



**DIOGO FRANCISCO ROSSONI**

**ANÁLISE DE VARIÂNCIA PARA  
EXPERIMENTOS COM DEPENDÊNCIA  
ESPACIAL**

**LAVRAS - MG**

**2011**

**DIOGO FRANCISCO ROSSONI**

**ANÁLISE DE VARIÂNCIA PARA EXPERIMENTOS COM  
DEPENDÊNCIA ESPACIAL**

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

Orientador

Dr. Renato Ribeiro de Lima

**LAVRAS - MG**

**2011**

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

Rossoni, Diogo Francisco.

Análise de variância para experimentos com dependência espacial / Diogo Francisco Rossoni. – Lavras : UFLA, 2011.  
108 p. : il.

Dissertação (mestrado) – Universidade Federal de Lavras, 2011.

Orientador: Renato Ribeiro de Lima.

Bibliografia.

1. Modelo autorregressivo. 2. Modelagem geoestatística. 3.  
ANOVA-AR. I. Universidade Federal de Lavras. II. Título.

CDD – 519.538

**DIOGO FRANCISCO ROSSONI**

**ANÁLISE DE VARIÂNCIA PARA EXPERIMENTOS COM  
DEPENDÊNCIA ESPACIAL**

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

APROVADA em 18 de fevereiro de 2011.

Dr. José Marcio de Mello UFLA

Dr. João Domingos Scalon UFLA

Dr. Renato Ribeiro de Lima  
Orientador

**LAVRAS - MG**

**2011**

*Ao meu pai (in memoriam) que me ensinou com seu jeito carismático e olhar sábio, a importância de ser ético, justo e detentor de caráter.*

*A minha mãe, contínua fonte de inspiração, que no ardor das batalhas da vida sempre se portou como uma excepcional guerreira, repleta de coragem.*

**DEDICO**

## AGRADECIMENTOS

Aos meus pais, Arlindo (*in memoriam*) e Valdira pela pessoa que sou. Em especial a minha querida *mamma* que sempre me guiou sabiamente, seja através de gestos de carinho ou palavras de repreensão.

A minha irmã Camila, pelas ótimas conversas que sempre temos, sejam elas diálogos bem humorados ou homéricas discussões com base em argumentação, mostrando que irmãos podem além de tudo serem amigos.

A minha amada noiva Giceli, suporte de toda essa jornada, seja me desejando sucesso e uma singela boa noite (quando a distância era um obstáculo); seja abdicando de conquistas próprias em prol de nossas conquistas conjuntas.

Ao meu amigo e irmão Luciano Sturm, mais conhecido como Fritz, por provar que a amizade supera o tempo e a distância.

Ao meu orientador Renato Ribeiro de Lima, por todo conhecimento compartilhado e pela liberdade de trabalho que me propiciou.

Aos professores Daniel Ferreira Furtado, João Domingos Scalon e Lucas Monteiro Chaves, por sempre estarem dispostos e de boa vontade para o esclarecimento de dúvidas.

Ao professor Marcelo Silva de Oliveira e aos colegas Gerson e João Lousada pelo conhecimento sobre Geoestatística continuamente compartilhado.

Aos colegas que se tornaram grandes amigos: Jair (O senhor dos A+), Lucas (Panda Cunha), Felipe (Biblioteca Ambulante – livros, artigos e jogos) e Crysttian (Crysttian “viajando” Paixão).

Aos amigos de república Renato, Ricardo, Matheus e Wanda pelos ótimos momentos de descontração.

Aos meus sogros João e Mari, pela confiança depositada em minha pessoa.

A todos os colegas e professores da Pós-graduação em Estatística e Experimentação Agropecuária da UFLA e aos funcionários do Departamento de Ciências Exatas (DEX): Josinha, Josão, Selminha, Edila, Maria e Miriam.

A FAPEMIG pela concessão da bolsa de estudos, essencial para essa conquista.

## RESUMO

A influência da heterogeneidade do solo na estimação das diferenças entre os tratamentos é uma preocupação para os pesquisadores de campo. Uma das alternativas para controlar esse fator e aumentar a precisão das comparações entre os tratamentos é o estudo da dependência espacial das variáveis. Com o conhecimento das posições relativas das amostragens (dados referenciados), a variabilidade espacial passa a ser utilizada como um fator positivo, colaborando com as conclusões experimentais. Para o desenvolvimento deste estudo foram utilizados dados gerados a partir de simulação. Foram efetuadas comparações da análise não espacial (que considera os erros independentes) com a análise espacial, utilizando a Geoestatística (para a modelagem de erros correlacionados) e a análise de variância considerando o modelo autorregressivo (ANOVA-AR). A utilização de ferramentas da estatística espacial na análise de experimentos tornou a análise mais precisa, haja vista que houve diminuição na variabilidade geral do experimento e/ou uma melhor modelagem do experimento.

Palavras-chave: Modelo autorregressivo. Modelagem geoestatística. ANOVA-AR.

## **ABSTRACT**

The influence of the soil heterogeneity in the estimation of the treatments effect is a concern for field researchers. The accuracy of comparisons between the treatments can be increased by considering the soil heterogeneity in the analysis. An alternative to this is considering the spatial dependence of the experimental plots. In this study, it was used simulated data. The comparison between non-spatial analysis (independent errors) and spatial analysis was implemented. In the spatial analysis, it was considered the Geostatistics and the ANOVA with autoregressive model (ANOVA-AR). The analyses by considering the spatial information increased the comparisons accuracy. Also, there was a decrease in the experiment error.

Keywords: Autoregressive model. Geostatistics modeling. ANOVA-AR.

## LISTA DE ILUSTRAÇÕES

Figura 1	Padrões de proximidade apresentados por Gumpertz et al. (1997), em que  representa uma parcela de referência,  representa as parcelas vizinhas consideradas na comparação e  representa as demais parcelas.....	43
Figura 2	Gríde regular com 9 observações. ....	45
Figura 3	Krigagem para o experimento (i) na configuração Gaus(0-1-6)	63
Gráfico 1	Amostragem em duas dimensões .....	27
Gráfico 2	Exemplo de Semivariograma .....	29
Gráfico 3	Modelos teóricos de semivariograma.....	32
Gráfico 4	Anisotropia Geométrica .....	33
Gráfico 5	Anisotropia Zonal.....	34
Gráfico 6	Anisotropia Combinada.....	34
Gráfico 7	Semivariograma do exemplo de abordagem Geoestatística.....	60
Gráfico 8	Histogramas do QME da configuração Gaus(0-1-6).....	68
Gráfico 9	Histogramas do QMT da configuração Gaus(0-1-6).....	69
Gráfico 10	Histogramas do QMB da configuração Gaus(0-1-6).....	71
Gráfico 11	Histogramas do AIC da configuração Gaus(0-1-6).....	73
Gráfico 12	Histogramas do QME para a configuração Gaus(0-1-2).....	81
Gráfico 13	Histogramas do QMT para a configuração Gaus(0-1-2).....	82
Gráfico 14	Histogramas do QMB para a configuração Gaus(0-1-2).....	83
Gráfico 15	Histogramas do AIC para a configuração Gaus(0-1-2).....	85

## LISTA DE TABELAS

Tabela 1	Análise de variância para o delineamento inteiramente casualizado (DIC).....	21
Tabela 2	Análise de variância para o delineamento em blocos casualizado (DBC) .....	23
Tabela 3	Padrões de proximidade em função do raio .....	44
Tabela 4	Análise de variância de modelo autorregressivo (ANOVA-AR).....	50
Tabela 5	Efeitos fixos dos parâmetros utilizados na simulação.....	53
Tabela 6	Parâmetros do semivariograma para a simulação do erro experimental.....	54
Tabela 7	Parâmetros do modelo com erro independente .....	58
Tabela 8	Parâmetros estimados dos modelos de semivariograma .....	60
Tabela 9	Parâmetros do modelo com erro correlacionado.....	61
Tabela 10	Critério de Akaike para o modelo 1 e modelo 2.....	62
Tabela 11	Estimativas de $\rho$ para os padrões de proximidade.....	64
Tabela 12	Análise de variância clássica (ANOVA).....	65
Tabela 13	Análise de variância com modelo autorregressivo de primeira ordem (ANOVA-AR1).....	65
Tabela 14	Análise de variância com modelo autorregressivo de segunda ordem (ANOVA-AR2).....	65
Tabela 15	Análise de variância com modelo autorregressivo de terceira ordem (ANOVA-AR3).....	66
Tabela 16	Intervalo de credibilidade da abordagem autorregressiva para Gaus(0-1-6), com 95% de probabilidade de conter o verdadeiro valor do parâmetro. ....	71

Tabela 17	Intervalo de credibilidade para o AIC da abordagem Geoestatística para Gaus(0-1-6), com 95% de probabilidade de conter o verdadeiro valor do parâmetro.....	74
Tabela 18	Intervalo de credibilidade da abordagem autorregressiva para todas as configurações de Gaus no alcance 6, com 95% de probabilidade de conter o verdadeiro valor do parâmetro.....	74
Tabela 19	Intervalo de credibilidade da abordagem autorregressiva para todas as configurações de Gaus no alcance 4, com 95% de probabilidade de conter o verdadeiro valor do parâmetro.....	76
Tabela 20	Intervalo de credibilidade da abordagem autorregressiva para todas as configurações de Gaus no alcance 2, com 95% de probabilidade de conter o verdadeiro valor do parâmetro.....	77
Tabela 21	Intervalo de credibilidade para o AIC de todas as configurações do experimento i, com 95% de probabilidade de conter o verdadeiro valor do parâmetro..	78
Tabela 22	Intervalo de credibilidade da abordagem autorregressiva para Gaus(0-1-2), com 95% de probabilidade de conter o verdadeiro valor do parâmetro. ....	84
Tabela 23	Intervalo de credibilidade para o AIC da abordagem Geoestatística para Gaus(0-1-2), , com 95% de probabilidade de conter o verdadeiro valor do parâmetro.....	86
Tabela 24	Intervalo de credibilidade da abordagem autorregressiva para todas as configurações de Gaus no alcance 2, , com 95% de probabilidade de conter o verdadeiro valor do parâmetro.....	87
Tabela 25	Intervalo de credibilidade da abordagem autorregressiva para todas as configurações de Gaus no alcance 2, com 95% de probabilidade de conter o verdadeiro valor do parâmetro.....	88

Tabela 26	Intervalo de credibilidade para o AIC de todas as configurações do experimento ii, com 95% de probabilidade de conter o verdadeiro valor do parâmetro.....	90
Tabela 27	Conjunto de dados utilizados no exemplo de abordagem Geoestatística. ....	96
Tabela 28	Conjunto de dados utilizados no exemplo da abordagem autorregressiva. ....	99

## SUMÁRIO

1	INTRODUÇÃO .....	15
2	REFERENCIAL TEÓRICO .....	17
2.1	Estatística experimental.....	17
2.1.1	Princípios básicos da experimentação .....	17
2.1.2	Análise de variância .....	19
2.1.3	Delineamentos experimentais.....	20
2.1.3.1	Delineamento inteiramente casualizado (DIC) .....	20
2.1.3.2	Delineamento em blocos casualizados (DBC) .....	21
2.2	Geoestatística .....	23
2.2.1	Descrição e modelagem da estrutura de variabilidade espacial. ....	25
2.2.2	Semivariância e semivariograma .....	27
2.2.3	Parâmetros do semivariograma .....	29
2.2.4	Métodos de estimação dos parâmetros do semivariograma .....	30
2.2.5	Modelos teóricos de semivariograma.....	31
2.2.6	Isotropia e Anisotropia .....	32
2.2.7	Krigagem .....	35
2.2.7.1	Krigagem ordinária.....	36
2.2.7.2	Determinação dos pesos na krigagem ordinária.....	37
2.2.8	Validação cruzada .....	39
2.2.9	Modelagem Geoestatística do erro.....	40
2.3	Modelo de resposta autorregressiva .....	42
2.3.1	Modelo espacial autorregressivo (SAR) .....	42
2.3.2	Estimação do parâmetro $\rho$ .....	48
2.3.3	Análise de variância de modelo autorregressivo (ANOVA-AR) .....	48
3	MATERIAL E MÉTODOS.....	51
3.1	Simulação de dados .....	51
3.2	Abordagem Geoestatística .....	56
3.3	Abordagem autorregressiva .....	57
4	RESULTADOS E DISCUSSÃO .....	58
4.1	Exemplo de abordagem Geoestatística.....	58
4.2	Exemplo de abordagem autorregressiva .....	64
4.3	Dados simulados .....	67
4.3.1	Experimento na configuração (i).....	67
4.3.1.1	Estudo da configuração Gaus (0-1-6) .....	67
4.3.1.2	Estudo dos demais casos .....	74
4.3.2	Experimento na configuração ii .....	80
4.3.2.1	Estudo da configuração Gaus (0-1-2) .....	80
4.3.2.2	Estudo dos demais casos .....	86
5	CONCLUSÃO .....	91

<b>REFERÊNCIAS .....</b>	<b>93</b>
<b>ANEXO A.....</b>	<b>96</b>
<b>ANEXO B.....</b>	<b>102</b>

## 1 INTRODUÇÃO

O maior desafio quando se conduz um experimento é o de comparar os tratamentos com a maior precisão possível, para se ter segurança nas inferências a serem feitas a partir dos resultados alcançados.

A precisão de um experimento está totalmente ligada a pequenas variações nas unidades experimentais, antes de serem aplicados os tratamentos, ou induzidas involuntariamente, durante a execução do experimento, que causam heterogeneidade entre as parcelas, também conhecida como variação do acaso, variação ambiental ou, simplesmente, erro experimental (STORCK et al., 2000).

Quando o erro experimental é muito grande, os testes estatísticos utilizados para comparar as médias de tratamentos podem ser influenciados, impedindo que algumas diferenças entre médias de tratamentos sejam detectadas. Tendo isso em vista, a experimentação agrícola requer cada vez mais a utilização de técnicas experimentais mais refinadas. Nesse refinamento estão envolvidos diversos fatores, tais como: escolha criteriosa do delineamento, locais de instalação dos experimentos, número de repetições, tamanho e forma da parcela e uma perfeita condução dos experimentos. Todos esses fatores influenciam de forma direta ou indireta na precisão experimental.

Com o desenvolvimento computacional, alternativas tornaram-se disponíveis aos pesquisadores, como por exemplo, os métodos de análise espacial de experimentos, que em certas situações podem proporcionar resultados satisfatórios. Esses métodos se baseiam na análise de vizinhança, ou seja, na informação proveniente de parcelas vizinhas ou modelam a dependência espacial em função da distância entre parcelas, visando o controle da heterogeneidade espacial.

As formas de análises baseadas em modelagem espacial são mais apropriadas para experimentos em que é detectada a dependência espacial entre os erros experimentais, pois, nesses casos, as eficiências dos estimadores dos contrastes dos tratamentos não dependerão apenas da variação residual, mas também das posições das parcelas.

Para o desenvolvimento deste estudo foram utilizados dados gerados a partir de simulação. Será efetuada uma comparação da análise não espacial (erros independentes) com a análise espacial, utilizando a Geoestatística (erros correlacionados) e a ANOVA com modelo autorregressivo (ANOVA-AR).

## **2 REFERENCIAL TEÓRICO**

O referencial teórico abordará três temas principais: estatística experimental, abordagem Geoestatística e abordagem autorregressiva.

### **2.1 Estatística experimental**

Para se compreender a estatística dita como experimental, é necessário inicialmente entender quais os princípios que regem sua aplicação.

#### **2.1.1 Princípios básicos da experimentação**

A pesquisa científica está constantemente se utilizando de experimentos para provar suas hipóteses. Mesmo havendo diferenças entre as pesquisas, todas compartilham de alguns princípios básicos, que são necessários para que as conclusões que venham a ser obtidas se tornem válidas.

Segundo Banzatto e Kronka (2006) podemos eleger três princípios básicos da experimentação:

**Repetição** - Ao compararmos, por exemplo, dois herbicidas (A e B), aplicados em duas parcelas perfeitamente iguais, apenas o fato do herbicida A ter apresentado melhor controle que o herbicida B não é suficiente para que possamos concluir que o mesmo é mais eficiente, pois esse seu melhor controle poderá ter ocorrido por simples acaso ou ter sido influenciado por fatores estranhos. Porém, se os dois herbicidas forem aplicados a várias parcelas e, ainda assim verificarmos que o herbicida A apresenta, em média, maior controle, existe já um indício de que ele seja mais eficiente.

O princípio da repetição consiste na reprodução do experimento básico e tem por finalidade propiciar a obtenção de uma estimativa do erro experimental.

Casualização - Mesmo reproduzindo o experimento básico, poderá ocorrer que o herbicida A apresentou maior controle por ter sido favorecido por qualquer fator, como por exemplo, ter todas as suas parcelas agrupadas numa faixa de menor infestação.

Para evitar que um dos herbicidas seja sistematicamente favorecido por qualquer fator externo, procede-se a casualização dos herbicidas às parcelas, isto é, eles são designados às unidades experimentais de forma totalmente casual.

O princípio da casualização tem por finalidade propiciar a todos os tratamentos a mesma probabilidade de ser designados a qualquer das unidades experimentais e conseqüentemente proporcionar uma estimativa válida para o erro experimental.

Princípio de controle local - Este princípio é frequentemente utilizado, mas não é de uso obrigatório, pois se podem realizar experimentos sem utilizá-lo. Ele consiste em aplicar os herbicidas sempre em pares de parcelas (para o caso citado anteriormente) os mais homogêneos possíveis com relação ao ambiente, podendo haver, inclusive, variação acentuada de um ambiente para outro. A cada par de parcelas denominamos bloco.

Quando se têm diversos tratamentos a comparar, cada bloco será constituído por um grupo de parcelas que deve ser múltiplo do número de tratamentos.

A finalidade do princípio de controle local é dividir um ambiente heterogêneo em sub-ambientes homogêneos e tornar o delineamento experimental mais eficiente, pela redução do erro experimental.

### 2.1.2 Análise de variância

A análise de variância (ANOVA), segundo Yandell (1997), é utilizada na estatística experimental para determinar a variabilidade da variável resposta, ocasionada pelo efeito dos níveis de um fator.

Os fatores incluem características da experiência que podem estar sob o controle do pesquisador, como drogas ou tratamentos dietéticos e características que ocorrem naturalmente e podem ou não ser de interesse direto, como sexo, espécie e localização. Fatores que geralmente têm vários níveis - tratamentos e medicamentos de controle, sexo masculino ou feminino, cultivares de uma espécie, etc. Às vezes é útil considerar uma combinação de fatores examinando um nível de cada fator.

A variabilidade do experimento que não pode ser explicada pelos fatores é atribuída ao erro ao acaso, isto é, a variabilidade do experimento que não pode ser controlada.

Para que esse erro seja mínimo, ou seja, para que a precisão do experimento seja a maior possível, alguns cuidados devem ser tomados pelo pesquisador.

Banzatto e Kronka (2006) sugerem alguns procedimentos que podem levar a um aumento dessa precisão:

- a) Escolha do material experimental;
- b) Seleção das unidades experimentais;
- c) Seleção dos tratamentos;
- d) Aumento do número de repetições;
- e) Técnicas mais refinadas.

### 2.1.3 Delineamentos experimentais

Experimentos (testes) são realizados por pesquisadores em todos os campos de investigação, usualmente para descobrir algo sobre um particular processo ou sistema.

Um experimento delineado é um teste ou séries de testes no quais propositadamente mudanças são feitas para as variáveis de entrada de um processo ou sistema tal que podemos observar e identificar razões para mudanças na variável resposta.

A seguir serão apresentados os dois principais delineamentos experimentais, sob a hipótese de dados balanceados, isto é, sem parcelas perdidas.

#### 2.1.3.1 Delineamento inteiramente casualizado (DIC)

Suponha que se tenha interesse, em geral, de se comparar  $i$  tratamentos. O modelo estatístico que aborda esse delineamento é dado por:

$$y_{ij} = \mu + t_i + e_{ij} \quad (1)$$

em que:

$y_{ij}$  é a observação do  $i$ -ésimo tratamento na  $j$ -ésima observação, com

$i = 1, 2, \dots, a$  e  $j = 1, 2, \dots, r$ ;

$\mu$  é uma constante inerente a cada observação;

$t_i$  é o efeito do  $i$ -ésimo tratamento;

$e_{ij}$  é o erro associado a cada observação.

Tabela 1 Análise de variância para o delineamento inteiramente casualizado (DIC).

