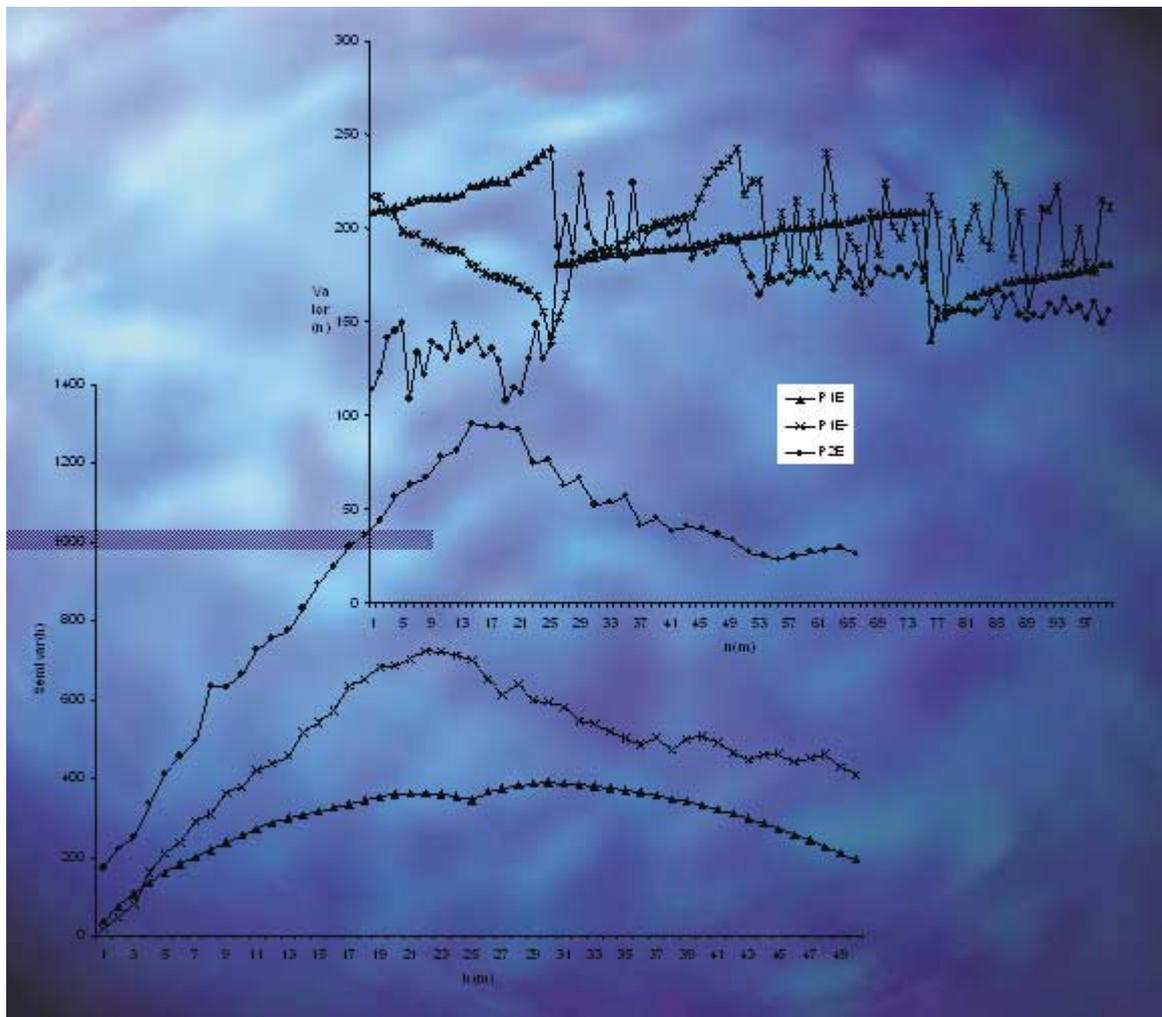


Aplicação da Geoestatística para Seleção de Unidades de Amostra Independentes



República Federativa do Brasil

Luís Inácio Lula da Silva

Presidente

Ministério da Agricultura, Pecuária e Abastecimento

Roberto Rodrigues

Ministro

Empresa Brasileira de Pesquisa Agropecuária - Embrapa

Diretoria Executiva da Embrapa

Clayton Campanhola

Diretor-Presidente

Gustavo Kauark Chianca

Herbert Cavalcante de Lima

Mariza Marilena T. Luz Barbosa

Diretores-Executivos

Embrapa Solos

Celso Vaiver Manzatto

Chefe Geral

David Dias Moreira Filho

Chefe Adjunto de Administração

Alúcio Granato de Andrade

Chefe Adjunto de Pesquisa e Desenvolvimento



*Empresa Brasileira de Pesquisa Agropecuária
Centro Nacional de Pesquisa de Solos
Ministério da Agricultura, Pecuária e Abastecimento*

ISSN 1517-2627

Dezembro, 2004

Documentos 59

Aplicação da Geoestatística para Seleção de Unidades de Amostra Independentes

Thomaz Corrêa e Castro da Costa
Elpídio Inácio Fernandes Filho

Rio de Janeiro, RJ
2004

Exemplares desta publicação podem ser adquiridos na:

Embrapa Solos

Rua Jardim Botânico, 1024 Jardim Botânico. Rio de Janeiro, RJ

Fone: (21) 2274.4999

Fax: (21) 2274.5291

Home page: www.cnps.embrapa.br

E-mail (sac): sac@cnps.embrapa.br

Supervisor editorial: *Jacqueline Silva Rezende Mattos*

Normalização bibliográfica: *Cláudia Regina Delaia*

Revisão de Português: *André Luiz da Silva Lopes*

Editoração eletrônica: *Saulo Stefano*

1ª edição

Todos os direitos reservados.

A reprodução não-autorizada desta publicação, no todo ou em parte, constitui violação dos direitos autorais (Lei nº 9.610).

Costa, Thomaz Corrêa e Castro da.

Aplicação da geoestatística para seleção de unidades de amostra independentes / Thomaz Corrêa e Castro da Costa e Elpídio Inácio Fernandes Filho. - Rio de Janeiro: Embrapa Solos, 2004.

34 p. -(Embrapa Solos. Documentos; n. 59)

ISSN 1517-2627

1.Geoestatística. 2.Geoestatística-Aplicação. I. Fernandes Filho, Elpídio Inácio. II. Embrapa Solos (Rio de Janeiro). III. Título. IV. Série.

CDD (21.ed.) 630.2515

© Embrapa 2004

Autores

Thomaz Corrêa e Castro da Costa

Pesquisador, Ds. Ciência Florestal, Embrapa Solos: R.
Jardim Botânico, 1.024 CEP 22.460-000, Rio de
Janeiro, RJ, thomaz@cnps.embrapa.br

Elpídio Inácio Fernandes Filho

Professor, Ds. Ciência do Solo, Universidade Federal de
Viçosa: Avenida P. H. Rolfs s/n - Campus UFV CEP
36.570-000 - Viçosa - MG, elpidio@ufv.br

Sumário

Introdução, 7

A Teoria de Variáveis Regionalizadas (Teoria da Geoestatística), 9
Pressuposições de Estacionaridade e Ergodicidade, 10

Metodologia, 15

Determinação da Influência de Unidades de Amostra em
Populações de Natureza Conhecida, 15

Resultados, 19

Considerações Finais, 25

Referências Bibliográficas, 25

Anexo - Amostras das populações, espaçadas de 1 metro, 27

Aplicação da Geoestatística para Seleção de Unidades de Amostra Independentes

Thomaz Corrêa e Castro da Costa
Elpídio Inácio Fernandes Filho

Introdução

A Geoestatística, também chamada de “Teoria de Variáveis Regionalizadas”, foi desenvolvida por Matheron (1963, 1971) citado por Vieira *et al.* (1983). Esta teoria foi originalmente aplicada na Geologia com o objetivo de exploração mineral, permitindo o estudo da variação dos dados, não considerados somente aleatórios como na estatística de Fischer, mas também com alguma relação de dependência espacial (Hamakawa, 1991).

A estatística clássica ou de Fischer, teoria mais utilizada de análise de dados na área agrária, reduz o conjunto de dados a dois parâmetros: média e dispersão, considerando a variação entre os dados, embora não se proponha a estudar a estrutura dessa variação. Esta teoria resolveu os maiores problemas de variações de solo em experimentos de campo, por meio da casualização e formação de blocos, destinando ao resíduo, as variáveis ambientais não controladas (Prevedello, 1987).

Por outro lado, o resíduo pode não ser totalmente aleatório, mas apresentar um componente de autocorrelação espacial, ou seja, valores vizinhos (duas unidades amostrais, por exemplo) podem estar correlacionados. Como uma análise de variância pressupõe a independência entre parcelas (Vieira *et al.*, 1983), e esta pressuposição é geralmente aceita quando os dados seguem a distribuição normal, se a autocorrelação estiver presente, pode afetar os resultados da análise e, conseqüentemente, provocar conclusões não confiáveis.

