

Estatística espacial

Geoestatística

Profa. Dra. Rúbia Gomes Morato

Estatística espacial

- Diferencia-se da estatística tradicional por considerar a localização, a posição no espaço.
- Estatística espacial
- Geoestatística
- As medidas estatísticas são selecionadas de acordo com o tipo de dado (pontos, área, superfície)

Estatística descritiva espacial

- Centro médio
- Centro médio ponderado
- Centro mediano
- Distância padrão
- Distância relativa
- Dado angular

Geoestatística

- Conceito e aplicação
- Exercício:
 1. Análise dos dados de entrada
 2. Cálculo do variograma experimental
 3. Modelagem do variograma
 4. Interpolação por krigagem

Centro médio

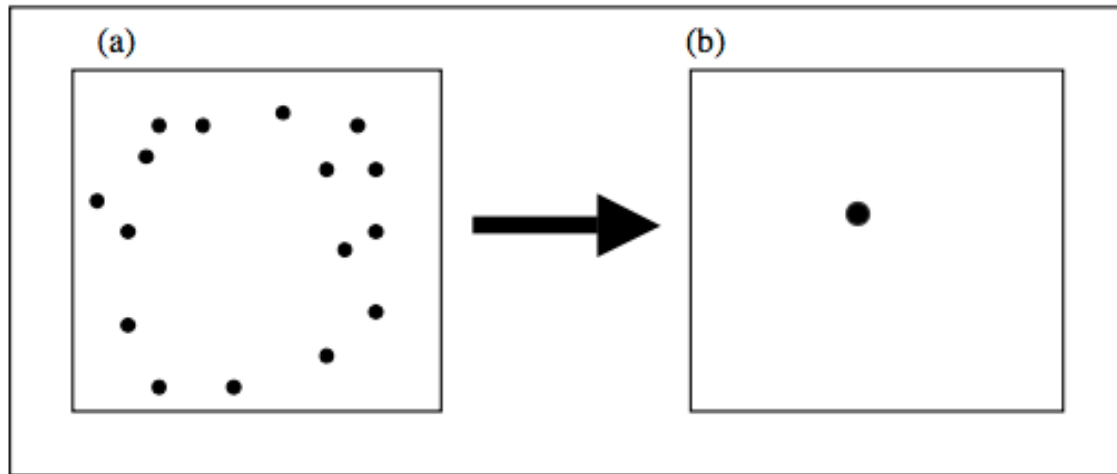
- Média das coordenadas X e Y de dados pontuais (ou dos centroides para polígonos).
- É análogo a média aritmética.

$$X = \frac{\sum x_i}{n}$$

$$Y = \frac{\sum y_i}{n}$$

- Exemplo: apontar o centro médio baseado nas coordenadas X e Y de determinada região.

Centro médio (Tartaruga, 2008)



Centro médio ponderado

- Média ponderada das coordenadas X e Y de dados pontuais (ou dos centroides para polígonos).
- É análogo a média ponderada.

$$X_w = \frac{\sum w_i X_i}{\sum w_i}$$

$$Y_w = \frac{\sum w_i Y_i}{\sum w_i}$$

- Exemplo: apontar o centro médio baseado nas coordenadas X e Y de determinada região, utilizando as respectivas populações como peso.

