

JOSMAR FURTADO DE CAMPOS

EFICIÊNCIA DA ANÁLISE ESTATÍSTICA ESPACIAL NA CLASSIFICAÇÃO DE
FAMÍLIAS DO FEIJOEIRO – ESTUDO VIA SIMULAÇÃO

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Estatística Aplicada e Biometria, para obtenção do título de *Magister Scientiae*.

VIÇOSA
MINAS GERAIS – BRASIL
2011

JOSMAR FURTADO DE CAMPOS

EFICIÊNCIA DA ANÁLISE ESTATÍSTICA ESPACIAL NA CLASSIFICAÇÃO DE
FAMÍLIAS DO FEIJOEIRO – ESTUDO VIA SIMULAÇÃO

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Estatística Aplicada e Biometria, para obtenção do título de *Magister Scientiae*.

APROVADA: 24 de fevereiro de 2011.

Prof.: Gerson Rodrigues dos Santos

Prof.: Fernando Luiz Finger

Prof.: Luiz Alexandre Peternelli
(Coorientador)

Prof.: Antônio Policarpo Souza Carneiro
(Orientador)

AGRADECIMENTOS

A Deus nosso criador e arquiteto do universo, pela vida;

À Universidade Federal de Viçosa, em especial ao Programa de Pós-Graduação em Estatística Aplicada e Biometria, pela oportunidade de concluir o curso nesta instituição pública de qualidade reconhecida;

Ao professor Antônio Policarpo Souza Carneiro, pela orientação, ajuda, paciência, apoio e confiança no meu trabalho;

Aos coorientadores José Eustáquio de Souza Carneiro, Luiz Alexandre Peternelli e Paulo Roberto Cecon.

Ao colega e amigo Fernando de Souza Bastos M.Sc., pela ajuda nos meus momentos de maiores dúvidas e dificuldades no decorrer desse curso, independente de horários e dias da semana;

A minha família, pelo total apoio e estímulo;

A todos aqueles que de alguma forma, direta ou indiretamente, contribuíram para o meu crescimento e para o desenvolvimento deste trabalho, meu muito obrigado.

SUMÁRIO

RESUMO	iv
ABSTRACT	vi
1. INTRODUÇÃO	1
2. OBJETIVO GERAL	4
2.1. Objetivos específicos	4
3. REVISÃO BIBLIOGRÁFICA.....	5
3.1. Dependência espacial.....	5
3.2. Variáveis regionalizadas	6
3.3. Variograma.....	8
3.4. Semivariograma	10
3.4.1. Parâmetros do semivariograma.....	11
3.5. Modelos teóricos de semivariograma.....	13
3.6. Análise Estatística Espacial.....	17
3.7. Eficiência da Análise Espacial.....	19
3.8. Acurácia Seletiva	19
4. MATERIAIS E MÉTODOS	23
5. RESULTADOS E DISCUSSÃO	31
6. CONCLUSÕES	50
7. REFERÊNCIAS BIBLIOGRÁFICAS.....	51

RESUMO

CAMPOS, Josmar Furtado de, M.Sc., Universidade Federal de Viçosa, fevereiro de 2011. **Eficiência da análise estatística espacial na classificação de famílias do feijoeiro - estudo via simulação.** Orientador: Antonio Policarpo Souza Carneiro. Coorientadores: Luiz Alexandre Peternelli, Paulo Roberto Cecon e José Eustáquio Souza Carneiro.

O objetivo deste trabalho foi avaliar a eficiência da análise Espacial, que considera erros dependentes espacialmente, para classificação de famílias de feijoeiro em relação às análises tradicionais em blocos casualizados e em látice que assumem erros independentes. Considerou-se diferentes graus de dependência espacial e de precisão experimental. Foram tomados como referência para simulação os resultados de sete ensaios instalados em látice quadrado simples para avaliação genética da produtividade de grãos (g/parcela) de famílias e cultivares de feijoeiro das safras de inverno e das águas de 2007 e 2008. A partir dos resultados apresentados na simulação, foi possível avaliar a qualidade dos respectivos experimentos com base nas diferentes análises (Bloco, Látice e Espacial) e médias simuladas das 100 famílias nos diferentes cenários para Dependência Espacial (DE) e Acurácia Seletiva (AS). No processo de simulação, a média de produção (645 g/parcela), bem como a variância residual (7744,00), foi definida com base nos resultados de análises em blocos de ensaios do programa de melhoramento do feijoeiro da UFV. Para a composição dos cenários simulados foram consideradas quatro magnitudes de dependência espacial (nula, baixa, média e alta), correspondendo aos alcances 0, 25, 50 e 100% da distância máxima entre parcelas. Também foram simuladas três classes de acurácia seletiva (0,95, 0,80 e 0,60), correspondente a precisão experimental muito alta, alta e média, respectivamente. A classificação real das famílias foi utilizada para avaliar a eficiência das metodologias de análise testadas através da correlação de Spearman aplicada às ordens de classificação genotípica e da Eficiência de Seleção entre classificações com base nas metodologias testadas e na classificação real, para a seleção de 10, 20 e 30% das melhores famílias. Para comparar a eficiência de ajuste dos modelos testados, foi utilizado o critério de Informação de Akaike (AIC), baseado em verossimilhança. A análise Espacial apresentou estimativas de variância residual muito próxima da variância residual simulada e maior acurácia seletiva estimada em todos os cenários, indicando maior precisão experimental. Com a redução na acurácia seletiva e aumento na dependência espacial, observou-se maior influência do tipo de análise sobre a classificação das

famílias, sendo que a análise espacial apresentou os melhores resultados, proporcionando seleção mais eficiente das famílias do feijoeiro do que as análises tradicionais em Látice e em Blocos casualizados, principalmente, para seleção de menor número de famílias. Os resultados para acurácia seletiva estimada em função da estatística F foram muito próximos aos obtidos para a correlação de Spearman entre médias estimadas e simuladas para as famílias, indicando que a acurácia seletiva deve ser utilizada como medida de precisão experimental nos ensaios de avaliação genética.

ABSTRACT

CAMPOS, Josmar Furtado de, M.Sc., Universidade Federal de Viçosa, February, 2011. **Efficiency of spatial statistical analysis in the classification of common bean families - the study via simulation** Advisor: Antonio Policarpo Souza Carneiro. Co-advisors: Luiz Alexandre Peternelli, Paulo Roberto Cecon and Jose Eustaquio de Souza Carneiro.

The aim of this study was to evaluate the efficiency of spatial analysis, which considers spatially dependent errors, for classification of common bean families in relation to traditional analysis in randomized blocks and lattice that assuming independent errors. Were considered different degrees of spatial dependence and experimental precision. Were taken as reference to simulate the results of seven experiments carried out in simple square lattice for genetic evaluation of yield (g/plot) of families and bean cultivars of winter crops and water used in 2007 and 2008. From the results presented in the simulation, it was possible to assess the quality of their experiments based on different analysis (Block, lattice and Spatial) and simulated average of 100 families in different scenarios for Spatial Dependence (DE) and Accuracy Selective (AS). In the process of simulation, the average yield (645 g/plot) and the residual variance (7744.00), was defined based on the analysis results of the tests in blocks of bean breeding program at UFV. To make up the four simulated scenarios were considered magnitude of spatial dependence (null, low, medium and high), corresponding to ranges of 0, 25, 50 and 100% of the maximum distance between plots. Were also simulated three classes of selective accuracy (0.95, 0.80 and 0.60), corresponding to the experimental precision very high, high and average, respectively. The actual classification of families was used to evaluate the efficiency of analysis methods tested by Spearman correlation applied to orders and genotypic classification of Selection Efficiency between classifications based on tested methodologies and the actual classification for the selection of 10, 20 and 30% of the best families. To compare the efficiency of adjustment of the models tested, was used the Akaike information criterion (AIC), based on likelihood. Spatial analysis has provided estimates of residual variance very close to the simulated residual variance and higher selective accuracy estimated in all scenarios, indicating greater experimental accuracy. With the reduction in the accuracy and selective increase in spatial dependence, there was greater influence of analysis on the classification of families, and the spatial analysis showed the best results, providing more efficient selection of bean families than traditional analysis of

randomized blocks and lattice, mainly for the selection of fewer families. The results for selective accuracy estimated on the basis of F statistics were very close to those obtained with the Spearman correlation between estimated and simulated averages for families, indicating that the accuracy should be used selectively as a measure of experimental precision tests of genetic evaluation.

1. INTRODUÇÃO

O feijoeiro comum (*Phaseolus vulgaris L.*) é uma planta cultivada a milhares de anos pelo homem. Sua origem até hoje constitui fonte de divergência entre os pesquisadores. Diversas hipóteses tentam explicar não somente a origem da planta, mas também de quando teria o homem começado a utilizá-la como uma cultura doméstica. Algumas evidências levam à hipótese de que o centro de origem da planta e sua domesticação como cultura teriam ocorrido na região da Mesoamérica, por volta de 7000 anos a.C., uma vez que cultivares selvagens, similares a variedades crioulas, foram encontrados nessa região, mais especificamente no México. Supõe-se que a partir dessa região, a cultura teria, posteriormente, disseminado para toda a América do Sul.

A importância do feijão na alimentação humana é comprovada em relatos que remontam aos primeiros registros históricos de que se tem notícia. O feijoeiro era cultivado no Antigo Egito e na Grécia, onde recebia cultos em sua homenagem, por ser considerado símbolo da vida. Já os antigos romanos usavam o feijão em suas festas e até mesmo como forma de pagamento para apostas.

O feijão comum (*Phaseolus vulgaris L.*) destaca-se nos hábitos alimentares nacionais. Possui grande importância na dieta nutricional dos brasileiros no provimento de proteínas, ferro e carboidratos, principalmente para aqueles com carência na ingestão de proteína de origem animal.

O Brasil destaca-se como o maior produtor mundial de feijão. O estado de Minas Gerais é o segundo maior produtor nacional, com aproximadamente 15% da produção total. O país é também o maior consumidor mundial de feijão-comum, com um consumo estimado em aproximadamente 16 kg/ano (CIF, 2011).

Nos programas de melhoramento de feijão, a fase inicial de seleção de famílias envolve grande número de genótipos a ser avaliado, o que exige grandes áreas experimentais. A busca de maior eficiência nos programas de melhoramento genético é constante, devido ao fato de precisar aumentar a produtividade e a estabilidade das espécies cultivadas.

Na fase inicial de seleção, a avaliação das famílias em experimentos com muitas repetições torna-se difícil, tendo em vista a pequena disponibilidade de sementes. Mesmo com pequeno número de repetições, os experimentos dependem de formas mais sofisticadas de planejamento e de análises que levem a estimativas precisas dos parâmetros de interesse. Na maioria das vezes o delineamento em blocos ao acaso torna-se inviável devido à existência de grande heterogeneidade dentro dos blocos.

Nesse caso, o delineamento em látice é muito utilizado em programas de melhoramento de feijão para aumentar a precisão experimental, em avaliações que envolvam grande número de genótipos. Na tentativa de se manter número adequado de repetições, sem sacrificar a precisão experimental, podem ser utilizadas várias alternativas pelos melhoristas, e entre elas destacam-se o método de análise espacial que leva em conta a posição das parcelas na área experimental.

Atualmente algumas formas de modelagens espaciais têm sido sugeridas em experimentos que apresentam dependência espacial por serem mais apropriadas.

A dependência espacial, em experimentos de campo, caracteriza-se pelo fato de que parcelas localizadas próximas umas das outras tendem a exibir resultados mais semelhantes, quanto à variável em estudo, do que parcelas distantes entre si. De acordo com CRESSIE e HARTFIELD (1996) essa dependência espacial tem sido reconhecida desde meados de 1930. Os mesmos autores afirmam que, em ensaios de campo, estimadores precisos de efeitos de tratamentos podem ser obtidos de duas maneiras: a primeira, e mais comum, é utilizar delineamentos experimentais complexos como os delineamentos em blocos, delineamento em linha-coluna, látice, etc.; e a segunda, que não é incompatível com a primeira, é utilizar modelos alternativos e análises que procurem modelar diretamente a heterogeneidade espacial.

Vários autores concluíram que a variabilidade espacial existe, mesmo em áreas consideradas homogêneas. Trabalhos de campos têm mostrado a importância do estudo das variações das condições do solo como aspecto fundamental para se implementar uma agricultura mais eficiente e rentável. Esses trabalhos mostram que a variabilidade do solo não é totalmente aleatória, apresentando autocorrelação espacial.

Segundo FERES (2009), em experimentos instalados em látice para avaliação genética da produtividade em feijoeiro, foi identificada dependência espacial entre resíduos de fraca a forte quando esses experimentos foram analisados em blocos. No entanto, na análise em látice não foi identificada dependência espacial entre resíduos. O tamanho dos experimentos não foi fator determinante da dependência espacial, uma vez que alguns experimentos com grande número de famílias apresentaram dependência fraca, enquanto experimentos com menor número de famílias apresentaram dependência moderada. Os modelos geoestatísticos esférico, exponencial e gaussiano com efeito pepita foram adequados para representar a estrutura de dependência espacial entre resíduos estimados na análise em blocos. Quanto ao ajuste, os três modelos foram bem similares, destacando que o alcance estimado no modelo gaussiano foi inferior aos demais (FERES, 2009).

De acordo com FERREIRA (2001), no melhoramento de plantas, o uso da simulação é justificado quando: (a) as soluções analíticas não existem ou o grau de dificuldade e o número de variáveis envolvidas não permitirem a realização de inferências adequadas sobre o problema; (b) pretende-se comparar a eficiência de um novo procedimento ou técnica em relação a outros já existentes e até mesmo consagrados; (c) os métodos que são rotineiramente empregados em algumas etapas de um programa de melhoramento propiciem qualquer melhoramento ou aumento de sua eficiência. Os primeiros trabalhos de simulação, utilizando recursos computacionais relacionados à genética e melhoramento de plantas, foram apresentados por FRASER (1957a, b).

2. OBJETIVO GERAL

O objetivo deste trabalho foi avaliar a eficiência da análise Espacial, que considera erros dependentes espacialmente, para classificação de famílias de feijoeiro em relação às análises tradicionais em blocos casualizados e em látice que assumem erros independentes.

2.1. Objetivos específicos

Avaliar a eficiência da análise Espacial em relação à precisão experimental.

Avaliar a eficiência da análise estatística espacial na classificação de famílias em ensaios de produtividade do feijoeiro em experimentos montados em látice.

Avaliar se a análise espacial proporciona maior percentagem de famílias selecionadas corretamente do que as análises tradicionais em Látice e em Blocos casualizados.

Verificar se a acurácia seletiva é uma medida eficiente da correlação entre genótipo verdadeiro e estimado.

3. REVISÃO BIBLIOGRÁFICA

Desde o início do século passado, o estudo da variabilidade espacial de algumas características do solo vem sendo uma das preocupações de pesquisadores, de acordo com Smith (1910), Montgomery (1913), Waynick e Sharp (1919), citados por Camargo (1997).

KRIGE (1951), trabalhando com dados de concentração de ouro, concluiu que somente a informação dada pela variância seria insuficiente para explicar o fenômeno em estudo e que se fazia necessário levar em consideração a distância entre as observações. Baseado nas observações de KRIGE (1951), MATHERON (1963, 1971) desenvolve a Teoria das Variáveis Regionalizadas, que é uma função numérica com distribuição espacial, que varia de um ponto a outro com continuidade aparente, onde as variações não podem ser representadas por funções matemáticas simples.

