



CRISTINA HENRIQUES NOGUEIRA

**ANÁLISE DE VARIÂNCIA COM
DEPENDÊNCIA ESPACIAL SOB UMA
ABORDAGEM GEOESTATÍSTICA**

LAVRAS - MG

2013

CRISTINA HENRIQUES NOGUEIRA

**ANÁLISE DE VARIÂNCIA COM DEPENDÊNCIA ESPACIAL
SOB UMA ABORDAGEM GEOESTATÍSTICA**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Mestre.

Orientador

Dr. Renato Ribeiro Lima

LAVRAS - MG

2013

**Ficha Catalográfica Elaborada pela Divisão de Processos
Técnicos da Biblioteca da UFLA**

Nogueira, Cristina Henriques.

Análise de variância com dependência espacial sob uma
abordagem geoestatística / Cristina Henriques Nogueira. – Lavras :
UFLA, 2013.

123 p. : il.

Dissertação (mestrado) – Universidade Federal de Lavras, 2013.

Orientador: Renato Ribeiro de Lima.

Bibliografia.

1. Covariância. 2. Análise de experimentos. 3. Semivariograma.
4. Erros dependentes. I. Universidade Federal de Lavras. II. Título.

CDD – 519.538

CRISTINA HENRIQUES NOGUEIRA

**ANÁLISE DE VARIÂNCIA COM DEPENDÊNCIA ESPACIAL
SOB UMA ABORDAGEM GEOESTATÍSTICA**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Mestre.

APROVADA em 27 de fevereiro de 2013.

Dr. José Márcio de Mello UFLA

Dr. Marcelo Silva de Oliveira UFLA

Dr. Renato Ribeiro Lima
Orientador

LAVRAS - MG

2013

Aos meus pais, que são as bases da minha vida, cuidando sempre com muito carinho e dedicação do meu crescimento pessoal e profissional.

À minha sobrinha Luiza, que, sempre me recebendo com abraços enforcantes e se despedindo com lágrimas nos olhos, conseguiu compreender que a minha ausência se fazia necessária.

DEDICO

AGRADECIMENTOS

Aos meus pais, por toda a confiança que depositaram em mim e cada palavra de incentivo, sem os quais eu não teria conseguido concluir esta etapa tão importante.

A minha irmã, Liliane, que fez do seu exemplo uma fonte inspiradora, pelo constante apoio e por estar sempre ao meu lado, nas horas mais difíceis.

Ao meu irmão, Rogério, que mesmo a distância, sempre torceu pelo meu sucesso. Aos demais familiares que me incentivaram.

Aos professores da Universidade Federal de São João del-Rei (UFSJ), por suas contribuições e ensinamentos durante toda a graduação em Matemática.

Ao professor Dr. Marcos Santos de Oliveira que, além de me apresentar à Estatística, permitiu que este sonho pudesse se tornar realidade, incentivando e acreditando no meu potencial até os dias atuais.

À Universidade Federal de Lavras e ao Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, pela oportunidade concedida para a realização deste trabalho.

À Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), pela concessão da bolsa de estudos.

Ao professor Dr. Marcelo Silva de Oliveira, pela confiança depositada, pelas sugestões e críticas, além do apoio e orientação durante o primeiro ano do curso.

Ao meu orientador, Dr. Renato Ribeiro de Lima, pelo incondicional e inestimável apoio, pela oportunidade de poder trabalhar ao seu lado, além dos preciosos ensinamentos compartilhados, meu sincero agradecimento.

A todos os professores do Departamento de Ciências Exatas (DEX), pela disponibilidade e pelos conhecimentos oferecidos durante o curso.

Aos demais funcionários do DEX, em especial à Josi, pela boa vontade e eficiência com que sempre me atenderam.

À Marcília Teixeira, minha melhor amiga durante esses dois anos, que, além do curso e da república, dividiu também alegrias, tristezas, medos e angústias. Sempre ao meu lado, possibilitou que essa jornada se tornasse mais prazerosa.

Às minhas amigas e companheiras de república, Sol Rivelí, Priscila Couto, Késia Teixeira e Marília Teixeira (*in memoriam*), pela amizade, apoio e convívio ao longo desses dois anos.

A todos meus amigos que, mesmo longe, se fizeram presentes, oferecendo sempre palavras de incentivo e apoio.

Aos membros participantes da banca examinadora, pelas sugestões e valiosas contribuições.

Finalmente, a todos aqueles que, de alguma forma, acreditaram e contribuíram para a realização deste sonho, meu muito obrigado.

RESUMO

Um dos princípios básicos da experimentação é a aleatorização, tendo como finalidade gerar uma distribuição de amostragem para os erros experimentais, cujo comportamento assintótico aproxima-se de distribuições normais não-correlacionadas. Entretanto, nem sempre a prática de aleatorização é suficiente para neutralizar os efeitos de correlação entre as parcelas adjacentes e, por isso, encontram-se experimentos cujos erros apresentam uma estrutura de dependência espacial definida. Uma maneira para contornar esse problema é utilizar uma abordagem espacial, em que é possível estimar e modelar a correlação espacial entre os erros. Por meio desses modelos espaciais a dependência entre os erros deixa de ser vista como uma inconveniência a ser evitada, passando a ser considerada como um verdadeiro benefício, de forma que sua utilização pode proporcionar resultados mais precisos. Assim, este trabalho foi realizado com o objetivo de descrever como implementar e executar a análise de variância com erros espacialmente correlacionados, na qual a matriz de covariância do erros, modelada por meio de uma abordagem geoestatística, é utilizada como ponderadora das somas de quadrados da análise de variância. Além disso, verificou-se, ainda, a aplicabilidade do teste F , o qual é utilizado para comparar a igualdade entre as médias dos fatores controlados no experimento. No intuito de avaliar a eficiência da abordagem geoestatística, foram analisados experimentos instalados sob os delineamentos inteiramente casualizado (DIC) e em blocos casualizados (DBC) com diferentes números de parcelas. Os dados foram obtidos via simulação, seguindo modelos de covariância esférico e exponencial, com diferentes configurações de dependência espacial. Os resultados obtidos mostraram que a modelagem da correlação espacial dos erros foi mais eficiente, uma vez que produziu valores maiores para a estatística F_0 , comparados aos valores obtidos pelo modelo que supunha independência espacial dos erros. Além disso, a análise espacial produziu melhores estimativas para as médias dos tratamentos. Com isso, pode-se concluir que a utilização de ferramentas geoestatísticas na análise de experimentos foi mais eficaz para detectar diferenças entre as médias dos fatores estudados.

Palavras-chave: Erros dependentes. Semivariograma. Covariância. Análise de experimentos.

ABSTRACT

One of the essential principles of the experimentation is the randomization which contributes towards the presuppositions which the errors should be both equally and identically distributed, are met. Nevertheless, not always the randomization practice is enough to counteract the effects of correlation among the neighboring plots and for that reason, experiments are found of which mistakes present a definite spatial dependence structure. A way to bypass that trouble is to utilize a spatial approach in which it is possible to estimate and model the spatial correlation among the mistakes. So, the objective of this work was to report how to implement and carry out the variance analysis with spatially correlated errors in which the error covariance matrix, modeled through a geostatistical modeling, was utilized as a weighting factor of the sums of squares of the variance analysis. With the purpose of evaluating the efficiency of the geostatistical approach, experiments established under the completely randomized designs and in randomized blocks with different numbers of plots. The data were obtained by simulation following spherical and exponential covariance models with different settings of spatial dependence. The results obtained showed that even in experiments presenting spatially correlated errors, it is possible to continue to take advantage of the variance analysis as a tool to compare the different sources of variability. Nevertheless for that to occur, it is necessary for the error correlation matrix to be regarded as a weighting factor of the sums of the squares. Besides, the spatial error correlation modeling proved more efficient, since it produced greater values to the F statistic, if compared with the values obtained by the model which supposed spatial error independence. With this, one can conclude that the use of geostatistic tools in the analysis of experiments was more effective to detect differences among the means of the investigated factors.

Key-words: Dependent errors. Semivariogram. Covariance. Analysis experiments.

LISTA DE ILUSTRAÇÕES

| | | |
|-----------|--|----|
| Figura 1 | <i>Grid</i> de um experimento em DIC, com 15 tratamentos e 8 repetições | 51 |
| Figura 2 | <i>Grid</i> de um experimento em DIC, com 6 tratamentos e 5 repetições | 51 |
| Figura 3 | <i>Grid</i> de um experimento em DBC, com 15 tratamentos e 8 blocos | 52 |
| Figura 4 | <i>Grid</i> de um experimento em DBC, com 6 tratamentos e 5 blocos | 53 |
| Figura 5 | <i>Grid</i> do experimento contendo informações sobre os erros referentes a cada ponto amostrado | 93 |
| Gráfico 1 | Relação entre as funções covariância e semivariância | 20 |
| Gráfico 2 | Exemplo de um semivariograma experimental com comportamento ideal | 22 |
| Gráfico 3 | Exemplos dos modelos teóricos esférico, exponencial e gaussiano | 26 |
| Gráfico 4 | Modelos linear e potência | 28 |
| Gráfico 5 | Modelo efeito pepita puro | 28 |
| Gráfico 6 | Semivariograma dos erros contendo o envelope simulado . . | 94 |
| Gráfico 7 | Ajuste do modelo esférico ao semivariograma dos erros . . | 96 |
| Gráfico 8 | (a) Semivariograma com ajuste do modelo teórico da componente espacial \hat{S} e (b) dos erros estimados ξ^0 | 99 |

LISTA DE TABELAS

| | | |
|-----------|--|----|
| Tabela 1 | Análise de variância para um delineamento inteiramente casualizado. | 38 |
| Tabela 2 | Análise de variância para um delineamento em blocos casualizados. | 42 |
| Tabela 3 | Valores fixos dos parâmetros para os experimentos em DIC com 15 tratamentos e 8 repetições. | 49 |
| Tabela 4 | Valores fixos dos parâmetros para os experimentos em DIC com 6 tratamentos e 5 repetições. | 49 |
| Tabela 5 | Valores fixos dos parâmetros para os experimentos em DBC com 15 tratamentos e 8 repetições. | 49 |
| Tabela 6 | Valores fixos dos parâmetros para os experimentos em DBC com 6 tratamentos e 5 repetições. | 49 |
| Tabela 7 | Valores dos parâmetros do semivariograma utilizados para simular o erro experimental. | 50 |
| Tabela 8 | Análise de variância de um DIC, considerando erros correlacionados espacialmente. | 68 |
| Tabela 9 | Análise de variância de um DIC, considerando $\Sigma = \mathbf{V}\sigma^2$, por meio de formas quadráticas. | 68 |
| Tabela 10 | Somas de quadrados da análise de variância de um DIC, obtidas por meio de modelos reduzidos. | 69 |
| Tabela 11 | Esperança dos quadrados médios de um DIC com erros correlacionados, a partir das somas de quadrados clássicas. | 71 |
| Tabela 12 | Esperança dos quadrados médios de um DIC considerando erros espacialmente correlacionados. | 74 |

| | | |
|-----------|--|----|
| Tabela 13 | Somas de quadrados para a análise de variância de um DBC, por meio de redução de modelos. | 79 |
| Tabela 14 | Análise de variância de um DBC, cujos erros são espacialmente correlacionados. | 80 |
| Tabela 15 | Esperança dos quadrados médios de um DBC cujos erros têm dependência dentro dos blocos. | 82 |
| Tabela 16 | Esperança dos quadrados médios de um DBC cujos erros têm dependência dentro e entre blocos. | 82 |
| Tabela 17 | Esperança dos quadrados médios da análise de variância de um DBC, considerando dependência espacial. | 83 |
| Tabela 18 | Avaliação da abordagem geoestatística de um DIC com 15 tratamentos e 8 repetições. | 88 |
| Tabela 19 | Avaliação da abordagem geoestatística de um DIC com 6 tratamentos e 5 repetições. | 88 |
| Tabela 20 | Avaliação da abordagem geoestatística de um DBC com 15 tratamentos e 8 blocos, em relação ao efeito de tratamento. | 90 |
| Tabela 21 | Avaliação da abordagem geoestatística de um DBC com 15 tratamentos e 8 blocos, em relação ao efeito de bloco. | 90 |
| Tabela 22 | Avaliação da abordagem geoestatística de um DBC com 6 tratamentos e 5 blocos, em relação ao efeito de tratamento. | 91 |
| Tabela 23 | Avaliação da abordagem geoestatística de um DBC com 6 tratamentos e 5 blocos, em relação ao efeito de bloco. | 92 |
| Tabela 24 | Parâmetros estimados do semivariograma a cada iteração. | 95 |
| Tabela 25 | Análise de variância, considerando erros correlacionados. | 97 |
| Tabela 26 | Análise de variância, considerando erros independentes. | 97 |

SUMÁRIO

| | | |
|-------|---|-----|
| 1 | INTRODUÇÃO..... | 13 |
| 2 | REFERENCIAL TEÓRICO..... | 15 |
| 2.1 | Geoestatística..... | 15 |
| 2.1.1 | Estacionariedade..... | 16 |
| 2.1.2 | Estimação do semivariograma..... | 21 |
| 2.1.3 | Modelos teóricos de semivariograma..... | 24 |
| 2.1.4 | Modelos espaciais lineares..... | 29 |
| 2.2 | Estatística Experimental..... | 33 |
| 2.2.1 | A análise de variância..... | 34 |
| 2.2.2 | Delineamento inteiramente casualizado (DIC)..... | 35 |
| 2.2.3 | Delineamento em blocos casualizados (DBC)..... | 39 |
| 2.2.4 | Abordagem espacial na experimentação..... | 43 |
| 3 | MATERIAIS E MÉTODOS..... | 48 |
| 3.1 | Simulação de dados..... | 48 |
| 3.2 | Estimação de parâmetros..... | 53 |
| 3.3 | Modelagem geoestatística do erro..... | 55 |
| 3.4 | O teste F da análise de variância..... | 59 |
| 4 | RESULTADOS E DISCUSSÕES..... | 62 |
| 4.1 | Estimação de parâmetros considerando $\Sigma = V\sigma^2$ | 62 |
| 4.2 | Delineamento inteiramente casualizado (DIC)..... | 64 |
| 4.2.1 | Estimação de parâmetros via modelos particionados.. | 64 |
| 4.2.2 | Análise de variância..... | 67 |
| 4.2.3 | Esperanças dos quadrados médios..... | 70 |
| 4.2.4 | Aplicabilidade do teste F | 74 |
| 4.3 | Delineamento em blocos casualizados (DBC)..... | 77 |
| 4.3.1 | Análise de variância..... | 77 |
| 4.3.2 | Esperanças dos quadrados médios..... | 81 |
| 4.3.3 | Aplicabilidade do teste F | 83 |
| 4.4 | Análise dos dados simulados..... | 87 |
| 4.5 | Exemplo de aplicação..... | 92 |
| 5 | CONCLUSÕES..... | 101 |
| | REFERÊNCIAS..... | 102 |
| | ANEXOS..... | 106 |

1 INTRODUÇÃO

A execução do planejamento clássico de um experimento baseia-se em três princípios básicos, estabelecidos por Ronald A. Fisher, que são: repetição, aleatorização e controle local. De forma sucinta, a repetição é responsável pela estimação do erro experimental, a aleatorização tem o objetivo de proporcionar uma estimativa válida para o erro, enquanto o princípio de controle local tem a finalidade de tornar o experimento mais eficiente, por meio da redução do erro experimental, quando o mesmo é instalado em condições ambientais heterogêneas.

A aleatorização dos tratamentos é considerada uma das principais contribuições de Fisher, uma vez que essa prática gera uma distribuição de amostragem para os erros experimentais cujo comportamento assintótico aproxima-se de distribuições normais não correlacionadas, proporcionando erros independentes e identicamente distribuídos com média zero e variância constante.

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

É importante ressaltar que a análise de variância clássica desconsidera a informação sobre as posições espaciais das unidades experimentais. Porém, uma vez sendo detectada a estrutura de correlação espacial, a utilização dessa informação na análise estatística fornece estimativas mais eficientes para as médias dos tratamentos, proporcionando resultados mais precisos ou tão bons quanto aqueles obtidos por meio da análise de variância clássica. Entretanto, se tal estrutura de variabilidade for desconsiderada, diferenças reais entre as médias comparadas podem não ser identificadas como significativas. Duarte (2000) ainda defende a utilização de uma análise que considere essa estrutura de dependência

espacial, afirmando que a consideração da informação espacial pode tornar-se um fator decisivo sobre as conclusões de um experimento.

Com isso, o presente trabalho foi realizado com o principal objetivo de desenvolver um material didático, descrevendo, detalhadamente, como implementar e executar a análise de variância com erros espacialmente correlacionados, na qual a matriz de covariância dos erros foi modelada por meio de uma abordagem geoestatística, em experimentos instalados sob os delineamentos inteiramente casualizados e em blocos casualizados. Nesse processo de desenvolvimento, foram ainda determinadas as esperanças dos quadrados médios, além de verificar a aplicabilidade do teste F para comparar a igualdade entre as médias dos fatores em estudo.

O semivariograma foi utilizado como principal ferramenta para detectar e descrever a estrutura de variabilidade espacial, de modo que os modelos teóricos ajustados ao mesmo foram utilizados para compor a matriz de covariância dos erros. Em síntese, a modelagem geoestatística do erro busca uma estimativa para a função geral de covariância, a qual pondera, além da solução para o sistema de equações normais (S.E.N), as somas de quadrados da análise de variância.

Quanto à estimação de parâmetros, foram utilizadas as teorias de quadrados mínimos generalizados e de máxima verossimilhança para a obtenção de estimadores para os parâmetros do modelo, levando em consideração a dependência espacial dos erros.

Com o objetivo de verificar a relação entre a eficiência da análise espacial e as configurações de dependência geoestatística, foram geradas diferentes combinações de valores dos parâmetros do semivariograma, seguindo modelos de semivariância esférico e exponencial.

Por fim, foi realizada uma comparação entre a análise clássica, cujos erros são considerados independentes e a análise espacial, em que a modelagem da correlação entre os erros foi feita por meio da Geoestatística. Nessa comparação avaliou-se o valor da estatística do teste F da análise de variância clássica e espacial, de modo que a análise que produziu maiores valores para essa estatística foi considerada a mais eficiente.

2 REFERENCIAL TEÓRICO

O referencial teórico é dividido em dois temas centrais, os quais abordam a Estatística Experimental e a Geoestatística.

2.1 Geoestatística

A necessidade do desenvolvimento e da utilização de ferramentas estatísticas, capazes de caracterizar a estrutura de variabilidade espacial de um determinado fenômeno, motivou o surgimento da teoria das variáveis regionalizadas ou, simplesmente, Geoestatística.

No início da década de 1950, Daniel G. Krige (engenheiro de minas que trabalhava com a mineração de ouro na África do Sul) observou que era preciso considerar a existência da variabilidade espacial para se obter métodos mais eficientes de estimação da concentração de ouro. Esses estudos despertaram o interesse de um engenheiro de minas francês chamado George Matheron. Já na década de 1960, Matheron conseguiu formalizar essas teorias relacionadas à variabilidade espacial, denominando-as de teoria das variáveis regionalizadas.

De acordo com Guerra (1988 citado por ROSSONI, 2011), as variáveis regionalizadas podem ser definidas como variáveis cujos valores estão relacionados de algum modo, com a posição espacial que ocupam, ou seja, que variam de um lugar a outro do espaço com uma certa continuidade. Complementando essa definição, Ribeiro Júnior (1995) entende que a Geoestatística baseia-se na utilização de um conjunto de procedimentos estatísticos aplicáveis a problemas nos quais os dados são espacialmente referenciados.

É natural pensar que, em fenômenos relacionados à geologia, ciências do solo, ciências florestais ou qualquer fenômeno cujos dados são espacialmente correlacionados, a associação de variáveis com dependência espacial em pontos distintos é maior à medida que se reduz a distância entre

elas, ou seja, variáveis amostradas em pontos mais próximos tendem a ser mais similares, entretanto, essa similaridade é menor à medida que aumenta a distância entre os pontos amostrados.

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

Existem várias medidas que se prestam à descrição e à modelagem das relações entre as distâncias e a associação espacial dessas variáveis, tais como a autocovariância, a autocorrelação e a semivariância. O prefixo ‘auto’ se justifica por se tratar da correlação de uma variável com ela mesma em pontos diferentes. Durante este texto, a utilização de tal prefixo será ignorada.

Na abordagem geoestatística, a medida normalmente utilizada é a semivariância, que pode ser definida como uma medida do grau de dependência espacial entre duas variáveis. É importante notar que, ao contrário da covariância e da correlação, a semivariância é uma medida de dissimilaridade, ou seja, é maior à medida que as variáveis estão menos associadas. Por exigir uma hipótese de estacionaridade menos restritiva em relação às demais, a semivariância é a medida mais utilizada pelos pesquisadores para descrever a estrutura de dependência espacial (RIBEIRO JÚNIOR, 1995).

2.1.1 Estacionariedade

A distribuição espacial de n valores amostrados de uma variável regionalizada Y pode ser modelada por um processo estocástico. Segundo

Oliveira (1991), um processo estocástico é definido como uma coleção $\{Y(x) : x \in R \subset \mathbb{R}^p\}$ de variáveis aleatórias reais, definidas sobre um mesmo espaço de probabilidade, indexadas em um subconjunto R do espaço vetorial p -dimensional \mathbb{R}^p , ou seja, $x \in R \subset \mathbb{R}^p$, em que \mathbb{R}^p é o espaço indexador das coordenadas x .

A Geoestatística baseia-se no pressuposto de que a distribuição espacial de uma variável Y em uma região R é uma realização $\{y(x) : x \in R \subset \mathbb{R}^p\}$ de um processo estocástico, em que $y(x)$ representa a realização de uma variável regionalizada $Y(x)$ na posição x sendo que, no caso bidimensional, x representa o vetor de coordenadas (u, v) de um ponto amostrado.

Conforme afirma Oliveira (1991), o espaço vetorial \mathbb{R}^p pode ser definido de tal maneira que seja possível representar variações aleatórias em espaços de qualquer dimensão. Por exemplo, o espaço vetorial $p = 1$ é utilizado para estudar principalmente as variações no tempo, já $p = 2$ estuda variações em superfícies, enquanto $p = 3$ é utilizado em variações que ocorrem no espaço tridimensional ou na superfície-tempo e $p = 4$ representa variações no espaço-tempo.

Ao contrário da análise estatística clássica, em que se trabalha com n observações de uma variável aleatória, na teoria das variáveis regionalizadas têm-se n variáveis aleatórias com apenas uma observação cada, ou seja, não existe a repetição de uma mesma variável. Essa restrição no número de repetições impede o estudo da distribuição dessas variáveis. Por esse motivo, é necessário supor que o processo obedeça a algum tipo de estacionaridade.

Segundo Guimarães (2004), um processo é estacionário se o desenvolvimento do mesmo, no tempo ou no espaço, ocorrer de forma homogênea, com pequenas oscilações aleatórias contínuas em torno de um valor médio. Desse modo, pode-se dizer que uma variável $Y(x)$ é estacionária de ordem k , se:

$$\begin{aligned}
E[Y(x)] &= m_1 = \text{constante } \forall x; \\
E[Y^2(x)] &= m_2 = \text{constante } \forall x; \\
&\vdots \\
E[Y^k(x)] &= m_k = \text{constante } \forall x;
\end{aligned} \tag{2.1}$$

em que m_k representa o k -ésimo momento da variável $Y(x)$.

Vale ressaltar que uma variável dita estacionária de ordem k é também estacionária para ordens inferiores a k , ou seja, é estacionária de ordem $k - 1, k - 2, \dots, 1$.

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

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

$$E[Y(x)] = \omega = \text{constante}; \tag{2.2}$$

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

$$C(Y(x+h), Y(x)) = E[Y(x+h) \cdot Y(x)] - \omega^2 = C(h). \tag{2.3}$$

Desse modo, pode-se definir o covariograma como sendo um gráfico no qual são plotados, no eixo das ordenadas, os valores da covariância em função das distâncias, que são representadas no eixo das abscissas.

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

$$Var(Y(x)) = C(Y(x+0), Y(x)) = C(0). \tag{2.4}$$

Apesar da pressuposição da estacionariedade de 2ª ordem ser viável em muitos casos, essa é uma hipótese muito restritiva, o que faz com que

nem sempre ela atendida. Por este motivo, é comum utilizar uma hipótese um pouco mais fraca, conhecida como hipótese intrínseca. Em estudos geoestatísticos, é necessário, no mínimo, que a variável em estudo obedeça a esta hipótese.

Segundo Cressie (1993), para que uma variável seja intrinsecamente estacionária é necessário que

- (i) a esperança da diferença da variável $Y(x)$ nas posições x e $x+h$ seja nula,

$$E[Y(x+h) - Y(x)] = 0; \quad (2.5)$$

- (ii) a variância da diferença $[Y(x+h) - Y(x)]$ exista e não dependa da posição espacial x ,

$$\frac{1}{2} \text{Var}[Y(x+h) - Y(x)] = \frac{1}{2} E\{ [Y(x+h) - Y(x)]^2 \} = \gamma(h). \quad (2.6)$$

em que $\gamma(h)$ representa a medida de dependência espacial denominada semivariância. A utilização do prefixo ‘semi’ se dá pelo fato de que $\gamma(h)$ representa a metade da variância. Por conveniência matemática, em relação à variância, é preferível trabalhar com a semivariância. Sendo assim, o semivariograma, um gráfico no qual são plotados os valores das semivariâncias em função das distâncias h , desponta como uma ferramenta fundamental na análise geoestatística.

Se um processo (ou variável) é estacionário de segunda ordem, ele também é intrínseco, mas o inverso nem sempre ocorre. De acordo com Journel e Huijbregts (1978), se a hipótese de estacionaridade de 2^a ordem for atendida, é possível estabelecer uma relação entre a covariância $C(h)$ e a semivariância $\gamma(h)$, que é dada por

$$C(h) = C(0) - \gamma(h). \quad (2.7)$$

Essa relação está representada graficamente no Gráfico 1. Nesse gráfico pode-se observar que, quando h tende a infinito, $C(h)$ tende a zero

e $\gamma(h)$ tende a $C(0)$.

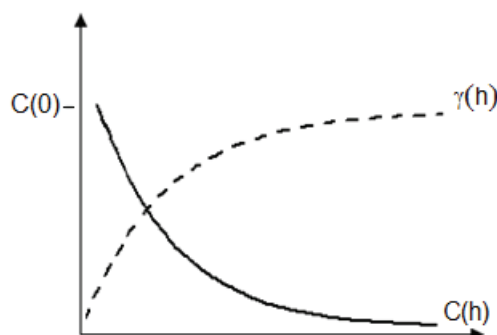


Gráfico 1 Relação entre as funções covariância e semivariância

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

$$\rho(h) = \frac{C(h)}{C(0)} = \frac{C(0) - \gamma(h)}{C(0)} = 1 - \frac{\gamma(h)}{C(0)}. \quad (2.8)$$

Segundo Guimarães (2004), esta função tem a vantagem de ser adimensional e estar limitada ao valor -1 e 1, permitindo comparações entre variáveis e também inferências sobre o grau de associação espacial. Porém, percebe-se que, na prática, a correlação de uma variável com ela mesma em pontos distintos varia entre 0 e 1, uma vez que, para $h = 0$, a correlação é máxima, $\rho(0) = 1$, e este valor vai decrescendo até o zero, ou seja, até uma distância em que não exista mais relação entre as observações.

Uma alternativa para analisar a dependência espacial é a construção do correlograma, que nada mais é do que um gráfico no qual são plotados os valores da correlação $\rho(h)$ em função das distâncias.

Com isso, conforme afirma Pontes (2002), se a hipótese de estacionariedade de 2^a ordem for atendida, o covariograma, o semivariograma e o correlograma tornam-se ferramentas equivalentes para caracterizar a dependência espacial. Porém, em situações cujo fenômeno em estudo apresenta capacidade infinita de dispersão, no qual somente a hipótese intrínseca está sendo satisfeita, apenas o semivariograma pode ser utilizado para descrever a dependência espacial. Por este motivo, o semivariograma se torna a ferramenta preferivelmente utilizada em uma análise geoestatística.

2.1.2 Estimação do semivariograma

Conforme afirma Journel (1987), a análise variográfica é uma arte, no melhor senso do termo. Segundo esse autor, artes requerem bons instrumentos, como neste caso, um bom programa interativo, além de experiência e habilidade do pesquisador para sintetizar e, muitas vezes, ir além dos dados.

Na busca por uma estimação cada vez mais eficaz, surgiram diversos estimadores de semivariograma. Segundo Ribeiro Júnior (1995), a escolha de um estimador é orientada por diversas propriedades, como não-tendenciosidade, precisão, resistência e robustez.

O estimador mais utilizado na literatura, proposto por Matheron (1962), é conhecido como estimador ‘clássico’ ou, simplesmente, estimador de Matheron. De acordo com Ribeiro Júnior (1995), esse estimador é obtido através do método dos momentos e tem, sob suposição que a variável atende à hipótese intrínseca, a qualidade de ser um estimador não tendencioso.

Conforme afirma Cressie (1993), o estimador clássico de semivariância pode ser definido por meio da seguinte expressão:

$$\hat{\gamma}(h) = \frac{1}{2n(h)} \sum_{i=1}^{n(h)} [y(x_i) - y(x_i + h)]^2, \quad (2.9)$$

em que

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

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

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

Entretanto, esse estimador apresenta como desvantagem o fato de ser influenciado por observações atípicas (*outliers*), o que pode ser justificado pelo termo ao quadrado que aparece no somatório. Sendo assim, se os dados apresentam *outliers*, outros estimadores de semivariância devem ser considerados para descrever a variabilidade espacial.

Um semivariograma típico com características ideais, cuja variável atende à estacionariedade de segunda ordem, pode ser ilustrado conforme o Gráfico 2.

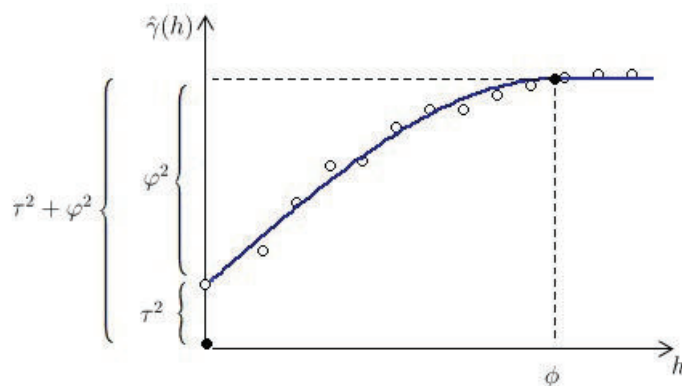


Gráfico 2 Exemplo de um semivariograma experimental com comportamento ideal

Analisando esse semivariograma, pode-se observar que, quanto mais próximos estiverem os pontos amostrados, maior será a semelhança entre eles e, portanto, menor a semivariância. Pode-se observar, ainda, que a semivariância $\gamma(h)$ cresce com o incremento de h , até atingir um valor constante e, a partir deste ponto, as variações tornam-se aleatórias, ou seja, as variações que ocorrem não podem ser justificadas pela semelhança de um ponto com outro.