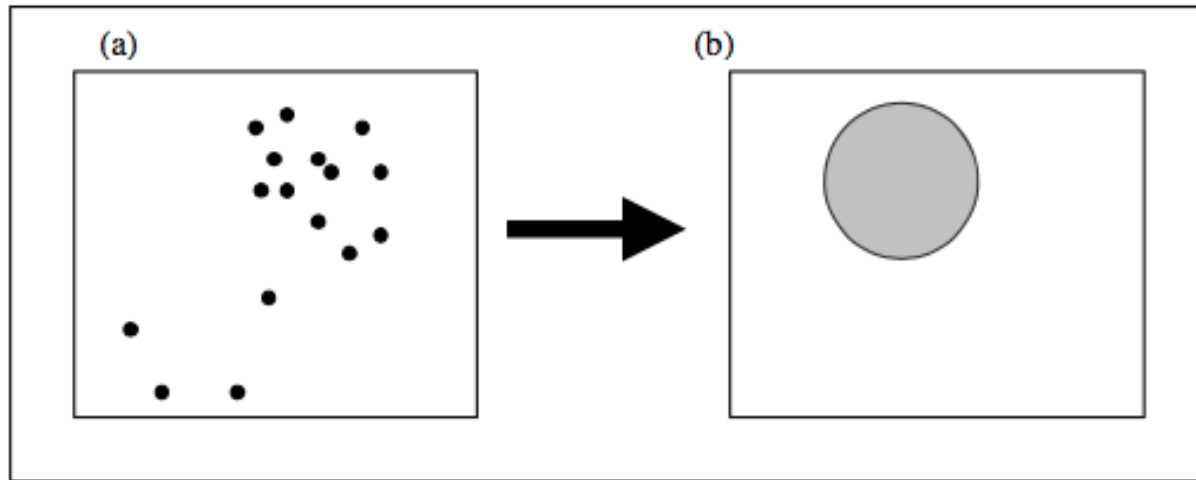
Centro mediano

- Localização que minimiza a soma das distâncias percorridas.
- Conceitualmente mais intuitivo, mas o cálculo é mais complexo, pois é feito iterativamente.
- A partir de um centro inicial, que pode ser o centro médio, por exemplo, as distâncias são recalculadas até o momento que não há diferença significativa entre um cálculo e o seguinte.
- Exemplo: Determinar uma posição de maior acessibilidade para instalação de um equipamento público.

Distância padrão

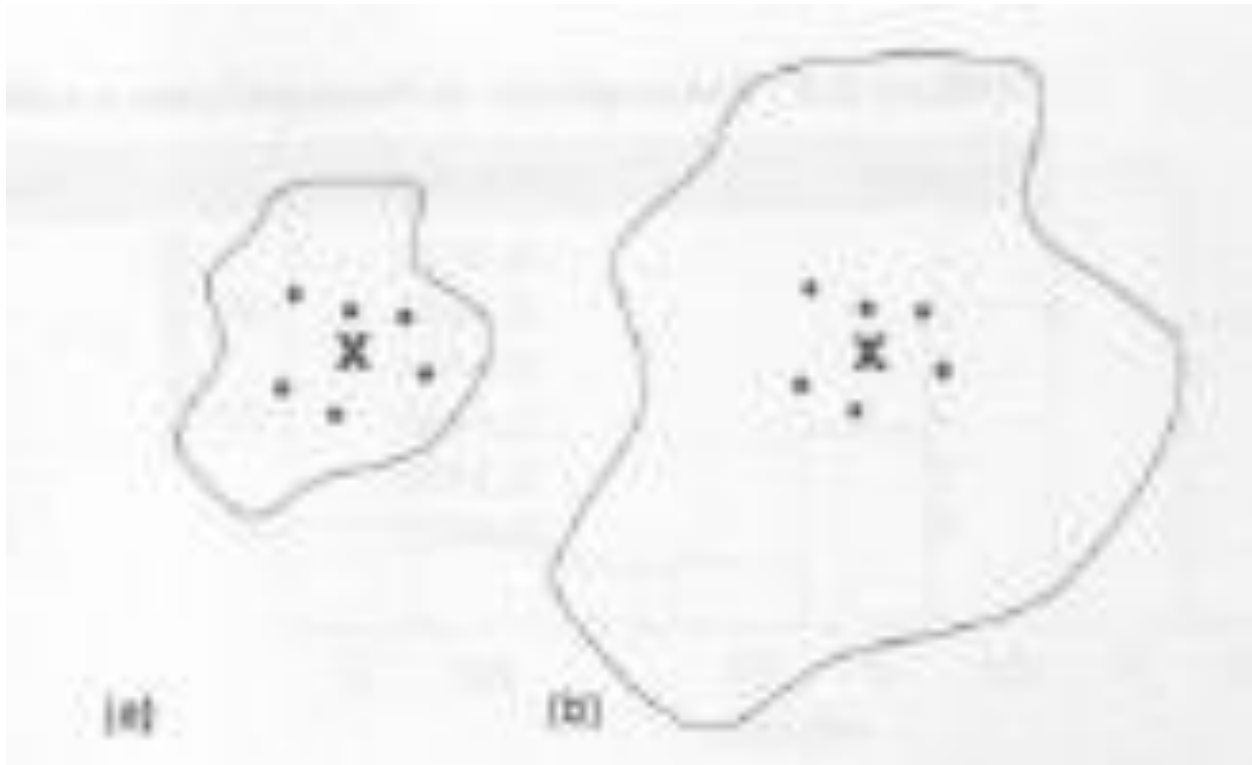
- Distância média das observações em relação ao centro médio (ou o centro ponderado).
- É uma medida do grau de concentração ou dispersão.
- O resultado é afetado pelo tamanho da área de estudo e conserva a unidade de medida (km).
- Exemplos: análise da distribuição da população, da renda ou PIB ao em determinada região. É possível identificar centralidades.

