

RESULTADOS E DISCUSSÃO

A superfície de resposta (1), estimada com os dados da **TABELA I** foi:

$$\hat{Y} = 4975{,}25+666{,}50x_1+165{,}16x_2-228{,}05x_1^2-36{,}06x_2^2+3{,}9176x_1x_2.$$

Esta superfície tem ponto de máximo, de coordenadas:

$$\hat{x}_1^* = 1{,}4817 \quad \text{e} \quad \hat{x}_2^* = 2{,}3706.$$

Caso consideremos as doses originais, com q = 40kg/ha, teremos:

$$\hat{X}_1^* = 99,3 \text{ kg/ha} \quad \text{e} \quad \hat{X}_2^* = 134,8 \text{ kg/ha}.$$

Verifica-se, pois, que essas coordenadas são extrapolações de maior dose usada no experimento, 80 kg/ha. Embora isto tenha ocorrido resolvemos manter o uso desses dados, pois desde que ocorra ponto de máximo, os métodos utilizados continuam válidos. Por outro lado, devemos salientar que essa equação decorre de dados de experimentos de campo e, pois, representa uma situação real.

A natureza dos pontos críticos obtidos consta da **TABELA II**:

TABELA II. Resultados obtidos com relação à natureza dos pontos críticos.

σ	$V(\hat{y})$	C.V. (%)	Nº de P. de máximo	Nº de P. de sela	Nº de P. de mínimo
10	100	0,21	200	0	0
50	2.500	1,04	172	28	0
100	10.000	2,08	142	58	0
200	40.000	4,17	107	92	1
300	90.000	6,25	83	105	12
400	160.000	8,33	73	118	9
500	250.000	10,42	68	109	23
600	360.000	12,50	67	115	18

Pela **TABELA II**, notamos que os pontos de mínimo apareceram a partir de $\sigma = 200$, isto é, a partir de um coeficiente de variação de apenas 4,17%. E para $\alpha = 500$ (C.V. = 10,42%) já tivemos apenas 68 pontos de máximo, ao lado de 109 de sela e de 23 pontos de mínimo, em 200 experimentos.

Foram obtidas as médias das coordenadas $\bar{x}_1^*$ e $\bar{x}_2^*$ e com elas os intervalos de confiança, usando as duas variâncias obtidas para cada coordenada $\sigma^2(\bar{x}_i^*)$ e $V(\hat{x}_i^*)$, onde $i = 1,2$. Os resultados estão expostos na **TABELA III** para x_1^* e na **TABELA IV** para x_2^*.

Vamos lembrar que os verdadeiros valores de x_1^* e x_2^*, já expostos são:

$$x_1^* = 1{,}4817 \qquad \text{e} \qquad x_2^* = 2{,}2706$$

Na **TABELA III** notamos que com $\hat{\sigma}^2(\hat{x}_1^*)$ apenas o intervalo obtido para $\alpha = 100$ não contém o valor real. Já os intervalos calculados com $\hat{V}(\hat{x}_1^*)$ todos contém x_1^*, mas são excessivamente amplos para $\sigma > 10$.

Para x_2^*, observamos na **TABELA IV** que os intervalos construídos com $\hat{\sigma}^2(\hat{x}_2^*)$ para $\sigma = 10$, 100, 400 e 500 não contêm o verdadeiro valor de x_2^*, este porém estando contido em todos os intervalos obtidos com $\hat{V}(\hat{x}_2^*)$. Mas, estes são muito amplos para $\sigma > 10$.

Com relação às médias calculadas na **TABELA III**, verificamos que os valores observados para $\bar{x}_1^*$ diferem muito do valor correto de x_1^*, chegando a 34,10 para $\sigma = 600$ e a -1,6010 para $\sigma = 400$. Para $\sigma = 10$ o valor é bem próximo do real.

Para $\bar{x}_2^*$, exposto na **TABELA IV**, há uma grande diferença entre os valores observados nos diferentes valores de σ, sendo o maior encontrado 2,4502, próximo do real, e o menor -14,93.

TABELA III. Relação das médias e intervalos de confiança calculados para a média da abscissa do ponto crítico x_1^*.

σ	$\bar{x}_1^*$	I. C. para x_1^* a 5%	
		Com $\hat{\sigma}^2(\bar{x}_1^*)$	Com $\overline{\hat{V}(\hat{x}_1^*)}$
10	1,4808	[1,473; 1,488]	[1,473; 1,488]
50	3,8127	[0,4402; 7,185]	[-7523,00 ; 7531,00]
100	1,7900	[1,492; 2,088]	[-7,750; 11,330]
200	11,4760	[-4,327; 27,279]	[-11266,50 ; 11289,50]
300	3,0788	[-2,936; 9,093]	[-3545,00 ; 3551,00]
400	-1,6010	[-5,949; 2,747]	[-1614,00 ; 1611,00]
500	0,5898	[-1,723; 2,902]	[-359,90 ; 361,10]
600	34,1040	[-41,66 ; 109,87]	[$-1,338\text{x}10^6$; $1,338\text{x}10^6$]

TABELA IV. Relação das médias e intervalos de confiança calculados para a média da ordenada do ponto crítico x_2^*.