<i>FV</i>	<i>GL</i>	<i>SQ</i>	<i>QM</i>	<i>F</i>
Tratamentos	$a - 1$	$SQ_{Trat}$	$QM_{Trat} = SQ_{Trat} / a$	$QM_{Trat} / QME$
Erro	$a(r - 1)$	$SQE$	$QME / ar - 1$	
Total	$ar - 1$	$SQT$		

Segundo Montgomery (2008), em um experimento em DIC tem-se interesse em comparar os efeitos dos tratamentos. Logo, a hipótese de interesse é:

$$\begin{aligned}
 H_0 : t_1 = t_2 = \dots = t_a = 0 \\
 H_1 : t_i \neq 0 \text{ para pelo menos um } i.
 \end{aligned}
 \tag{2}$$

A tabela de análise de variância (ANOVA) para o DIC é apresentada na Tabela 1, em que:

$$SQ_{T}(\text{Soma quadrada total}) = \sum_{i=1}^a \sum_{j=1}^r y_{ij}^2 - \frac{y_{..}^2}{ar}$$

$$SQ_{Trat}(\text{Soma quadrada do tratamento}) = \frac{1}{r} \sum_{i=1}^a y_{i.}^2 - \frac{y_{..}^2}{ar}$$

$$SQE(\text{Soma quadrada do erro}) = SQT - SQ_{Trat}$$

### 2.1.3.2 Delineamento em blocos casualizados (DBC)

Suponha que se tenha interesse, em geral, de se comparar  $a$  tratamentos e devido ao controle local seja necessária a construção de  $b$  blocos. O modelo estatístico que aborda esse delineamento é dado por:

$$y_{ij} = \mu + t_i + b_j + e_{ij} \quad (3)$$

em que:

$y_{ij}$  é a observação do  $i$ -ésimo tratamento na  $j$ -ésima observação, com

$i = 1, 2, \dots, a$  e  $j = 1, 2, \dots, b$ ;

$\mu$  é uma constante inerente a cada observação;

$t_i$  é o efeito do  $i$ -ésimo tratamento;

$b_j$  é o efeito do  $j$ -ésimo bloco;

$e_{ij}$  é o erro associado a cada observação.

Segundo Montgomery (2008), em um experimento em DBC tem-se interesse em comparar os efeitos dos tratamentos. Logo, a hipótese de interesse é

$$\begin{aligned} H_0 : t_1 = t_2 = \dots = t_a = 0 \\ H_1 : t_i \neq 0 \text{ para pelo menos um } i \end{aligned} \quad (4)$$

Pode-se também comparar os efeitos dos blocos, logo, a hipótese de interesse é

$$\begin{aligned} H_0 : b_1 = b_2 = \dots = b_b = 0 \\ H_1 : b_j \neq 0 \text{ para pelo menos um } j \end{aligned} \quad (5)$$

A tabela de análise de variância (ANOVA) para o DBC é apresentada na Tabela 2.

Tabela 2 Análise de variância para o delineamento em blocos casualizado (DBC).

<i>FV</i>	<i>GL</i>	<i>SQ</i>	<i>QM</i>	<i>F</i>
Blocos	$b-1$	$SQB$	$QMB = SQB/b$	$QMB/QME$
Tratamentos	$a-1$	$SQTrat$	$QMTrat = SQTrat/a$	$QMTrat/QME$
Erro	$(a-1)(b-1)$	$SQE$	$QME/(a-1)(b-1)$	
Total	$ab-1$	$SQT$		

em que

$$SQT(\text{Soma quadrada total}) = \sum_{i=1}^a \sum_{j=1}^b y_{ij}^2 - \frac{y_{..}^2}{ab}$$

$$SQB(\text{Soma quadrada dos blocos}) = \frac{1}{a} \sum_{j=1}^b y_{.j}^2 - \frac{y_{..}^2}{ab}$$

$$SQTrat(\text{Soma quadrada dos tratamentos}) = \frac{1}{b} \sum_{i=1}^a y_{i.}^2 - \frac{y_{..}^2}{ab}$$

$$SQE(\text{Soma quadrada dos erros}) = SQT - SQB - SQTrat$$

## 2.2 Geoestatística

A variabilidade espacial tem sido uma das preocupações de pesquisadores desde o início do século. Smith (1910) considerou a disposição de parcelas no campo, em experimentos de rendimento de variedades de milho como uma tentativa de eliminar o efeito de variações do solo. Waynick (1918) ressaltou a importância da variabilidade espacial no estudo em Ciências do Solo.

Os procedimentos usados na época baseavam-se na estatística clássica ou de Fisher, que se utilizava de parâmetros como média e desvio padrão para representar um fenômeno, assumindo a hipótese principal de que as variações de um local para outro eram aleatórias.

Trabalhando com dados de concentração de ouro, Krige (1951) concluiu que somente a informação dada pela variância é insuficiente para explicar o fenômeno em estudo. Para tal, é necessário levar em consideração a distância entre as observações, que leva em consideração a localização geográfica e a dependência espacial.

A teoria das variáveis regionalizadas, também denominada Geoestatística devido à sua aplicação nos estudos de Geologia e Mineralogia, foi um importante passo no conhecimento humano. Segundo Guerra (1988), entendem-se como variáveis regionalizadas (VR), uma função que varia de um lugar a outro no espaço com certa aparência de continuidade. São variáveis cujos valores estão relacionados de algum modo com a posição espacial que ocupam.

A continuidade atribuída às VR se manifesta, sobretudo, pela tendência de tomarem valores mais parecidos, quanto menos afastados estejam dois pontos amostrados.

As peculiaridades dessas variáveis ferem o pressuposto básico da técnica de estatística que não considera a distribuição espacial, pois pressupõem que existe independência entre as observações. Segundo Clark (1979), uma medida tomada em um ponto guarda relações de dependência com medidas tomadas em pontos adjacentes sugerindo uma estrutura de correlação.

Desse modo, a Geoestatística se propõe a estudar dois objetivos principais (JOURNAL; HUIJBREGTS, 1978): ser capaz de extrair da aparente desordem dos dados disponíveis, uma imagem da variabilidade dos mesmos e uma medida da correlação existente entre os valores tomados em dois pontos do espaço, que é feita por meio do semivariograma; e ser capaz de medir a precisão

de toda predição ou estimação através de dados fragmentados, ou seja, através de um processo discreto de amostragem, obter uma estimativa contínua da área amostrada através da técnica de krigagem.

### **2.2.1 Descrição e modelagem da estrutura de variabilidade espacial**

Para a aplicação da teoria espacial deve-se considerar que a associação das variáveis em pontos distintos é maior à medida que estes pontos estejam mais próximos um do outro, a fim de descrever e modelar estas relações entre distâncias e associação espacial (CLARK, 1979).

Várias medidas se prestam para a descrição dessa relação, tais como a autocovariância e autocorrelação, usuais na análise de séries temporais. Em Geoestatística a medida normalmente utilizada é a semivariância. Ao contrário da covariância e correlação, a semivariância é uma medida de dissimilaridade, ou seja, o valor da semivariância é maior à medida que as variáveis estão menos associadas.

Segundo Cressie (1993), a teoria Geoestatística supõe que uma, dentre três hipóteses, seja assumida. As hipóteses são: estacionaridade de 1ª ordem, estacionaridade de 2ª ordem e hipótese intrínseca.

Uma variável (ou processo) é estacionária se o desenvolvimento desse processo no tempo ou espaço ocorrer de maneira mais ou menos homogênea, com oscilações aleatórias contínuas em torno de um valor médio, em que nem a amplitude média e nem as oscilações mudam bruscamente no tempo ou no espaço.

Pode-se definir uma variável aleatória  $Z(x)$ , onde  $x$  denota uma posição em duas dimensões, com componentes  $(x_i, y_i)$ , como estacionária, se todos os momentos estatísticos são invariantes para toda mudança de origem.

Estatisticamente pode-se dizer que, se o processo é estacionário de

ordem  $k$ , então:

$$\begin{aligned} E[Z(x)] &= m_1(x) = \text{constante } \forall x \\ E[Z^2(x)] &= m_2(x) = \text{constante } \forall x \\ &\vdots \\ E[Z^k(x)] &= m_k(x) = \text{constante } \forall x \end{aligned} \quad (6)$$

em que  $m_k(x)$  é definido como o  $k$ -ésimo momento da variável aleatória  $Z(x)$ .

Se um processo é estacionário de ordem  $k$  ele também será estacionário para as ordens inferiores a  $k$ .

Para estudos geoestatísticos necessita-se, como restrição máxima, que o primeiro e o segundo momento em relação à origem sejam constantes, ou seja, exige-se no máximo estacionaridade de 2ª ordem.

Segundo Cressie (1993), a hipótese intrínseca é a hipótese mais frequentemente usada em Geoestatística, por ser menos restritiva do que a estacionaridade de 1ª ordem ou estacionaridade de 2ª ordem.

Na hipótese intrínseca tem-se:

- a) A esperança  $Z(x)$  existe e não depende do ponto  $x$ ;

$$E[Z(x)] = \mu \quad (7)$$

- b) Para todo  $h$ , em que  $h$  é um vetor distância, a variância da diferença  $[Z(x+h) - Z(x)]$  existe e não depende do ponto  $x$ , tal que:

$$\text{Var}[Z(x+h) - Z(x)] = E\left\{[Z(x+h) - Z(x)]^2\right\} = 2\gamma(h) \quad (8)$$

em que  $\gamma(h)$  é a semivariância definida na equação (9)

Se uma variável é estacionária de segunda ordem, então ela é também intrínseca, mas o inverso nem sempre ocorre.

### 2.2.2 Semivariância e semivariograma

A semivariância é a medida do grau de dependência espacial entre duas amostras. A magnitude da semivariância entre dois pontos depende da distância entre eles. O gráfico das semivariâncias em função da distância a um ponto é chamado de semivariograma. O semivariograma é uma ferramenta que permite representar quantitativamente a variação de um fenômeno regionalizado no espaço (JOURNEL; HUIJBREGTS, 1978).

Considere duas variáveis regionalizadas,  $X$  e  $Y$ , onde  $X = Z(x)$  e  $Y = Z(x+h)$  referentes ao mesmo atributo, mas medidas em duas posições diferentes conforme o Gráfico 1, em que  $h$  é um vetor distância que separa os pontos.

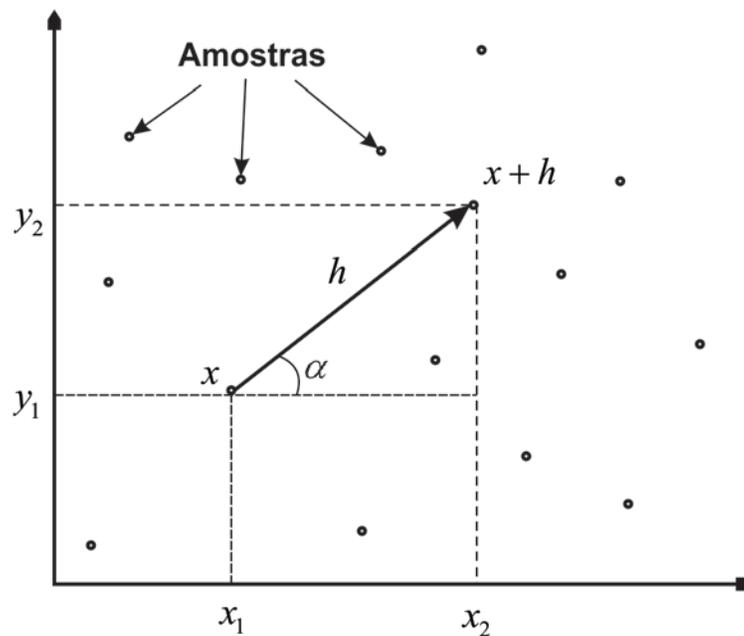


Gráfico 1 Amostragem em duas dimensões

O nível de dependência entre essas duas variáveis regionalizadas,  $X$  e  $Y$ , é representado pela variância,  $2\gamma(h)$ , a qual é definida como a esperança matemática do quadrado da diferença entre os valores de pontos no espaço, separados pelo vetor distância  $h$ , conforme apresentado na equação (8).

Por conveniência matemática o valor 2 a frente de  $\gamma(h)$  passou a dividir a expressão, sendo que  $\gamma(h)$  passou a ser chamada de semivariância.

De acordo com Cressie (1993), através de uma amostra  $Z(x_i)$ , com  $i = 1, 2, \dots, n$ , o estimador da semivariância é definido por:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(x_i) - z(x_i + h)]^2 \quad (9)$$

em que:

$\hat{\gamma}(h)$ : é estimador da semivariância;

$N(h)$ : é o número de pares de valores medidos,  $Z(x_i)$  e  $Z(x_i + h)$ , separados por um vetor distância  $h$ ;

$z(x_i)$  e  $z(x_i + h)$ : são realizações da variável aleatória  $Z(x)$ , isto é, valores da  $i$ -ésima observação da variável regionalizada, coletados nos pontos  $x_i$  e  $(x_i + h)$ , com  $i = 1, 2, \dots, n$ , separados pelo vetor  $h$ .

O gráfico das semivariâncias estimadas entre o vetor distância máximo e mínimo é chamado de semivariograma empírico.

### 2.2.3 Parâmetros do semivariograma

O Gráfico 2 ilustra um semivariograma empírico com características muito próximas do ideal. O seu padrão representa o que, intuitivamente, se espera de dados de campo, isto é, que as diferenças  $[Z(x_i) - Z(x_i + h)]$  decresçam à medida que  $h$  decresce. É esperado que observações mais próximas geograficamente tenham um comportamento mais semelhante entre si do que aquelas separadas por maiores distâncias. Desta maneira, é esperado que  $\gamma(h)$  aumente com a distância  $h$ .

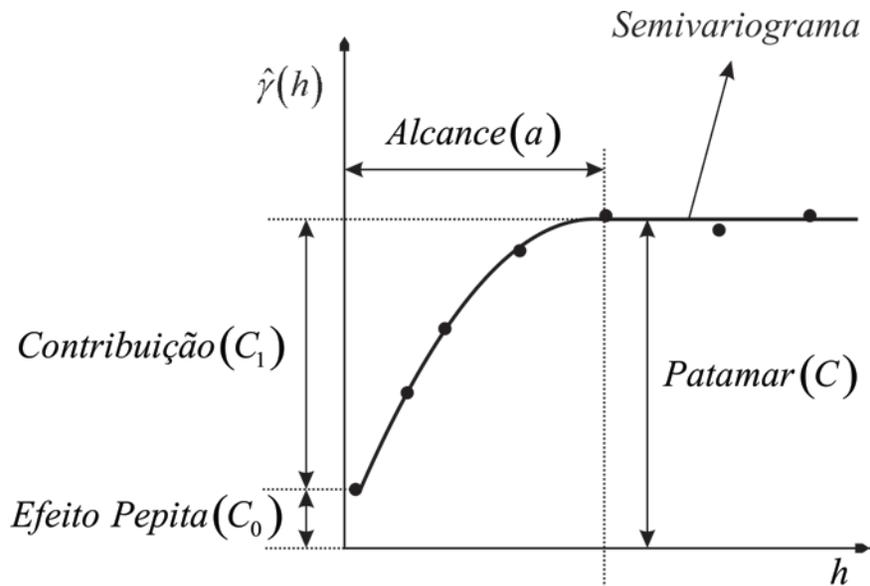


Gráfico 2 Exemplo de Semivariograma

Os parâmetros do semivariograma podem ser observados diretamente no Gráfico 2, conforme Isaacks e Srivastava (1989):

- a) Alcance( $a$ ): distância dentro da qual as amostras apresentam-se correlacionadas espacialmente. No Gráfico 2, o alcance ocorre próximo aos 25 m;

- b) Patamar ( $C$ ): é o valor da semivariância correspondente ao valor do alcance ( $a$ ), isto é,  $\gamma(a) = C$ . Desse ponto em diante, considera-se que não existe mais dependência espacial entre as amostras, porque a variância da diferença entre os pares de amostras ( $\text{Var}[Z(x) - Z(x+h)]$ ) torna-se invariante com a distância;
- c) Efeito Pepita ( $C_0$ ): teoricamente  $\gamma(0) = 0$ . Entretanto, na prática, a medida que  $h$  tende para 0 (zero),  $\gamma(h)$  se aproxima de um valor positivo chamado Efeito Pepita ( $C_0$ ), que revela a descontinuidade do semivariograma para distâncias menores que a menor distância entre as amostras. Parte desta descontinuidade pode ser também devida a erros de medição (ISSAKS; SRIVASTAVA, 1989), mas é impossível quantificar se a maior contribuição provém de erros de medição ou da variabilidade de pequena escala não captada pela amostragem;
- d) Contribuição ( $C_1$ ): é a diferença entre o patamar ( $C$ ) e o efeito pepita ( $C_0$ ).

#### 2.2.4 Métodos de estimação dos parâmetros do semivariograma

Existem diversos métodos de ajuste de modelos de semivariograma. McBratney e Webster (1986) descrevem sobre o ajuste de semivariogramas abordando os métodos de mínimos quadrados (ordinários, ponderados e generalizados) e de máxima verossimilhança. A escolha de determinado método está diretamente ligada à busca da melhor descrição do experimento através do modelo. Na Geoestatística uma das maneiras de se verificar qual modelo de semivariograma melhor se ajusta é através da técnica de validação cruzada.

### 2.2.5 Modelos teóricos de semivariograma

A sensibilidade dos semivariogramas para detectar a variabilidade espacial das amostras está diretamente ligada ao melhor ajuste de modelo teórico ao semivariograma empírico. Modelos teóricos de semivariogramas são superpostos à seqüência de pontos obtidos no semivariograma empírico, de modo que a curva que melhor se ajusta aos pontos representa a magnitude, o alcance e a intensidade da variabilidade espacial da variável estudada. Alguns modelos teóricos mais utilizados de semivariograma, segundo definição de Journel e Huijbregts (1978), são:

- a) Modelo Esférico (Sph): também chamado de modelo de Matheron é um dos modelos mais utilizados, definido por:

$$\gamma(h) = \begin{cases} 0 & , \quad |h| = 0 \\ C_0 + C_1 \left[ 1,5 \left( \frac{|h|}{a} \right) - 0,5 \left( \frac{|h|}{a} \right)^3 \right] & , \quad 0 < |h| \leq a \\ C_0 + C_1 & , \quad |h| > a \end{cases} \quad (10)$$

- b) Modelo Exponencial (Exp): outro modelo bastante utilizado, definido por:

$$\gamma(h) = \begin{cases} 0 & , \quad |h| = 0 \\ C_0 + C_1 \left[ 1 - \exp\left(-\frac{|h|}{a}\right) \right] & , \quad |h| \neq 0 \end{cases} \quad (11)$$

- c) Modelo Gaussiano (Gaus): um dos modelos utilizados no ajuste de fenômenos que demoram certa distância para aumentar:

$$\gamma(h) = \begin{cases} 0 & , \quad |h| = 0 \\ C_0 + C_1 \left[ 1 - \exp\left(-\frac{|h|^2}{a}\right) \right] & , \quad |h| \neq 0 \end{cases} \quad (12)$$

Os três modelos são apresentados no Gráfico 3.

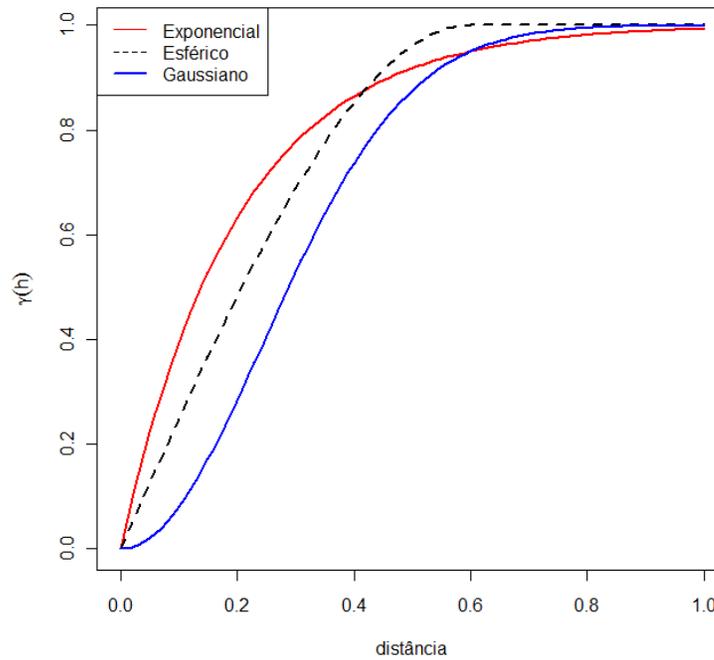


Gráfico 3 Modelos teóricos de semivariograma

### 2.2.6 Isotropia e Anisotropia

Caso a variável em estudo apresente o mesmo comportamento espacial em todas as direções, isto é, se construirmos semivariogramas direcionais e estes demonstrarem o mesmo comportamento, é dito que a variável é isotrópica.

Quando a variável em estudo apresenta diferentes padrões de dependência espacial, detectados através da obtenção de semivariogramas direcionais, diz-se que a variável é anisotrópica. Comumente calcula-se os semivariogramas em 4 direções: 0°, 45°, 90° e 135°.

Isaaks e Srivastava (1989) classificam a anisotropia em três tipos distintos:

- a) **Geométrica:** os semivariogramas direcionais apresentam o mesmo patamar ( $C$ ) com diferentes alcances ( $a$ ) (Gráfico 4);
- b) **Zonal:** os semivariogramas direcionais apresentam o mesmo alcance ( $a$ ) com diferentes patamares ( $C$ ) (Gráfico 5);
- c) **Combinada:** Os semivariogramas direcionais apresentam diferentes alcances ( $a$ ) e patamares ( $C$ ) (Gráfico 6).