Considerando um exemplo de autocorrelação espacial, podemos considerar o teor natural de cálcio (Ca) no solo como uma variável mensurada em diversos

pontos. Ver-se-á que esta apresenta variação devido às alterações ambientais não controladas que, se não existissem, resultariam em mensurações iguais de teor de Ca. Esta variação pode apresentar uma dependência espacial entre valores vizinhos. O teor de Ca, no caso, pode diminuir à medida que nos afastamos do material de origem calcário, até que a distância não permita mais a influência da rocha calcária. Podemos, então, medir o raio de influência de uma parcela por meio de uma medida de correlação entre valores vizinhos, afim de evitar a marcação de parcelas que sejam dependentes entre si. Esta é uma das finalidades da teoria da Geoestatística (McBratney & Webster, 1983).

Com relação à forma de distribuição destas parcelas ou unidades de amostra (desenho de amostragem), a amostra seletiva é a de menor representatividade, permitindo tendência nas estimativas de parâmetros e com maior risco de seleção de unidades de amostra dependentes. Já a amostra sistemática, para populações que apresentem periodicidade nos dados, pode não representar totalmente a variação populacional, sujeita à correlação espacial, provocando tendência nas estimativas dos parâmetros, neste caso (Stehman, 1992). Assim, o uso do desenho sistemático exige que se conheça a variação espacial da cena para definir a distância ideal de pontos na grade sistemática, afim de não selecionar pontos correlacionados. Desta forma, demanda-se um estudo de autocorrelação da cena quando for adotado o desenho sistemático.

A amostragem aleatória fornece estimativas consistentes, sendo a que oferece menor risco de correlação espacial. De qualquer forma, a intensidade amostral (número de unidades de amostra selecionadas em relação a toda a população) pode ser melhor definida, conhecendo-se a distância média entre unidades de amostra não correlacionadas, possível pela aplicação da geoestatística.

Em estudos de amostragem, McBratney & Webster (1983) demonstraram um método para determinar o tamanho de amostra, considerando a dependência espacial determinada pelo semivariograma. A variância total pode ser minimizada, para um dado tamanho de amostra, se a posição das observações forma uma grade sistemática. Exemplos mostraram que a relação entre erro padrão e tamanho de amostra pode ser alcançada para qualquer precisão desejada.

Outro estudo é apresentado em Entz & Chang (1991), que avaliaram delineamentos de amostragem usando análise geoestatística, semivariograma e "Kriging", um método interpolador que permite a estimação de valores de variáveis, em locais não mensurados, por meio das funções de autocorrelação (Burrough, 1986).

Em mapeamentos do uso/cobertura da terra pela interpretação de imagens de sensoriamento remoto, por exemplo, quando se objetiva medir a exatidão do produto cartográfico gerado, utiliza-se uma amostra de referência (uma rede de pontos com classes conhecidas) representando, assim, a verdade de campo. Como imagens de sensoriamento remoto apresentam forte correlação espacial, pois pixels mais próximos apresentam valores de brilho similares (demarcam uma feição), e à medida que se aumenta a distância entre pixels, o desvio entre valores de brilho tende a aumentar até que a variação não sofra mais a influência da distância (esta variação representa a dependência espacial entre valores vizinhos), pode-se medir o raio de influência de um ponto na imagem, por meio de uma função de correlação espacial entre valores vizinhos, afim de evitar a marcação de pontos que sejam dependentes entre si.

Funções de correlação espacial, como os semivariogramas, veem sendo utilizadas em sensoriamento remoto para identificar padrões espaciais de feições (Curran, 1988; Ramstein & Raffy, 1989; Cohen et al., 1990; Lacaze et al., 1994), para estimar erros provocados por ruídos em imagens de sensoriamento remoto (Atkinson, 1996; Atkinson, 1997), e para classificar imagens de radar (Miranda et al., 1992, 1996), entre outros.

Neste ensaio, será apresentada a teoria básica da Geoestatística que será então aplicada a dados fictícios, para verificar a consistência na seleção de unidades de amostra independentes, seguindo-se a sugestão de Curran (1988). Foram aplicados semivariogramas para determinar o raio de influência, ou alcance, entre unidades de amostra, indicando a distância ideal para seleção de unidades de amostra independentes, exigidas pela teoria de amostragem para definição de intervalos de confiança.

A Teoria de Variáveis Regionalizadas (Teoria da Geoestatística)

Um fenômeno que se modifica no espaço e apresenta alguma estrutura comporta-se como “regionalizado”. Uma variável regionalizada é uma função no espaço que varia de uma posição x para uma posição $x + h$, com uma “aparente” continuidade (Vieira *et al.*, 1983). É aparente porque pode ocorrer descontinuidade, como mudança abrupta na média entre dois subconjuntos de dados. Isto significa que a média pode não ser homogênea (ou constante) à medida em que aumentamos infinitamente, o tamanho da área amostrada.

Sendo $Z(x)$ o valor de uma variável regionalizada, Z , no ponto x , é possível descrever a variabilidade da função $Z(x)$ no espaço x (coordenadas) variando dependentemente do local da amostragem (Grossi sad, 1986, citado por Hamakawa, 1991). Seu comportamento pode ainda variar com a direção, ou seja, refletir a anisotropia do fenômeno. Observando a equação:

$$Z(x) = m(x) + V(x)_1 + V(x)_2 \quad (4)$$

sendo $m(x)$ a média, uma função determinística que corresponde ao componente estrutural de Z e $V(x)$ a sua dispersão. Pode-se dizer que a dispersão apresenta dois componentes: $V(x)_1$, aleatório, espacialmente dependente de $m(x)$ e $V(x)_2$, aleatório, espacialmente independente, ou termo de erro residual com média 0 e variância s^2 (Burrough, 1986). Note que, para a estatística clássica, o resíduo é expresso por apenas um componente, $V(x)$.

Este fenômeno de regionalização é associado ao conceito estocástico, considerando a média de uma amostra como uma dentre as possíveis médias amostrais de um processo aleatório, conhecida como média espacial ou média do conjunto (Vieira *et al.*, 1983). Com uma única realização, pode-se conhecer a função aleatória $Z(x)$. Se deseja-se estimar valores para locais não mensurados, pode-se introduzir a restrição de que a variável regionalizada deve ser, estatisticamente, homogênea e isotrópica, que nos permite fazer inferências estatísticas (Olea, 1975 citado por Vieira, 1983). A vantagem da aproximação estocástica é poder estimar outras propriedades estatísticas, além da média, como a variância e a covariância que caracteriza a correlação entre dois pontos contíguos. Pelo fato de termos apenas uma realização de um processo aleatório, algumas pressuposições são necessárias para sua consistência, que são Estacionaridade e Ergodicidade (Vieira *et al.*, 1983).

Pressuposições de Estacionaridade e Ergodicidade

Uma variável regionalizada é estacionária se as estatísticas de primeira, segunda, até grau de maior ordem da variável aleatória $Z(x+h)$ são as mesmas para qualquer vetor de distância h . Para uma variável que é estacionária de segunda ordem, por exemplo, todos os seus momentos até os de segunda ordem (média, variância e covariância) são constantes (Vieira *et al.*, 1983). A função de estacionaridade diz respeito, então, à continuidade da função $Z(x)$. Assume-se que as propriedades estatísticas (média, variância, covariância, momentos de maior ordem...) são estacionárias no espaço, isto é, não variam com a translação.

Ergodicidade implica que uma única realização se comporta, no espaço, com a mesma função densidade de probabilidade que o conjunto de possíveis realizações. Assim, com a observação da variação no espaço de uma propriedade, é possível determinar a função densidade de probabilidade para todas as realizações. Isto é chamado de inferência estatística da função densidade de probabilidade da função regionalizada $Z(x)$. Um fenômeno que é estacionário e ergódico é considerado homogêneo, ou uniforme.