σ	$\bar{x}_2^*$	I. C. para x_2^* a 5%	
		Com $\hat{\sigma}^2(\bar{x}_2^*)$	Com $\overline{\hat{V}(\hat{x}_2^*)}$
10	2,4502	[2,376 ; 2,524]	[2,368; 2,532]
50	-14,9300	[-105,05 ;75,19]	[-286577,00 ;286548,00]
100	0,7031	[-0,5788; 1,985]	[-41,03 ; 42,44]
200	-11,0100	[-31,56 ; 9,544]	[-37051,00 ; 37029,00]
300	0,3375	[-3,816 ; 4,491]	[-2138,00 ; 2139,00]
400	-0,1991	[-0,7948; 0,3966]	[-149,70 ; 149,30]
500	-0,2467	[-1,199 ; 0,7054]	[-80,60 ; 80,11]
600	-12,0000	[-36,68 ;12,68]	[-438325,00 ;438301,00]

CONCLUSÕES

1) Os pontos de mínimo aparecem para $\sigma \geq 200$.

2) Os intervalos de confiança calculados para a coordenada x_1^* em geral, tanto com $\sigma^2(\hat{x}_1^*)$ como com $\hat{V}(\hat{x}_1^*)$, contêm o verdadeiro valor de x_1^*, porém com esta última variância eles são excessivamente amplos para $\sigma > 10$.

3) Os intervalos de conciança calculados com $\hat{\sigma}^2(\hat{x}_2^*)$ não contêm o verdadeiro valor de x_2^* em 50% dos casos, porém todos os intervalos de confiança calculados com $\hat{V}(\hat{x}_2^*)$ contêm o verdadeiro valor, mas são excessivamente amplos para $\sigma > 10$.

4) Os valores médios de $\bar{x}_1^*$ e $\bar{x}_2^*$ apresentam dispersão muito grande, sendo que $\bar{x}_1^* \varepsilon [-1,60;34,10]$ e $\bar{x}_2^* \varepsilon [-14,93;2,45]$.

RESUMO

Consideramos neste trabalho um experimento fatorial 3^2 de milho com dois nutrientes, nitrogênio e fósforo, ambos com doses 0, 40, 80 kg/ha. Ajustamos aos dados de produção uma superfície de resposta de segundo grau, substituindo-se as variáveis por polinômios ortogonais, para que não haja correlação entre os coeficientes. A partir da equação teórica determinamos as coordenadas do ponto crítico x_1^* e x_2^* e, através de diferenciação em relação aos parâmetros, deduzimos fórmulas para $V(x_1^*)$ e $V(x_2^*)$. Testamos estas fórmulas com o auxílio da superfície:

$$\hat{Y} = 4975,25+666,50x_1+165,16x_2-228,05x_1^2 - 36,06x_2^2 + $$
$$+ 3,9176x_1x_2,$$

cujos parâmetros serviram de base para a simulação de séries de 200 experimentos com σ : 10, 50, 100, 200, 300, 400, 500 e 600. Com esses dados calculamos intervalos de confiança para as coordenadas x_1^* e x_2^* dos pontos críticos, de cada série de 200 ensaios gerados para cada σ, usando $\hat{\sigma}^2(\hat{x}_1^*)$ e $\hat{\sigma}^2(\hat{x}_2^*)$ com valores obtidos pela fórmula comum de variância e, $\overline{\hat{V}(\hat{x}_1^*)}$ e $\overline{\hat{V}(\hat{x}_2^*)}$, que são as médias dos valores calculados para $\hat{V}(\hat{x}_1^*)$ e $\hat{V}(\hat{x}_2^*)$.

Concluimos que:

1) Os intervalos de confiança calculados para a coordenada x_1^* em geral, contêm o verdadeiro valor x_1^* com ambas variâncias usadas. Porém com $\overline{\hat{V}(\hat{x}_1^*)}$ eles são excessivamente amplos para $\sigma > 10$.

2) Os intervalos de confiança calculados com $\hat{\sigma}^2(\bar{x}_2^*)$ não contêm o verdadeiro valor de x_2^* em 50% dos casos. Todos os intervalos obtidos com $\overline{\hat{V}(\hat{x}_2^*)}$ contêm esse valor, mas são excessivamente amplos a partir de $\sigma > 10$.

