

## APRIMORAMENTO DA ANÁLISE DE VARIÂNCIA: A INFLUÊNCIA DA PROXIMIDADE ESPACIAL

Cristina Henriques NOGUEIRA<sup>1</sup>  
Renato Ribeiro de LIMA<sup>2</sup>  
Marcelo Silva de OLIVEIRA<sup>2</sup>

- RESUMO: Um dos princípios básicos da experimentação é a aleatorização, que tem como finalidade gerar uma distribuição de amostragem para os erros experimentais, cujo comportamento assintótico aproxima-se de distribuições normais não-correlacionadas. Entretanto, nem sempre a prática de aleatorização é suficiente para neutralizar os efeitos de correlação entre as parcelas adjacentes e, por isso, encontram-se experimentos cujos erros apresentam uma estrutura de dependência espacial definida. Uma maneira para contornar esse problema é utilizar uma abordagem espacial, em que é possível estimar e modelar a correlação espacial entre os erros. Por meio desses modelos espaciais a dependência entre os erros deixa de ser vista como uma inconveniência a ser evitada, passando a ser considerada como um verdadeiro benefício, de forma que sua utilização pode proporcionar resultados mais precisos. Assim, este trabalho descreve como implementar a análise de variância com erros espacialmente correlacionados, na qual a matriz de covariância do erros, modelada por meio de uma abordagem geoestatística, é utilizada como ponderadora das somas de quadrados da análise de variância.
- PALAVRAS-CHAVE: Covariância; análise de experimentos; semivariograma; erros dependentes.

### 1 Introdução

A aleatorização dos tratamentos é considerada uma das principais contribuições de Fisher, uma vez que essa prática gera uma distribuição de amostragem para os erros experimentais cujo comportamento assintótico aproxima-se de

---

<sup>1</sup>Universidade Federal de Lavras – UFLA, Programa de Pós-graduação em Estatística e Experimentação Agronômica, CEP: 37200-000, Lavras, MG, Brasil. E-mail: *cris.hnogueira@yahoo.com.br*

<sup>2</sup>Universidade Federal de Lavras – UFLA, Departamento de Ciências Exatas, CEP: 37200-000, Lavras, MG, Brasil. E-mail: *rrlima@dex.ufla.br / marcelo.oliveira@dex.ufla.br*

distribuições normais não-correlacionadas, proporcionando erros independentes e identicamente distribuídos com média zero e variância constante.

Entretanto, o processo de aleatorização é insuficiente para garantir a independência dos erros  $e$ , por isso, Grondona e Cressie (1991) alegam que existem alguns casos em que, mesmo que a aleatorização seja realizada, as parcelas mais próximas são mais correlacionadas do que as mais distantes, justificando-se a utilização de alguma abordagem que modele essa dependência espacial.

Nesse mesmo contexto, Martinez (1994) argumenta que a proteção dada pela aleatorização não é suficiente, ressaltando, ainda, que mesmo que se cumpram os princípios básicos da experimentação, a correlação espacial implica em violações das pressuposições assumidas no modelo da análise não espacial e, conseqüentemente, em uma análise de variância menos precisa, uma vez que a mesma desconsidera a informação sobre as posições espaciais das unidades experimentais.

Uma vez sendo detectada a estrutura de correlação espacial, a utilização dessa informação na análise estatística fornece estimativas mais eficientes para as médias dos tratamentos, proporcionando resultados mais precisos ou tão bons quanto aqueles obtidos por meio da análise de variância clássica. Entretanto, se tal estrutura de variabilidade for desconsiderada, diferenças reais entre as médias comparadas podem não ser identificadas como significativas. Duarte e Vencovsky (2001) ainda defende a utilização de uma análise que considere essa estrutura de dependência espacial, afirmando que a consideração da informação espacial pode tornar-se um fator decisivo sobre as conclusões de um experimento.

Com isso, o presente trabalho foi realizado com o principal objetivo de descrever como executar a análise de variância com erros espacialmente correlacionados, para experimentos instalados sob o delineamento em blocos casualizados. Nesse processo de desenvolvimento, a matriz de covariância dos erros foi modelada por meio de uma abordagem geoestatística, na qual o semivariograma foi utilizado como principal ferramenta para detectar e descrever a estrutura de variabilidade espacial dos erros.

## 2 A análise geoestatística

Conforme salienta Ribeiro Júnior (1995), a Geoestatística não é um tipo especial, diferente ou alternativo de Estatística. A sua diferença está no fato de que cada observação é descrita não apenas pelo seu valor, mas também por informações de sua posição, expressa por um sistema de coordenadas. Além disso, a suposição de que as observações são independentes e igualmente distribuídas não é feita na Geoestatística, fazendo com que as medidas de correlação entre as observações assumam um papel fundamental na análise.