No Gráfico 2 estão apresentados ainda os principais parâmetros do semivariograma. Segundo Isaaks e Srivastava (1989), esses parâmetros podem ser definidos como:

Alcance (ϕ): distância dentro da qual as amostras apresentam-se correlacionadas espacialmente. Esse parâmetro representa o valor de h para o qual o semivariograma atinge seu patamar. Deste ponto em diante, considera-se que não existe mais dependência espacial entre os pontos amostrados e, com isso, $\gamma(h)$ torna-se invariante com a distância.

Patamar ($\tau^2 + \varphi^2$): é o valor da semivariância correspondente ao valor do alcance (ϕ), isto é, $\gamma(\phi) = \tau^2 + \varphi^2$. O patamar corresponde, teoricamente, à variância dos dados. Variáveis que atendem apenas à hipótese intrínseca não apresentam patamar definido.

Efeito pepita (τ^2): representa a descontinuidade do semivariograma para distâncias menores do que a menor distância observada na amostra. Teoricamente $\gamma(0) = 0$, porém, na prática percebe-se que a medida que h tende a zero, $\gamma(h)$ aproxima-se de um valor positivo chamado de efeito pepita (τ^2). Esse parâmetro está associado à variabilidade totalmente aleatória dos dados, ou seja, refere-se à variância do erro experimental.

Contribuição (φ^2): é a diferença entre o patamar e o efeito pepita. Esse valor está associado à variância estrutural, ou seja, à variabilidade regionalizada que pode ser explicada por um modelo com uma certa estrutura de dependência espacial.

É possível avaliar e quantificar a dependência espacial de uma variável. Uma das medidas utilizadas para tal avaliação é o grau de

dependência. Essa medida representa nada mais do que a razão entre o valor do efeito pepita e do patamar do semivariograma. Segundo Cambardella et al. (1994), pode-se classificar uma variável regionalizada, com relação ao seu grau de dependência espacial, da seguinte forma:

- (i) *variável com forte dependência espacial*: se o efeito pepita for menor ou igual a 25% do patamar, ou seja, $\tau^2/(\tau^2 + \varphi^2) \leq 0,25$;
- (ii) *variável com moderada dependência espacial*: se o efeito pepita representar entre 25% e 75% do patamar, isto é, $0,25 < \tau^2/(\tau^2 + \varphi^2) \leq 0,75$;
- (iii) *variável com fraca dependência espacial*: se a relação entre efeito pepita e patamar estiver entre 75% e 100% , ou seja, $0,75 < \tau^2/(\tau^2 + \varphi^2) < 1$;
- (iv) *variável independente espacialmente*: se a relação entre efeito pepita e patamar for igual a 100%. Neste caso, tem-se um semivariograma com efeito pepita puro.

2.1.3 Modelos teóricos de semivariograma

Conforme afirma Cressie (1993), a Geoestatística baseia-se, fundamentalmente, em dois conceitos: o semivariograma, cujo papel é descrever a estrutura da variabilidade espacial e a krigagem, que tem o objetivo de prever, não tendenciosamente e com variância mínima, os valores de um determinado atributo em pontos não amostrados. Desse modo, como afirma Rossoni (2011), por meio de um processo discreto de amostragem, é possível obter uma estimativa contínua de toda a área analisada.

Entretanto, Ribeiro Júnior (1995) reporta que, para assegurar que o sistema de krigagem tenha solução única, é necessário que a matriz de coeficientes desse sistema seja positiva definida. Consequentemente, é

necessário que o modelo de semivariância seja *negativo definido condicional*. De acordo com Cressie (1993), para que tal condição seja satisfeita, é preciso que

$$\sum_{i=1}^k \sum_{j=1}^k a_i a_j \gamma(x_i - x_j) \leq 0, \quad (2.10)$$

para qualquer posição espacial $x_1, x_2, \dots, x_k \in R$, e qualquer $a_1, a_2, \dots, a_k \in \mathbb{R}$ de tal modo que $\sum_{i=1}^k a_i = 0, \quad \forall k \in \mathbb{Z}^+$.

Diante dessa restrição, surge uma classe de modelos que são definidos como modelos autorizados. Alguns desses modelos, frequentemente utilizados na literatura, são os modelos esférico, exponencial e gaussiano, cujas semivariâncias são calculados por:

Modelo esférico:

$$\gamma(h) = \begin{cases} 0 & , \text{ se } h = 0; \\ \tau^2 + \varphi^2 \left[\frac{3}{2} \left(\frac{h}{\phi} \right) - \frac{1}{2} \left(\frac{h}{\phi} \right)^3 \right] & , \text{ se } 0 < h < \phi; \\ \tau^2 + \varphi^2 & , \text{ se } h \geq \phi; \end{cases} \quad (2.11)$$

Modelo exponencial:

$$\gamma(h) = \begin{cases} 0 & , \text{ se } h = 0; \\ \tau^2 + \varphi^2 \left[1 - \exp\left(\frac{-3h}{\phi}\right) \right] & , \text{ se } h \neq 0; \end{cases} \quad (2.12)$$

Modelo gaussiano:

$$\gamma(h) = \begin{cases} 0 & , \text{ se } h = 0; \\ \tau^2 + \varphi^2 \left[1 - \exp \left[- \left(\frac{\sqrt{3}h}{\phi} \right)^2 \right] \right] & , \text{ se } h \neq 0. \end{cases} \quad (2.13)$$

Exemplos desses modelos são apresentados no Gráfico 3. Os modelos citados são conhecidos como modelos “bem comportados”, uma vez que possuem patamares definidos, ou seja, a partir de um certo alcance, a variância se estabiliza. Porém, para os modelos exponencial e gaussiano, o patamar é obtido apenas assintoticamente. Para esses modelos, o alcance é determinado como sendo o valor da distância na qual se alcança 95% da contribuição (φ^2), ou seja, $\gamma(\phi) = 0,95\varphi^2$. Esse alcance, representado na expressão do modelo pela letra grega ϕ , é também chamado de alcance prático, ou seja, a partir desse ponto considera-se que não existe mais dependência espacial.

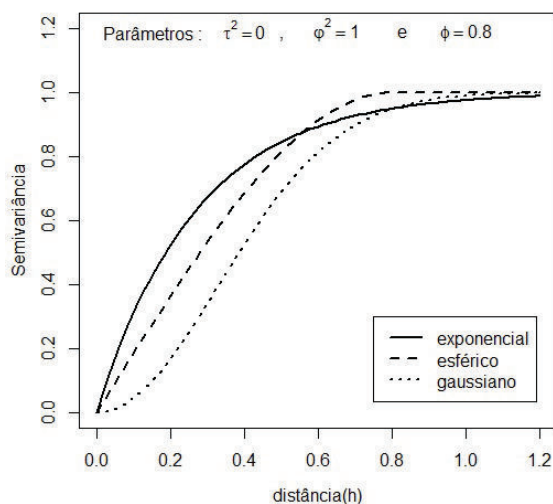


Gráfico 3 Exemplos dos modelos teóricos esférico, exponencial e gaussiano

No entanto, existem fenômenos em que o crescimento de $\gamma(h)$ é

contínuo com o incremento de h . Fenômenos desse tipo apresentam variância infinita, atendendo somente à hipótese intrínseca. Nesse tipo de situação, é comum ajustar o modelo potência, que se caracteriza por não apresentar patamar definido. A função de semivariância desse modelo é dada por:

Modelo potência:

$$\gamma(h) = \begin{cases} 0 & , \text{ se } h = 0; \\ \tau^2 + bh^\lambda & , \text{ se } h \neq 0; \end{cases} \quad (2.14)$$

com $\tau^2 > 0$, $b > 0$ e $0 \leq \lambda \leq 2$, sendo que τ^2 refere-se ao efeito pepita, b e λ representam, respectivamente, parâmetros de curvatura e intensidade de variação. Em modelos sem patamar, os parâmetros alcance (ϕ) e patamar ($\tau^2 + \varphi^2$) não têm interpretação prática.

O modelo linear é conhecido como um caso particular do modelo potência quando $\lambda = 1$. Nesse modelo, o valor da semivariância $\gamma(h)$ se caracteriza por crescer linearmente com o aumento da distância h . Exemplos desse modelo e do potência são apresentados no Gráfico 4.

Existem ainda casos em que o semivariograma se caracteriza por apresentar um conjunto de pontos oscilando aleatoriamente em torno de um valor constante de $\gamma(h)$. Em fenômenos dessa natureza é possível ajustar um modelo conhecido como modelo de efeito pepita puro. No Gráfico 5 pode-se observar o comportamento desse modelo. A sua função de semivariância é dada por

Modelo efeito pepita puro:

$$\gamma(h) = \begin{cases} 0 & , \text{ se } h = 0; \\ \tau^2 & , \text{ se } h \neq 0. \end{cases} \quad (2.15)$$

Se as informações presentes em um semivariograma podem ser explicadas por meio do ajuste de um modelo de efeito pepita puro, provavelmente a variável em estudo apresenta independência espacial, ou

seja, toda a variabilidade presente nos dados é aleatória. Em situações como essa, as próprias técnicas usuais da estatística clássica podem ser aplicadas para analisar os dados. Contudo, o pesquisador deve estar ciente de que, mesmo tendo ocorrido tal comportamento, pode ser que exista dependência espacial a distâncias menores do que a menor distância amostrada.

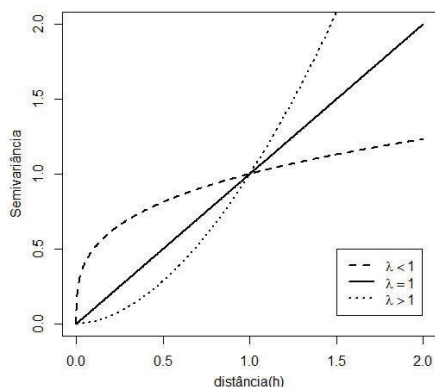


Gráfico 4 Modelos linear e potência

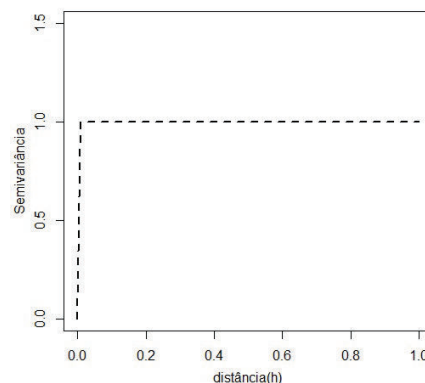


Gráfico 5 Modelo efeito pepita puro

Outros modelos teóricos de semivariograma podem ser encontrados em Journel e Huijbregts (1978). Além destes modelos básicos, Isaaks e Srisvatava (1989) salientam que qualquer combinação linear desses modelos atendem às condições de *negativo definido condicional*, o que amplia a quantidade de modelos válidos para ajustar um semivariograma.

No intuito de se ajustar um modelo teórico ao semivariograma, diversos métodos são apresentados na literatura. Segundo Ribeiro Júnior (1995), espera-se que o ajuste de um semivariograma seja eficiente principalmente para pequenas distâncias. Este é um dos motivos pelos quais o ajuste ‘*a sentimento*’ era amplamente utilizado. Por meio deste ajuste, as estimativas dos parâmetros são obtidas por meio da avaliação visual, sem qualquer critério formal de estimação. No entanto, pela subjetividade do ajuste, este método é comumente questionado. Com isso, devido ao avanço

dos recursos computacionais e à apresentação de métodos teoricamente fundamentados, outros métodos surgem como opção para o ajuste do semivariograma.

Entre os métodos frequentemente utilizados para o ajuste dos semivariogramas destacam-se os de quadrados mínimos (ordinários, ponderados e generalizados), que têm como base o estimador de semivariância. Além desses, destacam-se ainda os métodos de máxima verossimilhança e máxima verossimilhança restrita, que são aplicados diretamente ao conjunto de dados. Maiores detalhes sobre os métodos citados acima podem ser encontrados em Cressie (1993), Diggle e Ribeiro Júnior (2007) e McBratney e Webster (1986), entre outros.

2.1.4 Modelos espaciais lineares

Conforme se afirmou na seção 2.1.1, a distribuição espacial de uma variável $Y(x)$ pode ser modelada por meio de um processo estocástico $\{Y(x) : x \in R \subset \mathbb{R}^p\}$. Pressupondo que $Y(x)$ é estacionária de segunda ordem, pode-se representar essa variável por meio de sua decomposição em uma soma de três componentes: uma tendência constante, uma componente aleatória, espacialmente correlacionada e, por fim, um erro experimental. Desse modo, conforme afirma Burrough (1987), o valor da variável aleatória Y , na posição espacial x , é dada por:

$$Y(x) = \omega(x) + S(x) + \xi(x) \quad (2.16)$$

em que

$\omega(x)$ é um componente determinístico, uma tendência associada a um valor médio constante, de tal forma que $E[\omega(x)] = \omega(x)$ e $C[\omega(x_i), \omega(x_j)] = 0, \forall x_i, x_j \in R$;

$S(x)$ é uma variável aleatória com dependência espacial. É neste componente que está a contribuição da Geoestatística. Assume-se que

esta variável tem média zero, $E [S(x)] = 0$ e função de covariância $C[S(x_i), S(x_j)]$, parametrizada por ϕ , parâmetro de alcance do semivariograma. Se $i = j$, $C[S(x_i), S(x_j)] = \varphi^2$. Caso contrário, $C(\cdot)$ assume algum valor, que não é necessariamente zero;

$\xi(x)$ é uma variável aleatória não correlacionada, frequentemente denominada erro experimental. Essa variável é distribuída com média zero, $E [\xi(x)] = 0$, variância dada pelo parâmetro efeito pepita do semivariograma, $Var[\xi(x)] = \tau^2$ e, se $i \neq j$, $C[\xi(x_i), \xi(x_j)] = 0$, $\forall x_i, x_j \in R$.

As variáveis $S(x)$ e $\xi(x)$ são definidas de tal modo que $C[S(x), \xi(x)] = 0$, ou seja, $S(x)$ e $\xi(x)$ são espacialmente independentes. Com isso, para qualquer posição espacial $x_i, x_j \in R$, tem-se que a média e a covariância da variável aleatória $Y(x)$ são dadas por

$$\begin{aligned} E [Y(x)] &= \omega(x); \\ C[Y(x_i), Y(x_j)] &= \tau^2 + \varphi^2, \text{ se } i = j; \\ C[Y(x_i), Y(x_j)] &= C[S(x_i), S(x_j)], \text{ se } i \neq j. \end{aligned} \quad (2.17)$$

De modo geral, supondo-se distribuição gaussiana, a distribuição do vetor de erros aleatórios é dada por $\xi \sim N(\mathbf{0}, \mathbf{I}\tau^2)$, em que $\mathbf{0}$ corresponde a um vetor de zeros e \mathbf{I} representa uma matriz identidade de ordem n , sendo n o número de valores amostrados da variável Y . Da mesma forma, a distribuição do vetor S , referente à componente espacialmente estruturada, é dada por $S \sim N(\mathbf{0}, \mathbf{F}\varphi^2)$, sendo \mathbf{F} uma matriz quadrada de ordem n , cujo ij -ésimo elemento contém os valores da função $f(h_{ij})$, a qual corresponde aos modelos de covariância utilizados para descrever a variabilidade espacial da variável $S(x)$ nas posições espaciais x_i e x_j , de modo que $h_{ij} = \|x_i - x_j\|$ é a distância euclidiana entre x_i e x_j .

Com isso, é possível concluir que o vetor Y , constituído por $Y(x_i)$ em todas as posições espaciais $x_i \in R$ amostradas, segue distribuição gaussiana, de modo que $Y \sim N(\omega, \mathbf{I}\tau^2 + \mathbf{F}\varphi^2)$. No desenvolvimento deste texto, comumente, $(\mathbf{I}\tau^2 + \mathbf{F}\varphi^2)$ será representado por Σ . Para facilitar a

visualização e a interpretação dessa matriz de covariâncias, observe que:

$$\Sigma = \begin{bmatrix} C[Y(x_1), Y(x_1)] & C[Y(x_1), Y(x_2)] & \cdots & C[Y(x_1), Y(x_n)] \\ C[Y(x_2), Y(x_1)] & C[Y(x_2), Y(x_2)] & \cdots & C[Y(x_2), Y(x_n)] \\ \vdots & \vdots & \ddots & \vdots \\ C[Y(x_n), Y(x_1)] & C[Y(x_n), Y(x_2)] & \cdots & C[Y(x_n), Y(x_n)] \end{bmatrix}.$$

Sabe-se que, teoricamente, a variância dos dados corresponde ao patamar do semivariograma, e, com isso, tem-se que $Var[Y(x_i)] = C[Y(x_i), Y(x_i)] = \tau^2 + \varphi^2$, $\forall i = 1, \dots, n$. Além disso, assumindo que a variável $Y(x)$ é estacionária de segunda ordem, $C[Y(x_i), Y(x_j)] = C(h_{ij})$, em que $h_{ij} = \|x_i - x_j\|$. Por isso, a matriz Σ pode ser reescrita como

$$\Sigma = \begin{bmatrix} \tau^2 + \varphi^2 & C(h_{12}) & \cdots & C(h_{1n}) \\ C(h_{12}) & \tau^2 + \varphi^2 & \cdots & C(h_{2n}) \\ \vdots & \vdots & \ddots & \vdots \\ C(h_{1n}) & C(h_{2n}) & \cdots & \tau^2 + \varphi^2 \end{bmatrix}.$$

Considerando a decomposição dessa matriz como $\Sigma = \mathbf{I}\tau^2 + \mathbf{F}\varphi^2$, a matriz \mathbf{F} é, então, dada por

$$\mathbf{F} = \begin{bmatrix} 1 & \frac{C(h_{12})}{\varphi^2} & \cdots & \frac{C(h_{1n})}{\varphi^2} \\ \frac{C(h_{12})}{\varphi^2} & 1 & \cdots & \frac{C(h_{2n})}{\varphi^2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{C(h_{1n})}{\varphi^2} & \frac{C(h_{2n})}{\varphi^2} & \cdots & 1 \end{bmatrix}.$$

Tipicamente, em aplicações geoestatísticas, um dos principais interesses é fazer predições sobre os valores $S(\cdot)$ em toda a área analisada.

Nesses casos, segundo Diggle e Ribeiro Júnior (2007), o problema de predição reduz-se ao estudo da distribuição condicional de $S(\cdot)$ dadas as observações $y(x_i)$ como realização das variáveis aleatórias $Y(x_i)$, em que $x_i \in R$. Esses autores afirmam ainda que, considerando um único valor S_0 , referente à posição espacial x_0 , sob suposição de modelo gaussiano, a distribuição preditiva de $[S_0 | y]$ é gaussiana com média e variância dadas por

$$E [S_0 | y] = E [S(x_0)] + \varphi^2 v' (\mathbf{I}\tau^2 + \mathbf{F}\varphi^2)^{-1} (Y - \omega) \quad \text{e} \quad (2.18)$$

$$Var [S_0 | y] = \varphi^2 - \varphi^2 v' (\mathbf{I}\tau^2 + \mathbf{F}\varphi^2)^{-1} \varphi^2 v. \quad (2.19)$$

em que v' representa o vetor transposto de v , sendo que este, por sua vez, é um vetor cujo i -ésimo elemento é dado por $\rho(h_{0i})$, $\forall i = 1, \dots, n$, com $h_{0i} = \|x_0 - x_i\|$.

De maneira geral, dado um vetor S com média zero, a esperança do vetor $[S | y]$ é dada por

$$E [S | y] = \varphi^2 \mathbf{V}' (\mathbf{I}\tau^2 + \mathbf{F}\varphi^2)^{-1} (Y - \omega), \quad (2.20)$$

sendo \mathbf{V} uma matriz quadrada de ordem n , cujas colunas são formadas pelos vetores v_1, \dots, v_n , ou seja

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

É interessante notar que a matriz \mathbf{V} é simétrica, isto é, $\mathbf{V}' = \mathbf{V}$. Vale ressaltar ainda que, se o efeito pepita (τ^2) for nulo, as matrizes \mathbf{F} e \mathbf{V} são idênticas.

2.2 Estatística Experimental

A Estatística Experimental é uma ciência que tem como objetivo estudar experimentos cujas etapas incluem o planejamento, a execução, a coleta e a análise dos dados experimentais e interpretação dos resultados obtidos.

Em 1925, Fisher estabeleceu três princípios fundamentais para a análise de delineamentos experimentais, até hoje considerados o alicerce da experimentação, que são: a repetição, a aleatorização e o controle local.

(i) *Repetição*

A repetição refere-se ao número de parcelas que receberão um determinado tratamento. Ela é necessária para que se possa estimar o erro experimental, sem o qual seria impossível realizar qualquer tipo de inferência estatística. O uso de um número adequado de repetições possibilita uma boa estimativa do erro experimental, aumentando a precisão do experimento e, com isso, melhorando as estimativas de médias e diferenças entre médias.

(ii) *Aleatorização*

Esse princípio refere-se à distribuição aleatória dos tratamentos nas unidades experimentais, de tal modo que todos os tratamentos tenham a mesma probabilidade de serem alocados em qualquer uma das unidades experimentais, evitando, assim, que qualquer tratamento seja beneficiado pela unidade experimental a ele designada.

Segundo Kempthorne (1977), a aleatorização dos tratamentos foi uma das principais contribuições de Fisher, uma vez que a prática da aleatorização gera uma distribuição de amostragem para os erros experimentais, cujo comportamento assintótico aproxima-se de distribuições normais não correlacionadas. Por isso, Banzatto e Kronka (2006) afirmam que a aleatorização tem a finalidade de proporcionar uma estimativa válida para o erro experimental.

(iii) *Controle local*

O princípio de controle local tem a finalidade de tornar o experimento mais eficiente, por meio da redução do erro experimental. Este princípio é utilizado para atenuar problemas de heterogeneidade ambiental. Segundo Hinkelmann e Kempthorne (1994), a ideia básica do controle local é a partição do conjunto total heterogêneo de unidades experimentais em subconjuntos (blocos) que sejam os mais homogêneos possíveis. Com essa restrição, a aleatorização passa a ser feita dentro de cada bloco. Apesar de frequentemente utilizado, é importante ressaltar que este não é um princípio obrigatório, uma vez que, se o experimento for implantado em um ambiente homogêneo, não é necessária a sua aplicação.

2.2.1 A análise de variância

Uma das ferramentas para a análise e a interpretação de dados, proposta por Fisher, é a análise de variância, comumente denotada por ANOVA. De acordo com Banzatto e Kronka (2006), a análise de variância é uma técnica que consiste na decomposição da variância total (e dos graus de liberdade) em partes atribuídas às fontes de variação previamente definidas no planejamento do experimento, além de uma porção residual correspondente aos fatores não controlados no experimento, de natureza aleatória. O principal objetivo desta análise é avaliar as possíveis fontes de variação, verificando se existem diferenças significativas entre os níveis dos fatores controlados.

Cochran (1947) lembra que, para a construção de uma análise de variância, algumas pressuposições devem ser atendidas. Estas pressuposições são conhecidas como

1. *aditividade*: os efeitos dos fatores controlados e do erro devem ser aditivos;
2. *independência*: os erros devem ser independentes, ou seja, a

probabilidade de que o erro tenha um determinado valor não deve depender dos valores dos outros erros;

3. *normalidade*: os erros devem ser normalmente distribuídos;

4. *homogeneidade*: os erros devem apresentar variância comum.

Segundo Banzatto e Kronka (2006), as pressuposições citadas acima são necessárias para garantir a validade do teste F , que tem como finalidade comparar estimativas de variâncias. Por meio desse teste é possível realizar inferências sobre a hipótese de igualdade entre várias médias dos níveis de algum fator de interesse.

Frequentemente, o teste F surge pela consideração da decomposição da variabilidade de um conjunto de dados em termos de somas de quadrados. A estatística F_0 deste teste é obtida pela razão de quadrados médios que, por sua vez, são obtidos pela divisão da soma de quadrados de cada fator de interesse pelos seus respectivos graus de liberdade.

Os valores dos quadrados médios são estimativas de variâncias que refletem diferentes fontes de variabilidade. Essas fontes de variabilidade estão relacionadas com o delineamento experimental planejado para o experimento.

O delineamento experimental é o plano de distribuição dos tratamentos na área experimental, ou seja, são as formas de distribuição das parcelas experimentais na área do experimento. Os delineamentos experimentais mais utilizados na literatura são os delineamentos inteiramente casualizado (DIC) e em blocos casualizados (DBC).

2.2.2 Delineamento inteiramente casualizado (DIC)

O delineamento experimental inteiramente casualizado é utilizado quando a variabilidade entre as parcelas experimentais for muito pequena, isto é, praticamente inexistente. Devido a essa exigência, este delineamento

só deve ser utilizado se houver homogeneidade entre as unidades experimentais.

Nesse delineamento, os tratamentos são distribuídos nas parcelas de forma inteiramente aleatória, de modo que cada unidade experimental tenha a mesma probabilidade de receber qualquer um dos tratamentos estudados, sem nenhuma restrição no critério de aleatorização.

Por se fundamentar apenas nos princípios da aleatorização e da repetição, um experimento instalado em DIC deve ser implantado em locais cujas condições experimentais possam ser controladas, de modo que o princípio de controle local seja desnecessário. O DIC, normalmente, é mais utilizado em experimentos de laboratório ou em experimentos em vasos e bandejas em casa de vegetação, onde há possibilidade de controle das condições ambientais. A instalação do DIC no campo experimental exige homogeneidade das condições ambientais (tais como, a fertilidade do solo, a distribuição uniforme de água e de iluminação, etc).

Segundo Banzatto e Kronka (2006), as principais vantagens do delineamento inteiramente casualizado são:

- (i) é um delineamento bastante flexível, visto que o número de tratamentos e de repetições depende apenas do número de parcelas disponíveis;
- (ii) o número de repetições pode ser diferente de um tratamento para outro, embora o ideal seja que eles se apresentem igualmente repetidos;
- (iii) a análise estatística é simples, mesmo quando o número de repetições por tratamento é variável;
- (iv) o número de graus de liberdade para o resíduo é o maior possível;
- (v) é o delineamento mais simples de ser instalado e conduzido.

Este delineamento possui apenas duas causas ou fontes de variação: os tratamentos, que se referem ao efeito de um fator controlado e o resíduo, que reflete o efeito dos fatores não controlados. Desse modo, as variações de toda ordem, exceto apenas a variação que se atribui aos tratamentos,

compõem o erro experimental. Sendo assim, se o experimento em DIC for instalado em condições heterogêneas, altos valores para a estimativa da variância residual serão obtidos.

O modelo estatístico deste delineamento é dado por

$$y_{ij} = \mu + \alpha_i + e_{ij}, \quad (2.21)$$

em que

y_{ij} é a observação do i -ésimo tratamento na j -ésima repetição, com $i = 1, 2, \dots, t$ e $j = 1, 2, \dots, r$;

μ é uma constante inerente a cada observação que, sob restrição de que $\sum_{i=1}^t \alpha_i = 0$, é a média geral;

α_i é o efeito do i -ésimo tratamento;

e_{ij} é o erro aleatório associado à ij -ésima observação, em que se supõe que $e_{ij} \sim N(0, \sigma^2)$.

O objetivo da análise de experimentos em DIC é estimar as médias dos tratamentos e testar hipóteses apropriadas para compará-las. Segundo Montgomery (2008), para testar hipóteses sobre as médias dos tratamentos é necessário assumir que os erros são normalmente e independentemente distribuídos com média zero e variância comum, $e_{ij} \sim N(0, \sigma^2)$. Se α_i for um efeito fixo, isso implica que

$$Y_{ij} \sim N(\mu + \alpha_i, \sigma^2) \quad (2.22)$$

e que as observações são mutuamente independentes.

Este autor ainda afirma que, definindo $\mu_i = \mu + \alpha_i$ como a média do i -ésimo tratamento, as hipóteses apropriadas para testar a igualdade entre as médias são

$$\begin{aligned} H_0 : \mu_1 = \mu_2 = \dots = \mu_t \\ H_1 : \mu_i \neq \mu_j, \text{ para algum } i \neq j, \end{aligned} \quad (2.23)$$

ou, equivalentemente,

$$\begin{aligned} H_0 : \alpha_1 = \dots = \alpha_t = 0 \\ H_1 : \alpha_i \neq 0, \text{ para algum } i. \end{aligned} \quad (2.24)$$

O procedimento adequado para testar as hipóteses sobre a igualdade entre as médias dos tratamentos é a análise de variância. Antes de apresentar o quadro de análise de variância, é comum denotar $y_{i\cdot}$ como a soma das observações sujeitas ao i -ésimo tratamento e $\bar{y}_{i\cdot}$ como a média das observações que receberam o i -ésimo tratamento. Similarmente, $y_{\cdot\cdot}$ representa a soma geral de todas as observações do experimento e $\bar{y}_{\cdot\cdot}$ define a média geral de todas as observações. Simbolicamente, tem-se que

$$y_{i\cdot} = \sum_{j=1}^r y_{ij}, \quad \bar{y}_{i\cdot} = \frac{y_{i\cdot}}{t}, \quad y_{\cdot\cdot} = \sum_{i=1}^t \sum_{j=1}^r y_{ij} \quad \text{e} \quad \bar{y}_{\cdot\cdot} = \frac{y_{\cdot\cdot}}{tr}.$$

Desse modo, de acordo com Banzatto e Kronka (2006), a análise de variância para um delineamento inteiramente casualizado (DIC) com t tratamentos e r repetições é dada conforme a Tabela 1,

Tabela 1 Análise de variância para um delineamento inteiramente casualizado.

| FV | GL | SQ | QM | F_0 |
|------------|----------|-------------|--------------------------------------|------------------------------|
| Tratamento | $(t-1)$ | SQ_{Trat} | $QM_{Trat} = \frac{SQ_{Trat}}{t-1}$ | $\frac{QM_{Trat}}{QM_{Res}}$ |
| Resíduo | $t(r-1)$ | SQ_{Res} | $QM_{Res} = \frac{SQ_{Res}}{t(r-1)}$ | |
| Total | $(tr-1)$ | SQT | | |

em que

$$SQTrat = \frac{1}{r} \sum_{i=1}^t y_{i.}^2 - \frac{y_{..}^2}{tr}, \quad SQT = \sum_{i=1}^t \sum_{j=1}^r y_{ij}^2 - \frac{y_{..}^2}{tr} \quad e$$

$$SQRes = \sum_{i=1}^t \sum_{j=1}^r y_{ij}^2 - \frac{1}{r} \sum_{i=1}^t y_{i.}^2 = SQT - SQTrat.$$

Segundo Barbin (1998), considerando que α_i é um efeito fixo, ou seja, α_i representa constantes a serem estimadas, as esperanças dos quadrados médios da análise variância da Tabela 1 são dadas por

$$E[QMTrat] = \sigma^2 + \frac{r}{t-1} \sum_{i=1}^t \alpha_i^2 \quad e \quad E[QMRes] = \sigma^2. \quad (2.25)$$

2.2.3 Delineamento em blocos casualizados (DBC)

Em 1925, Ronald A. Fisher idealizou os delineamentos em blocos, cujo propósito era eliminar o efeito da heterogeneidade presente entre as unidades experimentais na comparação dos tratamentos (princípio de controle local). Fisher propôs que se fizessem grupos de parcelas homogêneas (blocos), os quais receberiam, cada um, uma repetição de cada tratamento. Esse delineamento foi denominado blocos completos casualizados, caracterizando-se pelo aparecimento de todos os tratamentos em cada bloco. A partir daí, principalmente em experimentação de campo, o delineamento em blocos casualizados tornou-se o mais utilizado, tendo em vista que, frequentemente, as condições ambientais sob as quais os experimentos são implantados são heterogêneas.

