UMA INTRODUÇÃO AO SEMIVARIOGRAMA: E SUA IMPORTÂNCIA DENTRO DA GEOESTATISTICA

Dionizio dos Santos[1]

Resumo: **O artigo em questão tem como foco principal introduzir um dos conceitos de grande importância da geoestatistica, o semivariograma.** O semivariograma é uma função chave na geoestatística, pois é utilizado para ajustar um modelo de correlação espacial do fenômeno observado. Além de ser um pré requisito para a krigagem, a Geoestatística tem sido difundida devido suas propriedades e também por causa de um grande número de softwares que realizam sua interpolação. A geoestatística é um modelo probabilístico e baseando-se-na aleatoriedade dos dados avalia a correlação espacial; entre o valor de uma variável em um local conhecido e o valor da mesma variável localizada em sua vizinhança (HUIJBREGTS, 1975).

Palavras-Chave: Coeficiente de correlação; variáveis regionalizadas; variograma.

INTRODUÇÃO

A Geoestatistica é uma ferramenta importante de analise das variáveis regionalizadas. Em 1951 Daniel G. Krige, na África do sul, analisando dados de concentração de ouro verificou que para encontrar sentido nas variâncias era preciso levar em conta as distancias das amostras, e assim desenvolveu empiricamente uma técnica própria de estimativa para o cálculo de reservas minerais. Já a formalização e a Geoestatistica propriamente dita é desenvolvida em 1963 por George Matheron na França, Dessa formalização surge a teoria das variáveis regionalizadas, tornando-se o fundamento da geoestatística.

Na prática a teoria das Variáveis Regionalizadas (VRs) é a esperança de que, na média, as amostras próximas, tanto no tempo quanto no espaço, terão mais similaridade entre si do que aquelas que se encontram mais distantes. Isto gera padrões que a Geoestatística procura reconhecer. Dessa forma a geoestatística possibilita a realização de interpolações, que ajudam na construção de mapas mais precisos.

Exemplos de Variáveis Regionalizadas (VRs), a precipitação anual da chuva, os teores de elementos ou substâncias, espessura de uma camada de rochas, taxas de doenças e até a demografia (densidade populacional). Ou seja, se há variáveis que obedeçam a essa dependência espacial, a principio poderá ser considerada um VR.

O variograma é uma ferramenta básica de suporte às técnicas de geoestatística, que permite representar quantitativamente a variação de um fenômeno regionalizado no espaço (Huijbregts, 1975). De outra maneira pode-se dizer que o variograma faz a analises estrutural dos dados e estatisticamente modela os dados no espaço. E a krigagem se utiliza do variograma para fazer a interpolação. O variograma não é parte da Krigagem, mas é um requisito.

MATERIAIS E MÉTODOS

[1] Graduando, Universidade Estadual de Londrina, dionizio10@uel.br

Variáveis regionalizadas (VRs)

Antes de apresentar o semivariograma, é importante definir as variáveis regionalizadas (VRs).

As VRs possuem um aspecto estruturado e um aleatório. O estruturado está ligado à distribuição global do fenômeno estudado. Já o aspecto aleatório está relacionado com o comportamento local do fenômeno estudado. Usando como exemplo de estudo uma área poluída (figura 1), têm-se zonas com médias maiores de concentrações de metal pesado do que outras, este é o chamado aspecto estruturado de uma variável regionalizada. Utilizando o mesmo exemplo, dentro de uma zona de concentração os teores do metal pesado apresentam flutuações, e esta característica é chamada de aspecto aleatório. Ou seja, localmente temos flutuações, mas o fenômeno estudado visto globalmente possui uma média que varia espacialmente. E é este comportamento global, que é importante no tratamento das VRs dentro da Geoestatistica.

