

**UNIVERSIDADE FEDERAL DA PARAÍBA  
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA  
DEPARTAMENTO DE ESTATÍSTICA**

RELATÓRIO FINAL DO

PROJETO DE PESQUISA:

**ANÁLISE ESPACIAL EM SISTEMAS DE  
INFORMAÇÃO GEOGRÁFICA**

ORIENTADOR: JOÃO AGNALDO DO NASCIMENTO

---

JOÃO AGNALDO DO NASCIMENTO

BOLSISTA: MÔNICA MARIA FERREIRA TELES

---

MÔNICA MARIA FERREIRA TELES

ORGÃO: PIBIC/CNPq/UFPB  
PERÍODO: 08/2000 a 07/2001

## Índice

1 – Introdução .....	1
2 – Objetivos .....	1
3 – Bases teóricas .....	2
3.1 – Introdução ao estudo de probabilidade .....	2
3.1.1 – Frequência relativa experimental.....	3
3.1.2 – Base matemática para experiências aleatórias.....	4
3.1.3 – Probabilidade condicional .....	4
3.1.4 – Independência estocástica.....	5
3.1.5 – Ensaios de Bernoulli.....	5
3.2 – Cadeias de Markov .....	9
3.3 – Análise dos dados .....	19
3.3.1 – Análise exploratória dos dados .....	19
3.3.2 – Análise exploratória espacial.....	20
3.3.3 – Variograma .....	20
3.3.4 – Anisotropia.....	27
3.3.5 – Krigeagem.....	28
4 – Metodologia .....	32
4.1. – Obtenção e organização dos dados .....	32
4.2. – Nomenclatura das variáveis.....	33
4.3. – Montagem do banco de dados .....	36
4.3.1 – Criando um banco de dados.....	36
4.3.2 – Criação do projeto.....	37
4.3.3 – Criação da categorias, cadastral, objeto e MNT .....	37

4.3.4 – Definição dos atributos da categoria objeto.....	38
4.3.5 – Importação dos dados.....	39
I – Importação dos arquivos de linhas PB_L2D. ....	39
II– Visualizando linhas importadas .....	40
III – Filtragem de linhas duplicadas .....	41
IV – Como criar a topologia do mapa .....	41
V – Importando arquivo de identificadores .....	41
VI – Visualizando os objetos municipais .....	42
VII – Importação de tabelas .....	42
VIII – Ligação entre tabelas. ....	43
4.4 – Estatísticas descritivas dos dados.....	43
4.5 – Geração do Variograma.....	48
4.6 – Aplicação da Krigeagem.....	53
5 – Considerações finais.....	57
– Bibliografia	
– Anexos	

## **Resumo**

Este é o relatório final de andamento do Projeto de Pesquisa – PIBIC – Análise espacial em Sistemas de Informação Geográfica. Na parte inicial deste relatório temos um estudo introdutório sobre probabilidades, onde discutimos a probabilidade condicional, a independência estocástica e os ensaios de Bernoulli. Em seguida fizemos um estudo sobre Cadeias de Markov e suas matrizes de transição. Estudamos os princípios e aplicação de análise exploratória espacial com análise de tendências e a introdução à representação do fenômeno através do variograma, como também estudamos de uma forma introdutória a técnica de interpolação denominada de Krigeagem. Explicamos como foi feita a coleta dos dados para a montagem do banco de dados, bem como os passos detalhados para a montagem do mesmo no sistema SPRING. Detalhamos as etapas realizadas para a obtenção da estatística descritiva dos dados, bem como a geração dos variogramas para as variáveis do banco de dados. Por fim, aplicamos a técnica da Krigeagem com a finalidade de aprendizado em duas das variáveis, gerando em seguida mapas com a variância da distribuição dos dados.

## **1 – Introdução**

O Sistemas de Informação Geográfica (SIG) são ferramentas computacionais que permitem realizar análises complexas, integrando dados de diversas fontes e indexando-os espacialmente (SMITH et alli., 1987). Os tipos de dados são referenciados geograficamente e associados à imagens ou mapas para compor planos de informação. Cada plano individualmente pode registrar espacialmente informações sobre uma variável, por exemplo: o número de habitantes das cidades, curvas de altitude, índices de produção agrícola, mapas de vegetação, etc. A própria imagem ou mapa passa a ser um plano de informação. Esses verdadeiros “bancos de dados estatísticos” podem compor espacialmente gráficos multi-dimensionais que não poderiam ser compostos de outra maneira.

Sobre este prisma, estamos montando um banco de dados semelhante ao já existente no Departamento de Estatística (TELES, 1999). Este banco de dados já existente, denominado de PB, está montado sobre o sistema SPRING (INPE, 1998) e contém um mapa cadastral do Estado da Paraíba, onde se encontram indexados diversos dados estatísticos obtidos do Censo Demográfico Brasileiro de 1991 (IBGE, 1991). Ao novo banco de dados serão introduzidas informações colhidas do Sistema DATASUS do Ministério da Saúde do Brasil (atualizado até 1999) associadas às respectivas cidades. Por mapa cadastral entende-se a associação de atributos não gráficos – os dados – a um atributo gráfico – um mapa político do estado.

A partir desses dados sobre a saúde do estado, pretendemos gerar alguns tipos de análises exploratórias espaciais, análise de tendências e variogramas, todas de caráter introdutório. Para isso estamos estudando os Processos Estocásticos (WEISS & YOSELOFF, 1978; CLARKE & DISNEY, 1979) por se tratarem de estudos indispensáveis, tanto para a compreensão, quanto para a realização das análises exploratórias espaciais.

## **2 - Objetivos**

Este projeto visa a confecção de um banco de dados semelhante ao já existente, denominado de PB, onde já se encontram disponíveis diversos dados sobre o estado da Paraíba. Para isso estamos seguindo as etapas de confecção de um banco de dados no Sistema SPRING. Procurando, com isso,

aperfeiçoar estes procedimentos e tentando a identificação de formas mais simples para montagem de um banco de dados

Com o novo banco de dados implementado, pretendemos aprofundar nossos estudos com a realização de análise exploratória espacial, que são técnicas descritivas que visam a descrição exploratória do fenômeno, além da análise da estrutura das dependências espaciais do Processo Estocástico através do variograma.

### **3 – Bases teóricas**

#### **3.1 – Introdução ao estudo de probabilidade**

O estudo da probabilidade diz respeito à observação e análise dos possíveis resultados de experiências aleatórias. (WEISS & YOSELOFF, 1978). Deste modo, tomando como exemplo uma eleição para diretor de uma grande escola, onde exista apenas um candidato concorrendo, não caberia consideração probabilística neste caso, pelo fato de que, só poderá haver um resultado para a eleição. Neste caso, a experiência (eleição) não poderá ser considerada como uma experiência aleatória.

Podemos dizer que uma experiência é aleatória, quando não podemos saber, com certeza, o seu resultado antes que a experiência ocorra. Quando, numa experiência, sabe-se que ocorrerá sempre um certo resultado, pode-se afirmar certamente que não existirá a incerteza ou a causalidade. Neste caso, logicamente, não haverá nenhum interesse quanto a verificação da probabilidade de resultados possíveis.

Podemos utilizar o mesmo exemplo da eleição para diretor da escola para ilustrar a experiência aleatória, só que desta vez o processo eleitoral terá quatro candidatos. Com esse número de candidatos, não se poderá ter certeza do resultado até que a eleição seja realizada. Essa experiência é considerada aleatória e possui quatro possíveis resultados.

Ao estudar experiências aleatórias é comum atribuir probabilidades aos vários possíveis resultados da experiência. Em muitos casos essa atribuição é claramente definida pelo contexto. No caso da eleição onde concorrerão quatro candidatos, poderíamos dizer que o contexto foi definido pela

atribuição do mesmo peso para os votos dos diferentes segmentos da escola (professores, funcionários e estudantes). Neste caso, implica dizer que a vitória de cada um dos quatro candidatos seria igualmente provável, ou seja, deve ser atribuído a cada candidato, uma mesma probabilidade (1/4 cada).

### 3.1.1 – Frequência relativa experimental

Existe um outro meio de se atribuir probabilidades que utiliza a frequência relativa experimental. Para explicarmos este tipo de atribuição de probabilidades, consideremos um fabricante de lâmpadas elétricas que anuncie que a probabilidade de que elas queimem em menos de um mês é de  $0,1 = 1/10$ . Essa afirmação quer dizer que, em média, uma em cada dez lâmpadas, ou 10% das lâmpadas queimarão em menos de um mês. Neste exemplo a experiência aleatória consiste em examinar uma lâmpada produzida por este fabricante e o evento ( $E$ ) é a lâmpada queimar-se em menos de um mês.

De acordo com a informação divulgada pelo fabricante de que a probabilidade de uma de suas lâmpadas queimem em menos de um mês seja  $0,1 = 1/10$ , podemos atribuir probabilidade por meio da frequência relativa experimental.

$$f_E(n) = \frac{n_E}{n}$$

Frequência relativa de ocorrência de  $E$  em  $n$  realizações da experiência aleatória

Onde,  $n_E$  = número de vezes que  $E$  ocorre

$n$  = número de realizações

Assim, podemos dizer que, a frequência relativa de ocorrência de  $E$  após um grande número de realizações da experiência aleatória, passa a ser a probabilidade de ocorrência de  $E$  em uma dada realização da experiência aleatória, isto é:

$$f_E(n) = p$$

### 3.1.2 – Base matemática para experiências aleatórias

Quando a experiência aleatória é muito simples não há necessidade de uma formulação matemática complexa para sua aplicação. Mas, algumas experiências aleatórias mais complexas que ocorrem, naturalmente necessitam de um modelo matemático para análise de sua estrutura.

Utilizando o exemplo de uma experiência aleatória, que consiste em jogar três vezes uma moeda perfeita. O primeiro procedimento a ser feito para análise desta experiência é determinar o conjunto de todos os possíveis resultados da experiência, ou seja:

$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$ , onde, a letra H representa o resultado cara e a letra T representa o resultado coroa. Este conjunto, representado por  $\zeta$ , é o conjunto dos possíveis resultados da experiência e é chamado *espaço amostral* da experiência aleatória. Cada possível resultado individual, ou seja, cada elemento do espaço amostral são chamados de *pontos amostrais*, e são representados pelo símbolo  $\omega$ . Por exemplo, HHT e THH são pontos amostrais ( $\omega$ ) pertencentes ao espaço amostral ( $\Omega$ ).

### 3.1.3 – Probabilidade condicional

A probabilidade atribuída à ocorrência de um evento deverá mudar se houver uma informação adicional sobre este evento. Em alguns casos, é bastante simples resolver questões que envolvem probabilidade condicional sem uma definição matemática complexa. Por exemplo, suponha que um dado perfeito seja jogado. Essa experiência possui seis resultados possíveis, sejam eles:  $\{1, 2, 3, 4, 5, 6\}$  considerados como o espaço amostral e representado por  $\Omega$ . A cada ponto amostral podemos atribuir a probabilidade de  $1/6$ . Então, por exemplo, a  $\Pr(\{6\}) = 1/6$ .

Agora se for adicionada a informação de que o número de pontos obtidos é ímpar, qual seria a probabilidade de que o número de pontos obtidos seja 5? Neste caso, a partir da informação fornecida, o espaço amostral se altera para  $\Omega = \{1, 3, 5\}$  e a probabilidade de ocorrer um cinco passa a ser  $1/3$ . Este exemplo ilustra o fato de que informações adicionais sobre uma experiência aleatória podem afetar as probabilidades.