Na análise geoestatística necessita-se, como restrição máxima, que a hipótese de estacionaridade de 2ª ordem seja atendida. Segundo Journel (1989), uma variável  $Y(x)$  pode ser considerada estacionária de 2ª ordem se:

(i) a esperança de  $Y(x)$  existe e não depende da posição  $x$ , ou seja:

$$E[Y(x)] = \omega = \text{constante}; \quad (1)$$

(ii) a função de covariância  $C(Y(x+h), Y(x))$  existe e depende apenas da distância ( $h$ ) entre os pontos. Dessa forma,

$$C(Y(x+h), Y(x)) = E[Y(x+h) \cdot Y(x)] - \omega^2 = C(h). \quad (2)$$

De acordo com a definição de estacionariedade de 2ª ordem, pode-se observar, ainda, que a variância  $Var(Y(x))$  de uma variável  $Y(x)$  é um caso particular da covariância quando  $h = 0$ :

$$Var(Y(x)) = C(Y(x+0), Y(x)) = C(0). \quad (3)$$

Na abordagem espacial, a medida normalmente utilizada para descrever e modelar as relações entre as distâncias e a associação espacial das variáveis é a semivariância, que pode ser definida como uma medida do grau de dependência espacial entre duas variáveis. De acordo com Journel e Huijbregts (1978), se a hipótese de estacionariedade de 2ª ordem for atendida, é possível estabelecer uma relação entre a covariância  $C(h)$  e a semivariância  $\gamma(h)$ , que é dada por

$$C(h) = C(0) - \gamma(h). \quad (4)$$

Se uma variável é estacionária de segunda ordem, uma outra medida que pode ser utilizada é a correlação ( $\rho$ ), a qual é definida como sendo a razão entre a covariância dos valores assumidos pela variável  $Y$ , nas posições  $x$  e  $x+h$  e a variância dessa variável, em função da distância  $h$ . Dessa forma, tem-se

$$\rho(h) = \frac{C(h)}{C(0)} = \frac{C(0) - \gamma(h)}{C(0)} = 1 - \frac{\gamma(h)}{C(0)}. \quad (5)$$

O estimador de semivariância mais utilizado na literatura, proposto por Matheron em 1962, é conhecido como estimador 'clássico' ou, simplesmente, estimador de Matheron. Conforme afirma Cressie (1993), este estimador pode ser definido por meio da seguinte expressão:

$$\hat{\gamma}(h) = \frac{1}{2n(h)} \sum_{i=1}^{n(h)} [y(x_i) - y(x_i+h)]^2, \quad (6)$$

em que

$\hat{\gamma}(h)$  é o estimador de semivariância;

$n(h)$  representa o número de pares de valores medidos separados por uma distância  $h$ ;

$y(x_i)$  e  $y(x_i + h)$  são realizações da variável aleatória  $Y$ , nas coordenadas  $x_i$  e  $x_i + h$ , de tal modo que esses pontos estão separados por uma distância  $h$ .

Um semivariograma típico com características ideais, cuja variável atende à estacionariedade de segunda ordem, pode ser ilustrado conforme a Figura 1.

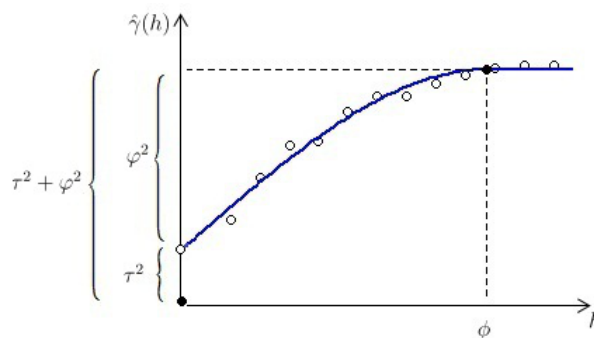


Figura 1 - Exemplo de um semivariograma experimental com comportamento ideal.

Na Figura 1 estão apresentados ainda os principais parâmetros do semivariograma. Segundo Journel e Huijbregts (1978), o alcance ( $\phi$ ) é distância dentro da qual as amostras apresentam-se correlacionadas espacialmente, o patamar ( $\tau^2 + \varphi^2$ ) é o valor da semivariância na distância ( $\phi$ ), isto é,  $\gamma(\phi) = \tau^2 + \varphi^2$ , que corresponde, teoricamente, à variância dos dados. Já o efeito pepita ( $\tau^2$ ) representa a descontinuidade do semivariograma para distâncias menores do que a menor distância observada na amostra. Está associado à variabilidade totalmente aleatória dos dados, ou seja, refere-se à variância do erro experimental. Por fim, a contribuição ( $\varphi^2$ ) é o valor associado à variância estrutural, ou seja, à variabilidade regionalizada que pode ser explicada por um modelo com uma certa estrutura de dependência espacial.

