

Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”

Estimação de contrastes de médias de tratamentos, de um experimento em blocos ao acaso, utilizando as análises clássica e espacial

Marina Rodrigues Maestre

Dissertação apresentada para obtenção do título de Mestre em Agronomia. Área de concentração: Estatística e Experimentação Agronômica

Piracicaba

2008

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

Marina Rodrigues Maestre
Bacharel em Estatística

Estimação de contrastes de médias de tratamentos, de um experimento em blocos ao acaso, utilizando as análises clássica e espacial

Orientador:

Prof. Dr. **DÉCIO BARBIN**

Dissertação apresentada para obtenção do título de Mestre em Agronomia. Área de concentração: Estatística e Experimentação Agronômica

Piracicaba

2008

**Dados Internacionais de Catalogação na Publicação (CIP)
DIVISÃO DE BIBLIOTECA E DOCUMENTAÇÃO - ESALQ/USP**

Maestre, Marina Rodrigues

Estimação de contrastes de médias de tratamentos, de um experimento em blocos ao acaso, utilizando as análises clássica e espacial / Marina Rodrigues Maestre. - - Piracicaba, 2008.

70 p. : il.

Dissertação (Mestrado) - - Escola Superior de Agricultura Luiz de Queiroz, 2008.
Bibliografia.

1. Estatísticas espaciais 2. Geoestatística 3. Mínimos quadrados 4. Modelagem de dados
5. Modelos matemáticos 6. Verossimilhança I. Título

CDD 519.5
M186e

“Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor”

DEDICATÓRIA

A

Deus

pela Presença e Força constantes em minha vida.

Aos meus pais,

Edilberto Rodrigues Maestre e

Angélica Rebola Rodrigues,

pelo amor incondicional,

pela confiança depositada e

por serem meu alicerce em tudo o que eu faço.

Aos meus sobrinhos,

Carlos Eduardo e Mariana.

Ao meu namorado, Alexsandro.

AGRADECIMENTOS

A Deus, pela vida, pelas bênçãos e pelas graças.

Aos meus pais, Edilberto Rodrigues Maestre e Angélica Rebola Rodrigues, que são a base de tudo na minha vida, ao apoio, incentivo, carinho e força.

Ao Prof. Dr. Décio Barbin, pela orientação, amizade, apoio, e crédito depositado neste trabalho.

Ao Prof. Dr. Paulo Justiniano Ribeiro Junior, do Departamento de Estatística da Universidade Federal do Paraná, aos ensinamentos, sugestões, apoio e amizade.

Ao Prof. Dr. Sérgio Oliveira Moraes pelo auxílio e atenção ao conseguir os dados para a realização deste trabalho.

Ao Prof. Dr. Cláudio Roberto Marciano, do Laboratório de Solos do Centro de Ciências e Tecnologias Agropecuárias da Universidade Estadual do Norte Fluminense Darcy Ribeiro, por disponibilizar os dados deste trabalho.

Ao Prof. Dr. Fernando Carvalho Oliveira, da Biossola Agricultura & Ambiente Ltda e do curso de Engenharia Ambiental da Faculdade Municipal Professor Franco Montoro de Mogi Guaçu, SP, por ser o mentor dos dados deste trabalho.

Aos professores do Departamento de Ciências Exatas da ESALQ/USP, pelos conhecimentos transmitidos e agradável convivência.

Aos funcionários do Departamento de Ciências Exatas da ESALQ/USP, que contribuíram para que o tempo de duração dos estudos fosse o melhor possível, pela atenção e dedicação.

Aos amigos do mestrado e doutorado, em especial, Júlio César Pereira, Lucimary Afonso dos Santos, Lúcio Borges de Araújo, Mirian Fernandes Carvalho Araújo, Renata Alcarde, Simone Daniela Sartório, Vanderly Janeiro e Wilson Alves de Oliveira, pela ajuda, pela amizade e pelos desabafos. E a todos os alunos do departamento que foram de suma importância para a boa convivência em Piracicaba.

Aos amigos muito especiais que fiz morando na Vila Estudantil da Pós-Graduação da ESALQ/USP, principalmente à Andreza, Deise, Edjane, Elenilson, Elton, Julieth, Luciélio, Maristela, Michele, Shirley e Wanderlei, pela amizade.

Às amigas que desde sempre são minhas companheiras, Caroline Sandi, Gilvana Rodrigues Maestre, Lidiane Vicente dos Anjos, Nayara Corrêa Farias, Patrícia Hobold Meurer e Roberta Pinto Dias.

Ao meu namorado, Alexsandro Cláudio dos Santos Almeida, pelo companheirismo e por estar sempre ao meu lado me incentivando e não me deixando desanimar nas horas mais difíceis.

À Coordenadoria para o Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela bolsa de estudos.

E a todos que, de forma direta ou indireta, contribuíram para a realização deste trabalho.

SUMÁRIO

RESUMO	8
ABSTRACT	9
LISTA DE FIGURAS	10
LISTA DE TABELAS	11
1 INTRODUÇÃO	12
2 DESENVOLVIMENTO	15
2.1 Revisão de Literatura	15
2.1.1 Modelagem de dados experimentais	15
2.1.2 Modelos Geoestatísticos	17
2.1.3 Métodos para estimação dos parâmetros do modelo Geoestatístico	20
2.1.3.1 Variograma seguido de Mínimos Quadrados Generalizados	20
2.1.3.2 Método da Máxima Verossimilhança	25
2.1.4 Estimação dos parâmetros do modelo de Papadakis	26
2.2 Material e Métodos	27
2.2.1 Material	27
2.2.2 Métodos	28
2.2.2.1 Análise de Variância Clássica (ANOVA)	28
2.2.2.2 Mínimos Quadrados Generalizados (MQG)	29
2.2.2.3 Estimação por Máxima Verossimilhança	31
2.2.2.4 Método de Papadakis	32
2.2.2.5 Medidas para comparar os métodos de estimação de contrastes entre tratamentos	34
2.2.2.5.1 Diferença absoluta	35
2.2.2.5.2 Medida de dispersão	35
2.2.2.5.3 Erros quadráticos	35
2.2.2.5.4 Intervalo de confiança	35
2.3 Resultados e Discussão	36
2.3.1 Efeitos dos Parâmetros	36
2.3.2 Análise da Variância	36
2.3.3 Testes de dependência espacial	37
2.3.3.1 Modelo Geoestatístico	37
2.3.3.2 Modelo de Papadakis	38

2.3.4 Estimativas dos Parâmetros	38
2.3.5 Estimativas dos Contrastes de Médias dos Tratamentos e Medida de Dispersão	40
2.3.6 Diferenças Absolutas	41
2.3.7 Estimativas dos Erros Quadráticos das estimativas dos Contrastes	42
2.3.8 Intervalos de Confiança	43
3 CONCLUSÕES	46
REFERÊNCIAS	47
APÊNDICES	49
ANEXOS	65

RESUMO

Estimação de contrastes de médias de tratamentos, de um experimento em blocos ao acaso, utilizando as análises clássica e espacial

Em um experimento, é comum ocorrerem fatores não controláveis, responsáveis pela heterogeneidade entre as parcelas. Mesmo executando os três princípios básicos da experimentação no planejamento (repetição, casualização e controle local), ainda assim, pode haver correlação nos erros e, portanto, dependência espacial na área estudada. Se for detectada essa estrutura de autocorrelação e se essa informação for utilizada na análise estatística, estimativas mais eficientes dos contrastes entre as médias dos tratamentos são garantidas, mas se tal estrutura for desconsiderada pode impedir que diferenças reais sejam detectadas. Neste trabalho, foram observadas as coordenadas dos centros das parcelas de um delineamento em blocos ao acaso. A variável resposta, deste experimento, é a concentração de carbono orgânico no solo, sendo as avaliações feitas no início do experimento, ou seja, antes da aplicação dos tratamentos, portanto, um ensaio em branco, um ano após a aplicação dos tratamentos e, novamente, depois de mais um ano. Para tanto, foram utilizadas as análises clássica e espacial na comparação dos métodos de estimação de contrastes de médias de tratamentos. O método estudado para a análise clássica, em que considera que os erros são não correlacionados, foi o dos mínimos quadrados ordinários. Já para a análise, levando em consideração a dependência espacial, foram utilizados o modelo geoestatístico, em que consiste na adição de um efeito aleatório com correlação, e o modelo de Papadakis, que consiste na adição de uma covariável construída a partir de observações em parcelas vizinhas. No modelo geoestatístico foi verificada a presença da dependência espacial através dos critérios de informação de Akaike e de informação Bayesiano ou de Schwarz e os métodos testados foram o do variograma seguido de mínimos quadrados generalizados e o da máxima verossimilhança. Para o modelo de Papadakis, foi testada a significância da covariável referente às médias dos resíduos entre as parcelas vizinhas e a própria parcela tanto no modelo em blocos ao acaso quanto no modelo inteiramente casualizado, e o teste não foi significativo em nenhum dos dois casos. Mesmo assim, os cálculos foram realizados para esse método, mostrando que para esse conjunto de dados, este método não é indicado. Fazendo uso de algumas medidas de comparação desses métodos, para os dados em questão, o método de estimação dos contrastes de médias de tratamentos que apresentou as medidas de comparação mais dispersas foi o do modelo de Papadakis e o menos disperso foi o da máxima verossimilhança. Ainda, pelos intervalos de confiança, observou-se que na análise espacial, outros contrastes diferiram de zero significativamente, além daqueles que foram observados na análise clássica, o que se conclui que quando é levada em consideração a autocorrelação dos erros, os contrastes são estimados com maior eficiência.

Palavras-chave: Mínimos Quadrados Ordinários; Variograma; Mínimos Quadrados Generalizados; Máxima Verossimilhança; Papadakis

ABSTRACT

Estimation of treatments means contrasts, in a random blocks model, using the classical and spatial analysis

Not controllable factors is common occur in experiments, they are responsible for the heterogeneity among parcels. Even executing the three experimentation basic principles in the design (repetition, randomization and local control), even so, may have correlation in errors and, therefore, spatial dependence in the area of study. If that autocorrelation structure is detected and if this information is used in statistical analysis, estimates more efficient of contrasts among treatments means are guaranteed, but if this structure is disregarded can prevent that real differences are detected. In this work, the coordinates of parcels centers in a design of random blocks were observed. The concentration of soil organic carbon is the response variable of this experiment, with the available made at the beginning of the experiment, ie, before the treatments application, therefore, a blank, a year after the treatments application and, again, after a year. Then, the classical and spatial analysis were used to compare the methods of estimation of treatments means contrasts. The method studied for the classical analysis, which considers that the errors are not correlated, was the ordinary least squares. For the analysis, considering the spatial dependence, were used the geostatistical model, where consists in the addition of a random effect with correlation, and the Papadakis model, which consists in the addition of a covariate built from observations in neighbouring. In geostatistical model was verified the spatial dependence through the Akaike and Bayesian or Schwarz criteria of information and the methods tested were the variogram followed by generalized least squares and the maximum likelihood. For the Papadakis model, was tested the significance of covariate referring to the average of residuals among neighbouring parcels and own parcel in the random blocks model and in the completely randomized model, and the test was not significant in any of both cases. Still, the calculus were made for this method, showing that for this data set, this method is not indicated. Using some measures to compare these methods, for these data, the method of estimation of treatments means contrasts which presented the measures of comparison more dispersed was the Papadakis model and the less dispersed was the maximum likelihood. Still, in the confidence intervals, it was observed that in spatial analysis other contrasts differed from zero significantly, besides of those which were observed in classical analysis, which concludes that when the autocorrelation of errors is considering, the contrasts are estimated with greater efficiency.

Keywords: Ordinary Least Squares; Variogram; Generalized Least Squares; Maximum Likelihood; Papadakis

LISTA DE FIGURAS

Figura 1 - Semivariograma Empírico Hipotético	23
Figura 2 - Croqui do experimento	27
Figura 3 - Semivariograma Empírico para os dados do carbono orgânico e a curva exponencial ajustada aos parâmetros espaciais	40
Figura 4 - Gráfico da transformação Box-Cox	68

LISTA DE TABELAS

Tabela 1 - Efeitos dos Parâmetros	36
Tabela 2 - Análise da Variância (ANOVA) Clássica	37
Tabela 3 - Critérios de Informação de Akaike e Bayesiano ou de Schwarz para o modelo geostatístico	37
Tabela 4 - Resumo das estimativas dos parâmetros para os métodos em estudo	39
Tabela 5 - Estimativas dos parâmetros para o método de Papadakis ($\widehat{\beta}_{PP}$), considerando o modelo inteiramente ao acaso	39
Tabela 6 - Valores verdadeiros dos contrastes (c_j), suas estimativas (\widehat{c}_j) para os métodos em estudo e medidas de dispersão (D)	41
Tabela 7 - Estimativas dos contrastes ($\widehat{\Lambda\beta}_{PP}$), e medida de dispersão (D) para o método de Papadakis considerando o modelo inteiramente ao acaso	41
Tabela 8 - Diferenças absolutas entre os contrastes estimados e os seus verdadeiros valores .	42
Tabela 9 - Diferenças absolutas para o método de Papadakis considerando o modelo inteira- mente ao acaso	42
Tabela 10 - Estimativas dos erros quadráticos dos contrastes para os métodos em estudo . . .	43
Tabela 11 - Estimativas dos erros quadráticos dos contrastes para o método de Papadakis ($\widehat{e}q_{PP}$), considerando o modelo inteiramente ao acaso	43
Tabela 12 - Intervalos de confiança para contrastes entre médias de tratamentos	44
Tabela 13 - Intervalos de confiança dos contrastes para o método de Papadakis considerando o modelo inteiramente ao acaso	45
Tabela 14 - Dados de concentrações de Carbono Orgânico no Solo (Blocos 4 e 3)	66
Tabela 15 - Dados de concentrações de Carbono Orgânico no Solo (Blocos 2 e 1)	67

1 INTRODUÇÃO

Fatores não controláveis em um experimento são responsáveis pelas variações casuais no mesmo. Essas variações são os conhecidos erros experimentais que causam heterogeneidade entre as parcelas. Quanto menor o erro experimental, melhor será a precisão dos testes estatísticos para comparações de médias.

A execução de um planejamento clássico, segundo Fisher (1935), baseia-se nos três princípios básicos da experimentação, que são repetição, casualização e controle local.

O princípio da repetição consiste em várias parcelas receberem o mesmo tratamento para garantir a eficiência deste. Pilon (2004) afirma que o uso de um número apropriado de repetições ajuda a aumentar a precisão das estimativas dos parâmetros e, também aumenta o poder dos testes estatísticos.

A casualização, segundo Barbin (2003), consiste em se distribuírem os tratamentos pelas parcelas através de sorteio. Com essa aleatorização, todas as unidades experimentais têm a mesma chance de receber determinado tratamento. Isso garante, até certo ponto, a independência dos erros.

Já o controle local refere-se a restringir a casualização dos tratamentos a grupos de unidades experimentais com pouca variabilidade entre si (NOGUEIRA, 2007). Estes grupos de parcelas homogêneas entre si, também são conhecidos como blocos.

Executando esses três princípios básicos, assume-se que os erros são independentes. Como a independência dos erros pode reduzir a correlação espacial entre as parcelas adjacentes (GRONDONA; CRESSIE, 1991), as posições das unidades experimentais no mapa de campo são, normalmente, ignoradas na abordagem clássica.

Entretanto, há vários problemas espaciais em muitos estudos estatísticos, como no caso em que o experimento é somente observável ao invés de planejado (ensaio em branco). Nesses ensaios, é impossível a realização dos procedimentos adequados a uma análise clássica. Assim, são encontrados muitos experimentos em branco com presença de dependência espacial, violando a suposição de independência dos erros, ou seja, os erros são correlacionados.

Os modelos que envolvem estrutura de correlação entre classes e estrutura de correlação serial são muito utilizados atualmente (PILON, 2004), porém são pouco apropriados para dados espaciais, pois a dependência no espaço está presente em todas as direções e tais modelos tornam-se frágeis quando as locações ficam mais dispersas.

É natural pensar que quanto mais próximas as parcelas, tanto no espaço quanto no

tempo, mais correlacionados são os dados delas advindos. E ainda, sabe-se que os dados no espaço são tridimensionais, enquanto que no tempo são unidimensionais.

Os modelos que levam em consideração a informação espacial tiveram seu uso iniciado em modelos físicos e em estudos de fenômenos sociais (PILON, 2004).

Para experimentos de campo, a dependência espacial foi considerada digna de ser estudada em meados de 1930, quando teve início a metodologia que utiliza os desvios de unidades experimentais vizinhas como uma covariável (Nearest Neighbour - NN - de Papadakis), como uma alternativa ao uso do controle local (DUARTE, 2000).