Distância padrão (Tartaruga, 2008)



$$\text{Distância padrão} = \sqrt{\frac{\sum (X_i - \text{Média de } X)^2 + \sum (Y_i - \text{Média de } Y)^2}{N}}$$

Distâncias padrão idênticas e áreas de estudos diferentes (Rogerson, 2012)



Distância relativa

- É obtida dividindo a distância padrão pelo raio de um círculo com área igual ao tamanho da área de estudo, tornando-se numa medida adimensional e padronizada, permitindo comparações.

Dado angular

- Pode ser utilizada para dados de direção dos ventos, ou lineamentos de rochas por exemplo.
- Os dados são divididos em direções.

Autocorrelação espacial

- A autocorrelação espacial é a correlação de uma determinada variável (atributo) z numa área i com os valores dessa mesma variável em áreas vizinhas. Por isso o prefixo AUTO.
- Permite avaliar se há dependência espacial na distribuição de determinada variável.

Índice de Moran

- O Índice de Moran é um coeficiente de medição da correlação espacial. Ele mede a relação do desvio padronizado de uma variável Z numa área i com o desvio padronizado das áreas vizinhas para a mesma variável Z . Numa matriz de vizinhança normalizada o índice de Moran é dado por:

$$I = \frac{\sum_{i \neq j}^n w_{ij} (z_i - \bar{z})(z_j - \bar{z})}{\sum_{i=1}^n (z_i - \bar{z})^2}$$

Índice de Moran

Significado do resultado:

- 0 ausência de correlação espacial (não há padrão espacial)
- 1 correlação espacial direta (valores altos vizinhos de valores altos e valores baixos vizinhos e valores baixos)
- -1 correlação espacial inversa (ocorrência rara)

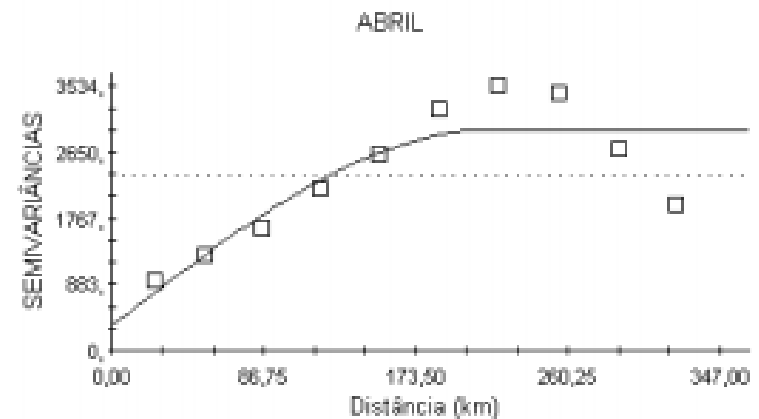
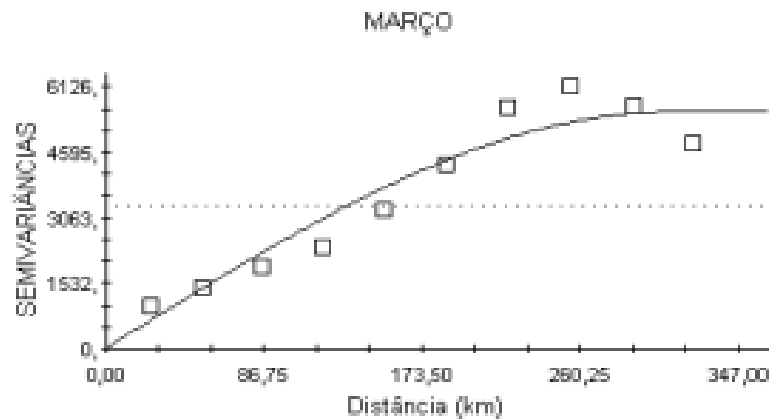
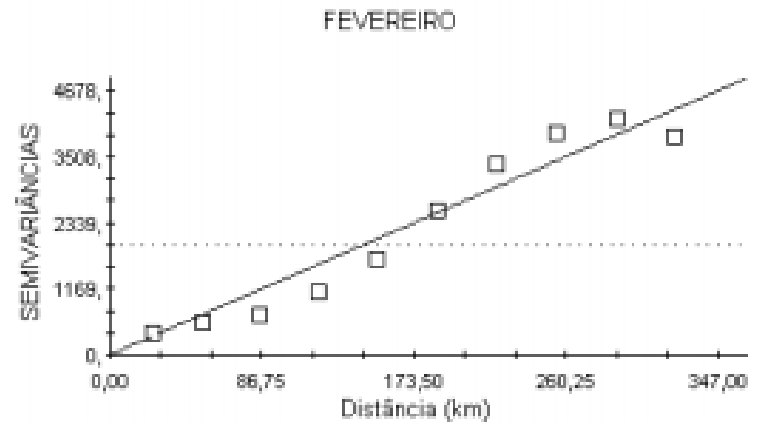
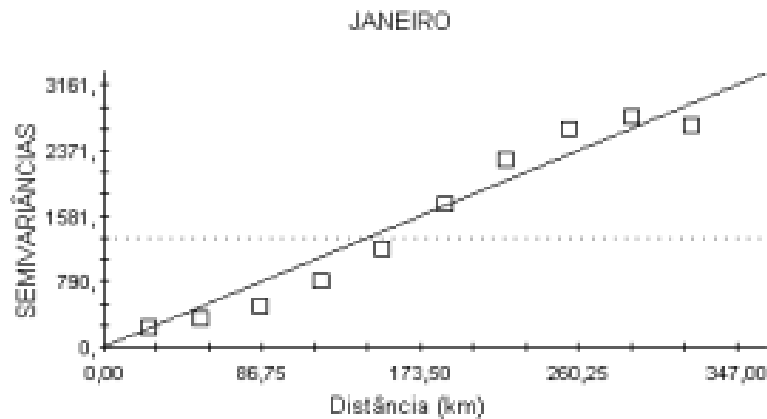
Variograma/Semivariograma