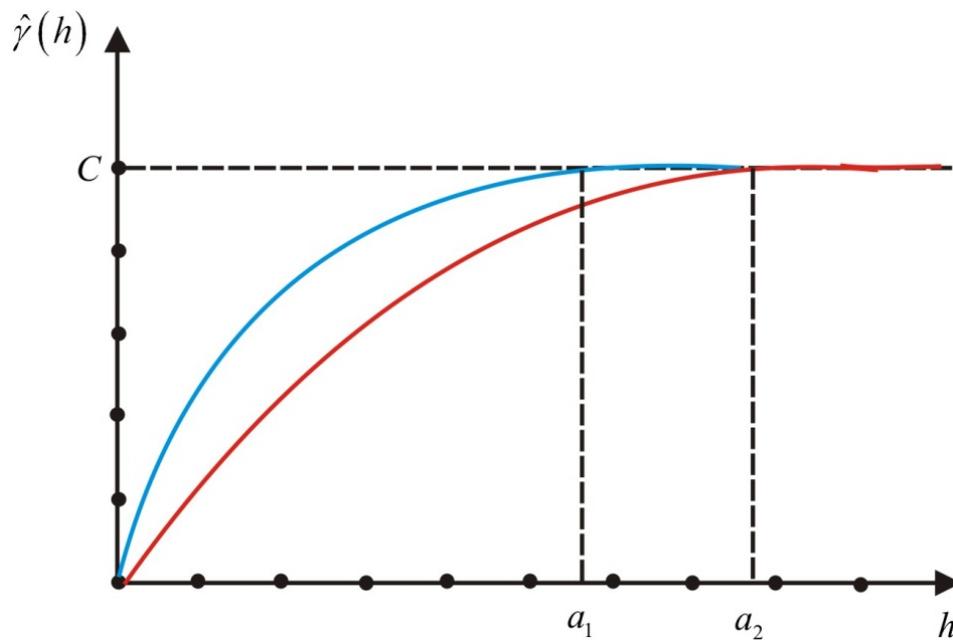
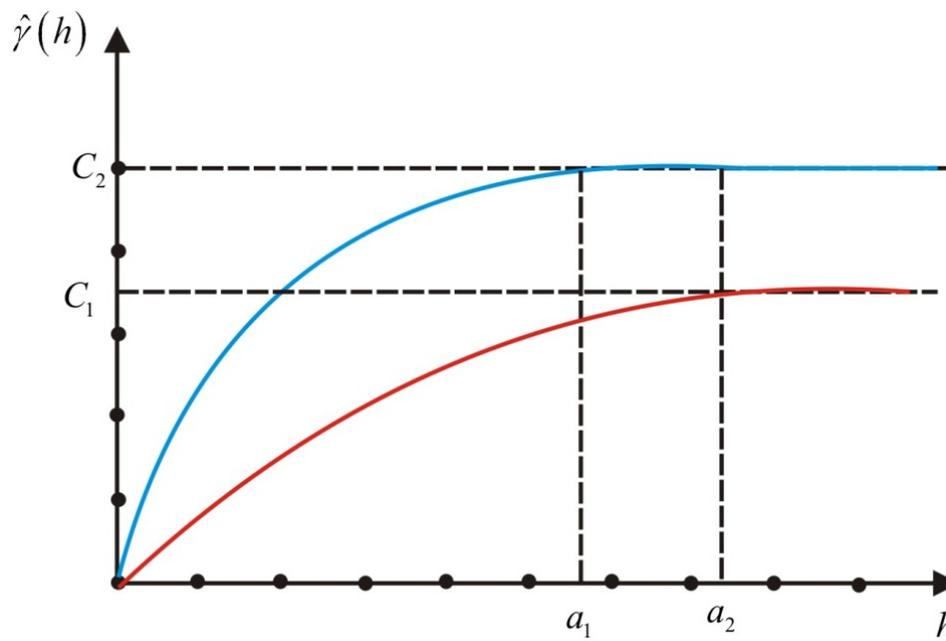
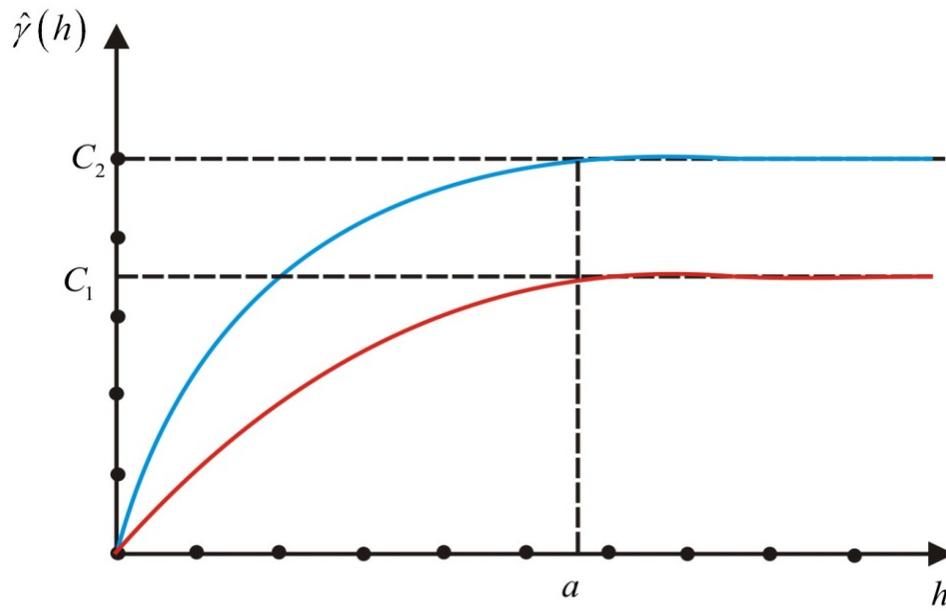


Gráfico 4 Anisotropia Geométrica



### 2.2.7 Krigagem

Muitas vezes além de um modelo de dependência espacial existe o interesse em se obter valores em pontos não amostrados ou obter uma malha de pontos interpolados que possibilite a visualização do comportamento da variável na região, através de mapas ou gráficos de superfície.

O termo krigagem é derivado do nome de Daniel G. Krige, que foi o primeiro a introduzir o uso de médias móveis para evitar a superestimação sistemática de reservas minerais. Inicialmente, o método de krigagem foi desenvolvido para solucionar problemas de mapeamentos geológicos, mas seu uso expandiu-se com sucesso no mapeamento de solos.

A diferença entre a krigagem e outros métodos de interpolação é a maneira como os pesos são atribuídos às diferentes amostras (CRESSIE, 1993). Os pesos são determinados a partir da análise espacial, baseada no semivariograma empírico que teve como ajuste um modelo teórico. Além disso, a krigagem fornece em média estimativas não tendenciosas e com variância mínima.

A krigagem engloba um conjunto de métodos de estimação, sendo que os principais são:

- a) Krigagem simples – é a krigagem em que se assume a função média do processo estocástico, logo, não é estimada a partir de realizações (BAILEY; GATRELL, 1995);
- b) Krigagem ordinária – segundo Chilès e Delfiner (1999) é a krigagem que envolve apenas o semivariograma, tornando-a a mais utilizada e cuja função média não é conhecida, portanto, estimada a partir dos dados;

- c) Krigagem universal – é a forma prevista originalmente para todos os casos em que a variável regionalizada não fosse estacionária. Este método é baseado inicialmente no ajuste de polinômios de baixo grau (1, 2 ou 3), que se encarregariam de remover a componente de deriva da variável regionalizada, e com os resíduos obter os variogramas que agora podem ser modelados, pois, em princípio, deveriam apresentar-se estacionários. A estimativa pela krigagem universal é feita por meio da equação da média ponderada, entretanto, os ponderadores são calculados simultaneamente com os coeficientes dos polinômios, o que exige maior tempo de computação;
- d) Krigagem disjuntiva – é a krigagem em que os pesos são encontrados por meio da projeção do valor desconhecido  $Z(x_0)$  no espaço vetorial intermediário gerado pela soma das “n” funções  $f(Z)$ .

Como esse trabalho teve maior enfoque na krigagem ordinária, a mesma será mais bem descrita a seguir.

### 2.2.7.1 Krigagem ordinária

Supondo que se tenha interesse em estimar  $Z(x_0)$ , em que  $Z(x_0)$  é uma variável a ser avaliada para qualquer local  $x_0$ , onde não se tem valores medidos. Um estimador, chamado de estimador de krigagem ordinária, é segundo Isaaks e Srivastava (1989), definido pela equação (13).

$$\hat{Z}(x_0) = \sum_{i=1}^N \lambda_i Z(x_i) \quad (13)$$

em que:

$N$  é o número de valores medidos;

$Z(x_i)$  são os valores medidos;

$\lambda_i$  são os pesos associados a cada valor medido.

Na krigagem os pesos variam de acordo com a variabilidade espacial expressa no semivariograma, sendo o estimador  $\hat{Z}(x_0)$  nada mais do que uma média móvel ponderada. O que torna a krigagem um interpolador ótimo é a maneira como os pesos são distribuídos, o que será descrito a seguir.

### 2.2.7.2 Determinação dos pesos na krigagem ordinária

Para que o estimador seja ótimo, ele não pode ser tendencioso e deve ter variância mínima, isto é,

$$E\{\hat{Z}(x_0) - Z(x_0)\} = 0 \quad (14)$$

e

$$\text{Var}\{\hat{Z}(x_0) - Z(x_0)\} = E\left\{\left[\hat{Z}(x_0) - Z(x_0)\right]^2\right\} \text{ seja mínima.} \quad (15)$$

As equações (14) e (15) representam as condições de não-tendenciosidade e de variância mínima, respectivamente. A condição de não-tendenciosidade significa que, em média, a diferença entre valores estimados e medidos para o mesmo ponto possa ser nula. A condição de variância mínima significa que, embora possam existir diferenças ponto por ponto entre o valor estimado e o medido, essas diferenças devem ser mínimas.

Substituindo-se a equação (13) na equação (14), tem-se:

$$E\{\hat{Z}(x_0) - Z(x_0)\} = E\left\{\sum_{i=1}^N \lambda_i Z(x_i) - Z(x_0)\right\} = 0 \quad (16)$$

Aplicando as propriedades da esperança, têm-se:

$$E\{Z(x_0) - Z(x_0)\} = \sum_{i=1}^N \lambda_i E\{Z(x_i)\} - E\{Z(x_0)\} = 0 \quad (17)$$

Substituindo a primeira condição de estacionaridade, ou seja,  $(E\{z(x_i)\} = \mu)$ , têm-se:

$$E\{\hat{Z}(x_0) - Z(x_0)\} = \sum_{i=1}^N \lambda_i \mu - \mu = \mu \left( \sum_{i=1}^N \lambda_i - 1 \right) = 0 \quad (18)$$

Para que a equação (18) seja verdade para qualquer valor de  $\mu$ , é necessário que:

$$\sum_{i=1}^N \lambda_i - 1 = 0, \quad (19)$$

ou seja,

$$\sum_{i=1}^N \lambda_i = 1. \quad (20)$$

Portanto, para que a estimativa seja não tendenciosa, é necessário que a soma dos pesos seja igual a 1, qualquer que seja a distribuição dos seus valores.

Journel e Huijbregts (1978) mostram que, minimizando a variância do erro  $(\text{Var}[\hat{Z}(x_0) - Z(x_0)])$  sob a condição dada pela equação (20), os pesos  $\lambda_i$  são obtidos a partir do seguinte sistema de equações, denominado sistema de krigagem ordinária:

$$\begin{cases} \sum_{j=1}^N \lambda_j \gamma(x_i, x_j) - \alpha = \gamma(x_i, x_0) & \text{para } i = 1, \dots, n \\ \sum_{j=1}^N \lambda_j = 1 \end{cases} \quad (21)$$

em que:

$\gamma(x_i, x_j)$  e  $\gamma(x_i, x_0)$  são, respectivamente, a semivariância entre os pontos  $x_i$  e  $x_j$  e entre os pontos  $x_i$  e  $x_0$

$\alpha$  é o multiplicador de Lagrange necessário para a minimização da variância do erro.

A correspondente variância minimizada do erro, denominada variância de krigagem ordinária ( $\sigma_{ko}^2$ ), é dada pela seguinte expressão (CRESSIE, 1993).

$$\begin{aligned}\sigma_{ko}^2 &= \text{Var}\left[Z(x_0) - \hat{Z}(x_0)\right] \\ &= \text{Cov}(x_0, x_0) - \sum_{i=1}^N \lambda_i \text{Cov}(x_i, x_0) - \alpha\end{aligned}\quad (22)$$

A krigagem ordinária é um interpolador exato no sentido de que, quando as equações (13), (21) e (22) forem usadas, os valores interpolados irão coincidir com os valores dos pontos amostrais.

### 2.2.8 Validação cruzada

Suponha que um modelo teórico de semivariograma tenha sido ajustado para um determinado conjunto de dados. O método de validação cruzada verifica se o fenômeno foi modelado de forma adequada. Segundo Isaaks e Srivastava (1989), a ideia básica consiste em excluir uma observação e através das observações remanescentes estimar a observação excluída. O valor estimado é então comparado com o valor observado.

Assim, a validação cruzada pode ser descrita nas seguintes fases:

- a) Suprimir do conjunto de dados, um ponto qualquer  $Z(x_i)$ ;
- b) Calcular, pelo método da krigagem,  $\hat{Z}(x_i)$ ;

- c) Calcular o erro de estimação através da expressão:

$$\frac{[Z(x_i) - \hat{Z}(x_i)]}{\sigma_{x_i}} \quad (23)$$

- d) Repetir as fases anteriores para todos os pontos;  
 e) Calcular o Erro Médio Reduzido (ER) e o Desvio Padrão do Erro Reduzido  $S_{ER}$ , dados por:

$$ER = \frac{1}{N} \sum_{i=1}^N \left( \frac{Z(x_i) - \hat{Z}(x_i)}{\sigma_{x_i}} \right) \quad (24)$$

e

$$S_{ER} = \sqrt{\frac{1}{N} \sum_{i=1}^N \left( \frac{Z(x_i) - \hat{Z}(x_i)}{\sigma_{x_i}} \right)^2} \quad (25)$$

em que:

$Z(x_i)$  é o valor observado no ponto  $x_i$ ;

$\hat{Z}(x_i)$  é o valor estimado para o ponto  $x_i$  e

$\sigma_{x_i}$  é o desvio padrão da krigagem no ponto  $x_i$ .

Segundo Cressie (1993), considera-se que o modelo foi bem ajustado se o Erro Médio Reduzido (equação (24)) for próximo de zero e o Desvio Padrão do Erro Reduzido (equação (25)) for próximo de um.

### 2.2.9 Modelagem Geoestatística do erro

Considere o modelo linear:

$$Y = X\beta + \varepsilon$$

em que:

$Y$  é um vetor  $n \times 1$  de valores observados;

$X$  é uma matriz  $n \times p$  das observações das variáveis independentes;

$\beta$  é um vetor  $p \times 1$  dos parâmetros;

$\varepsilon$  é um vetor  $n \times 1$  dos erros inerentes a cada observação.

O vetor de erros experimentais independentes e identicamente distribuídos é assumido como:  $\varepsilon \sim N(\phi, I\sigma^2)$ , em que  $\phi$  é um vetor nulo,  $I$  é a matriz identidade e  $\sigma^2$  a variância amostral.

Segundo Duarte (2000), no modelo de análise que admite autocorrelação espacial, o erro é assumido como:  $\varepsilon \sim N(\phi, R)$ , em que  $\phi$  é um vetor nulo e  $R$  a matriz de variâncias e covariâncias, definida por:

$$R = \begin{pmatrix} \sigma^2 & Cov(h_i) & \dots & Cov(h_{máx}) \\ Cov(h_i) & \sigma^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ Cov(h_{máx}) & \dots & \dots & \sigma^2 \end{pmatrix}.$$

em que  $Cov(h_i)$  é a covariância entre duas parcelas  $(x_i, x_i + h_i)$  separados por uma distância  $h_i$  e  $h_{máx}$  é a distância máxima considerada entre duas parcelas experimentais.

Considerando-se as pressuposições de estacionaridade de 1ª e 2ª ordem, pode-se obter  $Cov(h_i)$  através de:

$$Cov(h_i) = 1 - \gamma(h_i)$$

### 2.3 Modelo de resposta autorregressiva

O modelo autorregressivo é parte de um grupo de fórmulas de predição linear que tentam prever uma saída  $y_n$  de um sistema, baseado nos resultados anteriores  $(y_{n-1}, y_{n-2}, \dots)$ .

Um modelo que depende apenas das saídas anteriores do sistema é chamado de um modelo autorregressivo (AR), enquanto que um modelo que depende apenas das entradas para o sistema é chamado de um modelo de média móvel (MA) e um modelo baseado em entradas e saídas é um modelo autorregressivo de movimento da média (ARMA) (PINHEIRO; BATES, 2000).

Quando agregamos a esses modelos, não apenas informações anteriores, mas informações com localização anterior, frontal, superior, inferior, etc., passamos a ter um novo conjunto de modelos autorregressivos que passa a considerar a dependência espacial. Desses, os principais são o modelo CAR e SAR (GRIFFITH, 1988).

#### 2.3.1 Modelo espacial autorregressivo (SAR)

Segundo Griffith (1988), o modelo SAR (spatial auto regressive ou spatial lag model), descreve a variação espacial no vetor resposta  $Y_{n \times 1}$  e é definido por:

$$Y = \rho WY + X\beta + \varepsilon \quad (26)$$

em que:

$Y$  é um vetor  $n \times 1$  de valores observados;

$\rho$  é um parâmetro espacial autorregressivo;

$W$  é uma matriz  $n \times n$  com atribuições de peso da vizinhança espacial;

$X$  é uma matriz  $n \times p$  da incidência dos efeitos fixos;

$\beta$  é um vetor  $p \times 1$  dos parâmetros;

$\varepsilon$  é um vetor  $n \times 1$  dos erros inerentes a cada observação.

A matriz  $W$  é obtida através da multiplicação de duas outras matrizes  $D$  e  $C$  ( $W = D \times C$ ). A matriz  $C$  de dimensões  $n \times n$  é binária, e descreve a vizinhança das parcelas experimentais. Gumpertz et al. (1997) propõe um padrão de proximidade para definir a região de vizinhança. A apresenta quatro padrões de proximidade, considerando uma amostragem com gride regular.

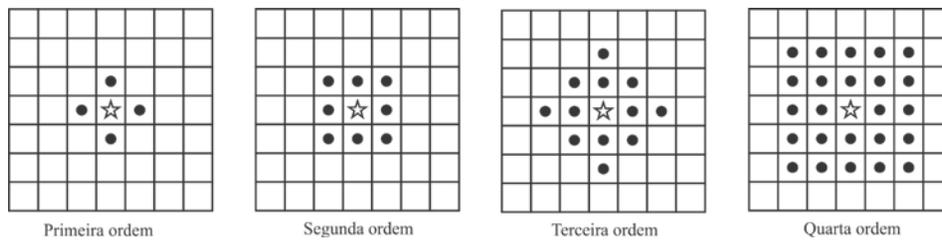


Figura 1 Padrões de proximidade apresentados por Gumpertz et al. (1997), em que  $\star$  representa uma parcela de referência,  $\bullet$  representa as parcelas vizinhas consideradas na comparação e  $\square$  representa as demais parcelas

Considere que na cada observação adjacente está distante uma da outra na proporção de uma unidade de medida ( $1 \text{ u.m.}$ ), logo, os centroides de cada parcela também distam  $1 \text{ u.m.}$  dos outros centroides adjacentes. Podemos então definir a vizinhança como as parcelas que são abrangidas pela circunferência de raio  $r$ , partindo da parcela de origem. Dessa maneira poderíamos reescrever os padrões de proximidade em função dos raios, centrado na parcela de referência, conforme apresentado na Tabela 3.

Tabela 3 Padrões de proximidade em função do raio.

<b>Padrão de proximidade</b>	<b>Raio da circunferência</b>
Primeira ordem	$1u.m.$
Segunda ordem	$1\sqrt{2}u.m.$
Terceira ordem	$2u.m.$
Quarta ordem	$2\sqrt{2}u.m.$
$\vdots$	$\vdots$
n-ésima ordem ímpar	$n u.m.$
n-ésima ordem par	$n\sqrt{2} u.m.$

Logo, cada elemento  $c_{ij}$  da matriz  $C$  pode ser definido por:

$$c_{ij} = \begin{cases} 1 & \text{para as parcelas contidas na circunferência de raio } r, \\ & \text{centrada em } c_{ij} \\ 0 & \text{caso contrário.} \end{cases}$$

Para uma melhor compreensão, considere um experimento com 9 observações apresentado na Figura 2.

1	2	3
4	5	6
7	8	9

Figura 2 Gride regular com 9 observações

Seja a parcela número 1 a parcela de referência. Considerando um padrão de proximidade de primeira ordem, tem-se que as parcelas 2 e 4 seriam consideradas vizinhas. Logo na matriz  $C$  os elementos  $c_{12}$  e  $c_{14}$  receberiam o valor de 1 e os demais  $c_{1i} = 0$ . Considerando-se a parcela 5 como referência, têm-se  $c_{52} = c_{54} = c_{56} = c_{58} = 1$  e  $c_{5i} = 0$  para  $i = 1, 3, 7$  e  $9$ .