Cressie (1993) mostra os efeitos da dependência espacial em planejamentos de experimentos. Mesmo que o princípio da casualização tenha sido realizado, acredita-se que as parcelas mais próximas sejam mais autocorrelacionadas do que as mais distantes. Diz ainda, que se for detectada a estrutura de autocorrelação e se essa informação for utilizada na análise estatística, estimativas mais eficientes dos contrastes entre as médias dos tratamentos são garantidas, mas se tal estrutura for desconsiderada pode impedir que diferenças reais sejam detectadas.

Esse mesmo autor afirma que métodos espaciais podem recuperar parte da eficiência perdida quando um método, não espacial, pouco eficiente foi utilizado, ou quando um experimento bem planejado foi prejudicado por alguma variável não esperada. Tal variável pode aumentar o ruído contido na informação. Exemplos dessa variável podem ser: variabilidade do solo não expressa no estágio de planejamento; ataques de insetos em direções preferenciais; desenvolvimento de doenças a partir de focos ou reboleiras, entre outras.

Para contornar esses problemas, uma maneira é utilizar uma aproximação espacial, em que é possível estimar e modelar a correlação a partir dos resíduos de uma análise clássica e usar, para estimar os efeitos do modelo, a teoria dos mínimos quadrados generalizados.

A utilização da estatística espacial foi ganhando espaço, com o avanço computacional, uma vez que mesmo conhecida há algum tempo, era pouco explorada devido à dificuldade de cálculos.

Assim, o desenvolvimento das técnicas geoestatísticas foi motivado por estudos referentes à exploração de minas de ouro na África do Sul (RIBEIRO JUNIOR, 1995). Ainda, segundo Andriotti (2003), a geoestatística foi criada em 1962 e quem primeiro utilizou este termo e a noção de variograma foi o engenheiro francês Georges Matheron que introduziu o conceito por meio do *Traité de Géostatistique Appliquée*.

Na análise geoestatística é considerada a distância entre as parcelas para modelar a

dependência espacial por meio de uma matriz de variâncias e covariâncias residuais (PILON, 2004).

Neste trabalho será utilizada a teoria de mínimos quadrados ordinários, em que não é levada em consideração a informação espacial. Considerando tal informação, serão utilizadas as teorias dos mínimos quadrados generalizados e da máxima verossimilhança, do modelo geoestatístico, e também a análise de vizinhança, do modelo de Papadakis.

O objetivo deste trabalho está em comparar esses métodos de estimação dos contrastes de médias de tratamentos para um conjunto de dados de um experimento em blocos ao acaso. Verificar se existe a variabilidade espacial dentro desse conjunto de dados. Se houver tal variabilidade, mostrar qual consequência que isso traz em relação a significância dos contrastes.

2 DESENVOLVIMENTO

2.1 Revisão de Literatura

2.1.1 Modelagem de dados experimentais

A análise experimental clássica baseia-se em três princípios básicos, estabelecidos por Fisher (1935), que são repetição, casualização e controle local, e são definidos por Barbin (2003) como:

- i. **repetição** consiste em se terem várias parcelas com o mesmo tratamento, objetivando confirmar a resposta que se tem a um determinado tratamento;
- ii. **casualização** é a distribuição dos tratamentos por sorteio nas parcelas, oferecendo a mesma chance a todos os tratamentos de ocuparem uma determinada posição ou parcela na área experimental e;
- iii. **controle local** é a subdivisão da área experimental (heterogênea) em áreas menores e homogêneas, sendo que em cada uma devem-se colocar todos os tratamentos, de preferência em igual número, o que é chamado de bloco.

Seja o modelo estatístico que representa o experimento em blocos ao acaso:

$$Y_{ij} = \mu + \tau_i + \omega_j + \varepsilon_{ij} \quad (1)$$

sendo, Y_{ij} a variável aleatória, μ a média geral, τ_i o efeito do i -ésimo tratamento, ω_j o efeito do j -ésimo bloco, e ε_{ij} o erro aleatório da (i, j) -ésima parcela, com possível autocorrelação entre os ε_{ij} .

A forma matricial do modelo fica:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (2)$$

em que, \mathbf{y} é o vetor de observações, \mathbf{X} é a matriz de delineamento, $\boldsymbol{\beta}$ é o vetor de parâmetros, e $\boldsymbol{\varepsilon}$ é o vetor de erros aleatórios supostos normalmente distribuídos, de média zero e variância constante.

Cargnelutti Filho, Storck e Lúcio (2003) afirmam que para reduzir o efeito do erro experimental, causado por pequenas variações nas unidades experimentais, e discriminar melhor as diferenças entre os tratamentos avaliados, deve-se aumentar o número de repetições, usar unidades experimentais homogêneas e manejo adequado do ensaio.

Segundo Grondona e Cressie (1991), a distribuição aleatória dos tratamentos produz erros experimentais aleatórios, que possuem autocorrelações muito pequenas e assim permitem o uso da teoria clássica.

Adotar unidades experimentais de pequeno tamanho, segundo Duarte (2000), aumenta a chance de violação da independência entre observações assumidas pelo método de mínimos quadrados ordinários (MQO). Esse fenômeno, denominado autocorrelação espacial, pode comprometer seriamente a comparação de tratamentos.

Na análise clássica, para estimar o vetor de contrastes de tratamentos $\Lambda\beta$, sendo Λ a matriz dos coeficientes dos contrastes e β o vetor de parâmetros, utilizam-se estimadores de mínimos quadrados ordinários (MQO) da seguinte maneira,

$$\widehat{\Lambda\beta}_{MQO} = \Lambda(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}, \quad (3)$$

com estimativa de variância dada por,

$$\widehat{var}(\widehat{\Lambda\beta}_{MQO}) = \widehat{\sigma}^2 \Lambda(\mathbf{X}'\mathbf{X})^{-1} \Lambda', \quad (4)$$

em que $\widehat{\sigma}^2$ é igual a $\widehat{\boldsymbol{\varepsilon}}'\widehat{\boldsymbol{\varepsilon}}/[n - \text{posto}(\mathbf{X})]$ e $\widehat{\boldsymbol{\varepsilon}}$ é o vetor de resíduos estimados, $\widehat{\boldsymbol{\varepsilon}} = \mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_{MQO}$, sendo que os erros quadráticos estimados para os contrastes são as raízes quadradas dos elementos da diagonal de $\widehat{var}(\widehat{\Lambda\beta}_{MQO})$.

Conforme Pilon (2004), se ignoradas todas as possíveis aleatoriedades, admitindo-se somente o que ocorreu, a variação natural nas parcelas faz com que os erros experimentais sejam autocorrelacionados, e argumenta que o problema então é modelar a autocorrelação e subseqüentemente usar essa informação para obtenção da melhor estimativa paramétrica.

Muitos autores como Grondona e Cressie (1991), Zimmerman e Harville (1991), Cullis e Gleeson (1991) entre outros, propõem uma abordagem que leve em conta a disposição espacial das unidades experimentais, por meio da Geoestatística. Já Cressie e Hartfield¹ (1996 apud PONTES; OLIVEIRA, 2004), defendem que, devido a sua maior eficiência no processo inferencial, é preferível que o pesquisador utilize a modelagem espacial direta. No entanto, Duarte (2000) afirma que esse tipo de análise pode ser utilizado alternativamente ou como complemento às análises usuais de blocos.

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

2.1.2 Modelos Geoestatísticos

Geoestatística é uma ferramenta fundamentada na teoria das variáveis regionalizadas, utilizada para estudar variabilidade espacial de variáveis nas quais seus valores estão de alguma maneira, relacionados à sua disposição espacial (VIEIRA et al., 1981²; VAUCLIN et al., 1983³ apud ORTIZ, 2004).

Dutilleul (1993) define heterogeneidade espacial como a variação nas densidades dos pontos (por exemplo, árvores numa floresta ou plantas numa comunidade) comparada com a variação esperada da expansão aleatória dos organismos.

Conforme Ribeiro Junior (1995), a geoestatística não é um tipo especial, diferente ou alternativo de estatística. Cada observação é descrita não apenas pelo seu valor, mas também por informações de sua posição, expressa por um sistema de coordenadas.

Para dados geoestatísticos univariados, basicamente o formato é (x_i, y_i) , $i = 1, \dots, n$, sendo que x_i identifica a posição espacial, que no contexto experimental representa o centro da parcela, tipicamente em um plano bidimensional, e y_i é uma variável de medição tomada na posição x_i .

Os métodos geoestatísticos consideram a dependência espacial das observações, lançando mão de medidas estatísticas que expressem esta estrutura de dependência tais como covariâncias, correlações e semivariâncias (RIBEIRO JUNIOR, 1995).

A suposição de que as observações são independentes e igualmente distribuídas, não é feita na geoestatística, o que faz com que essas medidas estatísticas assumam papel fundamental na análise, conforme esse mesmo autor. Completa ainda que a detecção da estrutura de autocorrelação espacial em áreas experimentais pode permitir estimação mais eficiente dos contrastes entre tratamentos.

O autor afirma que a geoestatística é construída a partir de conceitos básicos de estatística, em particular de processos estocásticos (ou funções aleatórias, termo usado na literatura francesa e por Matheron).

A forma básica de um modelo geoestatístico é um processo estocástico de valor real $\{Y(x) : x \in A \subset \mathbb{R}^2\}$ o qual é tipicamente considerado como uma realização parcial de um processo

²VIEIRA, S.R.; NIELSEN, D.R.; BIGGAR, J.W. Spatial variability of field-measured infiltration rate. **Soil Science Society of America Journal**, v.45, n.6, p.1040-1048, 1981.

³VAUCLIN, M.; VIEIRA, S.R.; VACHAUD, G.; NIELSEN, D.R. The use of cokriging with limited field soil observations. **Soil Science Society of America Journal**, v.47, p.175-184, 1983.

estocástico $\{Y(x) : x \in \mathbb{R}^2\}$.

O processo de medição Y_i pode ser freqüentemente registrado como uma versão modificada de uma variável aleatória subjacente $S(x_i)$, que é o valor na posição x_i do processo de interesse $\{S(x) : x \in A \subset \mathbb{R}^2\}$, sendo $S(x)$ chamado de sinal.

Pilon (2004) relata que um modelo geoestatístico básico é composto por dois componentes: um processo estocástico $S(x)$, e um modelo estatístico para a variável de medições, $y = (y_1, \dots, y_n)'$ condicional a $\{S(x) : x \in A \subset \mathbb{R}^2\}$. A confirmação é encontrada em Ribeiro Junior (1995) afirmando que, basicamente informações que possuem referencial de posição no espaço apresentam uma parte estruturada e uma parte aleatória.

Segundo Diggle e Ribeiro Jr. (2000), um processo estocástico $S(x)$ é dito *Gaussiano* se a distribuição conjunta de $S(x_1), \dots, S(x_n)$ é Gaussiana multivariada para todo n inteiro e um conjunto de posições x_i . Os autores afirmam que o processo é *estacionário* se a esperança e a variância de $S(x)$ são as mesmas para todo x , e a correlação entre $S(x_i)$ e $S(x_j)$ depende somente da distância entre as posições x_i e x_j . Completando, contam que o processo é estacionário e *isotrópico* se, adicionalmente, essa correlação depende somente da distância euclidiana, $u = \|x_i - x_j\|$, entre x_i e x_j . Citam as suposições que definem o modelo Gaussiano, estacionário e isotrópico para um conjunto de dados (x_i, x_j) :

1. $\{S(x) : x \in A \subset \mathbb{R}^2\}$ é um processo gaussiano estacionário com média μ , variância σ^2 , função de covariância $cov\{S(x_i), S(x_j)\} = \gamma(x_i, x_j) = \sigma^2\rho(u)$, e função de correlação $\rho(u) = corr\{S(x_i), S(x_j)\}$, em que $u = \|x_i - x_j\|$;
2. condicionalmente a $\{S(x) : x \in \mathbb{R}^2\}$, os y_i são realizações mutuamente independentes dos Y_i , normalmente distribuídos com média condicional $E[Y_i|S(\cdot)] = S(x_i)$ e variância condicional τ^2 .

Equivalentemente, esse modelo pode ser definido por:

$$Y_i = S(x_i) + Z_i, \quad i = 1, \dots, n, \quad (5)$$

em que $\{S(x_i) : x \in \mathbb{R}^2\}$ é definido pela suposição 1; Z_i são variáveis aleatórias mutuamente independentes $N(0, \tau^2)$.

Pilon (2004) mostra uma outra classe muito útil na prática, que são os processos gaussianos para os quais a média é variável, que podem depender de covariáveis e erros aleatórios, porém, com estrutura de covariância estacionária. Para tais situações, $Y(x) - \mu(x)$ é um processo estacionário Gaussiano com média zero.

A autora apresenta as características que, usualmente, o modelo da função de correlação $\rho(u)$ exige:

- i. $\rho(u)$ é monótona não crescente em u (as correlações entre duas medidas decrescem com o aumento da distância entre duas locações amostrais correspondentes),
- ii. $\rho(u) \rightarrow 0$ com $u \rightarrow \infty$ e
- iii. pelo menos um parâmetro no modelo controla a taxa com que $\rho(u)$ tende a zero.

Encontram-se em Diggle e Ribeiro Jr. (2000) alguns exemplos de famílias paramétricas que reúnem tais características e podem ser assumidas como função de correlação:

1. A família esférica, definida por

$$\rho(u; \phi) = \begin{cases} 1 - \frac{3}{2}(u/\phi) + \frac{1}{2}(u/\phi)^3, & 0 \leq u \leq \phi \\ 0, & u > \phi, \end{cases} \quad (6)$$

em que $\phi > 0$ determina a taxa para qual a correlação tende a zero ao se aumentar u . O nome se refere ao fato de que $\rho(u; \phi)$ tem uma interpretação geométrica como o volume da interseção de duas esferas das quais os centros estão a uma distância u . Como essa família depende somente do parâmetro ϕ , ela não é flexível ao modelo, e portanto, não recomendada.

2. A família exponencial poder, definida por

$$\rho(u) = \exp\{-(u/\phi)^\kappa\}, \quad (7)$$

com $\phi > 0$ e $0 < \kappa \leq 2$. O parâmetro κ , é chamado de ordem do modelo e determina a suavidade analítica do sinal $S(x)$. O processo correspondente é não-diferenciável se $\kappa < 2$, mas torna-se infinitamente diferenciável se $\kappa = 2$. A *função de correlação exponencial* corresponde ao caso em que $\kappa = 1$. O caso $\kappa = 2$ é, as vezes, denominado de *função de correlação Gaussiana*. O comportamento da função de $\kappa < 2$ para $\kappa = 2$ significa que a família exponencial poder bi-paramétrica não é tão flexível como parece ser à primeira vista.

3. A família Matérn, definida por

$$\rho(u; \phi, \kappa) = \{2^{\kappa-1}\Gamma(\kappa)\}^{-1}(u/\phi)^\kappa K_\kappa(u/\phi), \quad (8)$$

em que (ϕ, κ) são parâmetros descritos anteriormente e $K_\kappa(\cdot)$ denota a função Bessel modificada de terceiro tipo de ordem κ . A família é válida para $\phi > 0$ e $\kappa > 0$. O caso $\kappa = 0,5$

é o mesmo que a função de correlação exponencial. A função de correlação Gaussiana é o caso limitado para $\kappa \rightarrow \infty$. Essa família é, provavelmente, a melhor escolha devido à sua flexibilidade.

2.1.3 Métodos para estimação dos parâmetros do modelo Geoestatístico

Ao estimar os parâmetros do modelo Geoestatístico, os dados devem ter dependência espacial.

Para testar se o modelo é espacial (geoestatístico) ou não-espacial, uma alternativa é o teste da verossimilhança que consiste em minimizar os critérios de informação de máxima verossimilhança restrita. Assim, é selecionado o que tiver o menor valor para tais métodos (FREITAS, 2007). Os mais comuns são os critérios de Informação de Akaike (AIC) e de Informação Bayesiano ou de Schwarz (BIC), que podem ser expressos pelas seguintes fórmulas:

$$AIC = -2\ell + 2p \quad (9)$$

$$BIC = -2\ell + p \log(n - \text{posto}(\mathbf{X})) \quad (10)$$

em que ℓ é o logaritmo natural da função de verossimilhança, p é o número de parâmetros da matriz de covariâncias e n é o número de observações.

Vale salientar que, segundo a autora, o critério BIC penaliza os modelos com maior número de parâmetros e que nem sempre os dois critérios concordam quanto ao melhor modelo. Ainda, diz que, alguns autores afirmam que a performance do AIC é melhor que do BIC para identificar o melhor modelo.