### 3 O modelo linear

Um modelo linear, de forma geral, pode ser representado por:

$$Y = \mathbf{X}\theta + \varepsilon; \quad (7)$$

em que

$Y$  é o vetor de observações de ordem  $n \times 1$ ;

$\mathbf{X}$  é a matriz de incidência dos parâmetros do modelo de ordem  $n \times p$ , em que  $p$  é o número de parâmetros presentes no modelo;

$\theta$  é o vetor de parâmetros do modelo de ordem  $p \times 1$ ;

$\varepsilon$  é o vetor de erros aleatórios de ordem  $n \times 1$ , sendo que, frequentemente, assume-se que  $\varepsilon \sim N(\mathbf{0}, \mathbf{\Sigma})$ .

Neste trabalho, a matriz de covariância dos erros é dada por  $\mathbf{\Sigma} = \mathbf{V}\sigma^2$ , em que  $\sigma^2 = \tau^2 + \varphi^2$  representa o patamar do semivariograma ajustado aos erros do modelo e  $\mathbf{V}$  é uma matriz simétrica positiva definida conhecida.

Segundo Searle (1987), convenientemente, um modelo linear pode ser escrito de forma particionada. O modelo particionado para um delineamento em blocos casualizados é dado por:

$$Y = \mathbf{X}_1\mu + \mathbf{X}_2\alpha + \mathbf{X}_3\beta + \varepsilon, \quad (8)$$

sendo  $Y$  o vetor de  $n = tr$  observações,  $\mu$  uma constante inerente a cada observação,  $\alpha$  um vetor  $t \times 1$  que contém os efeitos de tratamento,  $\beta$  um vetor  $r \times 1$  contendo os efeitos de bloco,  $\mathbf{X}_1$  um vetor de uns,  $\mathbf{X}_2$  a matriz de incidência dos efeitos de tratamento de ordem  $n \times t$ ,  $\mathbf{X}_3$  a matriz de incidência dos efeitos de bloco de ordem  $n \times r$ , de tal modo que  $\mathbf{X} = [\mathbf{X}_1 : \mathbf{X}_2 : \mathbf{X}_3]$  e  $\theta' = [\mu : \alpha' : \beta']$  e, por fim,  $\varepsilon$  representa um vetor de erros.

Essa partição do modelo é útil para definir as somas de quadrados da análise de variância.

### 3.1 A análise de variância

Segundo Gotway e Cressie (1990), na presença de correlação espacial, os testes usuais para verificar a presença dos efeitos do tratamento podem não ser válidos e a aplicação dos mesmos pode levar a conclusões errôneas sobre o experimento. De acordo com esses autores, isso ocorre pelo fato de que, na presença de correlação espacial, a estatística obtida no teste  $F$  não segue uma distribuição  $F$  de Snedecor.

Para que fosse possível analisar experimentos com dependência espacial, esses autores propuseram e utilizaram uma análise de variância que considerava as informações sobre a correlação espacial entre as observações, de modo que, os procedimentos usuais de inferência para testar hipóteses de igualdade entre as médias dos tratamentos eram corretos.

Em experimentos balanceados, as somas de quadrados que compõem a análise de variância, considerando a existência de covariância não nula entre os erros, são dadas por:

$$SQMédia = Y'\mathbf{V}^{-1}\mathbf{X}_1(\mathbf{X}_1'\mathbf{V}^{-1}\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{V}^{-1}Y;$$

$$SQTrat = Y'\mathbf{R}_v\mathbf{X}_2(\mathbf{X}_2'\mathbf{R}_v\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{R}_vY;$$

$$SQBloco = Y'\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}Y - Y'\mathbf{R}_v\mathbf{X}_2(\mathbf{X}_2'\mathbf{R}_v\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{R}_vY + \\ - Y'\mathbf{V}^{-1}\mathbf{X}_1(\mathbf{X}_1'\mathbf{V}^{-1}\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{V}^{-1}Y;$$

$$SQRes = Y'V^{-1}Y - Y'V^{-1}X(X'V^{-1}X)^-X'V^{-1}Y;$$

$$SQTotal_{corrigida} = Y'V^{-1}Y - Y'V^{-1}X_1(X_1'V^{-1}X_1)^-X_1'V^{-1}Y;$$

$$\text{em que } R_v = V^{-1} - V^{-1}X_1(X_1'V^{-1}X_1)^-X_1'V^{-1},$$

em que a notação  $(X'V^{-1}X)^-$  representa uma inversa generalizada da matriz  $(X'V^{-1}X)$ .

Com isso, denotando o posto (*rank*) de uma matriz  $X$  por  $r(X)$ , com notação análoga aos demais postos das matrizes, tem-se que a análise de variância que considera a correlação espacial entre os erros pode ser representada, por meio de formas quadráticas, conforme mostra-se na Tabela 1.