A partir daí surge o conceito da geoestatística, que leva em consideração a localização geográfica e a dependência espacial. A teoria fundamental da geoestatística segundo ISAACS e SRIVASTAVA (1989), é a esperança de que na média, as amostras próximas no espaço e tempo sejam mais similares entre si do que as que estiverem mais distantes. É uma função que varia de um lugar a outro no espaço com certa tendência de continuidade e cujos valores das variáveis amostradas estão relacionados com a posição espacial que ocupam, assumindo assim valores diferentes em diferentes lugares de observação, com certa independência de um lugar a outro (MATHERON, 1965).

3.1. Dependência espacial

A dependência espacial é uma característica inerente à representação de dados através de subdivisões territoriais, ela pode ser entendida como a tendência a que o valor de uma variável em determinada localização assemelha-se mais ao valor de amostras vizinhas do que a demais valores do conjunto amostral. O conceito de vizinhança adotado pode ser formulado com base na propriedade de proximidade, zonas fronteiriças ou distância linear. A necessidade de quantificação da dependência espacial presente num conjunto de geodados levou ao desenvolvimento da chamada estatística espacial cuja principal característica é averiguar padrões espaciais de lugares e valores, identificando a associação espacial existente entre eles e a variação sistemática do fenômeno por localização (ANSELIN, 1992).

O estudo e a determinação da dependência espacial de variáveis aleatórias regionalizadas podem ser realizados utilizando a metodologia geoestatística. Dentro dessa metodologia o semivariograma é muito utilizado na determinação da estrutura de variabilidade espacial e da amplitude da dependência espacial das variáveis em estudo.

3.2. Variáveis regionalizadas

Variável regionalizada pode ser definida como uma variável aleatória que assume diferentes valores Z em função da posição x dentro de uma região S e que representa pares de coordenadas (x_i, y_i) (Figura 1), onde o ponto de referência para o sistema de coordenadas é arbitrário e fixado a critério do interessado (VIEIRA, 2000). Segundo Isaaks e Srivastava (1989), o conjunto de variáveis $Z(x)$ medidas em toda a área S pode ser considerado uma função aleatória $Z(x)$ uma vez que são variáveis aleatórias regionalizadas e assume-se que a dependência entre elas é especificada por algum mecanismo probabilístico.

Uma variável regionalizada (Figura-1) segundo BLAIS e CARLIER (1968) citados por OLEA (1975), é uma função numérica com distribuição espacial que varia de um ponto a outro com continuidade aparente e suas variações não podem ser representadas por uma função matemática simples.

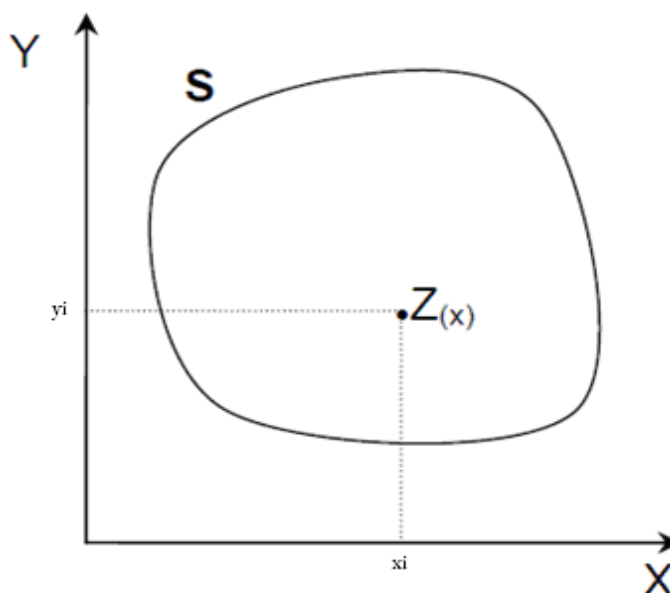


Figura 1 – Variável aleatória regionalizada $Z(x)$

Fonte: GENÚ, A. M.(2004), Geoestatística Multivariada

Ainda segundo OLEA (1975, 1977), as principais características de uma variável aleatória regionalizada são:

- Localização- a localização de uma variável regionalizada é numericamente definida por um valor o qual está associado a uma amostra de tamanho, forma e orientação específicos, sendo essas características geométricas denominadas suporte geométrico, que não necessariamente compreende volume, podendo ser usadas também a áreas e linhas. Quando o suporte geométrico tende a zero, tem-se um ponto ou amostra pontual e o suporte geométrico é imaterial. Por exemplo, para estudar a variação da saturação d'água no solo, são coletadas amostras de 10cm^3 . A variável regionalizada é a umidade do solo e o suporte geométrico é o volume da amostra (10cm^3). Note que, neste experimento, o conteúdo d'água no solo depende não somente da localização da amostra, mas também de seu tamanho, forma e orientação. Uma amostra de forma cilíndrica e longa, tomada na vertical, contém mais água que uma amostra de mesmo tamanho e forma tomada na direção horizontal em relação à superfície do solo. Se o volume da amostra for 10m^3 ao contrário de 10cm^3 , o resultado também será diferente. Resumindo, a teoria das variáveis regionalizadas considera a geometria das amostras, distintamente da estatística clássica onde a forma, o tamanho e a orientação não são considerados. Um experimento estatístico clássico como o lançamento de moedas têm resultados que são independentes se a moeda é grande ou pequena, leve ou pesada, e de como é lançada.
- Anisotropia- algumas variáveis regionalizadas são anisotrópicas, ou seja, apresentam variações graduais numa direção e rápidas ou irregulares em outra.
- Continuidade- dependendo do fenômeno observado, a variação espacial de uma variável regionalizada pode ser pequena ou grande, mas apesar da complexidade das flutuações, uma continuidade média geralmente está presente. Esta continuidade é exemplificada por OLEA (1975) em um caso hipotético, onde amostras de solo de mesmo tamanho, forma e orientação são coletadas em intervalos regulares ao longo de linhas imaginárias. Essas amostras podem originar duas séries distintas para a percentagem de H_2O (água) encontrada, conforme apresentado na (Tabela 1).

Tabela 1 - Percentagem de H₂O em duas amostras distintas A e B

A	5	10	15	20	25	20	15	10	5	% H ₂ O
B	10	25	15	10	20	5	15	5	20	% H ₂ O

(Fonte: OLEA, 1975 apud CAMARGO, 1997)

Nesta tabela, os valores individuais nas duas amostras são exatamente os mesmos. Portanto a média e a variância amostral, assim como o histograma de freqüência da variável observada nas amostras A e B, são rigorosamente idênticos. Qualquer análise que não leve em consideração outras estatísticas além da média, variância e histograma não diferenciará as duas séries. Esse exemplo enfatiza a importância da medida da continuidade espacial da variável regionalizada. Assim, é necessário considerar a posição espacial relativa de cada uma das observações nas duas amostras, para que as mesmas sejam diferenciadas. A continuidade espacial da variável regionalizada pode ser analisada a partir do variograma.

3.3. Variograma

O variograma é uma ferramenta básica de suporte às técnicas de geoestatística, que permite mostrar quantitativamente a variação de um fenômeno regionalizado no espaço (HUIJBREGTS, 1975). Considere duas variáveis regionalizadas, X e Y, onde $X = Z(\mathbf{x})$ e $Y = Z(\mathbf{x}+\mathbf{h})$ (Figura 2), onde \mathbf{x} denota uma posição em duas dimensões, com componentes (x_i, y_i) , e \mathbf{h} um vetor distância (módulo e direção) que separa os pontos.

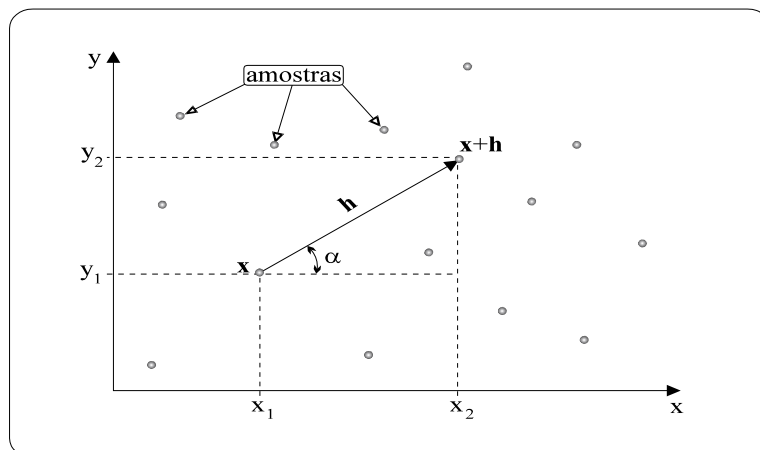


Fig. 2 - Amostragem em duas dimensões.

Fonte: CAMARGO, E. C. G. Geoestatística: fundamentos e aplicações.

O nível de dependência entre essas duas variáveis regionalizadas, X e Y, é representado pelo variograma, $2S(\mathbf{h})$, o qual é definido como a esperança matemática do quadrado da diferença entre os valores de pontos no espaço, separados pelo vetor distância \mathbf{h} , isto é,

$$2 S(\mathbf{h}) = E\{[Z(\mathbf{x})-Z(\mathbf{x}+\mathbf{h})]^2\} = \text{Var}[Z(\mathbf{x})-Z(\mathbf{x}+\mathbf{h})]$$

Através de uma amostra $z(\mathbf{x}_i)$, $i=1, 2, \dots, n$, o variograma pode ser estimado por:

$$2 \hat{S}(\mathbf{h}) = \frac{1}{N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} [z(\mathbf{x}_i) - z(\mathbf{x}_i + \mathbf{h})]^2$$

Em que:

$2\hat{S}(\mathbf{h})$ é o variograma estimado;

$N(\mathbf{h})$ é o número de pares de valores medidos, $z(\mathbf{x}_i)$ e $z(\mathbf{x}_i + \mathbf{h})$, separados por um vetor distância \mathbf{h} ;

$z(\mathbf{x}_i)$ e $z(\mathbf{x}_i + \mathbf{h})$ são valores da i -ésima observação da variável regionalizada, coletados nos pontos \mathbf{x}_i e $\mathbf{x}_i + \mathbf{h}$ ($i = 1, \dots, n$), separados pelo vetor \mathbf{h} .

O variograma é utilizado para calcular os valores de semivariância, para uma dada distância, os quais são necessários para a organização do sistema de equações de krigagem. O variograma substitui a distância euclidiana “h” pela distância “ $2S(h)$ ”, atributo específico do local em estudo. A distância dada pelo variograma mede o grau médio de similaridade entre um valor não amostrado e um valor conhecido vizinho.

3.4. Semivariograma

O semivariograma analisa o grau de dependência espacial entre amostras dentro de um campo experimental, além de definir parâmetros necessários para a estimativa de valores para locais não amostrados, através da técnica de krigagem (SALVIANO, 1996).

O semivariograma é definido como:

$$\hat{S}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(\mathbf{x}_i) - Z(\mathbf{x}_i + \mathbf{h})]^2$$

Um semivariograma experimental com características muito próximas do ideal é apresentado (figura 3) onde seu padrão representa o que, intuitivamente, se espera de dados de campo, isto é, que as diferenças $\{Z(\mathbf{x}_i) - Z(\mathbf{x}_i + \mathbf{h})\}$ decresçam à medida que \mathbf{h} , a distância que os separa decresce. É esperado que observações mais próximas geograficamente tenham um comportamento mais semelhante entre si do que aquelas separadas por maiores distâncias. Dessa maneira, é esperado que $S(\mathbf{h})$ aumente com a distância \mathbf{h} (figura 3).

Segundo DRUCK et al. (2004), o primeiro valor de semivariância calculado, $S(h)$, é denominado "efeito pepita" (C_0) e representa a variabilidade em distâncias menores do que aquelas amostradas. A dependência espacial existe quando há aumento da semivariância até determinada distância, denominada "alcance" (a), que é a distância dentro da qual as amostras encontram-se correlacionadas espacialmente. Após o alcance ocorre uma estabilização da semivariância em um valor denominado "patamar" (C_0+C_1), sendo que após este valor não existe mais dependência espacial entre as amostras.

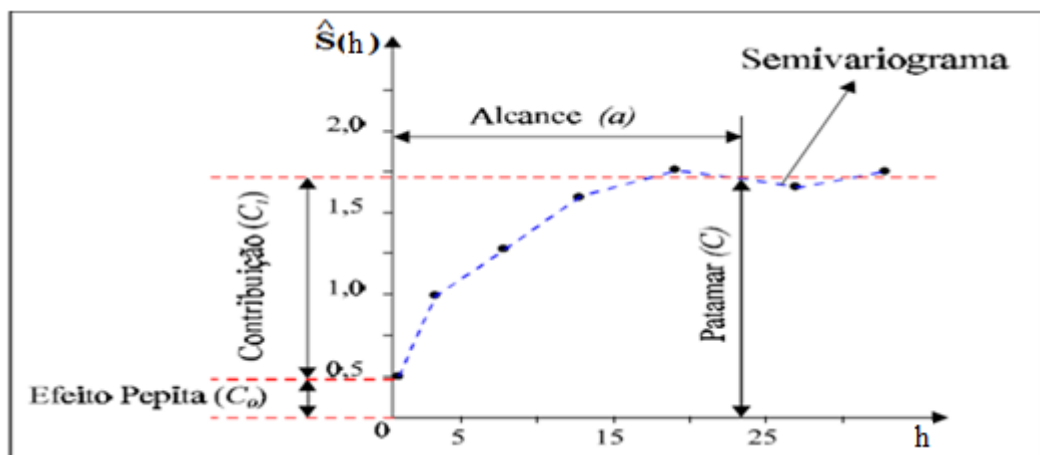


Figura 3 – Exemplo de semivariograma, $\hat{S}(h)$ = semivariância.

Fonte: CAMARGO, E.C.G. Geostatística: Fundamentos e Aplicações.

Assim, para observações separadas por distância d , o semivariograma é um procedimento que é utilizado para medir a dependência entre pontos amostrais separados pela distância d , distribuídos em um sistema espacial de referência. Não há consenso, na literatura, a respeito do número mínimo de pares de pontos amostrais suficientes para a estimativa da semivariância experimental a cada lag (*Lag refere-se a uma distância predefinida, a qual é utilizada no cálculo do semivariograma*). Segundo JOURNAL E HUIJBREGTS (1978) deve-se utilizar um mínimo de trinta pares, enquanto WEBSTER e OLIVER (1990) recomendam um mínimo de cem pares.

3.4.1. Parâmetros do semivariograma

Em um semivariograma (Figura 4), observa-se que as semivariâncias aumentam com o aumento da distância (h) indicando observações menos similares com maiores distâncias e se estabilizam a certa distância chamada de alcance (a), e todos os valores superiores ao alcance possuem distribuição aleatória, portanto são independentes entre si. O valor da $S(d)$ constante é chamado de patamar e é representado pela letra C no semivariograma.