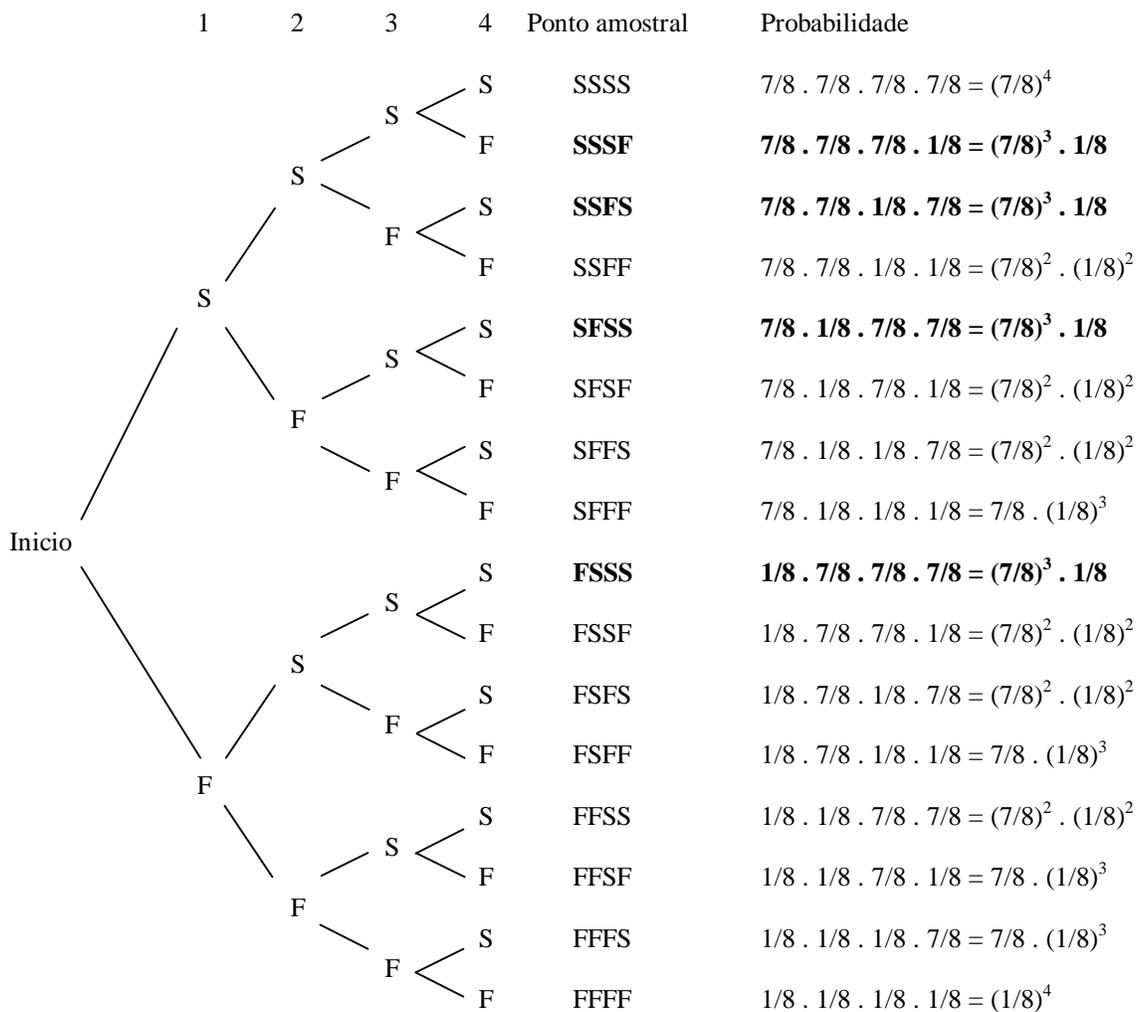
#### 3.1.4 – Independência estocástica e o princípio da multiplicação

Nesta parte trataremos um pouco sobre a independência dos resultados das experiências aleatórias. Para ilustrar a nossa explicação, vamos utilizar o exemplo da moeda sendo lançada duas vezes. Neste caso acreditamos que o evento de dar cara no segundo lançamento é “independente” do evento de dar cara no primeiro lançamento. Mas, o evento de dar cara nos dois lançamentos depende do resultado ser cara no primeiro lançamento.

“Mais precisamente, se E e F são eventos,  $\Pr(E/F)$ , em geral, difere de  $\Pr(E)$ . Intuitivamente, se E é ‘independente’ de F, o fato de que F tenha ocorrido não deve ter qualquer influência na probabilidade da ocorrência de E” (WEISS & YOSELOFF, 1978). Pode-se dizer que E é (estocasticamente) independente de F se,  $\Pr(E/F)$ , ou seja, a probabilidade de E, tal que F ocorra, é igual a probabilidade de E.

#### 3.1.5 – Ensaio de Bernoulli: a distribuição binomial

À seqüência de ensaios independentes, onde cada ensaio admite apenas dois resultados possíveis, sejam eles S (sucesso) e F (fracasso), damos o nome de Ensaio de Bernoulli, cujo interesse principal é determinar a probabilidade associada com o número de sucessos obtidos.



Fonte: ( WEISS & YOSELOFF, 1978).

Figura 1 – seqüência de ensaios com S sucessos e F fracassos

Para explicar estes ensaios iremos utilizar o exemplo dos parafusos defeituosos. Levemos em consideração a experiência de observar quatro parafusos no momento de sua fabricação e observar se são defeituosos, ou não (segundo algumas especificações de tolerância). É conhecido que a probabilidade de a máquina produzir um parafuso defeituoso é de 1/8. Qual a probabilidade de que três entre os quatro parafusos não sejam defeituosos?

Para resolver esta questão consideraremos a experiência aleatória como uma seqüência de quatro ensaios de Bernoulli, onde o S (sucesso) é o parafuso não ser defeituoso. Portanto o espaço

amostral consiste em 16 pontos amostrais, todos eles representados por seqüências de S (sucessos) e F (fracassos) de tamanho quatro, que estão ilustrados na figura 1.

Consideremos  $A$  como o evento de três dos quatro parafusos não serem defeituosos. A seguir observemos  $\text{card}(A)$ , ou seja, quantos pontos amostrais estão contidos em  $A$ , levando em consideração que um ponto amostral só pertencerá a  $A$  se for composto de 3 S (sucessos) e 1 F (fracasso). Ou seja, dos quatro parafusos observados, três serão perfeitos e um defeituoso. Neste caso, observando na figura 1 os pontos amostrais em negrito, podemos afirmar que  $A$  contém quatro pontos amostrais (SSSF, SSFS, SFSS e FSSS), ou seja:

$$\binom{4}{3} = 4$$

Onde cada ponto amostral possui a mesma probabilidade, sendo ela

$$\left(\frac{7}{8}\right)^3 \cdot \frac{1}{8}$$

Então,

$$\Pr(A) = \text{card}(A) \cdot \left(\frac{7}{8}\right)^3 \cdot \frac{1}{8} = \binom{4}{3} \cdot \left(\frac{7}{8}\right)^3 \cdot \frac{1}{8}$$

Para exemplo geral, a  $\Pr(A)$  pode ser escrita da seguinte forma:

$$\Pr(A) = \binom{4}{3} \cdot \left(\frac{7}{8}\right)^3 \cdot \left(1 - \frac{7}{8}\right)^{4-3}$$

Desta forma o resultado fundamental relativo aos Ensaio de Bernoulli demonstra que “O espaço amostral consiste em  $2^n$  pontos amostrais, todos eles seqüências de  $S$ s e  $F$ s de comprimento  $n$ . Seja  $A$  o evento de haver exatamente  $k$  sucessos nos  $n$  ensaios. Primeiramente calculamos  $\text{card}(A)$  (quantos pontos amostrais  $A$  contém). Este número é igual ao número de seqüências de  $S$ s e  $F$ s, de

comprimento  $n$ , com exatamente  $k$   $S$ s (e portanto  $n - k$   $F$ s). Isso, como se sabe, é o número de possíveis maneiras de selecionar  $k$  posições na seqüência de comprimento  $n$ , para colocar os  $S$ s.” (WEISS & YOSELOFF, 1978). Então, a probabilidade de exatamente  $k$  sucessos ( $0 \leq K \leq n$ ) em  $n$  ensaios é:

$$\binom{n}{k} p^k (1-p)^{n-k}$$

### 3.2 - Cadeias de Markov

Vamos definir um processo estocástico como uma seqüência de variáveis aleatórias, digamos,  $\{X_n, n = 0, 1, 2, \dots\}$ , ou  $\{X_n, n \geq 0\}$ . Cada variável aleatória da seqüência  $X_n$  terá espaço amostral igual ao conjunto dos naturais. Se,  $X_n = i$ , diremos que, no tempo  $n$ , o processo está no estado  $i$ . A passagem do estado  $i$  para o estado  $j$  ocorrerá com probabilidade  $P_{ij}$  cujo significado é uma probabilidade condicional de estar no estado  $j$  dado que estava no estado  $i$ .

Muitos processos ao mudar de estado são influenciados apenas pela última mudança, por exemplo, uma pessoa alcoólatra tem períodos de lucidez e embriaguez e, a chance dele estar lúcido depende apenas do que fez ultimamente. Um aluno que estuda, dia sim e dia não, em uma prova surpresa realizada com o assunto da aula anterior, obterá um resultado que depende de ter estudado na véspera. Ou mais simples, um aluno que comparece, aula sim e aula não, terá sua presença na aula de amanhã dependendo apenas de sua presença na aula de hoje. A variável aleatória  $X_n$  representa os possíveis estados deste processo no  $n$ -ésimo dia de aula: a presença de  $X_n = 1$  ou ausência de  $X_n = 0$ . Os possíveis estados de  $X_n$  são 0 ou 1, zero para ausência e 1 para presença.

A propriedade de que o resultado que irá um período à frente depende apenas do período anterior é comum em muitos fenômenos ou processos. Por isso, existe o interesse em estudar os processos que possuem esta propriedade. Nas Ciências Sociais, Biológicas e Tecnológicas, dentre outras, existem problemas que podem ser modelados por processos que possuem tal propriedade. Esta propriedade é conhecida como propriedade de Markov e o processo que tem tal propriedade de Processo Markoviano. Podemos escrever esta propriedade como:

$$P\{X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_1 = i_1, X_0 = i_0\} = P\{X_{n+1} = j | X_n = i\}. \quad (2)$$

para todo estado  $i_0, i_1, \dots, i_{n-1}, i, j$  e para todo  $N > 0$ . Note que neste capítulo as variáveis aleatórias  $X_n$  do processo estocástico são do tipo discreta e, como consequência, o espaço de estado será discreto finito ou infinito enumerável (por exemplo o conjunto dos Naturais). Observe que estaremos denotando a probabilidade de passar do estado  $i$  para o estado  $j$  por  $P_{ij}$  denominada de *probabilidade de transição*. O processo que no período  $n$  está no estado  $i$  ( $X_n = i$ ) e passa no período seguinte para o estado  $j$  tem a probabilidade de ocorrer esta passagem denotada por:

$$P[X_{n+1} = j | X_n = i] = P_{ij}.$$

**Definição:** Uma cadeia de Markov é um processo em que a probabilidade de estar em um certo estado em um tempo futuro, pode depender do estado do sistema no período atual, mas não dos estados dos períodos passados. Em uma cadeia de Markov, dado o estado atual do sistema, o próximo estado é independente do passado embora possa depender do presente.

**Exemplo 1:** Considere o sorteio com uma moeda. Após lançar esta moeda por dez vezes e anotar o resultado (estado) Cara = C e Coroa = O, obteve-se:

Lançamento	1	2	3	4	5	6	7	8	9	10
Estado	C	O	C	C	O	O	C	O	C	C

**Exemplo 2:** Mobilidade social. Suponha que o estado social ou financeiro de uma pessoa possa ser classificado em estados ou classes, 1 = Baixa, 2 = Média e 3 = Alta. Investigando-se a genealogia dos primogênitos de um indivíduo e, denominando-se a geração atual como pai, a anterior como avô e a futura como filho e, admitindo-se que a classe atual dependerá da classe da geração anterior, apresenta-se a seguir o resultado da investigação da árvore genealógica do indivíduo pesquisado:

Geração	0	1	2	3	4	5	6	7
Estado	2	1	3	2	3	2	2	1

Observe que  $P_{ij} \geq 0$  e que,

$$\sum_{j=1}^n P_{ij} = 1, \quad i = 0, 1, \dots$$

Vamos denotar por  $P$  a matriz de transição um passo ou período à frente, então:

$$P = \begin{bmatrix} P_{00} & P_{01} & P_{02} & \dots \\ P_{10} & P_{11} & P_{12} & \dots \\ \vdots & \vdots & \vdots & \vdots \\ P_{i0} & P_{i1} & P_{i2} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \quad (3)$$

Para o exemplo 2, a matriz  $P$  ou as probabilidades de transição foram as seguintes:

		<i>Filho</i>		
		<b>1</b>	<b>2</b>	<b>3</b>
<i>Pai</i>	<b>1</b>	0,6	0,3	0,1
	<b>2</b>	0,1	0,8	0,1
	<b>3</b>	0,1	0,2	0,7

isto é: o elemento da  $i$ -ésima linha e  $j$ -ésima coluna, o elemento  $ij$  fornece a probabilidade de transição do estado  $i$  para o estado  $j$  em uma geração e será denotada por  $P_{ij}$ . Note que  $P_{33} = 0,7$ , isto é: a probabilidade de que o filho ( $X_{n+1}$ ) seja da classe social alta dado que o pai ( $X_n$ ) seja de classe social alta é 0,7. Esta probabilidade pode ser escrita como:  $P[X_{n+1} = 3 | X_n = 3] = 0,7$ . Nota que a soma das linhas de  $P$  é igual a 1, pois a linha representa uma distribuição de probabilidade para o estado.

**Caminho aleatório:** Uma cadeia de Markov cujo espaço de estado pode ser os inteiros  $i = 0, 61, 62, \dots$  é um caminho aleatório se, para um número  $P$ ,  $0 < P < 1$ ,

$$P_{i,i+1} = P = 1 - P_{i,i-1}, \quad i = 0, 61, 62, \dots \quad (4)$$

é um modelo estocástico em que podemos associar a caminhada de um indivíduo sobre uma reta, dando um passo para a direita com probabilidade  $p$  e, um passo para a esquerda com probabilidade  $1-p$ .