Tabela 1 - Análise de variância de um DBC com erros espacialmente correlacionados

FV	GL	SQ	QM
Tratamento	$r(X_2) - r(X_1)$	$Y'P_2Y$	$SQTrat/GLTrat$
Bloco	$r(X) - r(X_2)$	$Y'(P - P_1 - P_2)Y$	$SQBloco/GLBloco$
Resíduo	$r(V^{-1}) - r(X)$	$Y'(V^{-1} - P)Y$	$SQRes/GLRes$
Total	$r(V^{-1}) - r(X_1)$	$Y'(V^{-1} - P_1)Y$	

em que

$$P = V^{-1}X(X'V^{-1}X)^-X'V^{-1},$$

$$P_2 = R_v X_2(X_2'R_v X_2)^-X_2'R_v \text{ e}$$

$$P_1 = V^{-1}X_1(X_1'V^{-1}X_1)^-X_1'V^{-1}.$$

É possível encontrar em Nogueira (2013), a demonstração detalhada de que

$$\frac{QMTreat}{QMRes} \sim F_{(\nu_1, \nu_2; \delta_t)} \text{ e } \frac{QMBloco}{QMRes} \sim F_{(\nu_3, \nu_2; \delta_b)},$$

em que  $\nu_1 = r(X_2) - 1$ ,  $\nu_2 = n - r(X)$ ,  $\nu_3 = r(X) - r(X_2)$ ,  $\delta_t = \frac{1}{2\sigma^2}\theta'X'P_2X\theta$

$$\text{e } \delta_b = \frac{1}{2\sigma^2}\theta'X'(R_v - P_2)X\theta.$$

Desse modo, tem-se que, ao utilizar as somas de quadrados ponderados pela matriz  $V$  de correlação, é possível garantir a validade do teste  $F$ , utilizado para comparar a igualdade entre as médias dos tratamentos e entre as médias dos blocos.

## 4 Materiais e Métodos

### 4.1 Simulação dos dados

Os dados analisados foram obtidos a partir de simulação, uma vez que, devido à incipiente difusão dessa abordagem, ainda é pouco frequente encontrar dados referentes a experimentos instalados em blocos casualizados, que possuem georreferenciação. Além disso, considera-se como vantagem de se trabalhar com dados simulados o fato de se conhecer os verdadeiros valores dos parâmetros e, com isso, ser capaz de avaliar a eficiência da abordagem geoestatística na detecção de diferenças entre os tratamentos, uma vez que os dados foram gerados de forma que tais diferenças existissem.

Desse modo, foi considerado um experimento instalado de acordo com um delineamento em blocos casualizados (DBC) composto por 26 tratamentos e 4 blocos. As parcelas experimentais foram distribuídas no campo de acordo com um *grid* irregular, conforme apresentado na Figura 2, sendo consideradas na estimação do semivariograma as distâncias entre os centroides das parcelas.

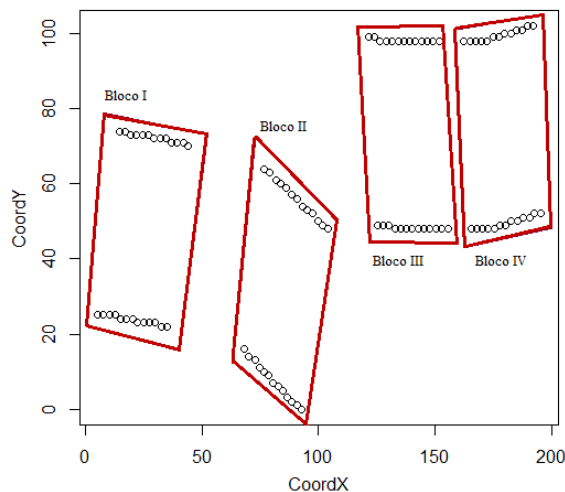


Figura 2 - *Grid* do experimento.

Na simulação dos dados, inicialmente fixou-se os valores dos parâmetros relacionados à constante e aos efeitos de tratamentos e blocos, de forma que o experimento apresentasse diferença entre as médias desses efeitos.

A esses valores fixos foi adicionado um erro aleatório com estrutura de dependência espacial definida, gerado por meio da função *GaussRF* do pacote *Random Fields* (SCHLATHER, 2012). Nesse processo de simulação, considerou-se o erro uma variável aleatória isotrópica, com média zero, seguindo modelo de semivariância esférico. Os valores dos parâmetros do semivariograma foram

definidos de forma que os erros apresentassem forte grau de dependência espacial com alcance médio.

Como a aleatorização não é um princípio obrigatório na análise geoestatística, realizou-se um arranjo sistemático dos tratamentos nas unidades experimentais, sem nenhum critério de aleatorização.

## 4.2 Modelagem geoestatística do erro

A modelagem geoestatística do erro consiste, primeiramente, em obter o vetor de erros estimados  $\hat{\varepsilon}$ , a partir do modelo que supõe independência entre os mesmos. Esses erros representam a diferença entre os valores observados e os estimados, podendo ser calculados, matricialmente, por:  $\hat{\varepsilon} = Y - \mathbf{X}\theta^0$ , em que  $\theta^0 = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'Y$ .