- Apresenta graficamente a relação entre as distâncias entre as áreas em estudo e a média dos desvios do atributo Z entre as áreas (dz)

$$d_z = \sum_{i=1}^n \sum_{j=1}^n w_{ij} (z_i - z_j)^2$$

$$d_z = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (z_i - z_j)^2}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}}$$

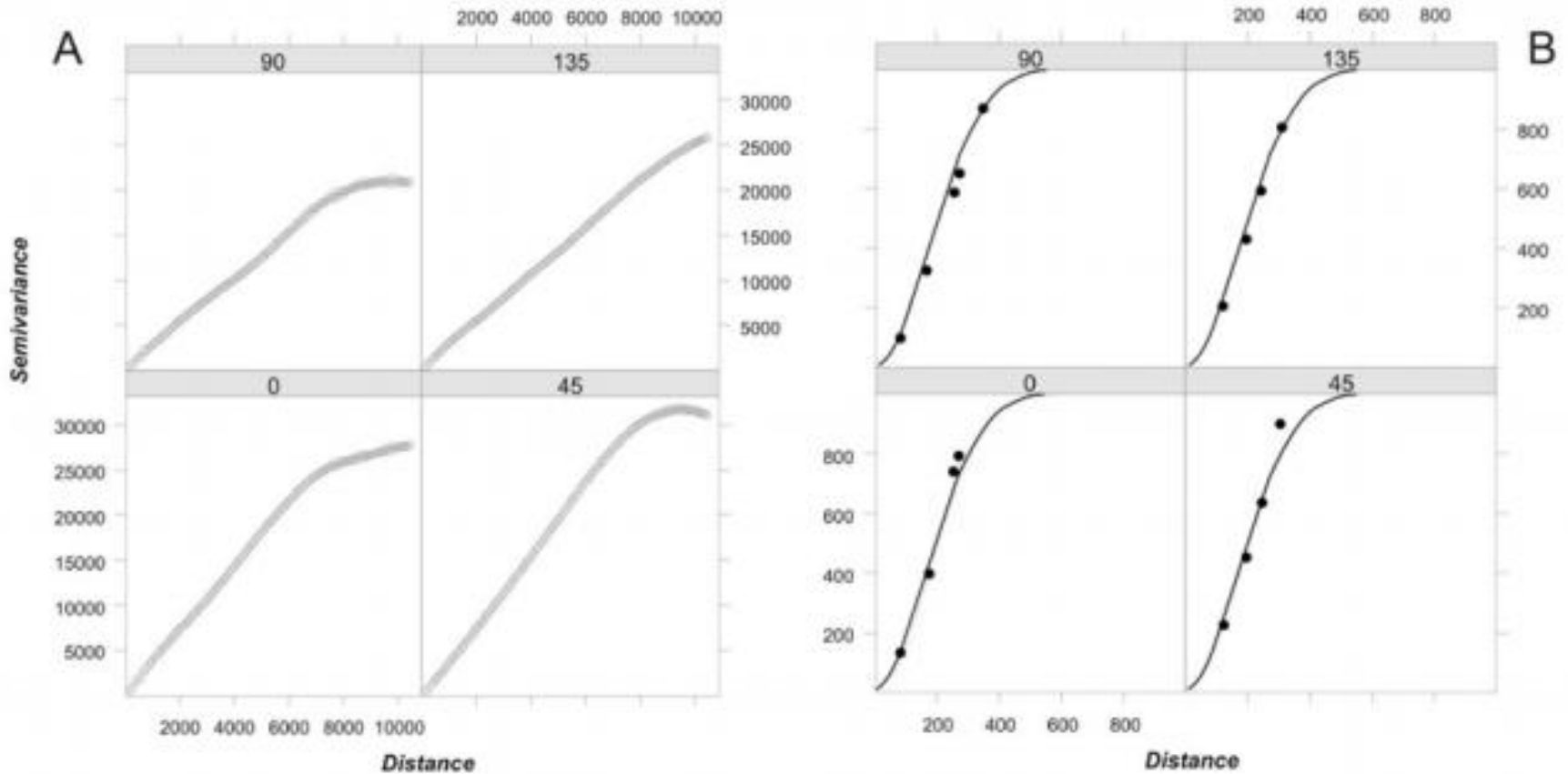
Precipitação pluviométrica na Paraíba (Gomes et al, 2011)