2.1.3.1 Variograma seguido de Mínimos Quadrados Generalizados

O variograma é uma técnica muito utilizada em modelagens espaciais que consiste em uma função que, segundo Ver Hoef e Cressie (1993), contém informação na autocorrelação espacial entre unidades experimentais.

A ferramenta mais indicada na estimativa da dependência entre amostras envolvendo duas direções (x, y) é o semivariograma, que analisa o grau desta dependência espacial dentro de um campo experimental.

Conforme Pires e Strieder (2006), descrever fenômenos geológicos com modelos matemáticos, é permitido utilizando o variograma que representa quantitativamente a variação do fenômeno regionalizado. Andriotti (2003) afirma que a análise de um variograma é a única técnica disponível para medir a variabilidade espacial de uma variável regionalizada.

A vantagem de se avaliar a dependência espacial utilizando o variograma é que, sob estacionaridade, este tem uma relação direta e simples com a função de autocovariância (DUARTE, 2000).

Quando a covariância existe, é dado o nome de Estacionaridade Estrita (ANDRIOTTI, 2003). Se ela não existir e a variável tiver uma capacidade infinita para dispersão, então assume-se que somente o variograma existe e é estacionário (Hipótese Intrínseca). Existência de covariância implica existência do variograma, mas o contrário não é válido. McBratney e Webster⁴ (1986 apud RIBEIRO JUNIOR, 1995), afirmam que sendo válida a hipótese intrínseca, o semivariograma, se conhecido, contém toda a informação sobre a variabilidade espacial do atributo em estudo.

A estacionaridade de segunda ordem e estrita coincidem quando a função aleatória é gaussiana, pois tal processo fica caracterizado pelas suas funções média e covariância (CRESSIE, 1989). Esta última serve como medida da associação espacial entre as observações, mas para isso é necessário que certas hipóteses restritivas sejam assumidas (no caso, estacionaridade de segunda ordem).

Em termos matemáticos, a estacionaridade de segunda ordem pode ser expressa por:

- a) $E\{Z(x)\} = m, \forall x;$
- b) $C(h) = E\{Z(x+h)Z(x)\} - m^2, \forall x;$
- c) $Var\{Z(x)\} = E\{[Z(x) - m]^2\} = C(0), \forall x.$

Nesse caso, $C(0) - C(h) = \frac{1}{2}E\{[Z(x+h) - Z(x)]^2\}, \forall x$ (existência de variograma).

Sob estacionaridade de segunda ordem, a semivariância $\gamma(h)$, a covariância $C(h)$ e a autocorrelação $\rho(h)$ podem ser relacionadas. Enquanto a covariância e a correlação são medidas de similaridade, a semivariância é medida de dissimilaridade (JOURNAL⁵, 1989, apud RIBEIRO JUNIOR, 1995).

Uma alternativa à covariância é a semivariância, que exige hipóteses menos restritivas (hipótese intrínseca) e, portanto, aplica-se a um universo maior de situações (RIBEIRO JUNIOR, 1995).

⁴McBRATNEY, A.B.; WEBSTER, R. Choosing functions for semivariograms of soil properties and fitting them to sample estimates. **Journal of Soil Science**, Oxford, v.37, p.617-39, 1986.

⁵JOURNAL, A.G. **Fundamentals of geostatistics in five lessons**. Washington, American Geophysical Union, 1989. 40p.

Pontes e Oliveira (2004) definem o semivariograma como sendo a ferramenta indispensável para caracterizar a dependência espacial. O valor de $\gamma(h)$ é denominado semivariância entre dois pontos separados pelo vetor h . Os autores denominam ainda de *variograma* a função $2\gamma(\cdot)$ e, conseqüentemente, de *semivariograma* a função $\gamma(\cdot)$, que é uma função real não-negativa, dada por:

$$\hat{\gamma}(u) = \frac{\sum_{M(u)} (Y(x_i) - Y(x_j))^2}{2N(u)}, \quad (11)$$

em que, $u = \|x_i - x_j\|$ é a distância euclidiana entre x_i e x_j ; $Y(x_i)$ e $Y(x_j)$ são as medidas de interesse nas posições x_i e x_j , respectivamente; $M(u) \equiv \{[x_i = (l, c), x_j = (l', c')] : \sqrt{(l - l')^2 + (c - c')^2} = u\}$, o conjunto de todos os pares de posições que estão separadas a uma distância euclidiana u , freqüentemente denominada “lag” e $N(u)$ é o número total de pares de posições em $M(u)$.

Citam as seguintes propriedades decorrentes da definição de semivariograma:

- i. $\gamma(0) = 0$;
- ii. $\gamma(x_i - x_j) = \gamma(x_j - x_i)$, $\forall x_i$ e $x_j \in \mathbb{R}$ e
- iii. $\sum_{i=1}^k \sum_{j=1}^k a_i a_j \gamma(x_i - x_j) \leq 0$, $\forall x_1, \dots, x_k \in \mathbb{R}$, $\forall a_1, \dots, a_k \in \mathbb{R}$, tal que $\sum_{i=1}^k a_i = 0$, $\forall k \in \mathbb{Z}^+$.

É importante se ter em mente que, ao se trabalhar com geoestatística e assumindo a hipótese intrínseca, a variável aleatória considerada é $Y(h) = Z(x + h) - Z(x)$. É a respeito da variabilidade de $Y(h)$ que se estará referindo ao efetuar os cálculos fundamentais como a estimação do semivariograma.

Os parâmetros do semivariograma, que podem ser observados na Figura 1, são os que seguem:

- i. O **efeito pepita** ou *nugget* (τ^2): reflete a variabilidade existente em pequenas distâncias e também incorpora erros analíticos.
- ii. O **patamar** ou *sill* ($\sigma^2 + \tau^2$): é o valor no qual a função semivariograma se estabiliza.
- iii. O **alcance** ou *range* (3ϕ): é que indica a distância além da qual os valores passam a não apresentar mais correlação.

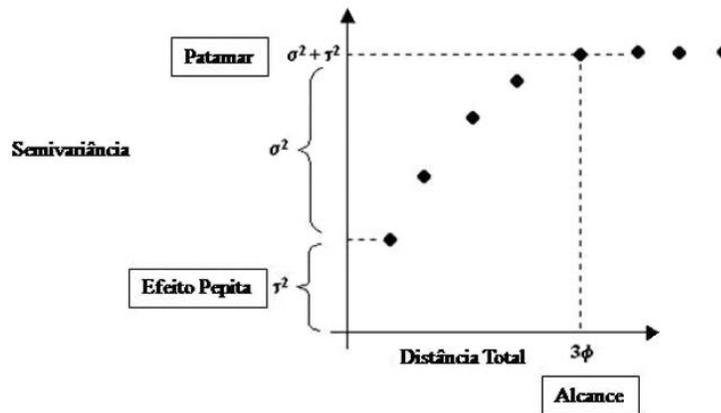


Figura 1 - Semivariograma Empírico Hipotético

O estimador (clássico) de Matheron vem sendo muito utilizado nas pesquisas em geoestatística. Nada mais sendo do que um estimador obtido pelo método dos momentos, tem, sob a hipótese intrínseca, a desejável qualidade de não-viesado. Sua expressão é:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{N(h)} [Z(s+h) - Z(s)]^2, \quad (12)$$

em que $N(h)$ é o número de pares de pontos separados pela distância h . Assim, para obtenção de tal estimador não é necessário o conhecimento da distribuição de $Z(s)$.

Outros estimadores têm sido propostos por diversos autores buscando uma maior qualidade de estimação. Os métodos de estimação do semivariograma podem ser agrupados em cinco grupos:

- 1) método dos momentos,
- 2) método dos mínimos quadrados,
- 3) método “jackknife”,
- 4) método de máxima verossimilhança e
- 5) método de validação cruzada.

Existem alguns modelos adequados descritos na literatura, que têm atendido a contento as necessidades de modelagem de variogramas. Alguns modelos básicos, usuais em geoestatística são:

Modelo esférico:

$$\gamma(h) = \begin{cases} t + u \left[\frac{3h}{2v} - \frac{1}{2} \frac{h^3}{v^3} \right], & 0 < h < v \\ t + u, & h \geq v. \end{cases} \quad (13)$$

Modelo exponencial:

$$\gamma(h) = t + u \left[1 - \exp \left\{ \frac{-3h}{v} \right\} \right], \quad h \geq 0. \quad (14)$$

Modelo gaussiano:

$$\gamma(h) = t + u \left[1 - \exp \left\{ \frac{-3h^2}{v^2} \right\} \right], \quad h \geq 0. \quad (15)$$

Nesses modelos, o vetor de parâmetros é $\boldsymbol{\theta} = (t, u, v)'$, cujos componentes são chamados de efeito pepita, patamar e alcance, respectivamente. Todos os parâmetros devem ser não negativos.

Modelo potência:

$$\gamma(h) = t + bh^\lambda, \quad h \geq 0. \quad (16)$$

Neste caso, o vetor de parâmetros é $\boldsymbol{\theta} = (t, b, \lambda)'$, em que $t \geq 0$, $b \geq 0$ e $0 \leq \lambda < 2$. O modelo linear é um caso particular deste modelo quando $\lambda = 1$.

Ribeiro Junior (1995) afirma que as expressões de outros modelos válidos podem ser encontradas na literatura como em Braga (1990), Valente (1982), entre outros.

Além destes modelos básicos, tem-se que qualquer combinação linear entre eles também atende às condições de negativo definido condicional, o que amplia as possibilidades para o ajuste.

Segundo esse mesmo autor, o método de ajuste de modelos variográficos mais comum é o chamado ajuste “a sentimento”, que uma vez obtido o variograma amostral, este serve de indicação visual sobre a forma e os parâmetros do modelo a ser adotado. Devido aos grandes avanços computacionais, esse método está sendo cada vez menos utilizado.

O método de ajuste de modelos de semivariograma por mínimos quadrados ordinários tem se mostrado insensível ao que se pretende no ajuste, porém, este é o de implementação mais fácil. Por outro lado, um procedimento mais criterioso é o uso de mínimos quadrados generalizados, o qual exige um processo iterativo, de implementação mais difícil. Uma forma intermediária é o uso de mínimos quadrados ponderados que podem ajustar bem modelos de variogramas ou, ao menos, servir de valor inicial para estimação por mínimos quadrados generalizados.

Assim, como o método de ajuste de modelos de semivariograma por mínimos quadrados generalizados é um bom procedimento, ele estima o vetor de estimativas dos contrastes de tratamentos,

$$\widehat{\Lambda\beta}_{MQG} = \Lambda(\mathbf{X}'\Gamma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Gamma^{-1}\mathbf{y}, \quad (17)$$

sendo que Γ^{-1} é a inversa da matriz contendo os valores do semivariograma entre todos os pares de locações que contêm os dados \mathbf{y} .

A variância é estimada por:

$$\widehat{var}(\widehat{\Lambda\beta}_{MQG}) = \Lambda(\mathbf{X}'\Gamma^{-1}\mathbf{X})^{-1}\Lambda', \quad (18)$$

logo, os erros quadráticos, estimados para os contrastes, são as raízes quadradas dos elementos da diagonal da matriz $\widehat{var}(\widehat{\Lambda\beta}_{MQG})$.

Segundo Kuan (2004), similar ao método dos mínimos quadrados ordinários, o de mínimos quadrados generalizados define o vetor de valores ajustados dado por:

$$\widehat{\mathbf{y}}_{MQG} = \mathbf{X}\widehat{\beta}_{MQG} = \mathbf{X}(\mathbf{X}'\Gamma^{-1}\mathbf{X})^{-1}\mathbf{X}'\Gamma^{-1}\mathbf{y}. \quad (19)$$

O vetor de resíduos estimados para este método é:

$$\widehat{\boldsymbol{\varepsilon}}_{MQG} = \mathbf{y} - \widehat{\mathbf{y}}_{MQG}. \quad (20)$$

Pode ser verificado que esse vetor de resíduos não é ortogonal a \mathbf{X} , isto é,

$$\widehat{\boldsymbol{\varepsilon}}_{MQG}'\mathbf{X} = \mathbf{y}'[\mathbf{I} - \Gamma^{-1}\mathbf{X}(\mathbf{X}'\Gamma^{-1}\mathbf{X})^{-1}\mathbf{X}']\mathbf{X} \neq \mathbf{0}. \quad (21)$$

Assim, os resíduos calculados para as iterações do método do variograma seguido de mínimos quadrados generalizados não somam zero.

2.1.3.2 Método da Máxima Verossimilhança

O método da máxima verossimilhança consiste na maximização da função de densidade de probabilidades, em relação aos efeitos fixos e aos componentes de variâncias dos efeitos aleatórios do modelo. É um método iterativo e fornece estimativas positivas de componentes de variância.

Segundo Zimmerman e Harville (1991), um inconveniente de usar a máxima verossimilhança (MV) para estimar os parâmetros dos modelos lineares de campo aleatório, particularmente no caso de experimentos espaciais (em que p/n pode ser relativamente grande, sendo p o

número de parâmetros e n o número de observações) é que o estimador de β pode ser viesado. Tal viés é atribuível à “perda nos graus de liberdade” devido à estimação dos efeitos fixos do modelo e pode ser otimamente reduzido adotando o método da máxima verossimilhança restrita (MVRE).

Conforme Pilon (2004), no método da MVRE, combinações lineares especiais dos dados são usadas no lugar das observações originais. Essas combinações são chamadas incrementos generalizados, que filtram as tendências e estimam somente os parâmetros de covariância, levando a um estimador não viesado.

2.1.4 Estimação dos parâmetros do modelo de Papadakis

Um método muito utilizado em modelagem espacial é o dos Vizinhos-Próximos, também conhecido por método de Papadakis, o qual analisa experimentos de campo examinando a dependência espacial usando informações observadas nas parcelas vizinhas (BESAG; KEMPTON, 1986).

Pilon (2004) afirma que na análise de vizinhança, a dependência espacial é modelada por uma análise de covariância, utilizando-se como covariável as médias dos resíduos entre a parcela e as parcelas vizinhas.

Tal método consiste em corrigir a resposta de cada parcela pelo efeito médio do tratamento aplicado à parcela, e então, a média dos valores corrigidos das parcelas vizinhas e a própria parcela é utilizada como covariável, seguindo-se com uma análise de covariância.

Procura-se dessa forma, eliminar grande parte das tendências existentes no solo das áreas experimentais, através do desempenho das parcelas vizinhas, com uma conseqüente diminuição no erro experimental.

Cargnelutti Filho, Storck e Lúcio (2003) afirmam que quanto maior o número de tratamentos, maior é a eficiência do método de Papadakis. Os autores, em seu trabalho, utilizam um experimento em blocos ao acaso, porém, para explicar a possível correlação existente na área, utilizam a covariável referente a média dos resíduos das parcelas vizinhas. Assim, para a análise clássica eles consideram os efeitos de blocos, e para a análise via Papadakis, desconsideram esses efeitos, assumindo o experimento inteiramente ao acaso.

Para se usar esse método é necessário que seja determinada uma estrutura de vizinhança. Esses autores calculam cinco formas de produtividade em cada unidade experimental em um delineamento inteiramente ao acaso.

Sejam, $F_{k(m,n)}$ o índice de produtividade, $k = 1, 2, \dots, 5$ formas, m o número da linha,

n o número da coluna e $R_{(i,j)}$ refere-se ao resíduo da parcela posicionada na linha i e coluna j . O $R_{(i,j)}$ é calculado da seguinte maneira:

$$R_{(i,j)} = Y_{(i,j)} - \bar{Y}_{(i,\cdot)}, \quad (22)$$

em que $Y_{(i,j)}$ é o valor observado na unidade experimental com o i -ésimo tratamento e j -ésima repetição e $\bar{Y}_{(i,\cdot)}$ é a média das unidades experimentais com o i -ésimo tratamento.

As formas que os autores utilizam são:

$$F_{1(m,n)} = [R_{(m,n)} + R_{(m,n-1)} + R_{(m,n+1)} + R_{(m-1,n)} + R_{(m+1,n)}]/5;$$

$$F_{2(m,n)} = [R_{(m,n)} + R_{(m,n-2)} + R_{(m,n-1)} + R_{(m,n+1)} + R_{(m,n+2)}]/5;$$

$$F_{3(m,n)} = [5F_{1(m,n)} + R_{(m,n-2)} + R_{(m,n+2)}]/7;$$

$$F_{4(m,n)} = [7F_{3(m,n)} + R_{(m-1,n-1)} + R_{(m-1,n+1)} + R_{(m+1,n-1)} + R_{(m+1,n+1)}]/11;$$

$$F_{5(m,n)} = [11F_{4(m,n)} + R_{(m-1,n-2)} + R_{(m-1,n+2)} + R_{(m+1,n-2)} + R_{(m+1,n+2)}]/15.$$

2.2 Material e Métodos

2.2.1 Material

Os dados a serem analisados no presente trabalho foram adaptados de Marciano (1999) que realizou um experimento em blocos casualizados, com quatro repetições e nove tratamentos, conduzido nos anos agrícolas de 1996/1997 e 1997/1998, em condições de campo, na área da Estação Experimental do Instituto Agrônomo de Campinas, em Piracicaba - SP.