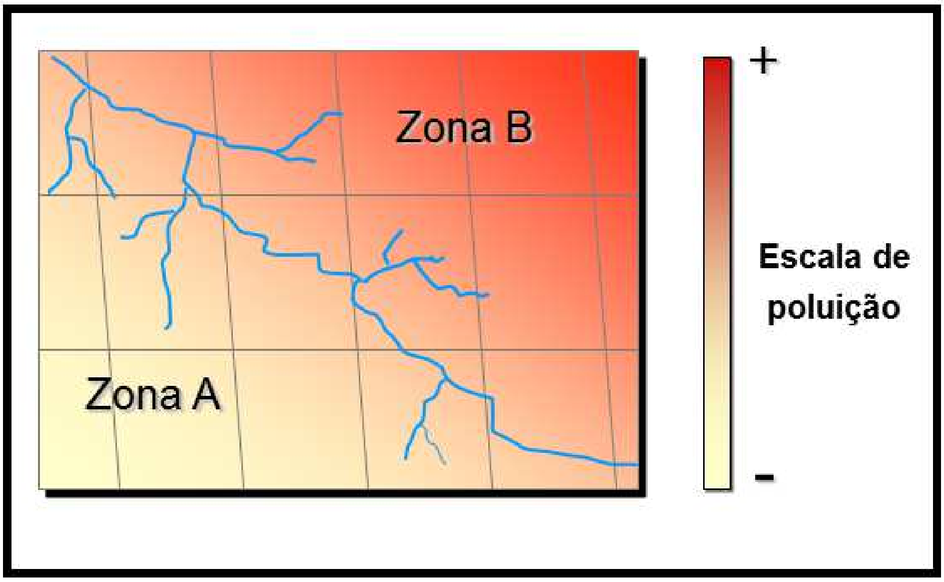

figura 1

Scaterplot, covariancia e coeficiente de correlação

Conhecer alguns conceitos da estatistica descritiva como por exemplo scatterplot, covariância e coeficiente de correlação também são imortantes também na comprrensão da construção do semivariograma.

Scatterplot

O Scaterplot é um gráfico de dispersão bivariado , plotado em um plano cartesiana em qual a abscissa e as ordenadas são quaisquer duas variáveis.

Os gráficos de dispersão são preparados para explorar a direção e força de associação entre dois atributos. Como pode ser visto a seguir no gráfico 1.

Quando a nuvem de dados está mais concentrada, mais associadas estão as váriaveis e quanto mais espalhada menor a associação entre as váriaveis. Também pode ser utilizado para perceber dados muito afastados da nuvem, que podem vir a ser caracterizados como outliers. Os outliers são pontos que foram ou coletados errados, ou valores errados inseridos, ou também pode ser uma varição local da zona amostrada.