Semivariograma

- Permite analisar a ocorrência de dependência espacial, inclusive em direções diferentes.
- É a base para métodos de interpolação mais sofisticados como a krigagem (Geoestatística).

Superfície topográfica da Serra do Japi (Grohmann e Steiner, 2006)



Índice de Geary

- É uma medida de covariância, que emprega do quadrado das diferenças entre pares de valores:

$$c = \frac{(n-1)}{2 \sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} (x_i - x_j)^2}{\sum_i (x_i - \bar{x})^2}$$

Índice de Geary

- O c de Geary varia entre 0 e 2.
- Valores mais baixos (entre 0 e 1) mostram autocorrelação espacial positiva.
- Valores mais altos (entre 1 e 2) indicam autocorrelação espacial negativa.

Geary x Moran

- O teste C de Geary difere do teste I de Moran por utilizar a diferença entre os pares, enquanto que Moran utiliza a diferença entre cada ponto e a média global (CAMARA et al, 2002).

Geoestatística

- Daniel Krige – década de 1950 (África do Sul)
- Avaliação de jazidas minerais (ouro)
- Georges Matheron – Teoria das variáveis regionalizadas, ou variáveis com condicionamento espacial.

Geoestatística

- Atualmente o termo Geoestatística acha-se consagrado como um tópico especial da estatística aplicada que trata de problemas referentes às variáveis regionalizadas, as quais têm um comportamento espacial mostrando características intermediárias entre as variáveis verdadeiramente aleatórias e as totalmente determinísticas (Landim, 2006).

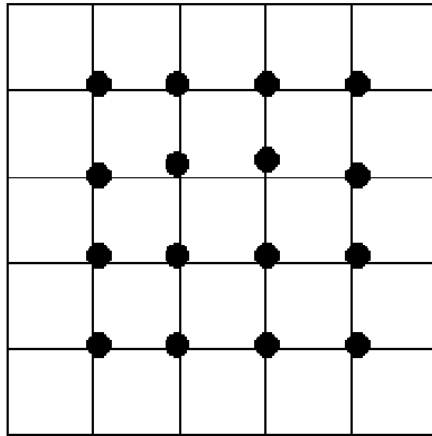
Geoestatística

- Uma das principais aplicações é a estimativa de valores de variáveis regionalizadas a partir de amostras bidimensionais ou tridimensionais.
- As estimativas podem ser feitas baseadas em uma série de parâmetros (mencionados a seguir) que ajudam a aumentar a qualidade dos resultados.

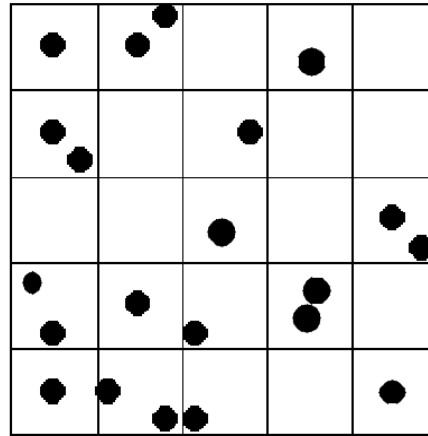
Principais tipos de amostragem

- Regular: espaçamentos sistemáticos entre os dados amostrados.
- Randômica (aleatória): não segue um padrão definido de distribuição.
- Híbrida: as amostras distribuídas aleatoriamente seguem uma regularidade fixada por um ponto em cada bloco.
- *Cluster*: a amostragem é concentrada.
- Transecto: muito utilizado para perfis de rios e praias.
- Curvas de nível: seguem contornos com o mesmo valor.