Hipótese de Estacionaridade

Hipótese de 2ª ordem: um fenômeno que se comporta, no espaço, com uma variância finita, ou seja, a partir de um certo raio de ação os valores mensurados fazem parte da distribuição de probabilidade representada pelos dados já medidos, apresenta estacionaridade de segunda ordem ou estacionaridade forte, que ocorre quando os dois primeiros momentos são estacionários. As hipóteses para comprovar este fenômeno são:

- 1) O valor esperado, $E[Z(x)] = E[Z(x_i + h)] = E[Z(x_i + 2h)] \dots = m$, não sendo uma função de x_i ;
- 2) A covariância, $C(h) = E[Z(x_i + h) * Z(x_i)] - m^2$, não sendo uma função de x_i , mas apenas do vetor de distância h (lag); e
- 3) A variância $C(0) = E[Z^2(x)] - m^2$.

m é a média populacional para uma área S onde ocorre o fenômeno e é constante, independente da regionalização do fenômeno. Sendo assim, desenvolvendo a covariância, $C(h)$, temos que:

$$C(h) = E\{[Z(x_i + h) - m(x_i + h)] * [Z(x_i) - m(x_i)]\} = E\{[Z(x_i + h) - m] * [Z(x_i) - m]\}$$

$$C(h) = E[Z(x_i + h) * Z(x_i)] - m * Z(x_i + h) - m * Z(x_i) + m^2$$

$$C(h) = E[Z(x_i + h) * Z(x_i)] - m * E[Z(x_i + h)] - m * E[Z(x_i)] + m^2 = E[Z(x_i + h) * Z(x_i)] - 2m^2 + m^2$$

$$C(h) = E[Z(x_i + h) * Z(x_i)] - m^2, \text{ conforme a hipótese (2).}$$

Fazendo-se $h = 0$, significando a inclusão de todos os dados contínuos, na expressão de $C(h)$, deduz-se a hipótese (3) para a variância. A hipótese (3) expressa a

variância populacional, com todo o conjunto de dados ocorrentes na área. Para verificar isto, deduz-se a fórmula da variância populacional

$$s^2(x_i) = \frac{\sum_{i=1}^n (x_i - m)^2}{n} = \frac{\sum_{i=1}^n (x_i^2 - 2mx_i + m^2)}{n} = \frac{\sum_{i=1}^n x_i^2 - 2m \sum_{i=1}^n x_i + nm^2}{n} = \frac{\sum_{i=1}^n x_i^2}{n} - 2m^2 + m^2 = E[Z(x_i)^2] - m^2$$

Hipótese Intrínseca: existem situações em que ocorre fraca estacionaridade, como em fenômenos físicos com infinita capacidade de dispersão, ou seja, à medida em que se aumenta a área amostrada, a variância não se estabiliza, aumentando proporcionalmente ao tamanho da área, infinitamente. Isto implica que o fenômeno não é estritamente estacionário, sendo necessário considerar a hipótese de fraca estacionaridade, ou hipótese intrínseca. Esta hipótese requer somente a hipótese de existência do semivariograma, sem a exigência de variância finita. Assim, a variância de $Z(x)$ não é finita, mas a variância do primeiro incremento de Z , $[Z(x+h)-Z(x)]$, é finita, e este incremento é fracamente estacionário, verificando que:

- 1) valor esperado para o primeiro incremento: $E[Z(x_i+h)-Z(x_i)] = m(h)$, não é função de x_i ;
- 2) A variância do primeiro incremento: $Var[Z(x_i+h)-Z(x_i)]$ é função de h somente, não de x_i .

Para esta hipótese, $m(h)$ pode ser zero ou não. A média agora é uma função de h , sem a exigência de homogeneidade espacial.

Funções Semivariância e Autocorrelação

Semivariância

A **Semivariância** é a função de autocorrelação mais utilizada para descrever a variabilidade de $Z(x)$, pois atende tanto a hipótese de forte quanto fraca estacionaridade. Semivariância é a média entre variâncias de pares de dados separados por um vetor distância h ou *lag*. Considere um par de valores $[Z(x), Z(x+h)]$, separados pela distância h . A variância S^2 para este par de valores é:

$$S^2 = \frac{[Z(x) - \bar{Z}]^2 + [Z(x+h) - \bar{Z}]^2}{2-1} = [Z(x) - \bar{Z}]^2 + [Z(x+h) - \bar{Z}]^2$$

Sendo $\bar{Z} = \frac{Z(x) + Z(x+h)}{2}$, e substituindo-se em S^2 , temos que:

$$S^2 = \frac{[Z(x) - Z(x+h)]^2}{2} = \text{SemiVar},$$

que é a semivariância para este par de valores. Se considerarmos k pares de valores separados por um mesmo h de n valores, a semivariância média será:

$$\text{SemiVar}(h) = \frac{E[Z(x+h) - Z(x)]^2}{2} = \frac{1}{2k} \sum_{i=1}^k [Z(x_i+h) - Z(x_i)]^2 \quad (5)$$

sendo $k = n-h$.

Se desenvolvermos a fórmula da semivariância e substituirmos as expressões da hipótese de estacionaridade de segunda ordem, conforme segue:

$2 * \text{SemiVar}(h) = E[Z(x_i+h) - Z(x_i)]^2 = E[Z(x_i+h)^2 + Z(x_i)^2 - 2 * Z(x_i+h) * Z(x_i)]$. Balanceando esta expressão com m^2 , temos:

$$\begin{aligned} 2 * \text{SemiVar}(h) &= E\{[Z(x_i+h)^2 - m^2] + [Z(x_i)^2 - m^2] - [2 * Z(x_i+h) * Z(x_i) + m^2]\} \\ &= E[Z(x_i+h)^2] - m^2 + E[Z(x_i)^2] - m^2 - 2E[Z(x_i+h) * Z(x_i)] + m^2. \end{aligned}$$

Substituindo as expressões das hipóteses nesta equação temos:

$$2 * \text{SemiVar}(h) = C(0) + C(0) - 2C(h)$$

Então, $\text{SemiVar}(h) = C(0) - C(h)$ para populações sob estacionaridade de 2ª ordem.

Autocorrelação

Esta função expressa a correlação linear de uma série de dados num espaço, quantificando a mudança de forte dependência para uma situação de independência dos valores de um parâmetro (Davis, 1973; Gajem *et al.*, 1981; Nielsen & Cassel, 1984 citados por Prevedello, 1987). Esta função é expressa pela equação:

$$\text{AutoCor}(h) = C(h)/C(0),$$

Sendo $C(0)$ a variância populacional do parâmetro x .

Segundo Vauclin *et al.* (1982), citado por Dourado Neto (1989), quando a hipótese estacionaridade de segunda ordem é satisfeita, as funções de autocorrelação e semivariância são equivalentes para caracterizar a correlação de uma variável espaçada de h . Quando apenas a hipótese intrínseca é satisfeita, só é válido o uso da função semivariância, isto porque a função covariância necessita da média para estimação e esta não é mais homogênea para todo o conjunto dos dados.

As funções de autocorrelação e semivariância representam dados discretos, gerando estimativas de dados contínuos ou populacionais. Os gráficos gerados por estas funções são chamados **autocorrelogramas** e **semivariogramas**, respectivamente. O Autocorrelograma é padronizado no intervalo $[-1,1]$, sendo o alcance (distância onde ocorre independência entre valores) determinado quando a curva atinge, aproximadamente, o valor zero. Já a curva do semivariograma, ao contrário do autocorrelograma, aumenta à medida que h cresce, atingindo um patamar, quando a semivariância é, aproximadamente, igual à variância da população (Clark, 1979, citado por Prevedello, 1987), embora isto não ocorra para populações que satisfazem apenas a hipótese intrínseca. O início deste patamar define o alcance ou raio de influência.

Para semivariogramas, é esperado que, para $h=0$, $SemiVar(h)$ também igual a zero que, na prática, não acontece. As curvas geralmente não passam pela origem, mas por um valor positivo. Este fenômeno é conhecido como “efeito pepita” ou *nugget*, que é explicado por flutuações no solo que ocorrem à distâncias mais curtas que o primeiro “lag” considerado na análise (Burgess & Webster, 1980 citado por Prevedello, 1987).

Modelos para ajuste dos semivariogramas

Os modelos mais utilizados para ajuste da curva da função de semivariância são (Prevedello, 1987):

Modelo linear : $SemiVar(h) = C_0 + Ch$

$C_0 + C$, para $h > a$

Modelo esférico: $SemiVar(h) = C_0 + C\{[(3/2) * (h/a)] - [(1/2) * (h/a)^3]\}$, para $0 \leq h \leq a$

Modelo exponencial: $SemiVar(h) = Co + C[1 - \exp(-h/a)]$

Modelo de Gauss: $SemiVar(h) = Co + C\{1 - \exp[-(h/a)^2]\}$

Todos estes modelos pressupõem o efeito pepita e a ocorrência de um patamar ("sill"): $h > a$.

Metodologia

Determinação da Influência de Unidades de Amostra em Populações de Natureza Conhecida

Na Tabela em anexo, é apresentado o rol de unidades de amostras de 10 populações, pela geração de números aleatórios, com natureza, distribuição e parâmetros pré-determinados (Tabela 1). Das 10 populações, seis são constituídas de elementos em ordem aleatória (P1, P2, P3, P4, P5 e P6, Figuras 1 e 2), com variabilidades diferentes, medidas pelo coeficiente de variação, e quatro são de natureza estratificada. Para as populações estratificadas, a composição dos estratos variou de população para população.

A população P1E é formada por três estratos, sendo o primeiro com 25 elementos, o segundo com 50 elementos e o terceiro com 25 elementos, todos com tendência linear crescente (Figura 1). Já para a população P1E-, o primeiro estrato com 25 elementos apresenta tendência linear decrescente, sendo o segundo estrato, também com 25 elementos, com tendência linear crescente e o terceiro estrato com 50 elementos em ordem aleatória. A população P2E constituiu-se de quatro estratos, todos com 25 elementos e em ordem aleatória. Finalmente, a população P5E constituiu-se em dois estratos, cada um com 50 elementos em ordem aleatória (Figura 2).