A eficiência do DBC depende da uniformidade dentro de cada bloco podendo, entretanto, haver heterogeneidade entre os mesmos. Segundo Gusmão (1986 citado por DUARTE, 2000), três são os aspectos necessários para se usufruir das vantagens do uso do controle local e permitir testes

de hipóteses corretos: homogeneidade dentro dos blocos, heterogeneidade entre blocos e nenhuma interação entre tratamentos e blocos.

No DBC, os blocos representam uma restrição na aleatorização, de modo que os tratamentos são aleatoriamente designados às unidades dentro de cada bloco.

Conforme afirmam Banzatto e Kronka (2006), a principal vantagem deste delineamento é o controle entre as diferenças que ocorrem nas condições experimentais, de um bloco para outro, além da obtenção de uma estimativa mais exata para a variância residual, uma vez que a variação ambiental entre blocos é isolada.

Em relação aos outros delineamentos, segundo estes mesmos autores, o delineamento em blocos casualizados apresenta as seguintes desvantagens:

- (i) pela utilização do princípio do controle local, há uma redução no número de graus de liberdade do resíduo;
- (ii) a exigência de homogeneidade das parcelas dentro de cada bloco limita o número de tratamentos, que não pode ser muito elevado.

O modelo estatístico deste delineamento, considerando-se um modelo fixo, é dado por

$$y_{ij} = \mu + \alpha_i + \beta_j + e_{ij}, \quad (2.26)$$

em que

y_{ij} é a observação do i -ésimo tratamento do j -ésimo bloco, com $i = 1, 2, \dots, t$ e $j = 1, 2, \dots, r$;

μ é uma constante inerente a cada observação que, sob restrição de que

$$\sum_{i=1}^t \alpha_i = 0 \text{ e } \sum_{j=1}^r \beta_j = 0, \text{ é a média geral;}$$

α_i é o efeito do i -ésimo tratamento;

β_j é o efeito relacionado ao bloco j ;

e_{ij} é o erro aleatório associado à ij -ésima observação, o qual supõe-se $e_{ij} \sim N(0, \sigma^2)$.

Assim como no delineamento inteiramente casualizado, para avaliar a igualdade entre as médias dos tratamentos, as hipóteses apropriadas são

$$\begin{aligned} H_0 : \mu_1 = \mu_2 = \dots = \mu_t \\ H_1 : \mu_i \neq \mu_j, \text{ para algum } i \neq j, \end{aligned} \quad (2.27)$$

em que $\mu_i = \mu + \alpha_i$ representa a média do i -ésimo tratamento. Ou, equivalentemente,

$$\begin{aligned} H_0 : \alpha_1 = \dots = \alpha_t = 0 \\ H_1 : \alpha_i \neq 0, \text{ para algum } i. \end{aligned} \quad (2.28)$$

Entretanto, conforme afirma Montgomery (2008), em um delineamento em blocos casualizados é possível ainda avaliar os efeitos dos blocos por meio das seguintes hipóteses:

$$\begin{aligned} H_0 : \beta_1 = \dots = \beta_r = 0 \\ H_1 : \beta_j \neq 0, \text{ para algum } j. \end{aligned} \quad (2.29)$$

Assim como no DIC, o procedimento adequado para testar a igualdade das t médias dos tratamentos e a igualdade entre as r médias dos blocos é a análise de variância. Denotando $y_{.j}$ como a soma das observações do j -ésimo bloco e $\bar{y}_{.j}$ como a média das observações do j -ésimo bloco, de acordo com Banzatto e Kronka (2006), a análise de variância de um DBC com t tratamentos e r blocos é dada conforme a Tabela 2,

Tabela 2 Análise de variância para um delineamento em blocos casualizados.

| FV | GL | SQ | QM | F_0 |
|------------|--------------|----------|--------------------------------|------------------------|
| Blocos | $(r-1)$ | SQB | $QMB = \frac{SQB}{r-1}$ | $\frac{QMB}{QMRes}$ |
| Tratamento | $(t-1)$ | $SQTrat$ | $QMTrat = \frac{SQTrat}{t-1}$ | $\frac{QMTrat}{QMRes}$ |
| Resíduo | $(t-1)(r-1)$ | $SQRes$ | $QMRes = \frac{SQRes}{t(r-1)}$ | |
| Total | $(tr-1)$ | SQT | | |

em que:

$$SQTrat = \frac{1}{r} \sum_{i=1}^t y_{i.}^2 - \frac{y_{..}^2}{tr}, \quad SQB = \frac{1}{t} \sum_{j=1}^r y_{.j}^2 - \frac{y_{..}^2}{tr},$$

$$SQRes = \sum_{i=1}^t \sum_{j=1}^r y_{ij}^2 - \frac{1}{r} \sum_{i=1}^t y_{i.}^2 - \frac{1}{t} \sum_{j=1}^r y_{.j}^2 + \frac{y_{..}^2}{tr} = SQT - SQTrat - SQB,$$

$$e \quad SQT = \sum_{i=1}^t \sum_{j=1}^r y_{ij}^2 - \frac{y_{..}^2}{tr}.$$

De acordo com Barbin (1998), se α_i e β_j são efeitos fixos, as esperanças dos quadrados médios de tratamento e resíduo da análise variância da Tabela 2 são exatamente iguais às esperanças dos quadrados médios de um DIC, apresentadas na expressão (2.25). Segundo este autor, a esperança do quadrado médio dos blocos é expressa por

$$E[QMB] = \sigma^2 + \frac{t}{r-1} \sum_{j=1}^r \beta_j^2. \quad (2.30)$$

2.2.4 Abordagem espacial na experimentação

Segundo Cressie (1993), a prática de aleatorização é insuficiente para garantir a independência espacial dos erros, justificando-se, assim, o uso de modelos espaciais. Dessa forma, as posições espaciais das unidades experimentais não devem ser ignoradas, uma vez que, em análises espaciais, quanto mais próximas as parcelas estão, mais correlacionados são os dados delas advindos.

Martinez (1994) argumenta que, mesmo que se cumpram os princípios básicos da experimentação, a correlação espacial implica em violações das pressuposições assumidas no modelo da análise não espacial e, conseqüentemente, em uma análise de variância menos precisa, ressaltando, ainda, que a proteção dada pela aleatorização não é suficiente.

Na experimentação agropecuária é comum encontrar, ainda, experimentos nos quais é feito o arranjo sistemático das parcelas, ocasionado pela impossibilidade de se executar o princípio de aleatorização. Na ausência do mesmo, não é possível analisar os experimentos por meio da metodologia clássica, que supõe erros independentes, uma vez que se podem encontrar experimentos cujos modelos especificados incluem explicitamente uma estrutura de dependência espacial, ou seja, de erros correlacionados.

De acordo com Stroup, Baenziger e Moltip (1994), tem-se constatado que as análises convencionais, frequentemente, não neutralizam de forma adequada os efeitos da variabilidade espacial, principalmente em experimentos com número elevado de tratamentos. Felizmente, avanços em estatística para dados espacialmente distribuídos têm fornecido métodos alternativos que consideram e modelam essa variabilidade, podendo ser mais eficientes nesse tipo de situação. Segundo Duarte (2000), esses métodos podem ser aplicados alternativamente ou em complementaridade aos métodos tradicionais. Esse autor entende também que a dependência espacial não deve ser considerada uma inconveniência estatística, mas um verdadeiro benefício que pode informar sobre locais não amostrados, a partir dos dados tomados em posições próximas aos pontos desejados, mostrando

a importância de uma abordagem que leve em conta a dependência espacial.

Nesse mesmo contexto, Cressie (1993) mostra os efeitos da correlação espacial em problemas de estimação, predição e de delineamentos experimentais, acrescentando que a detecção dessa estrutura de correlação e o uso dessa informação na análise estatística garantem estimativas mais eficientes dos contrastes de tratamentos. Por outro lado, sua desconsideração pode impedir que diferenças reais sejam levantadas. Diante dessas preocupações, uma grande quantidade de métodos tem sido proposta, no sentido de combinar as duas abordagens, clássica e espacial.

As possibilidades geradas pela análise geoestatística, desde o início, despertaram o interesse, mas também a sensação de inacessibilidade a profissionais de diversas áreas. A justificativa para tal sensação está ligada ao fato de a Geoestatística ter uma notação própria, sendo, muitas vezes, considerada uma teoria complexa, adicionada ao fato dessa análise exigir um processo computacional intenso e, até poucos anos atrás, ainda escasso. Journel (1987) reporta tal inacessibilidade afirmando que, durante muito tempo, a Geoestatística se apresentou como uma teoria difícil, um pouco esotérica, desenvolvida do outro lado do Atlântico e acessível apenas a pessoas altamente treinadas, ou aquelas afortunadas o bastante para ter acesso a algum programa computacional.

Gleeson (1997 citado por DUARTE, 2000), após demonstrar a necessidade da aplicação dos modelos espaciais na análise estatística de experimentos agrícolas, também reconhece que sem um software apropriado isto se torna realmente bastante difícil. No entanto, nos últimos anos, com o grande avanço dos métodos computacionais, destacando o *software* livre R (R DEVELOPMENT CORE TEAM, 2012), que tem implementado vários pacotes que se destinam à análise espacial, a utilização dessa abordagem surge como uma alternativa aos pesquisadores.

No intuito de difundir a análise geoestatística, Journel (1987) defende ainda que, uma vez que a teoria tenha sido desmistificada, é possível centrar a atenção nos verdadeiros problemas que tratam de sua implementação: como colocar uma ideia simples para funcionar.

Zimmerman e Harville (1991) propuseram uma abordagem na qual é feita a modelagem direta do efeito de parcela ($S(x) + \xi(x)$), de forma que as observações podem ser consideradas coletivamente como uma realização parcial de um campo aleatório. Nessa abordagem, assume-se que os efeitos de parcela distribuem-se de acordo com algum modelo de correlação espacial que descreve as tendências locais, análogos aos modelos de semivariograma usados em Geoestatística. Em síntese, o modelo busca uma estimativa da função geral de covariância, a qual pondera, entre outros resultados, a solução de quadrados mínimos generalizados dos efeitos fixos. Dada a sua concepção, os seus propositores chamaram-no de “modelo linear de campo aleatório” (*random field linear model* - RFLM). Porém, trata-se, essencialmente, de um modelo linear com erros espacialmente correlacionados (STROUP; BAENZIGER; MULITZE, 1994). Sua vantagem comparativa está no fato de aplicar-se a ensaios com dependência espacial em todas ou quaisquer direções, em associação com os diversos esquemas de definição dos blocos (ou nenhum).

Nesse mesmo contexto, Grondona e Cressie (1991) também modelaram diretamente o efeito de dependência espacial entre os erros, baseando suas análises na estimação de quadrados mínimos generalizados. Nesse procedimento, a matriz de covariâncias dos erros, que pondera o sistema de equações normais, é obtida diretamente do modelo de semivariância ajustado, o qual é obtido a partir da análise geoestatística dos resíduos, após a remoção dos efeitos dos fatores controlados no experimento.

Diferentemente de Grondona e Cressie (1991) e Zimmerman e Harville (1991), que examinaram experimentos em blocos casualizados, Hoef e Cressie (1993) analisaram experimentos simulados de acordo com um delineamento inteiramente casualizado. Nesse trabalho, os autores compararam os resultados obtidos da estimação dos efeitos dos tratamentos por meio da ANOVA clássica, cujos erros foram considerados independentes, com os resultados obtidos por meio de uma análise espacial, na qual a matriz de covariâncias foi modelada por meio do semivariograma ajustado. Nesse caso, a estimação do vetor de parâmetros foi feita por meio dos métodos

de quadrados mínimos generalizados, máxima verossimilhança e máxima verossimilhança restrita.

Sendo assim, autores como Grondona e Cressie (1991), Hoef e Cressie (1993) e Zimmerman e Harville (1991) foram unânimes em apontar os benefícios desse tipo de abordagem, uma vez que, ao adotarem uma abordagem espacial, encontraram uma redução significativa da estimativa do erro padrão, ao comparar com os resultados obtidos pela análise clássica, na qual os erros eram supostamente independentes.

Martínez (1994) também chegou a essas conclusões, uma vez que, ao analisar dados simulados com o objetivo de comparar algumas metodologias utilizadas para controlar a correlação espacial, observou que a análise por quadrados mínimos generalizados, utilizando o conceito de variograma, resultou também em menores erros padrões para as diferenças entre as médias dos tratamentos e que, além disso, a utilização dessa abordagem corrigiu significativamente o problema de super e subestimação dessas médias.

Resultados similares foram obtidos também por Reis e Miranda Filho (2003) que, trabalhando com um experimento instalado segundo um delineamento em blocos casualizados, seguindo um modelo misto, procuraram avaliar a presença de correlação espacial, bem como a influência da mesma nas estimativas dos parâmetros e testes de hipóteses relacionados. Segundo esses autores, a adoção da análise espacial resultou na redução das estimativas das variâncias residuais e, conseqüentemente, em melhorias nos ganhos esperados com a seleção dos tratamentos, concluindo que a análise espacial foi mais apropriada que a análise que considerava erros independentes.

Duarte e Vencovsky (2005), avaliando a eficiência da análise estatística espacial na seleção de genótipos de diferentes linhagens de soja, obtiveram resultados análogos aos encontrados por Reis e Miranda Filho (2003), ressaltando, ainda, que a análise espacial resultou em um diferente ordenamento das linhagens em relação à análise não espacial, levando a uma seleção menos influenciada pelos efeitos da variação local.

Rossoni (2011) também efetuou uma comparação da análise clássica (erros independentes) com a análise espacial, utilizando a Geostatística e a ANOVA com modelo espacial autorregressivo SAR. Diante dessa comparação, o autor concluiu que a utilização de ferramentas da estatística espacial tornou a análise mais precisa, uma vez que se observou uma diminuição na variabilidade geral do experimento. Além disso, o modelo espacial proporcionou uma redução nos valores do AIC, ao serem comparados com os valores obtidos pela modelagem clássica, indicando que a consideração da dependência espacial dos erros acarretou em um melhor ajuste.

Por fim, ainda em defesa da análise espacial, Campos (2011) concluiu que a utilização dessa análise proporcionou uma seleção mais eficiente das famílias do feijoeiro, comparadas às análises tradicionais para delineamentos em látice e em blocos casualizados, uma vez que apresentou estimativas de variância do erro muito próximas da variância simulada, além de uma maior acurácia seletiva, indicando maior precisão experimental.

3 MATERIAIS E MÉTODOS

Esta seção destina-se à descrição da metodologia utilizada para a simulação dos dados que foram analisados, a estimação de parâmetros de modelos lineares e a garantia da validade do teste F da análise de variância.

3.1 Simulação de dados

Para o desenvolvimento deste trabalho, os dados analisados foram obtidos via simulação, por meio do software R (R DEVELOPMENT CORE TEAM, 2012).

Foram propostos dois tipos configurações de experimentos para cada delineamento, DIC e DBC, o primeiro com um grande número de parcelas e o segundo com um pequeno número de parcelas, como a seguir:

- (i) Experimento em DIC com 15 tratamentos e 8 repetições;
- (ii) Experimento em DIC com 6 tratamentos e 5 repetições;
- (iii) Experimento em DBC com 15 tratamentos e 8 blocos;
- (iv) Experimento em DBC com 6 tratamentos e 5 blocos.

O objetivo de estabelecer tais configurações foi verificar se, mesmo em experimentos com pequenos números de parcelas, a abordagem geoestatística consegue captar a dependência espacial existente entre os erros.

No processo de simulação, os conjuntos de dados foram gerados inicialmente fixando-se os valores dos parâmetros, considerando-se modelos fixos. Nas Tabelas 3, 4, 5 e 6 são apresentados os valores da constante e efeitos de tratamentos e blocos, de acordo com cada configuração de experimento citada anteriormente.

Tabela 3 Valores fixos dos parâmetros para os experimentos em DIC com 15 tratamentos e 8 repetições.

| Parâmetro | μ | α_1 | α_2 | α_3 | α_4 | α_5 | α_6 | α_7 | α_8 | α_9 | α_{10} | α_{11} |
|------------|-------|------------|------------|------------|------------|------------|------------|------------|------------|------------|---------------|---------------|
| Valor fixo | 10 | 7 | 10 | 5 | 5 | 8 | 10 | 9 | 5 | 10 | 6 | 7 |

| Parâmetro | α_{12} | α_{13} | α_{14} | α_{15} |
|------------|---------------|---------------|---------------|---------------|
| Valor fixo | 7 | 6 | 6 | 9 |

Tabela 4 Valores fixos dos parâmetros para os experimentos em DIC com 6 tratamentos e 5 repetições.

| Parâmetro | μ | α_1 | α_2 | α_3 | α_4 | α_5 | α_6 |
|------------|-------|------------|------------|------------|------------|------------|------------|
| Valor fixo | 10 | 6 | 4,5 | 9 | 6,5 | 8 | 6,5 |

Tabela 5 Valores fixos dos parâmetros para os experimentos em DBC com 15 tratamentos e 8 repetições.

| Parâmetro | μ | α_1 | α_2 | α_3 | α_4 | α_5 | α_6 | α_7 | α_8 | α_9 | α_{10} | α_{11} |
|------------|-------|------------|------------|------------|------------|------------|------------|------------|------------|------------|---------------|---------------|
| Valor fixo | 10 | 4 | 6 | 8 | 9,5 | 5 | 7 | 6 | 10 | 8,5 | 4 | 9 |

| Parâmetro | α_{12} | α_{13} | α_{14} | α_{15} | β_1 | β_2 | β_3 | β_4 | β_5 | β_6 | β_7 | β_8 |
|------------|---------------|---------------|---------------|---------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| Valor fixo | 5,5 | 7 | 6 | 4 | 12 | 8 | 6 | 9 | 9 | 10 | 5 | 7,5 |

Tabela 6 Valores fixos dos parâmetros para os experimentos em DBC com 6 tratamentos e 5 repetições.

| Parâmetro | μ | α_1 | α_2 | α_3 | α_4 | α_5 | α_6 | β_1 | β_2 | β_3 | β_4 | β_5 |
|------------|-------|------------|------------|------------|------------|------------|------------|-----------|-----------|-----------|-----------|-----------|
| Valor fixo | 10 | 5,5 | 7 | 5,5 | 9,5 | 4 | 6 | 12 | 8 | 9,5 | 10 | 6 |

A esses valores fixos foi adicionado um erro aleatório com estrutura de dependência espacial definida, gerado por meio da função *GaussRF* do pacote *Random Fields* (SCHLATHER, 2012). Esses erros foram simulados com média zero e modelos de covariância esférico e exponencial. Com o objetivo de verificar a relação entre a eficiência da análise geoestatística e as configurações de dependência espacial, foram geradas

diferentes combinações de valores dos parâmetros do semivariograma. Essas configurações são apresentadas na Tabela 7. Vale ressaltar que as configurações para os experimentos em DIC e DBC, com o mesmo número de observações, são exatamente as mesmas.

Os valores do efeito pepita e patamar foram gerados de forma que os erros tenham forte e moderado grau de dependência, conforme definido na seção 2.1.2. Já os valores do alcance foram propostos no intuito de avaliar a dependência espacial de curto, médio e longo alcance.

Tabela 7 Valores dos parâmetros do semivariograma utilizados para simular o erro experimental.

| Experimento (<i>i</i>) | | | Experimento (<i>ii</i>) | | |
|--------------------------|---------|---------|---------------------------|---------|---------|
| efeito pepita | patamar | alcance | efeito pepita | patamar | alcance |
| 0,10 | 2,5 | 3 | 0,10 | 2,5 | 2 |
| 0,10 | 2,5 | 6 | 0,10 | 2,5 | 4 |
| 0,10 | 2,5 | 9 | 0,10 | 2,5 | 6 |
| 0,75 | 2,5 | 3 | 0,75 | 2,5 | 2 |
| 0,75 | 2,5 | 6 | 0,75 | 2,5 | 4 |
| 0,75 | 2,5 | 9 | 0,75 | 2,5 | 6 |

Os dados foram simulados de acordo com um *grid* regular, em que a distância entre os centroides das parcelas adjacentes foi definida como 1 *u.m* (unidade de medida). Em todo o processo de simulação, o erro foi considerado uma variável isotrópica, isto é, a dependência entre os erros varia de acordo com a distância que os separa, apresentando o mesmo comportamento espacial em todas as direções. Nas Figuras 1 e 2 estão apresentados os *grids* utilizados na simulação dos experimentos em DIC para as duas configurações propostas, de modo que α_i representa o efeito do tratamento *i* adicionado à unidade experimental correspondente.

Na análise geoestatística, a aleatorização não é um princípio obrigatório e, por isso, foi realizado um arranjo sistemático dos tratamentos nas unidades experimentais, sem utilização de qualquer critério de aleatorização.

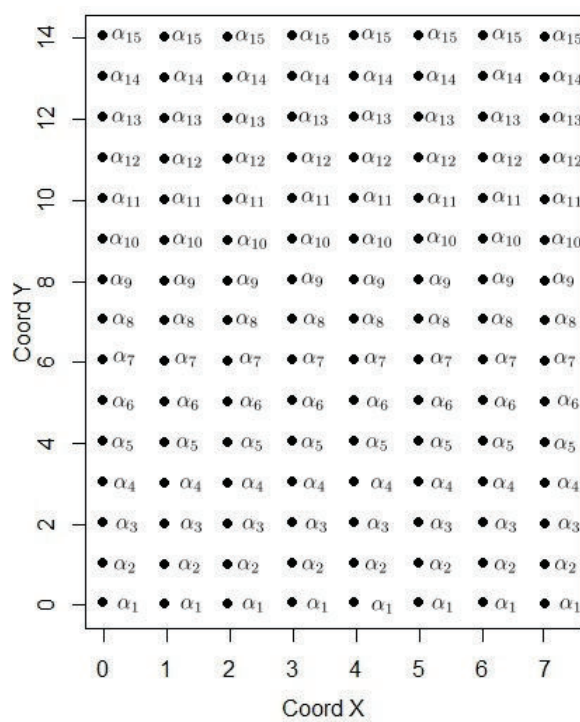


Figura 1 *Grid* de um experimento em DIC, com 15 tratamentos e 8 repetições

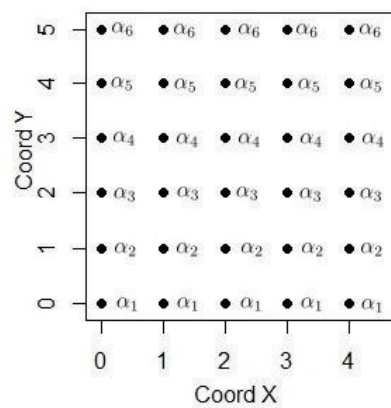


Figura 2 *Grid* de um experimento em DIC, com 6 tratamentos e 5 repetições

Já o *grid* utilizado na simulação dos experimentos em DBC com 15 tratamentos e 8 blocos é apresentado na Figura 3, enquanto na Figura 4 apresenta-se o croqui da simulação dos experimentos com 6 tratamentos e 5 blocos, especificando, além das coordenadas espaciais, os efeitos de tratamentos e blocos correspondentes a cada unidade experimental.

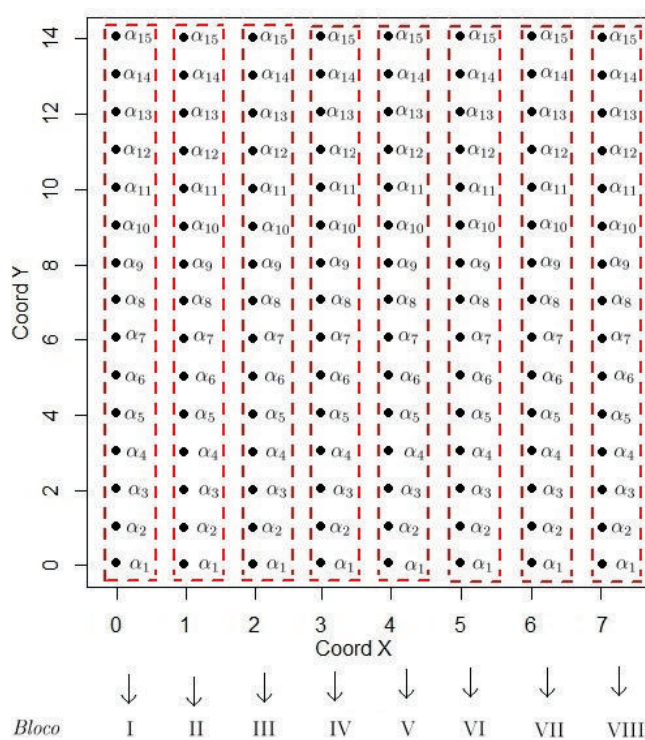


Figura 3 *Grid* de um experimento em DBC, com 15 tratamentos e 8 blocos

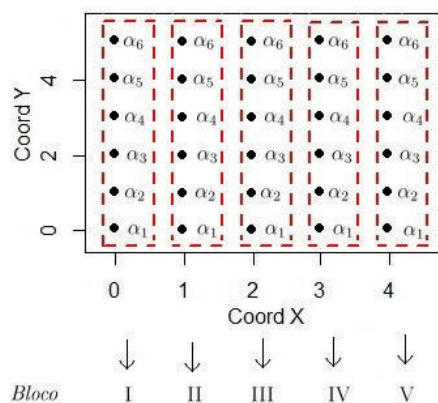


Figura 4 *Grid* de um experimento em DBC, com 6 tratamentos e 5 blocos

3.2 Estimação de parâmetros

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

$$Y = \mathbf{X}\theta + \varepsilon, \quad (3.1)$$

em que

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

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

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

ε é o vetor de erros aleatórios de ordem $n \times 1$, sendo que, frequentemente, assume-se que $\varepsilon \sim N(\mathbf{0}, \mathbf{I}\sigma^2)$.

Neste trabalho, os estimadores dos parâmetros do modelo considerando-se erros espacialmente dependentes foram obtidos pelos métodos de quadrados mínimos generalizados e máxima verossimilhança.

Entretanto, para a obtenção das somas de quadrados da análise de variância, por meio da redução de modelos, foi necessário obter estimadores dos parâmetros de acordo com os modelos particionados.

Segundo Searle (1987), convenientemente, um modelo linear pode

ser escrito de forma particionada. O modelo particionado para um delineamento inteiramente casualizado é dado por

$$Y = \mathbf{X}_1\mu + \mathbf{X}_2\alpha + \varepsilon, \quad (3.2)$$

em que Y é o vetor de observações, μ é uma constante inerente a cada observação, α é um vetor $t \times 1$ que contém os efeitos de tratamento, \mathbf{X}_1 é um vetor de uns, \mathbf{X}_2 é a matriz de incidência dos efeitos de tratamento de ordem $n \times t$, de tal modo que $\mathbf{X} = [\mathbf{X}_1 : \mathbf{X}_2]$ e $\theta' = [\mu : \alpha']$ e, por fim, ε representa um vetor de erros.

Desse modo, o modelo de um DIC com $t = 3$ tratamentos e $r = 2$ repetições pode ser representado, matricialmente, por

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \\ y_{31} \\ y_{32} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{12} \\ e_{21} \\ e_{22} \\ e_{31} \\ e_{32} \end{bmatrix};$$

ou, equivalentemente, de forma particionada,

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \\ y_{31} \\ y_{32} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \mu + \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{12} \\ e_{21} \\ e_{22} \\ e_{31} \\ e_{32} \end{bmatrix}.$$

Em um processo análogo, o modelo para um experimento em blocos casualizados pode ser particionado da seguinte forma:

$$Y = \mathbf{X}_1\mu + \mathbf{X}_2\alpha + \mathbf{X}_3\beta + \varepsilon, \quad (3.3)$$

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

Como forma de ilustração, o modelo de um DBC com $t = 3$ tratamentos e $r = 2$ blocos é dado por

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \\ y_{31} \\ y_{32} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{12} \\ e_{21} \\ e_{22} \\ e_{31} \\ e_{32} \end{bmatrix};$$

ou, por meio de sua forma particionada,

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \\ y_{31} \\ y_{32} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \mu + \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{12} \\ e_{21} \\ e_{22} \\ e_{31} \\ e_{32} \end{bmatrix}.$$

3.3 Modelagem geoestatística do erro

Os dados experimentais, obtidos via simulação, foram submetidos às análises estatísticas relacionadas a dois modelos: o primeiro assumindo

que os erros são espacialmente independentes e o segundo considerando a estrutura de dependência espacial entre os erros. Dessa forma, as seguintes estruturas da matriz de covariâncias dos erros foram consideradas:

- i. $\Sigma = \mathbf{I}\sigma^2$ (modelo com erros independentes);
- ii. $\Sigma = \mathbf{I}\tau^2 + \mathbf{F}\varphi^2 = \mathbf{V}\sigma^2$ (modelo com erros espacialmente dependentes),

de modo que $\sigma^2 = \tau^2 + \varphi^2$ corresponde ao patamar do semivariograma.

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

A partir da estimação de $\hat{\varepsilon}$, é possível verificar se existe ou não dependência entre os erros. Essa verificação foi realizada por meio da própria análise do semivariograma. Segundo Journel (1989), se os valores de semivariância distribuírem-se de forma aleatória em função das distâncias, significa que os erros são independentes. No entanto, se o semivariograma apresentar uma estrutura de dependência espacial, na qual os valores de semivariância aumentam com o aumento dos valores de h até uma determinada distância (ϕ), pode-se concluir que os erros possuem dependência espacial.

Como método auxiliar na análise do semivariograma, pode-se, ainda, construir envelopes, os quais são obtidos por meio da permutação dos valores observados nas posições espaciais. Se o envelope contiver todos os pontos do semivariograma, significa que os valores de semivariância distribuem-se de forma aleatória e, com isso, pode-se concluir que os erros são independentes. Entretanto, se algum valor de semivariância não estiver dentro do envelope, considera-se que o semivariograma apresenta uma estrutura de dependência espacial definida e, por isso, os erros são considerados dependentes espacialmente.

Uma vez detectada a estrutura de dependência espacial, a próxima etapa da modelagem da variabilidade espacial do erro consiste em ajustar

um modelo teórico ao semivariograma, sendo que, neste trabalho, foram ajustados apenas o modelo esférico e o exponencial, de acordo com o modelo utilizado no processo de simulação. Como o erro é uma variável isotrópica, assim definido na simulação, construiu-se o semivariograma *omnidirecional*, de forma que, para o cálculo de suas semivariâncias, foram consideradas apenas as distâncias, em qualquer direção, entre dois pontos amostrados. O método utilizado para ajustar os modelos teóricos ao semivariograma foi o de quadrados mínimos ordinários.