Os parâmetros do semivariograma apresentados nas figuras 4 e 5 são:

- Alcance (a): distância dentro da qual as amostras apresentam-se correlacionadas espacialmente. Nas figuras 4 e 5, o alcance ocorre próximo de 54m.
- Patamar (C): é o valor da semivariância correspondente ao seu alcance (a). Deste ponto em diante, considera-se que não existe mais dependência espacial entre as amostras, porque a variância da diferença entre pares de amostras ($\text{Var} [Z(x) - Z(x+h)]$) torna-se invariante com a distância.
- Efeito Pepita (C_0): à medida que d tende para 0 (zero), $S(h)$ se aproxima de um valor positivo chamado Efeito Pepita (C_0), que revela a descontinuidade do semivariograma para distâncias menores do que a menor distância entre as amostras. Parte desta descontinuidade pode ser também devida a erros de medição (ISAAKS E SRIVASTAVA, 1989), mas é impossível quantificar se a maior contribuição provém dos erros de medição ou da variabilidade de pequena escala não captada pela amostragem.
- Contribuição (C_1): é a diferença entre o patamar (C) e o Efeito Pepita (C_0).

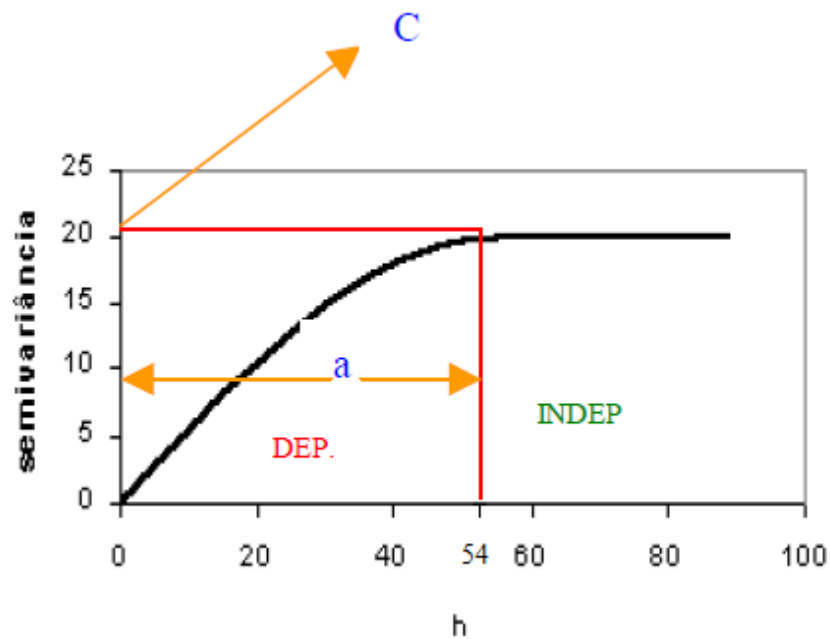


Figura 4 - Semivariograma sem o efeito pepita: C = patamar, a = alcance e h = distância entre observações ($C = 20$, $a = 54$).
 Fonte: GUIMARÃES (2004)

A utilização de dados amostrais na estimativa da semivariância e na construção do semivariograma revela que, freqüentemente, para $h=0$ a semivariância $S(0)$ difere de zero. A impossibilidade de se fazer reamostragem exatamente sobre o mesmo ponto já amostrado e erros como erros de amostragem, erros de análise de laboratório, etc., são justificativas desta descontinuidade na origem. Quando $S(0) \neq 0$ surge um novo termo no semivariograma chamado efeito pepita (C_0) e, neste caso, o patamar é dado por: C_1+C_0 , onde C_1 é a contribuição que é a diferença entre o patamar e o efeito pepita, (Figura 5), (GUIMARÃES, 2004).

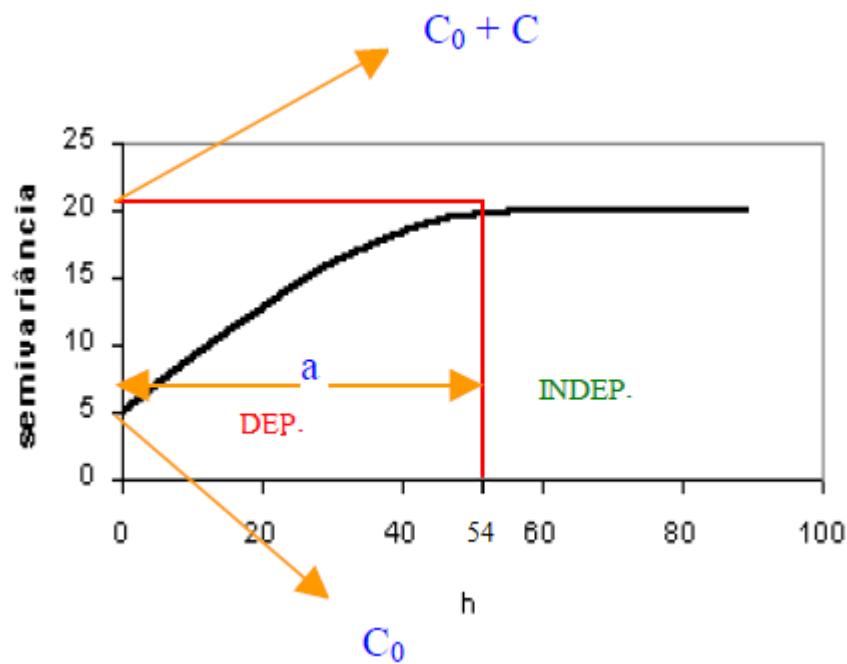


Figura 5 - Semivariograma experimental ou teórico com efeito pepita: C_0 =efeito pepita, C_1 = contribuição, $C+C_0$ = patamar, h = distância entre observações, a = alcance.
 Fonte: GUIMARÃES (2004)

Para a construção do semivariograma, deve-se ajustar uma função contínua ao semivariograma experimental, obtido a partir das amostras colhidas no campo, definindo os parâmetros contribuição (C_1), alcance (a) e efeito pepita (C_0). As funções de semivariogramas mais utilizadas são os modelos geoestatísticos: linear, exponencial, gaussiano e esférico (GUIMARÃES, 2004).

3.5. Modelos teóricos de semivariograma

ISAAKS E SRIVASTAVA (1989) apresentam os modelos básicos denominados modelos isotrópicos em que o semivariograma é idêntico para qualquer direção de h e estão divididos em dois tipos: com patamar e sem patamar. Modelos com patamar, que são referenciados na geoestatística como modelos transitivos onde alguns dos modelos atingem o patamar C assintoticamente e o alcance a é arbitrariamente definido como a distância correspondente a 95 % do patamar.

Modelos sem patamar são os que não atingem o patamar e que continuam aumentando à medida que a distância aumenta. Esses modelos são utilizados para modelar fenômenos que possuem infinita capacidade de dispersão.

Os modelos transitivos mais utilizados são: modelo esférico (Sph), modelo exponencial (Exp) e modelo gaussiano (Gau). Para efeito de comparação, estes modelos são apresentados com o mesmo alcance (a) (Figura 6).

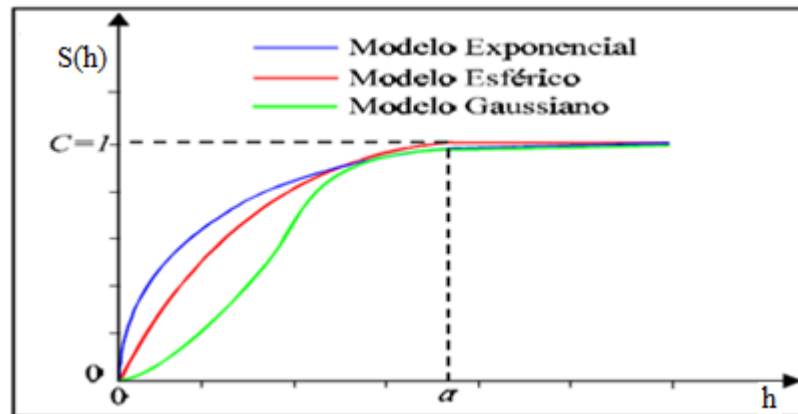


Figura 6 – Representação gráfica dos modelos geostatísticos.

Fonte modificada de Isaaks e Srivastava (1989), p.374.

O modelo esférico é um dos modelos mais utilizados. A equação desse modelo é:

$$S(h) = \begin{cases} 0 & , \quad h = 0 \\ C \left[1,5 \left(\frac{h}{a} \right) - 0,5 \left(\frac{h}{a} \right)^3 \right] & , \quad 0 < h \leq a \\ C & , \quad h > a \end{cases}$$

Outro modelo bastante utilizado é o modelo exponencial, o qual pode ser representado na equação a seguir:

$$S(h) = \begin{cases} 0 & , \quad h = 0 \\ C \left[1 - \exp\left(-\frac{h}{a}\right) \right] & , \quad h \neq 0 \end{cases}$$

Esse modelo atinge o patamar assintoticamente, com o alcance prático definido como a distância cujo valor é 95% do patamar.

O gaussiano é um modelo muitas vezes utilizado para modelar fenômenos extremamente contínuos, semelhante ao modelo exponencial. O modelo gaussiano

atinge o patamar assintoticamente e o parâmetro a é dito como o alcance prático ou à distância na qual o valor do modelo é 95% do patamar (ISAACS E SRIVASTAVA, 1989). Podendo ser representado a seguir pela equação:

$$S(h) = \begin{cases} 0 & , \quad h = 0 \\ C \left[1 - \exp\left(-\frac{h}{a}\right)^2 \right] & , \quad h \neq 0 \end{cases}$$

Na realidade, os modelos de semivariograma experimental possuem valores de efeito pepita (C_0) maiores que zero, (Figura 10).

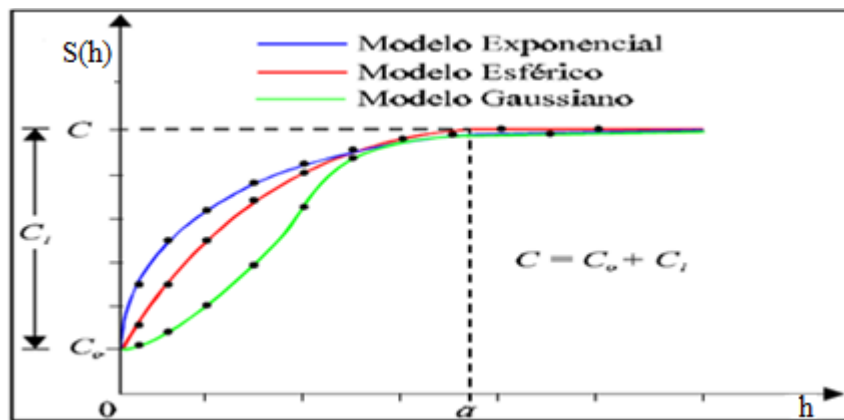


Figura 7 – Representação gráfica dos modelos geostatísticos com efeito pepita.
Fonte: CAMARGO, E. C. G. Geoestatística: fundamentos e aplicações.

Os modelos básicos de semivariograma com efeito pepita são definidos a seguir:

Modelo esférico de semivariograma com efeito

$$S(h) = \begin{cases} 0 & , \quad h = 0 \\ C_0 + C_1 \left[1,5 \left(\frac{h}{a} \right) - 0,5 \left(\frac{h}{a} \right)^3 \right] & , \quad 0 < h \leq a \\ C_0 + C_1 & , \quad h > a \end{cases}$$

Modelo exponencial de semivariograma com efeito

$$S(h) = \begin{cases} 0 & , \quad h = 0 \\ C_0 + C_1 \left[1 - \exp\left(-\frac{h}{a}\right) \right] & , \quad h \neq 0 \end{cases}$$

Modelo gaussiano de semivariograma com efeito

$$S(h) = \begin{cases} 0 & , \quad h = 0 \\ C_0 + C_1 \left[1 - \exp\left(-\frac{h}{a}\right)^2 \right] & , \quad h \neq 0 \end{cases}$$

3.6. Análise Estatística Espacial

A análise tradicional de experimentos de campo parte do princípio de que todas as observações tomadas em posições adjacentes (em plantas ou parcelas vizinhas) são não correlacionadas. Desta forma, a matriz de covariância residual ou dos erros é modelada como $R = I \sigma_e^2$, onde: I é uma matriz identidade e σ_e^2 é a variância residual. Assim, os erros são assumidos como independentes, também, a posição dos tratamentos no campo, ou seja, a distribuição espacial dos mesmos é ignorada. Segundo STEEL & TORRIE (1980), a casualização concorre para a neutralização dos efeitos da correlação espacial e, portanto, permite a geração de uma análise de variância fidedigna. Entretanto, conforme GRONDONA et al. (1996), embora a teoria da casualização enfatize a neutralização da correlação espacial, tal neutralização é mais eficiente quando se usam modelos espaciais. Também as formas de controle local baseadas em blocagem podem ser ineficientes para tratar de problemas de gradientes ambientais e mesmo os blocos incompletos não permitem uma avaliação completa dos efeitos espaciais. Além disso, a blocagem é realizada antes da implantação dos experimentos, de forma que se percebe muitas vezes (principalmente em espécies perenes), por ocasião da coleta dos dados experimentais, a presença de manchas ou gradientes ambientais dentro dos experimentos, os quais não foram considerados adequadamente pelos blocos delineados “a priori”. Nesta situação, somente as técnicas de análise espacial, permitem contornar a questão e propiciar uma seleção acurada, através de blocagem “a posteriori” ou através da flexibilização da matriz R baseados nos próprios dados experimentais.

A autocorrelação espacial advinda de observações com dependência espacial compromete as suposições de homogeneidade dentro de blocos e de ausência de interação genótipos x blocos, suposições estas essenciais para se obter eficiência no delineamento experimental e boa capacidade de teste. Assim, a constatação de ocorrência de autocorrelação espacial e a consideração de uma estrutura de erros correlacionados com $R=F$, em que F refere-se a uma matriz não diagonal, podem permitir a obtenção de estimativas e predições mais precisas, aumentando, assim, a eficiência da análise estatística. A consideração desse tipo de estrutura ($R = F$) é também relevante na análise de medidas repetidas tomadas sobre um mesmo indivíduo em plantas perenes, sendo uma opção aos modelos de repetibilidade, multivariado e de regressão aleatória. A autocorrelação espacial é positiva e aumenta com a diminuição da distância entre as observações. Dessa forma, modelos que consideram a correlação entre medidas repetidas como sendo a mesma para quaisquer pares ou combinações de idades

ou safras podem não ser os mais adequados. Assim sendo, os modelos de análise espacial podem ser vantajosos também para a análise de medidas repetidas, ou seja, para análise de dados com dependência temporal. A variabilidade ou heterogeneidade espacial associada à fertilidade e estrutura do solo, umidade, interceptação de luz e outros fatores ambientais contribuem para o aumento da variação residual. Assim, é importante controlar, por delineamento ou por análise, a variação residual (GRONDONA et al.; 1996).

Métodos clássicos de análise estatística de dados geralmente supõem que as realizações das variáveis aleatórias são independentes entre si, ou seja, que observações vizinhas não exercem influências umas sobre as outras.

Propostas para melhorar as estimativas obtidas a partir dos delineamentos em blocos aumentados foram apresentadas com o uso de modelos mistos (SCOTT e MILLIKEN; 1993), (WOLFINGER et al.;1997) e/ou emprego de métodos espaciais (SOUZA et al., 2000; DUARTE; 2000). Os autores supracitados comentam que com a adoção destes métodos de análises há um ganho na precisão das estimativas dos tratamentos regulares, no poder dos testes estatísticos etc. ZIMMERMAN e HARVILLE (1991) elaboraram um método de análise espacial chamado modelo linear de campo aleatório (RFLM - *Random Field Linear Model*), sendo uma metodologia similar à análise de Krigagem usado em geoestatística. Esta proporcionaria estimadores mais exatos que métodos usuais baseados em quadrados mínimos, ou outros métodos como Papadakis (MARTINEZ; 1994).