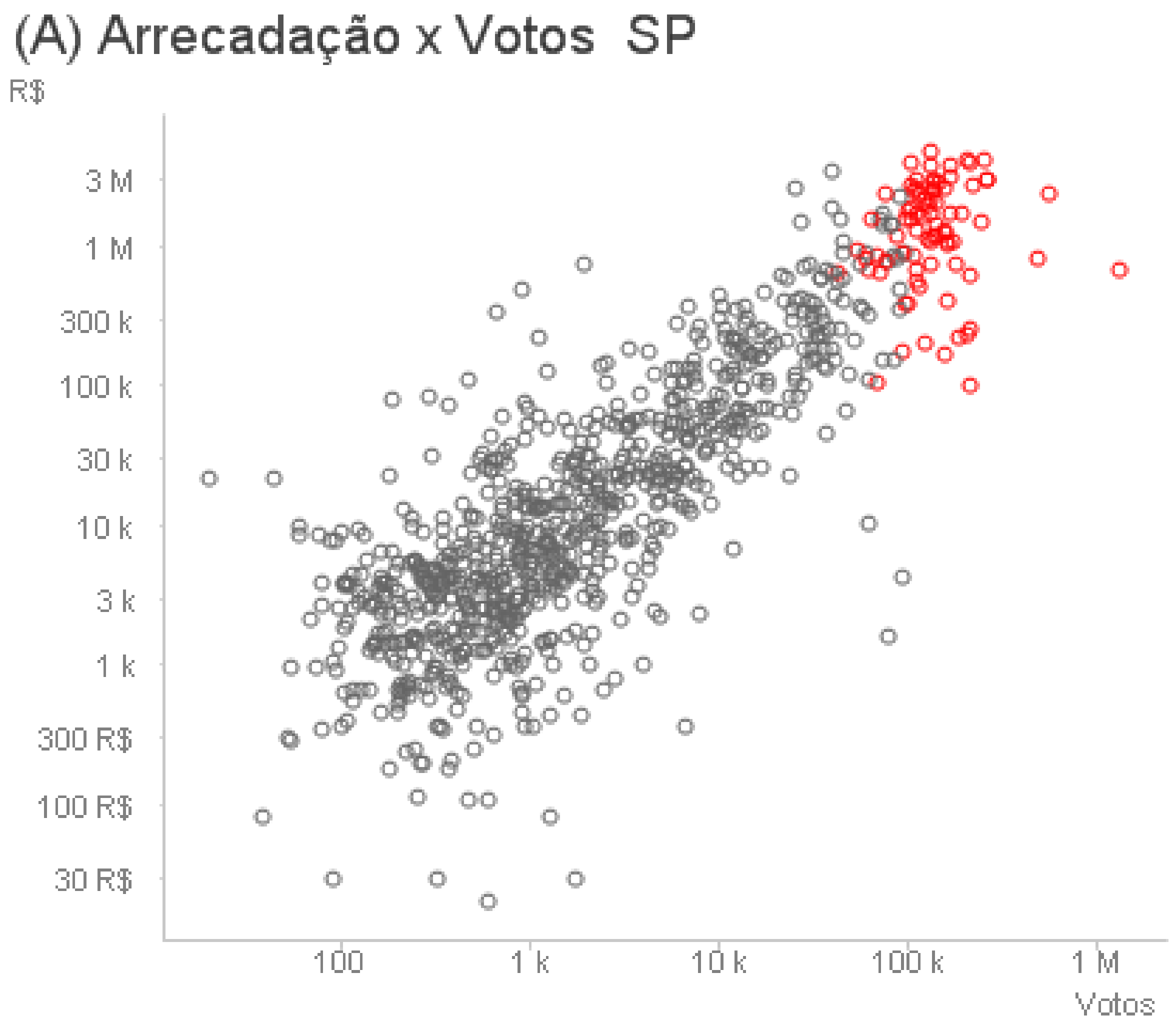

GRAFICO 1: Gráfico de dispersão feito pela revista Época em 2011. Através do uso de um gráfico de dispersão, identificou a relação entre o investimento na campanha e o resultado das urnas nas eleições para Deputado Federal de 2010: quanto maior a arrecadação (em reais), maior a chance de o candidato se eleger. Os pontos marcados em vermelho representam candidatos eleitos em 2010.

Covariância

A Covariância é utilizada para compreender a direção da relação entre as variáveis. Valores de covariância positivos indicam que valores acima da média de uma variável estão associados a valores médios acima da outra variável e abaixo dos valores médios são igualmente associados. Valores de covariância negativos indicam que valores acima da média de uma variável estão associados com valores médios abaixo da outra variável, ver gráfico 3. Mas não indica qual é intensidade de relação entre elas. A seguir a equação da Covariância.

$$Cov_{x,y} = E\left[(X - \mu_x)(Y - \mu_y)\right] \quad (3)$$

Nos Exemplos a seguir temos:

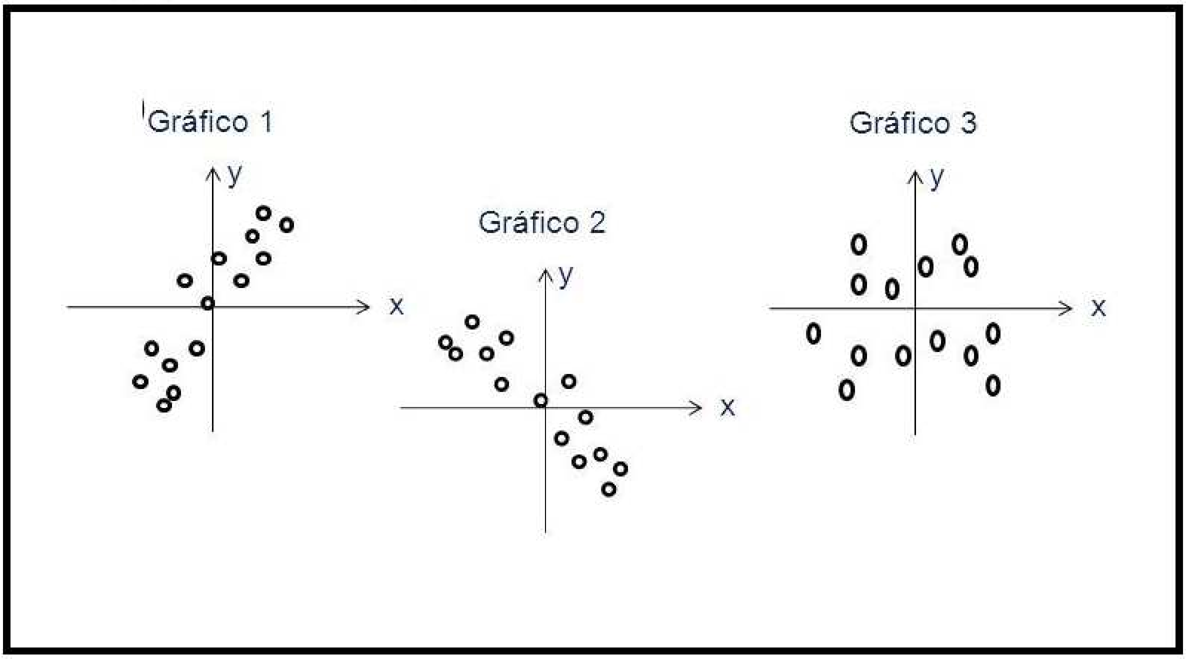

GRAFICO 2: exemplos gráficos da covariância. Se positiva indica relação linear positiva entre X e Y (graf.1). Se Negativa, indica relação linear negativa entre X e Y (graf. 2). E se próximo de zero, não há associação linear entre X e Y (graf.3)

Coeficiente de correlação

Já o coeficiente de correlação é o mais usado para resumir comparações bivariadas. Por ser padronizado, com valores que variam entre 1, 0 e -1. O que a correlação procura explicar como uma variável se comporta enquanto uma outra está variando, desta maneira pode-se identificar se existe alguma relação entre a variabilidade de ambas, o coeficiente de correlação quantifica a relação entre as variáveis.