**Um modelo de jogo com aposta entre duas pessoas:** O jogo é o seguinte: em cada jogada uma pessoa ganha ou perde R\$ 1,00 (um real). Ganha com probabilidade  $p$  e perde com probabilidade  $1-p$ . O jogo terminará quando uma das duas pessoas perder todo seu dinheiro (a ruína do jogador), ou

quando uma delas alcançar a quantia de R\$ N, 00 (N reais). Ao longo das jogadas haverá ganho e perdas e a quantia que um jogador terá em dado momento é uma variável aleatória, que possui a propriedade de uma cadeia de Markov, com probabilidades de transição dadas por:

$$P_{i, i+1} = p = 1 - P_{i, i-1}, i = 1, 2, \dots, N-1, \text{ e } P_{00} = 1. \quad (5)$$

Observe que a probabilidade de passar do estado 0 para o 0,  $P_{00} = 1$ , isto significa que uma vez no estado 0 não haverá chance de sair dele. O mesmo ocorre com o estado N, uma vez entrando nele infinitas vezes haverá retorno para sempre nele! Por esta razão, os estados 0 e N são chamados de *Estados absorventes*.

**Equações de Chapman-Kolmogorov:** Já conhecemos as probabilidades de transição  $P_{ij}$  para uma passo ou período à frente. Vamos conhecer agora as probabilidades de transição após  $n$  passos (períodos ou tempos) denotada por  $P^n_{ij}$  que é a probabilidade do processo no estado  $i$  chegar ao estado  $j$  após  $n$  realizações (ou após  $n$  períodos, após  $n$  jogadas, após  $n$  gerações). Esta probabilidade é escrita em termos de probabilidade condicional da seguinte maneira:

$$P^n_{ij} = P[X_{n+m} = j / X_m = i], \quad (6)$$

para todo  $n, m \geq 0$  e  $i, j \geq 0$ . Observe que,  $P^l_{ij} P_{ij}$ . As equações de Chapman-Kolmogorov permitem calcular as probabilidades de transição para  $n$  passos e são dadas por,

$$P^n_{ij} = \sum_{k=0}^{\infty} P^k_{ik} P^{n-k}_{kj}, \quad (7)$$

para todo  $n \geq 0$  e  $i, j \geq 0$ .

**Prova:** utilizando (6) com  $n + m$  no lugar de  $n$  temos que,

$$\begin{aligned}
 P_{ij}^{n+m} &= P [ X_{n+m} = j / X_0 = i ] \\
 &= \sum_{k=0}^{\infty} P [ X_{n+m} = j, X_n = k / X_0 = i ] \\
 &= \sum_{k=0}^{\infty} P [ X_{n+m} = j / X_n = k, X_0 = i ] P [ X_n = k / X_0 = i ] \\
 &= \sum_{k=0}^{\infty} P_{ik}^n P_{kj}^m
 \end{aligned}$$

Na forma matricial, a equação (7) pode ser escrita como:

$$P^{(n+m)} = P^{(n)} \cdot P^{(m)}, \quad (8)$$

para todo  $n, m \geq 0$ . Para  $m = n = 1$  temos que  $P^{(2)} = P^{(1+1)} = PP = P^{(2)}$  e por indução chegamos a:

$$P^{(n)} = P^{(n-1+1)} = P^{n-1} \cdot P = P^n,$$

ou seja: a matriz de transição em  $n$  passos é obtida multiplicando a matriz  $P$   $n$  vezes, ou  $P$  elevado à potência  $n$ .

**Exemplo 3: Previsão do tempo** – Suponha que choverá, ou não, amanhã, dependendo das condições climáticas do dia anterior (hoje). Suponha que se chove hoje, então choverá amanhã com probabilidade  $\alpha$ . Se não chover hoje, então choverá amanhã com probabilidade  $\beta$ . Denote o *estado 0* quando chove e, *estado 1*, quando não chove. Então temos uma cadeia de Markov com dois estados e matriz de transição  $P$  dada por:

		<i>Amanhã</i>	
		<b>0</b>	<b>1</b>
<i>Hoje</i>	<b>0</b>	$\alpha$	$1 - \alpha$
	<b>1</b>	$\beta$	$1 - \beta$

$$= \begin{bmatrix} \alpha & 1-\alpha \\ \beta & 1-\beta \end{bmatrix} = P.$$

se  $\alpha = 0,7$  e  $\beta = 0,4$  qual é a probabilidade de chover daqui a quatro dias, ou seja, calcule  $P[X_4 = 0 / X_0 = 0]$ . Note-se que o dia atual é representado por  $X_0$ . Observe que a matriz de transição para um período é igual a:

$$P = \begin{bmatrix} 0,7 & 0,3 \\ 0,4 & 0,6 \end{bmatrix}$$

Então

$$P^2 = P^2 = \begin{bmatrix} 0,7 & 0,3 \\ 0,4 & 0,6 \end{bmatrix} \cdot \begin{bmatrix} 0,7 & 0,3 \\ 0,4 & 0,6 \end{bmatrix} = \begin{bmatrix} 0,61 & 0,39 \\ 0,0,52 & 0,48 \end{bmatrix}$$

Podemos calcular

$$P^{(4)} = (P^2)^2 = \begin{bmatrix} 0,61 & 0,39 \\ 0,0,52 & 0,48 \end{bmatrix} \cdot \begin{bmatrix} 0,61 & 0,39 \\ 0,0,52 & 0,48 \end{bmatrix} = \begin{bmatrix} 0,5749 & 0,4251 \\ 0,5668 & 0,4332 \end{bmatrix}$$

Portanto,

$$P[X_4 = 0 / X_0 = 0] = P_{00}^4 = 0,5749.$$

As probabilidades consideradas até aqui foram probabilidades condicionais. As  $P_{ij}^n$  representam probabilidades de que no período  $n$  o processo esteja no estado  $j$  dado que no período inicial estava no estado  $i$ . Como faremos para determinar as probabilidades não condicionais no período  $n$ , digamos  $P[X_n = j]$ ? É necessário especificar as probabilidades no período inicial, isto é: especificar a distribuição de probabilidade do estado inicial. Vamos denotar

$$\alpha_i \equiv P[X_0 = i], i \geq 0, \quad \text{onde} \quad \sum_{i=1}^{\infty} \alpha_i = 1.$$

Agora podemos expressar todas as probabilidades condicionando sobre o estado inicial, isto é:

$$\begin{aligned}
 P[X_n = j] &= \sum_{i=1}^{\infty} P[X_n = j / X_0 = i] P[X_0 = i] \\
 &= \sum_{i=1}^{\infty} P_{ij}^n \alpha_i.
 \end{aligned} \tag{9}$$

**Exemplo 4:** Voltando ao exemplo 3. Informamos agora a distribuição inicial de  $X_0$ . Isto é:  $P[X_0 = 0] = \alpha_0 = 0,4$  e  $P[X_0 = 1] = \alpha_1 = 0,6$  (a soma)  $P[X_0 = 0] + P[X_0 = 1] = \alpha_0 + \alpha_1 = 1$ . Podemos calcular a probabilidade de chover nos próximos quatro dias (chover quatro dias adiante). Esta probabilidade é  $P[X_4 = 0]$  e seu cálculo é realizado com a aplicação da equação (9), ou seja:

$$P[X_4 = 0] = 0,4P_{00}^4 + 0,6P_{10}^4 = 0,4(0,5749) + 0,6(0,5668) = 0,5700.$$

### Classificação dos estados de uma Cadeia de Markov

Um estado  $j$  é *acessível* ao estado  $i$  se  $P_{ij}^n > 0$  para algum  $n \in \mathbb{N}$ . Note que  $j$  é acessível a  $i \Leftrightarrow$  partindo de  $i$  é possível chegar ao estado  $j$ . Vamos provar este resultado.

**Resultado.** Um estado  $j$  é acessível ao estado  $i$  se  $P_{ij}^n > 0$  para algum  $n \in \mathbb{N}$ .

**Prova.** Suponha que o espaço de estado seja o conjunto dos naturais. O estado  $j$  é acessível a  $i$  se, partindo de  $i$  atingir  $j$  ao menos uma vez. A probabilidade de atingir  $j$  ao menos uma vez partindo de  $i$  é dada por:  $P[\text{chegar a } j \text{ ao menos uma vez} / \text{estava em } i]$

$$\begin{aligned}
 &= P\left[\bigcup_{n=0}^{\infty} \{X_n = j\} / X_0 = i\right] \\
 &\leq \sum_{n=0}^{\infty} P[X_n = j / X_0 = i] \\
 &= \sum_{n=0}^{\infty} P_{ij}^n
 \end{aligned} \tag{10}$$

O lado direito de (10) é uma soma de números reais positivos e será não nula se  $P_{ij}^n > 0$ , enquanto que o lado esquerdo é a afirmativa de que o estado  $j$  se comunica com o estado  $i$ . Portanto,  $j$  é acessível a  $i$   $\Leftrightarrow P_{ij}^n > 0$ . Note que se  $j$  se comunica com  $i$  então o lado esquerdo de (10) é positivo força que alguma parcela do lado direito seja positivo e neste caso existe algum  $n \in \mathbb{N}$ , tal que  $P_{ij}^n > 0$ . Por outro lado, se  $P_{ij}^n > 0$ , então

$$\sum_{n=0}^{\infty} P_{ij}^n > 0$$

e implica que a probabilidade do lado esquerdo seja positiva e conclui-se que  $j$  se comunica com  $i$ . Fica provado o resultado.

**Definição.** Diremos que os estados  $i$  e  $j$  são comunicantes se  $j$  é acessível a  $i$  e  $i$  é acessível a  $j$ . Denotaremos que  $i$  e  $j$  são comunicantes por  $i \leftrightarrow j$  e que  $j$  é acessível a  $i$  por  $i \rightarrow j$ .

**Resultado:** A relação de comunicação entre os estados satisfaz às seguintes propriedades:

1. O estado  $i$  se comunica com ele mesmo.
2. Se  $i$  se comunica com  $j$  então  $j$  se comunica com  $i$ .
3. Se o estado  $i$  se comunica com o estado  $j$  e, o estado  $j$  se comunica com o estado  $k$ , então o estado  $i$  se comunica com o estado  $k$ .

**Prova.** As probabilidades 1 e 2 seguem da definição de comunicação dada acima. Para provar 3, suponha que  $i$  se comunica com  $j$  e  $j$  se comunica com  $k$ , então por definição existem inteiros  $n$  e  $m$ , tal que  $P_{jk}^m > 0$ . Utilizando a equação de Chapman-Kolmogorov, podemos escrever que:

$$P_{ik}^{n+m} = \sum_{r=0}^{\infty} P_{ir}^n P_{rk}^m \geq P_{ij}^n P_{jk}^m > 0.$$

Portanto o estado  $i$  é acessível ao estado  $k$ , pois existe um inteiro  $m + n$ , tal que,

$$P_{ik}^{n+m} > 0$$

O resultado acima mostra que os estados comunicantes formam uma classe de equivalência (possuem as propriedades: 1-Reflexiva, 2-Simétrica e 3-Transitiva). O conceito de *estado comunicante* classifica os estados de uma cadeia de Markov em comunicantes ou não comunicantes. Se todos os estados de uma cadeia de Markov são comunicantes então ela é dita irredutível. Isto é, em uma cadeia de Markov *irredutível* todos os estados se comunicam.

**Exemplo 4.** Considere a cadeia de Markov de estados 0, 1, 2 e matriz de transição

$$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{3} & \frac{2}{3} \end{bmatrix}$$

Observe em  $P$  que o *estado 1* é acessível ao *estado 0* e que o *estado 2* é acessível ao *estado 1*. Com a notação utilizada podemos denotar este fato como  $0 \square 1 \square 2$ , pois o *estado 0* se vai ao 1 com probabilidade  $\frac{1}{2}$  e do *estado 1* ao 2 com probabilidade  $\frac{1}{4}$ .

**Exemplo:** Considere a cadeia de Markov com quatro estados e matriz de transição  $P$  dada por,

$$P = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Os estados  $0$  e  $1$  são comunicantes, o estado  $3$  é absorvente, pois  $P_{33} = 1$ . O estado  $2$  não é acessível aos estados  $0$  e  $1$ .

Seja  $p_i$  a probabilidade de que o processo partindo do estado  $i$  volte novamente ao estado  $i$ . O estado  $i$  é *recorrente* se  $p_i = 1$  e é *transiente* se  $p_i < 1$ . Se o processo parte de  $i$  e,  $i$  é recorrente, então com probabilidade 1 ele voltará ao estado  $i$  e, como estamos tratando com cadeias irredutíveis (aquela em que todos os estados se comunicam), logo o processo estará no estado  $i$ . E voltará a  $i$  de novo. Repetindo este argumento repetidas vezes, conclui-se que *um estado recorrente  $i$  será visitado infinitas vezes*.