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

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

e, conforme apresentado na seção (2.1.4), assumindo que o erro é uma variável estacionária de segunda ordem, essa matriz Σ corresponde a

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

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

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

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

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

Os modelos teóricos de correlação esférico e exponencial são dados por

Modelo esférico:

$$\rho(h) = \begin{cases} 1 & , \text{ se } h = 0; \\ \frac{\varphi^2}{\tau^2 + \varphi^2} \left[1 - 1,5 \left(\frac{h}{\phi} \right) + 0,5 \left(\frac{h}{\phi} \right)^3 \right] & , \text{ se } 0 < h < \phi; \\ 0 & , \text{ se } h \geq \phi. \end{cases} \quad (3.4)$$

Modelo exponencial:

$$\rho(h) = \begin{cases} 1 & , \text{ se } h = 0; \\ \frac{\varphi^2}{\tau^2 + \varphi^2} \left[\exp \left(\frac{-3h}{\phi} \right) \right] & , \text{ se } h \neq 0. \end{cases} \quad (3.5)$$

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

$$\theta^0 = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}Y. \quad (3.6)$$

A fim de encontrar uma maior eficiência dos resultados, Pontes (2002) sugeriu um algoritmo no qual eram realizadas iterações até que a média dos erros padrão dos valores preditos atingisse convergência.

No processo iterativo desse algoritmo, a partir do vetor θ^0 , obtém-se uma nova estimativa para o vetor de erros, de modo que $\hat{\varepsilon}^* = Y - \mathbf{X}\theta^0$. No intuito de encontrar uma maior eficiência dos resultados, um novo semivariograma deve ser construído, com base na variável aleatória $\hat{\varepsilon}^*$. A esse semivariograma, novamente, ajusta-se um modelo teórico e obtêm-se novas estimativas para os seus parâmetros. Com essas estimativas é possível construir novamente a matriz de covariância dos erros Σ , a qual é utilizada para obter um novo vetor de parâmetros θ^0 . Pontes (2002) assumiu como critério de estabilização a convergência dos valores médios dos erros padrão dos valores preditos.

Neste trabalho, esse processo foi repetido até que as estimativas dos parâmetros do semivariograma se estabilizassem e, conseqüentemente, ocorresse a estabilização de todo o processo, sendo considerada uma tolerância de 10^{-4} . Os parâmetros encontrados na última iteração foram, então, utilizados para calcular a matriz de covariância dos erros (Σ), a qual assumiu um papel fundamental na obtenção das somas de quadrados da análise de variância.

É importante ressaltar que, durante o processo de modelagem da variabilidade espacial do erros, todas as análises geoestatísticas foram realizadas com o auxílio do pacote geoR (RIBEIRO JÚNIOR; DIGGLE, 2001), implementado no *software* R (R DEVELOPMENT CORE TEAM, 2012).

3.4 O teste F da análise de variância

Na teoria de probabilidade e estatística, a distribuição F de Snedecor é uma distribuição de probabilidade contínua que, frequentemente, é utilizada como a distribuição nula de uma estatística de teste que, em especial, é empregada na análise de variância.

Um dos objetivos deste trabalho foi demonstrar a validade do teste F da análise de variância que considera a informação espacial sobre os erros.

Para isso, foi necessário mostrar que a estatística F_0 deste teste vem de uma distribuição F de Snedecor.

Segundo Mood, Graybill e Boes (1974), uma variável aleatória com distribuição F é dada pela razão entre duas variáveis aleatórias com distribuição qui-quadrado (χ^2), divididas por seus respectivos graus de liberdade. Desse modo, considerando U uma variável aleatória que segue uma distribuição qui-quadrado com m graus de liberdade e V outra variável aleatória que tem distribuição qui-quadrado com n graus de liberdade, de tal forma que U e V sejam independentes, é possível afirmar que a variável aleatória

$$W = \frac{U/m}{V/n} \quad (3.7)$$

tem distribuição F central com m e n graus de liberdade.

Porém, conforme afirma Graybill (1976), se U é uma variável aleatória que tem distribuição qui-quadrado não central com m graus de liberdade e parâmetro de não centralidade δ e V é uma outra variável aleatória que tem distribuição qui-quadrado central com n graus de liberdade, ou seja, $U \sim \chi^2_{(m, \delta)}$ e $V \sim \chi^2_{(n)}$, de forma que U e V são independentes, então, a variável aleatória W , definida na equação (3.7), tem uma distribuição F não central com m e n graus de liberdade e parâmetro de não centralidade δ .

Na análise de variância, se os erros forem independentes e normalmente distribuídos com uma variância comum, a estatística F_0 , obtida pela razão de quadrados médios, segue uma distribuição F .

Por isso, para mostrar que o teste F continua sendo válido quando as somas de quadrados da análise de variância são ponderadas por uma matriz \mathbf{V} , em que \mathbf{V} modela a correlação entre os erros, deve-se mostrar que

$$\frac{SQ_{trat}}{\sigma^2} \sim \chi^2_{(r, \delta)} \quad \text{e} \quad \frac{SQ_{Res}}{\sigma^2} \sim \chi^2_{(k)}, \quad (3.8)$$

em que SQ_{trat} e SQ_{Res} representam, respectivamente, as soma de

quadrados de tratamentos e de resíduos. Além disso, σ^2 refere-se à variância do erro, r e k representam, respectivamente, os graus de liberdade do tratamento e do resíduo e δ é um parâmetro de não centralidade.

Dessa forma, se $SQtrat/\sigma^2$ e $SQRes/\sigma^2$ são independentes, então

$$F_0 = \frac{QMtrat}{QMRes} = \frac{(SQtrat/\sigma^2)/r}{(SQRes/\sigma^2)/k} \sim F(r, k; \delta), \quad (3.9)$$

sendo que $QMtrat$ e $QMRes$ referem-se, respectivamente, ao quadrado médio de tratamento e ao quadrado médio do resíduo.

Em um processo análogo, é possível mostrar a aplicabilidade do teste F na avaliação da igualdade entre as médias dos blocos.

4 RESULTADOS E DISCUSSÕES

Esta seção aborda, inicialmente, a construção e a utilização da análise de variância para experimentos cujos erros são espacialmente correlacionados. Posteriormente, foram realizadas análises referentes às simulações formuladas sob diferentes configurações espaciais. Por fim, um exemplo de aplicação foi apresentado, para uma melhor compreensão de como obter a análise de variância pela abordagem espacial.

4.1 Estimação de parâmetros considerando $\Sigma = V\sigma^2$

Um modelo linear com erros espacialmente correlacionados pode ser representado da seguinte forma:

$$Y = \mathbf{X}\theta + \varepsilon, \quad \text{com } \varepsilon \sim N(\mathbf{0}, \Sigma), \quad (4.1)$$

em que $\Sigma = V\sigma^2$, de modo que V é a matriz de correlação dos erros apresentada na seção (3.3).

Um dos métodos utilizados para a estimação de parâmetros é o método de máxima verossimilhança. De acordo com o modelo apresentado na expressão (4.1), para estimar o vetor de parâmetros θ por meio deste método é preciso assumir que Y segue alguma distribuição, sendo que, geralmente, assume-se que Y tem distribuição normal multivariada. Segundo Rencher e Schaalje (2008), ao supor que $Y \sim N(\mathbf{X}\theta, \Sigma)$, a função de verossimilhança é dada pela densidade de probabilidade de Y , ou seja,

$$l(\theta, \Sigma) = f(Y) = (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(Y - \mathbf{X}\theta)' \Sigma^{-1} (Y - \mathbf{X}\theta)\right], \quad (4.2)$$

em que n corresponde ao número de observações de Y .

A maximização da função de verossimilhança equivale à maximização do logaritmo neperiano dessa função. Além disso, segundo Rencher et al. (2008), se c é uma constante e \mathbf{A} é uma matriz quadrada de ordem n , então $|c\mathbf{A}| = c^n |\mathbf{A}|$. Desse modo, considerando-se que $\mathbf{\Sigma} = \mathbf{V}\sigma^2$, em que \mathbf{V} é uma matriz simétrica positiva definida conhecida e que o logaritmo neperiano da função de verossimilhança é dado por

$$\ln l(\theta, \sigma^2) = -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln(|\mathbf{V}|) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} (Y - \mathbf{X}\theta)' \mathbf{V}^{-1} (Y - \mathbf{X}\theta),$$

tem-se que os estimadores de máxima verossimilhança para θ e σ^2 são dados por

$$\frac{\partial \ln l(\theta, \sigma^2)}{\partial \theta} = -2\mathbf{X}'\mathbf{V}^{-1}Y + 2\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\theta^0 = 0,$$

$$\theta^0 = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}Y. \quad (4.3)$$

$$\frac{\partial \ln l(\theta, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2\hat{\sigma}^2} + \frac{1}{2(\hat{\sigma}^2)^2} (Y - \mathbf{X}\theta^0)' \mathbf{V}^{-1} (Y - \mathbf{X}\theta^0) = 0,$$

$$\hat{\sigma}^2 = \frac{(Y - \mathbf{X}\theta^0)' \mathbf{V}^{-1} (Y - \mathbf{X}\theta^0)}{n}. \quad (4.4)$$

Segundo Graybill (1976), um outro método para se obter um estimador para θ é através do método de quadrados mínimos generalizados. A partir desse método, é necessário considerar que \mathbf{L} é uma matriz quadrada de ordem n não singular, de modo que $\mathbf{V}^{-1} = \mathbf{L}\mathbf{L}'$. Pré-multiplicando o modelo (4.1) por \mathbf{L}' e definindo $W = \mathbf{L}'Y$, $\mathbf{Z} = \mathbf{L}'\mathbf{X}$ e $\eta = \mathbf{L}'\varepsilon$ tem-se um novo modelo que é dado por

$$W = \mathbf{Z}\theta + \eta; \quad (4.5)$$

de modo que

$$\begin{aligned}
E[\eta] &= E[\mathbf{L}'\varepsilon] = \mathbf{L}'E[\varepsilon] = \mathbf{0}, \\
Var[\eta] &= Var[\mathbf{L}'\varepsilon] = \mathbf{L}'Var[\varepsilon]\mathbf{L} = (\mathbf{L}^{-1}\mathbf{L})\mathbf{L}'\mathbf{V}\mathbf{L}\sigma^2 = \mathbf{L}^{-1}(\mathbf{L}\mathbf{L}')\mathbf{V}\mathbf{L}\sigma^2, \\
Var[\eta] &= \mathbf{I}\sigma^2.
\end{aligned}$$

Com isso, tem-se que $\eta \sim N(\mathbf{0}, \mathbf{I}\sigma^2)$ e, desse modo, o vetor de parâmetros θ pode ser estimado por

$$\theta^0 = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'W, \quad (4.6)$$

que, substituindo $W = \mathbf{L}'Y$ e $\mathbf{Z} = \mathbf{L}'\mathbf{X}$, equivale a

$$\theta^0 = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}Y; \quad (4.7)$$

que é exatamente o mesmo estimador encontrado através do método de máxima verossimilhança.

4.2 Delineamento inteiramente casualizado (DIC)

Nesta seção foram apresentados os estimadores de parâmetros, por meio de modelos particionados, para um delineamento inteiramente casualizado com erros espacialmente correlacionados, bem como sua análise de variância com as respectivas esperanças de seus quadrados médios e, por fim, foram ainda apresentadas as distribuições de probabilidade associadas a esses quadrados médios.

4.2.1 Estimação de parâmetros via modelos particionados

Como já foi dito anteriormente, para a obtenção da análise de variância por meio de formas quadráticas, é necessária a estimação dos parâmetros por meio de modelos particionados. Devido às suas propriedades, foi utilizada a inversa generalizada de Moore-Penrose.

Entretanto, antes de definir esses estimadores, é necessário apresentar algumas propriedades dessa inversa generalizada, a qual é denotada pelo expoente (+):

$$[P.1] : \mathbf{A}\mathbf{A}^+\mathbf{A} = \mathbf{A};$$

$$[P.2] : \mathbf{A}^+\mathbf{A}\mathbf{A}^+ = \mathbf{A}^+;$$

$$[P.3] : (\mathbf{A}\mathbf{A}^+)' = \mathbf{A}\mathbf{A}^+;$$

$$[P.4] : \mathbf{A}\mathbf{A}^+ = \mathbf{A}(\mathbf{A}'\mathbf{A})^+\mathbf{A}';$$

$$[P.5] : (\mathbf{A}')^+ = (\mathbf{A}^+)';$$

Conforme visto na seção (3.2), convenientemente, o modelo da equação (4.1) pode ser particionado de forma que $\mathbf{X} = [\mathbf{X}_1 : \mathbf{X}_2]$ e $\theta' = [\mu : \alpha']$. Com isso, o modelo transformado da expressão (4.5) terá a seguinte partição:

$$W = \mathbf{Z}_1\mu + \mathbf{Z}_2\alpha + \eta, \quad (4.8)$$

com $\mathbf{Z} = [\mathbf{Z}_1 : \mathbf{Z}_2]$, de forma que $\mathbf{Z}_1 = \mathbf{L}'\mathbf{X}_1$ e $\mathbf{Z}_2 = \mathbf{L}'\mathbf{X}_2$.

Os estimadores dos parâmetros μ e α foram obtidos a partir do sistema de equações normais (S.E.N) do modelo (4.5):

$$(\mathbf{Z}'\mathbf{Z})\theta^0 = \mathbf{Z}'W,$$

$$\begin{bmatrix} \mathbf{Z}'_1 \\ \mathbf{Z}'_2 \end{bmatrix} \begin{bmatrix} \mathbf{Z}_1 & \mathbf{Z}_2 \end{bmatrix} \begin{bmatrix} \mu^0 \\ \alpha^0 \end{bmatrix} = \begin{bmatrix} \mathbf{Z}'_1 \\ \mathbf{Z}'_2 \end{bmatrix} W,$$

$$\begin{bmatrix} \mathbf{Z}'_1\mathbf{Z}_1 & \mathbf{Z}'_1\mathbf{Z}_2 \\ \mathbf{Z}'_2\mathbf{Z}_1 & \mathbf{Z}'_2\mathbf{Z}_2 \end{bmatrix} \begin{bmatrix} \mu^0 \\ \alpha^0 \end{bmatrix} = \begin{bmatrix} \mathbf{Z}'_1W \\ \mathbf{Z}'_2W \end{bmatrix},$$

$$\begin{cases} \mathbf{Z}'_1\mathbf{Z}_1\mu^0 + \mathbf{Z}'_1\mathbf{Z}_2\alpha^0 = \mathbf{Z}'_1W, & \text{(I)} \\ \mathbf{Z}'_2\mathbf{Z}_1\mu^0 + \mathbf{Z}'_2\mathbf{Z}_2\alpha^0 = \mathbf{Z}'_2W. & \text{(II)} \end{cases}$$

Prémultiplicando-se a equação (I) por $\mathbf{Z}'_2\mathbf{Z}'_1^{+'}$ e lembrando que,

dadas duas matrizes \mathbf{A} e \mathbf{B} , $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$, tem-se que

$$\begin{aligned}\mathbf{Z}'_2\mathbf{Z}'_1\mathbf{Z}'_1\mathbf{Z}_1\mu^0 + \mathbf{Z}'_2\mathbf{Z}'_1\mathbf{Z}'_1\mathbf{Z}_2\alpha^0 &= \mathbf{Z}'_2\mathbf{Z}'_1\mathbf{Z}'_1W; \\ \mathbf{Z}'_2\mathbf{Z}_1\mathbf{Z}_1^+\mathbf{Z}_1\mu^0 + \mathbf{Z}'_2\mathbf{Z}_1\mathbf{Z}_1^+\mathbf{Z}_2\alpha^0 &= \mathbf{Z}'_2\mathbf{Z}_1\mathbf{Z}_1^+W; \\ \mathbf{Z}'_2\mathbf{Z}_1\mu^0 + \mathbf{Z}'_2\mathbf{Z}_1\mathbf{Z}_1^+\mathbf{Z}_2\alpha^0 &= \mathbf{Z}'_2\mathbf{Z}_1\mathbf{Z}_1^+W.\end{aligned}\quad (\text{III})$$

Subtraindo-se a equação (III) de (II) encontra-se um estimador para α que é dado por

$$\begin{cases} \mathbf{Z}'_2\mathbf{Z}_1\mu^0 + \mathbf{Z}'_2\mathbf{Z}_2\alpha^0 = \mathbf{Z}'_2W, \\ -\mathbf{Z}'_2\mathbf{Z}_1\mu^0 - \mathbf{Z}'_2\mathbf{Z}_1\mathbf{Z}_1^+\mathbf{Z}_2\alpha^0 = -\mathbf{Z}'_2\mathbf{Z}_1\mathbf{Z}_1^+W, \end{cases}$$

$$\mathbf{Z}'_2(\mathbf{I} - \mathbf{Z}_1\mathbf{Z}_1^+)\mathbf{Z}_2\alpha^0 = \mathbf{Z}'_2(\mathbf{I} - \mathbf{Z}_1\mathbf{Z}_1^+)W;$$

$$\alpha^0 = (\mathbf{Z}'_2\mathbf{R}\mathbf{Z}_2)^{-1}\mathbf{Z}'_2\mathbf{R}W,$$

com $\mathbf{R} = \mathbf{I} - \mathbf{Z}_1(\mathbf{Z}'_1\mathbf{Z}_1)^{-1}\mathbf{Z}'_1$.

Para encontrar um estimador para μ , basta substituir α^0 na equação (I),

$$\mu^0 = (\mathbf{Z}'_1\mathbf{Z}_1)^{-1}\mathbf{Z}'_1W - (\mathbf{Z}'_1\mathbf{Z}_1)^{-1}\mathbf{Z}'_1\mathbf{Z}_2\alpha^0;$$

$$\mu^0 = (\mathbf{Z}'_1\mathbf{Z}_1)^{-1}\mathbf{Z}'_1\mathbf{Q}W,$$

em que $\mathbf{Q} = \mathbf{I} - \mathbf{Z}_2(\mathbf{Z}'_2\mathbf{R}\mathbf{Z}_2)^{-1}\mathbf{Z}'_2\mathbf{R}$.

Por fim, substituindo-se $\mathbf{Z}_1 = \mathbf{L}'\mathbf{X}_1$, $\mathbf{Z}_2 = \mathbf{L}'\mathbf{X}_2$, $W = \mathbf{L}'Y$ e $\mathbf{V}^{-1} = \mathbf{L}\mathbf{L}'$, encontra-se que os estimadores para μ e α referentes ao modelo particionado, cujos erros são espacialmente correlacionados, são dados por

$$\alpha^0 = (\mathbf{X}'_2\mathbf{R}_v\mathbf{X}_2)^{-1}\mathbf{X}'_2\mathbf{R}_vY, \quad (4.9)$$

$$\mu^0 = (\mathbf{X}'_1\mathbf{V}^{-1}\mathbf{X}_1)^{-1}\mathbf{X}'_1\mathbf{Q}_vY, \quad (4.10)$$

de modo que $\mathbf{R}_v = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}_1(\mathbf{X}'_1\mathbf{V}^{-1}\mathbf{X}_1)^{-1}\mathbf{X}'_1\mathbf{V}^{-1}$ e

$$\mathbf{Q}_v = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}_2(\mathbf{X}'_2\mathbf{R}_v\mathbf{X}_2)^{-1}\mathbf{X}'_2\mathbf{R}_v.$$

4.2.2 Análise de variância

De acordo com o modelo transformado $W = \mathbf{Z}\theta + \eta$, de modo que $\eta \sim N(\mathbf{0}, \mathbf{I}\sigma^2)$, a soma de quadrados do erro é dada por $\eta'\eta = (W - \mathbf{Z}\theta)'(W - \mathbf{Z}\theta)$. Substituindo-se $W = \mathbf{L}'Y$, $\mathbf{Z} = \mathbf{L}'\mathbf{X}$, $\eta = \mathbf{L}'\varepsilon$ e $\mathbf{V}^{-1} = \mathbf{L}\mathbf{L}'$, tem-se que:

$$SQErro = \varepsilon'\mathbf{V}^{-1}\varepsilon = (Y - \mathbf{X}\theta)'\mathbf{V}^{-1}(Y - \mathbf{X}\theta),$$

$$SQErro = Y'\mathbf{V}^{-1}Y - 2\theta'\mathbf{X}'\mathbf{V}^{-1}Y + \theta'\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\theta.$$

Do sistema de equações normais tem-se que $\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\theta^0 = \mathbf{X}'\mathbf{V}^{-1}Y$. Desse modo, substituindo-se θ por θ^0 , obtém-se a soma de quadrados de resíduos que é dada por

$$SQRes = Y'\mathbf{V}^{-1}Y - 2\theta^{0'}\mathbf{X}'\mathbf{V}^{-1}Y + \theta^{0'}\mathbf{X}'\mathbf{V}^{-1}Y,$$

$$SQRes = Y'\mathbf{V}^{-1}Y - \theta^{0'}\mathbf{X}'\mathbf{V}^{-1}Y,$$

$$SQRes = SQTotal_{\text{n\~{a}o corrigida}} - SQPar\hat{a}metros.$$

De maneira análoga, é possível obter a soma de quadrados referente à média, comumente conhecida como correção para a média, que é dada por

$$SQM\acute{e}dia = \mu^{0'}\mathbf{Z}'_1W = W'\mathbf{Z}'_1(\mathbf{Z}'_1\mathbf{Z}_1)^{-1}\mathbf{Z}_1W,$$

$$SQM\acute{e}dia = \mu^{0'}\mathbf{X}'_1\mathbf{V}^{-1}Y = Y'\mathbf{V}^{-1}\mathbf{X}'_1(\mathbf{X}'_1\mathbf{V}^{-1}\mathbf{X}_1)^{-1}\mathbf{X}_1\mathbf{V}^{-1}Y.$$

Por fim, para obter a análise de variância para um delineamento inteiramente casualizado com erros espacialmente correlacionados, é necessário definir a soma de quadrados total (SQT) e a soma de quadrados de tratamentos ($SQTrat$), que são dadas por

$$SQTrat = SQPar\hat{a}metros - SQM\acute{e}dia = \theta^{0'}\mathbf{X}'\mathbf{V}^{-1}Y - \mu^{0'}\mathbf{X}'_1\mathbf{V}^{-1}Y;$$

$$SQT = W'W - SQMédia = Y'\mathbf{V}^{-1}Y - \mu^{0'}\mathbf{X}'_1\mathbf{V}^{-1}Y.$$

Na Tabela 8 é apresentado o esquema desta análise de variância, de forma que $r(\mathbf{A})$ denota o posto de uma matriz \mathbf{A} .

Tabela 8 Análise de variância de um DIC, considerando erros correlacionados espacialmente.

| FV | GL | SQ |
|------------|--|--|
| Tratamento | $r(\mathbf{X}) - r(\mathbf{X}_1)$ | $\theta^0\mathbf{X}'\mathbf{V}^{-1}Y - \mu^0\mathbf{X}'_1\mathbf{V}^{-1}Y$ |
| Resíduo | $r(\mathbf{V}^{-1}) - r(\mathbf{X})$ | $Y'\mathbf{V}^{-1}Y - \theta^0\mathbf{X}'\mathbf{V}^{-1}Y$ |
| Total | $r(\mathbf{V}^{-1}) - r(\mathbf{X}_1)$ | $Y'\mathbf{V}^{-1}Y - \mu^0\mathbf{X}'_1\mathbf{V}^{-1}Y$ |

Entretanto, ao substituir os estimadores θ^0 e μ^0 pelas suas expressões matriciais, as somas de quadrados desta análise de variância podem ser definidas por meio de formas quadráticas, conforme apresentado na Tabela 9, sendo $\mathbf{P} = \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}$ e $\mathbf{P}_1 = \mathbf{V}^{-1}\mathbf{X}_1(\mathbf{X}'_1\mathbf{V}^{-1}\mathbf{X}_1)^{-1}\mathbf{X}'_1\mathbf{V}^{-1}$.

Tabela 9 Análise de variância de um DIC, considerando $\Sigma = \mathbf{V}\sigma^2$, por meio de formas quadráticas.

| FV | GL | SQ |
|------------|--|---------------------------------------|
| Tratamento | $r(\mathbf{X}) - r(\mathbf{X}_1)$ | $Y'(\mathbf{P} - \mathbf{P}_1)Y$ |
| Resíduo | $r(\mathbf{V}^{-1}) - r(\mathbf{X})$ | $Y'(\mathbf{V}^{-1} - \mathbf{P})Y$ |
| Total | $r(\mathbf{V}^{-1}) - r(\mathbf{X}_1)$ | $Y'(\mathbf{V}^{-1} - \mathbf{P}_1)Y$ |

Um outro método para se obter a análise de variância é por meio da redução de modelos. Segundo Searle (1987), o modelo definido na equação (4.8) apresenta as seguintes reduções:

$$\begin{aligned} \text{(I)} \quad W &= \mathbf{Z}_1\mu + \eta; \\ \text{(II)} \quad W &= \mathbf{Z}_1\mu + \mathbf{Z}_2\alpha + \eta; \end{aligned}$$

sendo o estimador de μ para o modelo (I) dado por $\mu_I^0 = (\mathbf{Z}'_1\mathbf{Z}_1)^{-1}\mathbf{Z}'_1W$. Conforme visto em (4.1.1), os estimadores para μ e α , considerando o modelo (II), são definidos como

$$\mu_{II}^0 = (\mathbf{Z}'_1\mathbf{Z}_1)^{-1}\mathbf{Z}'_1\mathbf{Q}W \quad \text{e} \quad \alpha_{II}^0 = (\mathbf{Z}'_2\mathbf{R}\mathbf{Z}_2)^{-1}\mathbf{Z}'_2\mathbf{R}W,$$

em que $\mathbf{Q} = \mathbf{I} - \mathbf{Z}_2(\mathbf{Z}'_2\mathbf{R}\mathbf{Z}_2)^{-1}\mathbf{Z}'_2\mathbf{R}$ e $\mathbf{R} = \mathbf{I} - \mathbf{Z}_1(\mathbf{Z}'_1\mathbf{Z}_1)^{-1}\mathbf{Z}'_1$.

De acordo com Searle (1987), as somas de quadrados referentes aos modelos reduzidos são dadas por

$$\begin{aligned} SQ(\mu) &= SQParâmetros_I = \mu_I^{0'}\mathbf{Z}'_1W = W'\mathbf{Z}_1(\mathbf{Z}'_1\mathbf{Z}_1)^{-1}\mathbf{Z}'_1W, \\ SQ(\mu, \alpha) &= SQParâmetros_{II} = \mu_{II}^{0'}\mathbf{Z}'_1W + \alpha_{II}^{0'}\mathbf{Z}'_2W, \\ SQ(\mu, \alpha) &= W'\mathbf{Q}'\mathbf{Z}_1(\mathbf{Z}'_1\mathbf{Z}_1)^{-1}\mathbf{Z}'_1W + W'\mathbf{R}\mathbf{Z}_2(\mathbf{Z}'_2\mathbf{R}\mathbf{Z}_2)^{-1}\mathbf{Z}'_2W, \\ SQ(\mu, \alpha) &= W'\mathbf{Z}_1(\mathbf{Z}'_1\mathbf{Z}_1)^{-1}\mathbf{Z}'_1W + W'\mathbf{R}\mathbf{Z}_2(\mathbf{Z}'_2\mathbf{R}\mathbf{Z}_2)^{-1}\mathbf{Z}'_2\mathbf{R}W \\ SQ(\alpha|\mu) &= SQ(\mu, \alpha) - SQ(\mu) = W'\mathbf{R}\mathbf{Z}_2(\mathbf{Z}'_2\mathbf{R}\mathbf{Z}_2)^{-1}\mathbf{Z}'_2\mathbf{R}W. \end{aligned}$$

Segundo este autor, a análise de variância para um delineamento inteiramente casualizado, por meio das somas de quadrados obtidas por modelos reduzidos, é dada conforme a Tabela 10.

Tabela 10 Somas de quadrados da análise de variância de um DIC, obtidas por meio de modelos reduzidos.

| FV | GL | SQ |
|------------|---------------------------------------|--|
| Média | $r(\mathbf{X}_1)$ | $SQ(\mu)$ |
| Tratamento | $r(\mathbf{X}) - r(\mathbf{X}_1)$ | $SQ(\alpha \mu) = SQ(\mu, \alpha) - SQ(\mu)$ |
| Resíduo | $r(\mathbf{I}_{(n)}) - r(\mathbf{X})$ | $W'W - SQ(\mu, \alpha)$ |
| Total | $r(\mathbf{I}_{(n)})$ | $W'W$ |

Ao substituir $\mathbf{Z}_1 = \mathbf{L}'\mathbf{X}_1$, $\mathbf{Z}_2 = \mathbf{L}'\mathbf{X}_2$, $W = \mathbf{L}'Y$ e $\mathbf{V}^{-1} = \mathbf{L}\mathbf{L}'$, as somas de quadrados desta análise de variância são dadas por

$$SQT_{corrigida} = Y'\mathbf{V}^{-1}Y - Y'\mathbf{V}^{-1}\mathbf{X}_1(\mathbf{X}_1'\mathbf{V}^{-1}\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{V}^{-1}Y = Y'\mathbf{R}_v Y;$$

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

$$SQRes = SQT_{corrigida} - SQTrat = Y'[\mathbf{R}_v - \mathbf{R}_v\mathbf{X}_2(\mathbf{X}_2'\mathbf{R}_v\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{R}_v]Y;$$

em que $\mathbf{R}_v = \mathbf{V}^{-1} - \mathbf{P}_1 = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}_1(\mathbf{X}_1'\mathbf{V}^{-1}\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{V}^{-1}$.

4.2.3 Esperanças dos quadrados médios

Conforme visto na seção (2.2.2), o modelo estatístico de um delineamento inteiramente casualizado é representado por

$$y_{ij} = \mu + \alpha_i + e_{ij}, \quad \text{com } i = 1, 2, \dots, t \text{ e } j = 1, 2, \dots, r. \quad (4.11)$$

De acordo com este modelo, os somas de quadrados de tratamento e resíduos são dadas da seguinte maneira:

$$SQTrat = \frac{1}{r} \sum_{i=1}^t y_{i.}^2 - \frac{y_{..}^2}{tr} \quad \text{e} \quad SQRes = \sum_{i=1}^t \sum_{j=1}^r y_{ij}^2 - \frac{1}{r} \sum_{i=1}^t y_{i.}^2;$$

cujos respectivos quadrados médios são

$$QMTrat = \frac{SQTrat}{t-1} \quad \text{e} \quad QMRes = \frac{SQRes}{t(r-1)}.$$

Considerando que $E[e_{ij}] = 0$, $Var[e_{ij}] = \sigma^2$, $Cov[e_{ij}, e_{i'j'}] = E[e_{ij}e_{i'j'}] \neq 0$ para algum $ij \neq i'j'$ e definindo

$$T = \sum_j \sum_{i'>i} 2Cov(e_{ij}, e_{i'j}), \quad U = \sum_{i'>i} \sum_{j' \neq j} 2Cov(e_{ij}, e_{i'j'}) \quad \text{e}$$

$$V = \sum_i \sum_{j' > j} 2Cov(e_{ij}, e_{ij'}),$$

as esperanças desses quadrados médios podem ser estabelecidas conforme a Tabela 11.

De acordo com essa tabela, mesmo que a hipótese nula (H_0) de que $\alpha_i = 0$ seja verdadeira, $E[QMTrat] \neq E[QMRes]$ e, com isso, o teste F da análise de variância não é adequado para comparar as variâncias. Por este motivo, se os erros forem correlacionados, para que $E[QMTrat] = E[QMRes]$, sob H_0 , é necessário que essa correlação seja considerada como uma ponderadora das somas de quadrados.