O croqui do experimento encontra-se na figura 2.

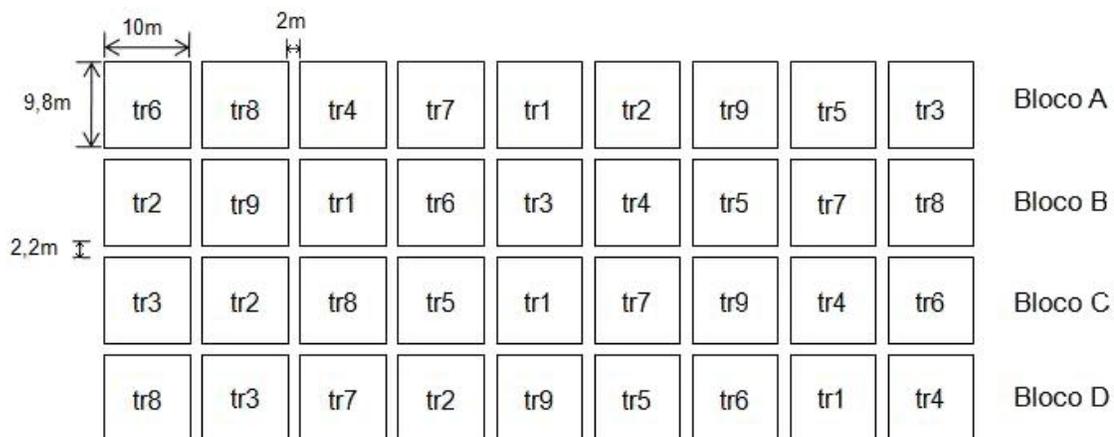


Figura 2 - Croqui do experimento

Foram coletados dados para caracterizar o solo no início do experimento, portanto,

sem a aplicação de tratamentos. A esses dados, de ensaio em branco, foram adicionados os efeitos de tratamentos e blocos, estimados para os dados um ano após a aplicação dos tratamentos no experimento. Esses dados encontram-se no ANEXO A.

Como planta-teste, foi utilizada a cana-de-açúcar (*Saccharum officinarum*) variedade RB-78-5148, implantada em ciclo anual no ano agrícola de 1996/1997 e cultivada como soqueira no ano agrícola de 1997/1998.

A variável em estudo, neste trabalho, foi a concentração do carbono orgânico no solo, sob o efeito de dois diferentes resíduos de origem urbana: o lodo de esgoto que, anaerobicamente digerido, foi proveniente da Estação de Tratamento de Esgotos de Barueri - SABESP, de Barueri - SP, e o composto de lixo, proveniente da Usina de Tratamento de Lixo da Cidade de São Jorge - Prefeitura Municipal de Santo André, Santo André - SP (MARCIANO, 1999).

Os tratamentos foram: (1) Calagem + Adubação mineral; (2) Testemunha para o lodo; (3) Lodo de esgoto, dose A; (4) Lodo de esgoto, dose B; (5) Lodo de esgoto, dose C; (6) Testemunha para o composto; (7) Composto de lixo, dose A; (8) Composto de lixo, dose B; (9) Composto de lixo, dose C.

A calagem e a adubação mineral do tratamento (1) seguiram as recomendações técnicas para a cultura da cana-de-açúcar. As doses de lodo de esgoto foram: dose A = 33 Mg ha⁻¹, dose B = 66 Mg ha⁻¹ e dose C = 99 Mg ha⁻¹ do material seco (respectivamente 100 Mg ha⁻¹, 200 Mg ha⁻¹ e 300 Mg ha⁻¹ do material úmido). As doses do composto de lixo foram: dose A = 20 Mg ha⁻¹, dose B = 40 Mg ha⁻¹ e dose C = 60 Mg ha⁻¹ do material seco (respectivamente 39 Mg ha⁻¹, 78 Mg ha⁻¹ e 117 Mg ha⁻¹ do material úmido).

O conteúdo de matéria orgânica do solo, alterado pelos tratamentos (principalmente pela aplicação de lodo de esgoto ou composto de lixo), foi acompanhado através da determinação dos teores de carbono oxidável do solo, realizada em amostras coletadas periodicamente na camada entre 0 e 0,20 m de profundidade. O C-oxidável foi determinado pela sua oxidação em solução 1N de dicromato de potássio (K₂Cr₂O₇) em meio ácido (H₂SO₄ concentrado) seguida de titulação com solução padronizada de sulfato ferroso amoniacal [Fe(NH₄)₂(SO₄)₂] (MARCIANO, 1999).

2.2.2 Métodos

2.2.2.1 Análise de Variância Clássica (ANOVA)

Sendo verificados os três princípios básicos de experimentação: repetição, casualização e controle local, será realizada a ANOVA clássica para estudar se os efeitos de trata-

mentos do conjunto de dados provenientes de um delineamento em blocos casualizados diferem entre si. Para tanto, será assumido que os erros são não correlacionados: $\varepsilon \sim N(0, \sigma^2)$.

Para estimar o vetor de contrastes $\mathbf{\Lambda}\boldsymbol{\beta}$, sendo $\mathbf{\Lambda}$ a matriz dos coeficientes dos contrastes, e $\boldsymbol{\beta}$ o vetor de parâmetros, a análise clássica utiliza estimadores de mínimos quadrados ordinários (MQO), conforme (3).

A estimativa de variância dos contrastes estimados ($\widehat{var}(\widehat{\mathbf{\Lambda}\boldsymbol{\beta}}_{MQO})$), é dada por (4), sendo que os erros quadráticos estimados, para esses contrastes, são as raízes quadradas dos elementos da diagonal desta matriz.

2.2.2.2 Mínimos Quadrados Generalizados (MQG)

Para contornar uma possível autocorrelação entre parcelas que recebem determinado tratamento dentro dos blocos em um experimento, utilizam-se os métodos de estimação que consideram a dependência espacial da área.

Um método que pode ser utilizado quando os dados forem referenciados espacialmente e os erros autocorrelacionados, ou seja, existe dependência espacial na área, é o método de mínimos quadrados generalizados (MQG) que utiliza uma média ponderada de modo a atribuir um peso menor àquelas parcelas que receberam o mesmo tratamento e se encontram próximas, pois, possivelmente possuem valores parecidos e assim, contribuindo com pequenas informações extras. Por outro lado, quando uma parcela recebe o mesmo tratamento que as outras, mas esta se encontra distante, esta terá um peso alto porque representa uma região maior. Esses pesos são também conhecidos como pesos ótimos e são obtidos pelos valores do semivariograma.

Esse método, que também estima o vetor de contrastes $\mathbf{\Lambda}\boldsymbol{\beta}$, segue um processo iterativo, no qual o primeiro passo consiste em encontrar os resíduos que são calculados por (23) e a Análise da Variância Clássica.

Para encontrar os resíduos, faz-se,

$$r_{ij} = y_{ij} - \hat{\tau}_i - \hat{\omega}_j, \quad (23)$$

sendo r_{ij} e y_{ij} , respectivamente, o resíduo e o valor observado da parcela que recebeu o tratamento i e está localizada no bloco j , $\hat{\tau}_i$ é o efeito estimado do i -ésimo tratamento e $\hat{\omega}_j$ é o efeito estimado do j -ésimo bloco.

O passo seguinte consiste em ajustar o semivariograma empírico, descrito na equação

(11), aos resíduos do passo anterior, descritos, agora, em função das posições.

$$\hat{\gamma}(u) = \frac{\sum_{M(u)} (R(x_i) - R(x_j))^2}{2N(u)}, \quad (24)$$

em que, $x_i = (l, c)$ e $x_j = (l', c')$ são as posições espaciais da i -ésima e j -ésima parcelas, respectivamente, $u = \|x_i - x_j\|$ é a distância euclidiana entre x_i e x_j , $R(x_i)$ e $R(x_j)$ são os resíduos da i -ésima e da j -ésima parcelas, respectivamente, $M(u) \equiv \{[x_i = (l, c), x_j = (l', c')] : \sqrt{(l - l')^2 + (c - c')^2} = u\}$, o conjunto de todos os pares de posições que estão separadas a uma distância euclidiana u , freqüentemente denominada “lag” e $N(u)$ é o número total de pares de parcelas em $M(u)$.

Para tanto, os valores empíricos calculados para $\hat{\gamma}(u)$ não devem ser usados na matriz $\mathbf{\Gamma}$ do semivariograma, porque esta deve satisfazer propriedades positivas definidas, o que não acontece.

O terceiro passo do processo garante tais propriedades positivas definidas ajustando um semivariograma teórico aos valores do semivariograma empírico. Um modelo de semivariograma teórico pode ser, por exemplo, o exponencial dado por:

$$\gamma(u; (\tau^2, \sigma^2, \phi)) = \begin{cases} \tau^2 + \sigma^2[1 - \exp(\frac{-u}{\phi})] & \text{se } 0 < u \leq \phi, \\ \tau^2 + \sigma^2 & \text{se } u > \phi, \\ 0 & \text{se } u = 0. \end{cases} \quad (25)$$

Em seguida, são obtidas as estimativas dos parâmetros desse semivariograma, $\hat{\tau}^2$, $\hat{\sigma}^2$ e $\hat{\phi}$ pelo método de mínimos quadrados ponderados.

O penúltimo passo consiste em inserir os valores das estimativas dos parâmetros de correlação no modelo do semivariograma teórico, e construir a matriz de covariância, $\mathbf{\Gamma}$ de dimensão $n \times n$.

Enfim, os efeitos de tratamentos e blocos são estimados por

$$\hat{\beta}_{MQG} = (\mathbf{X}'\mathbf{\Gamma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Gamma}^{-1}\mathbf{y}, \quad (26)$$

em que $\hat{\beta}$ é o vetor de estimativas dos parâmetros, \mathbf{X} e \mathbf{y} como descritos em (2) e $\mathbf{\Gamma}^{-1}$ como em (17).

As estimativas dos contrastes são obtidas conforme equação (17). E a variância é estimada segundo (18).

2.2.2.3 Estimação por Máxima Verossimilhança

Um outro método de estimação que considera a dependência espacial da área experimental é o da máxima verossimilhança. Esse método é iterativo e fornece estimativas positivas de componentes de variância.

Considere o modelo geoestatístico:

$$y(x_{ik}) = \mu(x_{ik}) + S(x_{ik}) + Z_i, \quad i = 1, \dots, n \text{ e } k = 1, \dots, p, \quad (27)$$

em que,

- i. x_{ik} é a posição espacial da i -ésima parcela que recebe o k -ésimo tratamento,
- ii. $y(x_{ik})$ é o valor observado da i -ésima parcela que recebe o k -ésimo tratamento,
- iii. $\mu(x_{ik})$ é a média determinada a partir de um modelo de regressão linear dado por

$$\mu(x_{ik}) = \sum_{k=1}^p f_k(x_i) \alpha_k, \quad (28)$$

sendo que $f_k(\cdot)$ são as observações espacialmente relacionadas às variáveis explicativas, $f_1(x_i), \dots, f_p(x_i)$ é a i -ésima linha da matriz \mathbf{X} e α_k é a média do tratamento k ,

- iv. $S(\cdot)$ é um processo estacionário Gaussiano, com média zero, variância σ^2 e função de correlação $\rho(u; \phi)$. Assim, $S \sim NMV\{0, \sigma^2 R(\phi)\}$, sendo $R(\phi)$ a matriz de covariância de dimensão $n \times n$ com o (i, j) -ésimo elemento dado por $\rho(u_{ij})$, em que $u_{ij} = \|x_i - x_j\|$ é a distância euclidiana entre os pontos x_i e x_j ,
- v. Z_i são variáveis aleatórias, independentes e normalmente distribuídas com média zero e variância τ^2 , ou seja, $Z_i \sim N.i.i.d.(0, \tau^2)$.

Para facilitar a derivada, o modelo (1) deve ser escrito na forma matricial conforme equação (2). Seja $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$, em que \mathbf{X} é a matriz do delineamento de dimensão $n \times p$. Também, tem-se que $\boldsymbol{\theta} = (\tau^2, \sigma^2, \phi)$, denotando o conjunto de parâmetros que define a matriz de covariância dos dados \mathbf{y} , sendo $Var(\mathbf{y}) = G(\boldsymbol{\theta})$. Note que $G(\boldsymbol{\theta}) = \tau^2 I + \sigma^2 R(\phi)$, em que I é a matriz identidade e $R(\phi)$ é a matriz de correlações da estrutura exponencial de dimensão $n \times n$, com o (i, j) -ésimo elemento dado por $r_{ij} = \rho(\|x_i - x_j\|; \phi)$.

Com essas notações, o modelo (27) implica que $\mathbf{y} \sim NMV\{\mathbf{X}\boldsymbol{\beta}, G(\boldsymbol{\theta})\}$, ou seja, segue uma distribuição multivariada gaussiana, em que a função de verossimilhança é expressa por:

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}; \mathbf{y}) = (2\pi)^{-\frac{n}{2}} |G(\boldsymbol{\theta})|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \{G(\boldsymbol{\theta})\}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right], \quad (29)$$

em que $G(\boldsymbol{\theta}) = \tau^2 I + \sigma^2 R(\phi)$, sendo $\nu^2 = \frac{\tau^2}{\sigma^2}$, portanto, tem-se que $G(\boldsymbol{\theta}) = \sigma^2 \{\nu^2 + R(\phi)\} = \sigma^2 \mathbf{V}$.

Segue que o logaritmo da função de verossimilhança para $(\boldsymbol{\beta}, \boldsymbol{\theta})$ é

$$\ell(\boldsymbol{\beta}, \boldsymbol{\theta}) = -\frac{1}{2} \{n \log(2\pi) + \log|G(\boldsymbol{\theta})| + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \{G(\boldsymbol{\theta})\}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\}. \quad (30)$$

Para maximizar a equação (30), primeiro elimina-se $\boldsymbol{\beta}$ da maximização numérica de $\ell(\cdot)$ notando que, para dado $\boldsymbol{\theta}$, o estimador de máxima verossimilhança de $\boldsymbol{\beta}$ é

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}) = \{\mathbf{X}'[G(\boldsymbol{\theta})]^{-1}\mathbf{X}\}^{-1}\mathbf{X}'[G(\boldsymbol{\theta})]^{-1}\mathbf{y}, \quad (31)$$

A substituição de (31) em (30) dá então a função de verossimilhança reduzida para $\boldsymbol{\theta}$,

$$L(\boldsymbol{\theta}) = \ell(\widehat{\boldsymbol{\beta}}(\boldsymbol{\theta}), \boldsymbol{\theta}).$$

A matriz de estimativas de variâncias dos contrastes estimados é dada pela fórmula:

$$\widehat{\text{var}}(\widehat{\boldsymbol{\Lambda}\boldsymbol{\beta}}_{MV}) = \widehat{\sigma}^2 \boldsymbol{\Lambda}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\boldsymbol{\Lambda}', \quad (32)$$

sendo,

$$\widehat{\sigma}^2 = n^{-1}(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}})' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}),$$

e os erros quadráticos das estimativas dos contrastes são dados pela raiz quadrada dos elementos da diagonal de $\widehat{\text{var}}(\widehat{\boldsymbol{\Lambda}\boldsymbol{\beta}}_{MV})$.

2.2.2.4 Método de Papadakis

Esse método busca controlar a heterogeneidade da área experimental acrescentando uma covariável ao modelo. Assim, nesse método, deve-se realizar uma análise de covariância, sendo que cada parcela deve ter o seu valor observado corrigido pelo efeito de uma covariável espacial adicionada ao modelo, que é a média dos resíduos entre as parcelas vizinhas e a própria parcela. Segundo Cargnelutti Filho, Storck e Lúcio (2003), essa covariável é denominada índice ambiental.

Para calcular a covariável índice ambiental, Z , deve-se estimar o resíduo de cada parcela. Tal resíduo pode ser calculado utilizando a equação (23).