Com os erros estimados, é possível verificar se existe ou não dependência entre os erros. Essa verificação pode ser realizada através da própria análise do semivariograma, de modo que estruturas semelhantes à ilustrada na Figura 1 indicam existência de dependência espacial. Como método auxiliar à análise do semivariograma, pode-se, ainda, construir envelopes, os quais são obtidos por meio da permutação dos valores observados nas posições espaciais. Se algum valor de semivariância não estiver dentro do envelope, considera-se que o semivariograma apresenta uma estrutura de dependência espacial definida e, por isso, os erros são considerados dependentes espacialmente.

Uma vez detectada a estrutura de dependência espacial, a próxima etapa da modelagem da variabilidade espacial do erro consiste em ajustar um modelo teórico ao semivariograma, sendo que, os modelos mais utilizados na literatura são: o esférico, o exponencial e o gaussiano.

A partir do ajuste do modelo teórico de semivariograma é possível, então, obter a matriz  $\Sigma$  de covariância dos erros, que é dada por

$$\Sigma = \begin{bmatrix} \text{Var}(e_1) & C(e_1, e_2) & \cdots & C(e_1, e_n) \\ C(e_2, e_1) & \text{Var}(e_2) & \cdots & C(e_2, e_n) \\ \vdots & \vdots & \ddots & \vdots \\ C(e_n, e_1) & C(e_n, e_n) & \cdots & \text{Var}(e_n) \end{bmatrix};$$

e, assumindo que o erro é uma variável estacionária de segunda ordem, essa matriz  $\Sigma$  corresponde a

$$\Sigma = \begin{bmatrix} \sigma^2 & C(h_{12}) & \cdots & C(h_{1n}) \\ C(h_{12}) & \sigma^2 & \cdots & C(h_{2n}) \\ \vdots & \vdots & \ddots & \vdots \\ C(h_{1n}) & C(h_{2n}) & \cdots & \sigma^2 \end{bmatrix};$$

de modo que  $\sigma^2 = \tau^2 + \varphi^2$  e  $h_{ij}$  é a distância entre os erros  $e_i$  e  $e_j$ , respectivamente alocados nas posições  $x_i$  e  $x_j$ , ou seja,  $h_{ij} = \|x_i - x_j\|$ .



Por meio das relações  $C(h) = C(0) - \gamma(h)$  e  $\rho(h) = \frac{C(h)}{C(0)}$ , em que  $C(0) = \sigma^2$ , é possível reescrever  $\Sigma$  como

$$\Sigma = \mathbf{V}\sigma^2 = \begin{bmatrix} 1 & \rho(h_{12}) & \cdots & \rho(h_{1n}) \\ \rho(h_{12}) & 1 & \cdots & \rho(h_{2n}) \\ \vdots & \vdots & \ddots & \vdots \\ \rho(h_{1n}) & \rho(h_{2n}) & \cdots & 1 \end{bmatrix} \sigma^2;$$

em que  $\mathbf{V}$  é conhecida como uma matriz de correlação.

De posse da matriz  $\mathbf{V}$ , é possível obter o vetor de parâmetros estimados, considerando-se a dependência espacial dos erros que, conforme afirma Graybill (1976), é dado por

$$\theta^0 = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}Y. \quad (9)$$

Uma alternativa para se encontrar resultados mais eficientes é utilizar um algoritmo proposto inicialmente por Pontes (2004), no qual são realizadas iterações até que a média dos erros padrão dos valores preditos atingisse convergência. Neste trabalho, esse processo foi repetido até que as estimativas dos parâmetros do semivariograma atinjissem convergência, sendo considerada uma tolerância de  $10^{-4}$  para o processo iterativo.

Basicamente, este algoritmo contém baseado nos passos apresentados na Figura 3.

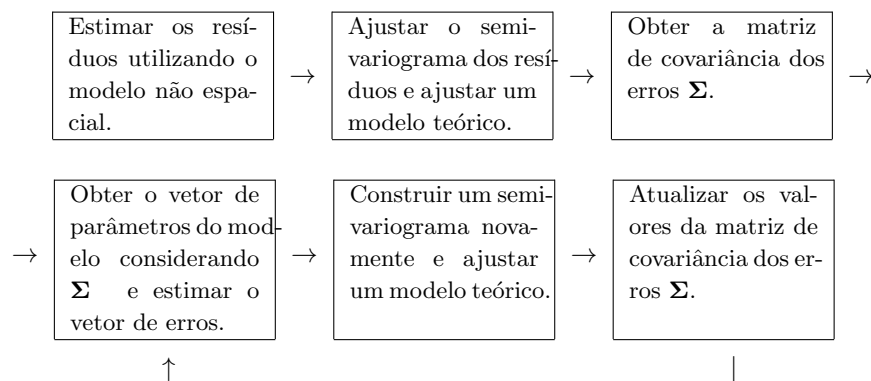


Figura 3 - Algoritmo para estimar os parâmetros do semivariograma.