No método para erros independentes assume-se que há dependência espacial entre os erros aleatórios e que tem distribuição normal com média zero e matriz de variância e covariâncias residuais (R), $\varepsilon_{ij} \sim N(0, R)$. Sendo a matriz de variâncias e covariâncias R definida por:

$$\text{i) } R = I \sigma_e^2 \text{ (modelo com erros independentes);}$$

$$\text{ii) } R = F \sigma_e^2 \text{ (modelo com erros dependentes).}$$

Nesse caso I é a matriz identidade, σ_e^2 é a variância residual no modelo com erros independentes e contribuição no modelo com erros dependentes e F é a matriz formada pelos elementos da função de distância, $f(h)$, que corresponde a modelos geoestatísticos usados para representar a variabilidade espacial.

DUARTE (2000) avaliou esta metodologia em delineamentos de blocos aumentados, utilizando resíduos provenientes de um modelo misto (efeitos aleatórios para tratamentos regulares). Os resultados demonstraram-se superiores à abordagem dos

modelos tradicionais que assumem erros independentes ($R = I \sigma_e^2$), porque proporcionou aumento no poder dos testes estatísticos, alargamento da amplitude das predições genotípicas, e redução do erro padrão das médias e preditores genotípicos. Entretanto, MATASSA et al (2001), a partir de estudo de simulação, demonstrou que recursos de análise espacial obtiveram pequenos ganhos em relação ao DBA apesar das vantagens teóricas apresentadas na literatura.

3.7. Eficiência da Análise Espacial

COSTA e SILVA et al. (2001) realizaram análises espaciais de 12 experimentos com espécies florestais (*Pinus* e *Picea*) empregando o software ASREML. Concluíram que houve melhoras no modelo com a inclusão dos parâmetros espaciais e que as estimativas dos parâmetros de autocorrelação apresentaram alta magnitude, KUSNADAR e GALWEY (2000) também constataram que um modelo espacial propiciou melhor ajuste que o delineamento de blocos incompletos para dados de crescimento individual em *Pinus pinaster*. A abordagem espacial tende a ser vantajosa, mas a sua eficiência não pode ser generalizada, pois a mesma depende de cada situação experimental e do caráter avaliado.

FERES (2009), utilizando dados de produtividade de grãos de família de feijão, safras de inverno 2006/2007 e seca 2007/2008, quando foram analisados os dados de 8 experimentos em látice quadrado simples para a avaliação da produtividade de grãos (g/parcela) em famílias de feijoeiro, concluiu que os métodos de análise estatística espacial avaliados, com o controle local para blocos, apresentaram melhor ajuste que a análise em látice e que esses métodos apresentaram razoável nível de concordância em relação às famílias selecionadas na análise em látice.

3.8. Acurácia Seletiva

A Acurácia de uma estimativa é uma medida da correlação entre o valor estimado e os valores das fontes de informação, ou seja, mede o quanto a estimativa que foi obtida esta relacionada com o "valor real" do parâmetro. Ela informa o quanto o valor estimado é "bom", ou seja, quanto o valor estimado é "próximo" do valor real e nos dá a "confiabilidade" daquela estimativa ou valor.

Segundo RESENDE (2002), nesse contexto, um dos parâmetros mais relevantes para a avaliação da qualidade de um experimento, embora ainda pouco utilizado, é a acurácia seletiva. Esta tem a propriedade de informar sobre o correto ordenamento das cultivares para fins de seleção e, também, sobre a eficácia da inferência acerca do valor genotípico da cultivar, isto é, do seu valor de cultivo e uso (VCU) e ainda quanto maior a acurácia, maior a precisão da seleção. Este parâmetro refere-se à correlação entre o valor genotípico verdadeiro do tratamento genético e aquele estimado ou predito a partir das informações dos experimentos. Como correlação de Spearman varia entre -1 e 1, quanto mais próximo estiver desses extremos, maior será a associação entre as variáveis, considerando que valores adequados de acurácia são aqueles próximos à unidade ou 100%. Logo, é natural que valores elevados de acurácia sejam almejados nos experimentos de avaliação de cultivares. A acurácia é tanto mais alta quanto menor forem os desvios absolutos entre os valores genéticos paramétricos e os valores genéticos estimados ou preditos (HENDERSON, 1984).

Tais desvios podem ser avaliados pela estatística erro quadrático médio (EQM). Este equivale à distância Euclidiana média entre as estimativas e os correspondentes parâmetros, um estimador ou preditor acurado do valor genotípico apresenta vício nulo ou pequeno, e alta precisão (baixa variância do erro de predição). Por isso, minimizar EQM significa maximizar a acurácia, e o método ideal de estimação ou predição é aquele que permite alcançar esse objetivo.

Valores de acurácia seletiva superiores a 80% ($F > 2,78$) devem ser buscados. Ensaios cujo valor de F para cultivares for inferior a 1,56 ($AS < 60\%$) não devem ser aproveitados. Para atingir uma acurácia seletiva (AS) ideal de 95% ou mais, os valores de F para cultivares devem ser iguais ou superiores a 10,26. Por conseguinte, este pode ser um valor de referência para experimentos de avaliação de VCU. Esse valor não depende da espécie cultivada e do caráter sob avaliação, podendo servir como padrão para avaliar experimentos dessa natureza, (RESENDE e DUARTE, 2007).

A estatística F facilita e agrega as informações mais relevantes sobre a qualidade experimental para os efeitos de tratamentos (famílias), que devem ser superiores a 10,26 para se atingir uma meta de acurácia acima de 95% com o valor de CVr (coeficiente de variação relativa) igual a 1,75 (considerando três repetições). A estatística F para os efeitos de cultivares e o coeficiente de variação relativa ($CVr = CVg / CVe$), em conjunto com o número de repetições (r), são indicados para inferir sobre a acurácia seletiva e a precisão em ensaios de avaliação genética.

A acurácia seletiva (AS) pode ser expressa como:

$$AS = \left[\frac{1}{1 + CPe^2 / CVg^2} \right]^{1/2} = \left[\frac{1}{1 + (CVe^2 / CVg^2) / r} \right]^{1/2} = \left[1 - \frac{1}{1 + r(CVr)^2} \right]^{1/2}$$

Em que:

$$CPe = \frac{\sigma_e}{(r)^{1/2} m} \times 100 = \frac{CVe}{(r)^{1/2}} : \text{coeficiente de precisão experimental};$$

σ_e^2 : variância residual ou variância dentro de cultivares;

r: número de repetições ou blocos (assumindo-se os ensaios típicos em blocos completos casualizados, usuais na avaliação de VCU);

m: média geral do experimento;

$$CVg = \frac{\sigma_g}{m} \times 100 : \text{coeficiente de variação genotípica, expresso em porcentagem};$$

σ_g^2 : variância genotípica entre cultivares;

$$CVe = \frac{\sigma_e}{m} \times 100 : \text{coeficiente de variação experimental, expresso em porcentagem};$$

$$CVr = \frac{CVg}{CVe} : \text{coeficiente de variação relativa.}$$

Entretanto, a acurácia seletiva (AS) fica dependente da proporção entre os coeficientes de variação experimental, variação genotípica e do número de repetições. VENCOVSKY (1987) usou da relação entre coeficientes de variação genética e residual para inferir sobre as possibilidades de sucesso no melhoramento genético, entretanto nessa proposta não se vinculou tal relação ao número de repetições.

Segundo RESENDE (2002), a acurácia seletiva (AS) ainda pode ser expressa como:

$$AS = \sqrt{\frac{F-1}{F}}$$

Em que F é a estatística F para efeito de tratamento (cultivares), associada à análise de variância (ANOVA). Assim, um parâmetro adequado para avaliar a qualidade dos experimentos de avaliação de cultivares pode ser sumarizado em termos de uma única estatística em que o coeficiente de variação experimental, o número de repetições e o coeficiente de variação genotípica, são usados simultaneamente.

Ainda podemos expressar o valor de F usando a seguinte expressão, onde os três parâmetros mencionados acima são usados:

$$F = 1 + r(CVr)^2$$

Onde $(CVr)^2 = \frac{CVg^2}{CVe^2}$ e r é o número de repetições

$$(CVr)^2 = \frac{(\sigma_g/m)^2}{(\sigma_e/m)^2}$$

$$(CVr)^2 = \frac{\sigma_g^2}{\sigma_e^2}$$

$$\sigma_g^2 = (CVr \times \sigma_e)^2$$

4. MATERIAIS E MÉTODOS

Neste trabalho a simulação foi realizada no ambiente R (R Development Core Team, 2009). Para a geração de dados de produtividade de feijoeiro, considerou-se diferentes graus de dependência espacial e de precisão experimental. Foram tomados como referência para simulação os resultados de sete ensaios instalados em látice quadrado simples para avaliação genética da produtividade de grãos (g/parcela) de famílias e cultivares de feijoeiro das safras de inverno e das águas de 2007 e 2008 (Tabela 2).

No processo de simulação, a média de produção (645 g/parcela), bem como a variância do erro (7744,00) foi definida com base nos resultados de análises em blocos de ensaios do programa de melhoramento do feijoeiro da UFV (Tabela 2).

Os experimentos usados como referência para a simulação foram conduzidos na estação experimental de Coimbra, pertencente ao Departamento de Fitotecnia da Universidade Federal de Viçosa (UFV). Estes ensaios fazem parte do programa de melhoramento genético do feijoeiro das instituições de pesquisa do estado de Minas Gerais (UFLA, UFV e EPAMIG) em parceria com a EMBRAPA, visando o desenvolvimento e a recomendação de cultivares para as diferentes regiões do estado de Minas Gerais.

Tabela 2 - Média, Quadrado médio de família ($QM_{\text{família}}$), Quadrado médio resíduo (QMR), Estatística F, Coeficiente de variação (CV), Coeficiente de variação relativo (CVr) e Herdabilidade (h^2) de 7 experimentos para a classificação de famílias do feijoeiro usados como referência no processo de simulação considerando a análise em bloco, com 3 repetições

Exp	Látice	Média	$QM_{\text{família}}$	QMR	F_{bloco}	$F_{\text{família}}$	CV	$\hat{\sigma}_F^2$	$\hat{\sigma}_F$	$\hat{\sigma}_r$	CV _r	h^2
1	7x7	709,95	38775,24	5634,09	10,47*	6,88*	10,57	11047,05	105,10	75,06	1,40	0,66
2	20x20	357,74	6126,03	2677,78	22,10*	2,29*	14,47	1149,42	33,90	51,75	0,66	0,30
3	10x10	680,26	41795,788	10763,36	4,35*	3,88*	15,25	10344,14	101,71	103,75	0,98	0,49
4	7x7	764,16	13979,89	5430,35	7,58*	2,57*	9,64	2849,85	53,38	73,69	0,72	0,34
5	6x6	774,36	46333,72	15085,09	8,4*	3,07*	15,86	10416,21	102,06	122,82	0,83	0,41
6	15x15	544,59	42595,39	9285,27	1,51 ^{ns}	4,59*	17,69	11103,37	105,37	96,36	1,09	0,54
7	10x10	680,83	16745,18	8582,33	14,32*	1,95*	13,61	2720,95	52,16	92,64	0,56	0,24
Média		645,00		7744,00						88,00		

Fonte: FERES (2009) - $\hat{\sigma}_F^2$; $\hat{\sigma}_F$ e $\hat{\sigma}_r$: Estimadores para variância e desvio padrão entre famílias e variância residual, respectivamente.

* - significativo a 5% de probabilidade

ns - não significativo

Para a composição dos cenários simulados foram consideradas quatro magnitudes de dependência espacial (nula, baixa, média e alta), correspondendo aos alcances 0, 25, 50 e 100% da distância máxima entre parcelas (Tabela 3). Também foram simuladas três classes de acurácia seletiva (0,60, 0,80 e 0,95), correspondente a precisão experimental muito alta, alta e média respectivamente (Tabela 6).

Tabela 3 -Alcances simulados sendo a distância máxima entre parcelas de 40 metros e respectivas dependências espaciais (DE) entre resíduos

¹ Alcance(%)	Alcance (m)	DE
0	0	nula
25	10	baixa
50	20	média
100	40	Alta

¹ Alcance em percentagem da distância máxima entre parcelas de experimento em látice 10x10 tomado como referência

No processo de simulação os alcances foram usados para obter a estrutura de covariâncias dos erros simulados da seguinte maneira:

$$\text{Cov}(e_i, e_j) = \sigma^2 \left(\exp - \frac{3d}{a_e} \right).$$

Em que:

$\text{Cov}(e_i, e_j)$ é a covariância entre erros estimados;

σ^2 é a variância;

a_e é o alcance efetivo, sendo $a_e = 3a$;

d é distância entre parcelas.

As distâncias entre parcelas (d) foram calculadas a partir do croqui experimental de um dos experimentos usado como referência (Figura 11) coordenadas x e y (Tabela 4) onde foi gerada uma matriz de distância entre parcelas de experimentos em látice 10x10 (Figura 12) pela seguinte equação,

$$d_{ij} = [(x_i - x_j)^2 + (y_i - y_j)^2]^{1/2}.$$

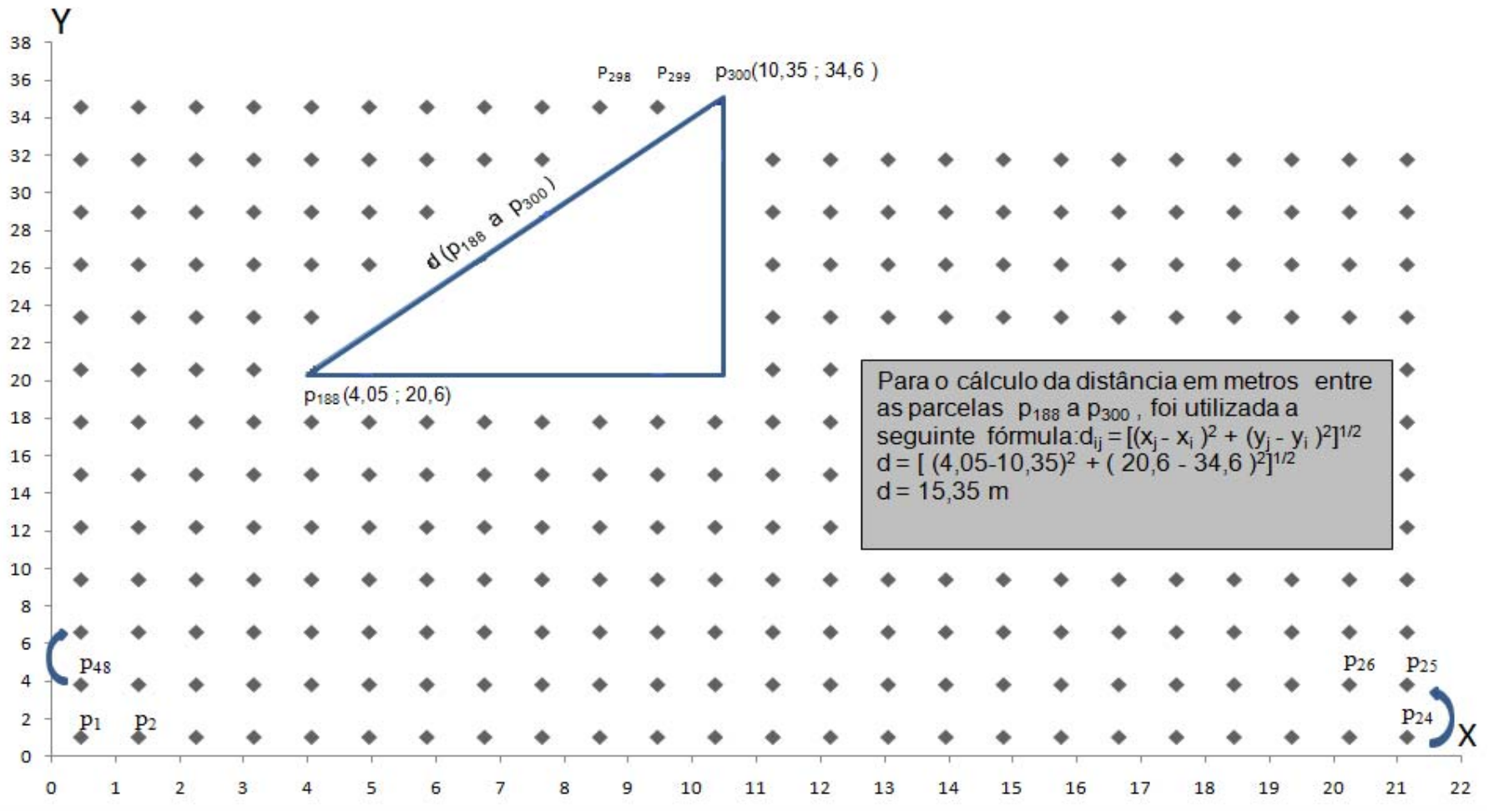


Figura 11-Croqui experimental indicando a posição do centro das parcelas p_1 a p_{300} e cálculo da distância entre parcelas (p_{188} a p_{300}), para experimento em látice 10x10 com 3 repetições, usado como referência no processo de simulação. Experimento conduzido em 2009 na estação experimental de Coimbra/MG, Universidade Federal de Viçosa (UFV).