O índice ambiental será calculado em cada parcela, da seguinte forma, pois o interesse está em verificar a dependência espacial dentro dos blocos.

$$Z_{(l,c)} = \frac{r_{(l,c-1)} + r_{(l,c)} + r_{(l,c+1)}}{3}, \quad (33)$$

sendo que cada unidade experimental é representada pelas coordenadas (l, c) , em que $l = 1, \dots, L$ é a coordenada que representa a l -ésima parte da área experimental dividida em L linhas e $c = 1, \dots, C$ é a coordenada que representa a c -ésima parte da área experimental dividida em C colunas.

Para as parcelas centrais será considerada uma estrutura de vizinhança de três parcelas, sendo a própria parcela, uma à esquerda e uma à direita. Já para as unidades experimentais das extremidades serão consideradas as duas parcelas disponíveis da estrutura citada.

Segue-se com a análise de covariância para o experimento em blocos ao acaso, de acordo com o modelo de regressão:

$$y_{ij} = \bar{y}_i + \bar{y}_j + \theta Z_{ij} + \varepsilon_{ij}, \quad (34)$$

sendo y_{ij} conforme (23), \bar{y}_i e \bar{y}_j são as médias do i -ésimo tratamento e j -ésimo bloco, respectivamente, θ é o coeficiente de regressão entre a variável resposta e a covariável Z , Z_{ij} é a covariável, índice ambiental, medida na parcela localizada no bloco j que recebeu o tratamento i e ε_{ij} é o erro experimental associado à observação y_{ij} .

Para estimar os contrastes entre as médias dos tratamentos será usado:

$$\widehat{\Lambda\beta}_{PP} = \Lambda(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}b, \quad (35)$$

sendo $\widehat{\Lambda\beta}_{PP}$, as estimativas dos contrastes das médias dos tratamentos, Λ , a matriz dos coeficientes dos contrastes, \mathbf{X} , a matriz do delineamento, \mathbf{y} , o vetor de observações, e b é um escalar calculado pela expressão:

$$b = \{\mathbf{Z}'[\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{Z}\}^{-1}\mathbf{Z}'[\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']\mathbf{y}, \quad (36)$$

com \mathbf{Z} sendo o vetor das covariáveis e \mathbf{I}_n a matriz identidade de dimensão $n \times n$.

A matriz de variâncias desses contrastes pode ser obtida por:

$$\widehat{var}(\widehat{\Lambda\beta}_{PP}) = \Lambda[(\mathbf{X}'\mathbf{X})^{-1} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{R}'_Z\mathbf{R}_Z)^{-1}\mathbf{Z}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}]\Lambda'\widehat{\sigma}^2, \quad (37)$$

em que, $\mathbf{R}_Z = (\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{Z}$, é o vetor de resíduos do ajustamento da covariável e

$$\widehat{\sigma}^2 = \frac{\mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} - \mathbf{y}'\mathbf{R}_Z(\mathbf{R}'_Z\mathbf{R}_Z)^{-1}\mathbf{R}'_Z\mathbf{y}}{[n - \text{posto}(\mathbf{X}) - \text{posto}(\mathbf{Z})]}. \quad (38)$$

Calculando a raiz quadrada dos elementos da diagonal de $\widehat{var}(\widehat{\Lambda\beta}_{PP})$ encontram-se os erros quadráticos das estimativas dos contrastes.

Conforme Cargnelutti Filho, Storck e Lúcio (2003), se forem desconsiderados os efeitos de blocos, ou seja, assumir que o experimento seja inteiramente casualizado, tem-se o modelo,

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \quad (39)$$

em que y_{ij} é o valor observado na j -ésima repetição que recebeu o i -ésimo tratamento, μ é a média geral, τ_i é o efeito do i -ésimo tratamento e ε_{ij} é o erro aleatório da (i, j) -ésima parcela, com possível autocorrelação entre os ε_{ij} .

O resíduo estimado de cada parcela, para este modelo é encontrado da seguinte forma,

$$r_{ij} = y_{ij} - \hat{\tau}_i, \quad (40)$$

em que r_{ij} e y_{ij} são, respectivamente, o resíduo e o valor observado da parcela com o i -ésimo tratamento e j -ésima repetição e $\hat{\tau}_i$ é o efeito estimado do i -ésimo tratamento.

O método de Papadakis, como sugerido pelos autores, pode então, ser usado neste caso, calculando a covariável índice ambiental Z da seguinte forma:

$$Z_{(l,c)} = \frac{r_{(l,c)} + r_{(l+1,c)} + r_{(l-1,c)} + r_{(l,c+1)} + r_{(l,c-1)}}{5} \quad (41)$$

sendo que cada unidade experimental é representada pelas coordenadas (l, c) , em que $l = 1, \dots, L$ é a coordenada que representa a l -ésima parte da área experimental dividida em L linhas e $c = 1, \dots, C$ é a coordenada que representa a c -ésima parte da área experimental dividida em C colunas.

Para as parcelas centrais será considerada uma estrutura de vizinhança de cinco parcelas, sendo a própria parcela, uma acima, uma abaixo, uma à esquerda e uma à direita. Já para as unidades experimentais das extremidades serão consideradas as parcelas disponíveis da estrutura citada.

Segue-se com a análise de covariância para o experimento inteiramente ao acaso, de acordo com o modelo de regressão:

$$y_{ij} = \bar{y}_i + \theta Z_{ij} + \varepsilon_{ij}, \quad (42)$$

sendo y_{ij} é o valor observado da j -ésima parcela que recebeu o i -ésimo tratamento, \bar{y}_i a média do i -ésimo tratamento, θ é o coeficiente de regressão entre a variável resposta e a covariável Z , Z_{ij} é a covariável Índice Ambiental medida na j -ésima parcela que recebeu o tratamento i e ε_{ij} o erro experimental associado à observação y_{ij} .

O vetor de estimativas dos contrastes de médias de tratamentos e a matriz de variâncias desses contrastes podem ser obtidos da mesma forma que em (35) e (37), respectivamente.

2.2.2.5 Medidas para comparar os métodos de estimação de contrastes entre tratamentos

As medidas utilizadas para comparar os métodos estudados serão:

- i. estimativas dos contrastes entre tratamentos;
- ii. diferenças absolutas;
- iii. medidas de dispersão;
- iv. erros quadráticos e
- v. intervalo de confiança.

2.2.2.5.1 Diferença absoluta

Para os dados em estudo serão calculadas as diferenças absolutas entre as estimativas dos contrastes, \hat{c}_j , e seus valores verdadeiros, c_j , dadas por

$$dif = |c_j - \hat{c}_j| \quad j = 1, \dots, 8. \quad (43)$$

2.2.2.5.2 Medida de dispersão

A medida de dispersão é calculada da seguinte maneira:

$$D = \sum_{j=1}^m (c_j - \hat{c}_j)^2, \quad (44)$$

em que, \hat{c}_j representa o valor do contraste estimado, c_j o verdadeiro valor do contraste e m o número de contrastes.

2.2.2.5.3 Erros quadráticos

Os erros quadráticos estimados serão calculados encontrando-se as raízes quadradas dos elementos das diagonais das matrizes de variâncias das estimativas dos contrastes entre tratamentos.

2.2.2.5.4 Intervalo de confiança

Os intervalos de confiança ao nível de 5% de significância para a análise clássica (ANOVA) e a análise de covariância proposta por Papadakis será expresso por:

$$IC(c_j) = \hat{c}_j \pm t_\alpha \sqrt{\widehat{var}(\hat{c}_j)}, \quad j = 1, \dots, 8, \quad (45)$$

em que, \hat{c}_j como em (44) e t_α é o t de *Student* tabelado ao nível α de significância.

Para os métodos de estimação dos parâmetros do modelo geoestatístico a distribuição a ser utilizada é a normal, sendo, portanto, o intervalo de confiança expresso por:

$$IC(c_j) = \hat{c}_j \pm z_\alpha \sqrt{\widehat{var}(\hat{c}_j)}, \quad j = 1, \dots, 8, \quad (46)$$

sendo, z_α o z tabelado da distribuição normal ao nível α de significância.

2.3 Resultados e Discussão

Para a comparação dos tratamentos do conjunto de dados deste estudo foram utilizados os métodos da análise da variância, do modelo geoestatístico e do modelo de Papadakis considerando o modelo em blocos casualizados e o método de Papadakis considerando que os dados fossem provenientes de um experimento inteiramente ao acaso.

A análise da variância (ANOVA Clássica) considera que os erros experimentais são independentes, o modelo geoestatístico modela a dependência espacial adicionando um componente de erros aleatórios dependentes e o modelo de Papadakis, realiza uma análise de covariância, sendo que a covariável, de natureza espacial, é a média dos resíduos entre as parcelas vizinhas e a própria parcela.

2.3.1 Efeitos dos Parâmetros

Utilizando os dados das concentrações de carbono orgânico (ANEXO A) coletados um ano após a implantação do experimento por Marciano (1999), foram estimados os efeitos de tratamentos e blocos utilizando o método de mínimos quadrados ordinários. Essas estimativas foram definidas como verdadeiros valores dos efeitos dos parâmetros, e foram adicionadas aos dados da caracterização inicial do solo, que é um ensaio em branco.

Esses efeitos estão apresentados na tabela 1.

Tabela 1 - Efeitos dos Parâmetros

τ_1	τ_2	τ_3	τ_4	τ_5	τ_6	τ_7	τ_8	τ_9	ω_1	ω_2	ω_3	ω_4
-1,27	-0,60	0,34	0,47	0,56	-1,07	-0,32	0,67	1,21	-0,41	0,10	0,06	0,25

2.3.2 Análise da Variância

Primeiramente, realizou-se a análise da variância clássica, para ver se os tratamentos são significativos. Tal análise encontra-se na tabela 2.

Como pode ser visto, os tratamentos são significativos, sendo necessária a comparação das médias dos tratamentos por meio de contrastes.

Tabela 2 - Análise da Variância (ANOVA) Clássica

Causa de variação	G.L.	SQ	QM	F	P-valor
Tratamentos	8	36,99	4,62	5,07	< 0,0001
Blocos	3	2,11	0,70	0,77	0,5217
Resíduos	24	21,90	0,91		
Total	35	61			

Observa-se, também, que o efeito de blocos não foi significativo o que se pode concluir que um bloco não difere significativamente do outro.

2.3.3 Testes de dependência espacial

Para analisar os dados deste trabalho, utilizando os métodos do variograma seguido de mínimos quadrados generalizados, da máxima verossimilhança e de Papadakis, é necessário que esses sejam dependentes quanto à posição das parcelas no experimento.

2.3.3.1 Modelo Geoestatístico

Utilizando os critérios de Informação de Akaike (AIC) e de Informação Bayesiano ou de Schwarz (BIC), testou-se a dependência espacial para o modelo geoestatístico (27). Esses métodos indicam como o melhor modelo aquele que apresentar o menor valor para o critério.

Os resultados encontrados, testando os modelos espacial (geoestatístico) e não-espacial, estão na tabela 3.

Tabela 3 - Critérios de Informação de Akaike e Bayesiano ou de Schwarz para o modelo geoestatístico

	AIC	BIC
Dependência Espacial	102,9	126,7
Não Dependência Espacial	110,3	130,9

Como observado na tabela, o melhor modelo, segundo os critérios de informação, é o modelo geoestatístico, mostrando que ao utilizar o componente aleatório deste modelo, que utiliza as estimativas dos parâmetros espaciais, é possível verificar dependência quanto as posições das

parcelas dentro dos blocos na área experimental. Assim, esse modelo pode ser usado para analisar os dados do experimento.

2.3.3.2 Modelo de Papadakis

Para o modelo (42), testou-se a significância do coeficiente θ , que é o coeficiente de regressão entre a variável resposta e a covariável Z . Esta indica se os dados têm ou não dependência espacial para o modelo de Papadakis.

O teste resultou num P-valor para a covariável que é a média dos resíduos entre a parcela e as parcelas vizinhas igual a 0,528, assim, esta covariável não é significativa para o modelo, ou seja, pelo método de Papadakis, os dados não têm dependência quanto as posições das parcelas dentro dos blocos.

De acordo com Cargnelutti Filho, Storck e Lúcio (2003), como blocos foi não significativo, pode-se considerar um experimento inteiramente ao acaso, apesar de o sorteio ter sido feito para um delineamento em blocos casualizados.

Novamente, foi testada a covariável para esse modelo, e o resultado do P-valor foi 0,73, indicando que a média dos resíduos entre a parcela e as parcelas vizinhas também não foi significativo ao modelo. Portanto, a correção feita para uma possível interferência das parcelas vizinhas também não tem efeito.

Mesmo assim, seguiu-se com a análise para ver o comportamento dos resultados e da comparação tanto para o modelo em blocos casualizados como para o modelo inteiramente ao acaso, tendo em vista a resposta de dependência obtida do modelo geoestatístico.

2.3.4 Estimativas dos Parâmetros

Utilizando os métodos de estimação estudados, foram obtidas as estimativas dos parâmetros que se encontram na tabela 4.

Observando esta tabela, pode-se notar que, para os dados das concentrações de Carbono Orgânico no solo, os efeitos de blocos são maiores que os efeitos de tratamentos para todos os métodos, e ainda, o método de Papadakis foi o que apresentou as menores estimativas dos parâmetros, seguido por mínimos quadrados ordinários, máxima verossimilhança e o método dos mínimos quadrados generalizados que foi o que apresentou as maiores estimativas. Os métodos ainda mostram certa coerência nos resultados, pois, em todos, as menores estimativas foram para os parâmetros τ_6 , referente ao efeito do tratamento 6 – Testemunha para o composto de lixo, e ω_1 ,

referente ao efeito do bloco 1, e as maiores estimativas para τ_9 , referente ao efeito do tratamento 9 – Composto de lixo, dose C, e ω_3 , referente ao efeito do bloco 3.

Tabela 4 - Resumo das estimativas dos parâmetros para os métodos em estudo

	$\hat{\beta}_{MQO}$	$\hat{\beta}_{MQG}$	$\hat{\beta}_{MV}$	$\hat{\beta}_{PP}$
$\hat{\tau}_1$	0,95	1,20	1,13	0,45
$\hat{\tau}_2$	2,25	2,00	1,97	1,07
$\hat{\tau}_3$	3,24	3,01	3,00	1,54
$\hat{\tau}_4$	2,92	3,34	3,27	1,39
$\hat{\tau}_5$	2,56	3,02	2,96	1,22
$\hat{\tau}_6$	0,65	0,75	0,71	0,31
$\hat{\tau}_7$	1,90	2,15	2,10	0,90
$\hat{\tau}_8$	3,42	3,03	3,02	1,62
$\hat{\tau}_9$	3,73	3,99	3,93	1,77
$\hat{\omega}_1$	5,16	5,37	5,27	2,45
$\hat{\omega}_2$	5,25	5,48	5,38	2,49
$\hat{\omega}_3$	5,79	6,02	5,92	2,75
$\hat{\omega}_4$	5,42	5,62	5,53	2,58

Para o caso em que foi considerado o modelo inteiramente ao acaso, as estimativas dos parâmetros para o método de Papadakis encontram-se na tabela 5.

Nesta tabela, observa-se que as estimativas para os efeitos de tratamentos, para esse método, foram sempre menores quando comparadas às estimativas para os métodos estudados no modelo em blocos ao acaso. Novamente, mostrando a coerência dos métodos, o de Papadakis, para esse modelo, também apresenta a menor estimativa para o parâmetro τ_6 – Testemunha para o composto de lixo, e a maior para o τ_9 – Composto de lixo, dose C.

Tabela 5 - Estimativas dos parâmetros para o método de Papadakis ($\hat{\beta}_{PP}$), considerando o modelo inteiramente ao acaso

$\hat{\tau}_1$	$\hat{\tau}_2$	$\hat{\tau}_3$	$\hat{\tau}_4$	$\hat{\tau}_5$	$\hat{\tau}_6$	$\hat{\tau}_7$	$\hat{\tau}_8$	$\hat{\tau}_9$
0,68	0,82	0,92	0,89	0,85	0,65	0,78	0,94	0,98

O método iterativo do variograma seguido de mínimos quadrados generalizados (MQG) convergiu a partir da segunda iteração e o semivariograma resultante é dado na figura

3, em que os parâmetros espaciais estimados são $\hat{\sigma}^2 = 1,7$, $\hat{\phi} = 192,7$ e $\hat{\tau}^2 = 0,3$, conforme descritos na figura 1.