Tabela 11 Esperança dos quadrados médios de um DIC com erros correlacionados, a partir das somas de quadrados clássicas.

| FV | E[QM] |
|------------|--|
| Tratamento | $\sigma^2 + \frac{r}{t-1} \sum_i \alpha_i^2 + \frac{1}{tr} V - \frac{1}{tr(r-1)}(T+U)$ |
| Resíduo | $\sigma^2 - \frac{1}{tr(r-1)} V$ |

Na análise de variância apresentada na Tabela 9 são definidas as somas de quadrados ponderadas pela matriz de correlação \mathbf{V} , representadas por meio de formas quadráticas. Segundo Searle (1971), se Y é um vetor aleatório com média $\mathbf{X}\theta$ e variância Σ e \mathbf{A} é uma matriz simétrica de constantes, então,

$$E[Y'\mathbf{A}Y] = tr(\mathbf{A}\Sigma) + \theta'\mathbf{X}'\mathbf{A}\mathbf{X}\theta. \quad (4.12)$$

Desse modo, a esperança da soma de quadrados de resíduos é dada por

$$E[SQRes] = E[Y'\mathbf{V}^{-1}Y] - E[Y'\mathbf{P}Y],$$

de modo que $\mathbf{P} = \mathbf{V}^{-1}\mathbf{X}'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}\mathbf{V}^{-1}$.

Entretanto, para os cálculos das somas de quadrados, considere-se que $(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}$ é uma inversa generalizada de Moore-Penrose, isto é, $(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{+}$. Além disso, é necessário apresentar algumas propriedades sobre traço e posto (*rank*) de matrizes:

$$[P.1] : tr(\mathbf{A}\mathbf{B}) = tr(\mathbf{B}\mathbf{A});$$

$$[P.2] : tr(\mathbf{A} \pm \mathbf{B}) = tr(\mathbf{A}) \pm tr(\mathbf{B});$$

$$[P.3] : tr(\mathbf{A}) = r(\mathbf{A}), \text{ se } \mathbf{A} \text{ é idempotente};$$

$$[P.4] : r(\mathbf{A}'\mathbf{A}) = r(\mathbf{A}\mathbf{A}') = r(\mathbf{A});$$

$$[P.5] : r(\mathbf{A}^{+}\mathbf{A}) = r(\mathbf{A}\mathbf{A}^{+}) = r(\mathbf{A})$$

$$[P.6] : \text{Se } \mathbf{B} \text{ é não singular de ordem } m \text{ e } \mathbf{A} \text{ é uma matriz } m \times n, \text{ então}$$

$$r(\mathbf{B}\mathbf{A}) = r(\mathbf{A}).$$

De acordo com essas propriedades, tem-se que

$$tr(\mathbf{P}\mathbf{V}) = tr(\mathbf{V}^{-1}\mathbf{X}'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{+}\mathbf{X}) = tr(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{+}),$$

$$tr(\mathbf{P}\mathbf{V}) = r(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{+}) = r(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}) = r((\mathbf{L}'\mathbf{X})'(\mathbf{L}'\mathbf{X})),$$

$$tr(\mathbf{P}\mathbf{V}) = r(\mathbf{L}'\mathbf{X}) = r(\mathbf{X}).$$

Com isso, a $E[SQRes]$ e, conseqüentemente, $E[QMRes]$ são dadas por

$$E[SQRes] = tr(\mathbf{V}^{-1}\mathbf{V}\sigma^2) + \theta'\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\theta - tr(\mathbf{P}\mathbf{V}\sigma^2) - \theta'\mathbf{X}'\mathbf{P}\mathbf{X}\theta;$$

$$E[SQRes] = \sigma^2(n - r(\mathbf{X})) + \theta'\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\theta - \theta'\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{+}\mathbf{X}\mathbf{V}^{-1}\mathbf{X}\theta;$$

$$E[SQRes] = \sigma^2(n - r(\mathbf{X}));$$

$$E[QMRes] = E\left[\frac{SQRes}{r(\mathbf{V}^{-1}) - r(\mathbf{X})}\right] = \frac{\sigma^2(n - r(\mathbf{X}))}{n - r(\mathbf{X})} = \sigma^2.$$

De maneira análoga, é possível encontrar a esperança da soma de quadrados de tratamento, de forma que

$$E[SQTrat] = E[Y'PY] - E[Y'P_1Y],$$

em que $\mathbf{P}_1 = \mathbf{V}^{-1}\mathbf{X}'_1(\mathbf{X}'_1\mathbf{V}^{-1}\mathbf{X}_1)^{-1}\mathbf{X}_1\mathbf{V}^{-1}$.

$$E[SQTrat] = \sigma^2 r(\mathbf{X}) + \theta'\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\theta - \sigma^2 tr(\mathbf{P}_1\mathbf{V}) - \theta'\mathbf{X}'\mathbf{P}_1\mathbf{X}\theta;$$

$$E[SQTrat] = \sigma^2(r(\mathbf{X}) - 1) + \underbrace{\theta'\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\theta - \theta'\mathbf{X}'\mathbf{P}_1\mathbf{X}\theta}_{(*)};$$

$$(*) = (\mu'\mathbf{X}'_1 + \alpha'\mathbf{X}'_2)\mathbf{V}^{-1}(\mathbf{X}_1\mu + \mathbf{X}_2\alpha) - (\mu'\mathbf{X}'_1 + \alpha'\mathbf{X}'_2)\mathbf{P}_1(\mathbf{X}_1\mu + \mathbf{X}_2\alpha);$$

$$\begin{aligned} (*) &= \\ & \mu^2\mathbf{X}'_1\mathbf{V}^{-1}\mathbf{X}_1 + \alpha'\mathbf{X}'_2\mathbf{V}^{-1}\mathbf{X}_2\alpha - \mu(\mathbf{X}'_1\mathbf{V}^{-1}\mathbf{X}_1)(\mathbf{X}'_1\mathbf{V}^{-1}\mathbf{X}_1)^{-1}\mathbf{X}'_1\mathbf{V}^{-1}\mathbf{X}_2\alpha \\ & + \mu\mathbf{X}'_1\mathbf{V}^{-1}\mathbf{X}_2\alpha + \alpha'\mathbf{X}'_2\mathbf{V}^{-1}\mathbf{X}_1\mu - \mu^2(\mathbf{X}'_1\mathbf{V}^{-1}\mathbf{X}_1)(\mathbf{X}'_1\mathbf{V}^{-1}\mathbf{X}_1)^{-1}(\mathbf{X}'_1\mathbf{V}^{-1}\mathbf{X}_1) \\ & - \alpha'\mathbf{X}'_2\mathbf{V}^{-1}\mathbf{X}_1(\mathbf{X}'_1\mathbf{V}^{-1}\mathbf{X}_1)^{-1}(\mathbf{X}'_1\mathbf{V}^{-1}\mathbf{X}_1)\mu - \alpha'\mathbf{X}'_2\mathbf{P}_1\mathbf{X}_2\alpha; \end{aligned}$$

$$(*) = \alpha'\mathbf{X}'_2(\mathbf{V}^{-1} - \mathbf{P}_1)\mathbf{X}_2\alpha.$$

Desse modo, a esperança do quadrado médio do tratamento é dada por

$$E[QMTrat] = E\left[\frac{SQTrat}{r(\mathbf{X}) - 1}\right] = \frac{1}{r(\mathbf{X}) - 1}E[SQTrat];$$

$$E[QMTrat] = \sigma^2 + \frac{1}{r(\mathbf{X}) - 1}\alpha'\mathbf{X}'_2(\mathbf{V}^{-1} - \mathbf{P}_1)\mathbf{X}_2\alpha.$$

Sob hipótese nula verdadeira, ou seja, se $\alpha = \mathbf{0}$, então, $E[QMTrat] = E[QMRes]$ e, com isso, a estatística F_0 da análise de variância é a razão entre dois estimadores não tendenciosos para σ^2 . Todos os cálculos efetuados nesta seção estão sintetizados na Tabela 12.

Tabela 12 Esperança dos quadrados médios de um DIC considerando erros espacialmente correlacionados.

| FV | E[QM] |
|------------|---|
| Tratamento | $\sigma^2 + \frac{1}{r(\mathbf{X}) - 1} \alpha' \mathbf{X}'_2 (\mathbf{V}^{-1} - \mathbf{P}_1) \mathbf{X}_2 \alpha$ |
| Resíduo | σ^2 |

4.2.4 Aplicabilidade do teste F

O teste F da análise de variância surge pela consideração da decomposição da variabilidade de um conjunto de dados em termos de somas de quadrados. Por meio deste teste é possível realizar inferências sobre a hipótese de igualdade entre várias médias dos níveis de algum fator de interesse.

Segundo Banzatto e Kronka (2006), para garantir a validade do teste F da análise de variância, é necessário que os erros sejam homocedásticos, independentes e normalmente distribuídos. Porém, a análise de variância apresentada na seção (4.1.2) parte do princípio de que os erros são espacialmente dependentes e, por isso, surge a necessidade de verificar se o teste F continua sendo válido nesta análise de variância.

De acordo com Rencher e Schaalje (2008), dado um vetor $Y \sim N(\mathbf{X}\theta, \Sigma)$ e $\mathbf{A}_{(n)}$ uma matriz real e simétrica de posto k , uma condição necessária e suficiente para $Y' \mathbf{A} Y \sim \chi^2_{(k, \delta)}$, em que $\delta = \frac{1}{2} \theta' \mathbf{X}' \mathbf{A} \mathbf{X} \theta$ é o parâmetro de não centralidade, é que $\mathbf{A} \Sigma$ seja idempotente.

Conforme visto anteriormente, se $Y \sim N(\mathbf{X}\theta, \Sigma)$, em que $\Sigma = \mathbf{V} \sigma^2$, então a soma de quadrados de resíduos é dada por

$$SQRes = Y' [\mathbf{V}^{-1} - \mathbf{V}^{-1} \mathbf{X} (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}^{-1}] Y.$$

Substituindo $\mathbf{V}^{-1} = \sigma^2 \Sigma^{-1}$, tem-se que

$$SQRes = Y' [\sigma^2 \Sigma^{-1} - \sigma^2 \Sigma^{-1} \mathbf{X} (\mathbf{X}' \sigma^2 \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}' \sigma^2 \Sigma^{-1}] Y;$$

$$\frac{SQRes}{\sigma^2} = Y'[\Sigma^{-1} - \Sigma^{-1}\mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X}) + \mathbf{X}'\Sigma^{-1}]Y.$$

Desse modo, definindo $\mathbf{A} = \Sigma^{-1} - \Sigma^{-1}\mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X}) + \mathbf{X}'\Sigma^{-1}$, é preciso mostrar que $\mathbf{A}\Sigma$ é idempotente:

$$\mathbf{A}\Sigma\mathbf{A}\Sigma = [\Sigma^{-1} - \Sigma^{-1}\mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X}) + \mathbf{X}'\Sigma^{-1}]\Sigma[\Sigma^{-1} - \Sigma^{-1}\mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X}) + \mathbf{X}'\Sigma^{-1}]\Sigma,$$

$$\mathbf{A}\Sigma\mathbf{A}\Sigma = [\Sigma^{-1} - \Sigma^{-1}\mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X}) + \mathbf{X}'\Sigma^{-1}][\mathbf{I} - \mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X}) + \mathbf{X}'\Sigma^{-1}]\Sigma,$$

$$\mathbf{A}\Sigma\mathbf{A}\Sigma = [\Sigma^{-1} - \Sigma^{-1}\mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X}) + \mathbf{X}'\Sigma^{-1}]\Sigma = \mathbf{A}\Sigma.$$

Logo, $\mathbf{A}\Sigma$ é idempotente e, por isso, $SQRes/\sigma^2 \sim \chi^2$ com graus de liberdade dado pelo posto de \mathbf{A} , que é calculado como

$$r(\mathbf{A}) = r(\Sigma^{-1} - \Sigma^{-1}\mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X}) + \mathbf{X}'\Sigma^{-1}) = r(\mathbf{I}_n - \Sigma^{-1}\mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X}) + \mathbf{X}');$$

$$r(\mathbf{A}) = tr(\mathbf{I}_n - \Sigma^{-1}\mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X}) + \mathbf{X}') = n - r(X).$$

Além disso, o parâmetro de não centralidade de $SQRes/\sigma^2$ é dado por

$$\delta = \frac{1}{2}\theta'\mathbf{X}'\mathbf{A}\mathbf{X}\theta = \frac{1}{2}\theta'\mathbf{X}'[\Sigma^{-1} - \Sigma^{-1}\mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X}) + \mathbf{X}'\Sigma^{-1}]\mathbf{X}\theta,$$

$$\delta = \frac{1}{2}\theta'[\mathbf{X}'\Sigma^{-1}\mathbf{X} - \mathbf{X}'\Sigma^{-1}\mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X}) + \mathbf{X}'\Sigma^{-1}\mathbf{X}]\theta = 0.$$

Portanto, $SQRes/\sigma^2$ tem distribuição χ^2 central com $(n - r(\mathbf{X}))$ graus de liberdade. Esse mesmo procedimento será utilizado para mostrar que $SQTrat/\sigma^2 \sim \chi^2$.

$$\frac{SQTrat}{\sigma^2} = Y'[\Sigma^{-1}\mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X}) + \mathbf{X}'\Sigma^{-1} - \Sigma^{-1}\mathbf{X}_1(\mathbf{X}'_1\Sigma^{-1}\mathbf{X}_1) + \mathbf{X}'_1\Sigma^{-1}]Y.$$

Dado $\mathbf{B} = \Sigma^{-1}\mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Sigma^{-1} - \Sigma^{-1}\mathbf{X}_1(\mathbf{X}'_1\Sigma^{-1}\mathbf{X}_1) + \mathbf{X}'_1\Sigma^{-1}$, é necessário mostrar que $\mathbf{B}\Sigma$ é idempotente:

$$\mathbf{B}\Sigma\mathbf{B}\Sigma = [\Sigma^{-1}\mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X}) + \mathbf{X}'\Sigma^{-1} - \Sigma^{-1}\mathbf{X}_1(\mathbf{X}'_1\Sigma^{-1}\mathbf{X}_1) + \mathbf{X}'_1\Sigma^{-1}] +$$

$$\begin{aligned}
& -\Sigma^{-1}\mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X})+\mathbf{X}'\Sigma^{-1}\mathbf{X}_1(\mathbf{X}'_1\Sigma^{-1}\mathbf{X}_1)^{-1}\mathbf{X}'_1\Sigma^{-1}+ \\
& -\Sigma^{-1}\mathbf{X}_1(\mathbf{X}'_1\Sigma^{-1}\mathbf{X}_1)^{-1}\mathbf{X}'_1\Sigma^{-1}\mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X})+\mathbf{X}'\Sigma^{-1}+ \\
& +\Sigma^{-1}\mathbf{X}_1(\mathbf{X}'_1\Sigma^{-1}\mathbf{X}_1)^{-1}\mathbf{X}'_1\Sigma^{-1}\mathbf{X}_1(\mathbf{X}'_1\Sigma^{-1}\mathbf{X}_1)^{-1}\mathbf{X}'_1\Sigma^{-1}]\Sigma.
\end{aligned}$$

Dado que $\mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X})+\mathbf{X}'\Sigma^{-1}\mathbf{X}_1 = \mathbf{X}_1$ e $\mathbf{X}'_1\Sigma^{-1}\mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X})+\mathbf{X}' = \mathbf{X}'_1$, encontra-se que

$$\mathbf{B}\Sigma\mathbf{B}\Sigma = [\Sigma^{-1}\mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X})+\mathbf{X}'\Sigma^{-1} - \Sigma^{-1}\mathbf{X}_1(\mathbf{X}'_1\Sigma^{-1}\mathbf{X}_1)^{-1}\mathbf{X}'_1\Sigma^{-1}] = \mathbf{B}\Sigma.$$

Logo, $\mathbf{B}\Sigma$ é idempotente e, desse modo, $SQTrat/\sigma^2 \sim \chi^2$ com graus de liberdade dado pelo posto de \mathbf{B} que é

$$\begin{aligned}
r(\mathbf{B}) &= r(\Sigma^{-1}\mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X})+\mathbf{X}'\Sigma^{-1} - \Sigma^{-1}\mathbf{X}_1(\mathbf{X}'_1\Sigma^{-1}\mathbf{X}_1)^{-1}\mathbf{X}'_1\Sigma^{-1}), \\
r(\mathbf{B}) &= r(\Sigma^{-1}\mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X})+\mathbf{X}' - \Sigma^{-1}\mathbf{X}_1(\mathbf{X}'_1\Sigma^{-1}\mathbf{X}_1)^{-1}\mathbf{X}'_1), \\
r(\mathbf{B}) &= tr(\Sigma^{-1}\mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X})+\mathbf{X}') - tr(\Sigma^{-1}\mathbf{X}_1(\mathbf{X}'_1\Sigma^{-1}\mathbf{X}_1)^{-1}\mathbf{X}'_1) = r(X) - 1.
\end{aligned}$$

Já o parâmetro de não centralidade de $SQTrat/\sigma^2$ é definido como

$$\begin{aligned}
\delta &= \frac{1}{2}\theta'(\mathbf{X}'\Sigma^{-1}\mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X})+\mathbf{X}'\Sigma^{-1}\mathbf{X} - \mathbf{X}'\Sigma^{-1}\mathbf{X}_1(\mathbf{X}'_1\Sigma^{-1}\mathbf{X}_1)^{-1}\mathbf{X}'_1\Sigma^{-1}\mathbf{X})\theta, \\
\delta &= \frac{1}{2}\theta'\mathbf{X}'(\Sigma^{-1} - \Sigma^{-1}\mathbf{X}_1(\mathbf{X}'_1\Sigma^{-1}\mathbf{X}_1)^{-1}\mathbf{X}'_1\Sigma^{-1})\mathbf{X}\theta, \\
\delta &= \frac{1}{2\sigma^2}\theta'\mathbf{X}'(\mathbf{V}^{-1} - \mathbf{P}_1)\mathbf{X}\theta = \frac{1}{2\sigma^2}\alpha'\mathbf{X}'_2(\mathbf{V}^{-1} - \mathbf{P}_1)\mathbf{X}_2\alpha.
\end{aligned}$$

Dessa forma, já que $SQRes/\sigma^2$ e $SQTrat/\sigma^2$ seguem distribuições qui-quadrado, para garantir a validade do teste F da análise de variância, basta mostrar que essas duas variáveis aleatórias são independentes. Segundo Graybill (1976), dado que $Y \sim N(\mathbf{X}\theta, \Sigma)$ e \mathbf{A} e \mathbf{B} são matrizes simétricas de ordem n , as formas quadráticas $Y'\mathbf{A}Y$ e $Y'\mathbf{B}Y$ são independentes se, e somente se, $\mathbf{A}\Sigma\mathbf{B} = \mathbf{0}$ (ou, equivalentemente, $\mathbf{B}\Sigma\mathbf{A} = \mathbf{0}$). Com isso, seja

$$\mathbf{A} = \Sigma^{-1} - \Sigma^{-1}\mathbf{X}(\mathbf{X}\Sigma^{-1}\mathbf{X})+\mathbf{X}\Sigma^{-1} = \Sigma^{-1} - \mathbf{M};$$

$$\mathbf{B} = \boldsymbol{\Sigma}^{-1} \mathbf{X}(\mathbf{X}\boldsymbol{\Sigma}^{-1}\mathbf{X})^+ \mathbf{X}\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1} \mathbf{X}_1(\mathbf{X}'_1\boldsymbol{\Sigma}^{-1}\mathbf{X}_1)^{-1} \mathbf{X}'_1\boldsymbol{\Sigma}^{-1} = \mathbf{M} - \mathbf{N};$$

$$\mathbf{A}\boldsymbol{\Sigma}\mathbf{B} = (\boldsymbol{\Sigma}^{-1} - \mathbf{M})\boldsymbol{\Sigma}(\mathbf{M} - \mathbf{N}) = \mathbf{M} - \mathbf{N} - \mathbf{M}\boldsymbol{\Sigma}\mathbf{M} + \mathbf{M}\boldsymbol{\Sigma}\mathbf{N}.$$

Porém, $\mathbf{M}\boldsymbol{\Sigma}\mathbf{M}$ e $\mathbf{M}\boldsymbol{\Sigma}\mathbf{N}$ são dadas por:

$$\mathbf{M}\boldsymbol{\Sigma}\mathbf{M} = \boldsymbol{\Sigma}^{-1} \mathbf{X}(\mathbf{X}\boldsymbol{\Sigma}^{-1}\mathbf{X})^+ \mathbf{X}\boldsymbol{\Sigma}^{-1} \mathbf{X}(\mathbf{X}\boldsymbol{\Sigma}^{-1}\mathbf{X})^+ \mathbf{X}\boldsymbol{\Sigma}^{-1},$$

$$\mathbf{M}\boldsymbol{\Sigma}\mathbf{M} = \boldsymbol{\Sigma}^{-1} \mathbf{X}(\mathbf{X}\boldsymbol{\Sigma}^{-1}\mathbf{X})^+ \mathbf{X}\boldsymbol{\Sigma}^{-1} = \mathbf{M} \quad \text{e}$$

$$\mathbf{M}\boldsymbol{\Sigma}\mathbf{N} = \boldsymbol{\Sigma}^{-1} \mathbf{X}(\mathbf{X}\boldsymbol{\Sigma}^{-1}\mathbf{X})^+ \mathbf{X}\boldsymbol{\Sigma}^{-1} \mathbf{X}_1(\mathbf{X}'_1\boldsymbol{\Sigma}^{-1}\mathbf{X}_1)^{-1} \mathbf{X}'_1\boldsymbol{\Sigma}^{-1},$$

$$\mathbf{M}\boldsymbol{\Sigma}\mathbf{N} = \boldsymbol{\Sigma}^{-1} \mathbf{X}_1(\mathbf{X}'_1\boldsymbol{\Sigma}^{-1}\mathbf{X}_1)^{-1} \mathbf{X}'_1\boldsymbol{\Sigma}^{-1} = \mathbf{N}.$$

Sendo assim, $\mathbf{A}\boldsymbol{\Sigma}\mathbf{B} = \mathbf{M} - \mathbf{N} - \mathbf{M} + \mathbf{N} = \mathbf{0}$ e, dessa forma, as variáveis $SQTrat/\sigma^2$ e $SQRes/\sigma^2$ são independentes. Com isso, pode-se concluir que

$$\frac{(SQTrat/\sigma^2)/(r(\mathbf{X}) - 1)}{(SQRes/\sigma^2)/(n - r(\mathbf{X}))} = \frac{SQTrat/(r(\mathbf{X}) - 1)}{SQRes/(n - r(\mathbf{X}))} = \frac{QMTrat}{QMRes} \sim F_{(\nu_1, \nu_2; \delta)},$$

em que $\nu_1 = r(\mathbf{X}) - 1$, $\nu_2 = n - r(\mathbf{X})$ e $\delta = \frac{1}{2\sigma^2} \alpha \mathbf{X}'_2(\mathbf{V}^{-1} - \mathbf{P}_1)\mathbf{X}_2\alpha$.

4.3 Delineamento em blocos casualizados (DBC)

Assim como para o delineamento inteiramente casualizado, além da análise de variância considerando erros espacialmente dependentes, foram apresentadas também as esperanças dos seus quadrados médios. Por fim, foi ainda verificada a validade do teste F para o delineamento em blocos casualizados.

4.3.1 Análise de variância

O modelo de um experimento instalado sob um delineamento em blocos casualizados, cujos erros são espacialmente correlacionados, é

definido por $Y = \mathbf{X}\theta + \varepsilon$, com $\varepsilon \sim N(\mathbf{0}, \mathbf{V}\sigma^2)$. Conforme visto na seção (3.2), este modelo pode ser convenientemente particionado como:

$$Y = \mathbf{X}_1\mu + \mathbf{X}_2\alpha + \mathbf{X}_3\beta + \varepsilon. \quad (4.13)$$

Prémultiplicando-se esse modelo por uma matriz \mathbf{L}' de ordem n , de modo que $\mathbf{V}^{-1} = \mathbf{L}\mathbf{L}'$, obtém-se um modelo transformado, que é dado por

$$W = \mathbf{Z}_1\mu + \mathbf{Z}_2\alpha + \mathbf{Z}_3\beta + \eta, \quad (4.14)$$

em que $W = \mathbf{L}'Y$, $\mathbf{Z}_1 = \mathbf{L}'\mathbf{X}_1$, $\mathbf{Z}_2 = \mathbf{L}'\mathbf{X}_2$, $\mathbf{Z}_3 = \mathbf{L}'\mathbf{X}_3$ e $\eta = \mathbf{L}'\varepsilon$. Conforme demonstrado na seção (4.1), $\eta \sim N(\mathbf{0}, \mathbf{I}\sigma^2)$.

A partir dessa partição é possível obter as somas de quadrados da análise de variância por meio da abordagem de modelos reduzidos. Conforme afirma Searle (1987), o modelo (4.14) apresenta as seguintes reduções:

- (I) $W = \mathbf{Z}_1\mu + \eta;$
- (II) $W = \mathbf{Z}_1\mu + \mathbf{Z}_2\alpha + \eta;$
- (III) $W = \mathbf{Z}_1\mu + \mathbf{Z}_2\alpha + \mathbf{Z}_3\beta + \eta.$

O estimador de μ para o modelo (I) é dado por $\mu_I^0 = (\mathbf{Z}'_1\mathbf{Z}_1)^{-1}\mathbf{Z}'_1W$, enquanto os estimadores para μ e α , considerando o modelo (II), são definidos como

$$\mu_{II}^0 = (\mathbf{Z}'_1\mathbf{Z}_1)^{-1}\mathbf{Z}'_1\mathbf{Q}W \quad \text{e} \quad \alpha_{II}^0 = (\mathbf{Z}'_2\mathbf{R}\mathbf{Z}_2)^{-1}\mathbf{Z}'_2\mathbf{R}W,$$

em que $\mathbf{Q} = \mathbf{I} - \mathbf{Z}_2(\mathbf{Z}'_2\mathbf{R}\mathbf{Z}_2)^{-1}\mathbf{Z}'_2\mathbf{R}$ e $\mathbf{R} = \mathbf{I} - \mathbf{Z}_1(\mathbf{Z}'_1\mathbf{Z}_1)^{-1}\mathbf{Z}'_1$.

Já o modelo apresentado em (III) corresponde ao modelo completo e, por isso

$$\theta^0 = \begin{bmatrix} \mu^0 \\ \alpha^0 \\ \beta^0 \end{bmatrix} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'W;$$

em que $\mathbf{Z} = [\mathbf{Z}_1 \dot{\vdots} \mathbf{Z}_2 \dot{\vdots} \mathbf{Z}_3]$.

Com base nesses estimadores, visando à análise de variância de um delineamento em blocos casualizados, é possível encontrar as seguintes somas de quadrados:

$$SQ(\mu) = SQParâmetros_I = \mu_I^{0'} \mathbf{Z}'_1 W = W' \mathbf{Z}_1 (\mathbf{Z}'_1 \mathbf{Z}_1)^{-1} \mathbf{Z}'_1 W;$$

$$SQ(\mu, \alpha) = SQParâmetros_{II} = \mu_{II}^0 \mathbf{Z}'_1 W + \alpha_{II}^0 \mathbf{Z}'_2 W;$$

$$SQ(\mu, \alpha) = W' \mathbf{Q}' \mathbf{Z}_1 (\mathbf{Z}'_1 \mathbf{Z}_1)^{-1} \mathbf{Z}'_1 W + W' \mathbf{R} \mathbf{Z}_2 (\mathbf{Z}'_2 \mathbf{R} \mathbf{Z}_2)^{-1} \mathbf{Z}'_2 W;$$

$$SQ(\mu, \alpha) = W' \mathbf{Z}_1 (\mathbf{Z}'_1 \mathbf{Z}_1)^{-1} \mathbf{Z}'_1 W + W' \mathbf{R} \mathbf{Z}_2 (\mathbf{Z}'_2 \mathbf{R} \mathbf{Z}_2)^{-1} \mathbf{Z}'_2 \mathbf{R} W;$$

$$SQ(\mu, \alpha, \beta) = SQParâmetros_{III} = \theta^{0'} \mathbf{Z}' W = W' \mathbf{Z} (\mathbf{Z}' \mathbf{Z})^{-1} \mathbf{Z}' W;$$

$$SQ(\alpha|\mu) = SQ(\mu, \alpha) - SQ(\mu) = W' \mathbf{R} \mathbf{Z}_2 (\mathbf{Z}'_2 \mathbf{R} \mathbf{Z}_2)^{-1} \mathbf{Z}'_2 \mathbf{R} W;$$

$$SQ(\beta|\mu, \alpha) = SQ(\mu, \alpha, \beta) - SQ(\mu, \alpha);$$

$$SQ(\beta|\mu, \alpha) = \theta^{0'} \mathbf{Z}' W - W' \mathbf{Z}_1 (\mathbf{Z}'_1 \mathbf{Z}_1)^{-1} \mathbf{Z}'_1 W - W' \mathbf{R} \mathbf{Z}_2 (\mathbf{Z}'_2 \mathbf{R} \mathbf{Z}_2)^{-1} \mathbf{Z}'_2 \mathbf{R} W.$$

A partir dessas somas de quadrados, como mostra Searle (1987), é possível obter a análise de variância para um delineamento em blocos casualizados, conforme apresentam-se na Tabela 13.

Tabela 13 Somas de quadrados para a análise de variância de um DBC, por meio de redução de modelos.

| FV | GL | SQ |
|------------|---------------------------------------|--------------------------------|
| Média | $r(\mathbf{X}_1)$ | $SQ(\mu)$ |
| Tratamento | $r(\mathbf{X}_2) - r(\mathbf{X}_1)$ | $SQ(\alpha \mu)$ |
| Bloco | $r(\mathbf{X}) - r(\mathbf{X}_2)$ | $SQ(\beta \mu, \alpha)$ |
| Resíduo | $r(\mathbf{I}_{(n)}) - r(\mathbf{X})$ | $W'W - SQ(\mu, \alpha, \beta)$ |
| Total | $r(\mathbf{I}_{(n)})$ | $W'W$ |

É comum obter as somas de quadrados de blocos e tratamentos, respectivamente, por $SQ(\beta|\mu)$ e $SQ(\alpha|\mu, \beta)$. Entretanto, como, neste trabalho, são estudados apenas experimentos balanceados, as duas formas de obtenção das somas de quadrados se equivalem.

Ao substituir $W = \mathbf{L}'Y$, $\mathbf{Z}_1 = \mathbf{L}'\mathbf{X}_1$, $\mathbf{Z}_2 = \mathbf{L}'\mathbf{X}_2$, $\mathbf{Z}_3 = \mathbf{L}'\mathbf{X}_3$ e $\mathbf{V}^{-1} = \mathbf{L}\mathbf{L}'$, essas somas de quadrados são dadas por

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

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

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

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

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

$$\text{em que } \mathbf{R}_v = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}_1(\mathbf{X}'_1\mathbf{V}^{-1}\mathbf{X}_1)^{-1}\mathbf{X}'_1\mathbf{V}^{-1}.$$

Com isso, a análise de variância que considera a correlação espacial entre os erros pode ser representada, por meio de formas quadráticas, conforme mostra a Tabela 14.