Tabela 4 - Coordenadas x e y relativa ao centro das parcelas 15 a 34 de experimento em látice 10x10 usado como referência na simulação.

Parc	x	y
p15	13,05	1,00
p16	13,95	1,00
p17	14,85	1,00
p18	15,75	1,00
p19	16,65	1,00
p20	17,55	1,00
p21	18,45	1,00
p22	19,35	1,00
p23	20,25	1,00
p24	21,15	1,00
p25	21,15	3,80
p26	20,25	3,80
p27	19,35	3,80
p28	18,45	3,80
p29	17,55	3,80
p30	16,65	3,80
p31	15,75	3,80
p32	14,85	3,80
p33	13,95	3,80
p34	13,05	3,80

	p15	p16	p17	p18	p19	p20	p21	p22	p23	p24	p25	p26	p27
p15	0,00	0,90	1,80	2,70	3,60	4,50	5,40	6,30	7,20	8,10	8,57	7,73	6,89
p16	0,90	0,00	0,90	1,80	2,70	3,60	4,50	5,40	6,30	7,20	7,73	6,89	6,08
p17	1,80	0,90	0,00	0,90	1,80	2,70	3,60	4,50	5,40	6,30	6,89	6,08	5,30
p18	2,70	1,80	0,90	0,00	0,90	1,80	2,70	3,60	4,50	5,40	6,08	5,30	4,56
p19	3,60	2,70	1,80	0,90	0,00	0,90	1,80	2,70	3,60	4,50	5,30	4,56	3,89
p20	4,50	3,60	2,70	1,80	0,90	0,00	0,90	1,80	2,70	3,60	4,56	3,89	3,33
p21	5,40	4,50	3,60	2,70	1,80	0,90	0,00	0,90	1,80	2,70	3,89	3,33	2,94
p22	6,30	5,40	4,50	3,60	2,70	1,80	0,90	0,00	0,90	1,80	3,33	2,94	2,80
p23	7,20	6,30	5,40	4,50	3,60	2,70	1,80	0,90	0,00	0,90	2,94	2,80	2,94
p24	8,10	7,20	6,30	5,40	4,50	3,60	2,70	1,80	0,90	0,00	2,80	2,94	3,33
p25	8,57	7,73	6,89	6,08	5,30	4,56	3,89	3,33	2,94	2,80	0,00	0,90	1,80
p26	7,73	6,89	6,08	5,30	4,56	3,89	3,33	2,94	2,80	2,94	0,90	0,00	0,90
p27	6,89	6,08	5,30	4,56	3,89	3,33	2,94	2,80	2,94	3,33	1,80	0,90	0,00
p28	6,08	5,30	4,56	3,89	3,33	2,94	2,80	2,94	3,33	3,89	2,70	1,80	0,90
p29	5,30	4,56	3,89	3,33	2,94	2,80	2,94	3,33	3,89	4,56	3,60	2,70	1,80
p30	4,56	3,89	3,33	2,94	2,80	2,94	3,33	3,89	4,56	5,30	4,50	3,60	2,70
p31	3,89	3,33	2,94	2,80	2,94	3,33	3,89	4,56	5,30	6,08	5,40	4,50	3,60
p32	3,33	2,94	2,80	2,94	3,33	3,89	4,56	5,30	6,08	6,89	6,30	5,40	4,50
p33	2,94	2,80	2,94	3,33	3,89	4,56	5,30	6,08	6,89	7,73	7,20	6,30	5,40
p34	2,80	2,94	3,33	3,89	4,56	5,30	6,08	6,89	7,73	8,57	8,10	7,20	6,30

Figura 12 – Dados parciais da matriz de distâncias entre parcelas de experimentos em látice 10x10 usado como referência na simulação.

Segundo HENDERSON (1984), no contexto da avaliação genotípica, o parâmetro estatístico mais importante é a acurácia seletiva (AS). Esse parâmetro refere-se à correlação entre o valor genotípico verdadeiro do tratamento genético e aquele estimado ou predito a partir das informações dos experimentos.

Os experimentos de avaliação do feijoeiro usados como referência no processo de simulação apresentaram precisão experimental de alta a muito alta com acurácia seletiva (AS) variando de 0,70 a 0,92 e CVr de 0,56 a 1,40% (Tabela 5). Assim, optou-se por simular AS de 0,60, 0,80 e 0,95.

Tabela 5 - Acurácia seletiva (AS) para os coeficientes de variação relativa (CVr), estatística F e classe de precisão de 7 experimentos para avaliação genética de famílias do feijoeiro usados como referência para a simulação

Exp	CV _r	F	AS	Classe de precisão
1	1,40	6,88	0,92	Muito alta
2	0,66	2,29	0,75	Alta
3	0,98	3,88	0,86	Alta
4	0,72	2,57	0,78	Alta
5	0,83	3,07	0,82	Alta
6	1,09	4,59	0,88	Alta
7	0,56	1,95	0,70	Alta

Fonte: FERES (2009) – Análise em bloco com 3 repetições

As variâncias devido a família simuladas (Tabela 6) foram obtidas a partir dos valores de AS e respectivos valores da estatística F para família e CVr, (RESENDE et al., 2007) conforme descrito a seguir:

$$F = \frac{1}{1 - AS^2}$$

$$CVr = \sqrt{\frac{F - 1}{r}}$$

$$\sigma^2_{família} = (CVr \times \sigma_e)^2$$

Tabela 6-Valores simulados de Acurácia Seletiva (AS) e respectivos valores da estatística F para efeito de tratamentos (famílias), coeficiente de variação relativa (CVr), variância devido a família ($\sigma_{família}^2$) e variância residual (σ_e^2)

Acurácia seletiva (AS)	Classe de precisão	Valor de F	CVr (r=3)	$\sigma_{família}^2$
0,95	Muito alta	10,26	1,75	23.893,88
0,80	Alta	2,78	0,77	4.589,03
0,60	Média	1,56	0,43	1.452,00

$\sigma_e^2 = 7.744,00$, r é o número de repetições

A partir da combinação de 4 classes de dependência espacial DE (Tabela 3) e 3 classes de AS (Tabela 6), foram obtidas 12 estruturas (cenário) de dados simuladas e cada cenário foi simulado 1000 vezes. Como o efeito das famílias foi simulado, a média de cada família é conhecida sem erro. Assim, a classificação real das famílias foi utilizada para avaliar a eficiência da metodologia de análise testada.

Os dados simulados foram analisados de acordo com os seguintes modelos estatísticos:

(blocos): $y_{ij} = \mu + b_j + g_i + e_{ij}$

(látice): $y_{ijk} = \mu + r_k + b_{j(k)} + g_i + e_{ijk}$

(Espacial): $y_{ij} = \mu + b_j + g_i + \varepsilon_{ij}$

Em que, y_{ij} é o valor observado para o genótipo i, no bloco j; μ é a constante associada a todas as observações; b_j é o efeito fixo do bloco j; g_i é o efeito fixo do genótipo i; e_{ij} , e_{ijk} e ε_{ij} são os erros aleatórios associados às observações, assumindo independência (e_{ij} e e_{ijk}) e dependência espacial (ε_{ij}) entre os erros e y_{ijk} é o valor observado para o genótipo i, no bloco j dentro da repetição k; r_k é o efeito fixo da repetição k; $b_{j(k)}$ é o efeito fixo do bloco j dentro da repetição k.

Os modelos Bloco e Látice assumem observações espacialmente independentes, de forma que $R = I\sigma_e^2$, em que R é a matriz de variâncias e covariâncias, I é a matriz identidade e σ_e^2 é a variância residual no modelo com erros independentes.

O modelo Espacial admite correlação espacial entre observações, de forma que $R = F\sigma_e^2$ em que F é a matriz formada pelos elementos da função de distância f(h), parcialmente apresentada (Figura 12), que corresponde a modelos geoestatísticos usados para representar a variabilidade espacial (STROUP et al., 1994; VIEIRA, 2000).

Nos modelos Bloco, Látice e Espacial, os efeitos dos genótipos avaliados serão considerados fixos. Para estimação dos parâmetros e das médias de famílias foi utilizada a metodologia de máxima verossimilhança, através do PROC MIXED (SAS, 2003).

A ordem de classificação e o processo de seleção das famílias avaliadas com base nas metodologias testadas (Bloco, Látice e Espacial) foram comparados com a classificação real das famílias, uma vez que no processo de simulação a média verdadeira das famílias é conhecida sem erro.

Assim, a classificação real das famílias foi utilizada para avaliar a eficiência das metodologias de análise testadas, através da correlação de Spearman e da percentagem de coincidência entre classificações com base nas metodologias testadas e na classificação real, para a seleção de 10, 20 e 30% das melhores famílias.

Para comparar a qualidade de ajuste dos modelos testados, foi utilizado o critério de Informação de Akaike (AIC), baseado em verossimilhança, que é expresso da seguinte forma:

$$AIC = -2 \ln L + 2p ,$$

em que: p é o número de parâmetros e L o valor da função de verossimilhança considerando as estimativas dos parâmetros. Pelo critério de Akaike, o melhor modelo é o que possuir menor valor absoluto de AIC.

5. RESULTADOS E DISCUSSÃO

Na Tabela 7, são apresentados os resultados relativos à variância do erro estimada nas diferentes análises (Bloco, Látice e Espacial) nos diferentes cenários para Dependência Espacial (DE) e Acurácia Seletiva (AS).

As médias das variâncias do erro para os cenários de alta AS (0,95) e DE nula foram bem próximas nas diferentes análises variando de 7726,66 a 7737,31, (Tabela 7). Com o aumento da DE, observou-se redução na média da variância do erro e aumento do CV, sendo que nas análises em Bloco e, principalmente em Látice, a variância do erro foi bastante subestimada em relação ao valor simulado (7744,00). Comparando os cenários com DE alta e nula, na análise em Bloco, Látice e Espacial a redução na média foi respectivamente de 57, 80 e 14%, enquanto os CV tiveram um aumento de 196,5; 89 e 384%, respectivamente nas análises em Bloco, Látice e Espacial.

Com a redução na AS, não houve praticamente nenhuma mudança nos valores de média da variância do erro. Por exemplo, as médias para análise em Bloco com AS de 0,95, 0,80 e 0,60 e alta DE foram, respectivamente, 3292,96; 3267,96 e 3265,06, e na análise em Látice as médias foram 1545,39; 1558,34 e 1553,56.

Apesar das análises em Bloco e em Látice apresentarem as menores estimativas para variância do erro, essas foram muito inferiores ao valor simulado, diferentemente da análise Espacial que apresentou estimativas muito próximas da variância simulada, mesmo para cenários de alta DE, enquanto variações na AS praticamente não afetaram as estimativas de variância do erro nas diferentes análises.

Tabela 7 – Média, CV (entre parênteses, %) e valor mínimo e máximo (entre colchetes) para variância do erro estimada nas diferentes análises e cenários simulados para Dependência Espacial (DE) e Acurácia Seletiva (AS)

Dep. Espacial	Bloco	Látice	Espacial
AS=0,95			
DE			
Nula	7730,59 (9,94) [5699,79 a 10437,00]	7726,66 (10,81) [5309,51 a 10907,0]	7737,31 (9,94) [5712,02 a 10437,0]
Baixa	6577,49 (17,91) [3794,26 a 10901,0]	4275,43 (16,53) [2404,44 a 6804,74]	7751,46 (22,37) [4311,82 a 14561,0]
Média	5114,72 (23,20) [2293,54 a 9895,93]	2778,43 (17,43) [1529,03 a 4775,07]	7597,43 (34,21) [2763,88 a 16501,0]
Alta	3292,96 (29,48) [1583,23 a 8583,88]	1545,39 (20,45) [834,05 a 2962,29]	6653,16 (48,14) [1776,91 a 17357,0]
AS=0,80			
Nula	7721,74 (9,86) [5417,34 a 10822,0]	7722,56 (10,54) [5148,38 a 11147,0]	7727,95 (9,86) [5423,25 a 10822,0]
Baixa	6574,38 (17,60) [3501,18 a 10935,0]	4330,90 (15,92) [2245,3 a 7259,47]	7765,24 (22,19) [3909,04 a 14091,0]
Média	4996,90 (23,36) [2671,46 a 10765,0]	2725,56 (18,62) [1541,98 a 4713,2]	7393,43 (34,93) [2961,19 a 16101,0]
Alta	3267,96 (27,93) [1637,01 a 7699,59]	1558,34 (19,34) [889,91 a 2914,73]	6421,38 (47,71) [1710,51 a 18258,0]
AS=0,60			
Nula	7760,01 (9,85) [5134,03 a 10487,0]	7756,43 (10,54) [4932,02 a 10874,0]	7766,61 (9,86) [5134,02 a 10489,0]
Baixa	6545,12 (17,77) [3802,09 a 11442,0]	4306,93 (15,39) [2540,44 a 6736,61]	7685,53 (21,80) [4074,84 a 15058,0]
Média	5069,54 (24,23) [2160,76 a 10159,0]	2738,60 (18,59) [1574,05 a 5085,11]	7496,21 (35,03) [2410,26 a 26,1596]
Alta	3265,06 (27,17) [1481,61 a 6889,86]	1553,56 (20,42) [844,41 a 2889,91]	6368,52 (46,42) [1745,39 a 17067,0]

As estimativas para variância do erro e alcance efetivo na análise espacial são apresentadas na Tabela 8. O alcance efetivo é a distância máxima, entre parcelas, onde os erros ainda apresentam-se correlacionados espacialmente.