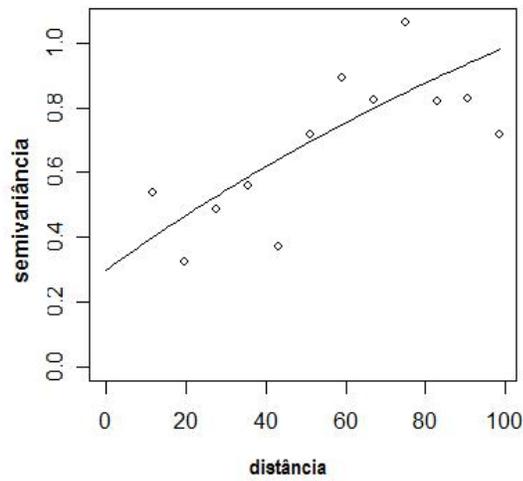


Figura 3- Semivariograma Empírico para os dados do carbono orgânico e a curva exponencial ajustada aos parâmetros espaciais

2.3.5 Estimativas dos Contrastes de Médias dos Tratamentos e Medida de Dispersão

A matriz de coeficientes (Λ) utilizada para estimar os contrastes é dada por:

$$\Lambda = \begin{bmatrix} 1 & -1/8 & -1/8 & -1/8 & -1/8 & -1/8 & -1/8 & -1/8 & -1/8 \\ 0 & 1/4 & 1/4 & 1/4 & 1/4 & -1/4 & -1/4 & -1/4 & -1/4 \\ 0 & 1 & -1/3 & -1/3 & -1/3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1/2 & -1/2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1/3 & -1/3 & -1/3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1/2 & -1/2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix}, \quad (47)$$

sendo que os contrastes são, a adubação mineral contra todos os outros tratamentos, a testemunha de cada tratamento contra as respectivas doses e os contrastes entre as diferentes doses de cada tratamento.

As estimativas dos contrastes podem ser comparadas aos verdadeiros valores dos mesmos pela medida de dispersão calculada por (44). Essas informações estão resumidas na tabela 6.

Observando os valores dessa tabela, verifica-se que o método que apresentou a maior

dispersão quanto às estimativas dos contrastes das médias de tratamentos foi o de Papadakis, em segundo está o de mínimos quadrados ordinários; logo em seguida o método do variograma seguido de mínimos quadrados generalizados e; por último, o método da máxima verossimilhança, que apresentou a menor dispersão dos métodos quanto às estimativas destes contrastes.

Tabela 6 - Valores verdadeiros dos contrastes (c_j), suas estimativas (\hat{c}_j) para os métodos em estudo e medidas de dispersão (D)

	$\Lambda\beta$	$\widehat{\Lambda\beta}_{MQO}$	$\widehat{\Lambda\beta}_{MQG}$	$\widehat{\Lambda\beta}_{MV}$	$\widehat{\Lambda\beta}_{PP}$
c_1/\hat{c}_1	-1,43	-1,63	-1,46	-1,49	-0,77
c_2/\hat{c}_2	0,08	0,32	0,36	0,36	0,15
c_3/\hat{c}_3	-1,06	-0,66	-1,13	-1,10	-0,31
c_4/\hat{c}_4	-0,18	0,50	-0,17	-0,12	0,24
c_5/\hat{c}_5	-0,09	0,36	0,32	0,31	0,17
c_6/\hat{c}_6	-1,59	-2,37	-2,31	-2,31	-1,12
c_7/\hat{c}_7	-1,27	-1,68	-1,35	-1,38	-0,80
c_8/\hat{c}_8	-0,54	-0,32	-0,96	-0,91	-0,15
D	—	1,73	0,95	0,91	1,82

No modelo inteiramente ao acaso, as estimativas dos contrastes para o método de Papadakis e a sua dispersão encontram-se na tabela 7.

Tabela 7 - Estimativas dos contrastes ($\widehat{\Lambda\beta}_{PP}$), e medida de dispersão (D) para o método de Papadakis considerando o modelo inteiramente ao acaso

\hat{c}_1	\hat{c}_2	\hat{c}_3	\hat{c}_4	\hat{c}_5	\hat{c}_6	\hat{c}_7	\hat{c}_8	D
-0,17	0,03	-0,07	0,05	0,04	-0,25	-0,18	-0,03	5,86

Pode-se observar que o método de Papadakis, considerando o experimento inteiramente ao acaso, apresentou dispersão muito maior que os métodos estudados para o modelo em blocos ao acaso.

2.3.6 Diferenças Absolutas

Na tabela 8, encontram-se as diferenças absolutas entre os contrastes estimados e os seus verdadeiros valores.

Tabela 8 - Diferenças absolutas entre os contrastes estimados e os seus verdadeiros valores

	MQO	MQG	MV	PP
$ c_1 - \hat{c}_1 $	0,20	0,04	0,06	0,65
$ c_2 - \hat{c}_2 $	0,24	0,29	0,28	0,08
$ c_3 - \hat{c}_3 $	0,40	0,07	0,04	0,75
$ c_4 - \hat{c}_4 $	0,68	0,00	0,06	0,41
$ c_5 - \hat{c}_5 $	0,45	0,41	0,40	0,26
$ c_6 - \hat{c}_6 $	0,77	0,71	0,72	0,47
$ c_7 - \hat{c}_7 $	0,41	0,09	0,11	0,47
$ c_8 - \hat{c}_8 $	0,22	0,42	0,37	0,39

Pode ser observado, nessa tabela, que o método de Papadakis é o que apresenta a maior variabilidade das diferenças absolutas para os contrastes. Dos oito contrastes estudados, este método contém três que apresentaram as maiores diferenças absolutas dos métodos em estudo e três que apresentaram as menores. A alta dispersão para o método de Papadakis, vista na seção anterior, na tabela 6, é justificada devido a essa variação.

Para o caso de experimento inteiramente ao acaso, as diferenças absolutas para o método de Papadakis, estão na tabela 9.

Tabela 9 - Diferenças absolutas para o método de Papadakis considerando o modelo inteiramente ao acaso

$ c_1 - \hat{c}_1 $	$ c_2 - \hat{c}_2 $	$ c_3 - \hat{c}_3 $	$ c_4 - \hat{c}_4 $	$ c_5 - \hat{c}_5 $	$ c_6 - \hat{c}_6 $	$ c_7 - \hat{c}_7 $	$ c_8 - \hat{c}_8 $
1,25	0,04	0,99	0,23	0,13	1,34	1,09	0,51

Nesta tabela, observa-se que as diferenças absolutas também têm variabilidade relativamente grande de um contraste para o outro, o que justifica a dispersão elevada apresentada na tabela 7.

2.3.7 Estimativas dos Erros Quadráticos das estimativas dos Contrastos

Encontraram-se as matrizes das variâncias estimadas dos contrastes estimados (ANEXO C) e, a partir delas, encontraram-se os valores dos erros quadráticos estimados que estão resumidos na tabela 10.

Tabela 10 - Estimativas dos erros quadráticos dos contrastes para os métodos em estudo

	$\hat{e}q_{MQO}$	$\hat{e}q_{MQG}$	$\hat{e}q_{MV}$	$\hat{e}q_{PP}$
\hat{c}_1	0,51	0,33	0,32	0,53
\hat{c}_2	0,34	0,21	0,21	0,35
\hat{c}_3	0,55	0,37	0,35	0,56
\hat{c}_4	0,58	0,41	0,39	0,67
\hat{c}_5	0,68	0,46	0,44	0,75
\hat{c}_6	0,55	0,37	0,35	0,56
\hat{c}_7	0,58	0,38	0,36	0,61
\hat{c}_8	0,68	0,46	0,43	0,68

Como pode ser observado nesta tabela, os menores valores foram para o método da máxima verossimilhança, logo em seguida está o método dos mínimos quadrados generalizados, na seqüência vem o dos mínimos quadrados ordinários e os maiores valores foram para o método de Papadakis.

Para o caso inteiramente ao acaso, encontrou-se a matriz das variâncias estimadas dos contrastes estimados (ANEXO D), e assim, os erros quadráticos para o método de Papadakis, que estão na tabela 11.

Tabela 11 - Estimativas dos erros quadráticos dos contrastes para o método de Papadakis ($\hat{e}q_{PP}$), considerando o modelo inteiramente ao acaso

\hat{c}_1	\hat{c}_2	\hat{c}_3	\hat{c}_4	\hat{c}_5	\hat{c}_6	\hat{c}_7	\hat{c}_8
0,51	0,34	0,56	0,60	0,69	0,56	0,59	0,68

Pode-se notar que esses erros quadráticos estão muito próximos dos erros quadráticos do método de Papadakis para o modelo em blocos casualizados. Isto mostra que as estimativas das variâncias dos contrastes estimados para o modelo em blocos não sofrem grandes alterações quando é retirado o efeito de blocos no modelo.

2.3.8 Intervalos de Confiança

Os intervalos com 95% de confiança para os contrastes entre médias de tratamentos encontram-se na tabela 12.

Tabela 12 - Intervalos de confiança para contrastes entre médias de tratamentos

	MQO		MQG		MV		PP	
	LI	LS	LI	LS	LI	LS	LI	LS
c_1	-2,67	-0,58	-2,11	-0,81	-2,11	-0,87	-1,87	0,32
c_2	-0,38	1,02	-0,06	0,78	-0,04	0,76	-0,58	0,88
c_3	-1,80	0,48	-1,85	-0,41	-1,79	-0,42	-1,47	0,85
c_4	-0,71	1,71	-0,97	0,63	-0,88	0,64	-1,14	1,62
c_5	-1,03	1,75	-0,58	1,22	-0,54	1,17	-1,39	1,73
c_6	-3,50	-1,23	-3,02	-1,59	-2,99	-1,62	-2,28	0,03
c_7	-2,88	-0,47	-2,10	-0,61	-2,09	-0,67	-2,05	0,46
c_8	-1,71	1,08	-1,86	-0,07	-1,76	-0,06	-1,56	1,26

Os métodos do modelo geoestatístico, que são, o variograma seguido de mínimos quadrados generalizados e a máxima verossimilhança, apresentaram contrastes de médias de tratamentos diferentes de zero, ao nível de 5% de significância em cinco dos oito intervalos de confiança. Com três intervalos de confiança indicando que os contrastes diferem de zero, está o método dos mínimos quadrados ordinários. O método de Papadakis não apresentou nenhum intervalo de confiança em que o zero não estivesse incluído. Portanto, os métodos do modelo geoestatístico foram os que se mostraram mais rigorosos em não rejeitar a igualdade dos contrastes a zero e o menos rigoroso foi o método de Papadakis. Este foi analisado e comparado como um modelo em que estivesse presente a dependência espacial, porém, a covariável que admite tal dependência, não se apresentou significativa quando testada.

Analisando a significância dos contrastes pelos intervalos de confiança, verifica-se que o método clássico, que não considera a dependência espacial dos dados, ou seja, pelo método dos mínimos quadrados ordinários, mostrou que os contrastes que diferem de zero, com 95% de confiança foram: o primeiro contraste, em que é testada a média da adubação mineral contra a média de todos os outros tratamentos; o sexto contraste, que testa a testemunha de composto de lixo contra as três doses utilizadas, e; o sétimo contraste, que testa a dose A do composto de lixo contra as suas outras duas doses.

Considerando que os dados têm dependência pelo modelo geoestatístico, analisando a significância dos contrastes pelos intervalos de confiança, verifica-se que os métodos espaciais detectaram que outros contrastes também diferem de zero. Além dos contrastes, que o método

de mínimos quadrados ordinários resultou em tal diferença, o método do variograma seguido de mínimos quadrados generalizados e o método da máxima verossimilhança concluíram também, que o terceiro contraste, que testa a média da testemunha para o lodo de esgoto contra as suas três doses e o oitavo contraste que testa a diferença entre as médias da dose B e da dose C do composto de lixo, diferem de zero, significativamente.

Portanto, conforme Cressie (1993), se for detectada a estrutura de autocorrelação e se essa informação for utilizada na análise estatística, estimativas mais eficientes dos contrastes entre as médias dos tratamentos são garantidas, mas se tal estrutura for desconsiderada pode impedir que diferenças reais sejam detectadas. Assim, nos dados do presente trabalho, quando não foi utilizada a informação espacial, um menor número de contrastes de médias de tratamentos foi significativamente diferente de zero do que quando foi utilizada tal informação.

O método de Papadakis não acusou essa diferença para nenhum contraste de médias de tratamentos, com 5% de significância, nem mesmo os que o método clássico acusou.

Finalmente, foram encontrados os intervalos de confiança dos contrastes com 95% de probabilidade para o método de Papadakis no modelo inteiramente ao acaso. Esses intervalos de confiança estão na tabela 13.

Tabela 13 - Intervalos de confiança dos contrastes para o método de Papadakis considerando o modelo inteiramente ao acaso

	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8
LI	-1,22	-0,66	-1,21	-1,18	-1,37	-1,39	-1,39	-1,43
LS	0,87	0,73	1,07	1,29	1,45	0,89	1,03	1,36

Nesta tabela, observa-se que, para esse caso, da mesma forma que no caso para o modelo em blocos casualizados, o zero também está incluído em todos os intervalos de confiança dos contrastes, e novamente, vale lembrar que a covariável média dos resíduos entre a parcela e as parcelas vizinhas não foi significativa.

3 CONCLUSÕES

De acordo com os objetivos deste trabalho, para os dados em estudo, foi detectada a presença de dependência espacial dentro dos blocos pelo modelo geoestatístico que é um modelo misto. Já pelo modelo de Papadakis, que é um modelo fixo, esta dependência não foi verificada tanto pelo modelo em blocos ao acaso quanto pelo modelo inteiramente ao acaso.

Utilizando a informação espacial dos dados, outros contrastes de médias de tratamentos, além dos encontrados pelo método dos mínimos quadrados ordinários, que não utiliza tal informação, foram indicados como significativamente diferentes de zero.

Isto não foi válido para o método de Papadakis, pois a covariável estudada, para os dados analisados, não foi significativa em nenhum modelo estudado.

Analisando as medidas utilizadas para comparar os métodos de estimação dos contrastes estudados, tem-se que, para os dados do carbono-orgânico, analisados neste trabalho, o método que apresentou maior discrepância em relação ao verdadeiro valor do contraste foi o da análise de covariância do método de Papadakis e a menor variação, foi para o método da máxima verossimilhança.

REFERÊNCIAS

- ANDRIOTTI, J.L.S. **Fundamentos de estatística e geoestatística**. São Leopoldo: UNISINOS, 2003. 165p.
- BARBIN, D. **Planejamento e Análise Estatística de Experimentos Agronômicos**. Arapongas: Midas, 2003. 194p.
- BESAG, J.; KEMPTON, R. Statistical Analysis of Field Experiments Using Neighbouring Plots. **Biometrics**, Arlington, v.42, n.2, p. 231-251, 1986.
- CARGNELUTTI FILHO, A.; STORCK, L.; LÚCIO, A.D. Ajustes de quadrado médio do erro em ensaios de competição de cultivares de milho pelo método de Papadakis. **Pesquisa agropecuária brasileira**, Brasília, v.38, n.4, p. 467-473, 2003.
- CRESSIE, N. Geostatistics. **The American Statistician**, Boston, v.43, n.4, p. 197-202, 1989.
- CRESSIE, N. **Statistics for Spatial Data**. New York: John Wiley, 1993. 900p.
- CULLIS, B.R.; GLEESON, A.C. Spatial Analysis of Field Experiments-An Extension to Two Dimensions. **Biometrics**, Arlington, v.47, n.4, p. 1449-1460, 1991.
- DIGGLE, P.J.; RIBEIRO Jr., P.J. **Model based geostatistics**. In: SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA, 14; Caxambu, 2000. 102p.
- DUARTE, J.B. **Sobre o emprego e a análise estatística do delineamento em blocos aumentados no melhoramento genético vegetal**. 2000. 293p. Tese (Doutorado) - Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2000.
- DUTILLEUL, P. Spatial Heterogeneity and the Design of Ecological Field Experiments. **Ecology**, Washington, v.74, n.6, p. 1646-1658, 1993.
- FISHER, R.A. **The design of experiments**. Edinburgh: Oliver and Boyd, 1935. 252p.
- FREITAS, E.G. **Análise de dados longitudinais em experimentos com cana-de-açúcar**. 2007. 74p. Dissertação (Mestrado) - Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2007.
- GRONDONA, M.O.; CRESSIE, N. Using spatial considerations in the analysis of experiments. **Technometrics**, Columbus, v.33, p. 381-392, 1991.
- KUAN, C.-M. **Statistics: Concepts and Methods**. 2nd. Taipei: Hua-Tai Publisher, 2004. 495p.
- MARCIANO, C.R. **Incorporação de resíduos urbanos e as propriedades físico-hídricas de um latossolo vermelho amarelo**. 1999. 93p. Tese (Doutorado) - Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 1999.
- NOGUEIRA, M.C.S. **Experimentação Agronômica I - conceitos, planejamento e análise estatística**. 1. Ed. Piracicaba: M.C.S. Nogueira, 2007. 463p.
- ORTIZ, J.L. **Geoestatística**. Disponível em: <<http://www.gpsglobal.com.br/Artigos/Geoestat.html>>. Acesso em: 10 mar. 2007.