A matriz de vizinhança  $C$  considerando o padrão de proximidade de primeira ordem, para o exemplo da Figura 2, seria:

$$C: \begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \end{pmatrix}$$



$$W = \begin{pmatrix} \frac{1}{2} & & & & & & & & & \\ & \frac{1}{3} & & & & & & & & \\ & & \frac{1}{2} & & & & & & & \\ & & & \frac{1}{3} & & & & & & \\ & & & & \frac{1}{4} & & & & & \\ & & & & & \frac{1}{3} & & & & \\ & & & & & & \frac{1}{2} & & & \\ & & & & & & & \frac{1}{3} & & \\ & & & & & & & & \frac{1}{2} & \end{pmatrix} \times \begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \end{pmatrix}$$

$$W = \begin{pmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{4} & 0 & \frac{1}{4} & 0 & \frac{1}{4} & 0 & \frac{1}{4} & 0 \\ 0 & 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{pmatrix}.$$

### 2.3.2 Estimação do parâmetro $\rho$

Uma das formas de se estimar o parâmetro  $\rho$  do modelo SAR é pelo método de máxima verossimilhança (MV). O MV é baseado no princípio de selecionar valores para os parâmetros que maximizem a probabilidade de ter obtido os dados observados. Smirnov e Anselin (2001) propuseram diversas técnicas para a estimação de  $\rho$ , como decomposição de autovalores, método de decomposição matricial através da decomposição de Cholesky e aproximação de funções.

A solução original da estimação por MV de um modelo espacial autorregressivo, originalmente proposta por Ord (1975), consiste em explorar a decomposição do Jacobiano  $|I - \rho W|$  em termos dos autovalores  $\omega_i$  (com  $i = 1, 2, \dots, N$ ) da matriz  $W$ , dada por:

$$|I - \rho W| = \prod_{i=1}^n (1 - \rho \omega_i) \quad (27)$$

ou

$$\ln |I - \rho W| = \sum_{i=1}^n \ln (1 - \rho \omega_i) \quad (28)$$

A partir da equação (27) ou (28) obtém-se um polinômio que não têm solução única e que deve ser solucionado iterativamente através de métodos computacionais.

### 2.3.3 Análise de variância de modelo autorregressivo (ANOVA-AR)

A análise de variância de modelo autorregressivo (ANOVA-AR), foi descrita por Long (1996). A idéia básica consiste em transformar observações

ditas autocorrelacionadas em observações não-correlacionadas. Para tanto, após estimar  $\rho$  deve-se proceder ao ajuste dos dados observados a partir de

$$Y_{adj} = Y - (\hat{\rho}WY - \hat{\rho}\beta_0) \quad (29)$$

em que

$Y$  é um vetor  $n \times 1$  de valores observados;

$\hat{\rho}$  é a estimativa do parâmetro espacial autorregressivo;

$W$  é a matriz  $n \times n$  com atribuições de peso da vizinhança espacial;

$\beta_0$  é a média dos valores observados;

$Y_{adj}$  é um vetor  $n \times 1$  de valores ajustados.

Com os valores de  $Y_{adj}$  obtém-se uma nova tabela de Análise de Variância. Esta tabela é utilizada como base para a construção da ANOVA-AR e está apresentado na Tabela 4.

O valor da Soma Quadrada referente ao parâmetro  $\rho$  ( $SQ\rho$ ) é obtido através da diferença entre a Soma de Quadrado Total ( $SQT$ ) da Análise de Variância dos dados não ajustados e a Soma de Quadrado Total Ajustada ( $SQT_{adj}$ ) da Análise de Variância com dados ajustados pela equação (29), isto é:

$$SQ\rho = SQT - SQT_{adj}$$

Tabela 4 Análise de variância de modelo autorregressivo (ANOVA-AR).

<i>FV</i>	<i>GL</i>	<i>SQ</i>	<i>QM</i>	<i>F</i>
Fator ( $\rho$ )	1	$SQ\rho$	---	---
Parâmetros	k	$SQP_{adj}$	$QMP_{adj} = SQP_{adj} / k$	$QMP_{Adj} / QME_{Adj}$
Resíduo	$n - k - 2$	$SQE_{adj}$	$QME_{adj}$	
Total	$n - 1$	$SQT$		

### 3 MATERIAL E MÉTODOS

Para o desenvolvimento deste estudo foram utilizados dados gerados a partir de simulação. Será efetuada uma comparação da análise não espacial (erros independentes) com a análise espacial, utilizando a Geoestatística (erros correlacionados) e a ANOVA com modelo autorregressivo (ANOVA-AR). Todas as análises foram feitas no *software* R (R DEVELOPMENT CORE TEAM, 2010).

#### 3.1 Simulação de dados

Os dados foram gerados em duas configurações diferentes:

- a) Experimento i - um experimento em blocos casualizados (DBC) com 18 tratamentos e 6 blocos ;
- b) Experimento ii - um experimento em blocos casualizados (DBC) com 5 tratamentos e 4 blocos.

As duas configurações foram propostas com o intuito de verificar se as metodologias apresentam os mesmos resultados em experimentos com um grande número de parcelas e experimentos com um pequeno número de parcelas.

Os experimentos foram construídos considerando-se um gride regular com o tamanho das parcelas definido em  $1 \text{ u.m}^2$ .

Os parâmetros iniciais foram fixados, sendo que a parte de aleatoriedade foi atribuída ao erro. A Tabela 5 apresenta os valores fixos para o experimento i e para o experimento ii.

O modelo estatístico para ambas as situações é dado por

$$y_{ij} = \mu + t_i + b_j + e_{ij}$$

em que

$y_{ij}$  é a observação do  $i$ -ésimo tratamento na  $j$ -ésima observação, sendo  $i=1,2,\dots,a$  e  $j=1,2,\dots,b$ , com  $a=18$  e  $b=6$  na configuração (i) e  $a=5$  e  $b=4$  na configuração (ii);

$\mu$  é uma constante inerente a cada observação;

$t_i$  é o efeito do  $i$ -ésimo tratamento;

$b_j$  é o efeito do  $j$ -ésimo bloco;

$e_{ij}$  é o erro associado a cada observação.

Os erros com característica aleatória foram gerados seguindo um modelo gaussiano (equação (12)). Segundo Clark (1979) os erros de experimentos de campo com dependência espacial tendem a ter erros que podem ser modelados por modelos gaussianos.

Para tanto, foram geradas várias configurações de erros dependentes, conforme a Tabela 6.

O intuito de se trabalhar com diferentes configurações de erro dependente foi o de verificar a eficácia das abordagens Geoestatística e autorregressiva quando confrontadas com diversas configurações de erro.

Diferentes alcances foram gerados para que o experimento (i) possuísse dependência espacial de curto, médio e longo alcance (2,4 e 6). O patamar em duas configurações indicou um decréscimo acentuado da variabilidade espacial (0,25) e brando (1). O efeito pepita em três configurações (0; 0,25 e 0,75) teve intuito de gerar um erro no qual a explicação por parte do modelo fosse total (efeito pepita 0), que apresenta-se uma pequena variabilidade não explicada entre a menor distância amostral (efeito pepita 0,25) e uma grande variabilidade não explicada entre a menor distância amostral (efeito pepita 0,75). O experimento (ii) foi construído de forma análoga.

Tabela 5 Efeitos fixos dos parâmetros utilizados na simulação.

<i>Experimento i</i>		<i>Experimento ii</i>	
Parâmetro	Valor Fixo	Parâmetro	Valor Fixo
$\mu$	100	$\mu$	100
$t_1$	33	$t_1$	33
$t_2$	40	$t_2$	37
$t_3$	33	$t_3$	35
$t_4$	36	$t_4$	34
$t_5$	35	$t_5$	38
$t_6$	37	$b_1$	17
$t_7$	32	$b_2$	15
$t_8$	34	$b_3$	18
$t_9$	33	$b_4$	20
$t_{10}$	34		
$t_{11}$	38		
$t_{12}$	39		
$t_{13}$	33		
$t_{14}$	39		
$t_{15}$	35		
$t_{16}$	35		
$t_{17}$	39		
$t_{18}$	35		
$b_1$	16		
$b_2$	19		
$b_3$	20		
$b_4$	18		
$b_5$	19		
$b_6$	20		

Tabela 6 Parâmetros do semivariograma para a simulação do erro experimental.

<i>Experimento i</i>				<i>Experimento ii</i>			
Nomenclatura	Efeito Pepita	Patamar	Alcance	Nomenclatura	Efeito Pepita	Patamar	Alcance
<i>Gaus(0-1-6)</i>	0	1	6	<i>Gaus(0-1-2)</i>	0	1	2
<i>Gaus(0,75-1-6)</i>	0,75	1	6	<i>Gaus(0,75-1-2)</i>	0,75	1	2
<i>Gaus(0,25-1-6)</i>	0,25	1	6	<i>Gaus(0,25-1-2)</i>	0,25	1	2
<i>Gaus(0-0,25-6)</i>	0	0,25	6	<i>Gaus(0-0,25-2)</i>	0	0,25	2
<i>Gaus(0-1-4)</i>	0	1	4	<i>Gaus(0-1-1)</i>	0	1	1
<i>Gaus(0,75-1-4)</i>	0,75	1	4	<i>Gaus(0,75-1-1)</i>	0,75	1	1
<i>Gaus(0,25-1-4)</i>	0,25	1	4	<i>Gaus(0,25-1-1)</i>	0,25	1	1
<i>Gaus(0-0,25-4)</i>	0	0,25	4	<i>Gaus(0-0,25-1)</i>	0	0,25	1
<i>Gaus(0-1-2)</i>	0	1	2				
<i>Gaus(0,75-1-2)</i>	0,75	1	2				
<i>Gaus(0,25-1-2)</i>	0,25	1	2				
<i>Gaus(0-0,25-2)</i>	0	0,25	2				

O erro dependente foi gerado utilizando-se a função *grf* do pacote *geoR* (RIBEIRO JR; DIGGLE, 2001). Segue a implementação para o experimento (i) na configuração *Gaus(0-1-6)*.

```
dep<-grf(108, grid="reg",ny=12, nx=9,xlims = c(0,8), ylims = c(0, 11),nugget  
= 0, cov.pars=c(1,6))
```

Para a validação da análise de variância, o erro deve possuir média zero e variância constante, logo sem perda de generalidade, o erro foi gerado como sendo  $N(0,1)$ , através da padronização dos valores obtidos em *dep*. Subtraiu-se  $\mu$  dos valores e dividiu-se por  $\sigma^2$ . Os valores  $\mu$  e  $\sigma^2$  foram obtidos através de uma reamostragem bootstrap com 1000 valores.

Após a padronização o erro foi adicionado aos efeitos fixos em cada observação.

Foi gerado um total de 20000 conjuntos de dados, perfazendo um total de 1000 simulações para cada configuração da Tabela 6. Para cada conjunto de dados aplicou-se a abordagem autorregressiva com padrões de proximidade de primeira, segunda e terceira ordem; e aplicou-se, também, a abordagem Geoestatística.

Foram utilizados três computadores para fazer a simulação. Cada configuração da Tabela 6 com um total de 1000 simulações demorou, em média, um total de 36 horas no Sistema Operacional Windows 7 Ultimate; e um total de 22 horas no Sistema Operacional Linux, com a distribuição Ubuntu 10.04, perfazendo um total aproximado de 600 horas de simulação ou 25 dias.

O computador 1 possui um processador Core2 Quad 9400, 4 Gb de RAM DDR2, 1000 Gb de HD, Placa Mãe Asus Formula Striker II e Placa de Vídeo Geforce GTX 9800+.

O computador 2 possui um processador Core2 Extreme X9770, 2 Gb de RAM DDR2, 500 Gb de HD e Placa Mãe Asus P5KPL-AM.

O computador 3 possui um processador Core i7 x980, 16 Gb de RAM DDR3, 1000 Gb de HD e Placa Mãe Intel DX58SO2.

Para a comparação de resultados entre a abordagem clássica e as abordagens autorregressiva e Geoestatística construíram-se intervalos de credibilidade visando à comparação dos parâmetros de cada abordagem.

Os parâmetros analisados na abordagem autorregressiva foram:  $\hat{\rho}$ , Fator ( $\rho$ ),  $QME$ ,  $QMT$  e  $QMB$ . Na abordagem Geoestatística analisou-se os intervalos de credibilidade para o critério de informação de Akaike (AIC) para o modelo com erros independentes e o modelo com erros dependentes.

### 3.2 Abordagem Geoestatística

- a) Calculam-se os resíduos do modelo que considera o erro como sendo independente;
- b) Ajusta-se um ou mais semivariogramas para os valores do resíduo;
- c) Através da técnica de validação cruzada escolhe-se o modelo de semivariograma que melhor se ajustou aos dados;
- d) Determinado o modelo geoestatístico do semivariograma, verificam-se os parâmetros fornecidos pelo mesmo: efeito pepita, patamar e alcance;
- e) Em seguida, calcula-se o  $Cov(h_i)$  para a construção da matriz  $R$ ;
- f) Tendo construído a matriz  $R$ , através de rotinas computacionais, atualiza-se o modelo linear com erros independentes para o modelo linear com erros correlacionados, utilizando mínimos quadrados generalizados;
- g) Através dos critérios de seleção de modelo (AIC), verifica-se se o incremento dos erros correlacionados contribuiu para um melhor ajuste do modelo, isto é, se o modelo com erros correlacionados

obteve um menor valor de AIC se comparado com o modelo com erros independentes.

### 3.3 Abordagem autorregressiva

A abordagem autorregressiva pode ser descrita nas seguintes etapas:

- a) Defini-se o padrão de proximidade a ser adotado conforme Gumpertz et al. (1997). Mais de um padrão de proximidade pode ser adotado, desde que as etapas seguintes sejam refeitas para cada um deles;
- b) Constrói-se a matriz  $C$  e conseqüentemente as matrizes  $D$  e  $W$ ;
- c) Calculam-se os autovalores da matriz  $W$ ;
- d) Estima-se o parâmetro  $\rho$  utilizando-se a equação (27);
- e) Tendo obtido  $\hat{\rho}$ , devem-se ajustar as observações autocorrelacionadas em observações não-correlacionadas através da equação (29);
- f) Com a obtenção do vetor  $Y_{adj}$  procede-se a construção da tabela de análise de variância de modelo autorregressivo (ANOVA-AR) conforme definido na Tabela 4;
- g) Através da análise do  $QME$  e  $QMP$  versus  $QME_{adj}$  e  $QMP_{adj}$  verifica-se se a abordagem autorregressiva contribuiu para a diminuição da variabilidade do experimento, isto é, constata-se se o  $QME_{adj}$  e  $QMP_{adj}$  obtiveram valores menores do que o  $QME$  e  $QMP$ .

## 4 RESULTADOS E DISCUSSÃO

Para uma melhor compreensão de cada uma das abordagens, este estudo apresentará, inicialmente, um caso particular de cada. Posteriormente abordar-se-á um conjunto de 1000 simulações para cada uma das configurações apresentadas na Tabela 6.

### 4.1 Exemplo de abordagem Geoestatística

Para exemplificar a abordagem Geoestatística foi escolhida uma simulação de configuração *Gaus(0-1-6)* do experimento (i). Os dados utilizados nesse exemplo se encontram na Tabela 27, no Anexo A.

A primeira análise consiste em considerar as observações como sendo espacialmente independentes, logo se procede com a construção do modelo linear com erros independentes.

Os parâmetros do modelo com erro independente estão apresentados na Tabela 7.

Tabela 7 Parâmetros do modelo com erro independente.

<i>Parâmetros</i>	<i>Valor</i>	<i>Erro Padrão</i>	<i>t -valor</i>	<i>p - valor</i>
Intercepto	151,65	1,03	147,34	0,00
$t_2$	8,83	1,29	6,86	0,00
$t_3$	1,87	1,29	1,45	0,15
$t_4$	5,44	1,29	4,22	0,00
$t_5$	1,24	1,29	0,97	0,34
$t_6$	3,27	1,29	2,54	0,01
$t_7$	-1,47	1,29	-1,14	0,26
$t_8$	1,74	1,29	1,35	0,18

Tabela 7, conclusão.

$t_9$	2,33	1,29	1,81	0,07
$t_{10}$	-0,43	1,29	-0,33	0,74
$t_{11}$	5,90	1,29	4,58	0,00
$t_{12}$	6,75	1,29	5,24	0,00
$t_{13}$	0,22	1,29	0,17	0,86
$t_{14}$	5,62	1,29	4,37	0,00
$t_{15}$	0,89	1,29	0,69	0,49
$t_{16}$	1,60	1,29	1,24	0,22
$t_{17}$	7,17	1,29	5,56	0,00
$t_{18}$	3,54	1,29	2,75	0,01
$b_2$	1,70	0,74	2,28	0,03
$b_3$	-0,26	0,74	-0,35	0,73
$b_4$	-2,91	0,74	-3,92	0,00
$b_5$	-1,78	0,74	-2,40	0,02
$b_6$	0,50	0,74	0,67	0,51

Os valores do resíduo obtidos foram utilizados para o cálculo do semivariograma empírico, para o qual foram ajustados dois modelos teóricos: Gaussiano e Esférico (Gráfico 7).

As estimativas dos parâmetros dos dois modelos estão apresentadas na Tabela 8.

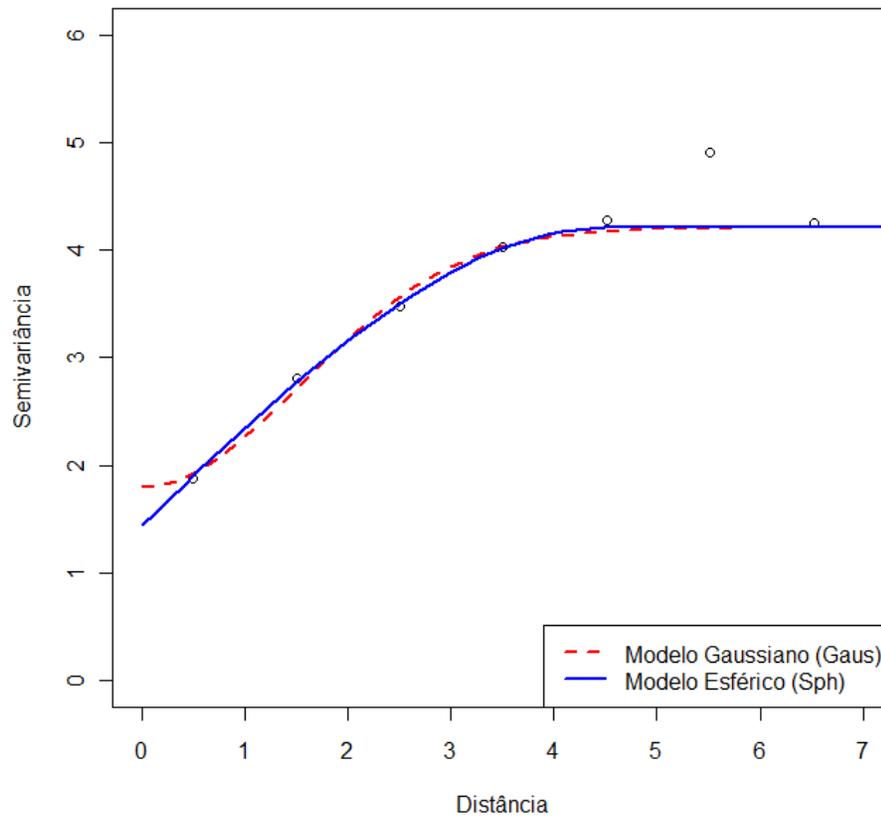


Gráfico 7 Semivariograma do exemplo de abordagem Geoestatística

Tabela 8 Parâmetros estimados dos modelos de semivariograma.

<i>Modelo</i>	<i>Efeito Pepita</i>	<i>Alcance</i>	<i>Patamar</i>
Gaussiano	1,21	2,01	3,53
Esférico	1,21	4,01	3,53

Tanto o modelo Gaussiano quanto o modelo Esférico apresentaram o mesmo valor para o efeito pepita e para o patamar, todavia, o modelo Gaussiano detectou a dependência espacial ao alcance de 2,01 *u.m* enquanto que o modelo Esférico detectou a dependência espacial ao alcance de 4,01 *u.m*.

Através da técnica de validação cruzada optou-se pela escolha do modelo Gaussiano, que obteve um melhor ajuste ao semivariograma empírico.

Os parâmetros do modelo com erro correlacionado estão apresentados na Tabela 9.

Tabela 9 Parâmetros do modelo com erro correlacionado.

<i>Parâmetros</i>	<i>Valor</i>	<i>Erro Padrão</i>	<i>t -valor</i>	<i>p - valor</i>
Intercepto	151,43	1,13	134,57	0,00
$t_2$	8,74	0,97	8,99	0,00
$t_3$	1,74	1,23	1,41	0,16
$t_4$	5,39	1,25	4,32	0,00
$t_5$	1,28	1,25	1,02	0,31
$t_6$	3,40	1,25	2,72	0,01
$t_7$	-1,18	1,25	-0,94	0,35
$t_8$	2,10	1,25	1,68	0,10
$t_9$	2,63	1,25	2,10	0,04
$t_{10}$	-0,37	0,68	-0,55	0,59
$t_{11}$	5,98	1,06	5,63	0,00
$t_{12}$	6,69	1,24	5,40	0,00
$t_{13}$	0,42	1,25	0,34	0,74
$t_{14}$	5,91	1,25	4,73	0,00
$t_{15}$	1,02	1,25	0,82	0,42
$t_{16}$	1,64	1,25	1,32	0,19
$t_{17}$	7,41	1,25	5,93	0,00

Tabela 9 conclusão.

$t_{18}$	3,71	1,25	2,97	0,00
$b_2$	1,89	0,97	1,95	0,05
$b_3$	0,02	1,03	0,01	0,99
$b_4$	-2,71	1,05	-2,57	0,01
$b_5$	-1,61	1,06	-1,51	0,13
$b_6$	0,54	1,09	0,50	0,62

O critério de seleção de modelos (AIC) para o modelo com erros correlacionados foi menor do que o modelo com erros independentes, mostrando que a abordagem Geoestatística obteve melhores resultados do que a convencional (Tabela 10).