Para aplicação das funções de correlação espacial neste conjunto de dados, deve-se considerar as populações em faixa, pois o exemplo restringe a análise em apenas uma direção e para toda a população; e os dados da primeira realização, distanciados inicialmente em 1 m de lag ($h = 1$).

Os lags aplicados aos dados para o cálculo da semivariância foram: $h = 1, 2, 3, \dots, 49, 50$. Para o cálculo da função de covariância os lags aplicados foram $h = 1, 5, 10, 15, 20, 25, 30, 35, 40, 45$ e 50 m, o que não altera a tendência dos correlogramas. Os pares são formados por distâncias fixas de 1, 5, ..., 50 m, o que

diferencia do procedimento para o cálculo de semivariogramas com abordagem espacial, em que o lag é determinado por intervalo: os pares formados pelo lag 5, por exemplo, serão todos os pares a uma distância entre [1m, 5m].

A metodologia empregada neste estudo seguiu a expressão básica da função semivariância, em que somente os pares a uma distância fixa são considerados. Não foram aplicados *lags* maiores que 50 m porque resultariam em número de pares de dados (*k*) menores que 50, com resultados não confiáveis.

Tabela 1- Características das Populações.

POP	Distrib.	Natureza	<i>n</i>	<i>m</i>	<i>s</i> ²	CV(%)
P1	Normal	Aleatória	100	195,13	426,9685	10,59
P1E	Normal	3 estratos	100	195,13	426,9685	10,59
P1E-	Normal	3 estratos	100	195,13	426,9685	10,59
P2	Normal	Aleatória	100	164,87	663,2187	15,62
P2E	Normal	4 estratos	100	164,87	663,2187	15,62
P3	Normal	Aleatória	150	244,28	2819,3936	21,74
P4	Normal	Aleatória	100	207,43	3683,6717	29,26
P5	Normal	Aleatória	100	205,62	8679,5876	45,31
P5E	Normal	2 estratos	100	205,62	8679,5876	45,31
P6	Normal	Aleatória	100	218,16	22608,6306	68,92

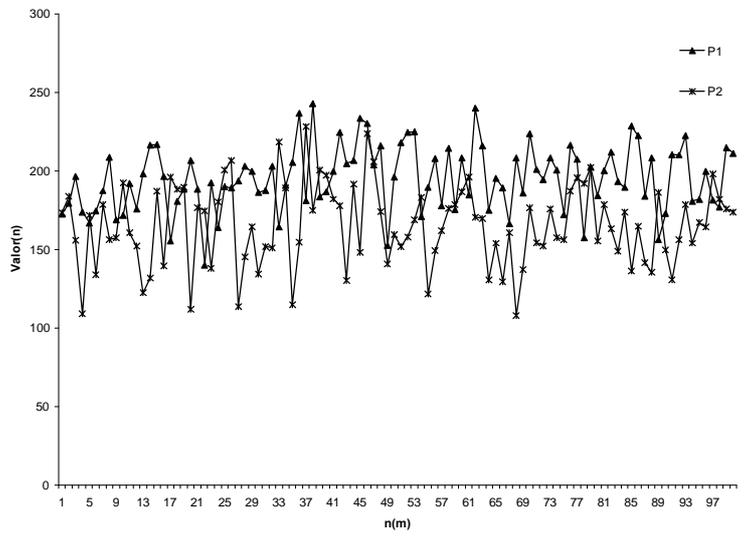
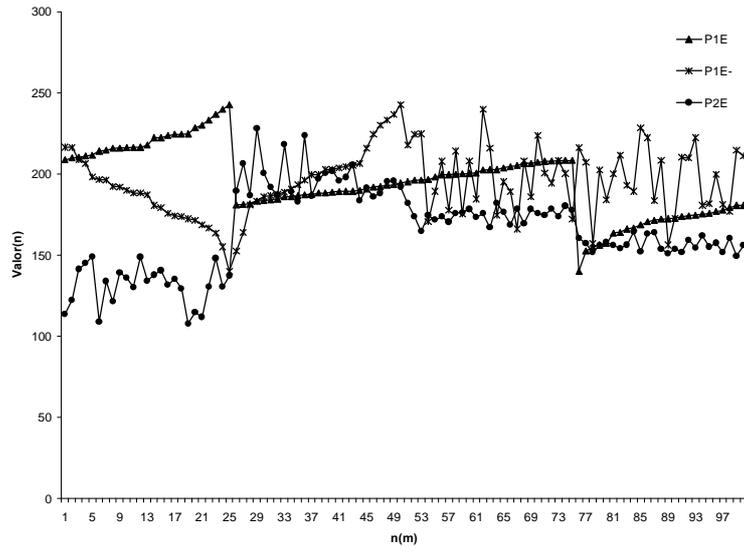


FIGURA 1 -Variação Sequencial das Populações P1E, P1E-, P2E, P1 e P2.

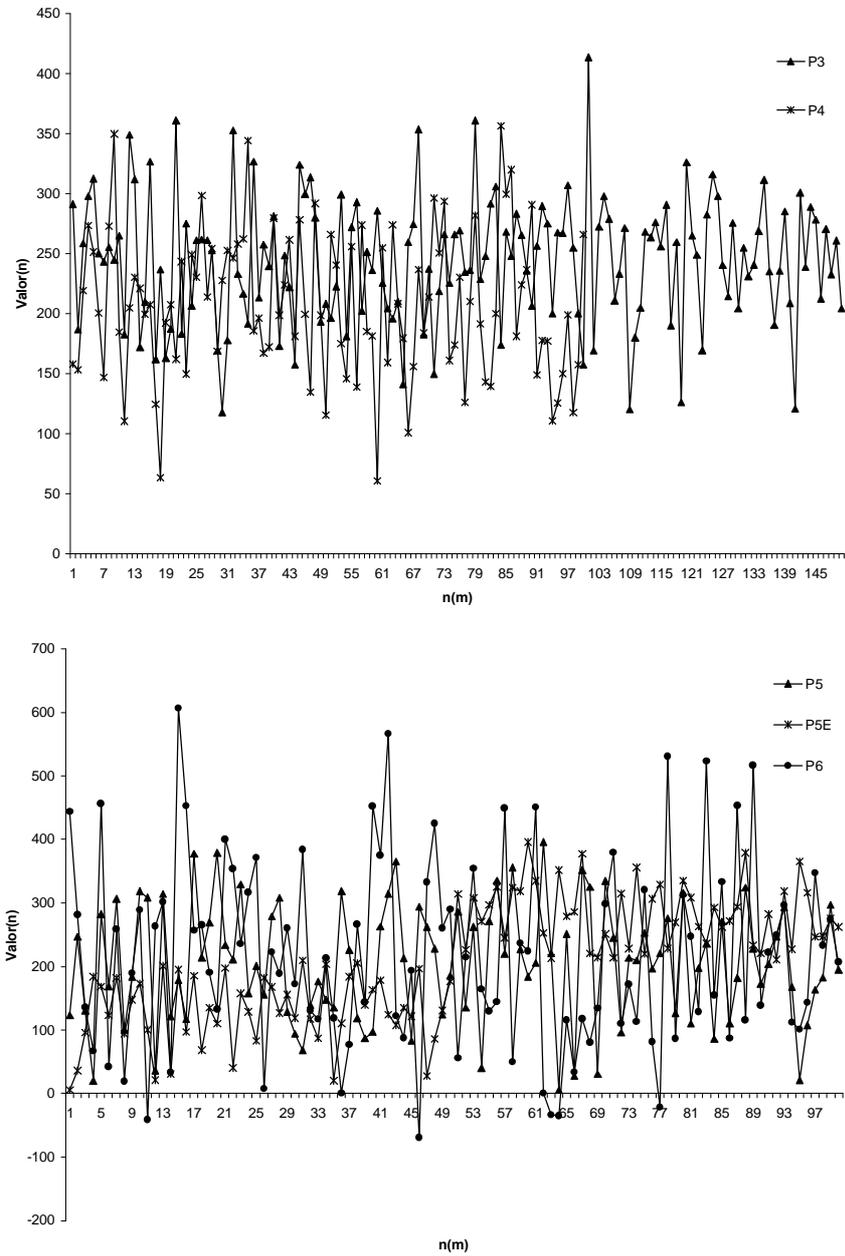


FIGURA 2 –Variação Seqüencial das Populações P3, P4, P5, P5E e P6.

Resultados

Como observado nos semivariogramas (Figuras 3 e 4), a semivariância não é função da distância para as populações aleatórias (P1, P2, P3, P4, P5 e P6), confirmando a forte estacionaridade destas populações. Nesta situação qualquer distância a partir de 1 m pode ser tomada para coleta de amostras independentes. O mesmo pode ser comprovado pelos autocorrelogramas (Figuras 5 e 6), pois o comportamento das curvas é uniforme em torno de zero, independente da distância entre pares de observação.