Tendo em mãos a matriz de covariância dos erros  $\Sigma$ , é possível calcular as somas de quadrados apresentadas na seção (3.1) e, com isso, obter a análise de variância que considera a dependência espacial entre os erros.

## 5 Resultados e discussões

Nesta seção são descritos os passos básicos necessários para a análise espacial de experimentos com erros dependentes, considerando-se os dados simulados conforme apresentado na seção 4.1. Toda a análise desse experimento foi realizada com a ajuda do software *R* (R DEVELOPMENT CORE TEAM, 2012).

### 5.1 Detecção da variabilidade espacial

O primeiro passo da análise espacial de um experimento consiste em verificar se existe variabilidade espacial dos erros. Conforme visto na seção anterior, a verificação da existência dessa dependência espacial pode ser feita por meio da análise gráfica do semivariograma. Esse semivariograma refere-se aos resíduos após a retirada de tendência que, neste caso, trata-se da tendência ocasionada pelos efeitos de tratamentos e blocos, ou seja, os resíduos obtidos na análise clássica.

Como o erro é uma variável isotrópica, assim definido na simulação, construiu-se o semivariograma *omnidirecional*, de forma que, para o cálculo de suas semivariâncias, foram consideradas apenas as distâncias, em qualquer direção, entre dois pontos amostrados.

Com o intuito de contornar os problemas de subjetividade para avaliar a existência de dependência espacial, sugeriu-se, ainda, que fossem construídos envelopes, obtidos por meio da permutação dos valores observados nas posições espaciais. A Figura 4 mostra o semivariograma dos erros e o envelope construído para o mesmo.

Observando-se o semivariograma da Figura 4 percebe-se que o mesmo apresenta uma estrutura de dependência espacial. Essa estrutura é confirmada com a construção do envelope, uma vez que é possível observar a existência de pontos do gráfico que não pertence ao intervalo criado.

### 5.2 Ajuste do semivariograma dos erros

De posse do semivariograma experimental, a etapa seguinte caracteriza-se pelo ajuste de um modelo teórico a esse semivariograma. Esse ajuste foi realizado pelo método de quadrados mínimos ordinários. O modelo ajustado ao semivariograma dos erros foi o modelo esférico, cuja função de correlação é dada por:

$$\rho(h) = \begin{cases} 1 & , \text{ se } h = 0; \\ \frac{\varphi^2}{\tau^2 + \varphi^2} \left[ 1 - 1,5 \left( \frac{h}{\phi} \right) + 0,5 \left( \frac{h}{\phi} \right)^3 \right] & , \text{ se } 0 < h < \phi; \\ 0 & , \text{ se } h \geq \phi. \end{cases} \quad (10)$$

Para se obter estimativas mais eficientes para os parâmetros do semivariograma, foram realizado um processo iterativo conforme representado esquematicamente na Figura 3.

Na Tabela 2 observam-se os valores estimados para os parâmetros do semivariograma a cada iteração, de modo que  $\varphi^2$  representa a contribuição,  $\tau^2$  é o efeito pepita e  $\phi$  representa o alcance.

Tabela 2 - Parâmetros estimados do semivariograma a cada iteração

Parâmetros	Iterações					
	1 <sup>a</sup>	2 <sup>a</sup>	3 <sup>a</sup>	4 <sup>a</sup>	5 <sup>a</sup>	6 <sup>a</sup>
$\varphi^2$	1,1062	1,8008	1,9458	1,9606	1,9616	1,9616
$\tau^2$	0,4647	0,1406	0,1073	0,1056	0,1055	0,1055
$\phi$	22,1173	22,5271	22,2722	22,2354	22,2327	22,2327

De acordo com essa tabela, é possível observar que, a partir da 5<sup>a</sup> iteração, os valores das estimativas dos parâmetros do semivariograma se estabilizaram e, portanto, os valores estimados correspondentes a essa iteração foram utilizados para a modelagem da matriz de covariância dos erros.

O ajuste do modelo teórico esférico, utilizando-se os valores dos parâmetros estimados na 6<sup>a</sup> iteração, pode ser observado na Figura 5.

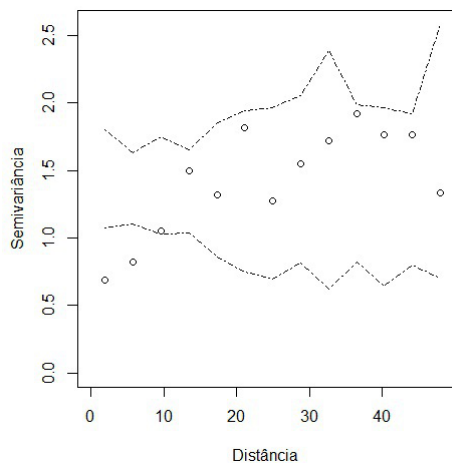


Figura 4 - Semivariograma dos erros contendo o envelope simulado.