Para uma melhor interpretação da variabilidade espacial foi construído um mapa de krigagem, mostrando que não houve homogeneidade dentro dos blocos, nem é possível, visualmente, diferir entre cada um dos seis blocos, conforme pode ser visto na Figura 3.

Tabela 10 Critério de Akaike para o modelo 1 e modelo 2.

	<i>AIC</i>
Modelo com erro independente	470,49
Modelo com erro correlacionado	433,96

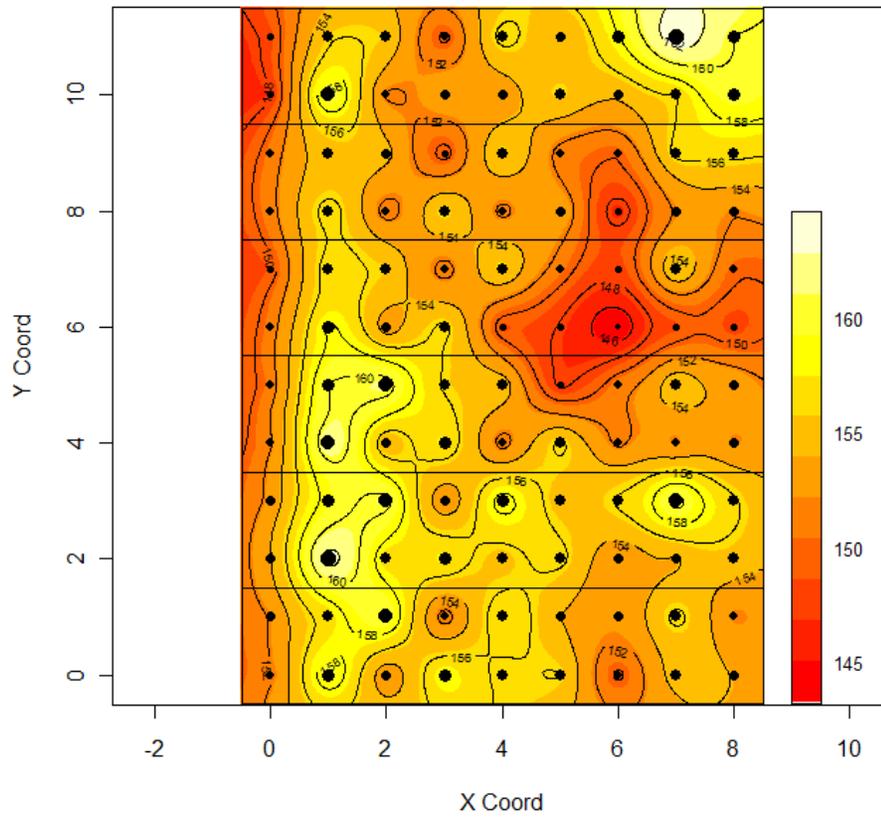


Figura 3 Krigagem para o experimento (i) na configuração Gaus(0-1-6)

## 4.2 Exemplo de abordagem autorregressiva

Para exemplificar a abordagem autorregressiva foi escolhida uma simulação de configuração  $Gaus(0-0,25-6)$  do experimento (i). Os dados utilizados nesse exemplo se encontram na Tabela 28, no Anexo A.

Foram aplicados três padrões de proximidade: primeira ordem, segunda ordem e terceira ordem.

Constatou-se que  $\hat{\rho}$  aumentou de acordo com o aumento dos padrões de proximidade (Tabela 11). Isso era esperado, visto que a configuração utilizada admite uma dependência espacial ao alcance de 6 *u.m.*

Logo, o padrão de proximidade que considera uma maior vizinhança conseguiu detectar melhor a dependência espacial.

As análises de variância com os dados ajustados se encontram nas Tabela 13, Tabela 14 e Tabela 15. A análise de variância clássica está na Tabela 12.

Tabela 11 Estimativas de  $\rho$  para os padrões de proximidade.

	<i>Primeira Ordem</i>	<i>Segunda Ordem</i>	<i>Terceira Ordem</i>
$\hat{\rho}$	0,2685	0,4811	0,8297

Tabela 12 Análise de variância clássica (ANOVA).

<i>FV</i>	<i>GL</i>	<i>SQ</i>	<i>QM</i>	<i>F-value</i>	<i>p-value</i>
Bloco	5	224,63	44,93	86,40	0,0001
Tratamento	17	833,61	49,03	94,28	0,0001
Erro	85	44,30	0,52		
Total	107	1102,54			

Tabela 13 Análise de variância com modelo autorregressivo de primeira ordem (ANOVA-AR1).

<i>FV</i>	<i>GL</i>	<i>SQ</i>	<i>QM</i>	<i>F-value</i>	<i>p-value</i>
Fator ( $\rho$ )	1	102,73	--	--	--
Bloco	5	135,62	27,12	61,63	0,0001
Tratamento	17	826,65	48,62	110,50	0,0001
Erro	84	37,54	0,44	--	--
Total	107	1102,54			

Tabela 14 Análise de variância com modelo autorregressivo de segunda ordem (ANOVA-AR2).

	<i>GL</i>	<i>SQ</i>	<i>QM</i>	<i>F-value</i>	<i>p-value</i>
Fator ( $\rho$ )	1	175,23	--	--	--
Bloco	5	94,78	18,95	43,06	0,0001
Tratamento	17	795,08	46,76	106,27	0,0001
Erro	84	37,45	0,44	--	--
Total	107	1102,54			

Tabela 15 Análise de variância com modelo autorregressivo de terceira ordem (ANOVA-AR3).

	<i>GL</i>	<i>SQ</i>	<i>QM</i>	<i>F-value</i>	<i>p-value</i>
Fator ( $\rho$ )	1	434,12	--	--	--
Bloco	5	51,71	10,34	21,10	0,0001
Tratamento	17	575,12	33,83	69,04	0,0001
Erro	84	41,59	0,49	--	--
Total	107	1102,54			

Comparando a ANOVA clássica com a ANOVA-AR1 notou-se uma redução de 15,38% do erro experimental, que é composto por variações devido a fatores não controlados no experimento (acaso), expressa através da redução do quadrado médio do erro (QME). Para a ANOVA-AR2 a redução foi de 15,38% e para a ANOVA-AR3 a redução foi de 5,76%. Logo, todas as abordagens autorregressivas obtiveram maior precisão do experimento.

O fator ( $\rho$ ) explicou 9,31% da variabilidade do experimento para a ANOVA-AR1; 15,89% para a ANOVA-AR2; 39,37% para a ANOVA-AR3.

Percebe-se que as estimativas de  $\rho$  (Tabela 11) estão diretamente ligadas à explicação atribuída ao fator ( $\rho$ ). Conforme houve aumento da região de vizinhança (padrão de proximidade), houve aumento na explicação atribuída ao fator ( $\rho$ ).

A variabilidade atribuída ao fator tratamento apresentou moderada redução com relação a ANOVA clássica, sendo que essa redução foi de 0,83% para a ANOVA-AR1; 4,62% para a ANOVA-AR2; e 30,95% para a ANOVA-AR3. Esse maior decréscimo do *QMT* da ANOVA-AR3 pode ser atribuído a variabilidade espacial interna do bloco.

A variabilidade atribuída ao fator bloco apresentou grande redução com relação a ANOVA clássica, sendo que essa redução foi de 39,63% para a ANOVA-AR1; 58,04% para a ANOVA-AR2; e 76,98% para a ANOVA-AR3.

A redução do *QMB* de até 76,98% mostra que a abordagem autorregressiva obteve uma excelente explicação da variabilidade espacial, que na ANOVA clássica é modelada apenas pela utilização do fator bloco.

### **4.3 Dados simulados**

Será feita uma abordagem de uma configuração específica do experimento i e do experimento ii.

#### **4.3.1 Experimento na configuração (i)**

Procedeu-se com a análise do experimento i na configuração *Gaus(0-1-6)*, isto é, com erro de configuração gaussiana, com efeito pepita zero, patamar um e alcance seis.

##### **4.3.1.1 Estudo da configuração Gaus (0-1-6)**

A configuração *Gaus(0-1-6)* representa um experimento em que a variabilidade amostral, incluindo a variabilidade espacial entre a menor distância observada, foi compreendida pelo modelo, visto que estamos trabalhando com o efeito pepita zero. Essa configuração também apresenta uma dependência espacial de grande alcance (6 *u.m.*).

Os histogramas apresentados no Gráfico 8. sugerem um decréscimo na variabilidade do *QME*. Esse decréscimo é um pouco mais acentuado quando

observada a ANOVA-AR3. Isso contribui para o experimento, pois sugere que a ANOVA-AR3 tornou o experimento mais preciso.

A mediana para ANOVA, ANOVA-AR1, ANOVA-AR2 e ANOVA-AR3, foi, respectivamente, 1,13; 1,02; 0,92 e 0,83, indicando que a abordagem autorregressiva concentrou 50% das observações em valores mais baixos do que os obtidos pela ANOVA.

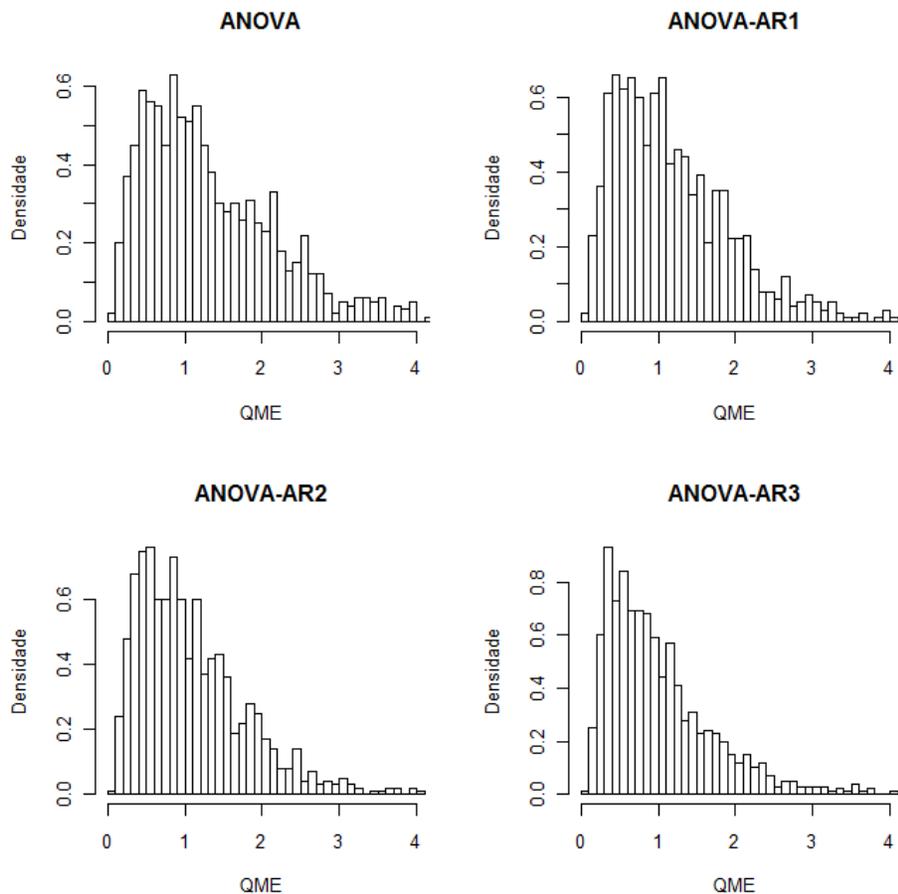


Gráfico 8 Histogramas do QME da configuração Gaus(0-1-6)

De maneira parecida com os histogramas do *QME*, vemos no Gráfico 9 um decréscimo na variabilidade do *QMT*. Aqui o decréscimo é certamente mais acentuado quando observada a ANOVA-AR3.

A mediana para ANOVA, ANOVA-AR1, ANOVA-AR2 e ANOVA-AR3, foi, respectivamente, 40,47; 41,82; 42,41 e 33,61, indicando que a abordagem autorregressiva concentrou 50% das observações de forma similar para a ANOVA, ANOVA-AR1 e ANOVA-AR2, sendo que valores mais baixos foram obtidos apenas na ANOVA-AR3.

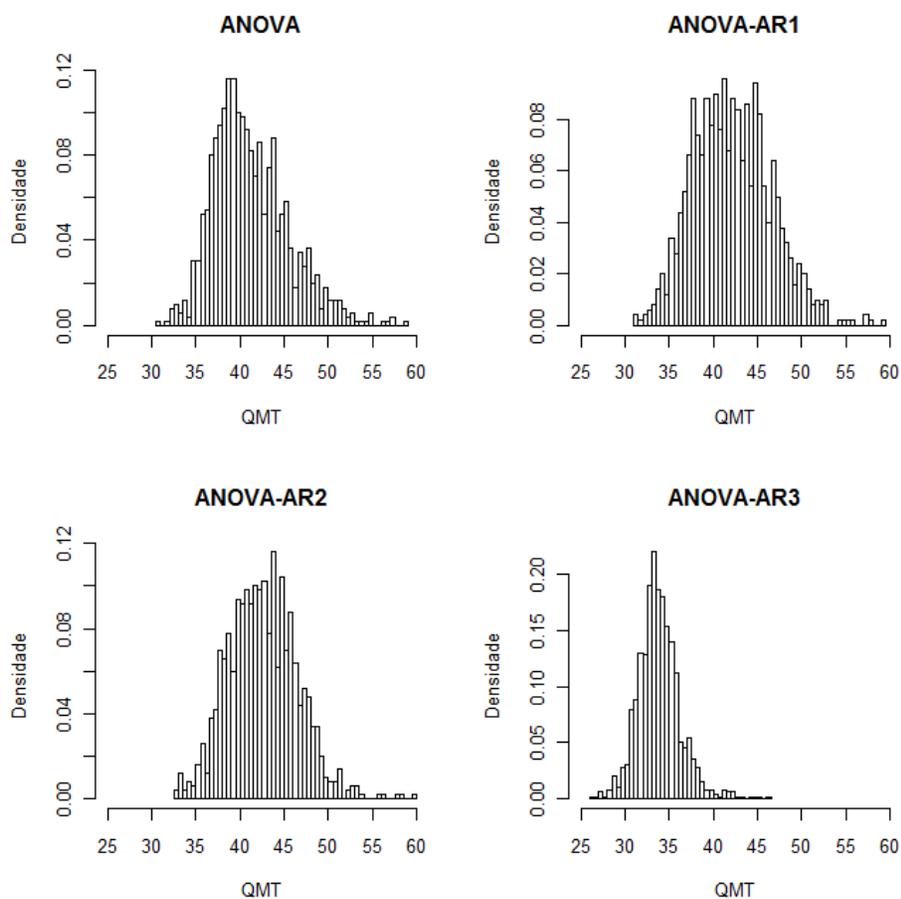


Gráfico 9 Histogramas do QMT da configuração Gaus(0-1-6)

Tabela 16 Intervalo de credibilidade da abordagem autorregressiva para Gaus(0-1-6), com 95% de probabilidade de conter o verdadeiro valor do parâmetro.

	ANOVA		ANOVA-AR1		ANOVA-AR2		ANOVA-AR3	
	Inferior	Superior	Inferior	Superior	Inferior	Superior	Inferior	Superior
$\hat{\rho}$	--	--	-0,11	0,34	-0,06	0,55	0,60	0,86
Fator ( $\rho$ )	--	--	-51,41	197,80	-26,50	304,07	122,20	576,93
<i>QME</i>	0,10	3,02	0,11	2,65	0,10	2,44	0,11	2,36
<i>QMT</i>	34,52	50,79	34,16	50,63	35,16	49,70	28,73	38,42
<i>QMB</i>	12,74	94,36	12,34	82,54	9,68	69,42	4,69	19,90

No Gráfico 10 pode-se verificar a real contribuição da metodologia proposta pela análise de variância com abordagem autorregressiva. Enquanto os histogramas do *QME* e *QMT* apresentaram uma pequena redução em sua abrangência, aqui podemos constatar uma imensa redução da variabilidade do fator bloco, representada pela diminuição considerável do *QMB*.

A mediana para ANOVA, ANOVA-AR1, ANOVA-AR2 e ANOVA-AR3, foi, respectivamente, 49,32; 36,79; 27,50 e 12,08, indicando que a abordagem autorregressiva concentrou 50% das observações em valores mais baixos do que os obtidos pela ANOVA.

Para uma melhor análise, deve-se analisar a Tabela 16 que contém os intervalos de credibilidade obtidos na configuração *Gaus(0-1-6)*.

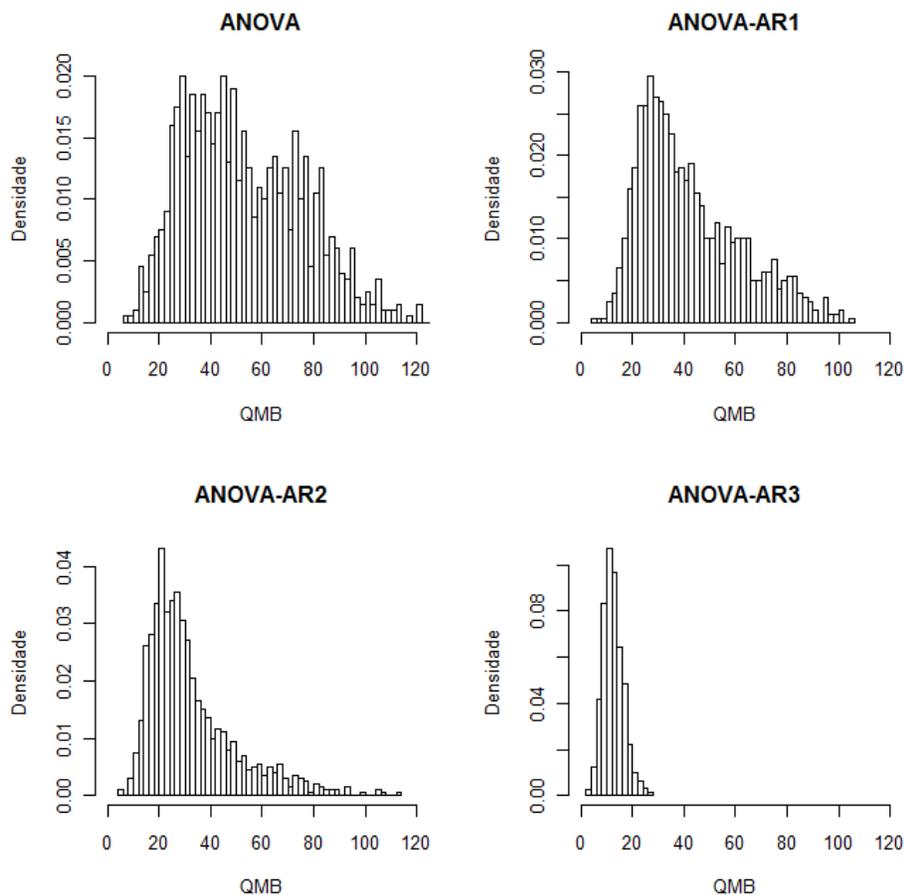


Gráfico 10 Histogramas do QMB da configuração Gaus(0-1-6)

Quanto mais distante de zero for o valor de  $\hat{\rho}$ , maior a variabilidade espacial detectada pelos padrões de proximidade. A variabilidade espacial detectada cresceu na medida em que aumentava a ordem do padrão de proximidade.

O fator ( $\rho$ ) conseguiu explicar a uma maior parte da variabilidade na ANOVA-AR3, embora na ANOVA-AR1 e ANOVA-AR2, sua explicação também foi significativa.

Como sugerido pelos histogramas do Gráfico 8, houve redução do *QME*, embora de forma branda. O mesmo pode ser dito com relação ao *QMT*, todavia, na ANOVA-AR3 houve uma redução mais acentuada, confirmando o que foi sugerido pelos histogramas do Gráfico 9.

O resultado mais interessante, contudo, se atribui a grande redução do *QMB*, sobretudo, quando se analisa a ANOVA versus ANOVA-AR3. O intervalo de credibilidade superior da ANOVA-AR3 ficou próximo do intervalo inferior da ANOVA, mostrando quão grande foi a redução da variabilidade desse fator.

Para todos os parâmetros analisados na Tabela 16 a construção do intervalo de credibilidade para as abordagens autorregressivas foi menor do que a da ANOVA clássica, indicando que essa abordagem obteve maior precisão.

Os histogramas para o AIC do modelo com erro independente e AIC do modelo com erro correlacionado (Gráfico 11), visualmente, não apresentaram muita diferença. A mediana do modelo com erro independente foi de 344,76, sendo que a mediana do modelo com erro dependente foi de 309,93, indicando um acúmulo de 50% das amostras em um valor um pouco inferior. Deve-se proceder com uma análise mais aprofundada através do intervalo de credibilidade.