Tabela 14 Análise de variância de um DBC, cujos erros são espacialmente correlacionados.

| FV | GL | SQ |
|------------|--|---|
| Tratamento | $r(\mathbf{X}_2) - r(\mathbf{X}_1)$ | $Y'\mathbf{P}_2Y$ |
| Bloco | $r(\mathbf{X}) - r(\mathbf{X}_2)$ | $Y'(\mathbf{P} - \mathbf{P}_1 - \mathbf{P}_2)Y$ |
| Resíduo | $r(\mathbf{V}^{-1}) - r(\mathbf{X})$ | $Y'(\mathbf{V}^{-1} - \mathbf{P})Y$ |
| Total | $r(\mathbf{V}^{-1}) - r(\mathbf{X}_1)$ | $Y'(\mathbf{V}^{-1} - \mathbf{P}_1)Y$ |

em que $\mathbf{P} = \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}$, $\mathbf{P}_2 = \mathbf{R}_v\mathbf{X}_2(\mathbf{X}'_2\mathbf{R}_v\mathbf{X}_2)^{-1}\mathbf{X}'_2\mathbf{R}_v$ e $\mathbf{P}_1 = \mathbf{V}^{-1}\mathbf{X}_1(\mathbf{X}'_1\mathbf{V}^{-1}\mathbf{X}_1)^{-1}\mathbf{X}'_1\mathbf{V}^{-1}$.

4.3.2 Esperanças dos quadrados médios

Conforme visto na seção (2.2.3), o modelo estatístico de um delineamento em blocos casualizados é representado por

$$y_{ij} = \mu + \alpha_i + \beta_j + e_{ij}, \quad \text{com } i = 1, 2, \dots, t \text{ e } j = 1, 2, \dots, r. \quad (4.15)$$

De acordo com este modelo, os somas de quadrados de tratamento, blocos e resíduos são dadas da seguinte maneira:

$$SQTrat = \frac{1}{r} \sum_{i=1}^t y_{i\cdot}^2 - \frac{y_{\cdot\cdot}^2}{tr}, \quad SQB = \frac{1}{t} \sum_{j=1}^r y_{\cdot j}^2 - \frac{y_{\cdot\cdot}^2}{tr} \text{ e}$$

$$SQRes = \sum_{i=1}^t \sum_{j=1}^r y_{ij}^2 - \frac{1}{r} \sum_{i=1}^t y_{i\cdot}^2 - \frac{1}{t} \sum_{j=1}^r y_{\cdot j}^2 + \frac{y_{\cdot\cdot}^2}{tr};$$

cujos respectivos quadrados médios são

$$QMTrat = \frac{SQTrat}{t-1}, \quad QMBloco = \frac{SQBloco}{r-1} \text{ e } QMRes = \frac{SQRes}{(t-1)(r-1)}.$$

Em um delineamento em blocos casualizados há duas formas de considerar a dependência espacial dos erros: dependência espacial apenas dentro dos blocos e dependência espacial dentro e entre os blocos. Considerando que existe dependência espacial apenas dentro dos blocos e mantendo as pressuposições de que os erros têm média zero e variância constante, tem-se que

$$E[e_{ij}] = 0,$$

$$Var[e_{ij}] = \sigma^2,$$

$$Cov[e_{ij}, e_{i'j}] = E[e_{ij}e_{i'j}] \neq 0 \text{ para algum } i \neq i' \text{ e}$$

$$Cov[e_{ij}, e_{i'j'}] = 0 \quad \forall \quad j' \neq j.$$

Desse modo, as esperanças dos quadrados médios apresentados são dadas conforme a Tabela 15.

Tabela 15 Esperança dos quadrados médios de um DBC cujos erros têm dependência dentro dos blocos.

| FV | E[QM] |
|------------|---|
| Tratamento | $\sigma^2 + \frac{r}{t-1} \sum_i \alpha_i^2 - \frac{1}{tr(r-1)} \sum_j \sum_{i'>i} 2Cov(e_{ij}, e_{i'j})$ |
| Blocos | $\sigma^2 + \frac{t}{r-1} \sum_j \beta_j^2 + \frac{1}{tr} \sum_j \sum_{i'>i} 2Cov(e_{ij}, e_{i'j})$ |
| Resíduo | $\sigma^2 - \frac{1}{tr(r-1)} \sum_j \sum_{i'>i} 2Cov(e_{ij}, e_{i'j})$ |

Entretanto, ao considerar que existe dependência espacial dentro e entre os blocos, tem-se que $Cov[e_{ij}, e_{i'j'}] \neq 0$ para algum $ij \neq i'j'$ e, com isso, as esperanças dos quadrados médios da análise de variância são dadas conforme a Tabela 16, de modo que T , U e V são as mesmas equações definidas na seção (4.2.3).

Tabela 16 Esperança dos quadrados médios de um DBC cujos erros têm dependência dentro e entre blocos.

| FV | E[QM] |
|------------|---|
| Tratamento | $\sigma^2 + \frac{r}{t-1} \sum_i \alpha_i^2 + \frac{1}{tr} V - \frac{1}{tr(r-1)} (T + U)$ |
| Blocos | $\sigma^2 + \frac{t}{r-1} \sum_j \beta_j^2 + \frac{1}{tr} T - \frac{1}{tr(r-1)} (U + V)$ |
| Resíduo | $\sigma^2 + \frac{1}{tr(r-1)} (V - T) + \frac{1}{tr(r-1)(t-1)} U$ |

Já as esperanças dos quadrados médios da análise de variância apresentada na Tabela 14, na qual a matriz de correlação \mathbf{V} é utilizada no cálculo da somas de quadrados, são dadas por:

Tabela 17 Esperança dos quadrados médios da análise de variância de um DBC, considerando dependência espacial.

| FV | E[QM] |
|------------|---|
| Tratamento | $\sigma^2 + \frac{(\alpha' \mathbf{X}'_2 \mathbf{P}_2 \mathbf{X}_2 \alpha + \beta' \mathbf{X}'_3 \mathbf{P}_2 \mathbf{X}_3 \beta + 2\alpha' \mathbf{X}'_2 \mathbf{P}_2 \mathbf{X}_3 \beta)}{r(\mathbf{X}_2) - 1}$ |
| Blocos | $\sigma^2 + \frac{(\alpha' \mathbf{X}'_2 \mathbf{G} \mathbf{X}_2 \alpha + \beta' \mathbf{X}'_3 \mathbf{G} \mathbf{X}_3 \beta + 2\alpha' \mathbf{G} \mathbf{X}_3 \beta)}{r(\mathbf{X}) - r(\mathbf{X}_2)}$ |
| Resíduo | σ^2 |

em que $\mathbf{G} = \mathbf{R}_v - \mathbf{P}_2 = \mathbf{R}_v - \mathbf{R}_v \mathbf{X}_2 (\mathbf{X}'_2 \mathbf{R}_v \mathbf{X}_2)^{-1} \mathbf{X}'_2 \mathbf{R}_v$.

4.3.3 Aplicabilidade do teste F

Para garantir a validade do teste F , utilizado para comparar a igualdade entre as médias dos tratamentos e a igualdade entre as médias dos blocos, primeiramente, é necessário mostrar que as variáveis aleatórias $SQTrat/\sigma^2$, $SQBloco/\sigma^2$ e $SQRes/\sigma^2$ seguem distribuições qui-quadrado (χ^2).

A soma de quadrados de resíduos tem a mesma expressão para os delineamentos inteiramente casualizados e em blocos casualizados. Desse modo, conforme demonstrado na seção (4.2.4), $SQRes/\sigma^2$ tem distribuição χ^2 central, com $n - r(\mathbf{X})$ graus de liberdade.

O mesmo procedimento da seção anterior será utilizado para mostrar que $SQTrat/\sigma^2 \sim \chi^2$.

$$\frac{SQTrat}{\sigma^2} = Y'[\mathbf{R}_\Sigma \mathbf{X}_2 (\mathbf{X}'_2 \mathbf{R}_\Sigma \mathbf{X}_2)^{-1} \mathbf{X}'_2 \mathbf{R}_\Sigma] Y.$$

em que $\mathbf{R}_\Sigma = \Sigma^{-1} - \Sigma^{-1}\mathbf{X}_1(\mathbf{X}'_1\Sigma^{-1}\mathbf{X}_1)^{-1}\mathbf{X}'_1\Sigma^{-1}$.

Definindo $\mathbf{D} = \mathbf{R}_\Sigma\mathbf{X}_2(\mathbf{X}'_2\mathbf{R}_\Sigma\mathbf{X}_2)^{-}\mathbf{X}'_2\mathbf{R}_\Sigma$, é necessário mostrar que $\mathbf{D}\Sigma$ é idempotente. Observe que

$$\mathbf{R}_\Sigma\Sigma\mathbf{R}_\Sigma = \Sigma^{-1} - 2\Sigma^{-1}\mathbf{X}_1(\mathbf{X}'_1\Sigma^{-1}\mathbf{X}_1)^{-1}\mathbf{X}'_1\Sigma^{-1} + \Sigma^{-1}\mathbf{X}_1(\mathbf{X}'_1\Sigma^{-1}\mathbf{X}_1)^{-1}\mathbf{X}'_1\Sigma^{-1};$$

$$\mathbf{R}_\Sigma\Sigma\mathbf{R}_\Sigma = \Sigma^{-1} - \Sigma^{-1}\mathbf{X}_1(\mathbf{X}'_1\Sigma^{-1}\mathbf{X}_1)^{-1}\mathbf{X}'_1\Sigma^{-1} = \mathbf{R}_\Sigma.$$

Desse modo

$$\mathbf{D}\Sigma\mathbf{D}\Sigma = \mathbf{R}_\Sigma\mathbf{X}_2(\mathbf{X}'_2\mathbf{R}_\Sigma\mathbf{X}_2)^{-}\mathbf{X}'_2\mathbf{R}_\Sigma\Sigma\mathbf{R}_\Sigma\mathbf{X}_2(\mathbf{X}'_2\mathbf{R}_\Sigma\mathbf{X}_2)^{-}\mathbf{X}'_2\mathbf{R}_\Sigma\Sigma;$$

$$\mathbf{D}\Sigma\mathbf{D}\Sigma = \mathbf{R}_\Sigma\mathbf{X}_2(\mathbf{X}'_2\mathbf{R}_\Sigma\mathbf{X}_2)^{+}\mathbf{X}'_2\mathbf{R}_\Sigma\mathbf{X}_2(\mathbf{X}'_2\mathbf{R}_\Sigma\mathbf{X}_2)^{+}\mathbf{X}'_2\mathbf{R}_\Sigma\Sigma;$$

$$\mathbf{D}\Sigma\mathbf{D}\Sigma = \mathbf{R}_\Sigma\mathbf{X}_2(\mathbf{X}'_2\mathbf{R}_\Sigma\mathbf{X}_2)^{+}\mathbf{X}'_2\mathbf{R}_\Sigma\Sigma = \mathbf{D}\Sigma.$$

Com isso, $SQTrat/\sigma^2$ tem distribuição χ^2 não central, com parâmetro de não centralidade $\delta_t = \theta'\mathbf{X}'\mathbf{P}_2\mathbf{X}\theta/2\sigma^2$ e graus liberdade dado por

$$r(\mathbf{D}) = r(\mathbf{P}_2) = r([\mathbf{I}_{(n)} - \mathbf{X}_1(\mathbf{X}'_1\Sigma^{-1}\mathbf{X}_1)^{-1}\mathbf{X}'_1\Sigma^{-1}]\mathbf{X}_2(\mathbf{X}'_2\mathbf{R}_\Sigma\mathbf{X}_2)^{+}\mathbf{X}'_2\mathbf{R}_\Sigma);$$

$$r(\mathbf{D}) = -tr(\mathbf{X}_1(\mathbf{X}'_1\Sigma^{-1}\mathbf{X}_1)^{-1}\mathbf{X}'_1\Sigma^{-1}\mathbf{X}_2(\mathbf{X}'_2\mathbf{R}_\Sigma\mathbf{X}_2)^{+}\mathbf{X}'_2\mathbf{R}_\Sigma) + \\ + tr(\mathbf{X}_2(\mathbf{X}'_2\mathbf{R}_\Sigma\mathbf{X}_2)^{+}\mathbf{X}'_2\mathbf{R}_\Sigma);$$

$$r(\mathbf{D}) = tr(\mathbf{X}'_2\mathbf{R}_\Sigma\mathbf{X}_2(\mathbf{X}'_2\mathbf{R}_\Sigma\mathbf{X}_2)^{+}) = r(\mathbf{X}'_2\mathbf{R}_\Sigma\mathbf{X}_2) = r(\mathbf{X}_2) - 1.$$

A próxima etapa é mostrar que $SQBloco/\sigma^2$ tem distribuição qui-quadrado χ^2 . Para isso, definindo $\mathbf{D} = \mathbf{R}_\Sigma\mathbf{X}_2(\mathbf{X}'_2\mathbf{R}_\Sigma\mathbf{X}_2)^{+}\mathbf{X}'_2\mathbf{R}_\Sigma$, $\mathbf{M} = \Sigma^{-1}\mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{+}\mathbf{X}'\Sigma^{-1}$ e $\mathbf{N} = \Sigma^{-1}\mathbf{X}_1(\mathbf{X}'_1\Sigma^{-1}\mathbf{X}_1)^{+}\mathbf{X}'_1\Sigma^{-1}$, basta mostrar que $(\mathbf{M} - \mathbf{D} - \mathbf{N})\Sigma$ é idempotente:

$$(\mathbf{M} - \mathbf{D} - \mathbf{N})\Sigma(\mathbf{M} - \mathbf{D} - \mathbf{N})\Sigma = [\mathbf{M}\Sigma\mathbf{M} - \mathbf{M}\Sigma\mathbf{D} - \mathbf{M}\Sigma\mathbf{N} - \mathbf{D}\Sigma\mathbf{M} + \mathbf{D}\Sigma\mathbf{D} + \\ \mathbf{D}\Sigma\mathbf{N} - \mathbf{N}\Sigma\mathbf{M} + \mathbf{N}\Sigma\mathbf{D} + \mathbf{N}\Sigma\mathbf{N}]\Sigma;$$

Facilmente, pode-se verificar que $\mathbf{M}\Sigma\mathbf{M} = \mathbf{M}$, $\mathbf{D}\Sigma\mathbf{D} = \mathbf{D}$ e $\mathbf{N}\Sigma\mathbf{N} = \mathbf{N}$. Além disso, substituindo $\mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{+}\mathbf{X}'\Sigma^{-1}\mathbf{X}_1 = \mathbf{X}_1$ e

$\mathbf{X}'_1 = \mathbf{X}'_1 \Sigma^{-1} \mathbf{X} (\mathbf{X}'_1 \Sigma^{-1} \mathbf{X})^+ \mathbf{X}'_1$, tem-se que $\mathbf{M} \Sigma \mathbf{N} = \mathbf{N} \Sigma \mathbf{M} = \mathbf{N}$. Pode-se notar ainda que

$$\mathbf{D} \Sigma \mathbf{N} = \mathbf{R}_\Sigma \mathbf{X}_2 (\mathbf{X}'_2 \Sigma^{-1} \mathbf{X}_2)^+ \mathbf{X}'_2 (\Sigma^{-1} - \mathbf{N}) \Sigma \mathbf{N};$$

$$\mathbf{D} \Sigma \mathbf{N} = \mathbf{R}_\Sigma \mathbf{X}_2 (\mathbf{X}'_2 \Sigma^{-1} \mathbf{X}_2)^+ \mathbf{X}'_2 (\mathbf{N} - \mathbf{N}) = \mathbf{0};$$

$$\mathbf{N} \Sigma \mathbf{D} = \mathbf{N} \Sigma (\Sigma^{-1} - \mathbf{N}) \mathbf{X}_2 (\mathbf{X}'_2 \Sigma^{-1} \mathbf{X}_2)^+ \mathbf{X}'_2 \mathbf{R}_\Sigma;$$

$$\mathbf{N} \Sigma \mathbf{D} = (\mathbf{N} - \mathbf{N}) \mathbf{X}_2 (\mathbf{X}'_2 \Sigma^{-1} \mathbf{X}_2)^+ \mathbf{X}'_2 \mathbf{R}_\Sigma = \mathbf{0}.$$

Dado $\mathbf{X} (\mathbf{X}' \Sigma^{-1} \mathbf{X})^+ \mathbf{X}' \Sigma^{-1} \mathbf{X}_2 = \mathbf{X}_2$ e $\mathbf{X}'_2 \Sigma^{-1} \mathbf{X} (\mathbf{X}' \Sigma^{-1} \mathbf{X})^+ \mathbf{X}' = \mathbf{X}'_2$, é possível mostrar que

$$\mathbf{D} \Sigma \mathbf{M} = \mathbf{R}_\Sigma \mathbf{X}_2 (\mathbf{X}'_2 \Sigma^{-1} \mathbf{X}_2)^+ \mathbf{X}'_2 (\Sigma^{-1} - \mathbf{N}) \Sigma \mathbf{M};$$

$$\mathbf{D} \Sigma \mathbf{M} = \mathbf{R}_\Sigma \mathbf{X}_2 (\mathbf{X}'_2 \Sigma^{-1} \mathbf{X}_2)^+ \mathbf{X}'_2 \mathbf{M} - \mathbf{R}_\Sigma \mathbf{X}_2 (\mathbf{X}'_2 \Sigma^{-1} \mathbf{X}_2)^+ \mathbf{X}'_2 \mathbf{N};$$

$$\mathbf{D} \Sigma \mathbf{M} = \mathbf{R}_\Sigma \mathbf{X}_2 (\mathbf{X}'_2 \Sigma^{-1} \mathbf{X}_2)^+ \mathbf{X}'_2 \Sigma^{-1} - \mathbf{R}_\Sigma \mathbf{X}_2 (\mathbf{X}'_2 \Sigma^{-1} \mathbf{X}_2)^+ \mathbf{X}'_2 \mathbf{N};$$

$$\mathbf{D} \Sigma \mathbf{M} = \mathbf{R}_\Sigma \mathbf{X}_2 (\mathbf{X}'_2 \Sigma^{-1} \mathbf{X}_2)^+ \mathbf{X}'_2 (\Sigma^{-1} - \mathbf{N}) = \mathbf{D}.$$

$$\mathbf{M} \Sigma \mathbf{D} = \mathbf{M} \Sigma (\Sigma^{-1} - \mathbf{N}) \mathbf{X}_2 (\mathbf{X}'_2 \Sigma^{-1} \mathbf{X}_2)^+ \mathbf{X}'_2 \mathbf{R}_\Sigma;$$

$$\mathbf{M} \Sigma \mathbf{D} = \mathbf{M} \mathbf{X}_2 (\mathbf{X}'_2 \Sigma^{-1} \mathbf{X}_2)^+ \mathbf{X}'_2 \mathbf{R}_\Sigma - \mathbf{N} \mathbf{X}_2 (\mathbf{X}'_2 \Sigma^{-1} \mathbf{X}_2)^+ \mathbf{X}'_2 \mathbf{R}_\Sigma;$$

$$\mathbf{M} \Sigma \mathbf{D} = \Sigma^{-1} \mathbf{X}_2 (\mathbf{X}'_2 \Sigma^{-1} \mathbf{X}_2)^+ \mathbf{X}'_2 \mathbf{R}_\Sigma - \mathbf{N} \mathbf{X}_2 (\mathbf{X}'_2 \Sigma^{-1} \mathbf{X}_2)^+ \mathbf{X}'_2 \mathbf{R}_\Sigma;$$

$$\mathbf{M} \Sigma \mathbf{D} = (\Sigma^{-1} - \mathbf{N}) \mathbf{X}_2 (\mathbf{X}'_2 \Sigma^{-1} \mathbf{X}_2)^+ \mathbf{X}'_2 \mathbf{R}_\Sigma = \mathbf{D}.$$

Sendo assim

$$(\mathbf{M} - \mathbf{D} - \mathbf{N}) \Sigma (\mathbf{M} - \mathbf{D} - \mathbf{N}) \Sigma = [\mathbf{M} - \mathbf{D} - \mathbf{N} - \mathbf{D} + \mathbf{D} + \mathbf{0} - \mathbf{N} + \mathbf{0} + \mathbf{N}] \Sigma;$$

$$(\mathbf{M} - \mathbf{D} - \mathbf{N}) \Sigma (\mathbf{M} - \mathbf{D} - \mathbf{N}) \Sigma = (\mathbf{M} - \mathbf{D} - \mathbf{N}) \Sigma.$$

Portanto, $SQBloco/\sigma^2$ tem distribuição χ^2 com $r(\mathbf{M} - \mathbf{D} - \mathbf{N}) = r(\mathbf{P} - \mathbf{P}_1 - \mathbf{P}_2) = r(\mathbf{X}) - r(\mathbf{X}_2)$ graus de liberdade e parâmetro de não centralidade dado por

$$\delta_b = \frac{1}{2\sigma^2} \theta' \mathbf{X}' (\mathbf{R}_v - \mathbf{P}_2) \mathbf{X} \theta.$$

Dessa forma, já que $SQRes/\sigma^2$, $SQTrat/\sigma^2$ e $SQBloco/\sigma^2$ seguem distribuições qui-quadrado, para garantir a validade do teste F para

comparar a igualdade entre as médias dos tratamentos basta mostrar que $SQRes/\sigma^2$ e $SQTrat/\sigma^2$ são variáveis aleatórias independentes. Do mesmo modo, para garantir a validade desse teste para comparar a igualdade entre as médias dos blocos, é necessário mostrar que $SQRes/\sigma^2$ e $SQBloco/\sigma^2$ são independentes. Para isso, dado que

$$\frac{SQRes}{\sigma^2} = Y' \mathbf{A} Y, \text{ com } \mathbf{A} = \Sigma^{-1} - \Sigma^{-1} \mathbf{X} (\mathbf{X}' \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}' \Sigma^{-1} = \Sigma^{-1} - \mathbf{M};$$

$$\frac{SQTrat}{\sigma^2} = Y' \mathbf{D} Y, \text{ com } \mathbf{D} = \mathbf{R}_\Sigma \mathbf{X}_2 (\mathbf{X}'_2 \mathbf{R}_\Sigma \mathbf{X}_2)^{-1} \mathbf{X}'_2 \mathbf{R}_\Sigma;$$

$$\frac{SQBloco}{\sigma^2} = Y' (\mathbf{M} - \mathbf{D} - \mathbf{N}) Y, \text{ com } \mathbf{M} = \Sigma^{-1} \mathbf{X} (\mathbf{X}' \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}' \Sigma^{-1} \text{ e}$$

$$\mathbf{N} = \Sigma^{-1} \mathbf{X}_1 (\mathbf{X}'_1 \Sigma^{-1} \mathbf{X}_1)^{-1} \mathbf{X}'_1 \Sigma^{-1},$$

para demonstrar a independência entre as variáveis de interesse, conforme afirma Graybill (1976), basta mostrar que $\mathbf{D} \Sigma \mathbf{A} = \mathbf{0}$ e $(\mathbf{M} - \mathbf{D} - \mathbf{N}) \Sigma \mathbf{A} = \mathbf{0}$:

$$\mathbf{D} \Sigma \mathbf{A} = \mathbf{R}_\Sigma \mathbf{X}_2 (\mathbf{X}'_2 \mathbf{R}_\Sigma \mathbf{X}_2)^{-1} \mathbf{X}'_2 (\Sigma^{-1} - \mathbf{N}) \Sigma (\Sigma^{-1} - \mathbf{M});$$

$$\mathbf{D} \Sigma \mathbf{A} = \mathbf{R}_\Sigma \mathbf{X}_2 (\mathbf{X}'_2 \mathbf{R}_\Sigma \mathbf{X}_2)^{-1} \mathbf{X}'_2 (\Sigma^{-1} - \mathbf{M} - \mathbf{N} + \mathbf{N} \Sigma \mathbf{M});$$

$$\mathbf{D} \Sigma \mathbf{A} = \mathbf{R}_\Sigma \mathbf{X}_2 (\mathbf{X}'_2 \mathbf{R}_\Sigma \mathbf{X}_2)^{-1} \mathbf{X}'_2 (\Sigma^{-1} - \mathbf{M});$$

$$\mathbf{D} \Sigma \mathbf{A} = \mathbf{R}_\Sigma \mathbf{X}_2 (\mathbf{X}'_2 \mathbf{R}_\Sigma \mathbf{X}_2)^{-1} (\mathbf{X}'_2 \Sigma^{-1} - \mathbf{X}'_2 \Sigma^{-1}) = \mathbf{0}.$$

$$(\mathbf{M} - \mathbf{D} - \mathbf{N}) \Sigma \mathbf{A} = (\mathbf{M} - \mathbf{D} - \mathbf{N}) \Sigma (\Sigma^{-1} - \mathbf{M});$$

$$(\mathbf{M} - \mathbf{D} - \mathbf{N}) \Sigma \mathbf{A} = \mathbf{M} - \mathbf{M} \Sigma \mathbf{M} - \mathbf{D} + \mathbf{D} \Sigma \mathbf{M} - \mathbf{N} + \mathbf{N} \Sigma \mathbf{M};$$

$$(\mathbf{M} - \mathbf{D} - \mathbf{N}) \Sigma \mathbf{A} = \mathbf{M} - \mathbf{M} - \mathbf{D} + \mathbf{D} - \mathbf{N} + \mathbf{N} = \mathbf{0}.$$

Com isso, pode-se concluir que

$$\frac{(SQTrat/\sigma^2)/r(\mathbf{x}_2) - 1}{(SQRes/\sigma^2)/n - r(\mathbf{X})} = \frac{QMTrat}{QMRes} \sim F_{(\nu_1, \nu_2; \delta_t)},$$

$$\frac{(SQBloco/\sigma^2)/r(\mathbf{X}) - r(\mathbf{X}_2)}{(SQRes/\sigma^2)/n - r(\mathbf{X})} = \frac{QMBloco}{QMRes} \sim F_{(\nu_3, \nu_2; \delta_b)},$$

de forma que $\nu_1 = r(\mathbf{X}_2) - 1$, $\nu_2 = n - r(\mathbf{X})$, $\nu_3 = r(\mathbf{X}) - r(\mathbf{X}_2)$,
 $\delta_t = \frac{1}{2\sigma^2}\theta'\mathbf{X}'\mathbf{P}_2\mathbf{X}\theta$ e $\delta_b = \frac{1}{2\sigma^2}\theta'\mathbf{X}'(\mathbf{R}_v - \mathbf{P}_2)\mathbf{X}\theta$.

4.4 Análise dos dados simulados

No intuito de avaliar a eficiência da análise de variância que considera a informação de dependência espacial dos erros, para as quatro configurações de experimentos apresentadas na seção (3.1), foram simulados 500 conjuntos de dados, de acordo com cada configuração de dependência espacial analisada. Essa avaliação foi realizada com base na comparação da estatística F_0 obtida nas abordagens espacial e clássica.

O que espera-se é que a abordagem espacial seja mais eficiente na detecção de diferenças entre as médias dos tratamentos, uma vez que os experimentos foram simulados de forma que essas diferenças existam. Para isso, espera-se que a estatística F_0 seja maior na análise de variância com abordagem espacial.

Na Tabela 18 apresentam-se os resultados para os experimentos instalados em um DIC com 15 tratamentos e 8 repetições. Nesta mesma tabela é possível observar que, para cada configuração (efeito pepita, patamar, alcance) são apresentadas a porcentagem das simulações nas quais a consideração da informação espacial resultou em um aumento do valor da estatística F_0 , a porcentagem das simulações nas quais as abordagens espacial e clássica obtiveram resultados iguais e, por fim, a porcentagem média de aumento da estatística F_0 , ao se considerar a dependência espacial dos erros. É interessante ressaltar que foram avaliados apenas o valor pontual de aumento e igualdade da estatística F_0 , nas duas abordagens consideradas.

Por exemplo, para a configuração (0,1; 2,5; 3), considerando o

modelo esférico, pode-se observar que a modelagem da dependência espacial dos erros resultou em um aumento da estatística F_0 em 86,0% dos conjuntos de dados simulados. Ainda nessa configuração, em 0,6% das simulações, os resultados das duas abordagens coincidiram. Porém, houve um aumento médio de 31,5% no valor da estatística F_0 , quando se utilizou a abordagem espacial.

Tabela 18 Avaliação da abordagem geoestatística de um DIC com 15 tratamentos e 8 repetições.

| Configuração | Modelo esférico | | | Modelo exponencial | | |
|----------------|-----------------|--------|-------|--------------------|--------|-------|
| | Aumento | Iguais | Média | Aumento | Iguais | Média |
| (0,10; 2,5; 3) | 86,0 | 0,6 | 31,50 | 78,6 | 8,0 | 39,77 |
| (0,10; 2,5; 6) | 93,6 | 0,0 | 88,48 | 84,6 | 6,2 | 56,68 |
| (0,10; 2,5; 9) | 90,4 | 0,2 | 96,41 | 86,0 | 4,4 | 62,94 |
| (0,75; 2,5; 3) | 56,4 | 2,2 | 7,75 | 49,0 | 22,4 | 9,44 |
| (0,75; 2,5; 6) | 82,4 | 0,8 | 22,81 | 52,0 | 19,0 | 9,94 |
| (0,75; 2,5; 9) | 64,2 | 1,6 | 21,08 | 45,2 | 21,2 | 7,39 |

Já na Tabela 19 é possível observar os resultados obtidos quando foram analisados experimentos instalados sob um DIC com 6 tratamentos e 5 repetições.

Tabela 19 Avaliação da abordagem geoestatística de um DIC com 6 tratamentos e 5 repetições.

| Configuração | Modelo esférico | | | Modelo exponencial | | |
|----------------|-----------------|--------|-------|--------------------|--------|-------|
| | Aumento | Iguais | Média | Aumento | Iguais | Média |
| (0,10; 2,5; 2) | 32,6 | 30,8 | 4,83 | 48,0 | 24,2 | 17,83 |
| (0,10; 2,5; 4) | 71,0 | 7,6 | 49,84 | 55,6 | 17,4 | 36,35 |
| (0,10; 2,5; 6) | 72,4 | 8,0 | 65,43 | 61,6 | 16,0 | 34,65 |
| (0,75; 2,5; 2) | 22,8 | 43,6 | 14,15 | 36,8 | 33,6 | 5,49 |
| (0,75; 2,5; 4) | 46,8 | 22,0 | 15,28 | 39,2 | 33,2 | 9,44 |
| (0,75; 2,5; 6) | 47,4 | 22,4 | 19,35 | 36,2 | 37,2 | 8,13 |

Analisando-se os resultados das Tabelas 18 e 19 é possível observar

que, em ambos os modelos, esférico e exponencial, a modelagem da dependência espacial dos erros apresentou resultados ainda melhores na presença de forte grau de dependência espacial, ou seja, em experimentos gerados com menor efeito pepita.

Com relação ao alcance, o que se pode notar é que, para maiores alcances, os resultados a favor da abordagem espacial são melhores. Além disso, observa-se, ainda, que quanto menor o alcance maior é a porcentagem de resultados iguais para as duas abordagens, o que pode ser justificado pelo fato de que, quanto menor o alcance maior é a chance de se ajustar o modelo de efeito pepita puro ao semivariograma dos erros e, com isso, a matriz \mathbf{V} reduz-se a uma matriz identidade, tornando idênticas as expressões das somas de quadrados das análises clássica e espacial.

Comparando-se os modelos esférico e exponencial, o que se pode perceber é que, para alcances maiores, a porcentagem média de aumento da estatística F_0 ao se considerar a dependência espacial dos erros é maior quando considerado o modelo esférico.