A Equação é:

$$\rho = \frac{Cov_{x,y}}{\sigma_x . \sigma_y} \quad (4)$$

Graficamente temos

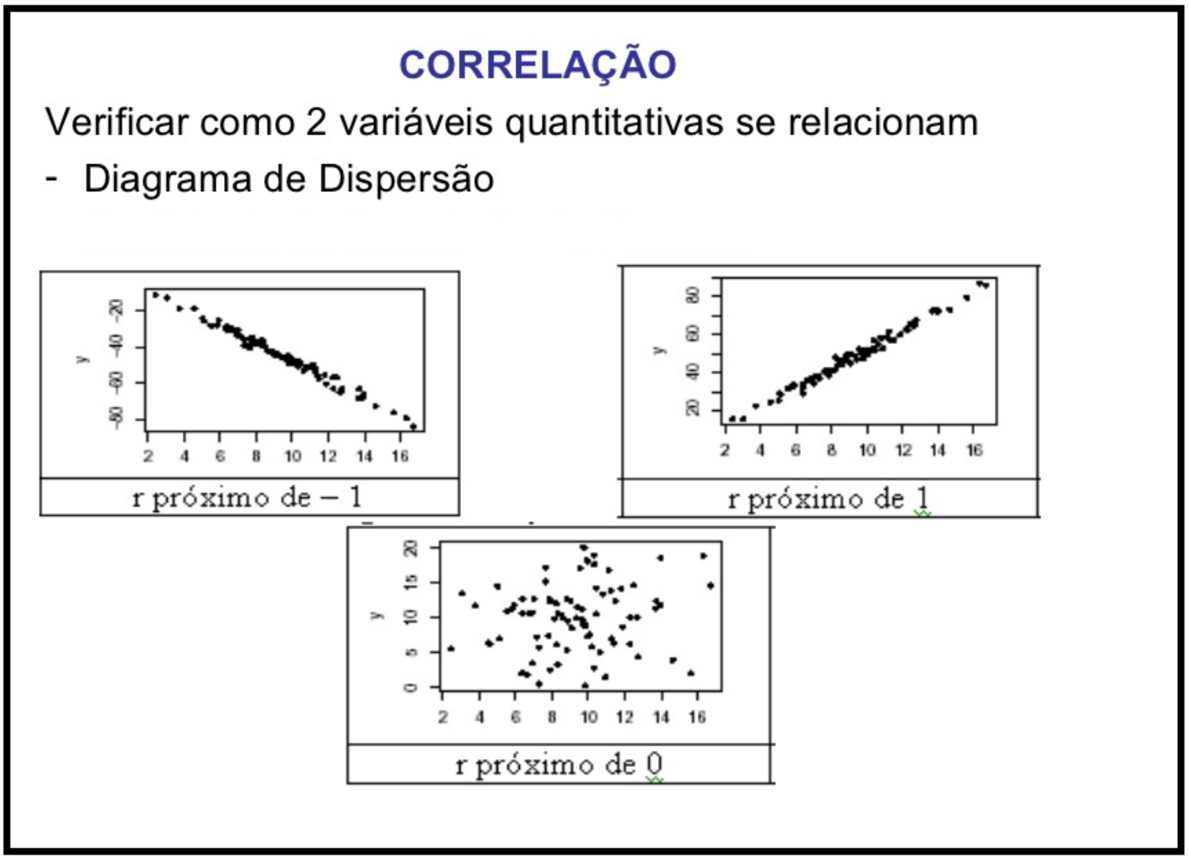

GRAFICO 3

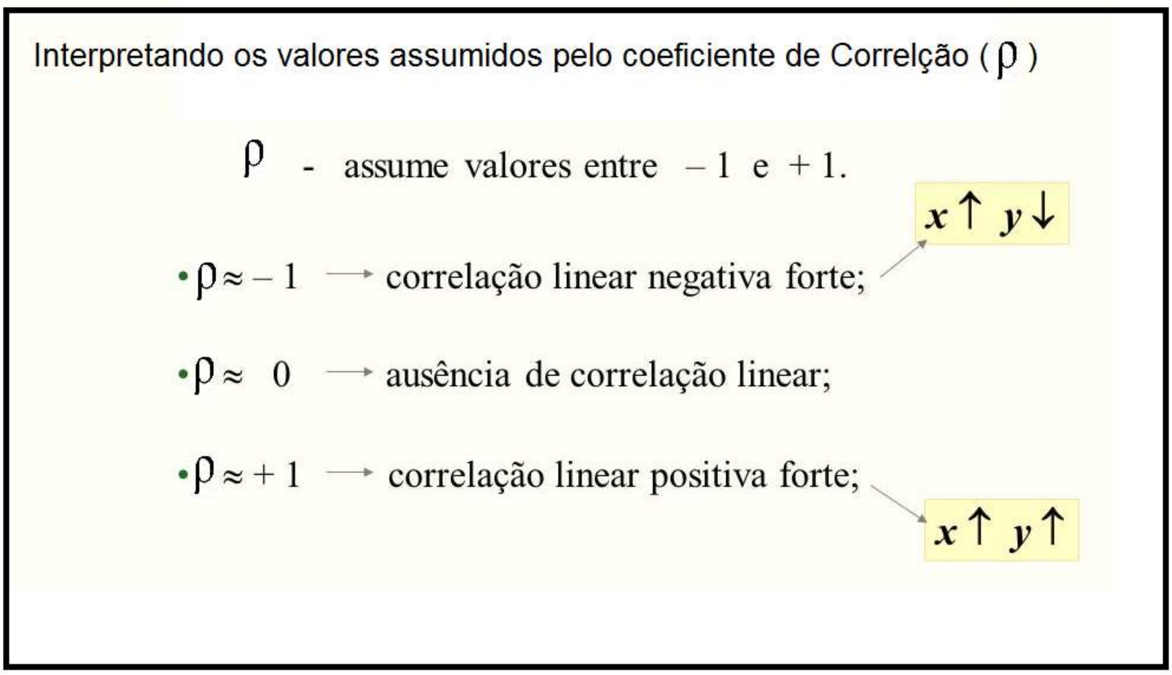

FIGURA1: Interpretação do Coeficiente Angular

A força do coeficiente de correlação pode ser dada em módulo da seguinte forma:

$|\varrho| \geq 0.8 \rightarrow$ Relação Linear Forte

$0.5 < |\varrho| < 0.8 \rightarrow$ Relação Linear Moderada

$|\varrho| \leq 0.5 \rightarrow$ Relação Linear Fraca

O semivariograma

Os dados que são usados para a construção do variograma, são as variáveis regionalizadas. O variograma é dado pela seguinte função:

$$2(\gamma) = \frac{1}{N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2 \qquad \textbf{(5)}$$

Normalmente na literatura, é usado semivariograma, só que para uma melhor praticidade, é chamado de variograma. Então a função descrita acima passa a ser:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2 \qquad \textbf{(5.1)}$$

Desta forma função variograma pode ser definida como sendo:

> "[...] a esperança matemática do quadrado da diferença entre os valores de pontos no espaço, separados por uma distância h [...]" (YAMAMOTO, 1991, p. 27).

Onde:

2γ(**h**)- é o variograma estimado.

N(**h**) - é o número de pares de valores medidos, z(**x**i) e z(**x**i+**h**), separados por um vetor de distancia **h**;

Z(**x**i) e Z(**x**i+**h**), - são valores da i-ésima observação da variável regionalizada, coletados nos pontos **x**i e **x**i+**h** (i = 1, ..., n), separados pelo vetor **h**.

A Figura nos da um exemplo de um variograma experimental. Apenas mostrando o alcance.

No presente artigo será considerado um fenômeno é isotrópico, ou seja, onde o comportamento é igual em todas as direções. Com isso temos que a determinação experimental do variograma depende apenas da distância entre as amostras e não da direção relativa entre elas. Dai tem-se o comportamento do variograma experimental, mostrado na figura 2. Espera-se que observações mais próximas geograficamente tenham um comportamento mais semelhante entre si do que aquelas separadas por maiores distâncias. Assim, o valor absoluto da diferença entre duas amostras z(x) e z(x+h) deveria crescer à medida que aumenta a distância entre elas, até um valor na qual os efeitos locais não teriam mais influência.