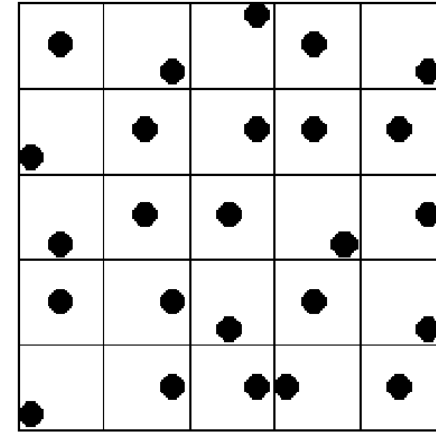
. Principais tipos de amostragens Burrough (1998)



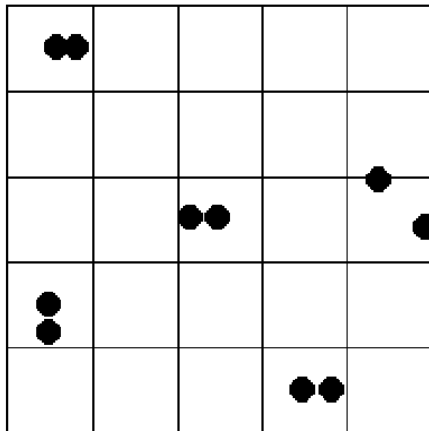
a) Regular



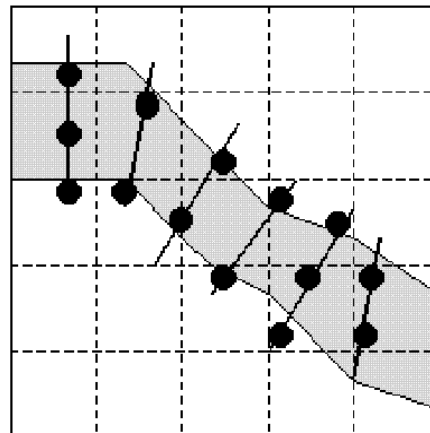
b) Randômica



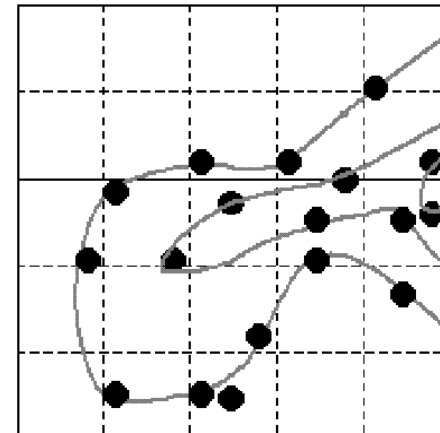
c) Randômica Estratificada



d) "Cluster"