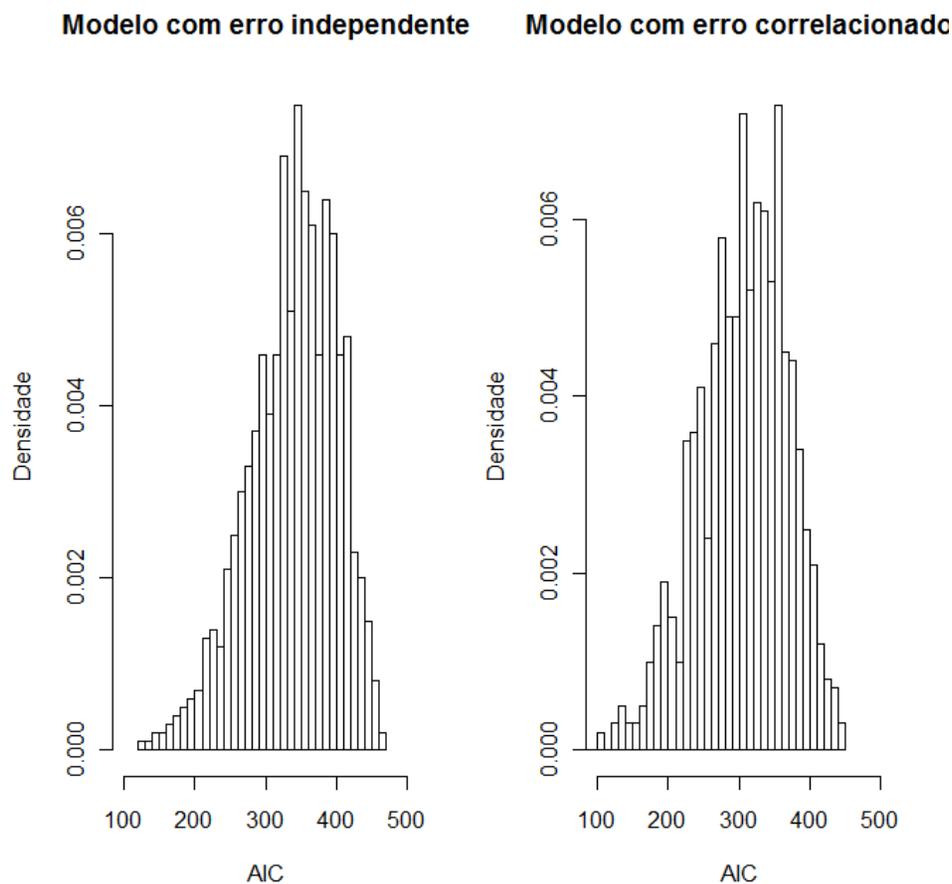


Gráfico 11 Histogramas do AIC da configuração Gaus(0-1-6)

Analisando a Tabela 17, constatamos que a utilização da abordagem Geoestatística propiciou um melhor ajuste do modelo ao conjunto de dados, representado por menores valores do intervalo superior e inferior do modelo com erro correlacionado, se comparado ao modelo com erro independente.

Temos então, que para a configuração *Gaus(0-1-6)*, tanto a abordagem autorregressiva, quanto a abordagem Geoestatística conseguiram explicar melhor o experimento.

Tabela 17 Intervalo de credibilidade para o AIC da abordagem Geoestatística para Gaus(0-1-6), com 95% de probabilidade de conter o verdadeiro valor do parâmetro.

	<i>AIC</i>			
	<b>Modelo com erro independente</b>		<b>Modelo com erro correlacionado</b>	
	<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>
<i>Gaus(0-1-6)</i>	219,07	450,51	185,88	425,54

#### 4.3.1.2 Estudo dos demais casos

Fixando o valor do alcance do semivariograma em 6 e trabalhando com outros valores para o efeito pepita e patamar, obtiveram-se resultados semelhantes ao já descrito. Houve redução pequena e moderada para o *QME* e *QMT*, respectivamente; e grande redução para o *QMB* (Tabela 18).

Tabela 18 Intervalo de credibilidade da abordagem autorregressiva para todas as configurações de Gaus no alcance 6, com 95% de probabilidade de conter o verdadeiro valor do parâmetro.

	<i>GAUS(0-1-6)</i>							
	<b>ANOVA</b>		<b>ANOVA-AR1</b>		<b>ANOVA-AR2</b>		<b>ANOVA-AR3</b>	
	<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>
$\hat{\rho}$	--	--	-0,11	0,34	-0,06	0,55	0,60	0,86
Fator ( $\rho$ )	--	--	-51,41	197,80	-26,50	304,07	122,20	576,93
<i>QME</i>	0,10	3,02	0,11	2,65	0,10	2,44	0,11	2,36
<i>QMT</i>	34,52	50,79	34,16	50,63	35,16	49,70	28,73	38,42
<i>QMB</i>	12,74	94,36	12,34	82,54	9,68	69,42	4,69	19,90

Tabela 18, conclusão.

		<b>GAUS(0,25-1-6)</b>							
		<b>ANOVA</b>		<b>ANOVA-AR1</b>		<b>ANOVA-AR2</b>		<b>ANOVA-AR3</b>	
		<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>
$\hat{\rho}$		--	--	-0,15	0,21	-0,12	0,43	0,57	0,82
Fator ( $\rho$ )		--	--	-33,93	89,44	-36,31	151,70	117,69	433,81
<i>QME</i>		0,21	1,90	0,16	1,82	0,23	1,83	0,19	1,81
<i>QMT</i>		34,18	47,18	33,63	48,26	34,20	48,15	29,51	39,49
<i>QMB</i>		21,21	77,97	19,14	72,75	14,64	63,03	6,94	19,64
		<b>GAUS(0,75-1-6)</b>							
		<b>ANOVA</b>		<b>ANOVA-AR1</b>		<b>ANOVA-AR2</b>		<b>ANOVA-AR3</b>	
		<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>
$\hat{\rho}$		--	--	-0,13	0,16	-0,13	0,33	0,59	0,78
Fator ( $\rho$ )		--	--	-31,07	37,59	-23,30	84,90	142,44	348,72
<i>QME</i>		0,28	1,09	0,27	1,08	0,23	1,04	0,30	1,05
<i>QMT</i>		34,08	44,29	33,69	44,59	34,40	45,10	29,57	37,78
<i>QMB</i>		25,65	62,61	26,95	65,05	21,72	57,96	8,09	16,78
		<b>GAUS(0-0,25-6)</b>							
		<b>ANOVA</b>		<b>ANOVA-AR1</b>		<b>ANOVA-AR2</b>		<b>ANOVA-AR3</b>	
		<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>
$\hat{\rho}$		--	--	0,06	0,60	0,18	0,75	0,62	0,90
Fator ( $\rho$ )		--	--	-1,75	793,02	-0,52	902,84	239,91	1214,50
<i>QME</i>		0,28	11,81	0,28	9,04	0,28	9,07	0,45	9,85
<i>QMT</i>		33,21	74,36	37,12	65,76	36,86	62,95	27,62	46,07
<i>QMB</i>		10,73	187,47	3,15	113,85	2,38	80,92	2,58	34,02

Fixando o valor do alcance do semivariograma em 4 e trabalhando com outros valores para o efeito pepita e patamar, obtiveram-se resultados semelhantes ao já descrito. Houve redução pequena e moderada para o *QME* e *QMT*, respectivamente; e grande redução para o *QMB* (Tabela 19).

Tabela 19 Intervalo de credibilidade da abordagem autorregressiva para todas as configurações de Gaus no alcance 4, com 95% de probabilidade de conter o verdadeiro valor do parâmetro.

		<i>GAUS(0-1-4)</i>							
		<b>ANOVA</b>		<b>ANOVA-AR1</b>		<b>ANOVA-AR2</b>		<b>ANOVA-AR3</b>	
		<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>
$\hat{\rho}$		--	--	-0,09	0,36	-0,04	0,57	0,60	0,86
Fator ( $\rho$ )		--	--	-42,51	193,07	-37,55	289,23	140,05	594,68
<i>QME</i>		0,12	3,13	0,12	2,76	0,12	2,57	0,16	2,52
<i>QMT</i>		34,46	51,62	34,44	51,69	34,75	50,99	29,14	39,30
<i>QMB</i>		14,13	94,30	11,41	80,41	10,35	71,71	6,00	20,10
		<i>GAUS(0,25-1-4)</i>							
		<b>ANOVA</b>		<b>ANOVA-AR1</b>		<b>ANOVA-AR2</b>		<b>ANOVA-AR3</b>	
		<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>
$\hat{\rho}$		--	--	-0,14	0,21	-0,14	0,41	0,59	0,81
Fator ( $\rho$ )		--	--	-31,59	74,58	-21,46	130,92	127,35	406,05
<i>QME</i>		0,20	1,61	0,20	1,59	0,21	1,55	0,19	1,55
<i>QMT</i>		34,66	46,29	34,17	47,26	34,64	47,90	30,05	38,96
<i>QMB</i>		21,52	70,75	22,13	70,59	18,14	61,07	7,68	18,68
		<i>GAUS(0,75-1-4)</i>							
		<b>ANOVA</b>		<b>ANOVA-AR1</b>		<b>ANOVA-AR2</b>		<b>ANOVA-AR3</b>	
		<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>
$\hat{\rho}$		--	--	-0,12	0,13	-0,12	0,30	0,60	0,77
Fator ( $\rho$ )		--	--	-20,96	37,11	-20,60	74,40	144,03	330,58
<i>QME</i>		0,29	1,01	0,30	1,02	0,29	1,01	0,28	1,00
<i>QMT</i>		34,20	43,98	33,92	44,47	34,76	45,54	29,72	37,86
<i>QMB</i>		25,94	60,23	26,37	59,35	20,73	52,47	8,58	16,17
		<i>GAUS(0-0,25-4)</i>							
		<b>ANOVA</b>		<b>ANOVA-AR1</b>		<b>ANOVA-AR2</b>		<b>ANOVA-AR3</b>	
		<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>
$\hat{\rho}$		--	--	0,02	0,54	0,16	0,74	0,61	0,90
Fator ( $\rho$ )		--	--	-3,58	562,97	-1,39	657,36	165,28	958,32
<i>QME</i>		0,69	9,34	0,40	7,46	0,41	7,24	0,30	7,48
<i>QMT</i>		33,03	66,75	35,48	61,50	36,29	59,52	27,77	44,38
<i>QMB</i>		7,71	151,60	5,84	101,69	2,63	70,29	3,04	30,68

Fixando o valor do alcance do semivariograma em 2 e trabalhando com outros valores para o efeito pepita e patamar obtiveram-se resultados semelhantes ao já descrito. Houve redução pequena e moderada para o *QME* e *QMT*, respectivamente; e grande redução para o *QMB* (Tabela 20)

Tabela 20 Intervalo de credibilidade da abordagem autorregressiva para todas as configurações de Gaus no alcance 2, com 95% de probabilidade de conter o verdadeiro valor do parâmetro.

		<i>GAUS(0-I-2)</i>							
		<b>ANOVA</b>		<b>ANOVA-AR1</b>		<b>ANOVA-AR2</b>		<b>ANOVA-AR3</b>	
		<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>
$\hat{\rho}$		--	--	-0,12	0,24	-0,11	0,43	0,60	0,81
Fator ( $\rho$ )		--	--	-24,05	83,42	-24,88	139,07	138,95	418,73
<i>QME</i>		0,31	1,73	0,29	1,67	0,28	1,61	0,23	1,58
<i>QMT</i>		34,99	46,44	34,30	47,32	35,19	48,13	30,05	38,51
<i>QMB</i>		21,68	71,53	20,25	64,46	16,34	53,75	7,31	18,77
		<i>GAUS(0,25-I-2)</i>							
		<b>ANOVA</b>		<b>ANOVA-AR1</b>		<b>ANOVA-AR2</b>		<b>ANOVA-AR3</b>	
		<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>
$\hat{\rho}$		--	--	-0,13	0,18	-0,10	0,39	0,60	0,80
Fator ( $\rho$ )		--	--	-22,16	55,87	-26,46	96,23	141,49	366,82
<i>QME</i>		0,28	1,27	0,28	1,27	0,27	1,23	0,26	1,22
<i>QMT</i>		34,78	44,89	34,23	45,64	34,69	46,55	30,25	38,37
<i>QMB</i>		25,71	65,25	24,83	61,98	19,99	52,31	8,01	17,39
		<i>GAUS(0,75-I-2)</i>							
		<b>ANOVA</b>		<b>ANOVA-AR1</b>		<b>ANOVA-AR2</b>		<b>ANOVA-AR3</b>	
		<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>
$\hat{\rho}$		--	--	-0,12	0,13	-0,09	0,32	0,62	0,78
Fator ( $\rho$ )		--	--	-19,52	31,39	-17,45	66,54	148,98	318,08
<i>QME</i>		0,28	0,84	0,29	0,83	0,28	0,83	0,26	0,82
<i>QMT</i>		35,12	43,65	34,46	44,02	35,24	45,18	30,34	37,72
<i>QMB</i>		28,22	56,93	28,51	57,22	23,79	49,25	8,57	15,85

Tabela 20, conclusão.

	<b>GAUS(0-0,25-2)</b>							
	<b>ANOVA</b>		<b>ANOVA-AR1</b>		<b>ANOVA-AR2</b>		<b>ANOVA-AR3</b>	
	<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>
$\hat{\rho}$	--	--	-0,01	0,45	0,05	0,61	0,53	0,86
Fator ( $\rho$ )	--	--	-10,72	333,64	-0,46	421,79	135,84	703,74
<i>QME</i>	1,22	6,93	0,87	6,02	1,03	6,13	0,87	6,22
<i>QMT</i>	33,10	57,58	33,95	57,08	35,25	57,94	28,09	45,53
<i>QMB</i>	12,42	112,70	9,19	82,27	6,79	61,36	3,21	29,11

Logo, para todas as configurações apresentadas na Tabela 6, para o experimento i, a abordagem autorregressiva apresentou os mesmos padrões de resposta, indiferente da forma a qual o erro espacial foi gerado.

Tabela 21 Intervalo de credibilidade para o AIC de todas as configurações do experimento i, com 95% de probabilidade de conter o verdadeiro valor do parâmetro.

	<b>AIC</b>			
	<b>Modelo com erro independente</b>		<b>Modelo com erro correlacionado</b>	
	<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>
<i>Gaus(0-1-6)</i>	219,07	450,51	185,88	425,54
<i>Gaus(0,25-1-6)</i>	233,32	400,54	221,06	391,80
<i>Gaus(0,75-1-6)</i>	236,64	352,65	233,64	346,71
<i>Gaus(0-0,25-6)</i>	330,32	560,73	303,12	543,07
<i>Gaus(0-1-4)</i>	222,33	445,67	188,68	421,31

Tabela 21, conclusão.

<i>Gaus(0,25-1-4)</i>	231,79	388,71	220,01	384,13
<i>Gaus(0,75-1-4)</i>	231,97	336,62	226,68	335,25
<i>Gaus(0-0,25-4)</i>	340,17	535,58	309,26	515,28
<i>Gaus(0-1-2)</i>	249,79	389,10	227,02	381,95
<i>Gaus(0,25-1-2)</i>	247,59	361,78	234,25	360,21
<i>Gaus(0,75-1-2)</i>	234,38	325,32	225,18	321,17
<i>Gaus(0-0,25-2)</i>	367,62	506,93	344,86	499,78

O critério de seleção de modelos AIC também apresentou melhores resultados para a abordagem Geoestatística, isto é, tanto os limites inferiores do intervalo de credibilidade, quanto superiores, para o modelo com erro correlacionado, apresentaram valores menores do que os do modelo com erro independente, indiferente da configuração de erro utilizada (Tabela 21).

### 4.3.2 Experimento na configuração ii

Procedeu-se com a análise do experimento ii na configuração *Gaus(0-1-2)*, isto é, com erro de configuração gaussiana, com efeito pepita zero, patamar um e alcance dois.

#### 4.3.2.1 Estudo da configuração Gaus (0-1-2)

A configuração *Gaus(0-1-2)* representa um experimento em que a variabilidade amostral, incluindo a variabilidade espacial entre a menor distância observada, foi compreendida pelo modelo, visto que estamos trabalhando com o efeito pepita zero. Essa configuração também apresenta uma dependência espacial de grande alcance ( $2 \text{ u.m.}$ ).

Diferentemente do que foram apresentados nas configurações do experimento i, os histogramas apresentados no Gráfico 12 não sugerem um decréscimo na variabilidade do *QME*, mas sim valores constantes para as quatro ANOVAs. A análise da mediana nos remete a mesma conclusão, visto que os valores obtidos foram de 0,15; 0,15; 0,15 e 0,13 para ANOVA, ANOVA-AR1, ANOVA-AR2 e ANOVA-AR3.

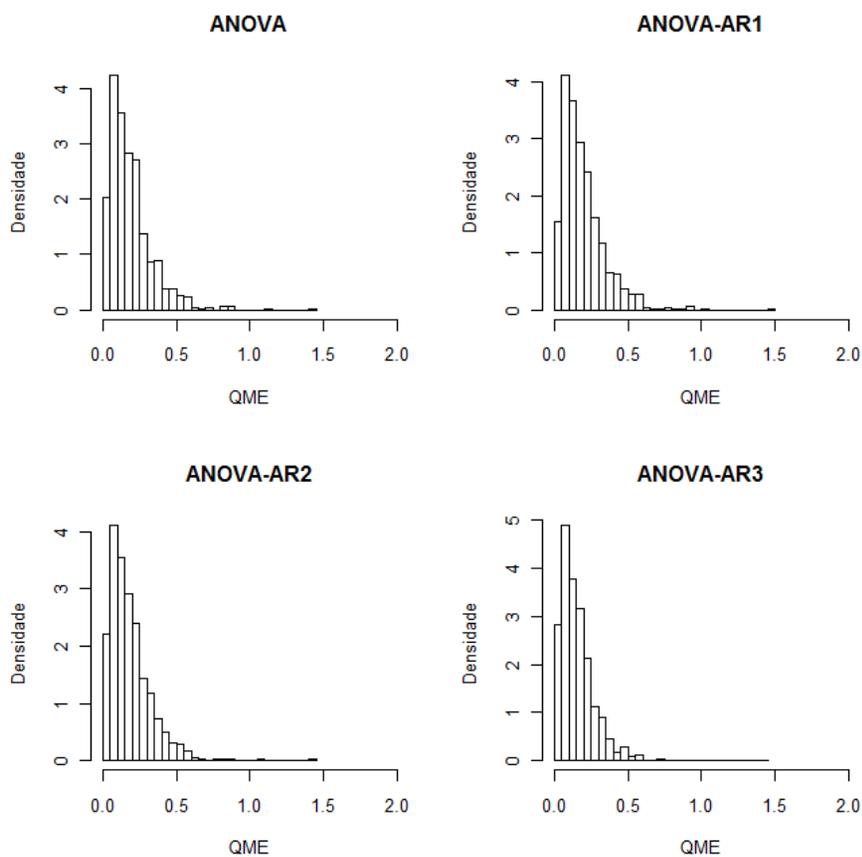


Gráfico 12 Histogramas do QME para a configuração Gaus(0-1-2)

Para os histogramas do  $QMT$ , apresentados no Gráfico 13, pode-se perceber uma pequena redução do  $QMT$ , na ANOVA-AR1, todavia, o  $QMT$  aumenta consideravelmente para a ANOVA-AR2 e ANOVA-AR3. Os valores da mediana foram de 4,66 para a ANOVA; 3,83 para a ANOVA-AR1; 4,23 para a ANOVA-AR2 e 5,04 para a ANOVA-AR3.

Diferente do que ocorreu no experimento (ii), onde com o aumento da vizinhança ocorria diminuição dos parâmetros, para o experimento (i) na configuração  $Gaus(0-1-2)$  isso não ocorre.

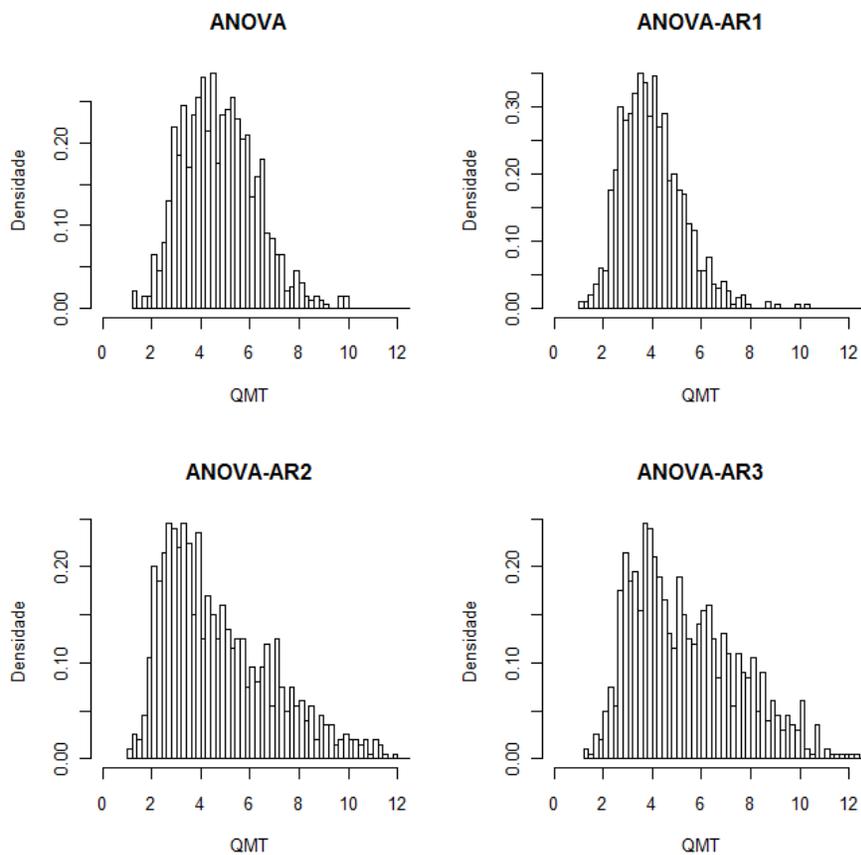


Gráfico 13 Histogramas do QMT para a configuração Gaus(0-1-2)

Com os histogramas do *QMB* (Gráfico 14) percebe-se algo similar ao apresentado pelos histogramas do *QMT*. Aqui, todavia, há uma redução mais acentuada para a ANOVA-AR1 e, posteriormente, um grande acréscimo com a ANOVA-AR2 e ANOVA-AR3. A mediana para a ANOVA foi de 13,30; para a ANOVA-AR1 foi de 8,72; para a ANOVA-AR2 foi de 21,99 e para a ANOVA-AR3 foi de 17,51.