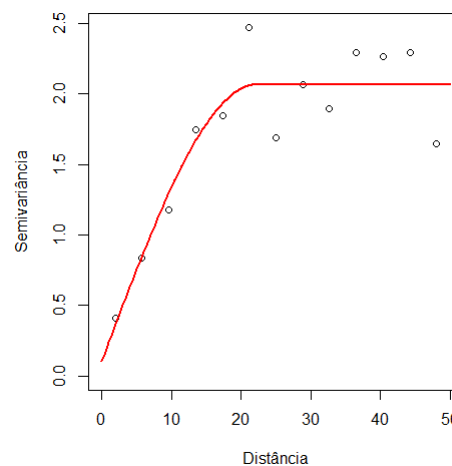


Figura 5 - Ajuste do modelo esférico ao semivariograma dos erros.

### 5.3 Obtenção da análise de variância

Com base no ajuste do modelo esférico de semivariância, apresentado na Figura 5, tem-se que:

$$\hat{\rho}(h) = \begin{cases} 1 & , \text{ se } h = 0; \\ 0,94896 - 0,06402h + 2,3493 \times 10^{-5}h^3 & , \text{ se } 0 < h < 22,2327; \\ 0 & , \text{ se } h \geq 22,2327. \end{cases} \quad (11)$$

Com isso, é possível estimar a matriz de correlação  $\hat{\mathbf{V}}$  e, conseqüentemente, calcular as somas de quadrados da análise de variância que considera a estrutura de dependência espacial dos erros.

Essa análise de variância, apresentada na Tabela 3, foi construída a partir das expressões contidas na Tabela 1.

Tabela 3 - Análise de variância, considerando erros correlacionados

FV	GL	SQ	QM	$F_0$	<i>p-valor</i>
Tratamento	25	114,294	4,572	1,728	0,0367
Blocos	3	6,207	2,069	0,782	0,5076
Resíduo	75	198,412	2,645		
Total	103	318,913			

Observando-se os dados da Tabela 3 é possível concluir que os efeitos de tratamentos foram significativos ao nível de 5% de probabilidade, o que implica dizer que as médias dos tratamentos não são todas iguais, ou seja, pode-se dizer que existe diferença entre as médias dos tratamentos com um nível de confiança de 95%. Já em relação ao fator blocos, o que se pode concluir é que não existe diferença significativa entre as médias dos blocos.

No intuito de comparar a abordagem clássica e a espacial, foi construída a análise de variância clássica, cujas somas de quadrados foram calculadas por meio das expressões usuais, desconsiderando a informação espacial dos dados. Na Tabela 4 apresenta-se essa análise de variância.

Tabela 4 - Análise de variância, considerando erros independentes

FV	GL	SQ	QM	$F_0$	$p$ -valor
Tratamento	25	30,508	1,2203	0,6428	0,8924
Blocos	3	20,674	6,891	3,630	0,017
Resíduo	75	142,372	1,8983		
Total	103	193,554			

De acordo com a análise de variância da Tabela 4, o que se pode observar é que, ao desconsiderar a informação sobre a dependência espacial dos erros, os efeitos de tratamentos foram considerados não significativos e, conseqüentemente, aceitava-se a hipótese de igualdade entre as médias dos tratamentos. Já os efeitos de blocos foram considerados significativos ao nível de 5% de probabilidade, o que significa dizer que as médias dos blocos apresentavam diferenças significativas.

Ao comparar os resultados obtidos, é possível observar que a utilização da informação espacial foi capaz de detectar diferenças significativas entre as médias dos tratamentos, as quais eram consideradas estatisticamente iguais quando se analisava um modelo que supunha independência dos erros. Desse modo, pode-se dizer que a modelagem espacial dos erros produziu estatísticas mais eficientes para avaliar a igualdade entre as médias dos tratamentos.

Além disso, pode-se notar ainda que a estatística  $F_0$ , utilizada para comparar as médias dos blocos, diminuiu quando a informação espacial dos erros foi considerada. Um dos motivos que podem justificar esse fato é que a modelagem da dependência espacial dos erros tenha conseguido explicar os efeitos referentes aos blocos e, desse modo, a variabilidade explicada por esse fator deixou de ser significativa, resultando em um valor menor para  $F_0$  quando se considerou a estrutura de dependência espacial.

Resultados similares foram encontrados Grondona e Cressie (1991) ao analisar um em experimento simulado em blocos casualizados com 20 tratamentos e blocos, também mostraram os benefícios da abordagem espacial, uma vez que a modelagem da dependência espacial dos seus dados proporcionou uma análise cerca de 30% mais eficiente com relação às diferenças entre os tratamentos.