Mas, se o estado  $i$  é *transiente*, então cada vez que o processo estiver em  $i$  terá uma probabilidade  $1 - p_i$  de não permanecer  $i$  (o sucesso é não permanecer em  $i$  em  $n$  ensaios independentes) e segue-se que o número de vezes que ele ficará em  $i$  após  $n$  períodos segue uma distribuição geométrica com parâmetro de sucesso igual a  $1 - p_i$ . Portanto esta probabilidade é igual a  $P_i^{n-1} (1 - p_i)$ . Conclui-se que, se  $i$  é transiente, partindo-se de  $i$ , o número de vezes que o processo estará no estado  $i$  tem distribuição geométrica com média finita  $p_i / (1 - p_i)$ . Isto quer dizer que, diferente do estado recorrente, um estado é transiente quando, uma vez estando nele, o processo voltará a ele um número finito de vezes. Vamos lembrar aqui que se  $X$  tem distribuição geométrica então  $X$  tem fdp  $f(x) = P[X = x] = (1-p)^{x-1} p$ ,  $x=0, 1, 2, \dots$  onde  $x$  é o número de vezes que experimentos independentes devem ser realizados até que o sucesso de probabilidade  $p$  seja alcançado. O valor esperado de  $X$  é dado por  $E(X) = 1/p$ . O estado transiente  $i$  permanecer  $n$  vezes em  $i$  até que saia de  $i$  realiza  $n$  experimentos independentes até atingir o sucesso dele sair de  $i$ , por isso a probabilidade de sucesso aqui é  $1 - p_i$ .

Portanto, um processo  $i$  é *recorrente*  $\Leftrightarrow$  partindo de  $i$ , o número esperado de vezes que se voltará a  $i$  é infinito. Vamos provar o seguinte resultado que esclarecerá a diferença entre estado *recorrente* e *transiente*.

**Resultado.** Um estado  $i$  é:

1. Recorrente, se

$$\sum_{n=1}^{\infty} P_{ii}^n = \infty$$

2. Transiente, se:

$$\sum_{n=1}^{\infty} P_{ii}^n < \infty$$

**Prova:** Seja

$$A_n = \begin{cases} 1 & \text{se } X_n = i \\ 0 & \text{se } X_n \neq i \end{cases}$$

Observe que

$$\sum_{n=0}^{\infty} A_n$$

é o número de vezes que o processo estará em  $i$  e seu valor esperado dado  $X_0 = i$  é dado por:

$$\begin{aligned} E\left[\sum_{n=0}^{\infty} A_n / X_0 = i\right] &= \\ &= \sum_{n=0}^{\infty} E[A_n / X_0 = i] \\ &= \sum_{n=0}^{\infty} P[X_n = i / X_0 = i] \\ &= \sum_{n=0}^{\infty} P_{ii}^n \end{aligned} \tag{11}$$

Para provar 1. Observe que o estado  $i$  sendo recorrente será visitado infinitas vezes e, portanto,  $(??)$  será uma soma de infinitos 1 e portanto igual a  $\infty$ . Para provar 2 observe que sendo o estado  $i$  transiente ele voltará a  $i$  um número finito de vezes e portanto a soma em  $(??)$  será finita isto é

$$\sum_{n=0}^{\infty} P_{ii}^n < \infty$$

### 3.3 - Análise dos dados

Devido a grande quantidade de informações disponíveis, torna-se essencial para os pesquisadores a utilização de técnicas que permitam a seleção e ordenação destas informações, para que se torne possível a manipulação e conseqüentemente a sua compreensão. Essas técnicas quantitativas permitem a redução e padronização das informações tornando-as manejáveis e interpretáveis. As possibilidades de análises mais profundas dos dados, a solução de problemas complexos e a exploração de novos campos impossíveis de serem explorados apenas com a observação de dados brutos, são possibilidades vislumbradas quando do emprego destas técnicas (GERARDI, 1981).

#### 3.3.1 - Análise exploratória dos dados

Por sua vez, “a análise exploratória dos dados é uma etapa de fundamental importância e obrigatória dentro de qualquer estudo geoestatístico. É uma etapa precursora introduzida para averiguar as propriedades estatísticas e matemáticas dos dados” (BURROUGH op cit CÂMARA, 1999)

O procedimento para sua realização passa pela coleta e organização dos dados sob a forma de tabelas que serão futuramente transformadas em matrizes. A partir daí, pode-se proceder a uma análise preliminar visual para verificar possíveis problemas como a falta de informação em alguma variável ou presença de valores muito grandes, além de outros tipos de problemas com os dados.

Após essa verificação passamos para a fase onde obtemos algumas estatísticas das amostras, como, cálculo da média, variância, desvio padrão, etc., além da obtenção de outros recursos visuais como o histograma. Um histograma trata-se de um gráfico de barras contíguas, onde a altura é proporcional à frequência e a base é constituída por um segmento cujos extremos representam os extremos da *i-ésima* classe. Com esses recursos visuais pode-se obter uma descrição das características do fenômeno, onde visualizamos se o tipo de distribuição que ocorre é assimétrica ou não entre outras características.

### 3.3.2 – Análise exploratória espacial

A análise espacial é a habilidade de se combinar dados espaciais de naturezas diversas, manipulando-os e aplicando-lhes técnicas (matemáticas, estatísticas, etc.) para se obter conclusões adicionais só evidentes com um estudo mais profundo e com uma espacialização dos dados. Segundo Marília de Sá Carvalho, “análise espacial é o estudo quantitativo de fenômenos que são localizados no espaço. Utiliza-se a expressão análise de dados espaciais em oposição à análise de dados em geral, quando as técnicas utilizadas consideram explicitamente a localização espacial” (CARVALHO, 1998).

A análise exploratória espacial permite-nos descrever e visualizar distribuições de dados espaciais, descobrir padrões de associação espacial, bem como identificar nos fenômenos, outros tipos de comportamento considerados atípicos.

### 3.3.3 – Variograma

Alguns fenômenos possuem propriedades espacialmente contínuas na superfície terrestre e alguns métodos de interpolação espacial possuem limitações na representação da variabilidade espacial, porque desconsideram a anisotropia (qualidade de um fenômeno se propagar no espaço de uma forma diferenciada) e a continuidade do fenômeno que se quer observar.

A interpolação espacial é um processo em que se determina o valor de uma função num ponto interno de um intervalo a partir dos valores da função nas fronteiras deste intervalo. Entre esses valores (amostras) situados nas fronteiras do intervalo são realizados cálculos em diversas direções (N, NE, S,

etc.) e, a partir dos resultados encontrados são deduzidos valores que indicarão uma similaridade espacial entre as amostras. É através deste procedimento que os métodos de interpolação espacial reconstruem uma superfície a partir de semelhanças encontradas entre algumas amostras existentes.

A representação quantitativa da variação de um fenômeno que possui dependência espacial é denominada de variograma. A análise da dependência espacial de um fenômeno é possível através desta representação, onde é medida a variância da diferença dos valores em todos os pontos (que apresentem uma distância constante) do espaço (CARVALHO, 1997). À medida que aumenta a distância entre os pontos, o variograma tende a igualar a variância total.

Considerando o processo estocástico  $Z(s)$ , o variograma é a variância da diferença entre a realização do processo em  $(s)$  e em  $(s+h)$ , onde  $\gamma(h)$  é o semi-variograma e  $h$  é o número de passos entre os pares

$$\text{Var}(Z(s+h) - Z(s)) = 2\gamma(h)$$

Existem diversos métodos utilizados para a modelagem de processos espacialmente distribuídos. O método clássico para estimar o variograma amostral é obtido pela equação:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{(i,j)} (v_i - v_j)^2$$

Onde:  $N$  é o número de pares de pontos;

$v$  é o valor da variável nos pontos  $i$  e  $j$

“Segundo Burrough (1987), a variação espacial de uma variável regionalizada pode ser expressa pela soma de três componentes: a) uma componente estrutural, associada a um valor médio constante ou a uma tendência constante; b) uma componente aleatória, espacialmente correlacionada; e c) um ruído aleatório ou erro residual” (CÂMARA, 1999; cap 12)

Se  $x$  representa uma posição em uma, duas ou três dimensões, então o valor da variável  $Z$ , em  $x$ , é dada por,

$$Z(x) = m(x) + \varepsilon'(x) + \varepsilon''$$

onde:

- $m(x)$  é uma função determinística que descreve a componente estrutural de  $Z$  em  $x$ ;
- $\varepsilon'(x)$  é um termo estocástico, que varia localmente e depende espacialmente de  $m(x)$ ;
- $\varepsilon''$  é um ruído aleatório não correlacionado, com distribuição normal com média zero e variância  $\sigma^2$ .

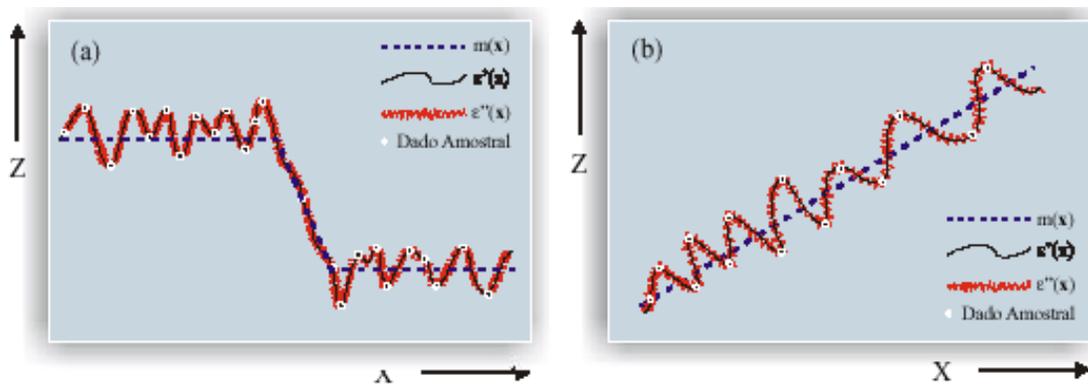


Figura 2a e 2b – Principais componentes da variação espacial

As Figuras 2a e 2b ilustram as três componentes principais da variação espacial. A Figura 2a apresenta uma componente determinística que varia abruptamente, enquanto a componente determinística na Figura 2b apresenta uma tendência constante.

Considere agora duas variáveis  $x$  e  $y$ , onde  $x = z(x)$  e  $y = z(x+h)$ . Neste caso, referem-se ao teor de zinco no solo medido em duas posições diferentes, conforme ilustra a figura 3.

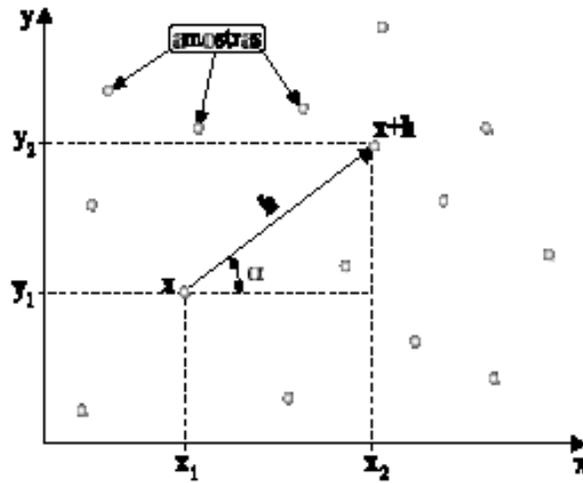


Figura 3 – Amostragem em duas dimensões

onde  $x$  denota uma posição em duas dimensões, com componentes  $(x_i, y_i)$ , e  $h$  um vetor distância (módulo e direção) que separa os pontos. O nível de dependência entre essas duas variáveis regionalizadas,  $X$  e  $Y$ , é representado pelo variograma,  $2\gamma(h)$ , o qual é definido como a esperança matemática do quadrado da diferença entre os valores de pontos no espaço, separados pelo vetor distância  $h$ , isto é,  $2\gamma(h) = E\{[Z(x)-Z(x+h)]^2\} = \text{Var}[Z(x)-Z(x+h)]$ .

Através de uma amostra  $z(x_i), i=1, 2, \dots, n$ , o variograma pode ser estimado por

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

Onde,

$$2\hat{\gamma}(h)$$

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 distância  $h$ ;

$$z(x_i) \text{ e } z(x_i+h)$$

são valores da  $i$ -ésima observação da variável regionalizada, coletados nos pontos  $x_i$  e  $x_i+h$  ( $i=1, \dots, n$ ), separados pelo vetor  $h$ .

De acordo com a afirmação que, os fenômenos localizados mais próximos têm comportamentos mais semelhantes entre si do que os fenômenos localizados mais distantes, espera-se dos padrões dos variogramas que as diferenças  $\{z(x_i)-(x_i+h)\}$  decresçam à medida que a distância ( $h$ ) que os separa diminua. Deste modo  $\gamma(h)$  deverá aumentar com a distância. Neste caso, considera-se que por se tratar de um processo contínuo, quando  $h$  tende a zero o variograma também tende a zero.

É importante destacar que o variograma aumenta com a distância até se estabilizar em um platô (*sill*) quando desaparece a dependência espacial. A autocovariância por sua vez decresce com a distância, à medida que diminui a covariância entre os pontos. Se houver uma tendência espacial em alguma direção o variograma não se estabiliza, neste caso é recomendável que se retire a tendência para que possibilite a modelagem em distâncias curtas.