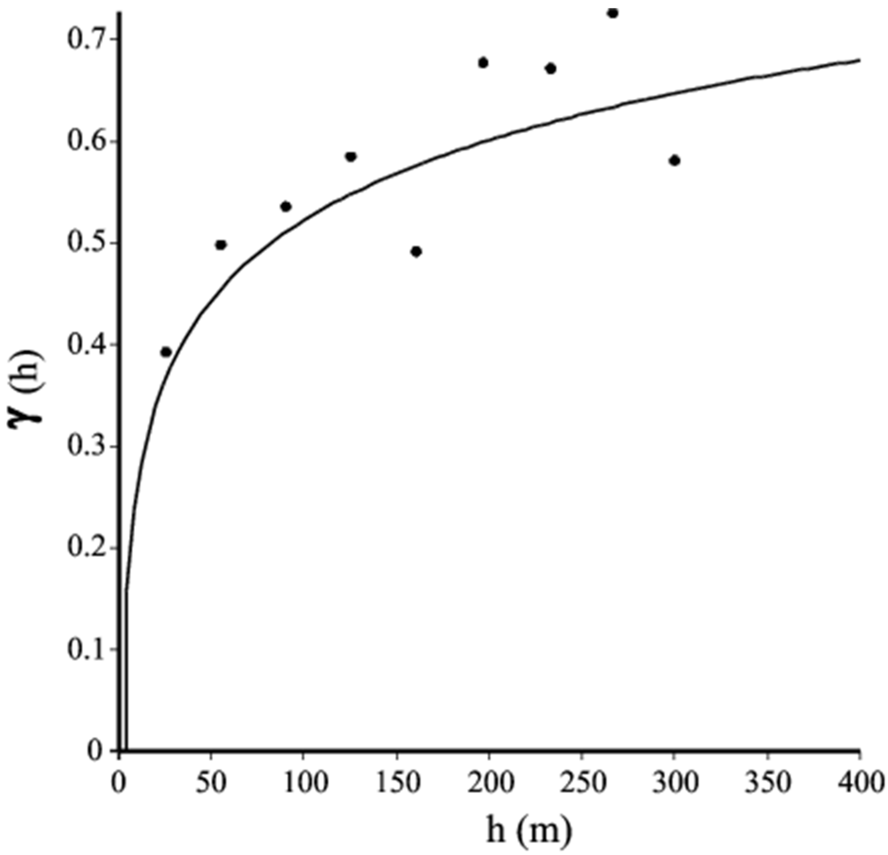

GRÁFICO 4: Exemplo de v**ariograma experimental. Mostrando o alcance.**

Nas propriedades do variograma (figura 2), temos o campo estruturado que também é chamado de zona de influência. Isto significa que, sendo o variograma uma função que cresce com o módulo do vetor h, quando esse módulo aumenta a variação média estre esses pares de amostras também aumenta, até uma certa distancia, que é chamada de Alcance. A partir dai os pares não mais apresentam dependência espacial. Esta parte onde não apresenta dependência espacial é chamada de patamar ou campo aleatório. Esse estado de independência significa que não há mais correlação espacial. E é ai que se encontra o campo de domínio da estatística clássica.