Constata-se algo similar ao ocorrido com o *QMT*. Para a ANOVA-AR1 50% dos valores amostrados se encontraram abaixo do que foi amostrado para a

ANOVA, mas ao se verificar a ANOVA-AR2 e ANOVA-AR3 há uma grande inflação do valor da mediana.

Para uma melhor compreensão, devem-se analisar os intervalos de credibilidade, conforme a Tabela 22.

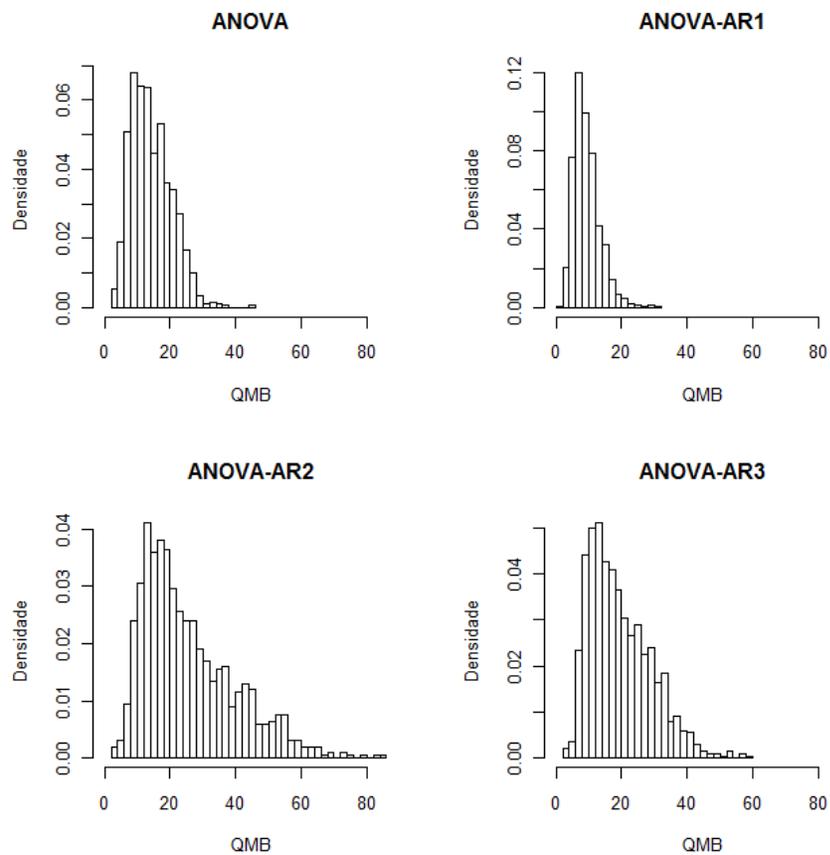


Gráfico 14 Histogramas do QMB para a configuração Gaus(0-1-2)

Tabela 22 Intervalo de credibilidade da abordagem autorregressiva para Gaus(0-1-2), com 95% de probabilidade de conter o verdadeiro valor do parâmetro.

	<i>GAUS(0-1-2)</i>							
	ANOVA		ANOVA-AR1		ANOVA-AR2		ANOVA-AR3	
	Inferior	Superior	Inferior	Superior	Inferior	Superior	Inferior	Superior
$\hat{\rho}$	--	--	0,04	0,57	-1,29	-0,22	-1,03	-0,07
Fator ( $\rho$ )	--	--	0,51	81,64	-173,68	11,80	-97,96	5,85
<i>QME</i>	0,01	0,46	0,02	0,44	0,01	0,44	0,01	0,37
<i>QMT</i>	1,91	7,40	1,78	6,45	1,33	9,31	2,14	10,06
<i>QMB</i>	4,59	25,83	3,17	17,58	5,87	54,42	6,16	38,46

Um fato interessante pode ser observado para  $\hat{\rho}$ . Em todos os casos ele conseguiu detectar a variabilidade espacial, todavia, apenas para a ANOVA-AR1 esse valor ficou com índice positivo. Para os demais casos ele apresentou índices negativos. Isso acabou inflacionando os valores do *QMT* e *QMB*, aumentando, conseqüentemente a variabilidade desses fatores, ao invés de diminuí-los. Isso pode ser percebido através do intervalo de credibilidade do fator ( $\rho$ ). Nesse intervalo o fator ( $\rho$ ) ficou com maior probabilidade de ocorrer com índice negativo, resultando no aumento do *QMT* e *QMB* ao invés do esperado que era a redução.

Podemos levantar como hipótese o fato do experimento (ii) conter apenas 20 observações. Dessa maneira, os padrões de proximidade de segunda e terceira ordem acabaram sobrepondo informações de vizinhança inúmeras vezes, devido ao seu raio ser maior do que o padrão de proximidade de primeira ordem (Tabela 3).

A ANOVA-AR1 apresentou diminuição da variabilidade para todos os fatores de interesse ( $QME$ ,  $QMT$  e  $QMB$ ), todavia de forma bem menos acentuada do que os apresentados no experimento (i).

Os histogramas para o AIC do modelo com erro independente e AIC do modelo com erro correlacionado (Gráfico 15), visualmente, não apresentaram muita diferença. A mediana para o modelo com erro independente foi de 63,27, enquanto que a do modelo com erro correlacionado foi de 62,99, indicando um acúmulo de 50% das amostras em um valor minimamente inferior. Deve-se proceder com uma análise mais aprofundada através do intervalo de credibilidade.

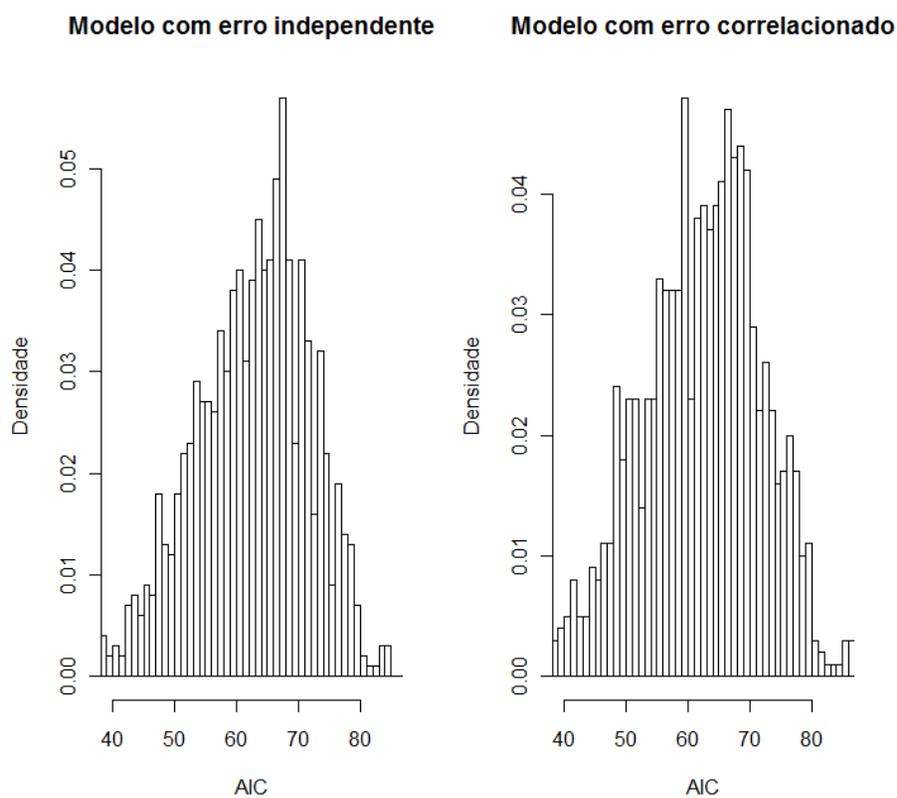


Gráfico 15 Histogramas do AIC para a configuração Gaus(0-1-2)

Tabela 23 Intervalo de credibilidade para o AIC da abordagem Geoestatística para  $Gaus(0-1-2)$ , com 95% de probabilidade de conter o verdadeiro valor do parâmetro.

	<i>AIC</i>			
	<b>Modelo com erro independente</b>		<b>Modelo com erro correlacionado</b>	
	<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>
<i>Gaus(0-1-2)</i>	43,87	79,08	43,34	79,95

De fato, os intervalos de credibilidade para os dois modelos (Tabela 23), são praticamente iguais, corroborando com os resultados obtidos na ANOVA-AR1, isto é, a abordagem espacial consegue melhorar a precisão do experimento, mas devido a pequena quantidade de observações, sua melhora no modelo é muito pequena, em contraste com o que se é obtido para experimentos maiores (experimento (ii)).

#### 4.3.2.2 Estudo dos demais casos

Fixando o valor do alcance do semivariograma em 2 e trabalhando com outros valores para o efeito pepita e patamar, obtivemos resultados semelhantes ao descrito para a configuração  $Gaus(0-1-2)$ . Houve pequena redução do  $QME$ ,  $QMT$  e  $QMB$  da ANOVA-AR1, e uma inflação dos valores do  $QMT$  e  $QMB$  para a ANOVA-AR2 e ANOVA-AR3 (Tabela 24).

Tabela 24 Intervalo de credibilidade da abordagem autorregressiva para todas as configurações de Gaus no alcance 2, com 95% de probabilidade de conter o verdadeiro valor do parâmetro.

		<i>GAUS(0-1-2)</i>							
		<b>ANOVA</b>		<b>ANOVA-AR1</b>		<b>ANOVA-AR2</b>		<b>ANOVA-AR3</b>	
		<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>
$\hat{\rho}$		--	--	0,04	0,57	-1,29	-0,22	-1,03	-0,07
Fator ( $\rho$ )		--	--	0,51	81,64	-173,68	11,80	-97,96	5,85
<i>QME</i>		0,01	0,46	0,02	0,44	0,01	0,44	0,01	0,37
<i>QMT</i>		1,91	7,40	1,78	6,45	1,33	9,31	2,14	10,06
<i>QMB</i>		4,59	25,83	3,17	17,58	5,87	54,42	6,16	38,46
		<i>GAUS(0,25-1-2)</i>							
		<b>ANOVA</b>		<b>ANOVA-AR1</b>		<b>ANOVA-AR2</b>		<b>ANOVA-AR3</b>	
		<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>
$\hat{\rho}$		--	--	0,03	0,57	-1,27	-0,34	-1,00	-0,16
Fator ( $\rho$ )		--	--	-0,38	75,67	-157,05	6,01	-84,34	0,17
<i>QME</i>		0,02	0,31	0,02	0,33	0,01	0,29	0,01	0,24
<i>QMT</i>		2,31	6,59	2,02	5,94	1,68	7,65	2,20	8,19
<i>QMB</i>		5,89	22,85	4,28	15,59	10,06	50,79	8,27	33,44
		<i>GAUS(0,75-1-2)</i>							
		<b>ANOVA</b>		<b>ANOVA-AR1</b>		<b>ANOVA-AR2</b>		<b>ANOVA-AR3</b>	
		<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>
$\hat{\rho}$		--	--	0,10	0,54	-1,27	-0,53	-0,92	-0,23
Fator ( $\rho$ )		--	--	3,16	61,67	-132,68	-8,82	-69,15	-7,05
<i>QME</i>		0,02	0,20	0,02	0,22	0,02	0,19	0,02	0,16
<i>QMT</i>		2,67	5,98	2,39	5,48	1,70	6,14	2,58	7,10
<i>QMB</i>		7,45	20,13	4,82	13,22	14,86	46,00	10,18	28,63
		<i>GAUS(0-0,25-2)</i>							
		<b>ANOVA</b>		<b>ANOVA-AR1</b>		<b>ANOVA-AR2</b>		<b>ANOVA-AR3</b>	
		<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>
$\hat{\rho}$		--	--	-0,22	0,61	-1,20	0,28	-1,33	0,13
Fator ( $\rho$ )		--	--	-31,96	137,03	-275,41	62,31	-225,89	56,33
<i>QME</i>		0,04	1,85	0,04	1,80	0,04	1,71	0,05	1,46
<i>QMT</i>		1,44	13,49	0,77	11,80	0,81	18,61	1,27	19,91
<i>QMB</i>		0,94	41,98	0,96	34,49	1,24	80,83	0,66	67,58

Fixando o valor do alcance do semivariograma em 1 e trabalhando com outros valores para o efeito pepita e patamar, obtivemos, novamente, resultados semelhantes ao descrito para a configuração *Gaus(0-1-2)*. Houve pequena redução do *QME*, *QMT* e *QMB* da ANOVA-AR1, e uma inflação dos valores do *QMT* e *QMB* para a ANOVA-AR2 e ANOVA-AR3 (Tabela 25).

Tabela 25 Intervalo de credibilidade da abordagem autorregressiva para todas as configurações de Gaus no alcance 2, com 95% de probabilidade de conter o verdadeiro valor do parâmetro.

		<i>GAUS(0-1-1)</i>							
		ANOVA		ANOVA-AR1		ANOVA-AR2		ANOVA-AR3	
		Inferior	Superior	Inferior	Superior	Inferior	Superior	Inferior	Superior
$\hat{\rho}$		--	--	0,04	0,55	-1,32	-0,35	-1,02	-0,14
Fator ( $\rho$ )		--	--	-1,60	69,82	-153,32	4,96	-86,77	0,56
<i>QME</i>		0,02	0,34	0,03	0,36	0,01	0,32	0,02	0,28
<i>QMT</i>		2,20	6,84	1,96	6,04	1,68	7,58	2,06	8,34
<i>QMB</i>		5,99	23,81	4,18	16,40	9,21	50,24	6,95	33,73
		<i>GAUS(0,25-1-1)</i>							
		ANOVA		ANOVA-AR1		ANOVA-AR2		ANOVA-AR3	
		Inferior	Superior	Inferior	Superior	Inferior	Superior	Inferior	Superior
$\hat{\rho}$		--	--	0,08	0,57	-1,29	-0,47	-0,94	-0,23
Fator ( $\rho$ )		--	--	2,75	66,07	-143,90	-9,04	-75,86	-4,20
<i>QME</i>		0,02	0,27	0,03	0,28	0,02	0,25	0,01	0,22
<i>QMT</i>		2,31	6,36	2,00	5,54	1,78	6,90	2,35	7,64
<i>QMB</i>		6,76	21,47	3,92	14,84	11,59	48,21	9,31	31,23
		<i>GAUS(0,75-1-1)</i>							
		ANOVA		ANOVA-AR1		ANOVA-AR2		ANOVA-AR3	
		Inferior	Superior	Inferior	Superior	Inferior	Superior	Inferior	Superior
$\hat{\rho}$		--	--	0,12	0,52	-1,28	-0,59	-0,88	-0,26
Fator ( $\rho$ )		--	--	10,55	62,79	-129,34	-12,98	-64,09	-6,05
<i>QME</i>		0,02	0,18	0,02	0,20	0,01	0,17	0,02	0,15
<i>QMT</i>		2,64	5,87	2,17	5,02	2,14	6,02	2,54	6,59

Tabela 25, conclusão.

<i>QMB</i>	7,95	19,98	4,92	13,19	14,58	44,21	11,59	29,39
	<i>GAUS(0-0,25-1)</i>							
	<b>ANOVA</b>		<b>ANOVA-AR1</b>		<b>ANOVA-AR2</b>		<b>ANOVA-AR3</b>	
	Inferior	Superior	Inferior	Superior	Inferior	Superior	Inferior	Superior
$\hat{\rho}$	--	--	-0,25	0,62	-1,21	0,11	-1,28	0,04
Fator ( $\rho$ )	--	--	-37,20	108,51	-227,50	48,18	-194,89	23,42
<i>QME</i>	0,08	1,34	0,09	1,30	0,08	1,24	0,08	1,14
<i>QMT</i>	1,07	10,98	0,88	10,12	0,73	14,12	1,10	15,53
<i>QMB</i>	0,89	36,64	1,29	31,98	0,78	72,74	1,45	58,26

O critério de seleção de modelos AIC não apresentou melhores resultados para a abordagem Geoestatística, isto é, tanto os limites inferiores do intervalo de credibilidade, quanto superiores, para o modelo com erro correlacionado, apresentaram valores equivalentes aos do modelo com erro independente, indiferente da configuração de erro utilizada (Tabela 26).

Tabela 26 Intervalo de credibilidade para o AIC de todas as configurações do experimento ii, com 95% de probabilidade de conter o verdadeiro valor do parâmetro.

	<i>AIC</i>			
	<b>Modelo 1</b>		<b>Modelo 2</b>	
	<b>Inferior</b>	<b>Superior</b>	<b>Inferior</b>	<b>Superior</b>
<i>Gaus(0-1-2)</i>	43,87	79,08	43,34	79,95
<i>Gaus(0,25-1-2)</i>	45,89	75,23	45,70	75,90
<i>Gaus(0,75-1-2)</i>	43,37	68,42	44,26	70,45
<i>Gaus(0-0,25-2)</i>	60,51	95,71	59,98	96,58
<i>Gaus(0-1-1)</i>	46,03	74,14	46,31	75,90
<i>Gaus(0,25-1-1)</i>	45,10	71,78	46,12	74,18
<i>Gaus(0,75-1-1)</i>	42,85	66,66	43,60	68,23
<i>Gaus(0-0,25-1)</i>	63,23	91,98	63,60	93,98

## 5 CONCLUSÃO

A utilização de ferramentas da estatística espacial, na análise de experimentos, torna a análise mais precisa, haja vista que há diminuição na variabilidade geral do experimento e/ou uma melhor modelagem do experimento, conforme apresentado nas tabelas de AIC.

No experimento (i) todas as configurações apresentaram melhores resultados quando abordadas por análises que consideram a informação espacial. Vale ressaltar que a redução da variabilidade do bloco foi muito expressiva, o que leva a crer que as análises que consideram a informação espacial conseguem “compreender” a variabilidade espacial do solo, mais do que apenas a utilização do fator bloco.

Fica a sugestão de se utilizar um delineamento inteiramente casualizado (DIC) agregado a abordagem autorregressiva, para quando houver heterogeneidade do solo. Dessa maneira, “perder-se-ia” apenas um grau de liberdade para o fator ( $\rho$ ), e não uma quantidade maior que geralmente é utilizada para o fator bloco. Dessa maneira o GL do erro ficaria maior e conseqüentemente se diminuiria ainda mais a variabilidade geral do experimento.

No experimento (ii) os resultados não foram tão expressivos quando no experimento (i). A abordagem autorregressiva de primeira ordem apresentou resultados satisfatórios, sendo que as demais ordens devem ser ignoradas, devido a inflacionarem a variabilidade do experimento, ao invés do contrário.

Logo se sugere, para experimentos semelhantes ao (ii) a utilização apenas da ANOVA-AR1.

Em suma, tivemos que as análises que fazem uso da informação espacial (modelo autorregressivo e modelo Geoestatístico), quando deparadas com

dependências espaciais em diversas configurações, apresentaram melhores resultados do que a usual abordagem clássica.

## REFERÊNCIAS

BAILEY, T. C.; GATRELL, A. C. **Interactive spatial data analysis**. London: Longman Scientific and Technical, 1995. 413 p.

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

CHILÈS, J.; DELFINER, P. **Geostatistics: modeling spatial uncertainty**. New York: Wiley Interscience, 1999. 695 p.

CLARK, I. **Practical geostatistics**. London: Applied Science, 1979. 129 p.

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

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

GRIFFITH, D. A. Estimating spatial autoregressive model parameters with commercial statistical package. **Geographical Analysis**, Ohio, v. 20, n. 1, p. 176-186, Feb. 1988.

GUERRA, P. A. G. **Geostatística operacional**. Brasília: Ministério das Minas e Energia, 1988. 145 p.

GUMPERTZ, M. L. et al. Autologistic model of spatial pattern of phytophthora epidemic in bell pepper: effects of soil variables on disease presence. **Journal of Agricultural, Biological and Environmental Statistics**, New York, v. 2, n. 1, p. 131-156, Mar. 1997.

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

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

KRIGE, D. G. A. Statistical approach to some basic mine valuation problems on the witwatersrand. **Journal of the Chemical, Metallurgical and Mining Society of South Africa**, Johannesburg, v. 6, n. 52, p. 119-139, June 1951.

LONG, D. S. **Practical handbook of spatial statistics**. London: CRC, 1996. 307 p.

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

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

ORD, J. K. Estimation methods for models of spatial interaction. **Journal of the American Statistical Association**, Alexandria, v.70, n. 3, p. 120-126, Nov. 1975.

PINHEIRO, J. C.; BATES, D. M. **Mixed-effects models in S and S-PLUS**. New York: Springer-Verlag, 2000. 521 p.

R DEVELOPMENT CORE TEAM. **R: A language and environment for statistical computing**. Versão 12.2.1. Vienna: R Foundation for Statistical Computing, 2010. Disponível em: <<http://www.r-project.org>>. Acesso em: 21 jun. 2010.