As médias para os alcances estimados foram muito próximas aos valores simulados 0, 10, 20 e 40 m. Os alcances estimados foram em média 0,45; 9,95; 19,20 e 32,25 respectivamente, para os cenários de DE nula, baixa, média e alta.

Para os cenários de alta, média e baixa AS, respectivamente 0,95, 0,80 e 0,60, e para DE nula, as médias para alcance efetivo variaram entre 0,45 a 0,48. Com o aumento da DE, como esperado, houve aumento na média para o alcance estimado, variando entre 32,54 e 34,25m e que esses valores estão bem próximos aos valores simulados, considerando a distância máxima entre parcelas de 40 metros em experimento. Os coeficientes de variação CV para as diferentes AS tiveram praticamente pouco aumento em relação a AS, variando entre 56,84 a 60,76%, observou-se que com o aumento da DE, houve diminuição no CV de 58% em média nas diferentes AS.

Com a redução na AS, não houve praticamente nenhuma mudança nos valores de média do alcance efetivo, por exemplo, as médias para AS=0,95, 0,80 e 0,60 com DE nula foram respectivamente 0,48, 0,48 e 0,45, já para DE média os valores do alcance efetivo foram 10,00, 10,00 e 9,94, valores bem próximos aos simulados, a AS praticamente não afetou os resultados para as estimativas do alcance efetivo nas diferentes análises.

Tabela 8 - Média, CV (entre parênteses, %) e valores mínimos e máximos (entre colchetes) para variância do erro e alcance efetivo estimados na análise Espacial nos cenários simulados para Dependência Espacial (DE) e Acurácia Seletiva (AS)

Dep.Espacial	Variância	Alcance efetivo (m)
AS=0,95		
DE		
Nula	7737,31 (9,94) [5712,02 a 10437,00]	0,48 (113,73) [0,00 a 3,13]
Baixa	7751,46 (22,37) [4311,82 a 14561,00]	10,08 (25,40) [4,55 a 17,53]
Média	7597,43 (34,21) [2763,88 a 16501,00]	19,42 (35,65) [7,67 a 38,70]
Alta	6653,16 (48,14) [1776,91 a 17357,00]	34,25 (48,54) [8,51 a 77,94]
AS=0,80		
Nula	7727,95 (9,86) [5423,25 a 10822,00]	0,48 (112,64) [0,00 a 2,31]
Baixa	7765,24 (22,19) [3909,04 a 14091,00]	10,01 (39,51) [4,55 a 17,87]
Média	7393,43 (34,93) [2961,19 a 16101,00]	19,10 (37,18) [5,79 a 38,98]
Alta	6421,38 (47,71) [1710,51 a 18258,00]	32,91 (48,61) [7,91 a 76,93]
AS=0,60		
Nula	7766,61 (9,86) [5134,02 a 10489,00]	0,45 (122,01) [0,00 a 2,51]
Baixa	7685,53 (21,80) [4074,84 a 15058,00]	9,94 (25,48) [3,93 a 17,96]
Média	7496,21 (35,03) [2410,26 a 15967,00]	19,36 (36,34) [7,03 a 38,93]
Alta	6368,52 (46,42) [1745,39 a 17067,00]	32,54 (47,87) [8,24 a 77,99]

Para avaliação da qualidade de ajuste dos modelos, utilizou-se o Critério de Informação de Akaike (AIC). Na Tabela 9, são apresentados os resultados relativos ao AIC para as diferentes análises (Bloco, Látice e Espacial). O critério AIC se baseia no valor que maximiza o logaritmo da verossimilhança ajustado para o número de parâmetros do modelo, sendo preferidos os com menores valores absolutos de AIC.

As médias para AIC nos cenários de alta AS (0,95) nas análises em Bloco e Espacial (Tabela 9) foram mais altas do que a análise em Látice, variando de 2056,75 a 2454,00 e de 1901,80 a 2179,43 (Látice). Os coeficientes de variação (CV) para as análises em Bloco e Látice variaram entre 0,84 a 1,79 %, já na análise Espacial foram baixos, variando de 0,80 a 0,98% assim como os valores de máximo e mínimo. Observou-se que com o aumento da DE houve diminuição na média AIC, o que melhorou o ajuste, principalmente para análise Espacial, onde houve diminuição de 16% na média e o menor aumento do CV, oscilando entre 22,5% (AS=0,95) a 25,31% (AS=0,60).

Com a redução na AS, não houve praticamente nenhuma mudança nos valores de AIC, por exemplo, as médias de AIC para AS=0,95, 0,80 e 0,60 com alta DE são respectivamente 2277,75, 2276,80 e 2276,81, portanto a AS não afetou a qualidade do ajuste.

Segundo FERES (2009), de acordo com os critérios de ajuste AIC a utilização de metodologias espaciais promoveu grande redução no valor da função de verossimilhança e tiveram melhor ajuste que as análises em látice e em blocos, indicando melhor ajuste destas metodologias.

Tabela 9 - Média, CV (entre parênteses, %) e valores mínimos e máximos (entre colchetes) para o Critério de Informação de Akaike (AIC) nas diferentes análises e cenários simulados para Dependência Espacial (DE) e Acurácia Seletiva (AS)

Dep.Espacial	Bloco	Látice	Espacial
DE	AS=0,95		
Nula	2453,57 (0,80) [2394,20 a 2514,0]	2179,43 (0,84) [2116,30 a 2239,40]	2454,00 (0,80) [2393,90 a 2514,00]
Baixa	2419,45 (1,45) [2313,60 a 2522,60]	2076,91 (1,35) [1980,8 a 2158,70]	2300,00 (0,87) [2227,8 a 2359,40]
Média	2367,60 (1,90) [2214,25 a 2503,40]	2002,98 (1,47) [1903,40 a 2098,10]	2187,09 (0,89) [2129,60 a 2249,10]
Alta	2277,75 (2,41) [2140,60 a 2475,30]	1901,80 (1,79) [1799,80 a 2016,50]	2056,75 (0,98) [1998,10 a 2120,70]
	AS=0,80		
Nula	2453,36 (0,79) [2384,10 a 2521,20]	2179,38 (0,83) [2111,00 a 2243,10]	2453,84 (0,79) [2386,00 a 2521,20]
Baixa	2419,41 (1,44) [2297,70 a 2523,20]	2079,31 (1,29) [1969,10 a 2169,80]	2301,49 (0,85) [2222,20 a 2366,50]
Média	2363,00 (1,90) [2244,20 a 2520,10]	1999,35 (1,57) [1904,80 a 2095,90]	2185,07 (0,90) [2116,60 a 2248,90]
Alta	2276,80 (2,34) [2147,20 a 2453,80]	1903,62 (1,67) [1810,80 a 2013,70]	2057,49 (0,98) [1996,80 a 2124,60]
	AS=0,60		
Nula	2454,33 (0,80) [2373,50 a 2514,90]	2180,12 (0,83) [2103,70 a 2238,90]	2454,69 (0,79) [2373,50 a 2516,90]
Baixa	2418,54 (1,43) [2314,25 a 2532,20]	2078,47 (1,26) [1990,20 a 2157,00]	2300,47 (0,85) [2238,50 a 2360,20]
Média	2365,53 (1,95) [2202,20 a 2508,60]	2000,23 (1,54) [1908,40 a 2108,90]	2185,11 (0,89) [2107,50 a 2234,90]
Alta	2276,81 (2,32) [2127,40 a 2431,80]	1902,69 (1,79) [1801,90 a 2012,30]	2058,15 (0,99) [1996,20 a 2120,20]

Na Tabela 10, são apresentados os resultados relativos à correlação de Spearman entre médias estimadas nas diferentes análises (Bloco, Látice e Espacial) e médias simuladas das 100 famílias nos diferentes cenários. A correlação de Spearman foi utilizada para comparar a classificação das famílias com base nas médias estimadas com a classificação verdadeira obtida com base nas médias simuladas para cada família.

Para os cenários de alta AS (0,95), as médias para as correlações de Spearman foram altas e apresentaram valores bem próximos, variando de 0,9333 a 0,9934. Os coeficientes de variação (CV) foram baixos, variando de 0,18 a 1,62, assim como os valores mínimos e máximos para as correlações foram bem próximos. Pequenos valores de CV, bem como pequenas diferenças entre valores mínimos e máximos refletem pequena variação entre os resultados obtidos para diferentes repetições de um mesmo cenário simulado. Com o aumento da DE, houve ligeiro aumento na média e ligeira redução no CV para a correlação de Spearman. Embora a análise Espacial tenha apresentado maiores médias e menores CV, estes foram muito próximos aos resultados obtidos nas análises em Bloco e Látice. Assim, para experimentos com alta AS, onde a variância devido a famílias é alta, a classificação das famílias será bem próxima da classificação verdadeira independente do tipo de análise e do grau de DE.

Com a redução na AS de 0,95 para 0,80 e 0,60, observou-se maior influência do tipo de análise sobre a classificação das famílias, sendo que a análise Espacial apresentou os melhores resultados. As médias para as correlações de Spearman foram de 0,8854, 0,9326 e 0,9721, respectivamente para as análises em Bloco, Látice e Espacial nos cenários com AS de 0,80 e alta DE. Para cenários com baixa AS (0,60) e alta DE, a eficiência da análise Espacial foi ainda maior, sendo as médias de 0,7371, 0,8287 e 0,9232, respectivamente para as análises em Bloco, Látice e Espacial (Tabela 10). Nos cenários com baixa DE, a análise Espacial também apresentou as maiores médias, porém, as magnitudes das diferenças foram menores (0,6079, 0,6603 e 0,7773, respectivamente para as análises em Bloco, Látice e Espacial nos cenários com AS de 0,60).

O comportamento do CV e dos valores mínimos e máximos das correlações de Spearman foi semelhante ao observado para as médias (Tabela 10).

Tabela 10 - Média, CV (entre parênteses, %), valores mínimos e máximos (entre colchetes) para correlação de Spearman entre médias estimadas nas diferentes análises e médias simuladas das famílias nos cenários para Dependência Espacial (DE) e Acuraria Seletiva (AS)

Dep. Espacial	Bloco	Látice	Espacial
DE	AS=0,95		
Nula	0,9403 (1,44) [0,8812 a 0,9731]	0,9333 (1,62) [0,8756 a 0,9739]	0,9400 (1,45) [0,8870 a 0,9731]
Baixa	0,9490 (1,39) [0,8902 a 0,9828]	0,9610 (1,04) [0,9154 a 0,9838]	0,9789 (0,52) [0,9592 a 0,9915]
Média	0,9597 (1,27) [0,8924 a 0,9832]	0,9740 (0,71) [0,9471 a 0,9877]	0,9883 (0,29) [0,9738 a 0,9951]
Alta	0,9722 (0,99) [0,9149 a 0,9912]	0,9842 (0,47) [0,9621 a 0,9933]	0,9934 (0,18) [0,9853 a 0,9971]
	AS=0,80		
Nula	0,7808 (5,39) [0,6280 a 0,8771]	0,7615 (6,19) [0,5847 a 0,8756]	0,7799 (5,46) [0,6084 a 0,8753]
Baixa	0,8017 (5,82) [0,5885 a 0,9057]	0,8394 (4,45) [0,6798 a 0,9279]	0,9076 (2,47) [0,7927 a 0,9623]
Média	0,8420 (4,82) [0,6292 a 0,9346]	0,8925 (2,86) [0,7824 a 0,9541]	0,9491 (1,27) [0,8962 a 0,9806]
Alta	0,8854 (3,87) [0,7582 a 0,9336]	0,9326 (2,03) [0,8526 a 0,9672]	0,9721 (0,73) [0,9443 a 0,9863]
	AS=0,60		
Nula	0,5757 (12,54) [0,3085 a 0,7734]	0,5515 (13,41) [0,2870 a 0,7233]	0,5748 (12,56) [0,2980 a 0,7734]
Baixa	0,6079 (11,94) [0,3347 a 0,7897]	0,6603 (10,07) [0,4048 a 0,8252]	0,7773 (5,93) [0,5928 a 0,9009]
Média	0,6600 (10,72) [0,3653 a 0,8679]	0,7454 (7,35) [0,4875 a 0,8785]	0,8648 (3,39) [0,7545 a 0,9375]
Alta	0,7371 (8,80) [0,4189 a 0,8721]	0,8287 (4,81) [0,6312 a 0,9246]	0,9232 (1,96) [0,8478 a 0,9619]

A análise Espacial apresentou os menores valores de CV e as menores diferenças entre mínimo e máximo, em relação às análises em Bloco e Látice. Com o aumento da DE e redução da AS, observou-se maiores diferenças entre os resultados das análises testadas. Para os cenários com alta DE e AS de 0,60, os CV foram de 8,80, 4,81 e 1,96%, respectivamente para as análises em Bloco, Látice e Espacial.

Nos cenários com DE nula e AS 0,60 (Tabela 10), os três tipos de análises apresentaram resultados muito próximos, com médias variando de 0,5515 a 0,5757 e CV de 12,56 a 13,42% para AS de 0,60. Revelando que mesmo na ausência de DE a análise Espacial apresentou a mesma eficiência de classificação das famílias que as demais análises.

Na Tabela 11, são apresentados os valores da estatística F para o efeito de família relacionado ao nível de acurácia obtido nas diferentes análises e cenários simulados para Dependência Espacial (DE) e Acurácia Seletiva (AS). Segundo RESENDE, et al. (2007), a estatística F, assim como a acurácia seletiva é uma medida que facilita e agrega as informações mais relevantes sobre a precisão experimental para avaliação genotípica.

Valores de acurácia seletiva superiores a 80% ($F > 2,78$) devem ser buscados. Ensaio cujo valor de F para cultivares for inferior a 1,56 ($AS < 60\%$) não devem ser aproveitados. Para atingir uma acurácia seletiva (AS) ideal de 95% ou mais, os valores de F para cultivares devem ser iguais ou superiores a 10,26, por conseguinte, este pode ser um valor de referência para experimentos de avaliação de VCU (RESENDE, et al., 2007).

Observou-se que as médias da estatística F para cenários com DE nula foram bem próximas nas diferentes análises, variando de 9,49 a 10,41, de 2,08 a 2,87, e de 1,53 a 1,59, respectivamente, para cenários com AS de 0,95, 0,80 e 0,60 (Tabela 11). Com a redução na AS simulada observou-se, também, grandes reduções nos valores da estatística F. Esse resultado era esperado, uma vez que os cenários simulados com baixa AS se referem a cenários com baixa variância devido a família.

Com o aumento da DE, observou-se aumento na média da estatística F em toda a análise e cenários de AS. Comparando os cenários com DE nula e alta, esse aumento para AS 0,95 foi de 136,24, 370,6 e 1206,14%, respectivamente nas análises em Bloco, Látice e Espacial enquanto para AS de 0,60 esses aumentos foram de menor magnitude (55,41, 139,86 e 759%). Sendo as maiores variações observadas para a análise Espacial (Tabela 11).