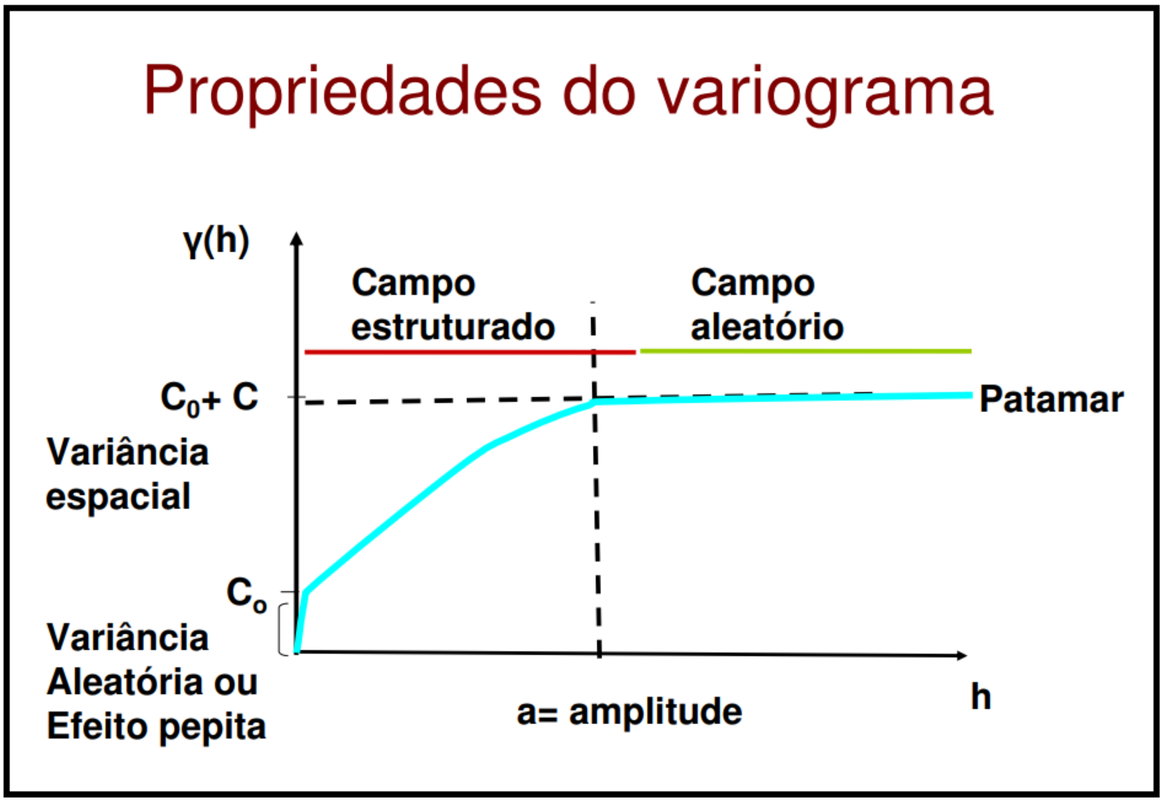

GRÁFICO 5: Propriedades do Variograma

O variograma fornece um significado preciso do conceito de zona de influência de uma amostra. É uma função dada por uma curva (Figura 2). E é crescente com o aumento da distância h que separa pares de amostras, de tal forma que quanto mais distantes as amostras entre si, maior a diferença entre seus teores, e, portanto, menor a continuidade, ou dependência espacial, entre as mesmas (MATHERON, 1963) (MATHERON, 1971).

Os Métodos estatísticos não sofrem influência da localização espacial das amostras, considerando-as independentes espacialmente, levando apenas em conta a variabilidade do conjunto de dados. Já os métodos geoestatísticos levam em conta a correlação espacial entre as amostras.

A seguir para responder essa pergunta, será utilizado um exemplo didático dado pelos professores Maranhão, 1983 (pg. 101) Yamamoto, 2001 (cap. 4).

Considerando os dados de o gráfico a seguir

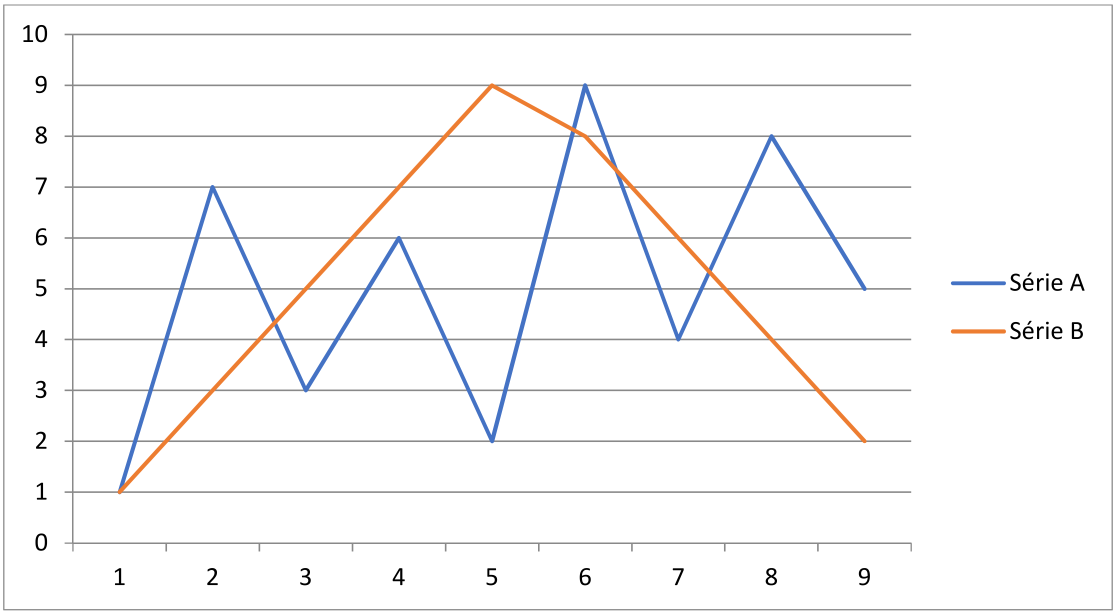

GRÁFICO 6: Série A (1 , 7, 3, 6, 2, 9, 4, 8, 5). Série B (1, 3, 5, 7, 9, 8, 6, 4, 2)

Ao se calcular a média e a variância para os dados da Série A e B, não se obtém nenhuma diferença, muito embora graficamente nota-se que são totalmente diferentes.

TABELA 1

SÉRIE	MÉDIA	VARIÂNCIA
A	5	6,67
B	5	6.67

As características estatísticas são idênticas, entretanto as duas séries são bem diferentes.

A estatística clássica não consegue reconhecer a diferença entre as duas séries. Porque considera as amostras independentes entre si. Mas ao se considerar a posição espacial relativa de cada amostra, pode-se distinguir as duas séries.