Em relação ao número de parcelas do experimento, o que se nota é que a modelagem da dependência espacial dos erros é mais eficiente, considerando-se o aumento do valor da estatística F_0 , em experimentos com maior número de parcelas.

As Tabelas 20 e 21 referem-se à avaliação da abordagem espacial com relação aos efeitos de tratamento e bloco, respectivamente, de experimentos instalados de acordo com um DBC com 15 tratamentos e 8 blocos.

Analisando-se os dados da Tabela 20 é possível notar que, assim como nos experimentos em DIC, os melhores resultados a favor da abordagem espacial ocorreram na presença de forte grau de dependência espacial.

Comparando-se os modelos esférico e exponencial, pode-se verificar que a porcentagem das simulações nas quais as abordagens clássica e espacial coincidiram é maior no modelo exponencial.

Tabela 20 Avaliação da abordagem geoestatística de um DBC com 15 tratamentos e 8 blocos, em relação ao efeito de tratamento.

| Configuração | Modelo esférico | | | Modelo exponencial | | |
|----------------|-----------------|--------|-------|--------------------|--------|-------|
| | Aumento | Iguais | Média | Aumento | Iguais | Média |
| (0,10; 2,5; 3) | 41,8 | 0,2 | 1,81 | 40,6 | 32,6 | 7,09 |
| (0,10; 2,5; 6) | 74,4 | 0,6 | 33,89 | 52,6 | 28,2 | 12,86 |
| (0,10; 2,5; 9) | 69,2 | 0,2 | 29,44 | 55,2 | 25,4 | 14,09 |
| (0,75; 2,5; 3) | 19,6 | 0,4 | -8,67 | 16,6 | 48,8 | -1,18 |
| (0,75; 2,5; 6) | 41,8 | 4,4 | 0,50 | 16,6 | 49,0 | -2,18 |
| (0,75; 2,5; 9) | 26,2 | 10,2 | -7,54 | 11,6 | 59,0 | -1,92 |

Pode-se verificar, ainda, que, em geral, a distâncias maiores, a porcentagem média de aumento da estatística F_0 ao se considerar a dependência espacial dos erros também aumentou.

Tabela 21 Avaliação da abordagem geoestatística de um DBC com 15 tratamentos e 8 blocos, em relação ao efeito de bloco.

| Configuração | Modelo esférico | | | Modelo exponencial | | |
|----------------|-----------------|--------|-------|--------------------|--------|-------|
| | Aumento | Iguais | Média | Aumento | Iguais | Média |
| (0,10; 2,5; 3) | 51,0 | 0,2 | 6,04 | 41,8 | 32,6 | 6,54 |
| (0,10; 2,5; 6) | 75,0 | 0,6 | 52,02 | 49,4 | 28,2 | 10,54 |
| (0,10; 2,5; 9) | 73,4 | 0,2 | 30,12 | 47,0 | 25,2 | 11,94 |
| (0,75; 2,5; 3) | 28,8 | 5,4 | -5,15 | 15,2 | 48,4 | -2,15 |
| (0,75; 2,5; 6) | 38,6 | 4,4 | -0,44 | 11,8 | 49,0 | -3,25 |
| (0,75; 2,5; 9) | 26,2 | 10,2 | -4,9 | 8,0 | 59,0 | -2,76 |

De acordo com os dados Tabela 21, pode-se dizer que, ao gerar dados com moderada dependência espacial, isto é, com efeito pepita 0,75, a consideração da informação espacial na construção da análise de variância forneceu, em média, valores para F_0 menores do que os encontrados por meio da análise que supunha independência espacial dos erros.

Com relação ao parâmetro alcance, observa-se que, considerando o efeito pepita 0,1, o número de simulações nas quais a abordagem espacial resultou em valores maiores para F_0 , comparando-se aos valores obtidos pela

abordagem espacial, é maior quando se consideram maiores valores para o parâmetro alcance.

As Tabelas 22 e 23 referem-se à avaliação da abordagem espacial com relação aos efeitos de tratamento e bloco, respectivamente, de experimentos instalados de acordo com um DBC com 6 tratamentos e 5 blocos.

Tabela 22 Avaliação da abordagem geoestatística de um DBC com 6 tratamentos e 5 blocos, em relação ao efeito de tratamento.

| Configuração | Modelo esférico | | | Modelo exponencial | | |
|----------------|-----------------|--------|-------|--------------------|--------|-------|
| | Aumento | Iguais | Média | Aumento | Iguais | Média |
| (0,10; 2,5; 2) | 28,2 | 60,6 | 6,90 | 29,6 | 57,2 | 5,18 |
| (0,10; 2,5; 4) | 52,4 | 33,4 | 25,99 | 37,0 | 48,0 | 7,38 |
| (0,10; 2,5; 6) | 51,2 | 36,6 | 23,13 | 37,2 | 49,6 | 8,63 |
| (0,75; 2,5; 2) | 21,8 | 64,4 | 4,50 | 19,4 | 63,8 | 2,00 |
| (0,75; 2,5; 4) | 33,2 | 50,8 | 10,60 | 16,6 | 67,6 | 0,73 |
| (0,75; 2,5; 6) | 26,6 | 58,2 | 6,75 | 19,4 | 66,0 | 1,84 |

De acordo com os dados Tabela 22 é possível observar que houve alta porcentagem de simulações nas quais as abordagens espacial e clássica obtiveram os mesmos resultados, ou seja, houve alta porcentagem de simulações, nas quais se ajustou um modelo de efeito pepita puro ao semivariograma dos erros, significando que o semivariograma não detectou dependência espacial entre os erros. Esse resultado pode ser justificado pelo fato de se estar analisando um experimento com pequeno número de parcelas.

Com relação à porcentagem de aumento da estatística F_0 , o que se pode notar é que a abordagem espacial apresentou melhores resultados quando se considerou o modelo esférico, de modo que, nesse modelo, o valor de alcance 4 propiciou os melhores resultados em favor da abordagem espacial.

Por fim, pode-se observar, ainda, que, assim como ocorreu nas configurações de experimento analisadas anteriormente, ao se considerar o

valor de efeito pepita 0,1 a abordagem espacial foi mais eficiente do que se considerando o valor de efeito pepita 0,75.

Na Tabela 23 é possível observar que a estatística F_0 , utilizada para comparar as médias dos blocos, em média, diminuiu quando a informação espacial dos erros foi considerada. Um dos motivos que podem justificar esse fato é que a modelagem da dependência espacial dos erros tenha conseguido explicar os efeitos referentes aos blocos e, desse modo, a variabilidade explicada por esse fator deixou de ser significativa, resultando em valores para F_0 menores quando se considerou a estrutura de dependência espacial.

Por fim, pode-se verificar, ainda, que, na maioria das configurações, a maior parte das simulações resultou nos mesmos valores para F_0 comparando-se as abordagens espacial e clássica.

Tabela 23 Avaliação da abordagem geoestatística de um DBC com 6 tratamentos e 5 blocos, em relação ao efeito de bloco.

| Configuração | Modelo esférico | | | Modelo exponencial | | |
|----------------|-----------------|--------|-------|--------------------|--------|-------|
| | Aumento | Iguais | Média | Aumento | Iguais | Média |
| (0,10; 2,5; 2) | 3,6 | 60,6 | -3,99 | 8,2 | 57,2 | -5,39 |
| (0,10; 2,5; 4) | 18,0 | 33,4 | -3,26 | 8,4 | 48,0 | -6,41 |
| (0,10; 2,5; 6) | 14,2 | 36,6 | -5,72 | 7,2 | 49,6 | -6,11 |
| (0,75; 2,5; 2) | 2,4 | 64,4 | -5,54 | 5,0 | 63,8 | -5,42 |
| (0,75; 2,5; 4) | 8,2 | 50,8 | -6,25 | 2,6 | 67,6 | -5,90 |
| (0,75; 2,5; 6) | 7,4 | 58,2 | -5,46 | 3,6 | 66,0 | -5,62 |

4.5 Exemplo de aplicação

Nesta seção, o intuito foi ilustrar a aplicação da análise espacial de experimentos com erros dependentes, discutida durante todo o trabalho. Os dados analisados foram obtidos a partir de simulação, sendo consideradas as distâncias entre os centroides das parcelas, de modo que os erros foram gerados seguindo modelo de semivariância esférico. Esse experimento foi instalado de acordo com um delineamento em blocos casualizados (DBC) composto por 26 tratamentos e 4 blocos. Esses dados podem ser visualizados

no ANEXO A.

Toda a análise desse experimento foi implementada no software *R* (R DEVELOPMENT CORE TEAM, 2012), de modo que as rotinas utilizadas nesse tipo de análise podem ser vistas no ANEXO C.

A seguir serão descritos os passos básicos necessários para a construção da análise de variância, considerando-se dependência espacial dos erros.

1º passo: Detecção da variabilidade espacial

O primeiro passo da análise espacial de um experimento é verificar se existe variabilidade espacial dos erros. Na Figura 5 observa-se o *grid* desse experimento, no qual são especificados os blocos que o compõem. Os pontos identificados nesse gráfico referem-se aos erros obtidos a partir de uma análise clássica. Esses erros foram divididos em quatro grupos, de acordo com o quartil a que pertencem, conforme apresentado na Figura 5.

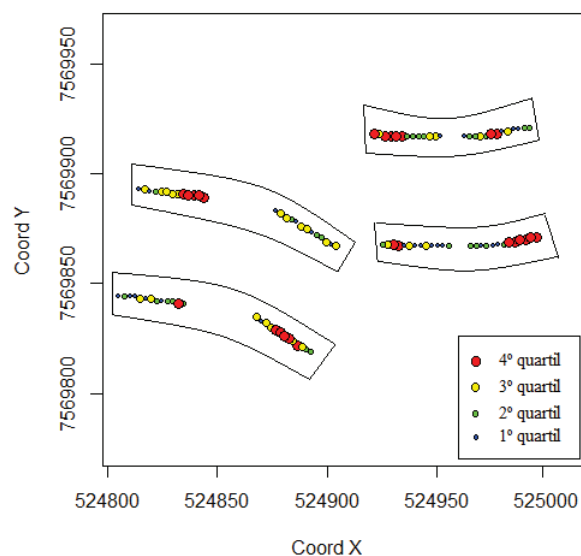


Figura 5 *Grid* do experimento contendo informações sobre os erros referentes a cada ponto amostrado

Por meio Figura 5 é possível perceber que os pontos mais próximos apresentam maior similaridade do que pontos mais distantes, indicando que pode existir uma dependência espacial entre os erros. Conforme visto na seção (3.3), a verificação da existência dessa dependência espacial pode ser feita por meio da análise gráfica do semivariograma. Esse semivariograma é contruído a partir dos resíduos após a retirada de tendência que, neste caso, refere-se à tendência ocasionada pelos efeitos de tratamentos e blocos, ou seja, os resíduos obtidos na análise clássica. Na seção citada, no intuito de contornar os problemas de subjetividade, sugeriu-se, ainda, que fossem construídos envelopes, obtidos por meio da permutação dos valores observados nas posições espaciais, para auxiliar na detecção da estrutura de dependência espacial. O Gráfico 6 mostra o semivariograma dos erros e o envelope construído para o mesmo.

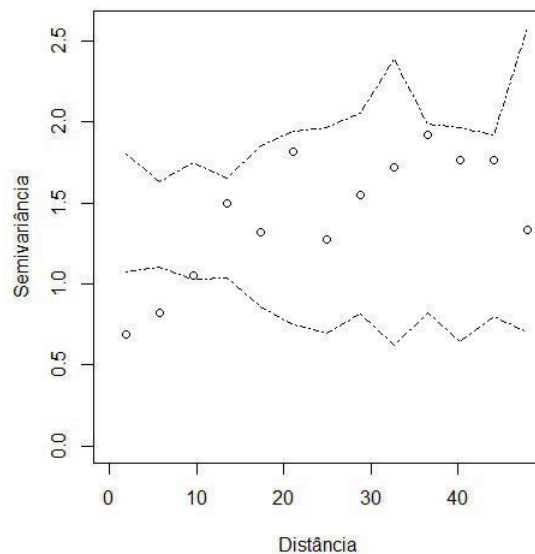


Gráfico 6 Semivariograma dos erros contendo o envelope simulado

Observando-se o semivariograma do Gráfico 6 percebe-se que o mesmo apresenta uma estrutura de dependência espacial definida, em que os

valores de semivariância aumentam de acordo com o aumento dos valores da distância (h) até um determinado valor, a partir do qual a semivariância se estabiliza. A construção do envelope confirma a existência de dependência espacial, uma vez que é possível observar a existência de pontos do gráfico que não pertence ao intervalo criado.

2º passo: Ajuste do semivariograma dos erros

De posse do semivariograma experimental, a etapa seguinte caracterizou-se pelo ajuste de um modelo teórico a esse semivariograma. Esse ajuste foi realizado pelo método de quadrados mínimos ordinários. O modelo ajustado ao semivariograma dos erros foi o modelo esférico, de modo que esse ajuste foi realizado pelo método de quadrados mínimos ordinários.

Para se obter estimativas mais eficientes para os parâmetros do semivariograma, foram realizadas várias iterações, até que os valores dessas estimativas se estabilizassem. Nesse processo iterativo, os parâmetros estimados a cada iteração foram utilizados para modelar a matriz de correlação \mathbf{V} , utilizada na ponderação da solução do sistema de equações normais (S.E.N), a partir da qual obtém-se a estimação do vetor de parâmetros dos efeitos fixos do modelo. Com esse novo vetor de parâmetros estimados foi possível atualizar o vetor de erros, construir novamente o semivariograma e, com isso, atualizar as estimativas dos parâmetros.

Na Tabela 24 observam-se os valores estimados para os parâmetros do semivariograma a cada iteração, de modo que φ^2 representa a contribuição, τ^2 é o efeito pepita e ϕ é o alcance prático.

Tabela 24 Parâmetros estimados do semivariograma a cada iteração.

| Parâmetros | Iterações | | | | | |
|-------------|-----------|---------|---------|---------|---------|---------|
| | 1ª | 2ª | 3ª | 4ª | 5ª | 6ª |
| φ^2 | 1,1062 | 1,8008 | 1,9458 | 1,9606 | 1,9616 | 1,9616 |
| τ^2 | 0,4647 | 0,1406 | 0,1073 | 0,1056 | 0,1055 | 0,1055 |
| ϕ | 22,1173 | 22,5271 | 22,2722 | 22,2354 | 22,2327 | 22,2327 |

De acordo com essa tabela, é possível observar que, a partir da 5ª iteração, esses valores estabilizaram e, portanto, os valores estimados correspondentes a essa iteração foram utilizados para a modelagem da matriz de covariância dos erros.

O ajuste do modelo esférico, utilizando-se os valores dos parâmetros estimados na 5ª iteração, pode ser observado no Gráfico 7.

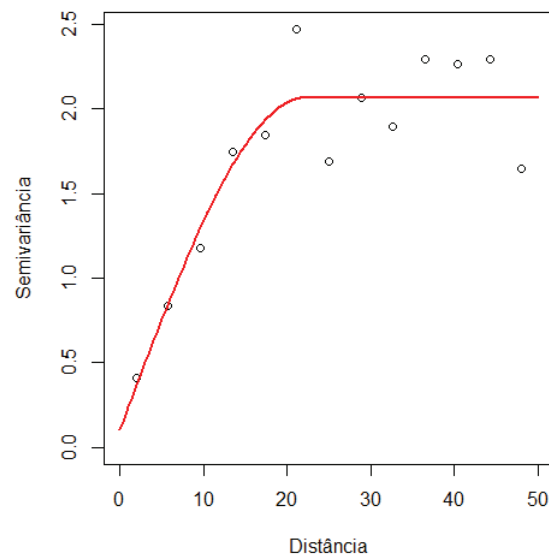


Gráfico 7 Ajuste do modelo esférico ao semivariograma dos erros

3º passo: Obtenção da análise de variância

Após ajustar o modelo teórico adequado ao semivariograma e, conseqüentemente, calcular a matriz de covariâncias dos erros, a etapa seguinte consistiu em construir a análise de variância que considerasse a estrutura de dependência espacial dos erros. Essa análise de variância, apresentada na Tabela 25, foi construída a partir das expressões contidas na Tabela 14, apresentada na seção 4.3.1.

Tabela 25 Análise de variância, considerando erros correlacionados.

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

Observando-se os dados da Tabela 25 é possível concluir que os efeitos de tratamentos foram significativos ao nível de 5% de probabilidade, o que implica dizer que as médias dos tratamentos não são todas iguais. Já em relação ao fator blocos, o que se pode concluir é que não existe diferença significativa entre as médias dos blocos.

No intuito de comparar a abordagem clássica e a espacial, foi construída a análise de variância clássica, cujas somas de quadrados foram calculadas por meio das expressões usuais, desconsiderando a informação espacial dos dados. Na Tabela 26 apresenta-se essa análise de variância.

Tabela 26 Análise de variância, considerando erros independentes.

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

De acordo com essa análise, o que se pode observar é que, ao desconsiderar a informação sobre a dependência espacial dos erros, os efeitos de tratamentos foram considerados não significativos e, conseqüentemente, aceitava-se a hipótese de igualdade entre as médias dos tratamentos. Já os efeitos de blocos foram considerados significativos ao nível de 5% de

O arquivo de dados desse comando é um *data.frame* que contém as coordenadas nas colunas 1 e 2, os dados observados na coluna 3 e os efeitos de tratamento e bloco nas colunas 4 e 5. Além disso, para a análise de experimentos, os tratamentos e blocos devem ser transformados em fatores. Vale ressaltar ainda que a função *likfit* ajusta o modelo teórico ao semivariograma por meio dos métodos de máxima verossimilhança (*default*) ou máxima verossimilhança restrita.

Conforme visto na seção 2.1.4, dado o modelo apresentado em (2.16), assume-se que $S(x)$ tem variância φ^2 e função de covariância $C(\cdot)$ parametrizada por ϕ (alcance). De acordo com esse modelo, assume-se, ainda, que $\xi(x)$ é uma variável aleatória independente e identicamente distribuída com média zero e variância τ^2 .

Afim de verificar, intuitivamente, se essas suposições foram satisfeitas, foi contruído o semivariograma da componente espacial (\hat{S}) e do erro experimental (ξ^0), apresentados no Gráfico 8. A esses semivariogramas foi ajustado o modelo esférico a partir do método de quadrados mínimos ordinários.

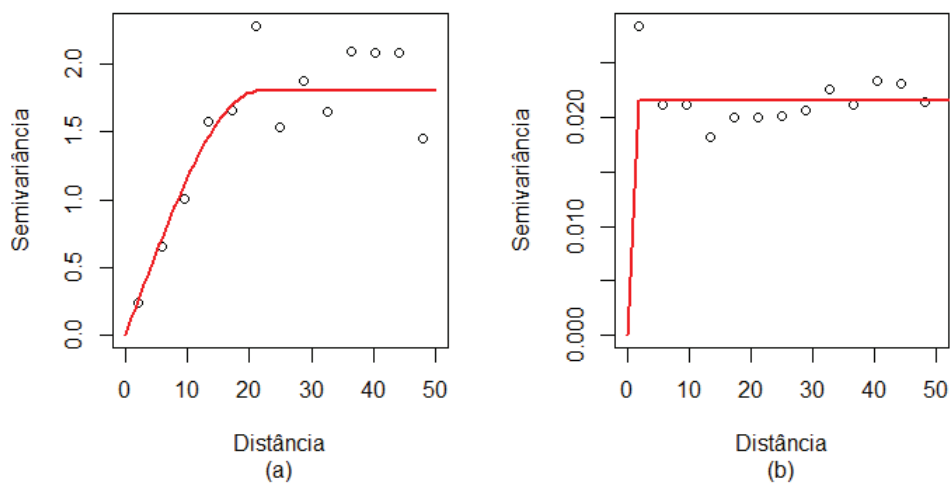


Gráfico 8 (a) Semivariograma com ajuste do modelo teórico da componente espacial \hat{S} e (b) dos erros estimados ξ^0

A partir da análise desses gráficos é possível observar que o semivariograma da componente espacial descreve toda a variabilidade espacial presente nos dados. Por sua vez, ao semivariograma dos erros ajustou-se o modelo de efeito pepita puro, o que significa dizer que, a partir da menor distância amostrada, não existe dependência espacial entre os erros estimados e, portanto, a variabilidade dos erros é totalmente decorrente do acaso. Esses resultados comprovam, subjetivamente, que as suposições em relação às distribuições dos componentes do modelo foram atendidas.

5 CONCLUSÕES

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

É importante ressaltar que a análise de variância apresentada neste trabalho não tem como enfoque reduzir a variabilidade residual e, sim obter estimadores dessa variabilidade que considerem a estrutura de dependência espacial, de modo que a análise de variância continue sendo uma ferramenta apropriada para avaliar a igualdade entre as médias dos fatores controlados no experimento,

Na abordagem espacial, o ajuste do semivariograma exerce um papel fundamental e determinante para a eficiência dessa análise. Diante disso, a modelagem do semivariograma deve ser uma etapa cautelosa, a qual exige, além de conhecimentos geoestatísticos, o conhecimento do fenômeno que se deseja modelar,

Em experimentos com pequeno número de parcelas não é possível estimar precisamente os parâmetros do semivariograma, resultando, muitas vezes, no ajuste do efeito pepita puro ao semivariograma. Ao ajustar o modelo de efeito pepita puro, a matriz de correlação dos erros reduz-se a uma matriz identidade, fazendo com que a análise de variância considerando a dependência espacial dos erros seja exatamente a mesma análise de variância obtida por meio da abordagem clássica,

Em geral, em experimentos com maior número de parcelas, a modelagem da correlação espacial dos erros tornou a análise mais eficiente, uma vez que, na maioria das vezes, produziu maiores valores para a estatística F_0 , com isso, aumentando a chance de se detectar diferenças entre as médias dos tratamentos,

Diante do que foi exposto, entende-se que é realmente necessária

a difusão da abordagem da análise espacial de experimentos, uma vez que a consideração da informação sobre a dependência espacial dos erros pode alterar completamente as conclusões sobre um determinado experimento.

REFERÊNCIAS

BANZATTO, D. A.; KRONKA, S. N. **Experimentação agrícola**. 4. ed. Jaboticabal: FUNEP, 2006. 237 p.

BARBIN, D. **Componentes de variância: teoria e aplicações**. 2 ed. Piracicaba: FEALQ, 1998. 120 p.

BURROUG, P.; **Principles of geographical information systems for land resources assessment**. Oxford: Clarendon, 1987. 193 p.

CAMBARDELLA, C. A. et al. Field scale variability of soil properties in central Iowa soils. **Soil Science Society of America Journal**, Madison, v. 58, n. 5, p. 1501-1511, 1994.

CAMPOS, J. F. **Eficiência da análise estatística espacial na classificação de famílias do feijoeiro: estudo via simulação**. 2011. 62 p. Dissertação (Mestrado em Estatística Aplicada a Biometria) – Universidade Federal de Viçosa, Viçosa, MG, 2011.

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

COCHRAN, W. G. Some consequences when the assumptions for the analysis of the variance are not satisfied. **Biometrics**, Raleigh, v. 3, n. 1, p. 3-22, 1947.

DIGGLE, P. J.; RIBEIRO JÚNIOR, P. J. **Model-based geostatistics**. New York: Springer, 2007. 230 p.

DUARTE, J. B. **Sobre o emprego e a análise estatística do delineamento em blocos aumentados no melhoramento genético vegetal**. 2000. 293 p. Tese (Doutorado em Agronomia) - Escola Superior de Agricultura “Luiz de Queiroz”, Piracicaba, 2000.

DUARTE, J. B.; VENCOSKY, R. Spatial statistical analysis and selection of genotypes in plant breeding. **Pesquisa Agropecuária Brasileira**, Brasília, v. 40, n. 2, p. 107-114, fev. 2005.

GRAYBILL, F. A. **Theory and application of the linear model.** Boston: Duxbury, 1976. 704 p.

GRONDONA, M. O.; CRESSIE, N. Using spatial considerations in the analysis of experiments. **Technometrics**, Alexandria, v. 33, n. 4, p. 381-392, 1991.

GUIMARÃES, E. C. **Geoestatística básica e aplicada.** Uberlândia: UFU/FAMAT, 2004. 77 p. Apostila.

HINKELMANN, K.; KEMPTHORNE, O. **Designs and analysis of experiments: introduction to experimental design.** New York: J. Wiley, 1994. v.1, 495 p.

HOEF, J. M. ver; CRESSIE, N. Spatial statistics: analysis of field experiments. In SCHEINER S.; GUREVITCH, J. (Ed.). **Design and analysis of ecological experiments.** London: Chapman and Hall, 1993. p. 319-341.

ISAAKS, E. H.; SRIVASTAVA, R. M. **An introduction to applied geostatistics.** New York: Oxford University, 1989. 561 p.

JOURNEL, A. G. **Fundamentals of geostatistics in five lessons.** Washington: American Geophysical Union, 1989. 40 p.

_____. **Geostatistics for the environmental sciences.** Las Vegas: Report of United States Environmental Protection Agency, 1987. 135 p.

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

KEMPTHORNE, O. Why Randomize? **Journal of Statistical Planning and Inference**, Amsterdam, v. 1, p. 1-25, 1977.

MARTINEZ, R. Control de la correlación espacial en experimentos de campo em el sector agrícola. **Agronomía Colombiana**, Bogotá, v. 11, p. 83-89, 1994.

MATHERON, G. **Traite de geostatistique applique**, Tome I. Mémoires du Bureau de Recherches Géologiques et Minières, Paris: Technip, 1962. 333 p.

MCBRATNEY, A. B.; WEBSTER, R. Choosing functions for semivariograms of soil properties and fitting them to sampling estimates. **Journal of Soil Science**, New Brunswick, v. 37, n. 6, p. 617-639, 1986.

MONTGOMERY, D. C. **Design and analysis of experiments**. 7th ed. Hoboken: J. Wiley, 2008. 656 p.

MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. **Introduction to the theory of statistics**. 3rd ed. Tokio: McGraw-Hill Kogakusha, 1974. 564 p.

OLIVEIRA, M. S. de. **Planos amostrais para variáveis espaciais utilizando geoestatística**. 1991. 100 p. Dissertação (Mestrado em Estatística) – Universidade Estadual de Campinas, Campinas, 1991.

PONTES, J. M. **Geoestatística: aplicações em experimentos de campo**. 2002. 82 p. Dissertação (Mestrado em Agronomia - Estatística e Experimentação Agropecuária) – Universidade Federal de Lavras, Lavras, 2002.

R DEVELOPMENT CORE TEAM. **R: a language and environment for statistical computing**. Version 2.14.2. Vienna: R Foundation for Statistical Computing, 2012. Disponível em: <<http://www.r-project.org>>. Acesso em: 10 mar. 2012.

REIS, A. J. dos S.; MIRANDA FILHO, J. B. de. Autocorrelação espacial na avaliação de compostos de milho para resistência à lagarta do cartucho (*Spodoptera frugiperda*). **Pesquisa Agropecuária Tropical**, Goiânia, v. 33, n. 2, p. 65-72, 2003.

RENCHEER, A. C.; SCHAALJE, G. B. **Linear models in statistics**. 2nd ed. Hoboken: Wiley-Interscience, 2008. 672 p.

RIBEIRO JÚNIOR, P. J. **Métodos geoestatísticos no estudo da variabilidade espacial de parâmetros do solo**. 1995. 99 f. Dissertação (Mestrado em Agronomia) - Escola Superior de Agricultura “Luiz de Queiroz”, Piracicaba, 1995.

RIBEIRO JÚNIOR, P. J.; DIGGLE, P. J. GeoR: a package for geoestatistical analysis. **R-News**, London, v. 1, n. 2, p. 15-18, June 2001.

ROSSONI, D. F. **Análise de variância para experimentos com dependência espacial**. 2011. 108 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) – Universidade Federal de Lavras , Lavras, 2011.

SCHLATHER, M. **Random Fields**: simulation and analysis of random fields. R package version 2.0.59. Disponível em: <<http://CRAN.R-project.org/package=RandomFields>>. Acesso em: 10 nov. 2012.

SEARLE, S. R. **Linear models**. New York: J. Wiley, 1971. 536 p.

_____. **Linear models for unbalanced data**. New York: J. Wiley, 1987. 536 p.

STROUP, W. W.; BAENZIGER, P. S.; MULITZE, D. K. Removing spatial variation from wheat yield trials: a comparison of methods. **Crop Science**, Madison, v. 34, n. 1, p. 62-66, Jan. 1994.

ZIMMERMAN, D. I.; HARVILLE, D. A. A random field approach to the anlysis of field-plot experiments and other spatial experiments. **Biometrics**, Raleigh, v. 47, n. 1, p. 223-239, 1991.