e) Transversal

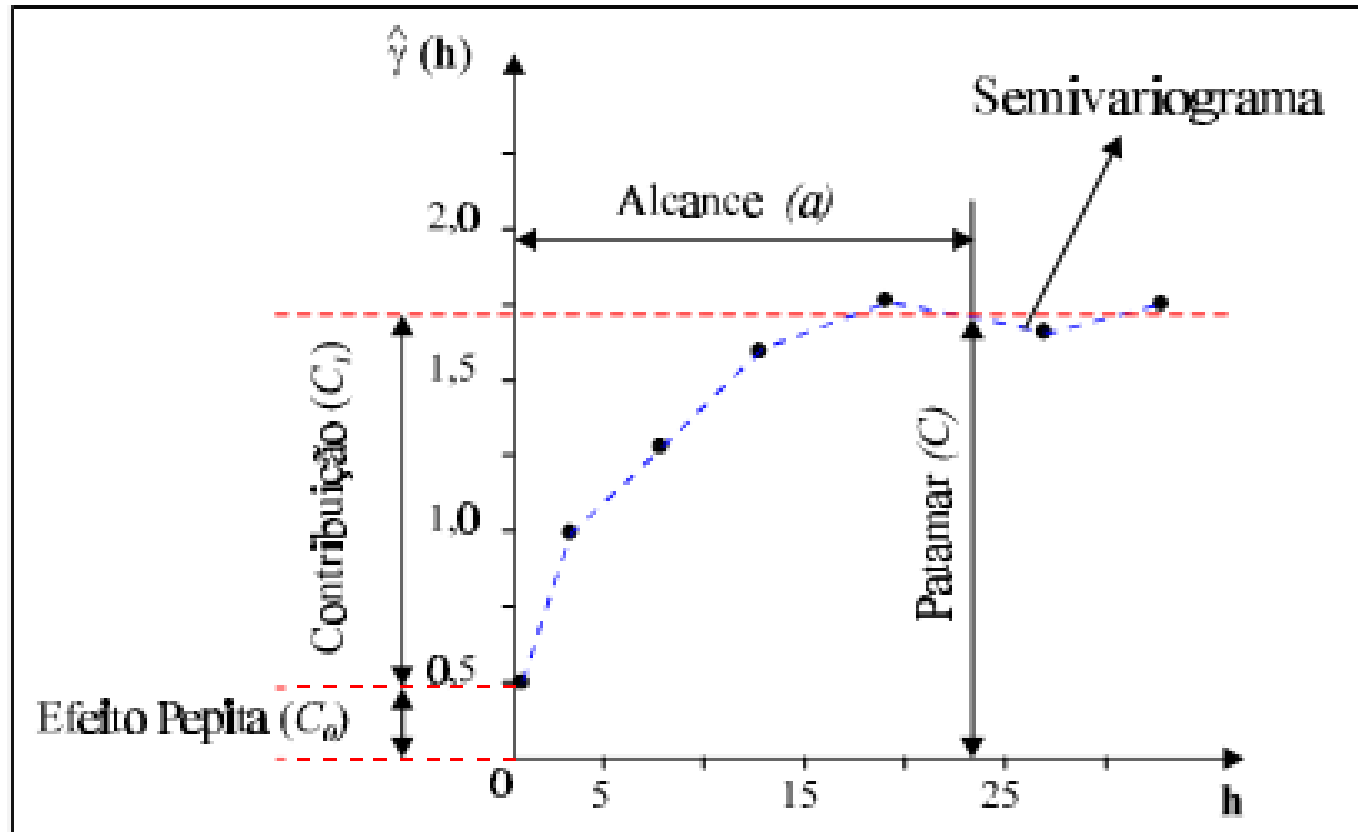


f) Contorno

Etapas da estimativa por Krigagem

1. Análise dos dados de entrada
2. Cálculo do variograma experimental
3. Modelagem do variograma
4. Interpolação por krigagem

Exemplo de semivariograma (Camargo, 1998)



Parâmetros do variograma (Camargo, 1998)

- **Alcance (a):** distância dentro da qual as amostras apresentam-se correlacionadas espacialmente. Na Figura 5.3, o alcance ocorre próximo de 25m.
- **Patamar (C):** é o valor do semivariograma correspondente a seu alcance (a). Deste ponto em diante, considera-se que não existe mais dependência espacial entre as amostras, porque a variância da diferença entre pares de amostras torna-se invariante com a distância.

Parâmetros do variograma (Camargo, 1998)

- **Efeito Pepita (C_0):** idealmente, $g(0)=0$. Entretanto, na prática, à medida que h tende para 0 (zero), $g(h)$ se aproxima de um valor positivo chamado Efeito Pepita (C_0), que revela a descontinuidade do semivariograma para distâncias menores do que a menor distância entre as amostras. Parte desta descontinuidade pode ser também devida a erros de medição (Isaaks e Srivastava, 1989), mas é impossível quantificar se a maior contribuição provém dos erros de medição ou da variabilidade de pequena escala não captada pela amostragem.
- **Contribuição (C_1):** é a diferença entre o patamar (C) e o Efeito Pepita (C_0).

Exercício

Dados do capítulo Spatial data
analysis: geostatistical Tools do ILWIS
2.1 Applications Guide.

Etapas da estimativa por Krigagem

1. Análise dos dados de entrada
2. Cálculo do variograma experimental
3. Modelagem do variograma
4. Interpolação por Krigagem

1. Análise dos dados de entrada

- O mapa C14 apresenta medidas de Carbono 14 na água em Botsuana. O mapa de pontos está ligado a tabela de atributos de mesmo nome.
- Clicar duas vezes na tabela C14, selecionar a opção Attribute, selecionar c14, confirmar com ok e visualizar os dados.
- Clicar duas vezes no arquivo de pontos C14 e visualizar os dados. Avalie a distribuição dos pontos!

1. Análise dos dados de entrada

- Com o compasso, verificar a distância entre os pontos. Também é possível inserir um grid com a distância desejada a partir do menu Layers, Add grid lines, ok (10000 m é um valor razoável!).
- Estime o tamanho do mapa, as distâncias mínimas e máximas entre pares de pontos.

2. Correlação espacial e semivariograma empírico

- Faça a correlação espacial novamente, com 5000.
- Clicar com o lado direito do mouse no mapa de pontos C14, selecionar Statistics, em seguida. Spatial Correlation. Especificar um espaçamento de 5000 m e chamar a tabela de saída de Omnidirectional_5.
- Clique no ícone Create Graph no menu. Selecione a coluna Distance column para o eixo X e SemiVar para o eixo Y, e clique em ok.
- Explique as diferenças para os valores referentes distância 20000 nas duas tabelas.