Nesse mesmo contexto, ao analisar dados de infiltração de água no solo, Gotway e Cressie (1990) mostraram as diferenças encontradas no teste  $F$  da análise de variância, considerando os modelos clássico e espacial, e concluíram que, na presença de dependência espacial, a utilização do teste  $F$  supondo independência dos erros pode levar a conclusões equivocadas sobre os resultados do experimento.

Esses autores ainda avaliaram contrastes entre os tratamentos e, puderam observar que, ao adotar a modelagem clássica na análise da infiltração de água no solo, estariam propensos a concluir diferenças de tratamento de média significativas quando, na verdade, tais diferenças não existiam.

## Conclusões

Na análise de experimentos, mesmo que os erros apresentem-se espacialmente correlacionados, é possível continuar desfrutando da análise de variância como ferramenta para comparar as diferentes fontes de variabilidade. No entanto, para que isso ocorra, é necessário que a matriz de correlação dos erros seja considerada como ponderadora das somas de quadrados.

Diante do que foi exposto, entende-se que é realmente necessária a difusão da abordagem da análise espacial de experimentos, uma vez que a consideração da informação sobre a dependência espacial dos erros pode alterar completamente as conclusões sobre um determinado experimento.

## Agradecimentos

Agradecimento à Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) pelo apoio financeiro.

NOGUEIRA, C. H.; LIMA, R. R.; OLIVEIRA, M. S. Use of information spatial in the analysis of variance. *Rev. Bras. Biom.*, São Paulo, v.31, n.3, p.408-422, 2013.

■ **ABSTRACT:** *One of the essential principles of the experimentation is the randomization which contributes towards the presuppositions which the errors should be both equally and identically distributed, are met. Nevertheless, experiments are found of which errors present a definite spatial dependence structure. A way to bypass that trouble is to utilize a spatial approach in which it is possible to estimate and model the spatial correlation among the errors. So, the objective of this work was to report how to carry out the variance analysis with spatially correlated errors in which the error covariance matrix, modeled through a geostatistical approach, was utilized as a weighting factor of the sums of squares of the variance analysis. The results obtained showed that spatial error correlation modeling proved more efficient, since it produced greater values to the F statistic, if compared with the values obtained by the model which supposed spatial error independence.*

■ **KEYWORDS:** *Covariance; analysis experiments; semivariogram; dependent errors.*

## Referências

CRESSIE, N. *Statistics for spatial data*. 2<sup>nd</sup> ed. New York: J. Wiley, 1993. 900p.

DUARTE, J. B.; VENCOVSKY, R. Estimação e predição por modelo linear misto com ênfase na ordenação de médias de tratamentos genéticos. *Scientia Agrícola*, v.58, n.1, p.109-117, 2001.

- GOTWAY, C. A.; CRESSIE, N. *A Spatial Analysis of Variance Applied to Soil-Water Infiltration*, Water Resources Research, 26, 2695-2703, 1990.
- GRAYBILL, F. A. *Theory and application of the linear model*. Boston: Duxbury, 1976. 704p.
- GRONDONA, M. O.; CRESSIE, N. *Using spatial considerations in the analysis of experiments*. Technometrics, Alexandria, v. 33, n. 4, p. 381-392, 1991.
- JOURNEL, A. G. *Fundamentals of geostatistics in five lessons*. Washington: American Geophysical Union, 1989. 40p.
- JOURNEL, A. G.; HUIJBREGTS, C. J. *Mining geostatistics*. London: Academic Press, 1978. 600 p.
- MARTINEZ, R. *Control de la correlación espacial en experimentos de campo em el sector agrícola*. Agronomía Colombiana, Bogotá, v. 11, p. 83-89, 1994.
- NOGUEIRA, C. H. *Análise de variância com dependência espacial sob uma abordagem geoestatística*. 2013. 123 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) – Universidade Federal de Lavras, Lavras, 2013.
- PONTES, J. M.; OLIVEIRA, M. S. *Uma proposta alternativa para a análise de experimentos de campo utilizando a geoestatística*. Ciênc. agrotec. vol.28, n.1, p. 135-141, 2004.
- R DEVELOPMENT CORE TEAM. *R: a language and environment for statistical computing*. Version 2.14.2. Vienna: R Foundation for Statistical Computing, 2012. Disponível em: <<http://www.r-project.org>>. Acesso em: 10 mar. 2012.
- RIBEIRO JÚNIOR, P. J. *Métodos geoestatísticos no estudo da variabilidade espacial de parâmetros do solo*. 1995. 99 f. Dissertação (Mestrado em Agronomia) - Escola Superior de Agricultura "Luiz de Queiroz", Piracicaba, 1995.
- SCHLATHER, M. *Random Fields: simulation and analysis of random fields*. R package version 2.0.59. Disponível em: <<http://CRAN.R-project.org/package=RandomFields>>. Acesso em: 10 nov. 2012.
- SEARLE, S. R. *Linear models for unbalanced data*. New York: J. Wiley, 1987. 536p.

Recebido em 08.05.2013.

Aprovado após revisão em 14.11.2013.