Para a população P1E, que é formada por três estratos com correlação linear crescente, o primeiro com 25 elementos, o segundo com 50 elementos e o terceiro com 25 elementos, a curva do semivariograma apresenta um ponto de máximo na distância de 25 m, tornando a decair com o aumento do *lag* (obtenção de unidades de amostra correlacionadas reduz a semivariância), pois para esta população, a melhor opção para coleta de amostras independentes é a distância maior que 25 m, obtendo-se uma unidade de amostra no primeiro estrato, duas no segundo estrato (correlacionadas) e uma no terceiro estrato. Observando seu correlograma, o inverso do semivariograma, seu ponto de mínimo ocorre para 25 m, com autocorrelação de aproximadamente -0,28. Pela teoria, esta população estaria sob estacionaridade de segunda ordem, com variância finita, onde a função covariância é simétrica à função semivariância.

A amostra para a população P1E- (primeiro estrato com tendência linear decrescente, o segundo estrato com tendência linear crescente, ambos com 25 elementos, e o terceiro estrato com 50 elementos em ordem aleatória) apresenta semivariograma com estabilidade da semivariância aproximadamente igual à variância populacional, formando o patamar (*sill*). O ponto inicial deste patamar é também o *lag* de 25 m porque, a partir desta distância, consegue-se quatro unidades de amostra independentes nesta população.

O comportamento do autocorrelograma foi análogo ao que aconteceu para a população P1E. Aqui, ocorrem duas evidências para hipótese de estacionaridade de segunda ordem: estabilização da variância igual à variância populacional e simetria do autocorrelograma em relação ao semivariograma.

Para a população P2E (quatro estratos com elementos em ordem aleatória, todos com 25 elementos), como em P1E-, seu semivariograma apresentou também

patamar *sill* a partir de 25 m, distância que torna possível a coleta de quatro unidades de amostra independentes. Da mesma forma, seu autocorrelograma é simétrico ao semivariograma. Note que não ocorre estabilização em autocorrelação zero, mas sim em valores negativos, que para esta população foi de -0,4, isto porque a população é estratificada. Os pares representam valores altos e baixos, seqüencialmente, apresentando um pequeno nível de correlação negativa. Esta população está também sob estacionaridade de 2ª ordem, pelas mesmas razões apresentadas em P1E-.

A população P5E apresenta a seguinte estrutura, pela sua amostra: dois estratos de 50 elementos, ambos em ordem aleatória. Observando seu semivariograma e autocorrelograma poderíamos, equivocadamente, constatar que esta população apresenta variância infinita sendo totalmente correlacionada, crescendo ou diminuindo, indefinidamente, estando sob a hipótese intrínseca. Com o conhecimento da estrutura, podemos afirmar que a partir do *lag* 50, o último medido por razões de precisão, tanto o semivariograma quanto o autocorrelograma irão estabilizar para *lags* maiores que 50 m. Isto porque só serão formados pares com valores de cada estrato aleatório. Estes pares, independente da distância entre os valores, apresentarão, aproximadamente, a mesma variância.

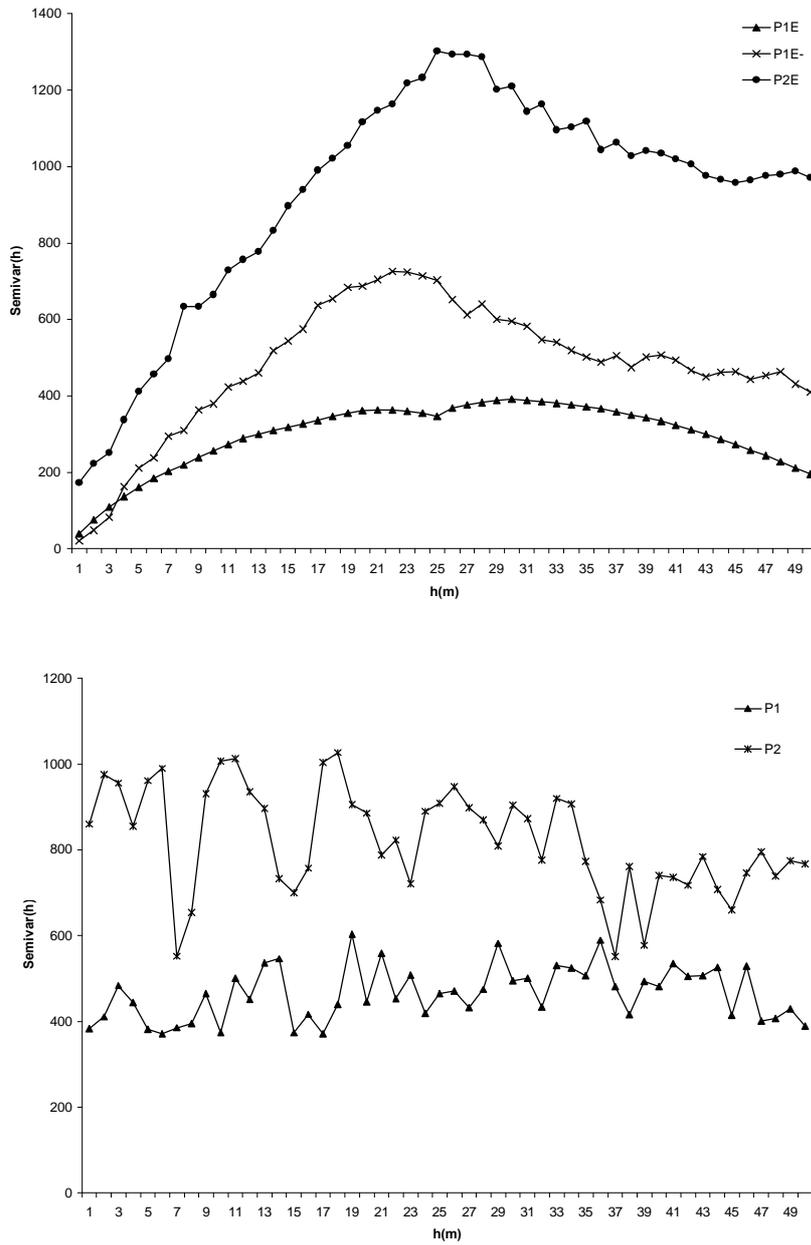


FIGURA 3 - Semivariância para as Populações P1E, P1E-, P2E, P1 e P2.

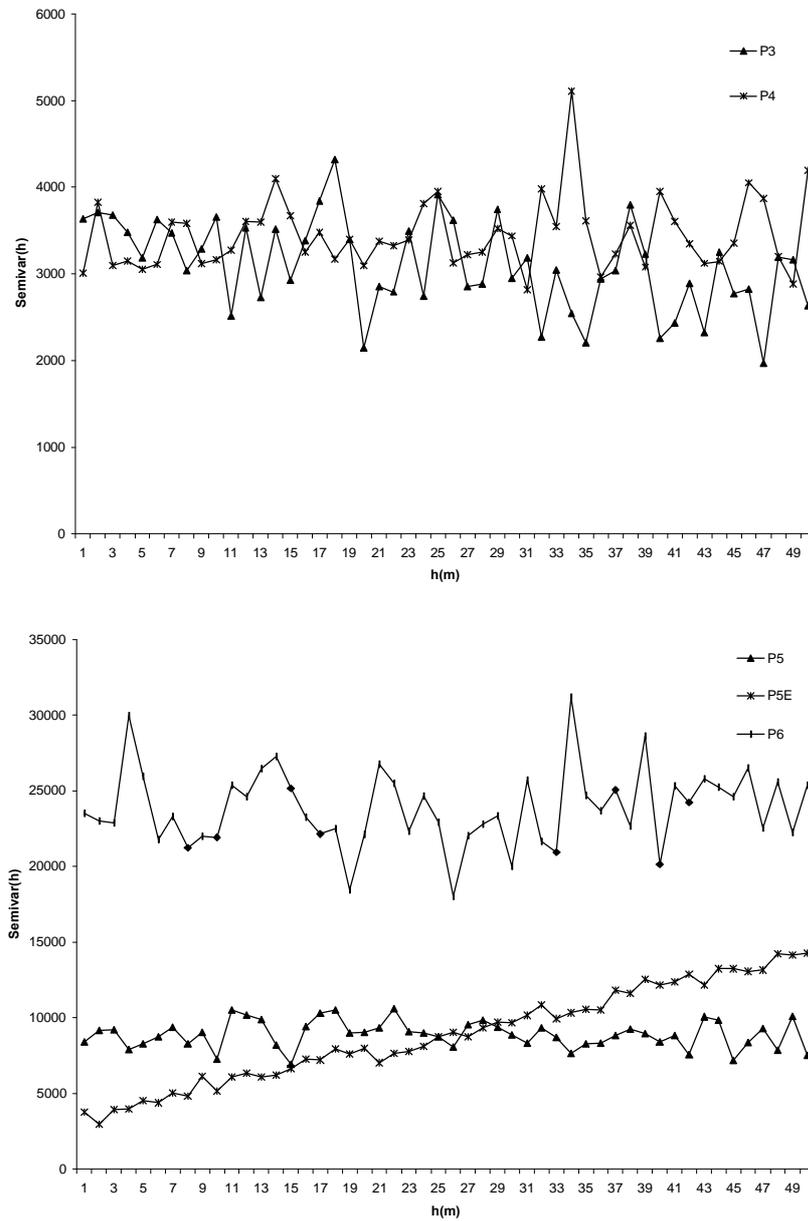


FIGURA 4 - Semivariância para as Populações P3, P4, P5, P5E e P6.