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

SMIRNOV, O.; ANSELIN, L. Fast maximum likelihood estimation of very large spatial autoregressive models: a characteristic polynomial approach. **Computational Statistics and Data Analysis**, Amsterdam, v. 35, n. 8, p. 301-319, Sept. 2001.

SMITH, L. H. Plot arrangement for variety experiment with corn. **Proceedings of the American Society of Agronomy**, New York, v. 107, n. 9, p. 84-89, Nov. 1910.

STORCK, L. et al. **Experimentação vegetal**. Santa Maria: Editora UFSM, 2000. 198 p.

WAYNICK, D. D. Variability in soils an its significance to past and future soil investigations: statistical study of nitrification in soil. **University of California Publications Agronomy Science**, California, v. 3, n. 9, p. 243-270, Nov. 1918.

YANDELL, B. S. **Practical data analysis for designed experiments**. Madison: University of Wisconsin, 1997. 437 p.

### ANEXO A – Dados utilizados para exemplos

Tabela 27 Conjunto de dados utilizados no exemplo de abordagem Geoestatística.

<i>Observação</i>	<i>Variável Resposta</i>	<i>Tratamento</i>	<i>Bloco</i>	<i>X</i>	<i>Y</i>
1	150,651718246264	1	1	0	0
2	157,845075414125	2	1	1	0
3	150,813339626476	3	1	2	0
4	153,019659327854	4	1	3	0
5	151,153846932120	5	1	4	0
6	152,421590420875	6	1	5	0
7	147,321952876790	7	1	6	0
8	149,375938674545	8	1	7	0
9	148,756257826914	9	1	8	0
10	150,994328818460	10	1	0	1
11	155,927345228249	11	1	1	1
12	155,669953881167	12	1	2	1
13	149,377028409306	13	1	3	1
14	154,583081892265	14	1	4	1
15	149,981744306475	15	1	5	1
16	150,125250036556	16	1	6	1
17	155,293014098574	17	1	7	1
18	150,452386515782	18	1	8	1
19	153,415357187250	1	2	0	2
20	160,837782791488	2	2	1	2
21	152,924523323838	3	2	2	2
22	155,293558692994	4	2	3	2
23	153,204822733013	5	2	4	2
24	155,364521258396	6	2	5	2
25	150,188002395061	7	2	6	2
26	153,531470979913	8	2	7	2
27	151,673975292373	9	2	8	2
28	155,172648708707	10	2	0	3
29	159,012656054645	11	2	1	3
30	157,697630847368	12	2	2	3
31	152,046848446654	13	2	3	3
32	157,841616557526	14	2	4	3
33	154,017386792786	15	2	5	3

Tabela 27, continua

34	154,029112916157	16	2	6	3
35	157,651436511895	17	2	7	3
36	154,635813595538	18	2	8	3
37	155,316942463886	1	3	0	4
38	160,781866055587	2	3	1	4
39	152,302923665430	3	3	2	4
40	156,080362554492	4	3	3	4
41	154,281102289959	5	3	4	4
42	156,793078428489	6	3	5	4
43	151,974624086266	7	3	6	4
44	153,336532192763	8	3	7	4
45	152,665125394684	9	3	8	4
46	155,927556704818	10	3	0	5
47	158,892694063945	11	3	1	5
48	159,992255367574	12	3	2	5
49	152,135064342321	13	3	3	5
50	158,917562790506	14	3	4	5
51	154,819075319569	15	3	5	5
52	154,334863315566	16	3	6	5
53	157,082571863851	17	3	7	5
54	154,314019227931	18	3	8	5
55	153,483002035481	1	4	0	6
56	159,675917322092	2	4	1	6
57	152,054262114768	3	4	2	6
58	152,672029540863	4	4	3	6
59	152,423049622778	5	4	4	6
60	154,712711823826	6	4	5	6
61	148,275428839200	7	4	6	6
62	151,520309657024	8	4	7	6
63	149,562912481045	9	4	8	6
64	153,495293601589	10	4	0	7
65	156,997472934280	11	4	1	7
66	156,627251129416	12	4	2	7
67	150,234687980755	13	4	3	7
68	156,092026380867	14	4	4	7
69	151,501263646153	15	4	5	7
70	152,136735979974	16	4	6	7
71	155,552869662763	17	4	7	7
72	150,389069609204	18	4	8	7
73	153,110111386989	1	5	0	8

Tabela 27, conclusão.

74	159,362559788266	2	5	1	8
75	151,668049785135	3	5	2	8
76	154,055434629420	4	5	3	8
77	152,633467163891	5	5	4	8
78	154,833352839189	6	5	5	8
79	149,202567963763	7	5	6	8
80	150,871352250772	8	5	7	8
81	149,060965707106	9	5	8	8
82	155,254334544092	10	5	0	9
83	158,122237063899	11	5	1	9
84	157,202143323700	12	5	2	9
85	150,868870705211	13	5	3	9
86	157,220737419379	14	5	4	9
87	153,857872877892	15	5	5	9
88	153,378692077618	16	5	6	9
89	155,930893443133	17	5	7	9
90	150,524665160679	18	5	8	9
91	155,731750781010	1	6	0	10
92	161,303233723153	2	6	1	10
93	153,835357489732	3	6	2	10
94	156,635914426920	4	6	3	10
95	155,310566804044	5	6	4	10
96	157,678474478469	6	6	5	10
97	152,762714780094	7	6	6	10
98	152,869431915234	8	6	7	10
99	149,801974750551	9	6	8	10
100	156,208982144513	10	6	0	11
101	159,520801651024	11	6	1	11
102	160,623767850982	12	6	2	11
103	154,082403034213	13	6	3	11
104	160,238397572068	14	6	4	11
105	155,331553483485	15	6	5	11
106	154,223429984755	16	6	6	11
107	159,442669365203	17	6	7	11
108	153,583075526307	18	6	8	11

Tabela 28 Conjunto de dados utilizados no exemplo da abordagem autorregressiva.

<i>Observação</i>	<i>Variável Resposta</i>	<i>Tratamento</i>	<i>Bloco</i>	<i>X</i>	<i>Y</i>
1	151,525337445104	1	1	0	0
2	159,763799014337	2	1	1	0
3	152,338936307044	3	1	2	0
4	158,186683073137	4	1	3	0
5	155,536942386609	5	1	4	0
6	156,107336439772	6	1	5	0
7	149,502058860602	7	1	6	0
8	154,900894052852	8	1	7	0
9	153,924581611179	9	1	8	0
10	152,192289031264	10	1	0	1
11	156,461707603784	11	1	1	1
12	160,277751807983	12	1	2	1
13	150,915774126714	13	1	3	1
14	157,457799447474	14	1	4	1
15	153,979725965802	15	1	5	1
16	152,611911508279	16	1	6	1
17	156,647583938890	17	1	7	1
18	151,953798099551	18	1	8	1
19	153,864121890160	1	2	0	2
20	163,003765175493	2	2	1	2
21	156,368954363996	3	2	2	2
22	157,587768504289	4	2	3	2
23	154,973005925329	5	2	4	2
24	156,549912513333	6	2	5	2
25	152,167755903934	7	2	6	2
26	153,757821357473	8	2	7	2
27	155,112750424662	9	2	8	2
28	152,620897141399	10	2	0	3
29	158,489733006389	11	2	1	3
30	160,696501007099	12	2	2	3
31	152,220870882417	13	2	3	3
32	159,116656610729	14	2	4	3
33	154,841068277038	15	2	5	3
34	156,216514706755	16	2	6	3
35	160,949072639899	17	2	7	3
36	156,259973842633	18	2	8	3
37	151,718009756284	1	3	0	4

Tabela 28, continua

38	161,758497647935	2	3	1	4
39	154,248117359234	3	3	2	4
40	157,473232112555	4	3	3	4
41	151,152027102773	5	3	4	4
42	156,485719593551	6	3	5	4
43	151,554071864075	7	3	6	4
44	151,996920014755	8	3	7	4
45	152,118663537096	9	3	8	4
46	150,987718223699	10	3	0	5
47	160,085207308261	11	3	1	5
48	160,923944758307	12	3	2	5
49	155,692855950421	13	3	3	5
50	155,225718709957	14	3	4	5
51	147,508647687038	15	3	5	5
52	151,353264296928	16	3	6	5
53	155,868495530081	17	3	7	5
54	153,484407509933	18	3	8	5
55	151,170375453382	1	4	0	6
56	158,279348941547	2	4	1	6
57	152,792749023930	3	4	2	6
58	156,526531577343	4	4	3	6
59	148,810218004905	5	4	4	6
60	147,832273656438	6	4	5	6
61	144,114146270153	7	4	6	6
62	148,792584512183	8	4	7	6
63	148,390938709571	9	4	8	6
64	149,091417754204	10	4	0	7
65	156,917193416599	11	4	1	7
66	155,642885221829	12	4	2	7
67	151,145189835187	13	4	3	7
68	156,126818797696	14	4	4	7
69	151,134557781054	15	4	5	7
70	148,495004889435	16	4	6	7
71	156,307093799530	17	4	7	7
72	150,314970927940	18	4	8	7
73	151,398198885536	1	5	0	8
74	156,876080622312	2	5	1	8
75	151,009714600261	3	5	2	8
76	156,312062443978	4	5	3	8

Tabela 28, conclusão.

77	151,414053683147	5	5	4	8
78	153,867179581576	6	5	5	8
79	146,770224791986	7	5	6	8
80	152,191460230212	8	5	7	8
81	152,112073591827	9	5	8	8
82	151,025252826715	10	5	0	9
83	154,869095563756	11	5	1	9
84	154,556986311729	12	5	2	9
85	149,118707604496	13	5	3	9
86	156,094400148250	14	5	4	9
87	151,269801800266	15	5	5	9
88	150,169575431382	16	5	6	9
89	156,644545493500	17	5	7	9
90	156,474108499910	18	5	8	9
91	147,485446270982	1	6	0	10
92	160,446492644220	2	6	1	10
93	151,607325066436	3	6	2	10
94	153,717885821424	4	6	3	10
95	152,741333869814	5	6	4	10
96	155,945993494939	6	6	5	10
97	154,240659473952	7	6	6	10
98	155,963316214814	8	6	7	10
99	159,494712243845	9	6	8	10
100	148,689682666205	10	6	0	11
101	155,708915167183	11	6	1	11
102	155,542424765754	12	6	2	11
103	149,410994641192	13	6	3	11
104	156,871768903210	14	6	4	11
105	153,766549276388	15	6	5	11
106	157,923353328915	16	6	6	11
107	163,739538155492	17	6	7	11
108	159,928213455188	18	6	8	11

## ANEXO B – Script da simulação

### SIMULAÇÃO PARA O EXPERIMENTO i

```
#eftrat<-sample(seq(32,40), 18, replace = T)
eftrat<-c(33,40,33,36,35,37,32,34,33,34,38,39,33,39,35,35,39,35)

#efbloco<-sample(seq(15,20), 6, replace = T)
efbloco<-c(16,19,20,18,19,20)

n<-1000
x1<-data.frame(rho=rep(0,n),rhoest=rep(0,n),
SQE=rep(0,n),SQE_AR=rep(0,n),QMEpor=rep(0,n),
SQT=rep(0,n),SQT_AR=rep(0,n),SQTpor=rep(0,n),SQB=rep(0,n),SQB_AR=rep(0,n),
SQBpor=rep(0,n))

x2<-data.frame(rho=rep(0,n),rhoest=rep(0,n),
SQE=rep(0,n),SQE_AR=rep(0,n),QMEpor=rep(0,n),
SQT=rep(0,n),SQT_AR=rep(0,n),SQTpor=rep(0,n),SQB=rep(0,n),SQB_AR=rep(0,n),
SQBpor=rep(0,n))

x3<-data.frame(rho=rep(0,n),rhoest=rep(0,n),
SQE=rep(0,n),SQE_AR=rep(0,n),QMEpor=rep(0,n),
SQT=rep(0,n),SQT_AR=rep(0,n),SQTpor=rep(0,n),SQB=rep(0,n),SQB_AR=rep(0,n),
SQBpor=rep(0,n))

y1<-data.frame(AIC_Mod1=rep(0,n),AIC_Mod2=rep(0,n))

for (k in 1:n)
{

### SIMULANDO DADOS COM ERRO DEPENDENTE

dados<-data.frame(y=100, trat=rep(1:18), bloco=rep(1:6,each=18))

  for (i in 1:18)
  for (j in 1:6)
  dados$y[dados$trat==i & dados$bloco==j]<-
  dados$y[dados$trat==i & dados$bloco==j] + eftrat[i] + efbloco[j]

  ### Erro espacial dependente

  library(geoR)
```

```

dep<-grf(108, grid="reg",ny=12, nx=9,xlims = c(0,8), ylims = c(0, 11),nugget =
0, cov.pars=c(1,6))
# configuração Gaus(0-1-6)

dep_pad<-((dep$data -mean(dep$data))/var(dep$data))

dados$y<-dados$y+dep_pad

### Abordagem auto regressiva
library(MASS)
library(spdep)

coords<-dep$coords

#####
#####          ANOVA_AR1          #####
#####

nb_dados<-dnearneigh(coords,0,1)
# raio de tamanho 1 unidade de medida

w<-nb2mat(nb_dados, style="W",zero.policy=F)

#plot(nb_dados, coords)

listw<-nb2listw(nb_dados, glist=NULL, style="W",zero.policy=FALSE)

### Modelo SAR

attach(dados)
SAR<-lagsarlm(y~trat+bloco, dados, listw,
method="eigen",quiet=T,tol.solve=1e-15)
ajuste<-summary(SAR)

rho<-ajuste["rho"]
rho<-rho[[1]][[1]]

### Ajustando os dados

rho
beta<-mean(dados[,1])

Y_ajus<-dados[,1]-(rho*w%*%dados[,1]-rho*beta)

```

```

d.ajus<-dados
d.ajus$y<-Y_ajus
### Comparação entre ANOVA e ANOVA_AR

attach(dados)
reg_dep<-lm(y~factor(trat)+factor(bloco), dados)

attach(d.ajus)
reg<-lm(y~factor(trat)+factor(bloco), d.ajus)

aovd<-anova(reg_dep); aovd

aov<-anova(reg); aov

rhoest<-sum(aovd[[2]])-sum(aov[[2]]);rhoest

QMEpor<-(aov[3,2]/aovd[3,2])*100
SQTpor<-(aov[1,2]/aovd[1,2])*100
SQBpor<-(aov[2,2]/aovd[2,2])*100

x1[k,]<-c(rho, rhoest,aovd[3,2],aov[3,2],QMEpor,aovd[1,2], aov[1,2],SQTpor,
aovd[2,2],aov[2,2], SQBpor)

### ANÁLISE GEOESTATÍSTICA

library(nlme)

# anova padrão : Erros independentes
Mod1<- gls(y~factor(trat)+factor(bloco), dados)
Mod1
anova(Mod1)

resMod1 <- as.geodata(data.frame(coords,Mod1$residuals))
summres<-summary(resMod1)
x_max<-summres$distances.summary[[2]]

#plot(vario.res <- variog(resMod1,max.dist=x_max))
#ols1 <- variofit(vario.res,cov.model="gaus",fix.nug=F)
#lines(ols1, col="red", lty=1, lwd=2)

# Modelo com estrutura gaussiana

Mod2 <- update(Mod1, corr = corGaus(form = ~ coords, nugget = F))
anova(Mod2)

y1[k,]<-c(AIC(Mod1), AIC(Mod2))

```

```
#####
#####          ANOVA_AR2          #####
#####

nb_dados<-dnearneigh(coords,0,sqrt(2))
# raio de tamanho raiz de 2 unidades de medida

nb_dados

class(nb_dados)

w<-nb2mat(nb_dados, style="W",zero.policy=F)
w

#plot(nb_dados, coords)

listw<-nb2listw(nb_dados, glist=NULL, style="W",zero.policy=FALSE)
listw

### Modelo SAR

attach(dados)
SAR<-lagsarlm(y~trat+bloco, dados, listw,
method="eigen",quiet=T,tol.solve=1e-15)
ajuste<-summary(SAR)

rho<-ajuste["rho"]
rho<-rho[[1]][[1]]

### Ajustando os dados

rho
beta<-mean(dados[,1])

Y_ajus<-dados[,1]-(rho*w%*%dados[,1]-rho*beta)

d.ajus<-dados
d.ajus$y<-Y_ajus

### Comparação entre ANOVA e ANOVA_AR

attach(dados)
reg_dep<-lm(y~factor(trat)+factor(bloco), dados)
```

```

attach(d.ajus)
reg<-lm(y~factor(trat)+factor(bloco), d.ajus)

aovd<-anova(reg_dep); aovd

ov<-anova(reg); aov

rhoest<-sum(aovd[[2]])-sum(aov[[2]);rhoest

QMEpor<-(aov[3,2]/aovd[3,2])*100
SQTpor<-(aov[1,2]/aovd[1,2])*100
SQBpor<-(aov[2,2]/aovd[2,2])*100

x2[k,]<-c(rho, rhoest,aovd[3,2],aov[3,2],QMEpor,aovd[1,2], aov[1,2],SQTpor,
aovd[2,2],aov[2,2], SQBpor)

#####
#####          ANOVA_AR3          #####
#####

nb_dados<-dnearneigh(coords,0,2)
#raio de tamanho 2 unidades de medida

w<-as.fractions(nb2mat(nb_dados, style="W",zero.policy=F))

#plot(nb_dados, coords)

listw<-nb2listw(nb_dados, glist=NULL, style="W",zero.policy=FALSE)
listw

### Modelo SAR

attach(dados)
SAR<-lagsarlm(y~trat+bloco, dados, listw,
method="eigen",quiet=T,tol.solve=1e-15)
ajuste<-summary(SAR)

rho<-ajuste["rho"]
rho<-rho[[1]][[1]]

### Ajustando os dados

rho
beta<-mean(dados[,1])

```

```

Y_ajus<-dados[,1]-(rho*w%*%dados[,1]-rho*beta)
Y_ajus

d.ajus<-dados
d.ajus$y<-Y_ajus

### Comparao entre ANOVA e ANOVA_AR

attach(dados)
reg_dep<-lm(y~factor(trat)+factor(bloco), dados)

attach(d.ajus)
reg<-lm(y~factor(trat)+factor(bloco), d.ajus)

aovd<-anova(reg_dep); aovd

aov<-anova(reg); aov

rhoest<-sum(aovd[[2]])-sum(aov[[2]]);rhoest

QMEpor<-(aov[3,2]/aovd[3,2])*100
SQTpor<-(aov[1,2]/aovd[1,2])*100
SQBpor<-(aov[2,2]/aovd[2,2])*100

x3[k,]<-c(rho, rhoest,aovd[3,2],aov[3,2],QMEpor,aovd[1,2], aov[1,2],SQTpor,
aovd[2,2],aov[2,2], SQBpor)

}

x1
x2
x3
y1

```

## **SIMULAÇÃO PARA O EXPERIMENTO ii**

Nesta configuração de experimento, basta substituir a parte inicial da simulação para o experimento i, pela apresentada abaixo.

```

#eftrat<-sample(seq(32,40), 5, replace = T)
#para gerar novos efeitos fixos
eftrat<-c(33,37,35,34,38)

#efbloco<-sample(seq(15,20), 4, replace = T)

```

```

#para gerar novos efeitos fixos
efbloco<-c(17,15,18,20)

n<-1
# número de simulações

x1<-data.frame(rho=rep(0,n),rhoest=rep(0,n),
QME=rep(0,n),QME_AR=rep(0,n),QMEpor=rep(0,n),
SQT=rep(0,n),SQT_AR=rep(0,n),SQTpor=rep(0,n),SQB=rep(0,n),SQB_AR=rep(0,n),
SQBpor=rep(0,n))

x2<-data.frame(rho=rep(0,n),rhoest=rep(0,n),
QME=rep(0,n),QME_AR=rep(0,n),QMEpor=rep(0,n),
SQT=rep(0,n),SQT_AR=rep(0,n),SQTpor=rep(0,n),SQB=rep(0,n),SQB_AR=rep(0,n),
SQBpor=rep(0,n))

x3<-data.frame(rho=rep(0,n),rhoest=rep(0,n),
QME=rep(0,n),QME_AR=rep(0,n),QMEpor=rep(0,n),
SQT=rep(0,n),SQT_AR=rep(0,n),SQTpor=rep(0,n),SQB=rep(0,n),SQB_AR=rep(0,n),
SQBpor=rep(0,n))

y1<-data.frame(AIC_Mod1=rep(0,n),AIC_Mod2=rep(0,n))

for (k in 1:n)
{
### SIMULANDO DADOS COM ERRO DEPENDENTE
dados<-data.frame(y=100, trat=rep(1:5), bloco=rep(1:4,each=5))
dados

  for (i in 1:5)
  for (j in 1:4)
  dados$y[dados$trat==i & dados$bloco==j]<-
  dados$y[dados$trat==i & dados$bloco==j] + efrat[i] + efbloco[j]

  ### Erro espacial dependente

  dep<-grf(20, grid="reg",ny=4, nx=5,xlims = c(0,4), ylims = c(0, 3),nugget = 0,
cov.pars=c(1,2))
  ## Configuração Gaus(0-1-2)

  dep_pad<-((dep$data -mean(dep$data))/var(dep$data)

  mean(dep_pad)
  var(dep_pad)

  dados$y<-dados$y+dep_pad

```