Quando as distâncias são muito pequenas, procura-se dimensionar o grau de continuidade através de uma análise da variação em pequena escala. Assim, considerando-se que há uma descontinuidade de valores entre as amostras e que existem valores abaixo da frequência de amostragem, faz-se necessário incorporar essas descontinuidades de valores através de uma técnica denominada “efeito pepita” ( $C_0$ ).

Esta técnica serve para modelar as descontinuidades que se apresentam. Ou seja, quando o  $h=0$ , o valor do variograma na sua origem é realmente zero. Porém, em muitos variogramas, o  $h$  apresenta apenas uma *tendência* a ser zero, e neste caso o valor da origem *pode ser* maior que zero, apresentando esta espécie de descontinuidade na origem. Quando houver esta descontinuidade faz-se necessário, como podemos observar na figura 4, que se acrescente esta constante ( $C_0$ ) ao modelo.

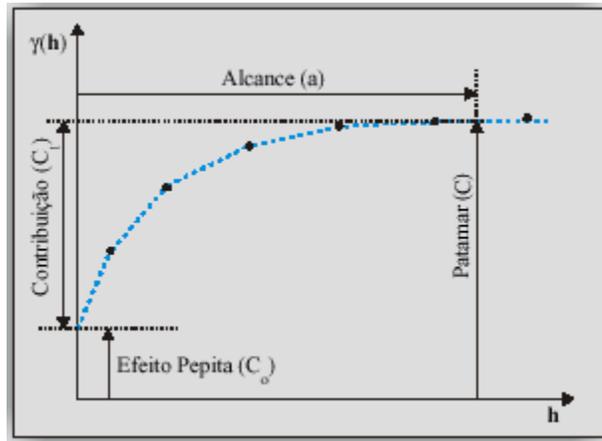


Figura 4 – Exemplo de semi variograma experimental

*Alcance (a)*: distância dentro da qual as amostras apresentam-se correlacionadas espacialmente.

*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 ( $\text{Var}[Z(x) - Z(x+h)]$ ) torna-se invariante com a distância.

*Efeito Pepita (C<sub>0</sub>)*

*Contribuição (C<sub>1</sub>)*: é a diferença entre o patamar (C) e o Efeito Pepita (C<sub>0</sub>).

O variograma  $\gamma(h)$  é formado por uma série de valores, sobre os quais se objetiva ajustar uma função que represente uma tendência de  $\gamma(h)$  em relação a  $h$ . Este ajuste não é automático, sendo necessário realizar um primeiro ajuste observando se há uma adequação do modelo utilizado. Dependendo do ajuste obtido pode-se ou não redefinir o modelo até a escolha de um considerado satisfatório.

Vamos expor aqui alguns modelos considerados básicos tomando como referência CÂMARA (1999). Esses modelos são denominados isotrópicos e estão divididos em dois tipos: modelos que apresentam patamar e modelos sem patamar. Nos modelos que possuem um patamar, seu alcance ( $a$ ) é arbitrariamente definido como a distância correspondente a 95% do patamar. Os modelos que não atingem um patamar e continuam aumentando enquanto a distância aumenta são utilizados para modelar fenômenos que tem uma capacidade infinita de dispersão. Os modelos de variogramas mais

utilizados são o Esférico, o Exponencial e o Gaussiano. Podemos observar na figura 5 a representação gráfica dos modelos citados.

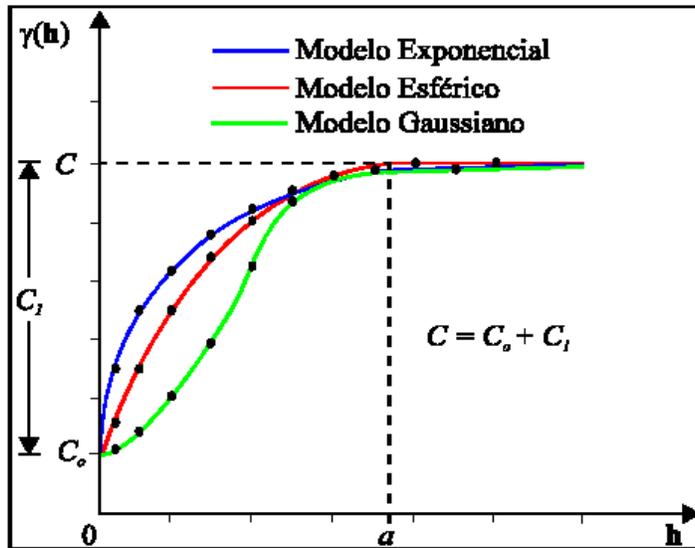


Figura 5 – Representação gráfica de semivariogramas experimentais e modelos teóricos

- O modelo esférico é um dos modelos mais utilizados. A equação normalizada deste modelo é:

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

- A equação normalizada do modelo Exponencial é:

$$Exp(h) = \begin{cases} 0 & , |h| = 0 \\ 1 - \exp\left(-\frac{|h|}{a}\right) & , |h| \neq 0 \end{cases}$$

- O modelo gaussiano é um modelo cuja formulação é dada por:

$$Gau(h) = \begin{cases} 0 & , |h| = 0 \\ 1 - \exp\left(-\frac{|h|^2}{a}\right) & , |h| \neq 0 \end{cases}$$

### 3.3.4. – Anisotropia

Entende-se como anisotropia a qualidade de certos fenômenos de se distribuírem no espaço de uma forma diferenciada segundo a direção de propagação. Deste modo, nos fenômenos distribuídos espacialmente, a dependência espacial (intensidade, distância, ciclicidade) pode variar de acordo com a direção escolhida para a realização da análise. A anisotropia dos fenômenos estudados pode ser identificada a partir dos variogramas obtidos sendo consideradas direções diferentes. Temos na figura 6 a demonstração das direções mais utilizadas em Geoestatística.

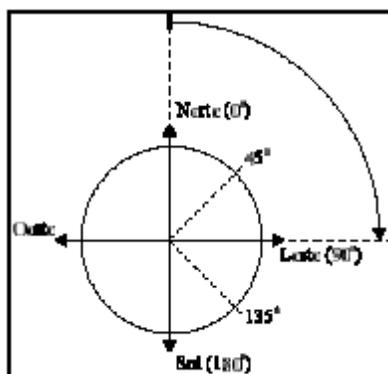


Figura 6 - Convenções direcionais utilizadas na Geoestatística

Na distribuição isotrópica os fenômenos se distribuem de uma maneira uniforme. Quando a distribuição de um fenômeno possui esta característica, os semivariogramas possuem similaridades entre si. Neste tipo de comportamento do fenômeno é possível descrever sua variabilidade espacial com a utilização de um único modelo. Na figura 7 temos a representação dos variogramas obtidos para 0°, 45°, 90° e 135° onde podemos observar um comportamento isotrópico de um fenômeno.

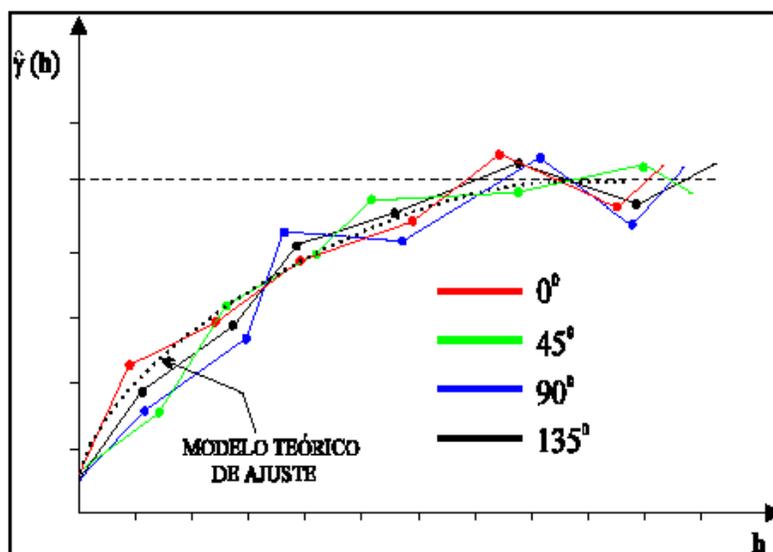


Figura 7 - Representação gráfica de semivariogramas isotrópicos

### 3.3.5.-. Krigeagem

A *krigeagem* é um método de interpolação que possui a propriedade de levar em consideração a continuidade do fenômeno. A base conceitual da Krigeagem está fundamentada na teoria das variáveis regionalizadas, (...) “onde a variável regionalizada é uma variável distribuída no espaço (ou tempo). A teoria diz que alguma medida pode ser vista como uma realização de uma função aleatória (ou processo aleatório, ou campo aleatório, ou processo estocástico). Esta teoria consolida o alicerce da geoestatística” (CÂMARA, 1999; cap 12). A geoestatística diz respeito à uma classe de técnicas usada para analisar e inferir valores de uma variável distribuída no espaço e ou no tempo. As técnicas de krigeagem no ambiente de Geoprocessamento têm como finalidade a análise de dados amostrais, desde que as amostras estejam indexadas ao espaço de forma pontual.

Neste modelo para se determinar o valor da função num ponto interno de um intervalo entre duas amostras, atribui-se os pesos a essas amostras por meio de uma análise espacial, baseada no variograma. Em outros métodos, por exemplo o de interpolação linear simples, os pesos atribuídos são todos iguais a  $1/N$  ( $N$  = número de amostras). Na Krigeagem há estimadores exatos com propriedades de não tendenciosidade e com variância mínima. Um estimador é um valor de parâmetro de uma população determinado por uma função dos valores observados na amostra. Estimativas não

tendenciosas são aquelas que em média, não apresentam diferença entre os valores verdadeiros e os estimados para cada amostra. A variância mínima significa que estes estimadores possuem menor variância dentre todos os estimadores não tendenciosos.

Existem vários tipos de Krigagem: krigagem ordinária, simples, universal, Co-krigagem, krigagem disjuntiva, etc. Vamos aqui explicar apenas um tipo, a Krigagem Ordinária.

Vamos considerar uma distribuição no espaço de algum fenômeno ( $Z$ ), em  $n$  pontos diferentes,. Teremos então um conjunto de pontos distribuídos no espaço, representados por pares de coordenadas  $(x, y)$ , onde  $x_i$  indica uma posição em duas dimensões do fenômeno. Para que se deduza o valor de  $Z(x_0)$ , ou seja o valor de  $Z$  no ponto  $0$ , pode-se realizar uma combinação linear dos  $n$  valores observados, adicionado a um parâmetro  $\lambda_0$ .

$$Z_{x_0}^* = \lambda_{.0} + \sum_{i=1}^n \lambda_i . Z(x_i)$$

O estimador que se vai utilizar será não tendencioso, portanto

$$E[Z_{x_0} - Z_{x_0}^*] = 0$$

Aplicando-se uma equação na outra, obtém-se

$$E[Z_{x_0}] = E\left[\lambda_0 + \sum_{i=1}^n \lambda_i . Z(x_i)\right] \Rightarrow m\lambda_0 + \sum_{i=1}^n \lambda_i m$$

Para que esta igualdade seja satisfeita é necessário que

$$\lambda_0 = 0$$

e

$$\sum_{i=1}^n \lambda_i = 1$$

Desta forma o estimador da Krigagem ordinária é:

$$z_{x_0}^* = \sum_{i=1}^n \lambda_i z(x_i)$$

Minimizando a variância do erro ( $\text{var}[Z_{X_0} - Z_{X_0}^*]$ ) sob a condição de que

$$\sum_{i=1}^n \lambda_i = 1$$

os pesos  $\lambda_i$  são obtidos a partir do seguinte sistema de equações, denominado sistema de krigagem ordinária:

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

onde,

- $C(X_i, X_j)$  e  $C(X_i, X_0)$  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.

Os valores interpolados a partir das equações da Krigagem coincidem com os valores das amostras. Sendo assim, este método pode ser considerado confiável para a determinação de valores de pontos a partir de algumas amostras.

#### 4 - Metodologia

No decorrer do projeto tomamos a decisão de montar um novo banco de dados, ao contrário de ampliarmos o banco de dados PB (já existente). Essa decisão foi tomada, ao verificarmos que este banco de dados já se constituía em um arquivo muito grande e que a sua ampliação acabaria por inviabilizar a manipulação dos dados nos computadores do Departamento de Estatística da UFPB. Além disso, observamos a existência de uma incompatibilidade temporal entre os dados antigos e os novos, já que os dados do banco PB são do Censo de 1991 e os dados escolhidos para a sua ampliação são referentes aos anos de 1998, 1999 e 2000. Além da carência de equipamentos e da incompatibilidade temporal percebemos que as informações inseridas no banco de dados PB não seriam necessárias para o desenvolvimento do projeto atual e não haveria sentido em trabalhar com arquivos tão pesados.