- PILON, A.A. **Métodos para incorporação da dependência espacial na análise de dados experimentais**. 2004. 128p. Dissertação (Mestrado) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2004.
- PIRES, C.A.F.; STRIEDER, A.J. Modelagem geoestatística de dados geofísicos, aplicada a pesquisa de Au no prospecto Volta Grande (Complexo intrusivo Lavras do Sul, RS, Brasil). **Geomática**, Santa Maria, v.1, n.1, p. 43-55, 2006.
- PONTES, J.M.; OLIVEIRA, M.S. de Uma proposta alternativa para a análise de experimentos de campo utilizando a geoestatística. **Ciência agrotécnica**, Lavras, v.28, n.1, p. 135-141, 2004.
- R DEVELOPMENT CORE TEAM. R: A language and environment for statistical computing. **R Foundation for Statistical Computing**, Vienna, 2008.
- RIBEIRO JR, P.J.; DIGGLE, P.J. geoR: A package for geostatistical analysis. **R-News**, Vienna, v.1, n.2, p. 15-18, 2001.
- RIBEIRO JUNIOR, P.J. **Métodos estatísticos no estudo da variabilidade espacial de parâmetros do solo**. 1995. 99p. Dissertação (Mestrado) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 1995.
- VER HOEF, J.M.; CRESSIE, N. Spatial statistics: Analysis of field experiments. In SHEINER, S.M.; GUREVITCH, J. (Ed.). **Design and analysis of ecological experiments**. London: Chapman and Hall, 1993. p. 319-341.
- ZIMMERMAN, D.L.; HARVILLE, D.A. A Random Field Approach to the Analysis of Field-Plot Experiments and Other Spatial Experiments. **Biometrics**, Arlington, v.47, n.1, p. 223-239, 1991.

APÊNDICES

APÊNDICE A - Subrotina R para organização dos dados do Carbono Orgânico para o modelo em blocos ao acaso

```

rm(list = ls(all = TRUE))
require(geoR)
require(MASS)

# Leitura dos dados
todos.dados <- read.table("D:/.../c.organico.txt", head = T)

# Médias dos tratamentos e blocos
m.trat <- with(todos.dados, tapply(todos.dados$c.org.1o, trat, mean))
m.bloco <- with(todos.dados, tapply(todos.dados$c.org.1o, bloco, mean))

# Efeitos dos parâmetros dos dados do 1º ano
ef.trat <- m.trat[1:9] - mean(todos.dados$c.org.1o)
ef.bloco <- m.bloco[1:4] - mean(todos.dados$c.org.1o)
efeitos <- matrix(c(ef.trat, ef.bloco), c(13, 1))
todos.dados$ef.t <- ef.trat[todos.dados$trat]
todos.dados$ef.b <- ef.bloco[todos.dados$bloco]
todos.dados$ef.t <- as.matrix(todos.dados$ef.t)
todos.dados$ef.b <- as.matrix(todos.dados$ef.b)

# Dados do Ensaio em Branco (Caracterização Inicial do Solo) mais efeitos verdadeiros
todos.dados$dados <- todos.dados$c.org.ini + todos.dados$ef.t + todos.dados$ef.b
todos.dados$dados <- as.matrix(todos.dados$dados)

# Médias dos tratamentos e dos blocos dos novos dados
m.trat.dados <- with(todos.dados, tapply(todos.dados$dados, trat, mean))
m.bloco.dados <- with(todos.dados, tapply(todos.dados$dados, bloco, mean))

todos.dados$m.trat.dados <- m.trat.dados[todos.dados$trat]
todos.dados$m.bloco.dados <- m.bloco.dados[todos.dados$bloco]
todos.dados$m.trat.dados <- as.matrix(todos.dados$m.trat.dados)
todos.dados$m.bloco.dados <- as.matrix(todos.dados$m.bloco.dados)

```

```

# Matriz dos Coeficientes dos Contrastos
c.1 <- rbind(c( 1, -1/8, -1/8, -1/8, -1/8, -1/8, -1/8, -1/8, -1/8, 0, 0, 0, 0),
             c( 0, 1/4, 1/4, 1/4, 1/4, -1/4, -1/4, -1/4, -1/4, 0, 0, 0, 0),
             c( 0, 1, -1/3, -1/3, -1/3, 0, 0, 0, 0, 0, 0, 0, 0),
             c( 0, 0, 1, -1/2, -1/2, 0, 0, 0, 0, 0, 0, 0, 0),
             c( 0, 0, 0, 1, -1, 0, 0, 0, 0, 0, 0, 0, 0),
             c( 0, 0, 0, 0, 0, 1, -1/3, -1/3, -1/3, 0, 0, 0, 0),
             c( 0, 0, 0, 0, 0, 0, 1, -1/2, -1/2, 0, 0, 0, 0),
             c( 0, 0, 0, 0, 0, 0, 0, 1, -1, 0, 0, 0, 0))

```

```
dim(c.1)
```

```
c1.ginv <- ginv(c.1)
```

```
# Valores Verdadeiros dos Contrastos
```

```
cj <- c.1 %*% efeitos
```

```
round(cj, dig = 2)
```

```
# Lendo tratamentos e blocos como Fatores
```

```
todos.dados$trat <- as.factor(todos.dados$trat)
```

```
todos.dados$bloco <- as.factor(todos.dados$bloco)
```

APÊNDICE B - Subrotina R para análise clássica dos dados do Carbono Orgânico

```
# Modelo de Blocos ao Acaso sem Intercepto
```

```
mod1 <- lm(dados ~ trat + bloco - 1, data = todos.dados)
```

```
anova(mod1)
```

```
summary(mod1)
```

```
res <- residuals(mod1)
```

```
# Matriz do Delineamento
```

```
X <- model.matrix(mod1)
```

```
col <- matrix(c(rep(0, 27), rep(1, 9)), c(36, 1))
```

```
colnames(col) <- c("bloco1")
```

```
X1 <- cbind(X[, 1:9], col, X[, 10:12])
```

```
X1
```

```

# Estimativa de Mínimos Quadrados Ordinários
betaMQO <- ginv(crossprod(X1)) %*% crossprod(X1, todos.dados$dados)
round(betaMQO, dig = 2)

# Estimativa dos Contrastes de Mínimos Quadrados Ordinários
cBetaMQO <- c.1 %*% betaMQO
round(cBetaMQO, dig = 2)

# Diferença Absoluta
difMQO <- abs(cj - cBetaMQO)
round(difMQO, dig = 2)

# Medida de Dispersão (MQO)
dispersaoMQO <- sum(difMQO ^ 2)
round(dispersaoMQO, dig = 2)

# Resíduos 1
todos.dados$dados <- matrix(c(todos.dados$dados), c(36, 1))
todos.dados$res1 <- todos.dados$dados - X1 %*% betaMQO
round(sum(todos.dados$res1), dig = 2))
sig.MQO <- crossprod(todos.dados$res1)/(36 - mod1$rank)
draw0 <- NULL
sig.MQO <- c(draw0, sig.MQO)

# Variâncias Estimadas (MQO)
var.MQO <- sig.MQO * c.1 %*% ginv(crossprod(X1)) %*% t(c.1)
round(var.MQO, dig = 2)

# Erro Quadrático Estimado (MQO)
erro.quad.MQO <- sqrt(diag(var.MQO))
round(erro.quad.MQO, dig = 2)

# Intervalo de Confiança (MQO)
linfMQO <- cBetaMQO + qt(0.025, 24) * erro.quad.MQO
lsupMQO <- cBetaMQO + qt(0.975, 24) * erro.quad.MQO
cbind(round(linfMQO, dig = 2), round(lsupMQO, dig = 2))

```

APÊNDICE C - Subrotina R para testar a dependência espacial pelo modelo Geostatístico e normalidade dos dados do Carbono Orgânico

```
# Teste da Dependência Espacial
names(todos.dados)
c.organico.gd <- as.geodata(todos.dados, coords.col = 1:2, data.col = 9, covar.col = 3:4)

# Normalidade
boxcox(c.organico.gd, trend = ~ trat + bloco - 1, lam = seq(-2, 3, l = 1500))
shapiro.test(todos.dados$dados)

# Estimação dos parâmetros espaciais utilizados como valor inicial no ajuste da verossimilhança
v0 <- variog(c.organico.gd, trend = ~ trat + bloco - 1)
plot(v0)
li0 <- likfit(c.organico.gd, ini = c(0.7, 3), nug = 0.3, trend = ~ trat + bloco - 1)
li0
summary(li0)    # É Espacial (Modelo Geoestatístico)!
```

APÊNDICE D - Subrotina R para analisar os dados do Carbono Orgânico levando em consideração a dependência espacial do modelo Geoestatístico

```
# Método dos Mínimos Quadrados Generalizados (1ª Iteração)
names(todos.dados)
dad1 <- as.geodata(todos.dados, coords.col = 1:2, data.col = 12, covar.col = 3:4)
v1 <- variog(dad1, trend = ~ trat + bloco - 1)
plot(v1)
vf1 <- variofit(v1, nug = 0.3)
h <- as.matrix(dist(dad1$coords))
names(vf1)
theta0 <- vf1$nugget
theta1 <- vf1$cov.pars[1]
theta2 <- 1/vf1$cov.pars[2]
Cov <- theta0 * diag(36) + theta1 * exp( - theta2 * h)
```

```

# Estimativas dos Parâmetros (MQG1)
betaMQG1 <- ginv((crossprod(X1, solve(Cov))) %*% X1) %*% ((crossprod(X1,
      solve(Cov))) %*% todos.dados$dados)
round(betaMQG1, dig = 2)

# Estimativas dos Contrastes (MQG1)
cBetaMQG1 <- c.1 %*% betaMQG1
round(cBetaMQG1, dig = 2)

# Diferença Absoluta (MQG1)
difMQG1 <- abs(cj - cBetaMQG1)
round(difMQG1, dig = 2)

# Medida de Dispersão (MQG1)
dispersaoMQG1 <- sum(difMQG1^2)
round(dispersaoMQG1, dig = 2)

# Resíduos 2
todos.dados$res2 <- todos.dados$dados - X1 %*% betaMQG1
round(sum(todos.dados$res2), dig = 2)

# Variâncias Estimadas (MQG1)
var.MQG1 <- c.1 %*% ginv(crossprod(X1, solve(Cov)) %*% X1) %*% t(c.1)
round(var.MQG1, dig = 2)

# Erros Quadráticos Estimados (MQG1)
erro.quad.MQG1 <- sqrt(diag(var.MQG1))
round(erro.quad.MQG1, dig = 2)

# Intervalos de Confiança (MQG1)
linfMQG1 <- cBetaMQG1 + qnorm(0.025) * erro.quad.MQG1
lsupMQG1 <- cBetaMQG1 + qnorm(0.975) * erro.quad.MQG1
cbind(round(linfMQG1, dig = 2), round(lsupMQG1, dig = 2))

# Mínimos Quadrados Generalizados (2ª Iteração)
names(todos.dados)

```

```

dad2 <- as.geodata(todos.dados, coords.col = 1:2, data.col = 13, covar.col = 3:4)
v2 <- variog(dad2, trend = ~ trat + bloco - 1)
plot(v2)
vf2 <- variofit(v2, nug = 0.3)
names(vf2)
theta0 <- vf2$nugget
theta1 <- vf2$cov.pars[1]
theta2 <- 1/vf2$cov.pars[2]
Cov <- theta0 * diag(36) + theta1 * exp( - theta2 * h)

# Estimativas dos Parâmetros (MQG2)
betaMQG2 <- ginv((crossprod(X1, solve(Cov))) %*% X1) %*% ((crossprod(X1,
      solve(Cov))) %*% todos.dados$dados)
round(betaMQG2, dig = 2)

# Estimativas dos Contrastes (MQG2)
cBetaMQG2 <- c.1 %*% betaMQG2
round(cBetaMQG2, dig = 2)

# Diferenças Absolutas (MQG2)
difMQG2 <- abs(cj - cBetaMQG2)
round(difMQG2, dig = 2)

# Medida de Dispersão (MQG2)
dispersaoMQG2 <- sum(difMQG2^ 2)
round(dispersaoMQG2, dig = 2)

# Resíduos 3
todos.dados$res3 <- todos.dados$dados - X1 %*% betaMQG2
round(sum(todos.dados$res3), dig = 2)

# Variâncias Estimadas (MQG2)
var.MQG2 <- c.1 %*% ginv(crossprod(X1, solve(Cov)) %*% X1) %*% t(c.1)
round(var.MQG2, dig = 2)

# Erros Quadráticos Estimados (MQG2)

```

```

erro.quad.MQG2 <- sqrt(diag(var.MQG2))
round(erro.quad.MQG2, dig = 2)

# Intervalos de Confiança (MQG2)
linfMQG2 <- cBetaMQG2 + qnorm(0.025) * erro.quad.MQG2
lsupMQG2 <- cBetaMQG2 + qnorm(0.975) * erro.quad.MQG2
cbind(round(linfMQG2, dig = 2), round(lsupMQG2, dig = 2))

# Método da Máxima Verossimilhança (MV)
names(todos.dados)
gd <- as.geodata(todos.dados, data.col = 9, covar.col = 3:4)
ml <- likfit(gd, trend = ~ trat + bloco - 1, ini = c(1.5, 2.5))
ml
summary(ml)
theta0 <- ml$nugget
theta1 <- ml$cov.pars[1]
theta2 <- 1/ml$cov.pars[2]
G <- theta0 * diag(36) + theta1 * exp( - theta2 * h)
V <- G/theta1

# Estimativas dos Parâmetros (MV)
betaMV <- ginv(crossprod(X1, solve(G)) %*% X1) %*% crossprod(X1,
  solve(G)) %*% todos.dados$dados
round(betaMV, dig = 2)

# Estimativas dos Contrastes (MV)
cBetaMV <- c.1 %*% betaMV
round(cBetaMV, dig = 2)

# Diferenças Absolutas (MV)
difMV <- abs(cj - cBetaMV)
round(difMV, dig = 2)

# Medida de Dispersão (MV)
dispersaoMV <- sum(difMV^ 2)

```

```

round(dispersaoMV, dig = 2)

# Variabilidade
sig2.MV <- t(todos.dados$dados - X1 %*% betaMV) %*% solve(h) %*%
      (todos.dados$dados - X1 %*% betaMV)/36
draw1 <- NULL
sig2.MV <- c(draw1, sig2.MV)

# Variâncias Estimadas (MV)
var.MV <- sig2.MV * c.1 %*% ginv(crossprod(X1, solve(h)) %*% X1) %*% t(c.1)
round(var.MV, dig = 2)

# Erros Quadráticos Estimados (MV)
erro.quad.MV <- sqrt(diag(var.MV))
round(erro.quad.MV, dig = 2)

# Intervalos de Confiança (MV)
linfMV <- cBetaMV + qnorm(0.025) * erro.quad.MV
lsupMV <- cBetaMV + qnorm(0.975) * erro.quad.MV
cbind(round(linfMV, dig = 2), round(lsupMV, dig = 2))

```

APÊNDICE E - Subrotina R para testar a dependência espacial pelo método de Papadakis para o modelo em blocos casualizados

```

# Croqui do experimento
croqui <- matrix(c(todos.dados$dados[c(28:36, 19:27, 10:18, 1:9)]), c(4, 9), byrow=T)
croqui
res <- res[c(28:36, 19:27, 10:18, 1:9)]
layout1 <- matrix(c(res), c(4, 9), byrow=T)
layout1

# Médias dos resíduos das parcelas vizinhas
c1 <- matrix(c(rep(0, 4)), c(4, 1))
matrizres <- cbind(c1, layout1, c1)
nl <- 4

```

```

nc <- 9
i <- 1:nl
j <- 2:(nc + 1)
resi <- (matrizres[i, j - 1] + matrizres[i, j] + matrizres[i, j + 1])
somares <- matrix(resi, c(4, 9))
somares
A <- somares[i, 1]/2
B <- somares[i, 3:(nc - 1)]/3
C <- somares[i, nc]/2
media <- cbind(A, B, C)
medias <- rbind(media[4, ], media[3, ], media[2, ], media[1, ])
medias
draw2 <- NULL
med <- c(draw2, medias)
todos.dados$medias <- med

# Teste de Dependência Espacial do Modelo de Papadakis
mod11 <- lm(todos.dados$dados ~ - 1 + trat + bloco + medias, data = todos.dados)
summary(mod11) # Não é Espacial! (Modelo de Papadakis) – Blocos ao Acaso