Na presença de DE, a análise Espacial apresentou as maiores médias para estatística F em relação às análises em Látice e Bloco, respectivamente, principalmente para cenários com alta DE. Indicando a maior precisão experimental da análise Espacial em relação às análises tradicionais.

A análise Espacial foi a que apresentou os menores CV para estatística F, 18,41, enquanto Bloco e Látice foram 22,36 e 22,11 respectivamente (Tabela 11).

Tabela 11 - Média, CV (entre parênteses, %) e valores mínimos e máximos (entre colchetes) para estatística F para efeito de família nas diferentes análises e cenários simulados para Dependência Espacial (DE) e Acurácia Seletiva (AS)

Dep.Espacial	Bloco	Látice	Espacial
DE	AS=0,95		
Nula	10,3237 (17,63) [5,75 a 18,29]	9,4974 (18,46) [5,05 a 17,08]	10,4102 (17,80) [5,75 a 19,78]
Baixa	12,2872 (21,50) [6,28 a 22,45]	16,7084 (21,95) [8,58 a 36,82]	36,2411 (19,09) [18,46 a 63,9]
Média	15,8538 (26,07) [6,17 a 35,16]	25,2498 (22,13) [12,86 a 48,94]	68,8276 (17,77) [34,08 a 107,15]
Alta	24,3851 (29,80) [9,01 a 54,93]	44,6622 (23,91) [19,17 a 81,85]	135,9741 (18,48) [76,75 a 224,08]
	AS=0,80		
Nula	2,8450 (17,78) [1,60 a 4,89]	2,6885 (18,01) [1,46 a 4,39]	2,8715 (17,74) [1,68 a 4,89]
Baixa	3,1552 (20,15) [1,65 a 5,64]	3,9595 (20,04) [2,14 a 6,90]	7,6484 (18,97) [4,26 a 13,53]
Média	3,9185 (22,89) [1,79 a 7,11]	5,7794 (22,15) [2,36 a 10,27]	14,1450 (19,38) [7,84 a 24,23]
Alta	5,52 (26,26) [2,64 a 10,75]	9,3153 (22,78) [4,39 a 18,28]	26,8164 (18,62) [13,24 a 51,16]
	AS=0,60		
Nula	1,5767 (17,50) [0,93 a 2,62]	1,5302 (18,09) [0,79 a 2,54]	1,5943 (17,75) [0,93 a 2,63]
Baixa	1,6794 (17,96) [0,95 a 2,85]	1,9361 (19,07) [1,03 a 3,55]	3,1007 (18,61) [1,51 a 5,4]
Média	1,9204 (18,97) [1,03 a 3,55]	2,5156 (19,39) [1,25 a 4,28]	5,1719 (18,62) [2,80 a 9,86]
Alta	2,4486 (22,36) [0,98 a 5,0]	3,6769 (22,16) [1,86 a 6,99]	9,1831 (18,42) [4,55 a 15,25]

Outra medida alternativa para quantificar a precisão experimental e com maior facilidade de interpretação é a acurácia seletiva, que foi estimada em função da própria estatística F. A acurácia seletiva refere-se à correlação entre o valor genotípico verdadeiro do tratamento genético e aquele estimado ou predito a partir das informações do experimento. Ensaio com maior acurácia seletiva, em geral, são aqueles com menor variância residual e maior variância genotípica. Na Tabela 12, são apresentados os resultados relativos à acurácia seletiva estimada (ASE) nas diferentes análises (Bloco, Látice e Espacial), diferentes cenários para Dependência Espacial (DE) e Acurácia Seletiva (AS).

Para os cenários de alta AS (0,95), as médias para acurácia seletiva estimada foram altas e apresentaram valores bem próximos, variando de 0,9439 a 0,9962. Os coeficientes de variação (CV) foram baixos, variando de 0,07 a 1,12%, assim como os valores mínimos e máximos foram bem próximos. Pequenos valores de CV, bem como pequenas diferenças entre valores mínimos e máximos refletem em ensaios com maior acurácia seletiva. Aumentando a DE de nula a alta, houve pequeno aumento na média e redução no CV, principalmente na análise Espacial onde o CV reduziu em 92,82%, enquanto que nas análises em Bloco e Látice essa redução foi de 13,30 e 71,14% respectivamente.

Com a redução na AS de 0,95 para 0,80 e 0,60, observou-se a análise Espacial apresentou os melhores resultados. As médias para a acurácia seletiva estimada foram de 0,8936, 0,9404 e 0,9806, respectivamente para as análises em Bloco, Látice e Espacial nos cenários com AS de 0,80 e alta DE. Para cenários com baixa AS (0,60) e alta DE, a acurácia seletiva estimada da análise Espacial foi menor em relação as AS 0,95 e 0,80, sendo as médias de 0,7459, 0,8417 e 0,9423, respectivamente para as análises em Bloco, Látice e Espacial (Tabela 10).

A análise Espacial apresentou os menores valores de CV e as menores diferenças entre mínimo e máximo em alta DE, revelando que mesmo na ausência de DE a análise Espacial apresentou acurácia seletiva estimada maiores que as demais análises.

Com redução da AS, houve diminuição nos valores de média da acurácia seletiva, por exemplo, as médias para análise em Bloco com AS de 0,95, 0,80 e 0,60 e alta DE foram, respectivamente, 0,9764, 0,8936, 0,7459; na análise em Látice, as médias foram 0,9878, 0,9404, 0,8417.

Tabela 12 - Média, CV (entre parênteses, %) e valores mínimos e máximos (entre colchetes) para ¹acurácia seletiva estimada (ASE) nas diferentes análises e cenários simulados para Dependência Espacial (DE) e Acurácia Seletiva (AS)

Dep.Espacial	Bloco	Látice	Espacial
AS=0,95			
DE			
Nula	0,9487 (0,99) [0,9089 a 0,9723]	0,9439 (1,12) [0,8655 a 0,9703]	0,9491 (0,97) [0,9089 a 0,9744]
Baixa	0,9561 (1,04) [0,9056 a 0,9775]	0,9681 (0,72) [0,9399 a 0,9863]	0,9856 (0,28) [0,9725 a 0,9921]
Média	0,9650 (1,01) [0,9154 a 0,9857]	0,9788 (0,50) [0,9603 a 0,9897]	0,9925 (0,14) [0,9852 a 0,9960]
Alta	0,9764 (0,81) [0,9274 a 0,9909]	0,9878 (0,32) [0,9736 a 0,9939]	0,9962 (0,07) [0,9935 a 0,9978]
AS=0,80			
Nula	0,7971 (5,33) [0,6124 a 0,8919]	0,7832 (6,00) [0,5613 a 0,8788]	0,7993 (5,21) [0,6362 a 0,8919]
Baixa	0,8169 (5,17) [0,6059 a 0,9070]	0,8578 (3,64) [0,7299 a 0,9247]	0,9298 (1,50) [0,8748 a 0,9666]
Média	0,8514 (4,69) [0,5886 a 0,9270]	0,9031 (2,63) [0,7591 a 0,9501]	0,9627 (0,76) [0,9340 a 0,9791]
Alta	0,8936 (3,65) [0,7567 a 0,9524]	0,9404 (1,61) [0,8788 a 0,9723]	0,9806 (0,38) [0,9615 a 0,9902]
AS=0,60			
Nula	0,5789 (18,97) [0,0 a 0,7863]	0,5572 (22,43) [0,0 a 0,7787]	0,5848 (18,43) [0,0 a 0,7873]
Baixa	0,6112 (17,14) [0,0 a 0,8057]	0,6767 (12,14) [0,1707 a 0,8475]	0,8152 (5,06) [0,5812 a 0,9027]
Média	0,6707 (12,55) [0,1707 a 0,8475]	0,7363 (7,13) [0,4472 a 0,8754]	0,8944 (2,35) [0,8018 a 0,9479]
Alta	0,7459 (9,62) [0,0 a 0,8944]	0,8417 (4,57) [0,6800 a 0,9257]	0,9423 (1,19) [0,8833 a 0,9681]

¹ASE = $(1-1/F)^{0,5}$ para $F \geq 1$; e AS = 0 para $F < 1$, em que F é a estatística para o efeito de família

Resultados para ASE (Tabela 12) foram próximos aos obtidos para correlação de spearman entre médias estimadas e simuladas para família (Tabela 10), isso reflete que a AS é uma medida eficiente da correlação entre genótipo verdadeiro e estimado, devendo ser utilizada como medida de precisão experimental nos ensaios de avaliação genética.

As propriedades favoráveis da estatística acurácia seletiva são alternativas adequadas para a classificação da precisão experimental de ensaios de competição de genótipos de milho, feijão e soja, Resende & Duarte (2007).

A Tabela 13 apresenta resultados para Eficiência de Seleção (ES), considerando seleção de 10% das melhores famílias, nas diferentes análises e cenários simulados para Dependência Espacial (DE) e Acurácia Seletiva (AS). A ES foi calculada como porcentagem de famílias selecionadas corretamente em cada análise em comparação à classificação verdadeira das famílias.

Para os cenários com DE nula, as análises apresentaram resultados bastante semelhantes quanto a ES. As médias da ES variaram de 0,7552 a 0,7655 (AS=0,95), de 0,5398 a 0,56 (AS=0,80) e de 0,3623 a 0,3815 (AS=0,60). Porém, observou-se que com o aumento da DE houve aumento na média e redução do CV da ES para todas as análises, sendo que a análise Espacial apresentou os melhores resultados. Considerando alta DE e AS de 0,95, as médias para ES das análises em Bloco, Látice e Espacial foram 0,8433; 0,8823 e 0,9262, respectivamente. A análise Espacial também apresentou os menores CV para ES, 9,94; 8,51 e 6,78%, respectivamente para baixa, média e alta DE (Tabela 13). Esses resultados indicam que em média, das 10 famílias selecionadas utilizando análise Espacial, 9,262 famílias foram selecionadas corretamente, ou seja, pertenciam realmente ao grupo das 10 famílias com maiores médias simuladas.

Assim, nos cenários de alta DE e baixa AS, das 10 famílias selecionadas utilizando análise Espacial em média, 7,33 famílias foram selecionadas corretamente. Enquanto para as análises em Látice e Bloco, respectivamente, apenas 6,00 e 5,08 famílias foram selecionadas corretamente. A diferença entre as análises quanto ao número de famílias selecionadas corretamente parece irrelevante. Porém, cada família selecionada incorretamente, representa uma família com maior valor genético, que foi descartada e perdida.

Tabela 13 - Média, CV (entre parênteses, %) e valores mínimos e máximos (entre colchetes) para Eficiência de Seleção (ES), considerando seleção de 10% das melhores famílias, nas diferentes análises e cenários simulados para Dependência Espacial (DE) e Acurácia Seletiva (AS)

Dep.Espacial	Bloco	Látice	Espacial
DE	AS=0,95		
Nula	0,7654 (13,05) [0,4 a 1]	0,7552 (13,37) [0,4 a 1]	0,7646 (13,03) [0,4 a 1]
Baixa	0,7840 (12,44) [0,5 a 1]	0,8049 (11,59) [0,5 a 1]	0,8591 (9,94) [0,5 a 1]
Média	0,8099 (12,03) [0,5 a 1]	0,8439 (10,53) [0,5 a 1]	0,8980 (8,51) [0,6 a 1]
Alta	0,8433 (10,53) [0,5 a 1]	0,8823 (8,71) [0,6 a 1]	0,9262 (6,78) [0,7 a 1]
	AS=0,80		
Nula	0,5600 (22,32) [0,1 a 0,9]	0,5398 (23,56) [0,1 a 0,9]	0,5592 (22,40) [0,1 a 0,9]
Baixa	0,5831 (22,03) [0,2 a 0,9]	0,6237 (20,20) [0,2 a 0,9]	0,7124 (15,55) [0,4 a 1]
Média	0,6204 (19,89) [0,2 a 0,9]	0,6856 (17,66) [0,3 a 1]	0,7850 (12,59) [0,5 a 1]
Alta	0,6784 (18,02) [0,2 a 1]	0,7520 (13,82) [0,3 a 1]	0,8394 (10,93) [0,6 a 1]
	AS=0,60		
Nula	0,3803 (34,05) [0 a 0,7]	0,3623 (35,99) [0 a 0,7]	0,3815 (34,39) [0 a 0,7]
Baixa	0,4051 (33,10) [0,0 a 0,8]	0,4474 (30,06) [0,0 a 0,9]	0,5443 (23,44) [0,1 a 0,9]
Média	0,4407 (29,68) [0,1 a 0,8]	0,5135 (24,75) [0,1 a 0,9]	0,6515 (17,68) [0,3 a 0,9]
Alta	0,5083 (26,42) [0,1 a 0,9]	0,6005 (20,84) [0,3 a 0,9]	0,7330 (15,34) [0,3 a 1]

Em todas as análises (Tabela 13), a redução da AS proporcionou aumento dos CV e das amplitudes para ES. Mesmo assim, o CV para ES na análise Espacial (15,34%) foi inferior ao das análises em Látice (20,84%) e em Bloco (26,42%).

Na Tabela 14, são apresentados os resultados para Eficiência de Seleção (ES), considerando seleção de 20% das melhores famílias, nas diferentes análises e cenários simulados para Dependência Espacial (DE) e Acurácia Seletiva (AS). A ES foi calculada como percentagem de famílias selecionadas corretamente em cada análise em comparação à classificação verdadeira das famílias.

As análises apresentaram resultados bastante semelhantes quanto a ES para os cenários com DE nula. Em que médias variaram de 0,8068 a 0,8176 (AS=0,95), de 0,6224 a 0,6391 (AS=0,80) e de 0,4780 a 0,4921 (AS=0,60). Porém, observou-se que com o aumento da DE houve aumento na média e redução do CV da ES para todas as análises, sendo que a análise Espacial apresentou os melhores resultados. Considerando alta DE e AS de 0,95, as médias para ES das análises em Bloco, Látice e Espacial foram 0,8766, 0,9084 e 0,9421, respectivamente. A análise Espacial também apresentou os menores CV para ES, 5,81, 4,94 e 4,18%, respectivamente para baixa, média e alta DE (Tabela 14). Esses resultados indicam que em média, das 20 famílias selecionadas utilizando análise Espacial, 9,421 famílias foram selecionadas corretamente, ou seja, pertenciam realmente ao grupo das 20 famílias com maiores médias simuladas.

Com a redução da AS, a análise Espacial foi ainda mais eficiente que as análises tradicionais, para os cenários de alta DE e baixa AS, das 20 famílias selecionadas utilizando análise Espacial em média, 7,89 famílias foram selecionadas corretamente. Enquanto para as análises em Látice e Bloco, respectivamente, apenas 6,82 e 6,02 famílias foram selecionadas corretamente.

Em todas as análises, a redução da AS proporcionou aumento dos CV e das amplitudes para ES. Mesmo assim, o CV para ES na análise Espacial (8,72%) foi inferior ao das análises em Látice (11,50%) e em Bloco (15,20%), na baixa AS e alta DE.