Desta forma, ao invés da ampliação, seguimos as etapas de confecção de um banco de dados no Sistema SPRING e procuramos, com isso, aperfeiçoar estes procedimentos, tentando a identificação de formas mais simples para montagem de um banco de dados.

##### 4.1 – Obtenção e organização dos dados

Os dados selecionados para este projeto foram obtidos no Sistema DATASUS do Ministério da Saúde do Brasil, que pode ser acessado no endereço <http://www.datasus.gov.br/>. Neste mesmo endereço está disponível o *software* TABWIN que faz tratamento georreferenciado de dados e que oferece também a opção de gravação dos dados no formato do SPRING.

Os dados escolhidos são de cinco tipos, número de hospitais por município (lei), quantidade de internações hospitalares por município (mor), recursos do SUS destinados a cada município (rec), quantidade de atendimentos ambulatoriais (amb) e população estimada para cada município (pop) e compreendem no total 147 variáveis trabalhadas. Apenas os dados de população são anuais, as demais variáveis são mensais referentes ao período dos anos de 1998, 1999 e 2000.

A nossa intenção, a princípio, era organizar uma única tabela no sistema TABWIN com os dados dos três anos selecionados. Em seguida faríamos a importação dessa tabela para o nosso banco

de dados no sistema SPRING. Devido a um grande número de dados, nos deparamos com uma limitação do sistema TABWIN quanto ao número de colunas permitidas em cada tabela, pois precisaríamos de uma tabela com 147 colunas, o que não foi possível fazer.

Entramos em contato com a equipe do INPE responsável pelo SPRING, e obtivemos a informação que em breve estaria saindo a versão 3.5 (a previsão fornecida era maio de 2001). Nesta versão iria haver a possibilidade de juntar duas tabelas diferentes dentro do sistema. A partir de então optamos em organizar os dados em tabelas separadas. Os dados ficaram organizados da seguinte maneira: uma primeira tabela com todas as variáveis dos anos de 1998 e 1999; e uma segunda tabela com todas as variáveis do ano de 2000.

#### 4.2 – Nomenclatura das variáveis

No Banco de dados só é permitido inserir nomes de variável com no máximo 8 caracteres. Deste modo fizemos uma espécie de abreviação em alguns dos nomes das variáveis para ser possível a sua inserção. Iremos explicar em seguida como fizemos essa abreviação para ficar mais clara a manipulação desses dados durante as análises. As variáveis que se apresentam no Banco de Dados são as seguintes:

1ª variável – **Geoid** Trata-se de um identificador geográfico próprio do SPRING. É um número de identificação que relaciona os municípios aos seus respectivos polígonos no mapa digital dentro do sistema Spring e que não deve ser alterado.

2ª variável – **Codigo** Nas tabelas do sistema TABWIN do DATASUS, cada município apresenta um número que serve como um tipo de identificador para o Ministério da Saúde. Este dado é exportado automaticamente pelo TABWIN quando fazemos a transferência da tabela para o SPRING.

3ª variável – **NomeMuni** São os nomes dos objetos geográficos, que no nosso caso são os municípios do estado. Esta variável é nomeada por nós e poderia ter outro nome como por exemplo, municípios, cidades, etc.

4ª variável – **Numero** Esta variável é o número que cada município possui no IBGE. Não é colocada de forma automática, para o nosso trabalho optamos por inserir na tabela esta variável cujos valores foram digitados para cada um dos municípios.

5ª variável – lei Número de hospitais para cada município, tomados mensalmente (leitos hospitalares). Abaixo encontram-se relacionadas as abreviaturas utilizadas para cada um dos meses e anos desta variável.

<u>Abreviatura</u>	<u>Ano – 1998</u>	<u>Abreviatura</u>	<u>Ano - 1999</u>	<u>Abreviatura</u>	<u>Ano - 2000</u>
<i>lei_0198</i>	Janeiro	<i>lei_0199</i>	Janeiro	<i>lei_0100</i>	Janeiro
	Fevereiro		Fevereiro		Fevereiro
<i>lei_0298</i>		<i>lei_0299</i>		<i>lei_0200</i>	
<b>lei_0398</b>	Março	<b>lei_0399</b>	Março	<b>lei_0300</b>	Março
<b>lei_0498</b>	Abril	<b>lei_0499</b>	Abril	<b>lei_0400</b>	Abril
<b>lei_0598</b>	Maio	<b>lei_0599</b>	Maio	<b>lei_0500</b>	Maio
<b>lei_0698</b>	Junho	<b>lei_0699</b>	Junho	<b>lei_0600</b>	Junho
<b>lei_0798</b>	Julho	<b>lei_0799</b>	Julho	<b>lei_0700</b>	Julho
<b>lei_0898</b>	Agosto	<b>lei_0899</b>	Agosto	<b>lei_0800</b>	Agosto
<b>lei_0998</b>	Setembro	<b>lei_0999</b>	Setembro	<b>lei_0900</b>	Setembro
<b>lei_1098</b>	Outubro	<b>lei_1099</b>	Outubro	<b>lei_1000</b>	Outubro
<b>lei_1198</b>	Novembro	<b>lei_1199</b>	Novembro	<b>lei_1100</b>	Novembro
<b>lei_1298</b>	Dezembro	<b>lei_1299</b>	Dezembro	<b>lei_1200</b>	Dezembro

6ª variável – mor Quantidade de internações hospitalares por município, tomados mensalmente (morbidade). Abaixo encontram-se relacionadas as abreviaturas utilizadas para cada um dos meses e anos desta variável.

<u>Abreviatura</u>	<u>Ano – 1998</u>	<u>Abreviatura</u>	<u>Ano - 1999</u>	<u>Abreviatura</u>	<u>Ano - 2000</u>
<i>mor_0198</i>	Janeiro	<i>mor_0199</i>	Janeiro	<i>mor_0100</i>	Janeiro
	Fevereiro		Fevereiro		Fevereiro
<i>mor_0298</i>		<i>mor_0299</i>		<i>mor_0200</i>	
<b>mor_0398</b>	Março	<b>mor_0399</b>	Março	<b>mor_0300</b>	Março
<b>mor_0498</b>	Abril	<b>mor_0499</b>	Abril	<b>mor_0400</b>	Abril
<b>mor_0598</b>	Maio	<b>mor_0599</b>	Maio	<b>mor_0500</b>	Maio
<b>mor_0698</b>	Junho	<b>mor_0699</b>	Junho	<b>mor_0600</b>	Junho
<b>mor_0798</b>	Julho	<b>mor_0799</b>	Julho	<b>mor_0700</b>	Julho

<b>mor_0898</b>	Agosto	<b>mor_0899</b>	Agosto	<b>mor_0800</b>	Agosto
<b>mor_0998</b>	Setembro	<b>mor_0999</b>	Setembro	<b>mor_0900</b>	Setembro
<b>mor_1098</b>	Outubro	<b>mor_1099</b>	Outubro	<b>mor_1000</b>	Outubro
<b>mor_1198</b>	Novembro	<b>mor_1199</b>	Novembro	<b>mor_1100</b>	Novembro
<b>mor_1298</b>	Dezembro	<b>mor_1299</b>	Dezembro	<b>mor_1200</b>	Dezembro

7ª variável – **Rec**

Recursos do SUS destinados mensalmente a cada município.

Abaixo encontram-se relacionadas as abreviaturas utilizadas para cada um dos meses e anos desta variável.

<u>Abreviatura</u>	<u>Ano – 1998</u>	<u>Abreviatura</u>	<u>Ano - 1999</u>	<u>Abreviatura</u>	<u>Ano - 2000</u>
<i>rec_0198</i>	Janeiro Fevereiro	<i>rec_0199</i>	Janeiro Fevereiro	<i>rec_0100</i>	Janeiro Fevereiro
<i>rec_0298</i>		<i>rec_0299</i>		<i>rec_0200</i>	
<b>rec_0398</b>	Março	<b>rec_0399</b>	Março	<b>rec_0300</b>	Março
<b>rec_0498</b>	Abril	<b>rec_0499</b>	Abril	<b>rec_0400</b>	Abril
<b>rec_0598</b>	Maio	<b>rec_0599</b>	Maio	<b>rec_0500</b>	Maio
<b>rec_0698</b>	Junho	<b>rec_0699</b>	Junho	<b>rec_0600</b>	Junho
<b>rec_0798</b>	Julho	<b>rec_0799</b>	Julho	<b>rec_0700</b>	Julho
<b>rec_0898</b>	Agosto	<b>rec_0899</b>	Agosto	<b>rec_0800</b>	Agosto
<b>rec_0998</b>	Setembro	<b>rec_0999</b>	Setembro	<b>rec_0900</b>	Setembro
<b>rec_1098</b>	Outubro	<b>rec_1099</b>	Outubro	<b>rec_1000</b>	Outubro
<b>rec_1198</b>	Novembro	<b>rec_1199</b>	Novembro	<b>rec_1100</b>	Novembro
<b>rec_1298</b>	Dezembro	<b>rec_1299</b>	Dezembro	<b>rec_1200</b>	Dezembro

8ª variável – **amb**

Quantidade de atendimentos ambulatoriais realizados mensalmente por

município. Abaixo encontram-se relacionadas as abreviaturas utilizadas para cada um dos meses e anos desta variável.

<u>Abreviatura</u>	<u>Ano – 1998</u>	<u>Abreviatura</u>	<u>Ano - 1999</u>	<u>Abreviatura</u>	<u>Ano - 2000</u>
<i>amb_0198</i>	Janeiro Fevereiro	<i>amb_0199</i>	Janeiro Fevereiro	<i>amb_0100</i>	Janeiro Fevereiro
<i>amb_0298</i>		<i>amb_0299</i>		<i>amb_0200</i>	
<b>amb_0398</b>	Março	<b>amb_0399</b>	Março	<b>amb_0300</b>	Março
<b>amb_0498</b>	Abril	<b>amb_0499</b>	Abril	<b>amb_0400</b>	Abril
<b>amb_0598</b>	Maio	<b>amb_0599</b>	Maio	<b>amb_0500</b>	Maio
<b>amb_0698</b>	Junho	<b>amb_0699</b>	Junho	<b>amb_0600</b>	Junho
<b>amb_0798</b>	Julho	<b>amb_0799</b>	Julho	<b>amb_0700</b>	Julho
<b>amb_0898</b>	Agosto	<b>amb_0899</b>	Agosto	<b>amb_0800</b>	Agosto
<b>amb_0998</b>	Setembro	<b>amb_0999</b>	Setembro	<b>amb_0900</b>	Setembro
<b>amb_1098</b>	Outubro	<b>amb_1099</b>	Outubro	<b>amb_1000</b>	Outubro
<b>amb_1198</b>	Novembro	<b>amb_1199</b>	Novembro	<b>amb_1100</b>	Novembro
<b>amb_1298</b>	Dezembro	<b>amb_1299</b>	Dezembro	<b>amb_1200</b>	Dezembro

9ª variável – **pop**

População anual estimada para cada município.

**Pop\_98** Estimativa da população residente em cada município no ano de 1998

**Pop\_99** Estimativa da população residente em cada município no ano de 1999

**Pop\_00** Estimativa da população residente em cada município no ano de 2000

#### 4.3 – Montagem do banco de dados

Um banco de dados no SPRING corresponde fisicamente a um diretório onde serão armazenados suas definições de Categorias e Classes e os projetos pertencentes ao banco. Entende-se por categorias do banco de dados, arquivos com estruturas próprias, como por exemplo, mapas temáticos, mapas cadastrais, redes, imagens de satélite, modelos numérico de terreno e objeto, que são tipos de dados em Geoprocessamento.