Construindo um variograma.

Variância espacial para 1 intervalo de amostragem:

A= [(1-7)2+(7-3)2+(3-6) 2+(6-2)2+(2-9)2+(9-4)2+(4-8)2+(8-5)2]/8

A = 22

B= [(1-3)2+(3-5)2+(5-7)+(7-9)2+(9-8)2+(8-6)2+(6-4)2 +(4-2)2]/8

B= 3,63

Variância espacial para 2 intervalo de amostragem:

A: [(1-3) 2+(3-2) 2+(2-4)+(4-5) 2+(7-6) 2+(6-9) 2+(9-8) 2]/7

A=3

B: [(1-5) 2+(5-9) 2+(9-6)+(6-2) 2+(3-7) 2+(7-8) 2+(8-4) 2]/7

B= 12,86

TABELA 1: Calculando-se a variância espacial até 4 intervalos de amostragem tem-se:

Intervalo de amostragem	Variância espacial A	Variância espacial B
1	22,0	3,63
2	3,0	12,86
3	23,67	23,83
4	3,80	29,60

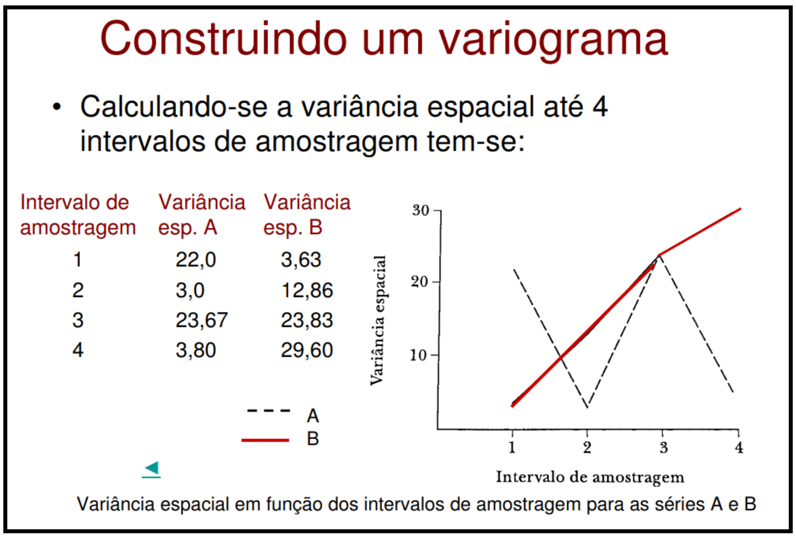

Gráfico 5: Variograma gerado a partir dos dados anteriores

Pode-se observar através desse exemplo didático que o gráfico de B, se aproxima mais do gráfico do variograma, visto anteriormente na figura 2. Já o mesmo não ocorre com o gráfico de A, que apresenta valores erráticos, e nenhuma similaridade com a figura 2, que representa a função de um variograma esperado.

RESULTADOS E DISCUSSÃO

O estudo do variograma nos ajuda a entender o alcance da geoestatistica , e também suas limitações quando variograma é errático, e quando isso acontece não se pode usar a geoestatistica. Se o variograma possui muitos valores erráticos, isto mostra que os dados não possuem correlação espacial, portanto deixam de serem variáveis regionalizadas.

De outra maneira pode-se dizer que o variograma é uma ferramaneta básica de suporte a geoestatistica

Ao ser contatado que se tem variáveis regionalizadas, o variograma poderá ser feito e servirá de base para a Krigagem, um tema a ser explorado em um próximo artigo.

CONSIDERAÇÕES FINAIS

Posteriormente, depois de realizada a análise geoestatística, o modelo produzido da variabilidade espacial dos dados, será usado na krigagem. Isso produzirá mapas ,mais precisos.

REFERÊNCIAS

YAMAMOTO, J.K.; LANDIM, P.M.B. Geoestatística: conceitos e aplicações. São Paulo: Oficina de textos, 2013. 215p.

LANDIM P.M.B. 2006. Sobre Geoestatística e mapas. Terræ Didática, 2(1):19-33

ISAAKS; SRIVASTAVA. An Introduction to Applied GEOostatistics. New York: Oxford University Press, 1989.

CAMARGO, Eduardo Celso Gerbi. Geoestatistica Fundamentos e aplicações: Geoprocessamento para Projetos Ambientais, Apresentado nos congressos GIS Brasil (96,97 e 98) e Simpósio Brasileiro de Sensoriamento Remoto (96 e 98).

SOUSA, António Jorge. Análise Geoestatistica de Dados, CVRM / Centro de Geo-sistemas. Universidade do Algarve. Portugal, 2000