Palavras-chave: Superfície de resposta, Intervalos de confiança, Coordenadas do ponto crítico.

SUMMARY

CONFIDENCE INTERVALS FOR THE COORDINATES OF THE CRITICAL POINT OF THE QUADRATIC REGRESSION EQUATION WITH TWO INDEPENDENT VARIABLES

In this paper a 3^2 factorial experiment with two nutrients (nitrogen and phosphorus) was used, assuming levels 0, 40, 80 kg/ha. A second degree response surface was fitted, using orthogonal polynomials in order to have estimates independent among one another.

From the theoric equation we obtained the coordinates x_1^* and x_2^* of the critical point. Their differentiation with respect to the coefficients was carried out, in order to calculate formulas for $V(x_1^*)$ and $V(x_2^*)$. These formulas were tested using the surface:

$$\hat{Y} = 4975,25 + 666,50x_1 + 165,16x_2 - 228,05x_1^2 - 36,06x_2^2 + + 3,9176x_1x_2,$$

the parameters of which were used to simulate sets of 200 experiments, with σ: 10, 50, 100, 200, 300, 400, 500 and 600. With these data confidence intervals were calculated for x_1^* and x_2^* critical point coordinates, of the 200 experiments generated for each α, using $\hat{\sigma}^2(\hat{x}_1^*)$ and $\hat{\sigma}^2(\hat{x}_2^*)$, calculated by the common formula of variance, and $\overline{\hat{V}(\hat{x}_1^*)}$ and $\overline{\hat{V}(\hat{x}_2^*)}$, which are the means of the values calculated for $\hat{V}(\hat{x}_1^*)$ and $\hat{V}(\hat{x}_2^*)$.

Conclusions:

1) The confidence intervals calculated for the coordinate x_1^*, include, generally, the real value of x_1^*, with the variances $\hat{\sigma}^2(\hat{x}_1^*)$. But with $\overline{\hat{V}(\hat{x}_1^*)}$ they are excessively large for σ > 10.

2) The confidence intervals calculated with $\hat{\sigma}^2(\hat{x}_2^*)$ do not include the real value of x_2^* in 50% of the cases. All of them include it when calculated with $\overline{\hat{V}(\hat{x}_2^*)}$, but are excessively large for σ > 10.

Key words: Response surface, Confidence intervals, Critical intervals, Point coordinates.

LITERATURA CITADA

CAMPOS, H., 1967. Aspectos da Aplicação das Superfícies de Resposta a Ensaios Fatoriais 3^2 de Adubação. Piracicaba, 82p. (Tese - ESALQ/USP).

GOMES, Marli De Bem, 1989. Intervalos de Confiança para Coordenadas do Ponto Crítico da Equação Quadrática de Regressão a Duas Variáveis Independentes. Piracicaba, 70p. (Tese - ESALQ/USP).

JORGE, Joassy P.N. & A. CONAGIN, 1977. Estudos em um Grupo Especial de Delineamento $(1/5)5^3$. **Bragantia**, **36**: 59-88.

MALHEIROS, E.B. & D. PERECIN, 1983. Posição das Doses de Nutrientes na Análise de Uma Superfície de Resposta à Adubação. **R. Mat. Estat., 1**: 69-78.

MORAES, R.S., 1969. Superfície Polinomial de Resposta num Ensaio de Adubação com Níveis Não Equivalentes. Piracicaba, 58p. (Tese - ESALQ/USP).

PIMENTEL-GOMES, F., 1969. Novos Aspectos do Estudo Econômico de Ensaios de Adubação. **Fertilité, 34**: 3-9.

PIMENTEL-GOMES, F. & A. CONAGIN, 1987. **Experimentos de Adubação**: Planejamento e Análise Estatística. Londrina, Universidade Estadual de Londrina, 102p.

SANCHES, A., 1986. Superfícies de Respostas em Experimentos de Adubação: O Problema dos Pontos de Sela. Piracicaba, 91p. (Mestrado - ESALQ/USP).

ZAGATTO, A.G. & F. PIMENTEL-GOMES, F., 1967. Aspectos Econômicos da Adubação. In: MALAVOLTA, E. **Manual de Química Agrícola**: Adubos e Adubação. 2.ed. São Paulo, Ed. Agronômica Ceres. p.560-586.

ZIMMERMANN, F.J.P. & A. CONAGIN, 1986. Ajuste de Modelos Polinomiais de 2º Grau em Pesquisas com Fertilizantes. **Pesquisa Agropecuária Brasileira, 21**: 971-978.