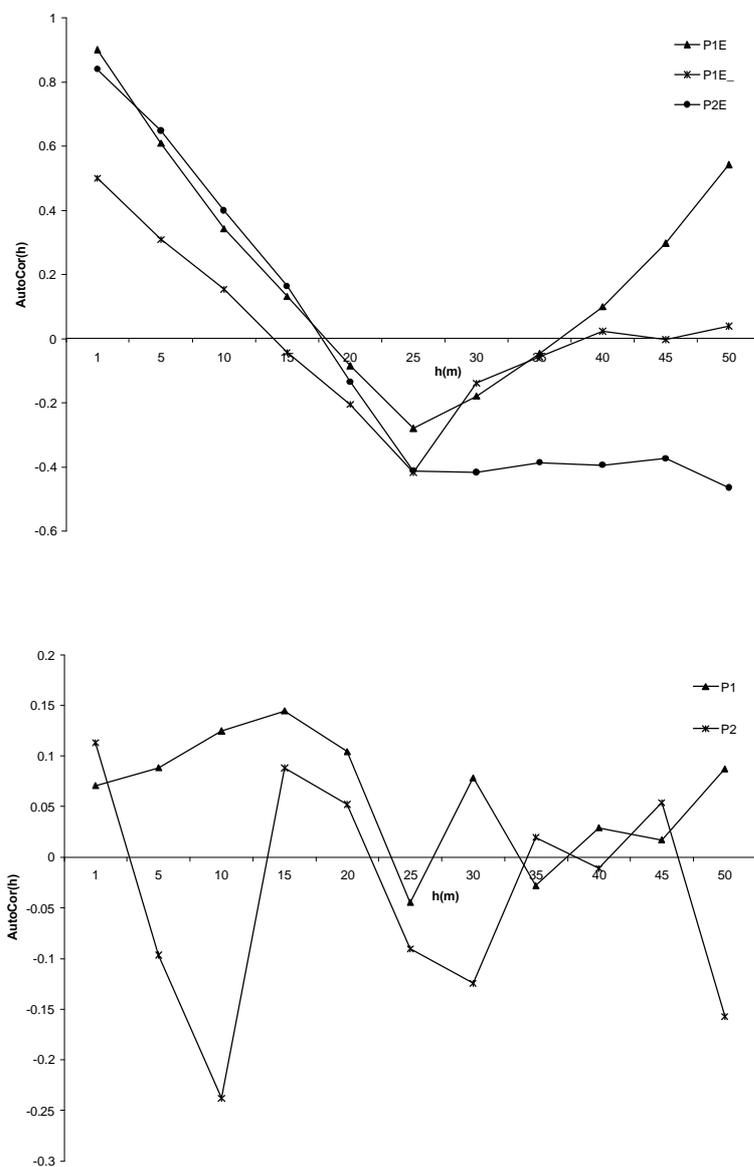


FIGURA 5 - Autocorrelograma para as Populações P1, P1E, P1E-, P2 e P2E.

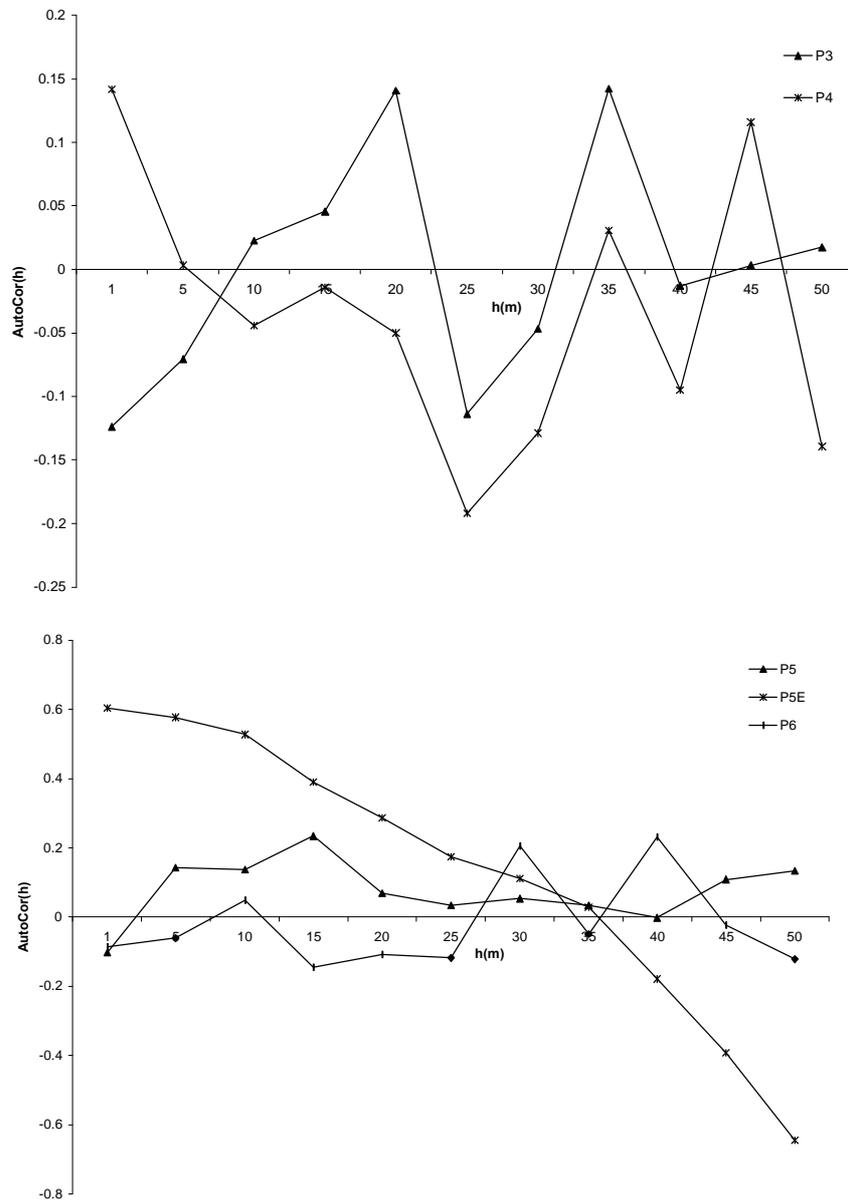


FIGURA 6 - Autocorrelograma para as Populações P3, P4, P5, P5E e P6.

Considerações Finais

A revisão da geoestatística básica, acompanhada de um ensaio com populações fictícias teve como finalidade confrontar as hipóteses e pressuposições descritas pela teoria com dados simulando correlações espaciais pré-determinadas, verificando se os conceitos básicos se reproduzem nas diferentes variações criadas, de forma a permitir maior segurança na aplicação. Pelos resultados, a análise de correlação espacial aplicada para a determinação de unidades de amostra independentes foi exata para os pontos de inflexão esperados, podendo auxiliar na montagem de experimentos e delineamentos de amostragem, afim de evitar resultados sob efeito da violação de independência entre amostras.

Referências Bibliográficas

- ATKINSON, P. M. On estimating measurement error in remotely-sensed images with the variogram. **International Journal of Remote Sensing**, Basingstoke, v. 18, n. 14, p. 3075-3084, 1997.
- ATKINSON, P. M.; DUNN, R.; HARRISON, A . R. Measurement error in reflectance data and its implications for regularizing the variogram. **International Journal of Remote Sensing**, Basingstoke, v. 17, n. 18, p. 3735-3750, 1996.
- BURROUGH, P. A. **Principles of geographic information systems for land resources assessment**. Oxford: Oxford University Press, 1986. 193 p.
- COHEN, W. B.; SPIES, T. A .; BRADSHAW, G. A . Semivariograms of digital imagery for analysis of conifer canopy structure. **Remote Sensing of Environment**, New York, n. 34, p.167-178, 1990.
- CURRAN, P. J. The semivariogram in remote sensing: an introduction. **Remote Sensing of Environment**, New York, v. 24, p.493-507, 1988.
- DOURADO NETO, D. **Variabilidade espacial das alturas de chuva e irrigação e de potenciais da solução do solo**. Piracicaba: ESALQ, 1989. 180 p.
- ENTZ, T.; CHANG, C. Evaluation of soil sampling schemes for geostatistical analysis: a case study for soil bulk density. **Canadian Journal Soil Science**, Ottawa, v. 71, p.165-176, 1991.