##### 4.3.1 - Criando um banco de dados:

```
#spring
[Arquivo] [Banco de Dados]
[Diretório: /<usuário>/springdb]
```

(Nome: *Saude*)

(nome qualquer com 8 letras no máximo)

[Criar]

[Ativar]

#### 4.3.2 - Criação do projeto

Para a criação de um Projeto deve ser fornecido um nome, projeção (ver glossário) e retângulo envolvente. Será criado um sub-diretório, debaixo do diretório correspondente ao banco, e todos os dados referentes a uma determinada região serão armazenados nele. Os projetos do banco de dados são armazenados em sub-diretórios juntamente com seus arquivos de dados, pontos, linhas, objetos, textos, etc. Para definirmos um projeto fazemos o seguinte:

[Arquivo] [Projeto]

(Nome: *SUS*)

(nome com 8 letras no máximo)

[Projeção]

[Sistemas ] UTM]

[Modelos da Terra ] SAD 69]

(Origem - Long: *o 39 0 0.00*)

[Executar]

[Coordenadas - Geográficas]

(Long 1: *o 38 55 0.00*) (Long 2: *o 34 35 0.00*)

(Lat 1: *s 8 25 0.00*) (Lat 2: *s 5 55 0.00*)

[Criar]

[Ativar]

#### 4.3.3.- Criação das categorias cadastral, objeto e modelo numérico de terreno

Neste tópico, explicaremos como criamos a estrutura para o recebimento dos dados (mapa digitalizado, informações, etc.). Inicialmente definimos uma categoria *Cadastral* que será denominada por *Estado*. Em seguida definiremos a categoria *Objeto* que serão os *municípios*. Por fim iremos necessitar da categoria *modelo numérico* que será utilizada para a saída dos novos dados gerados, que serão do tipo *pontos amostrais*. Para definirmos modelos dos dados procedemos da seguinte maneira:

[Arquivo] [Modelos de dados]

(Categorias - Nome: *Estado*)

(Modelo - Cadastral)

(Categorias - Criar)

[Modelos de dados]

(Categorias - Nome: *Município*)

(Modelo - Objeto)

(Categorias - Criar)

[Modelos de dados]

(Categorias - Nome: *PontosA*)

(Modelo - Objeto)

(Categorias - Criar)

(Executar)

(para armazenar as categorias criadas)

#### 4.3.4 – Definição dos atributos (informações) da categoria objeto

Esta etapa da confecção do banco de dados prepara a estrutura para o recebimento das informações sobre os objetos (municípios). É na verdade a inserção das nomenclaturas das variáveis que serão utilizadas no banco. Os passos para a definição dos atributos são os seguintes:

[Arquivo] [Modelo de dados]

(Categorias – *Município*)

(Atributos) do tipo texto

(Nome: *NomeMuni*)

(Tipo - texto)

(Inserir)

(selecionar a categoria municípios)

(clique em atributos)

(definir o tamanho: número de letras)

(Atributos) do tipo inteiro

(Nome: *Numero*)

(Tipo - inteiro)

(Inserir)

(Atributos) do tipo inteiro

(Nome: *lei*)

(Tipo - Inteiro)

(Inserir)

(Atributos) do tipo inteiro

(Nome: *mor*)

(Tipo - inteiro)

(Inserir)

(Atributos) do tipo real  
(Nome: *rec*)  
(Tipo - Real)  
(Inserir)  
(Executar) (esses passos deverão ser feitos para todas as variáveis que irão fazer parte do seu banco)

Uma vez definido todos os atributos pressiona-se o botão *Executar*. Esta ação irá gravar na estrutura de dados do Spring todos os atributos do Objeto "Municipios". Uma vez gravado **não tem volta**; isto é; o que foi definido não pode ser **apagado** e nem **renomeado**. Portanto, antes de pressionar o botão executar deve-se certificar que os nomes estão corretos. Depois de efetuados os procedimentos dos tópicos anteriores, definindo o modelo cadastral e objeto, como também os atributos da categoria objeto estaremos prontos para efetuar a importação dos dados (mapas, tabelas, etc.)

#### 4.3.5 – Importação dos dados

O SPRING aceita a importação de arquivos em alguns formatos, como por exemplo ASCII, ARC/INFO, DXF, TIFF, RAW e SITIM. No caso do nosso banco teremos que importar as linhas e polígonos dos mapas, bem como seus identificadores. Além destes arquivos temos que importar também as tabelas com os dados. A importação dos dados pode ser feita através dos seguintes procedimentos:

I - Importação do arquivo de linhas: PB\_L2D. spr

[Arquivo] [Importar]  
(Diretório...c:\arquivos de programa\ spring\springdb)  
(Modelo - ASCII)  
(Selecionar: *PB\_L2D.spr*)  
(Entidade: *linhas sem ajuste*)  
(Unidade: *Graus*)  
(Escala: *100.000*)

(Projeção)  
(Projeção: *UTM SAD 69*)  
(Longitude: *o 39 0 0.00*)  
(Executar)  
(Retângulo envolvente: cursor: *não* PI: *ativo*)

(Coordenadas: *Geográficas*)

<u>Long</u> 1: <i>o 38 55 0.00</i>	<u>Long</u> 2: <i>o 34 35 0.00</i>
Lat 1: <i>s 8 25 0.00</i>	Lat 2: <i>s 5 55 0.00</i>

(Projeto: *SUS*)  
(Categoria: *Estado*)  
(PI: *Municipios*)  
(Executar)  
(Fechar)

Com estes procedimentos estão importadas as linhas do mapa. Devemos agora visualizar o mapa para nos certificar que o mesmo foi importado com sucesso.

## II - Visualizando as linhas importadas

[Painel de controle]  
(Selecionar as categorias: *Estado e Municipios*)  
(Selecionar: *linhas*)  
Clicar no lápis para desenhar o mapa na tela



Depois de realizada a importação do arquivo de linhas temos que filtrar essas linhas além de criar uma topologia para estas linhas. A filtragem e a criação da topologia devem ser feitas porque os mapas provenientes do DATASUS, assim como de outros sistemas vem com as linhas de fronteiras ou divisas duplicadas e também não possuem topologia. Resumidamente, topologia significa o relacionamento entre os polígonos ou regiões de um mapa. Por exemplo, como podemos ver na figura 8, o polígono “A” faz fronteira (ou divisa) com o polígono “B” através da linha “L1”.

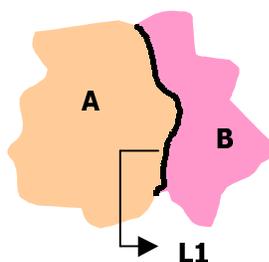


Figura 8 – Topologia entre polígonos

Deste modo temos que filtrar ou limpar as linhas duplicadas e depois criar a topologia. Uma vez a topologia criada, vamos importar os identificadores dos polígonos (o Arquivo Pb\_LAB.spr) e finalmente o arquivo das Tabelas (no caso de haver mais de uma).

### III - Passos para realizar a filtragem de linhas duplicadas

[Painel de controle]  
(categorias: *Estado e municipios*) (selecionar)  
[Ferramentas – Limpar vetor]  
(*Eliminar fragmentos duplicados*) (selecionar)  
(*Topologia automática*)  
(Fator de ajuste: *0 m*)  
(Executar)

### IV – Como criar a topologia do mapa

[Cadastral – Edição topológica]  
(Edição gráfica)  
(Editar: *linhas*)  
(Modo: *contínuo*)  
(Topologia: *Manual*)  
(Fator de digit.: *0*)  
(Operação: Criar linhas)  
(Ajustar)  
  
(***Poligonalizar***)  
(Fechar)

Até agora foi criada a topologia do mapa importado. Agora vamos importar identificadores para cada um dos polígonos (municípios).

### V – Importando o arquivo de identificadores

[Arquivo] [Importar]  
(Diretório...c:\arquivos de programa\ spring\springdb)  
(Modelo - ASCII)  
(Selecionar: *PB\_LAB.spr*)  
(Entidade: *Identificadores*)  
(Unidade: *Graus*)  
(Escala: *100.000*)  
  
(Projeção)  
(Projeção: *UTM SAD 69*)  
(Longitude: *o 39 0 0.00*)  
(Executar)  
(Retângulo envolvente: cursor: *não* PI: *ativo*)  
(Coordenadas: *Geográficas*) Long 1: *o 38 55 0.00* Long 2: *o 34 35 0.00*  
Lat 1: *s 8 25 0.00* Lat 2: *s 5 55 0.00*  
  
(Projeto: *SUS*)

(Categoria: *Estado*)  
(PI: *Municipios*)  
(Executar)  
(Fechar)

A partir de agora os objetos Municípios estão identificados e associados à uma localização no mapa e podem ser visualizados seguindo os passos descritos na seqüência.

#### VI – Visualizando os objetos Municípios

[Editar – Objeto]	(selecionar)
(Categorias: <i>Estado e Municipios</i> )	
(Modo de seleção: <i>Tela</i> )	(Pressione o cursor sobre um polígono qualquer do mapa e observe que aparece seu nome)

Nesta etapa o banco de dados já está pronto para receber as tabelas com os atributos dos objetos. Isto quer dizer que, depois da importação da tabela, cada município do mapa terá associado ao seu polígono os atributos definidos anteriormente.

#### VII – Importação de tabelas

Na versão 3.5 do SPRING podemos importar várias tabelas para um banco de dados, mas apenas uma tabela pode ser importada como *tabela de objeto*, ou seja, tabela georreferenciada. As outras tabelas importadas deverão ser *tabelas não espaciais*, as quais não poderão ter seus dados consultados através do mapa. Para importá-las devemos fazer o seguinte:

[Arquivo] [Importar tabela]  
(Diretório...c:\Tabwin\ Temp\  
(OK)  
(Formato: DBASE)  
(Selecionar a tabela a ser importada – no nosso caso a tabela importada foi : *finalX.dbf*)  
(Operação Spring: *Criar nova tabela de objetos*)  
(Executar)

Com isso teremos importada a tabela dos atributos espaciais. Para a importação das outras tabelas iremos proceder da mesma forma, só que a “operação do Spring” que iremos selecionar será *criar nova tabela não espacial*.

## VIII – Ligação entre tabelas

Para a realização de consultas e de análises no banco de dados é preciso que os dados estejam arranjados em uma só estrutura em forma de matriz (tabela). Para que dados organizados separadamente possam ser lidos como um só, existe na versão 3.5 do SPRING uma ferramenta que permite que estas tabelas sejam ligadas. Assim, somente com a realização de uma ligação entre as tabelas, o sistema reconhecerá as duas como sendo uma só no momento de fazer as consultas espaciais ou as análises de dados. Para ligar duas tabelas podemos fazer da seguinte forma;

[Editar – Ligação de tabelas]                      (aqui temos uma janela com a tabela de objeto e tabela não espacial que importamos. É necessário que selecionemos a tabela que queremos associar ao objeto, bem como o atributo que queremos que seja associado

(Tabela não espacial: *AfinalX*)  
(Atributo da tabela não espacial: *Amb\_1299*)  
(Conectar)  
(Fechar)

### 4.4 – Estatísticas descritivas dos dados

Realizamos a análise exploratória dos dados, onde efetuamos a estatística descritiva para todas as variáveis do banco de dados. Temos nas tabelas 1a, 1b, 1c, e 1d exemplos da estatística de algumas amostras das variáveis. Selecionamos apenas alguns exemplos para ilustrar o comportamento das variáveis, pois, em geral, eles se apresentam de maneira semelhante, tendo sua distribuição muito longe de ser considerada como normal

Tabela 1 – Estatística descritiva da variável Amb\_1299

	Amostra Amb_1299_LAB
Número de amostras	224
Nº. de amostras válidas	224
Média	8772.384
Variância	1469139613.362
Desvio Padrão	38329.357
Coeficiente de Variação	4.369
Coeficiente de Assimetria	12.130
Coeficiente de Curtose	160.784
Valor Mínimo	327.000
Quartil Inferior ..	1808.500
Mediana	3346.500
Quartil Superior .	6267.000
Valor Máximo	534136.000

Tabela 2 - Estatística descritiva da variável Lei\_0198\_LAB

	Amostra Lei_0198_LAB
Número de amostras	224
Número de amostras válidas	224
Média	52.580
Variância	97321.886
Desvio Padrão	311.965
Coeficiente de Variação	5.933
Coeficiente de Assimetria	10.307
Coeficiente de Curtose	111.491
Valor Mínimo	0.000
Quartil Inferior	0.000
Mediana	0.000
Quartil Superior	25.000
Valor Máximo	3693.000

**Tabela 3** - Estatística descritiva da variável Mor\_0198\_LAB

	Amostra Mor_0198_LAB
Número de amostras	224
Número de amostras válidas	224
Média	88.746
Variância	208435.708
Desvio Padrão	456.548
Coeficiente de Variação	5.144
Coeficiente de Assimetria	9.558
Coeficiente de Curtose	97.258
Valor Mínimo	0.000
Quartil Inferior	0.000
Mediana	0.000
Quartil Superior	40.500
Valor Máximo	4808.000

**Tabela 4** - Estatística descritiva da variável Rec\_0198\_LAB

	Amostra Rec_0198_LAB
Número de amostras	224
Número de amostras válidas	224
Média	47001.176
Variância	133339780698.547
Desvio Padrão	365157.200
Coeficiente de Variação	7.769
Coeficiente de Assimetria	12.136
Coeficiente de Curtose	157.395
Valor Mínimo	0.000
Quartil Inferior	1693.655
Mediana	4094.290
Quartil Superior	12301.405
Valor Máximo	5003715.500

Podemos ver através dos valores tabela 1 a que na variável **amb\_1299** a média dos valores está em torno de 8.772 apresentando um desvio padrão muito alto, o que indica uma disparidade na distribuição dos atendimentos ambulatoriais em dezembro de 1999 no estado. Na tabela 2 da variável **lei\_0198**, a diferença entre os valores máximo e mínimo é grande, o desvio padrão também se apresenta alto com relação aos valores. As demais variáveis do banco de dados possuem um comportamento semelhante aos que estão ilustrados. Nos gráficos de distribuição normal e histogramas das figuras 9, 10, 11 e 12 podemos observar a distribuição não normal das variáveis de uma forma mais ilustrativa.