```

APÊNDICE F - Subrotina R para analisar os dados do Carbono Orgânico levando em consideração a dependência espacial do método de Papadakis do modelo em blocos casualizados

```

# Modelo com a covariável Médias dos resíduos das parcelas vizinhas
names(todos.dados)
gd.pp <- as.geodata(todos.dados, data.col = 9, coords.col = 1:2, covar.col = 15)
y <- todos.dados$dados
n <- nl * nc
P <- diag(n) - (X1 %*% (ginv(crossprod(X1))) %*% t(X1))
b <- solve(crossprod(todos.dados$medias, P) %*% todos.dados$medias) %*%
  crossprod(todos.dados$medias, P) %*% y

# Estimativas dos Parâmetros (PP)

```

```

betaPP <- betaMQO %*% b
round(betaPP, dig = 2)

# Estimativas dos Contrastes (PP)
cBetaPP <- c.1 %*% betaPP
round(cBetaPP, dig = 2)

# Diferenças Absolutas (PP)
difPP <- abs(cj - cBetaPP)
round(difPP, dig = 2)

# Medida de Dispersão (PP)
dispersaoPP <- sum(difPP^ 2)
round(dispersaoPP, dig = 2)

# Vetor de resíduos do ajuste da covariável
Rz <- P %*% todos.dados$medias

# Variabilidade
sig2.PP <- (crossprod(y) - crossprod(y, X1) %*% ginv(crossprod(X1)) %*% crossprod(X1,
y) - crossprod(y, Rz) %*% solve(crossprod(Rz)) %*% crossprod(Rz, y))/(n -
mod1$rank - qr(todos.dados$medias)$rank)
draw3 <- NULL
sig2.PP <- c(draw3, sig2.PP)

# Variâncias Estimadas (PP)
var.PP <- c.1 %*% (ginv(crossprod(X1)) + ginv(crossprod(X1)) %*% crossprod(X1,
med) %*% solve(crossprod(Rz)) %*% crossprod(med,
X1) %*% ginv(crossprod(X1))) %*% t(c.1) * sig2.PP
round(var.PP, dig = 2)

# Erros Quadráticos Estimados (PP)
erro.quad.PP <- sqrt(diag(var.PP))
round(erro.quad.PP, dig = 2)

# Intervalos de Confiança (PP)

```

```

linfPP <- cBetaPP + qt(0.025, 24) * erro.quad.PP
lsupPP <- cBetaPP + qt(0.975, 24) * erro.quad.PP
cbind(round(linfPP, dig = 2), round(lsupPP, dig = 2))

```

APÊNDICE G - Subrotina R para organização dos dados do Carbono Orgânico para o modelo inteiramente ao acaso

```

rm(list=ls(all=TRUE))
require(geoR)
require(MASS)

# Leitura dos dados
todos.dados <- read.table("D:/.../c.organico.txt", head = T)

# Médias dos tratamentos
m.trat <- with(todos.dados, tapply(todos.dados$c.org.1o, trat, mean))

# Efeitos dos parâmetros dos dados do 1º ano
ef.trat <- m.trat[1:9] - mean(todos.dados$c.org.1o)
efeitos <- matrix(c(ef.trat), c(9, 1))
todos.dados$ef.t <- ef.trat[todos.dados$trat]
todos.dados$ef.t <- as.matrix(todos.dados$ef.t)

# Dados do Ensaio em Branco (Caracterização Inicial do Solo) mais efeitos verdadeiros
todos.dados$dados <- todos.dados$c.org.ini + todos.dados$ef.t
todos.dados$dados <- as.matrix(todos.dados$dados)

# Médias dos tratamentos dos novos dados
m.trat.dados <- with(todos.dados, tapply(todos.dados$dados, trat, mean))
todos.dados$m.trat.dados <- m.trat.dados[todos.dados$trat]
todos.dados$m.trat.dados <- as.matrix(todos.dados$m.trat.dados)

# Matriz dos Coeficientes dos Contrastes
c.1 <- rbind(c( 1, -1/8, -1/8, -1/8, -1/8, -1/8, -1/8, -1/8, -1/8),
             c( 0, 1/4, 1/4, 1/4, 1/4, -1/4, -1/4, -1/4, -1/4),
             c( 0, 1, -1/3, -1/3, -1/3, 0, 0, 0, 0),

```

```

c( 0, 0, 1, -1/2, -1/2, 0, 0, 0, 0),
c( 0, 0, 0, 1, -1, 0, 0, 0, 0),
c( 0, 0, 0, 0, 0, 1, -1/3, -1/3, -1/3),
c( 0, 0, 0, 0, 0, 0, 1, -1/2, -1/2),
c( 0, 0, 0, 0, 0, 0, 0, 1, -1))

dim(c.1)
c1.ginv <- ginv(c.1)

# Valores Verdadeiros dos Contrastes
cj <- c.1 %*% efeitos
round(cj, dig = 2)

# Lendo tratamentos como Fatores
todos.dados$trat <- as.factor(todos.dados$trat)

# Modelo Inteiramente ao Acaso sem Intercepto
mod1 <- lm(dados ~ trat - 1, data = todos.dados)
anova(mod1)
summary(mod1)

# Resíduos
res <- residuals(mod1)

# Matriz do Delineamento
X <- model.matrix(mod1)
X

```

APÊNDICE H - Subrotina R para testar a dependência espacial pelo método de Papadakis para o modelo inteiramente ao acaso

```

# Croqui do experimento com os resíduos
res <- res[c(28:36, 19:27, 10:18, 1:9)]
layout1 <- matrix(c(res), c(4, 9), byrow=T)
layout1

# Médias dos resíduos das parcelas vizinhas

```

```

c1 <- matrix(c(rep(0, 4)), c(4, 1))
meio <- cbind(c1, layout1, c1)
l1 <- matrix(c(rep(0, 11)), c(1, 11))
matrizres <- rbind(l1, meio, l1)
nl <- 4
nc <- 9
i <- 2:(nl + 1)
j <- 2:(nc + 1)
resi <- (matrizres[i + 1, j] + matrizres[i - 1, j] + matrizres[i, j] + matrizres[i,
      j - 1] + matrizres[i, j + 1])
somaes <- matrix(resi, c(4, 9))
somaes

A <- somaes[1, 1]/3
B <- somaes[1, 2:(nc - 1)]/4
C <- somaes[1, nc]/3
D <- somaes[2:(nl - 1), 1]/4
E <- somaes[2:(nl - 1), 2:(nc - 1)]/5
F <- somaes[2:(nl - 1), nc]/4
G <- somaes[nl, 1]/3
H <- somaes[nl, 2:(nc - 1)]/4
I <- somaes[nl, nc]/3

meio1 <- matrix(c(D, E, F), c((nl - 2), nc))
media <- rbind(c(A, B, C), meio1, c(G, H, I))

medias <- rbind(media[4, ], media[3, ], media[2, ], media[1, ])
medias

draw0 <- NULL
med <- c(draw0, medias)
todos.dados$medias <- med

# Teste de Dependência Espacial do Modelo de Papadakis
mod11 <- lm(todos.dados$dados ~ - 1 + trat + medias, data = todos.dados)

```

```
summary(mod11) # Não é Espacial! (Modelo de Papadakis) – Inteiramente ao Acaso
```

APÊNDICE I - Subrotina R para analisar os dados do Carbono Orgânico levando em consideração a dependência espacial do método de Papadakis do modelo inteiramente ao acaso

```
# Modelo com a covariável Média dos resíduos das parcelas vizinhas
names(todos.dados)
gd.pp <- as.geodata(todos.dados, data.col = 8, coords.col = 1:2, covar.col = 10)
y <- todos.dados$dados
n <- nl * nc
P <- diag(n) - (X %*% (ginv(crossprod(X))) %*% t(X))
b <- solve(crossprod(todos.dados$medias, P) %*% todos.dados$medias) %*%
  crossprod(todos.dados$medias, P) %*% y

# Estimativas dos Parâmetros (PP)
betaMQO <- ginv(crossprod(X)) %*% crossprod(X, todos.dados$dados)
betaPP <- betaMQO %*% b
round(betaPP, dig = 2)

# Estimativas dos Contrastes (PP)
cBetaPP <- c.1 %*% betaPP
round(cBetaPP, dig = 2)

# Diferenças Absolutas (PP)
difPP <- abs(cj - cBetaPP)
round(difPP, dig = 2)

# Medida de Dispersão (PP)
dispersaoPP <- sum(difPP^ 2)
round(dispersaoPP, dig = 2)

# Vetor de resíduos do ajuste da covariável
Rz <- P %*% todos.dados$medias

# Variabilidade
```

```

sig2.PP <- (crossprod(y) - crossprod(y, X) %*% ginv(crossprod(X)) %*% crossprod(X,
  y) - crossprod(y, Rz) %*% solve(crossprod(Rz)) %*% crossprod(Rz, y))/(n -
  mod1$rank - qr(todos.dados$medias)$rank)

draw1 <- NULL
sig2.PP <- c(draw1, sig2.PP)

# Variâncias Estimadas (PP)
var.PP <- c.1 %*% (ginv(crossprod(X)) + ginv(crossprod(X)) %*% crossprod(X,
  med) %*% solve(crossprod(Rz)) %*% crossprod(med, X) %*%
  ginv(crossprod(X))) %*% t(c.1) * sig2.PP
round(var.PP, dig = 2)

# Erros quadráticos Estimados (PP)
erro.quad.PP <- sqrt(diag(var.PP))
round(erro.quad.PP, dig = 2)

# Intervalos de Confiança (PP)
linfPP <- cBetaPP + qt(0.025, 27) * erro.quad.PP
lsupPP <- cBetaPP + qt(0.975, 27) * erro.quad.PP
cbind(round(linfPP, dig = 2), round(lsupPP, dig = 2))

```

ANEXOS

ANEXO A - Dados analisados no presente trabalho, referentes às concentrações de Carbono Orgânico no Solo (MARCIANO, 1999)

Tabela 14- Dados de concentrações de Carbono Orgânico no Solo (Blocos 4 e 3)

x	y	Tratamento	Bloco	Caracterização Inicial	1º Ano
5	4,9	8	4	8,8	11,93
17	4,9	3	4	8,1	10,68
29	4,9	7	4	8	10,12
41	4,9	2	4	7,1	9,31
53	4,9	9	4	8,2	10,85
65	4,9	5	4	7,7	9,52
77	4,9	6	4	6,3	8,66
89	4,9	1	4	6,9	8,27
101	4,9	4	4	7,1	10,14
5	16,9	3	3	9,5	10,05
17	16,9	2	3	10,6	9,24
29	16,9	8	3	7,6	8,56
41	16,9	5	3	8,7	12,22
53	16,9	1	3	7,8	8,33
65	16,9	7	3	7,1	8,56
77	16,9	9	3	7,1	11,35
89	16,9	4	3	7,5	10,74
101	16,9	6	3	7,3	8,72

Tabela 15 - Dados de concentrações de Carbono Orgânico no Solo (Blocos 2 e 1)

x	y	Tratamento	Bloco	Caracterização Inicial	1º Ano
5	28,9	2	2	8,2	9,7
17	28,9	9	2	8,2	11,37
29	28,9	1	2	8,7	8,8
41	28,9	6	2	6,5	8,31
53	28,9	3	2	7,8	9,83
65	28,9	4	2	7,6	9,9
77	28,9	5	2	6	9,7
89	28,9	7	2	7,1	9,41
101	28,9	8	2	7,8	11,17
5	40,9	6	1	8,4	8,8
17	40,9	8	1	8,4	9,8
29	40,9	4	1	9,2	9,9
41	40,9	7	1	8,3	9,39
53	40,9	1	1	7,1	8,31
65	40,9	2	1	7,1	8,13
77	40,9	9	1	8,2	10,05
89	40,9	5	1	7,2	9,6
101	40,9	3	1	7,8	9,6

ANEXO B - Gráfico Box-Cox dos dados do Carbono Orgânico

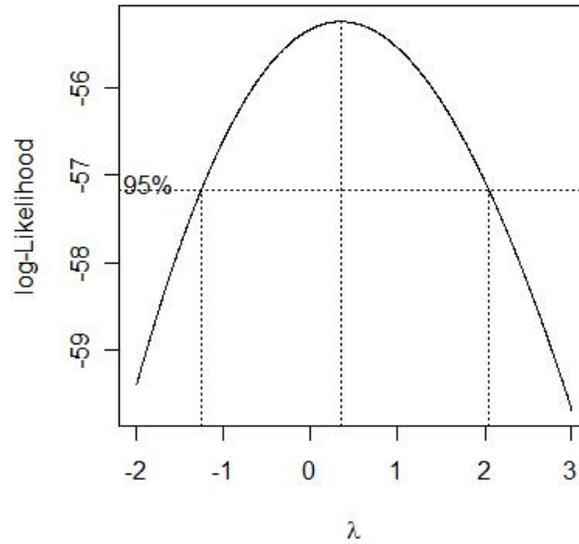


Figura 4 - Gráfico da transformação Box-Cox

ANEXO C - Matriz de variâncias estimadas dos contrastes estimados dos métodos em estudo considerando o modelo em blocos ao acaso para os dados do Carbono Orgânico

$$\widehat{var}(\widehat{\Lambda\beta}_{MQO}) = \begin{bmatrix} 0,26 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0,11 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0,3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0,34 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0,46 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0,3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,34 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,46 \end{bmatrix}$$

$$\widehat{var}(\widehat{\Lambda\beta}_{MQG}) = \begin{bmatrix} 0,11 & 0 & 0 & -0,01 & 0,01 & 0,01 & 0 & -0,01 \\ 0 & 0,05 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0,13 & 0,01 & 0 & 0 & -0,01 & 0 \\ -0,01 & 0 & 0,01 & 0,17 & -0,01 & -0,01 & -0,01 & 0,02 \\ 0,01 & 0 & 0 & -0,01 & 0,21 & 0,01 & 0 & 0,01 \\ 0,01 & 0 & 0 & -0,01 & 0,01 & 0,13 & 0 & 0 \\ 0 & 0 & -0,01 & -0,01 & 0 & 0 & 0,14 & -0,01 \\ -0,01 & 0 & 0 & 0,02 & 0,01 & 0 & -0,01 & 0,21 \end{bmatrix}$$

$$\widehat{var}(\widehat{\Lambda\beta}_{MV}) = \begin{bmatrix} 0,10 & 0 & 0 & -0,01 & 0,01 & 0 & 0 & 0 \\ 0 & 0,04 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0,12 & 0,01 & 0 & 0 & -0,01 & 0 \\ -0,01 & 0 & 0,01 & 0,15 & 0 & -0,01 & -0,01 & 0,02 \\ 0,01 & 0 & 0 & 0 & 0,19 & 0,01 & 0 & 0,01 \\ 0 & 0 & 0 & -0,01 & 0,01 & 0,12 & 0 & 0 \\ 0 & 0 & -0,01 & -0,01 & 0 & 0 & 0,13 & 0 \\ 0 & 0 & 0 & 0,02 & 0,01 & 0 & 0 & 0,19 \end{bmatrix}$$

$$\widehat{var}(\widehat{\Lambda\beta}_{PP}) = \begin{bmatrix} 0,20 & 0,01 & 0,01 & 0 & 0 & -0,01 & 0 & -0,01 \\ 0,01 & 0,10 & 0,02 & 0,01 & 0 & -0,01 & 0 & -0,03 \\ 0,01 & 0,02 & 0,26 & 0,01 & -0,01 & -0,02 & 0 & -0,05 \\ 0 & 0,01 & 0,01 & 0,26 & 0 & -0,01 & 0 & -0,02 \\ 0 & 0 & -0,01 & 0 & 0,34 & 0 & 0 & 0,01 \\ -0,01 & -0,01 & -0,02 & -0,01 & 0 & 0,24 & 0 & 0,03 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,26 & 0,01 \\ -0,01 & -0,03 & -0,05 & -0,02 & 0,01 & 0,03 & 0,01 & 0,42 \end{bmatrix}$$

ANEXO D - Matriz de variâncias estimadas dos contrastes estimados do método de Papadakis considerando o modelo inteiramente ao acaso para os dados do Carbono Orgânico

$$\widehat{var}(\widehat{\Lambda\beta}_{PP}) = \begin{bmatrix} 0,26 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0,12 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0,31 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0,36 & -0,01 & 0 & 0 & 0 \\ 0 & 0 & 0 & -0,01 & 0,47 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0,31 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0,35 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,46 \end{bmatrix}$$

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)