HAMAKAWA, P.J. **Variabilidade espacial de alguns componentes de produção de uma cultura de feijoeiro**. Piracicaba: ESALQ, 1991. 114 p.

LACAZE, B.; RAMBAL, S.; WINKEL, T. Identifying spatial patterns of Mediterranean landscapes from geostatistical analysis of remotely-sensed data. **International Journal of Remote Sensing**, Basingstoke, v. 15, n.12, p. 2437-2450, 1994.

McBRATNEY, A. B.; WEBSTER, R. How many observations are needed for regional estimation of soil properties? **Soil Science**, Baltimore, v. 135, n. 3, p. 177-183, 1983.

MIRANDA, F. P.; FONSECA, L. E. N.; CARR, J. R.; TARANIK, J. V. Analysis of JERS-1 (Fuyo-1) SAR data for vegetation discrimination in northwestern Brazil using the semivariogram textural classifier (STC). **International Journal of Remote Sensing**, Basingstoke, v. 17, n. 17, p. 3523-3529, 1996.

MIRANDA, F. P.; MacDONALD, J. A.; CARR, J. R. Application of the semivariogram textural classifier (STC) for vegetation discrimination using SIR-B data of Borneo. **International Journal of Remote Sensing**, Basingstoke, v. 13, n. 12, p. 2349-2354, 1992.

PREVEDELLO, B. M. S. **Variabilidade espacial de parâmetros de solo e planta**. Piracicaba: ESALQ, 1987. 166 p.

RAMSTEIN, G; RAFFY, M. Analysis of the structure of radiometric remotely-sensed images. **International Journal of Remote Sensing**, Basingstoke, v. 10, p.1049-1073, 1989.

STEHMAN, S. V. Comparison of systematic and random sampling for estimating the accuracy of maps generated from remotely sensed data. **Photogrammetric Engineering and Remote Sensing**, Bethesda, v. 58, n. 9, p. 1343-1350, 1992.

VIEIRA, S. R.; HATFIELD, J. L.; NIELSEN, D. R.; BIGGAR, J. W. Geostatistical theory and application to variability of some agronomical properties. **Hilgardia**, Berkeley, v. 51, n. 3, p.1-75, 1983.

ANEXO

Amostras das populações

Tabela - Amostras das populações, espaçadas de 1 metro.

N	P1	P1E	P1E-	P2	P2E	P3	P4	P5	P5E	P6
1	1725711	208.7960	216.6416	173.3969	113.7074	291.1899	157.9616	122.8672	5.3988	444.1395
2	179.2014	209.9938	216.3817	183.5815	122.3352	186.8509	163.4021	247.0856	36.1087	281.3905
3	196.4355	210.3948	208.7960	155.8922	141.3794	258.3719	219.2050	130.7466	96.3940	135.6632
4	173.8407	211.1471	206.6257	108.8890	144.9692	297.6890	273.5942	20.4378	184.2745	66.7384
5	166.8164	211.5698	198.1849	171.8195	149.2179	312.5280	251.3481	282.9092	168.5719	456.7663
6	174.3791	214.2096	196.5957	133.8526	108.8890	250.1409	200.5641	168.5719	122.8672	42.2361
7	187.2879	214.7321	196.4355	178.3487	133.8526	243.5015	147.0866	306.2272	182.0523	259.2207
8	208.7960	215.7857	192.3318	156.2198	121.4489	255.4968	272.8826	100.1195	93.9883	19.1108
9	168.5929	215.9570	191.9474	157.2463	139.2660	244.8776	349.8349	184.1214	147.8875	189.1471
10	171.5375	216.2563	189.9220	192.1458	136.1359	264.8153	184.5469	318.2405	173.1119	288.7666
11	191.9474	216.3817	188.4224	160.4544	130.2040	182.8211	110.3155	308.6822	100.1195	-41.5248
12	175.7662	216.6416	188.2878	151.9053	148.7601	349.0040	204.9120	36.1087	21.0684	262.9241
13	198.1849	217.9330	187.2879	122.3352	134.2665	311.9549	230.3161	314.2794	200.7817	301.5899
14	216.3817	222.3199	180.7014	131.7838	137.8596	172.2054	220.9371	121.3907	30.5462	33.3926
15	216.6416	222.4858	179.2014	186.9102	140.6617	209.8105	199.3217	178.5919	195.1244	606.5155
16	196.5957	223.6412	175.7662	139.2660	131.7838	326.5825	207.5521	117.2607	97.2677	452.7989
17	155.1226	224.5335	174.3791	195.7324	135.3016	161.9369	124.5243	377.3102	185.6084	256.8051
18	180.7014	224.5927	173.8407	188.1153	129.5918	236.8019	63.3903	214.3610	66.2039	265.3711
19	188.4224	224.8389	172.5711	189.6060	107.8126	163.1528	192.3385	269.5915	135.1145	190.2741
20	206.6257	228.4140	171.5375	111.8193	114.8175	187.2521	207.6946	379.2787	109.9881	132.4298
21	188.2878	230.1505	168.5929	176.4493	111.8193	360.9075	161.9757	233.4330	197.2755	400.3139
22	140.0271	233.2895	166.8164	174.5391	130.5740	182.9861	243.5423	211.2236	40.2736	353.6231
23	192.3318	236.8423	163.8207	137.8596	147.9984	274.9965	149.8166	328.8303	157.9088	236.0273
24	163.8207	239.8531	155.1226	180.2522	130.4777	206.1786	249.2858	157.9088	128.9286	316.2360
25	189.9220	242.7216	140.0271	200.6294	137.1452	261.1015	230.1952	200.7817	82.9104	371.0998

N	P1	P1E	P1E-	P2	P2E	P3	P4	P5	P5E	P6
26	188.9202	181.0830	152.5066	206.5678	189.6060	261.7495	298.7178	155.7237	182.6111	7.4654
27	193.4933	181.2382	164.0289	113.7074	206.5678	261.0941	213.9963	278.8240	167.8896	222.6159
28	202.6882	181.8691	181.0830	144.9692	186.9102	253.1203	254.0528	307.9619	126.7272	188.7350
29	199.6879	183.4464	183.4464	164.3340	228.2879	169.0559	168.6606	128.9286	155.7237	260.4024
30	186.1261	183.5757	186.1261	134.2665	200.4730	117.7992	227.9063	93.9883	118.6853	172.6801
31	187.7044	184.1792	186.6576	151.7015	191.7203	177.9824	252.7341	68.2039	209.6190	384.0914
32	202.6785	184.7083	187.7044	151.0450	187.0066	352.6824	246.3555	139.5315	117.2607	130.1922
33	164.0289	185.9234	188.9202	218.2523	218.2523	233.1625	258.0061	176.0036	87.0889	117.3416
34	190.8091	186.1261	190.8091	189.1532	189.1532	216.4676	262.3076	147.8875	203.5873	213.1857
35	205.2009	186.6576	193.4933	114.8175	183.0394	191.5657	344.3725	135.1145	20.4378	119.0475
36	236.8423	187.2879	196.1935	154.6162	223.7609	326.8411	185.4933	318.7293	110.4310	0.9969
37	181.0830	187.7044	199.5798	228.2879	186.6023	213.3676	196.1149	226.3541	184.1214	77.2743
38	242.7216	188.2878	199.6879	174.6899	197.0276	257.6295	167.3472	118.6853	205.9176	266.1083
39	183.4464	188.4224	202.6785	200.4730	200.6294	239.4871	172.1983	87.0889	139.5315	143.3857
40	186.6576	188.9202	202.6882	197.0276	201.9589	281.1245	279.5153	97.2677	163.7726	452.5548
41	199.5798	189.2680	203.8838	181.8864	195.9241	173.0630	198.6786	262.7912	178.5919	374.7672
42	224.5335	189.5028	204.4943	177.7663	197.9255	248.4516	224.0004	314.7602	124.1035	566.0703
43	204.4943	189.5080	205.2009	130.2040	205.8062	222.0692	261.3997	365.2037	107.9692	121.9523
44	206.5501	189.9220	206.5501	191.4739	183.5815	157.3072	181.1746	212.9352	135.3167	87.9197
45	233.2895	190.8091	215.7857	147.9984	191.4739	323.9353	278.2894	82.9104	121.3907	183.5110
46	230.1505	191.9474	224.5335	223.7609	186.0977	299.7394	199.4508	294.3120	196.6708	-69.0611
47	203.8838	192.3318	230.1505	205.8062	188.1153	313.7192	134.3541	262.3281	27.5072	332.6058
48	215.7857	193.0922	233.2895	174.0023	195.4714	280.0992	291.9319	228.5145	85.9155	424.7789
49	152.5066	193.4933	236.8423	140.6617	195.7324	193.4341	198.6164	124.1035	130.7466	260.3081
50	196.1935	194.5717	242.7216	159.3663	192.1458	208.1544	115.4792	185.6084	176.0036	290.1643