ANEXO A - Conjunto de dados

Tabela 27: Conjunto de dados usados no exemplo de aplicação.

| Obs | CoordX | CoordY | trat | bloco | y |
|-----|----------|---------|------|-------|-------------|
| 1 | 524822,3 | 7569842 | 1 | 1 | 2,72965762 |
| 2 | 524886,6 | 7569822 | 1 | 2 | 3,85774578 |
| 3 | 524932,7 | 7569867 | 1 | 3 | 4,88034591 |
| 4 | 524975,5 | 7569867 | 1 | 4 | 3,32146498 |
| 5 | 524831,9 | 7569891 | 2 | 1 | 2,61821296 |
| 6 | 524897,5 | 7569871 | 2 | 2 | 2,12708780 |
| 7 | 524929,1 | 7569917 | 2 | 3 | 3,73484167 |
| 8 | 524972,4 | 7569917 | 2 | 4 | 2,98415509 |
| 9 | 524807,7 | 7569844 | 3 | 1 | 2,06788330 |
| 10 | 524874,5 | 7569830 | 3 | 2 | 1,84389411 |
| 11 | 524952,8 | 7569867 | 3 | 3 | -0,06236613 |
| 12 | 524985,4 | 7569869 | 3 | 4 | 5,45928759 |
| 13 | 524816,9 | 7569893 | 4 | 1 | 2,63575713 |
| 14 | 524883,8 | 7569879 | 4 | 2 | 0,99627626 |
| 15 | 524949,2 | 7569917 | 4 | 3 | 2,88394902 |
| 16 | 524982,4 | 7569919 | 4 | 4 | 3,02898110 |
| 17 | 524834,8 | 7569841 | 5 | 1 | 2,27472516 |
| 18 | 524890,6 | 7569820 | 5 | 2 | 1,82149956 |
| 19 | 524942,7 | 7569867 | 5 | 3 | 1,08235752 |
| 20 | 524995,4 | 7569871 | 5 | 4 | 6,28655463 |
| 21 | 524844,0 | 7569889 | 6 | 1 | 4,88439380 |
| 22 | 524901,8 | 7569868 | 6 | 2 | 0,76569249 |
| 23 | 524939,1 | 7569917 | 6 | 3 | 2,20142496 |
| 24 | 524992,2 | 7569921 | 6 | 4 | 2,79772211 |
| 25 | 524824,8 | 7569842 | 7 | 1 | 2,44809917 |
| 26 | 524870,2 | 7569833 | 7 | 2 | 1,68835769 |
| 27 | 524945,3 | 7569867 | 7 | 3 | 3,82280712 |
| 28 | 524990,4 | 7569870 | 7 | 4 | 5,91716355 |
| 29 | 524834,4 | 7569891 | 8 | 1 | 3,72192438 |
| 30 | 524879,1 | 7569882 | 8 | 2 | 2,73725198 |
| 31 | 524941,6 | 7569917 | 8 | 3 | 2,52002699 |
| 32 | 524987,3 | 7569920 | 8 | 4 | 1,74483556 |
| 33 | 524812,6 | 7569844 | 9 | 1 | 1,57146954 |
| 34 | 524876,5 | 7569829 | 9 | 2 | 3,77942758 |
| 35 | 524937,7 | 7569867 | 9 | 3 | 3,15777696 |
| 36 | 524967,8 | 7569867 | 9 | 4 | 2,88581649 |
| 37 | 524821,9 | 7569892 | 10 | 1 | 1,62297704 |
| 38 | 524886,0 | 7569878 | 10 | 2 | 0,58095273 |

Tabela 27, continua

| Obs | CoordX | CoordY | trat | bloco | y |
|-----|----------|---------|------|-------|------------|
| 39 | 524934,1 | 7569917 | 10 | 3 | 3,93060176 |
| 40 | 524964,9 | 7569917 | 10 | 4 | 2,76032432 |
| 41 | 524829,9 | 7569842 | 11 | 1 | 1,23415962 |
| 42 | 524892,8 | 7569819 | 11 | 2 | 1,27569133 |
| 43 | 524940,2 | 7569867 | 11 | 3 | 0,64414488 |
| 44 | 524987,7 | 7569870 | 11 | 4 | 4,90664105 |
| 45 | 524839,4 | 7569890 | 12 | 1 | 4,9624938 |
| 46 | 524904,2 | 7569867 | 12 | 2 | 2,49568381 |
| 47 | 524936,6 | 7569917 | 12 | 3 | 2,69825241 |
| 48 | 524984,8 | 7569920 | 12 | 4 | 2,25309339 |
| 49 | 524810,3 | 7569844 | 13 | 1 | 0,61759147 |
| 50 | 524884,6 | 7569824 | 13 | 2 | 1,69706664 |
| 51 | 524930,2 | 7569868 | 13 | 3 | 4,24812604 |
| 52 | 524980,4 | 7569868 | 13 | 4 | 2,30390779 |
| 53 | 524819,5 | 7569892 | 14 | 1 | 1,06595803 |
| 54 | 524895,1 | 7569872 | 14 | 2 | 1,73003741 |
| 55 | 524926,4 | 7569917 | 14 | 3 | 3,6466667 |
| 56 | 524977,4 | 7569918 | 14 | 4 | 4,03946483 |
| 57 | 524827,3 | 7569842 | 15 | 1 | 1,66912363 |
| 58 | 524867,9 | 7569835 | 15 | 2 | 1,89804673 |
| 59 | 524955,4 | 7569867 | 15 | 3 | 1,68328249 |
| 60 | 524972,7 | 7569867 | 15 | 4 | 2,4872148 |
| 61 | 524836,9 | 7569890 | 16 | 1 | 4,42154672 |
| 62 | 524876,6 | 7569883 | 16 | 2 | 0,74670515 |
| 63 | 524951,6 | 7569917 | 16 | 3 | 1,32730755 |
| 64 | 524969,9 | 7569917 | 16 | 4 | 3,56429478 |
| 65 | 524815,1 | 7569843 | 17 | 1 | 3,00862428 |
| 66 | 524899,6 | 7569869 | 17 | 2 | 2,44022547 |
| 67 | 524927,7 | 7569868 | 17 | 3 | 3,52500662 |
| 68 | 524970,5 | 7569867 | 17 | 4 | 2,3721857 |
| 69 | 524824,5 | 7569892 | 18 | 1 | 2,92101937 |
| 70 | 524888,6 | 7569821 | 18 | 2 | 2,25920415 |
| 71 | 524924,0 | 7569918 | 18 | 3 | 3,0215446 |
| 72 | 524967,4 | 7569917 | 18 | 4 | 3,16494288 |
| 73 | 524805,3 | 7569844 | 19 | 1 | 1,26152386 |
| 74 | 524878,7 | 7569828 | 19 | 2 | 3,76978953 |
| 75 | 524935,2 | 7569867 | 19 | 3 | 2,45535037 |
| 76 | 524992,7 | 7569871 | 19 | 4 | 6,07665732 |
| 77 | 524814,5 | 7569893 | 20 | 1 | 0,30127024 |
| 78 | 524888,3 | 7569876 | 20 | 2 | 1,8400352 |
| 79 | 524931,5 | 7569917 | 20 | 3 | 5,09727619 |

Tabela 27, conclusão.

| Obs | CoordX | CoordY | trat | bloco | y |
|-----|----------|---------|------|-------|-------------|
| 80 | 524989,7 | 7569921 | 20 | 4 | 2,43358911 |
| 81 | 524820,1 | 7569843 | 21 | 1 | 2,32981675 |
| 82 | 524882,7 | 7569825 | 21 | 2 | 3,08488474 |
| 83 | 524950,4 | 7569867 | 21 | 3 | 0,52179681 |
| 84 | 524977,7 | 7569868 | 21 | 4 | 1,41315988 |
| 85 | 524829,5 | 7569891 | 22 | 1 | 2,98262907 |
| 86 | 524893,0 | 7569873 | 22 | 2 | 1,42669735 |
| 87 | 524946,6 | 7569917 | 22 | 3 | 3,22036924 |
| 88 | 524974,9 | 7569918 | 22 | 4 | 4,43438716 |
| 89 | 524832,4 | 7569841 | 23 | 1 | 2,92981601 |
| 90 | 524872,6 | 7569832 | 23 | 2 | 1,15087931 |
| 91 | 524947,8 | 7569867 | 23 | 3 | -0,16823921 |
| 92 | 524965,5 | 7569867 | 23 | 4 | 2,01154347 |
| 93 | 524841,9 | 7569890 | 24 | 1 | 5,84492159 |
| 94 | 524881,6 | 7569880 | 24 | 2 | 2,86438622 |
| 95 | 524944,1 | 7569917 | 24 | 3 | 2,64892336 |
| 96 | 524962,4 | 7569917 | 24 | 4 | 1,72124409 |
| 97 | 524817,6 | 7569843 | 25 | 1 | 1,73945554 |
| 98 | 524880,5 | 7569826 | 25 | 2 | 3,41730436 |
| 99 | 524925,2 | 7569868 | 25 | 3 | 2,95138179 |
| 100 | 524982,7 | 7569869 | 25 | 4 | 4,94352491 |
| 101 | 524826,9 | 7569892 | 26 | 1 | 3,17731669 |
| 102 | 524890,7 | 7569875 | 26 | 2 | 3,47098189 |
| 103 | 524921,4 | 7569918 | 26 | 3 | 4,0546775 |
| 104 | 524979,8 | 7569919 | 26 | 4 | 2,74822518 |

ANEXO B - Componentes do modelo

Tabela 28: Componentes do modelo espacial apresentado no exemplo de aplicação.

| Obs | Tendência | Espacial | Resíduos |
|-----|-----------|-------------|--------------|
| 1 | 3,851434 | -1,02313432 | -0,098641825 |
| 2 | 3,322108 | 0,49108276 | 0,044555399 |
| 3 | 3,939470 | 0,94800931 | -0,007133396 |
| 4 | 4,793174 | -1,5329293 | 0,061219822 |
| 5 | 2,029570 | 0,62446574 | -0,035822537 |
| 6 | 1,500244 | 0,44862356 | 0,178220616 |
| 7 | 2,117606 | 1,6349501 | -0,017714424 |
| 8 | 2,971310 | 0,13752829 | -0,124683655 |
| 9 | 2,390178 | -0,50672006 | 0,184425232 |
| 10 | 1,860852 | 0,08842287 | -0,105380751 |
| 11 | 2,478214 | -2,37065891 | -0,169921581 |
| 12 | 3,331919 | 2,03649166 | 0,0908771 |
| 13 | 3,130319 | -0,75861946 | 0,264058005 |
| 14 | 2,600992 | -1,30603002 | -0,29868617 |
| 15 | 3,218355 | -0,34433149 | 0,009925691 |
| 16 | 4,072059 | -1,06778066 | 0,024702474 |
| 17 | 2,033903 | 0,30016479 | -0,059342562 |
| 18 | 1,504577 | 0,29827039 | 0,018652377 |
| 19 | 2,121939 | -0,91859302 | -0,120988623 |
| 20 | 2,975644 | 3,1492322 | 0,161678808 |
| 21 | 1,916927 | 2,88727431 | 0,08019294 |
| 22 | 1,387600 | -0,41369422 | -0,208213703 |
| 23 | 2,004963 | 0,17497715 | 0,021485031 |
| 24 | 2,858667 | -0,16748087 | 0,106535732 |
| 25 | 3,511136 | -1,06739323 | 0,004355952 |
| 26 | 2,981810 | -1,12420233 | -0,169250285 |
| 27 | 3,599173 | -0,18442768 | 0,408062122 |
| 28 | 4,452877 | 1,7074542 | -0,243167789 |
| 29 | 2,509134 | 1,25556754 | -0,042776914 |
| 30 | 1,979808 | 0,54917986 | 0,20826451 |
| 31 | 2,597170 | -0,05069612 | -0,026446882 |
| 32 | 3,450874 | -1,56699817 | -0,139040714 |
| 33 | 3,107599 | -1,36284209 | -0,173287345 |
| 34 | 2,578273 | 1,12186184 | 0,079292912 |
| 35 | 3,195635 | -0,11874319 | 0,080884948 |
| 36 | 4,049340 | -1,17663266 | 0,013109485 |
| 37 | 2,391470 | -0,72348715 | -0,045005627 |
| 38 | 1,862144 | -1,167149 | -0,114041947 |

Tabela 28, continua

| Obs | Tendência | Espacial | Resíduos |
|-----|-----------|-------------|--------------|
| 39 | 2,479506 | 1,40486493 | 0,046230782 |
| 40 | 3,333211 | -0,68570299 | 0,112816793 |
| 41 | 1,557801 | -0,30670474 | -0,016936348 |
| 42 | 1,028475 | 0,22887606 | 0,018340694 |
| 43 | 1,645837 | -0,88160306 | -0,120089007 |
| 44 | 2,499541 | 2,28841498 | 0,11868466 |
| 45 | 2,492006 | 2,49348139 | -0,022993299 |
| 46 | 1,962680 | 0,39282336 | 0,140180872 |
| 47 | 2,580042 | 0,27708709 | -0,158876625 |
| 48 | 3,433746 | -1,22234207 | 0,041689052 |
| 49 | 1,999684 | -1,27269018 | -0,109402129 |
| 50 | 1,470358 | 0,36060141 | -0,133892412 |
| 51 | 2,087720 | 1,84013088 | 0,320275142 |
| 52 | 2,941424 | -0,56053609 | -0,076980601 |
| 53 | 2,508662 | -1,28430003 | -0,158403477 |
| 54 | 1,979335 | -0,2176819 | -0,031616089 |
| 55 | 2,596698 | 1,05064825 | -0,000679322 |
| 56 | 3,450402 | 0,39836371 | 0,190698888 |
| 57 | 2,987332 | -1,16970026 | -0,148508595 |
| 58 | 2,458006 | -0,62505981 | 0,065100191 |
| 59 | 3,075369 | -1,4575344 | 0,065448172 |
| 60 | 3,929073 | -1,45981861 | 0,017960232 |
| 61 | 2,111305 | 2,20406228 | 0,106179431 |
| 62 | 1,581979 | -0,63562107 | -0,199652653 |
| 63 | 2,199341 | -0,84011821 | -0,031915485 |
| 64 | 3,053046 | 0,38586037 | 0,125388707 |
| 65 | 3,018049 | -0,24855358 | 0,239129136 |
| 66 | 2,488723 | -0,04287814 | -0,005618972 |
| 67 | 3,106085 | 0,5731527 | -0,154231029 |
| 68 | 3,959789 | -1,50832458 | -0,079279135 |
| 69 | 2,517736 | 0,21865528 | 0,184628516 |
| 70 | 1,988409 | 0,30412924 | -0,033334523 |
| 71 | 2,605772 | 0,52671814 | -0,110945346 |
| 72 | 3,459476 | -0,25418474 | -0,040348647 |
| 73 | 2,394062 | -1,0150329 | -0,117505628 |
| 74 | 1,864736 | 1,74517908 | 0,159874206 |
| 75 | 2,482099 | 0,08945643 | -0,116204676 |
| 76 | 3,335803 | 2,66701814 | 0,073836098 |
| 77 | 2,485123 | -1,9391421 | -0,244711041 |
| 78 | 1,955797 | -0,14659151 | 0,030829463 |
| 79 | 2,573160 | 2,28314805 | 0,240968522 |

Tabela 28, conclusão.

| Obs | Tendência | Espacial | Resíduos |
|-----|-----------|-------------|--------------|
| 80 | 3,426864 | -0,96618802 | -0,027086944 |
| 81 | 2,582984 | -0,40841446 | 0,155247171 |
| 82 | 2,053658 | 0,97872205 | 0,052504788 |
| 83 | 2,671020 | -2,17523172 | 0,026008257 |
| 84 | 3,524725 | -1,87780464 | -0,233760216 |
| 85 | 2,779478 | 0,23298067 | -0,029829172 |
| 86 | 2,250151 | -0,60813684 | -0,215317241 |
| 87 | 2,867514 | 0,20740201 | 0,145453432 |
| 88 | 3,721218 | 0,61347591 | 0,099692981 |
| 89 | 2,050101 | 0,68066527 | 0,199049927 |
| 90 | 1,520775 | -0,4031131 | 0,033217739 |
| 91 | 2,138137 | -2,03710806 | -0,269268197 |
| 92 | 2,991842 | -1,01729857 | 0,037000531 |
| 93 | 2,845095 | 2,92301263 | 0,076814035 |
| 94 | 2,315769 | 0,3078128 | 0,240804631 |
| 95 | 2,933131 | -0,18993259 | -0,094275207 |
| 96 | 3,786836 | -1,84224807 | -0,223343459 |
| 97 | 2,627035 | -0,75309951 | -0,134479849 |
| 98 | 2,097709 | 1,32441261 | -0,004817011 |
| 99 | 2,715071 | 0,26274309 | -0,02643243 |
| 100 | 3,568776 | 1,20902003 | 0,16572929 |
| 101 | 3,207282 | 0,02646902 | -0,056433997 |
| 102 | 2,677956 | 0,54304301 | 0,249983357 |
| 103 | 3,295318 | 0,69897947 | 0,060380132 |
| 104 | 4,149022 | -1,14686769 | -0,253929493 |

ANEXO C - Rotinas em R

```
#####
##### ANÁLISE DE UM DELINEAMENTO INTEIRAMENTE CASUALIZADO #####
#####

### Pacotes utilizados ###
require(geoR)
require(fBasics)
require(RandomFields)

dados<-read.table("Exemplo.txt",sep=";", header=TRUE)

t=4 # número de tratamentos
r=3 # número de repetições

coords<-cbind(dados$CoordX, dados$CoordY)

#####
## exemplo do arquivo de entrada ##
## de dados de um DIC com t=4 e r=3 ##
## ##
## > dados ##
## CoordX CoordY trat y ##
## 1 0 0 1 16.92792 ##
## 2 1 0 1 17.66208 ##
## 3 2 0 1 17.57933 ##
## 4 0 1 2 19.58477 ##
## 5 1 1 2 22.44035 ##
## 6 2 1 2 22.65466 ##
## 7 0 2 3 16.94244 ##
## 8 1 2 3 17.68582 ##
## 9 2 2 3 17.00374 ##
## 10 0 3 4 18.49544 ##
## 11 1 3 4 18.46935 ##
## 12 2 3 4 19.94619 ##
#####

### Construindo a matriz X ###
X1<-matrix(c(rep(1,r*t)),ncol=1)
X2<-kron(diag(t),c(rep(1,r)))
X<-cbind(X1,X2)
```

```

### Calculando o erro ###
Y<-c(dados$y)
theta<-ginv(t(X)%*%X)%*%t(X)%*%Y
erro<- data.frame(Y- X)%*%theta)

### Análise Geoestatística ###
geodados <- as.geodata(data.frame(coords,erro),coords.col = 1:2,
                        data.col = 3)

geodados
h_max<- summary(geodados)[[3]][[2]]
dist<-0.5*h_max+2
vario.res <- variof(geodados,max.dist=dist)

### Envelope ###
geo.env <- variof.mc.env(geodados, obj.var = vario.res)
plot(vario.res, envelope = geo.env)

### Ajustando o modelo teórico ###
ols<-variofit(vario.res,cov.model="sph",weights="equal",
              max.dist=dist)
plot(vario.res)
lines(ols, col="blue", lty=1, lwd=2)

phi<-summary(ols)[[10]][[3]]      # alcance
sill<-summary(ols)[[10]][[2]]    # contribuição
nugget<-summary(ols)[[10]][[1]]  # ef. pepita

### Obtendo a matriz de covariâncias ###
sigma<-varcov.spatial(coords = coords, cov.model = "sph", nugget =
                       nugget, kappa = 0.5,cov.pars = c(sill,phi))$varcov

V<- sigma*(1/(as.numeric(nugget+sill)))
i.V<-solve(V)
theta_dep<-ginv(t(X)%*%i.V)%*%X)%*%t(X)%*%i.V)%*%Y
erro<- data.frame(Y- X)%*%theta_dep)

### Iterações ###
repeat
{

```

```

geodados <- as.geodata(data.frame(coords,erro),coords.col = 1:2,
                        data.col = 3)
vario.res <- variog(geodados,max.dist=dist)
ols<-variofit(vario.res, cov.model="sph", weights="equal",
              max.dist=dist)

phi1<-summary(ols)[[10]][[3]]
sill1<-summary(ols)[[10]][[2]]
nugget1<-summary(ols)[[10]][[1]]

sigma<-varcov.spatial(coords = coords, cov.model = "sph", nugget =
                      nugget1, kappa = 0.5, cov.pars = c(sill1,phi1))$varcov

i.V<-solve(sigma*(1/(as.numeric(nugget1+sill1))))

theta_dep<-ginv(t(X)%*%i.V%*%X)%*%t(X)%*%i.V%*%Y
erro<- data.frame(Y- X%*%theta_dep)

e1=phi-phi1
e2=sill-sill1
e3=nugget-nugget1

phi=phi1
sill=sill1
nugget=nugget1
if ((e1<0.0001) && (e2<0.0001) && (e3<0.0001)) break
}

### Obtendo as somas de quadrados ###
P1<-i.V%*%X1%*%solve(t(X1)%*%i.V%*%X1)%*%t(X1)%*%i.V
P<-i.V%*%X%*%ginv(t(X)%*%i.V%*%X)%*%t(X)%*%i.V

SQT<-t(Y)%*%(i.V-P1)%*%Y
SQRes<-t(Y)%*%(i.V-P)%*%Y
SQTrat<- t(Y)%*%(P-P1)%*%Y

### Soma de quadrados de tratamento via modelos reduzidos ###
R<-i.V - P1
SQtrat<-t(Y)%*%R%*%X2%*%ginv(t(X2)%*%R%*%X2)%*%t(X2)%*%R%*%Y

## Graus de liberdade ##
glt<-rk(i.V-P1)

```

```

glres<-rk(i.V-P)
gltrat<-rk(P-P1)

### Construindo a análise de variância ###

QMTrat<-SQTrat/gltrat
QMRes<-SQRes/glres

F.calc<-QMTrat/QMRes
pval<- 1-pf(F.calc,gltrat,glres)

anova<-data.frame(FV = c("Tratamento","Residuo","Total"),
  GL = c(gltrat,glres,glt), SQ=round(c(SQTrat,SQRes,SQT),3),
  QM = c(as.character(round(c(QMTrat,QMRes),3))," "),
  F = c(as.character(round(F.calc,2)), " ", " "),
  Pvalue = c(as.character(pval), " ", " "))

anova

### Análise de variância supondo erros independentes ###
reg<-lm(y~factor(trat),dados)
anova(reg)

### Obtendo os componentes do modelo espacial ###
theta_dep<-ginv(t(X)%*%i.V%*%X)%*%t(X)%*%i.V%*%Y
trend<-X%*%theta_dep
erro<-Y-X%*%theta_dep

sig0<-varcov.spatial(coords, cov.model = "sph", nugget = 0,
  cov.pars = c(sill, phi))$varcov
sigma<-varcov.spatial(coords, cov.model = "sph", nugget = nugget,
  cov.pars = c(sill, phi))$varcov
i.sig<-solve(sigma)

spatial.comp <- sig0%*%i.sig%*%erro
residuals<-erro - spatial.comp

comp.model<-cbind(trend,spatial.comp, residuals)

```

```
#####
##### ANÁLISE DE UM DELINEAMENTO EM BLOCOS CASUALIZADOS #####
#####
```

```
dados<-read.table("Exemplo.txt",sep=";", header=TRUE)
```

```
t=4 # número de tratamentos
r=3 # número de blocos
```

```
coords<-cbind(dados$CoordX, dados$CoordY)
```

```
#####
##      exemplo do arquivo de entrada      ##
##      de dados de um DBC com t=4 e r=3      ##
##      ##                                     ##
##      CoordX CoordY trat bloco           y      ##
##  1         0         0     1     1 24.47424  ##
##  2         1         0     1     2 19.82192  ##
##  3         2         0     1     3 20.56174  ##
##  4         0         1     2     1 25.91073  ##
##  5         1         1     2     2 22.77189  ##
##  6         2         1     2     3 23.97054  ##
##  7         0         2     3     1 26.25216  ##
##  8         1         2     3     2 25.88862  ##
##  9         2         2     3     3 26.42087  ##
## 10         0         3     4     1 29.91367  ##
## 11         1         3     4     2 26.52883  ##
## 12         2         3     4     3 26.53225  ##
#####
```

```
### Construindo a matriz X ###
X1<-matrix(c(rep(1,r*t)),ncol=1)
X2<-kron(diag(t),c(rep(1,r)))
X3<-kron(c(rep(1,t)),diag(r))
X<-cbind(X1,X2,X3)
```

```
### Calculando o erro (análise clássica) ###
Y<-c(dados$y)
theta<-ginv(t(X)%*%X)%*%t(X)%*%Y
erro<- data.frame(Y- X%*%theta)
```

```

### Análise Geoestatística ###
geodados <- as.geodata(data.frame(coords,erro),coords.col = 1:2,
                        data.col = 3)

geodados
h_max<- summary(geodados)[[3]][[2]]
dist<-0.5*h_max+2
vario.res <- variog(geodados)

### Envelope ###
geo.env <- variog.mc.env(geodados, obj.var = vario.res)
plot(vario.res, envelope = geo.env)

### Ajustando o modelo teórico ###
plot(vario.res, xlim=c(0,dist))
ols<- variofit(vario.res,cov.model="sph", max.dist=dist)
lines(ols, col="red", lty=1, lwd=2)
phi<-summary(ols)[[10]][[3]]      # alcance
sill<-summary(ols)[[10]][[2]]    # contribuicao
nugget<-summary(ols)[[10]][[1]]  # ef. pepita

### Obtendo a matriz de covariancias ###
sigma<-varcov.spatial(coords = coords, cov.model = "sph", nugget =
                       nugget, kappa = 0.5, cov.pars = c(sill,phi))$varcov

V<- sigma*(1/(as.numeric(sill+nugget)))
i.V<-solve(V)

### Iterações ###
repeat
{
geodados <- as.geodata(data.frame(coords,erro),coords.col = 1:2,
                        data.col = 3)
vario.res <- variog(geodados,max.dist=dist)
ols<-variofit(vario.res, cov.model="sph", weights="equal",
              max.dist=dist)

phi1<-summary(ols)[[10]][[3]]
sill1<-summary(ols)[[10]][[2]]
nugget1<-summary(ols)[[10]][[1]]

```

```

sigma<-varcov.spatial(coords = coords, cov.model = "sph", nugget =
      nugget1, kappa = 0.5, cov.pars = c(sill1,phi1))$varcov

i.V<-solve(sigma*(1/(as.numeric(nugget1+sill1))))

theta_dep<-ginv(t(X)%*%i.V)%*%X)%*%t(X)%*%i.V)%*%Y
erro<- data.frame(Y- X)%*%theta_dep)

e1=phi-phi1
e2=sill-sill1
e3=nugget-nugget1

phi=phi1
sill=sill1
nugget=nugget1
if ((e1<0.0001) && (e2<0.0001) && (e3<0.0001)) break
}

### Somas de quadrados ###
P1<-i.V)%*%X1)%*%solve(t(X1)%*%i.V)%*%X1)%*%t(X1)%*%i.V
P<-i.V)%*%X)%*%ginv(t(X)%*%i.V)%*%X)%*%t(X)%*%i.V
R<-i.V - P1
P2<-R)%*%X2)%*%ginv(t(X2)%*%R)%*%X2)%*%t(X2)%*%R

SQtrat<-t(Y)%*%P2)%*%Y
SQRes<-t(Y)%*%(i.V-P)%*%Y
SQT<-t(Y)%*%(i.V - P1)%*%Y
SQBloco<-t(Y)%*%(P-P1-P2)%*%Y

## Graus de liberdade ##
gl<-rk(i.V-P1)
glres<-rk(i.V-P)
gltrat<-rk(P2)
glbloco<-rk(P-P1-P2)

## Construindo a análise de variância
QMTrat<-SQtrat/gltrat
QMRes<-SQRes/glres
QMBloco<-SQBloco/glbloco

F.calc<-QMTrat/QMRes
F.calc2<-QMBloco/QMRes

```



```

pval<- 1-pf(F.calc,gltrat,glres)
pval2<- 1-pf(F.calc2,glbloco,glres)

anova<-data.frame(FV = c("Tratamento","Bloco","Residuo","Total"),
  GL = c(gltrat,glbloco,glres,gltr),
  SQ=round(c(SQtrat,SQBloco,SQRes,SQT),3),
  QM = c(as.character(round(c(QMTrat,QMBloco,QMRes),3))," "),
  F = c(as.character(round(c(F.calc,F.calc2),3)), " ", " "),
  Pvalue = c(as.character(round(c(pval,pval2),3)), " ", " "))

anova

### Análise de variância supondo erros independentes ###
reg<-lm(y~factor(trat)+factor(bloco),dados)
anova(reg)

### Obtendo os componentes do modelo espacial ###
theta_dep<-ginv(t(X)%*%i.V%*%X)%*%t(X)%*%i.V%*%Y
trend<-X%*%theta_dep
erro<-Y-X%*%theta_dep

sig0<-varcov.spatial(coords, cov.model = "gaus", nugget = 0,
  cov.pars = c(sill, phi))$varcov
sigma<-varcov.spatial(coords, cov.model = "gaus", nugget = nugget,
  cov.pars = c(sill, phi))$varcov
i.sig<-solve(sigma)

spatial.comp <- sig0%*%i.sig%*%erro
residuals<-erro - spatial.comp

comp.model<-cbind(trend,spatial.comp, residuals)

```

```
#####
#####          Simulação de dados          #####
#####

### Simulação de um DIC com 15 trat. e 8 rep. ###

m=500 # número de simulações

result<-data.frame(F.dep=rep(0,m),F.ind=rep(0,m),porc.F=rep(0,m),
                   phi=rep(0,m),sill=rep(0,m),nugget=rep(0,m))

t=15 # numero de tratamentos
r=8  # numero de repetições
v=15 # eixo y
u=8  # eixo x
model<-"sph"
eftrat<-c(7, 10, 5, 5, 8, 10, 9, 5, 10, 6, 7, 7, 6, 6, 9)
coords<- expand.grid(x= seq(0,(u-1),1),y=seq(0,(v-1),1))

X1<-matrix(c(rep(1,r*t)),ncol=1)
X2<-kron(diag(t),c(rep(1,r)))
X<-cbind(X1,X2)

for(k in 1:m)
{
  dados<-data.frame(y=10, trat=rep(1:t, each=r))
  for (i in 1:t)
  dados$y[dados$trat==i]<-
  dados$y[dados$trat==i] + eftrat[i]

  ## Gerando erros dependentes
  dep<-GaussRF(coords,model=model, param=c(0,2.5,0,6))
  dados$y<-dados$y + dep ## adicionando os erros aos efeitos fixos

  Y<-c(dados$y)
  theta<-ginv(t(X)%*%X)%*%t(X)%*%Y
  erro<- data.frame(Y- X)%*%theta)

  ### Análise geoestatística ###
  geodados <- as.geodata(data.frame(coords,erro),coords.col = 1:2,
                        data.col = 3)
  h_max<- summary(geodados)[[3]][[2]]
  dist<-0.5*h_max+2
}
```

```

vario.res <- variog(geodados,max.dist=dist)
ols<-variofit(vario.res,cov.model=model,weights="equal",
              max.dist=dist)
phi<-summary(ols)[[10]][[3]]
sill<-summary(ols)[[10]][[2]]
nugget<-summary(ols)[[10]][[1]]

sigma<-varcov.spatial(coords = coords, cov.model = model, nugget =
                      nugget,kappa = 0.5,cov.pars = c(sill,phi))$varcov

V<- sigma*(1/(as.numeric(nugget+sill)))
i.V<-solve(V)
theta_dep<-ginv(t(X)%*%i.V%*%X)%*%t(X)%*%i.V%*%Y
erro<- data.frame(Y- X%*%theta_dep)

### Iterações ###
repeat
{
geodados <- as.geodata(data.frame(coords,erro),coords.col = 1:2,
                      data.col = 3)
vario.res <- variog(geodados,max.dist=dist)
ols<-variofit(vario.res, cov.model="sph", weights="equal",
              max.dist=dist)

phi1<-summary(ols)[[10]][[3]]
sill1<-summary(ols)[[10]][[2]]
nugget1<-summary(ols)[[10]][[1]]

sigma<-varcov.spatial(coords = coords, cov.model = "sph", nugget =
                      nugget1, kappa = 0.5, cov.pars = c(sill1,phi1))$varcov

i.V<-solve(sigma*(1/(as.numeric(nugget1+sill1))))
theta_dep<-ginv(t(X)%*%i.V%*%X)%*%t(X)%*%i.V%*%Y
erro<- data.frame(Y- X%*%theta_dep)

e1=phi-phi1
e2=sill-sill1
e3=nugget-nugget1

phi=phi1
sill=sill1
nugget=nugget1
if ((e1<0.0001) && (e2<0.0001) && (e3<0.0001)) break
}

```

```

### Análise considerando erros dependentes ###
P1<-i.V%*%X1%*%solve(t(X1)%*%i.V%*%X1)%*%t(X1)%*%i.V
P<-i.V%*%X%*%ginv(t(X)%*%i.V%*%X)%*%t(X)%*%i.V
QMTrat_dep<-t(Y)%*(P-P1)%*%Y/rk(P-P1)
QMRes_dep<-t(Y)%*(i.V-P)%*%Y/rk(i.V-P)
F.dep<-QMTrat_dep/QMRes_dep

### Análise considerando erros independentes ###
reg<-lm(y~factor(trat),dados)
QMTrat_ind<-anova(reg)[[3]][1]
QMRes_ind<-anova(reg)[[3]][2]
F.ind<-anova(reg)[[4]][1]

### Porcentagem de aumento do valor F ao considerar dependência ###
porc.F<-as.numeric(100*F.dep/F.ind -100)

result[k,]<-c(F.dep,F.ind,porc.F,phi,sill,nugget)
}

result

```