Tabela 14 - Média, CV (entre parênteses, %) e valores mínimos e máximos (entre colchetes) para Eficiência de Seleção (ES), considerando seleção de 20% das melhores famílias, nas diferentes análises e cenários simulados para Dependência Espacial (DE) e Acurácia Seletiva (AS)

Dep.Espacial	Bloco	Látice	Espacial
DE	AS=0,95		
Nula	0,8174 (8,11) [0,6 a 1]	0,8068 (8,32) [0,6 a 1]	0,8176 (8,08) [0,6 a 1]
Baixa	0,8287 (7,72) [0,6 a 1]	0,8531 (7,09) [0,65 a 1]	0,8922 (5,81) [0,7 a 1]
Média	0,8480 (7,21) [0,65 a 1]	0,8786 (6,28) [0,7 a 1]	0,9205 (4,94) [0,75 a 1]
Alta	0,8766 (6,61) [0,7 a 1]	0,9084 (5,44) [0,75 a 1]	0,9421 (4,18) [0,8 a 1]
	AS=0,80		
Nula	0,6390 (13,30) [0,35 a 0,85]	0,6246 (13,56) [0,35 a 0,9]	0,6391 (13,30) [0,35 a 0,85]
Baixa	0,6612 (12,70) [0,35 a 0,9]	0,6958 (11,64) [0,4 a 0,95]	0,7701 (9,06) [0,55 a 0,95]
Média	0,6924 (11,68) [0,45 a 0,95]	0,7491 (10,13) [0,45 a 0,95]	0,8281 (7,64) [0,5 a 1]
Alta	0,7442 (10,93) [0,45 a 0,95]	0,8065 (8,65) [0,55 a 1]	0,8755 (6,48) [0,65 a 1]
	AS=0,60		
Nula	0,4921 (18,63) [0,2 a 0,75]	0,4780 (18,55) [0,2 a 0,75]	0,4910 (18,47) [0,2 a 0,75]
Baixa	0,5106 (18,48) [0,25 a 0,8]	0,5454 (16,92) [0,25 a 0,80]	0,6363 (13,48) [0,35 a 0,9]
Média	0,5501 (16,94) [0,25 a 0,80]	0,6106 (15,01) [0,30 a 0,85]	0,7190 (10,58) [0,5 a 0,9]
Alta	0,6029 (15,20) [0,30 a 0,85]	0,6823 (11,50) [0,4 a 0,9]	0,7896 (8,72) [0,55 a 0,95]

São apresentados na Tabela 15, os resultados para Eficiência de Seleção (ES), considerando seleção de 30% das melhores famílias, nas diferentes análises e cenários simulados para Dependência Espacial (DE) e Acurácia Seletiva (AS). A ES foi calculada como percentagem de famílias selecionadas corretamente em cada análise em comparação à classificação verdadeira das famílias.

Nos cenários com DE nula, as análises apresentaram resultados bastante semelhantes quanto a ES, em que suas médias variaram de 0,8404 a 0,8485 (AS=0,95), de 0,6866 a 0,6999 (AS=0,80) e de 0,5575 a 0,5707 (AS=0,60). Com o aumento da DE, observou-se que houve aumento na média e redução do CV da ES para todas as análises, sendo que a análise Espacial apresentou os melhores resultados.

Considerando alta DE e AS de 0,95, as médias para ES das análises em Bloco, Látice e Espacial foram 0,8971; 0,9250 e 0,9519, respectivamente. A análise Espacial também apresentou os menores CV para ES, 4,23; 3,54 e 3,17%, respectivamente para baixa, média e alta DE (Tabela 15). Esses resultados indicam que em média, das 30 famílias selecionadas utilizando análise Espacial, 9,51 famílias foram selecionadas corretamente, ou seja, pertenciam realmente ao grupo das 30 famílias com maiores médias simuladas.

Com a redução da AS, a análise Espacial foi ainda mais eficiente que as análises tradicionais, para os cenários de alta DE e baixa AS, das 30 famílias selecionadas utilizando análise Espacial em média, 8,26 famílias foram selecionadas corretamente. Enquanto para as análises em Látice e Bloco, respectivamente, apenas 7,40 e 6,66 famílias foram selecionadas corretamente.

Tabela 15 - Média, CV (entre parênteses, %) e valores mínimos e máximos (entre colchetes) para Eficiência de Seleção (ES), considerando seleção de 30% das melhores famílias, nas diferentes análises e cenários simulados para Dependência Espacial (DE) e Acurácia Seletiva (AS)

Dep.Espacial	Bloco	Látice	Espacial
DE	AS=0,95		
Nula	0,8485 (5,66) [0,7 a 0,96]	0,8404 (5,64) [0,7 a 0,9667]	0,8479 (5,64) [0,7 a 0,9667]
Baixa	0,8587 (5,46) [0,7000 a 0,9667]	0,8762 (5,03) [0,7333 a 1]	0,9094 (4,23) [0,7667 a 1]
Média	0,8752 (5,43) [0,6667 a 1]	0,9001 (4,39) [0,7667 a 1]	0,9346 (3,54) [0,7667 a 1]
Alta	0,8971 (4,68) [0,7333 a 1]	0,9250 (3,91) [0,7331 a 1]	0,9519 (3,17) [0,7667 a 1]
	AS=0,80		
Nula	0,6999 (9,10) [0,53 a 0,9]	0,6866 (9,43) [0,5 a 0,86]	0,6993 (8,98) [0,5 a 0,9]
Baixa	0,7167 (9,30) [0,4667 a 0,9333]	0,7473 (8,22) [0,5333 a 0,9333]	0,8115 (6,44) [0,6 a 0,9667]
Média	0,7473 (8,61) [0,5 a 0,9333]	0,7933 (7,08) [0,6 a 0,9667]	0,8608 (5,61) [0,6667 a 0,9667]
Alta	0,7891 (7,47) [0,5667 a 0,9667]	0,8389 (5,94) [0,6667 a 0,9667]	0,8987 (4,33) [0,7333 a 1]
	AS=0,60		
Nula	0,5706 (12,49) [0,2667 a 0,8]	0,5575 (12,52) [0,3 a 0,7667]	0,5707 (12,45) [0,2667 a 0,8]
Baixa	0,5915 (12,34) [0,3667 a 0,8333]	0,6204 (11,44) [0,4 a 0,8333]	0,6983 (9,23) [0,4667 a 0,9]
Média	0,6200 (11,45) [0,4 a 0,8333]	0,6749 (10,06) [0,4 a 0,9]	0,7660 (7,54) [0,6 a 0,9333]
Alta	0,6663 (10,77) [0,4 a 0,8667]	0,7404 (8,11) [0,5333 a 0,9]	0,8262 (6,17) [0,6667 a 0,9667]

6. CONCLUSÕES

Em relação à precisão experimental, a análise Espacial apresentou estimativas de variância do erro muito próximas da variância simulada (7744,00), apresentou também maior acurácia seletiva estimada em todos os cenários, indicando maior precisão experimental.

Comparando a classificação das famílias com base nas médias estimadas com a classificação verdadeira (médias simuladas), a análise Espacial foi mais eficiente para classificação das famílias.

A análise Espacial proporcionou maior percentagem de famílias selecionadas corretamente do que as análises tradicionais em Látice e em Blocos casualizados, principalmente, para seleção de menor número de famílias.

A acurácia seletiva, calculada em função da estatística F, é uma medida eficiente da correlação entre valor genotípico verdadeiro e estimado para famílias, devendo ser utilizada como medida de precisão experimental nos ensaios de avaliação genética.

7. REFERÊNCIAS BIBLIOGRÁFICAS

ANSELIN, L. SpaceStat tutorial: a workbook for using SpaceStat in the analysis of spatial data. Santa Barbara, NCGIA (National Center for Geographic Information and Analysis), 1992.

BLAIS RA, CARLIER PA (1968) Applications Of Geostatistics In Ore Evaluation. En Ore Reserve Estimation And Grade Control. Special Vol. 9. Canadian Institute of Mining and Metalurgy. Montreal, Canada. pp. 41-48.

CAMARGO, E. C. G. Geoestatística: fundamentos e aplicações.

CAMARGO, E.C.G. Módulo de procedimentos geoestatísticos para o sistema de processamento de informações geo-referenciadas (SPRING) baseado em técnicas de krigagem. 1997. 108 f. Tese (Mestrado) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos.

CIF - CENTRO DE INTELIGÊNCIA DO FEIJÃO. Histórico. Acedido em 01 de Março de 2011, em: <http://www.cifeijao.com.br>.

COSTA e SILVA, J.; DUTKOWSKI, G. W.; GILMOUR, A. R. Analysis of early tree height in forest genetic trials is enhanced by including a spatially correlated residual. *Silvae Genet.*, v. 31, p.1887-1893, 2001.

CRESSIE, N.; HARTFIELD, M.N. Conditionally specified gaussian models for spatial statistical analysis of field trials. *Journal of Agricultural, Biological, and Environmental Statistics*, v.1, p.60-77, 1996.

DRUCK, S.; CARVALHO, M. S.; CÂMARA, G.; MONTEIRO, A. V. M. Análise Espacial de Dados Geográficos. Brasília: Embrapa, 2004. 209 p.

DUARTE, J.B (2000). Sobre o emprego e a análise estatística do delineamento em blocos aumentados no melhoramento genético vegetal. Piracicaba .293p. ESALQ/USP.

FERES, A.L.G. Análise estatística na avaliação de produtividade no melhoramento genético do feijoeiro.2009.Dissertação (Mestrado) – Universidade Federal de Viçosa , Viçosa , MG , 2009.

FERREIRA, D. F. Uso de simulação no melhoramento. In: NASS, L. L.; VALOIS, A. C. C.; MELO, I. S. de; VALADARES-INGLIS, M. C. (Ed). Recursos genéticos e melhoramento de plantas. Rondonópolis: Fundação-MT, 2001. p. 1119-1141.

FRASER, A. S. Simulation of genetics systems by automatic digital computers. I: Introduction. Australian Journal of Biological Science, Melbourne, v. 10, p. 484-491, 1957a.

FRASER, A. S. Simulation of genetics systems by automatic digital computers. II: Effect of linkage on rates of advance under selection. Australian Journal of Biological Science, Melbourne, v. 10, p. 492-499, 1957b.

GENÚ, A. M. Geoestatística Multivariada.Universidade de São Paulo, Escola superior de agricultura “luiz de queiroz”, 2004.

GRONDONA, M. O.; CROSSA, J.; FOX, P. N.; PFEIFFER, W. H. Analysis of variety yield trials using two-dimensional separable ARIMA processes. Biometrics, v.52, p.763-770, 1996.

GUIMARÃES, E.C. Geoestatística básica e aplicada. Uberlândia - MG UFU, 2004. 78p.

HENDERSON, C.R. 1984. Applications of linear models in animal breeding. University of Guelph, Guelph. 462 p.

HUIJBREGTS, C. J. Regionalized variables and quantitative analysis of spatial data. In: DAVIS, J. C.; McCULLAGH, M. J. (Ed.). Display and analysis of spatial data. New York: John Wiley, 1975. p.38-53.

ISAAKS, E. H.; SRIVASTAVA, R. M. An Introduction to applied geostatistics. New York: Oxford University Press, 1989. 560 p.

JOURNEL, A. G. e C. J. HUIJBREGTS. Mining geostatistics . London: Academic Press. 1978. 600 p.

KRIGE, D. G. A statistical approach to some basic mine evaluation problems on the witwatersrand. Journal of South African Institution of Mining and Metallurgy, Johannesburg, v. 52, p. 119-139, 1951.

KUSNADAR, D.; GALWEY, N. A proposed method for estimation of genetic parameters on forest trees without raising progeny: critical evaluation and refinement. *Silvae Genet.*, v.49, p.15-21, 2000.

MARTINEZ, R.(1994) Control de la correlacion espacial en experimentos de campo en el sector agricola. *Agronomia Colombiana*, v. XI, p. 83-89.

MATASSA, V.; BASFORD, K.E.; CHAN, B.; ECCLESTON, J. (2001) Comparisons of augmented and spatial unreplicated experimental designs in sugarcane breeding under a spatial model – a simulation study. In: Australasian Biometrics and New Zealand Statistical Association Joint Conference 2001, 2001. Christchurch, New Zealand. Disponível em: < <http://at.yorku.ca/cgi-bin/amca/caic-12>> Acesso em 03 de janeiro 2011.

MATHERON, G. Les variables régionalisées et leur estimation. Paris: Masson et Cie, 1965. 305 p.

MATHERON, G. Principles of geostatistics. *Economic Geology*, Lancaster,v. 58, p. 1246-1266, 1963.

MATHERON, G. The theory of regionalized variables and its application. *Les Cahiers du Centre de Morphologie Mathématique*, n. 5, 1971.

OLEA, R. A. Measuring spatial dependence with semivariograms. Kansas, Kansas Geological Survey, 1977. (Series on Spatial Analysis 3).

OLEA, R. A. Optimum mapping techniques using regionalized variable theory. Kansas, Kansas Geological Survey, 1975. (Series on Spatial Analysis, 2).

R DEVELOPMENT CORE TEAM (2009). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, <http://www.R-project.org>.

RESENDE, M.D.V. de; DUARTE, J.B. Precisão e controle de qualidade em experimentos de avaliação de cultivares. Pesquisa Agropecuária Tropical, v.37, p.182-194, 2007.

RESENDE, M.D.V. de. 2002. Genética biométrica e estatística no melhoramento de plantas perenes. Embrapa Informação Tecnológica, Brasília. 975 p.

SALVIANO, A.A.C. Variabilidade de atributos de solo e de *Crotalaria juncea* em solo degradado do município de Piracicaba-SP. Piracicaba, 1996. 91p. Tese (Doutorado) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo.

SAS Institute Inc. Statistical Analysis System user's guide. Version 9.1 ed. Cary: SAS Institute, USA, 2003. Licenciado para a Universidade Federal de Viçosa, 2010.

SCOTT, R.A. e MILLIKEN, G.A. (1993) A SAS program for analyzing augmented randomized complete-block designs. Crop Science, v.33, p. 865-867.

SOUZA. A.P; GERALDI, I.O.; RAMALHO, M.A.P. Alternativas experimentais na avaliação de famílias em programas de melhoramento genético do feijoeiro. Pesq. agropec. bras. v.35,n9, p. 1765-1771. 2000.

STEEL, R. G. D.; TORRIE, J. H. Principles and procedures of statistics. 2th. ed. New York: Mac Graw-Hill, 1980. 633p.

STROUP, W.W.; BAENZIGER, P.S.; MULITZE, D.K. Removing spatial variation from wheat yield trials: a comparison of methods. *Crop Science*, Madison, v.86, p.62-66, 1994.

VENCOVSKY, R. 1987. Herança quantitativa. p. 137-214. In E. Paterniani & G.P. Viegas (Ed.). *Melhoramento e produção de milho*. 2.ed. Fundação Cargill, Campinas. v. 1. 795 p.

VIEIRA, S. R. Geoestatística em estudos de variabilidade espacial do solo. In: NOVAIS, R. F. de; ALVAREZ V., V. H.; SCHAEFER, C. E. G. R. (Ed.). *Tópicos em ciência do solo*. Viçosa: Sociedade Brasileira de Ciência do Solo, 2000. v. 1, p. 1-54.

WEBSTER, R.; OLIVER M, A, *Statistical methods in soil and land resource survey*, Oxford Univ, Press, New York, 1990.

WOLFINGER, R.D.; FEDERER, W.T.; CORDERO-BRANA, O. (1997) Recovering information in augmented designs, using SAS PROC GLM and PROC MIXED. *Agronomy J.*; v. 89, p 856-859.

ZIMMERMAN, D.I.; HARVILLE, D.A. (1991) A random field approach to the analysis of field-plot experiments and other spatial experiments *Biometrics*, v. 47, p. 223-239.