N	PI	PIE	PIE-	P2	P2E	P3	P4	P5	P5E	P6
51	217.9330	195.2741	217.9330	151.7091	181.8864	196.4854	265.9268	285.7301	314.2794	55.5609
52	224.5927	196.1935	224.5927	157.8382	173.8236	222.6987	240.3027	135.3167	226.3541	214.7780
53	224.8389	196.4355	224.8389	168.7869	164.6967	299.0966	174.8454	261.9938	307.9619	354.4259
54	170.6378	196.5957	170.6378	183.0394	174.5391	181.1891	145.5743	40.2736	271.1853	164.1951
55	189.5080	198.1849	189.5080	121.4489	171.8195	272.2490	255.8316	271.7362	296.9819	129.9864
56	207.7584	199.5798	207.7584	149.2179	173.8484	292.7954	138.9118	335.6112	325.4578	143.8412
57	177.8603	199.6879	177.8603	161.9212	170.5417	202.3532	273.8143	219.9329	245.5112	449.1960
58	214.2096	199.7988	214.2096	175.9553	175.9553	251.4895	185.2624	355.7023	324.4217	49.9955
59	175.4164	200.2431	175.4164	178.2618	176.2724	236.1926	181.4882	227.2528	318.2405	236.9758
60	207.9818	200.4444	207.9818	186.6023	178.2618	285.9173	60.7783	184.2745	396.0257	223.9754
61	184.7083	200.6506	184.7083	195.9241	173.3969	225.7305	254.7707	205.9176	334.6607	450.7778
62	239.8531	202.5983	239.8531	170.5417	175.8155	204.0052	159.3068	396.0257	252.8325	0.5078
63	215.9570	202.6785	215.9570	169.5218	167.0978	195.8383	273.9486	220.6707	212.9352	-33.3270
64	174.7596	202.6882	174.7596	130.4777	181.9660	210.1293	208.2879	5.3958	351.7052	-35.0668
65	195.2741	203.8838	195.2741	153.9444	176.4483	141.3459	179.5299	250.9653	278.8240	115.5797
66	189.2680	204.4943	189.2680	129.5918	168.7869	259.6396	100.9478	27.5072	285.7301	33.9721
67	166.1364	205.2009	166.1364	160.4421	178.5869	274.2274	155.9917	351.7052	377.3102	117.6669
68	208.1507	206.5501	208.1507	107.8126	169.5218	353.6019	236.5602	325.4578	220.2981	80.5369
69	185.9234	206.6257	185.9234	137.1452	178.3487	182.7254	183.8803	30.5462	214.3610	133.8561
70	223.6412	207.3266	223.6412	176.2724	175.8246	237.1891	213.9715	334.6607	250.9653	288.5658
71	200.6506	207.7584	200.6506	154.2916	174.6899	149.8441	296.1822	245.5112	213.7154	379.7980
72	194.5717	207.9818	194.5717	152.2997	178.5287	218.5724	250.4428	96.3940	314.7602	109.6509
73	208.4410	208.1507	208.4410	175.8246	174.0023	266.0848	293.7198	213.7154	226.5145	171.9162
74	200.4444	208.2927	200.4444	157.5721	180.2522	225.3223	160.8944	209.6190	355.7023	112.9814
75	172.1467	208.4410	172.1467	156.0317	177.7663	265.8800	173.7191	252.8325	219.9329	320.5076

N	P1	P1E	P1E-	P2	P2E	P3	P4	P5	P5E	P6
76	216,2563	140,0271	216,2563	187,0066	160,4421	269,1434	230,2538	196,6708	306,2272	81,6191
77	207,3266	152,5066	207,3266	195,4714	157,2463	234,5791	125,8891	220,2981	328,8303	-21,1488
78	157,1746	155,1226	157,1746	191,7203	151,7015	236,1060	210,0804	275,8009	228,4104	530,7730
79	202,5983	156,3317	202,5983	201,9589	155,8922	361,0497	281,5874	126,7272	269,5915	86,6630
80	184,1792	157,1746	184,1792	155,1969	157,8382	228,8166	191,4639	315,7865	335,6112	312,8083
81	200,2431	163,8207	200,2431	178,5869	156,1042	247,6974	143,1611	109,9881	308,6822	247,4911
82	211,5698	164,0289	211,5698	163,2149	154,2916	291,4559	139,3546	197,2755	262,7912	128,9857
83	193,0922	166,1364	193,0922	148,7601	156,2198	305,7247	199,9371	236,1004	236,1004	522,9630
84	189,50284	166,8164	189,5028	173,8484	164,334	173,7861	356,5591	85,9155	293,0019	154,7355
85	228,4140	168,5929	228,4140	136,1359	152,2997	268,4085	299,5462	271,1853	261,9938	333,1446
86	222,4858	170,6378	222,4858	164,6967	163,2149	248,1672	319,8491	110,4310	271,7362	87,1894
87	183,5757	171,5375	183,5757	141,3794	164,1084	283,2689	181,2532	182,0523	294,3120	463,2446
88	208,2927	172,1467	208,2927	135,3016	153,9711	265,6057	223,7296	324,4217	379,2787	115,5445
89	156,3317	172,5711	156,3317	186,0977	151,0450	235,7725	236,7692	228,4104	233,4330	516,4807
90	172,6216	172,6216	172,6216	149,5886	153,9444	206,2608	290,7334	173,1119	220,6707	138,6910
91	210,3948	173,8407	210,3948	130,5740	151,9053	256,6343	148,8058	203,5873	282,9092	221,8104
92	209,9938	174,3791	209,9938	156,1042	159,3663	289,7384	177,6072	247,2826	211,2236	250,1694
93	222,3199	174,7596	222,3199	178,5287	154,6162	275,1532	177,3185	293,0019	318,7293	236,1352
94	180,6626	175,4164	180,6626	153,9711	161,9212	200,1000	110,6438	167,8896	227,2528	111,6964
95	181,8691	175,7662	181,8691	167,0978	155,1969	267,5007	125,7954	21,0884	365,2037	100,6652
96	199,7988	176,9076	199,7988	164,1084	157,5721	267,3438	149,8987	107,9692	315,7865	143,2138
97	181,2382	177,8603	181,2382	197,9255	151,7091	306,8033	199,0197	163,7726	247,0856	347,7179
98	176,9076	179,2014	176,9076	181,9660	160,4544	254,5132	117,3572	182,6111	247,2826	233,0720
99	214,7321	180,6626	214,7321	175,8155	149,5886	200,2476	157,6168	296,9819	275,8009	274,3619
100	211,1471	180,7014	211,1471	173,8236	156,0317	157,2997	265,8878	195,1244	262,3281	207,1982

N	P1	P1E	P1E-	P2	P2E	P3	P4	P5	P5E	P6
101						413.5915				
102						169.1016				
103						272.4676				
104						288.0113				
105						278.8984				
106						210.8092				
107						233.4232				
108						271.4258				
109						120.1734				
110						179.9893				
111						204.9475				
112						268.1812				
113						263.5383				
114						275.8350				
115						256.1066				
116						290.7237				
117						190.1531				
118						259.4978				
119						126.1250				
120						326.1760				
121						264.9111				
122						248.7922				
123						169.2876				
124						282.5729				
125						315.8924				

N	P1	P1E	P1E-	P2	P2E	P3	P4	P5	P5E	P6
126						298.1787				
127						240.3771				
128						214.3078				
129						275.5238				
130						204.0085				
131						254.8518				
132						231.1906				
133						240.8201				
134						268.7479				
135						311.3860				
136						235.3708				
137						190.5317				
138						235.7756				
139						285.2536				
140						208.8255				
141						120.7441				
142						300.7169				
143						238.7454				
144						288.7973				
145						278.2215				
146						212.6397				
147						270.4352				
148						232.8350				
149						260.8545				
150						204.0324				

N	P1	P1E	P1E-	P2	P2E	P3	P4	P5	P5E	P6
Média	195.1268	195.1268	195.1268	164.8724	164.8724	244.2789	207.4271	205.6230	205.6230	218.1624
VarPop	426.96848	426.96849	426.9685	663.2187	663.2187	2819.394	3683.6717	8679.588	8679.588	22608.63
V(Estr1)	344.271	91.456416	344.271	715.558	153.0643	3625.01	3596.6578	9902.178	3695.977	25119.48
V(Estr2)	463.47318	14.920521	463.4732	1085.627	143.8438	2895.84	2992.8745	7742.176	2541.423	22973.58
V(Estr3)	400.00789	18.614927	400.0079	446.9685	19.07948	2533.576	3691.1523	12135.4	3073.387	19404.6
V(Estr4)	391.89011	100.13692	391.8901	418.8383	17.61132	3790.632	4662.1582	5878.227	2013.618	22840.52
V(Estr5)						1919.677				
C_w	-0.0456	-0.1970	-0.2025	-0.1092	-0.2898	-0.0654	-0.0767	0.0637	-0.2515	-0.1305
C_{wst}	-0.0165	0.7332	-0.1892	-0.1034	0.0239	-0.0664	-0.0723	0.0653	-0.0747	-0.1240