2. Correlação espacial e semivariograma empírico

- A krigagem admite a correlação espacial entre os dados de entrada.
- Clicar com o lado direito do mouse no mapa de pontos C14, selecionar Statistics, em seguida. Spatial Correlation. Especificar um espaçamento de 10000 m e chamar a tabela de saída de Omnidirectional_10.
- Clique no ícone Create Graph no menu. Selecione a coluna Distance column para o eixo X e SemiVar para o eixo Y, e clique em ok.
- Avalie a distribuição dos pontos!

3. Modelagem do semivariograma

- No menu no gráfico, selecione Edit, Add New Graph, Semivariogram model. Selecione o modelo Spherical, use a distância de 200 m para nugget (efeito pepita), 800 m para sill (contribuição) e 40000 m para range (alcance), e clique em ok.
- Clique em Spherical model na barra lateral para reabrir as opções de modelagem. Altere os valores para melhorar o ajuste dos pontos.

3. Modelagem do semivariograma

- Crie outros modelos (No menu no gráfico, selecione Edit, Add New Graph, Semivariogram model, escolha modelo e defina os parâmetros).
- Teste os parâmetros e compare os modelos. Qual tem melhor ajuste? **Anote os dados do melhor para a etapa seguinte.**

4. Estimativa por Krigagem

Dependendo da versão do Ilwis esta operação pode ser necessária:

- Clique com o lado direito no mapa de pontos C14, selecione Vector Operations e Attribute Map. Na janela do mapa de atributos selecione C14 para point map, C14 para table, C14 para Attribute e preencha C14K para o mapa de saída (output point map).

4. Estimativa por Krigagem

- Recupere o modelo de semivariograma com melhor ajuste e seus parâmetros:
- tipo (esférico, circular, exponencial, etc.)
- distância para nugget (efeito pepita)
- sill (contribuição)
- range (alcance),

4. Estimativa por Krigagem

- Na janela principal do Ilwis, selecione Kriging na (operation-list – lista de comandos do esquerdo). Na janela Kriging selecione o mapa C14K, o tipo de modelo de semivariograma e os parâmetros do semivariograma que apresentaram o melhor ajuste.
- A distância limite pode ser 25000 (25 km), o número mínimo de pontos 4 e máximo 16, nomeie o arquivo como Ord_krig2, selecione a georreferência C14, habilite o mapa de erros e clique em show.

4. Estimativa por Krigagem

- Avalie a interpolação e os erros. Compare com colegas que utilizaram modelos diferentes!
- Se tiver interesse, refaça utilizando outro modelo de semivariograma e novos parâmetros.

Referências bibliográficas

- BURROUGH, P. A.; MCDONNELL, R. Principles of Geographical Information Systems. Oxford: University Press, 1998, 333p.
- CÂMARA, G. et al. Análise espacial de áreas. In: FUCKS, S. D. et al. Análise espacial de dados geográficos. 1ª ed. São José dos Campos: INPE, 2002. Disponível em: <http://www.dpi.inpe.br/gilberto/tutoriais/analise/cap2.pdf>
- CAMARGO, E.C.G. Geoestatística: fundamentos e aplicações. In: CÂMARA, G. & MEDEIROS, J.S. eds. Geoprocessamento para projetos ambientais. São José dos Campos: INPE, 1998. Disponível em: http://www.dpi.inpe.br/gilberto/tutoriais/gis_ambiente/5geoest.pdf.
- GOMES, O. M.; SOUZA, F. A. S.; SANTOS, C. A. C.; PAIVA, W. Análise geoestatística da precipitação pluvial do Estado da Paraíba. Revista Brasileira de Geografia Física, Recife, v. 4, n. 4, p. 692-702, 2011. Disponível em: <<https://periodicos.ufpe.br/revistas/rbgfe/article/view/232712/26719>>.
- Grohmann, C.H.; Steiner, S.S. 2008. SRTM resample with Short Distance-Low Nugget Kriging. International Symposium on Terrain Analysis and Digital Terrain Modelling. Nanjing, China.
- Landim, P. M. B. Sobre geoestatística e mapas. Terra e Didática, v.2, p.19-33, 2006. Disponível em: <http://www.ppegeo.igc.usp.br/index.php/TED/article/view/8334>
- TARTARUGA, I. G. P. Análise espacial da centralidade e da dispersão da população e da riqueza gaúchas de 1970 a 2000: metodologias e notas preliminares. Porto Alegre: FEE, 2008. (Textos para discussão FEE, n. 43). Disponível em: <<http://www.fee.rs.gov.br/sitefee/download/tds/043.pdf>>.