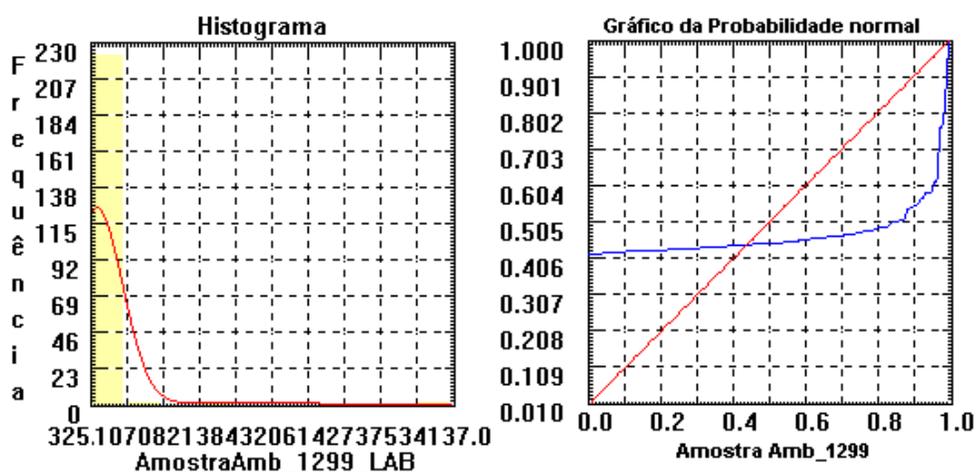


Figura 9 – Histograma e gráfico da probabilidade normal da variável Amb\_1299

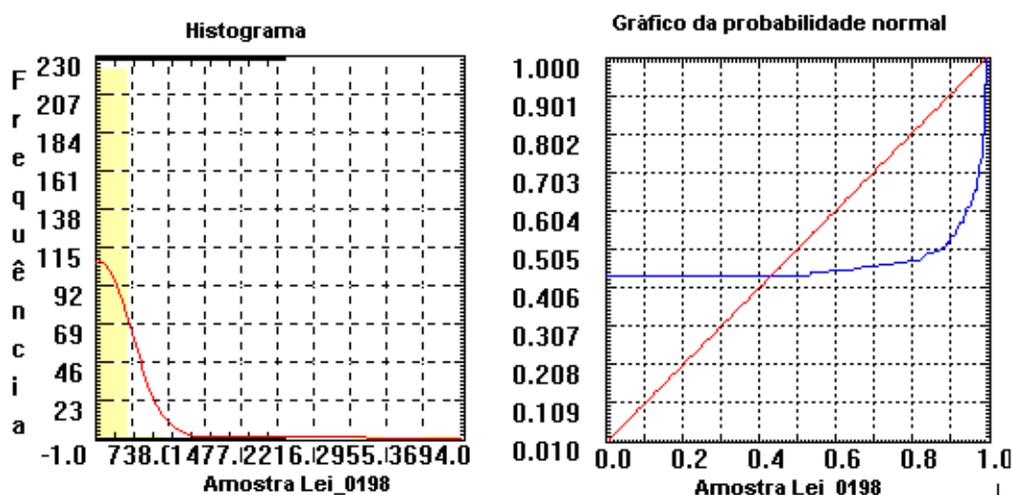


Figura 10 – Histograma e gráfico da probabilidade normal da variável Lei\_0198

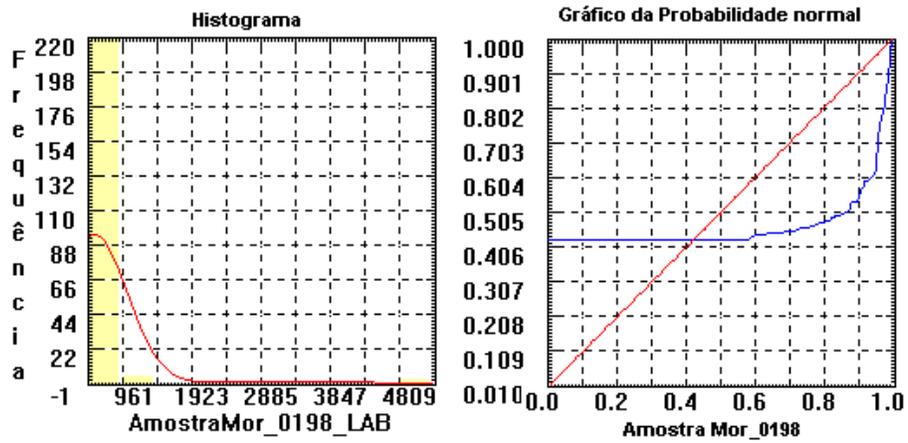


Figura 11 – Histograma e gráfico da probabilidade normal da variável Mor\_0198

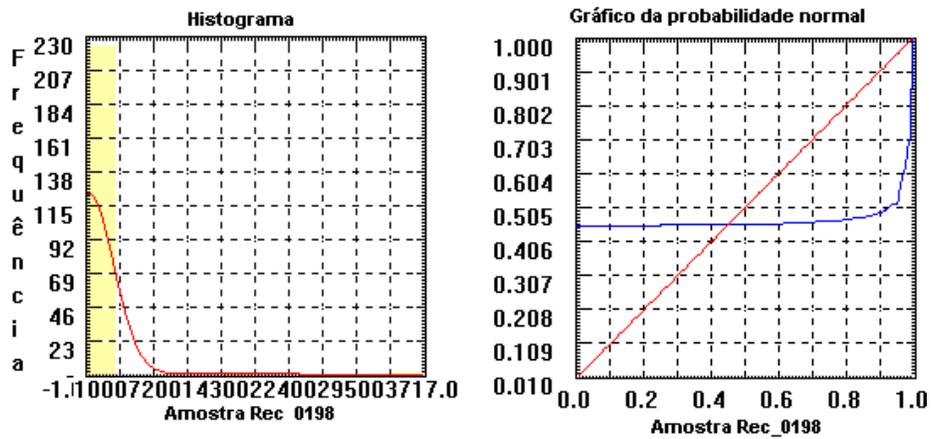


Figura 12 – Histograma e gráfico da probabilidade normal da variável Rec\_0198

Apesar dos gráficos da estatística descritiva demonstrarem claramente um comportamento não normal da distribuição dos dados, iremos efetuar a geração do variograma, bem como efetuar a técnica da Krigeagem, de acordo com o proposto inicialmente no projeto. No entanto queremos deixar claro que esse procedimento será feito somente para fins de aprendizado, já que o procedimento correto neste caso seria a suspensão da continuidade da análise, buscando transformações para adequar a distribuição das variáveis à distribuição normal. Porém, como o sistema SPRING não permite transformações das variáveis diretamente no banco de dados, tal procedimento levaria a refazer totalmente o banco de dados desde o princípio. Assim sendo, apesar de que, do ponto de vista teórico as análises e procedimentos a seguir não têm nenhum sentido prático, eles serão efetuados levando-se em conta, única e exclusivamente, o aprendizado das técnicas.

#### 4.5 – Geração do variograma

A variação dos fenômenos que possuem uma distribuição contínua no espaço é representada pelo variograma, onde é medida a variância da diferença dos valores em todos os pontos do espaço. Desta forma, para a realização da geração do variograma é necessário que os valores do nosso banco de dados estejam na forma de pontos amostrais.

No nosso caso temos um banco de dados com um plano de informação cadastral (Estado), representado por entidades do tipo polígonos (Municípios). Iremos criar um outro Plano de informação do tipo numérico porém com entidades do tipo pontos. Temos na ferramenta “geração de pontos amostrais” a possibilidade para essa transformação que pode ser feita da seguinte maneira:

[Painel de controle]  
(selecionar categorias: *Estado*)  
(selecionar PI: *Municípios*)  
(Ferramenta: *Geração de pontos amostrais*)  
(Categorias de saída: *PontosA*)  
(Executar)  
(PI de saída: *AmostraRec0198*)  
(Categoria: *Municípios*)  
(Atributos: *Rec0198*)  
(Executar)  
(Sim – Criar novos pontos no centro de massa)

De maneira semelhante foram criadas as outras amostras das variáveis do banco de dados. Cada mês de cada uma das variáveis se torna uma amostra. Para a geração do variograma da amostraAmb\_1299 fizemos da seguinte maneira:

- (Análise – Geoestatística)
- (Geração de semivariograma)
- (PI ativo: *AmostraAmb\_1299\_LAB*)
- (Análise: *Unidirecional*)
- (Amostragem: *Irregular*)
- (Opções: *Semivariograma*)
- Parâmetros de Lag
- (No Lag: 3)
- (Incremento e Tolerância: *o que vem no sistema*)
- (Dir: 2)
- (Executar)

Temos na figura 13 o variograma gerado. Ao lado temos os resultados numéricos, como o número de lags, o incremento, a tolerância, direção, etc., que servirão de base para realização do ajuste do variograma que iremos realizar em seguida.

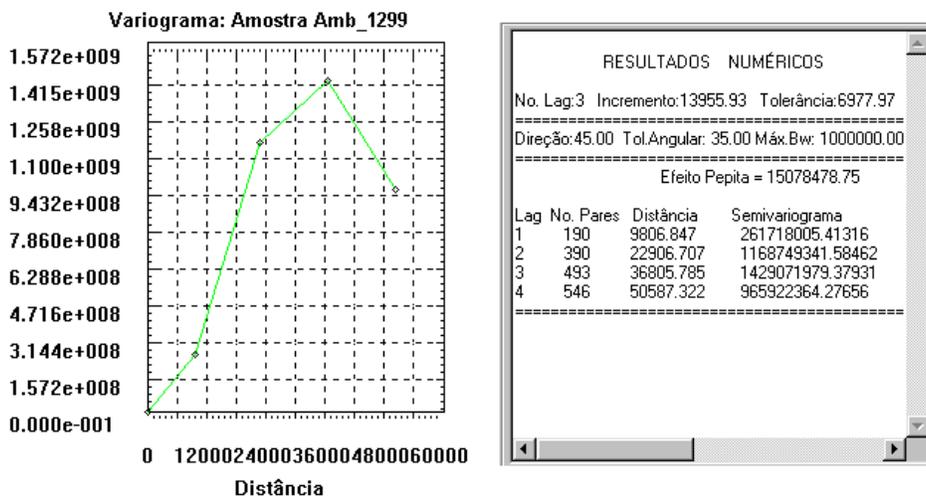


Figura 13– Variograma da amostra Amb\_1299

O ajuste do variograma pode ser feito utilizando os modelos Exponencial, Gaussiano e o Esférico. Para a escolha do modelo que mais se adequa ao nosso caso, tomamos como base a figura demonstrativa dos modelos ilustrada na figura 5 e selecionamos aquele que apresenta uma curva que

mais se assemelha com a curva do nosso variograma. Para a realização do ajuste do variograma fizemos o seguinte:

- (Análise – Geoestatística)
- (Ajuste do semivariograma)
- (Ajuste: Automático)
- (Número de estruturas: 1)
- (Modelo: Gaussiano)
- (Executar)
- (Verificar ajustes)

Para visualizar o ajuste do variograma clicar sobre o nome da amostra: amb\_1299. Além do gráfico, o sistema mostra também um relatório de dados do ajuste do variograma, que servirá para o preenchimento dos “parâmetros estruturais” que existe na janela do ajuste. Este passo (parâmetros estruturais) é fundamental para a realização da técnica da Krigeagem. A figura 14 nos mostra o gráfico do ajuste do variograma da amostra Amb\_1299.

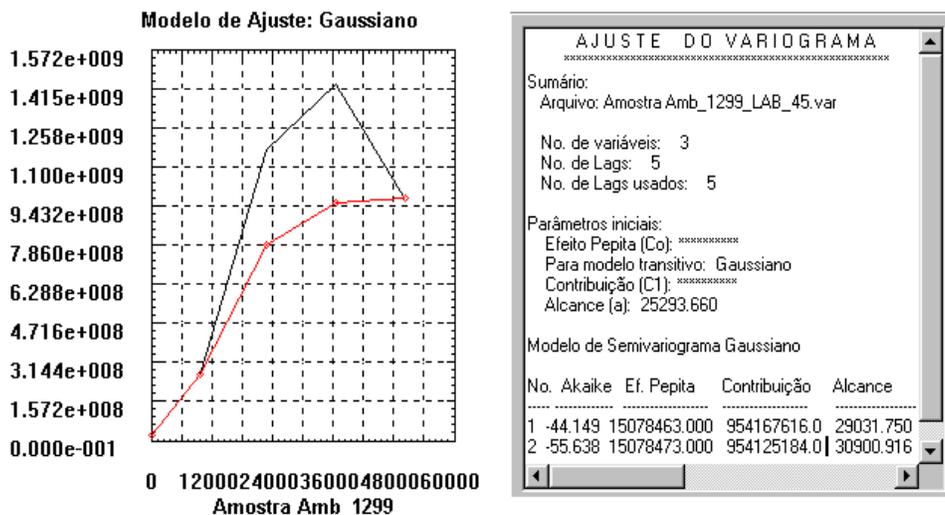


Figura 14 – Gráfico do ajuste do variograma da amostra Amb\_1299

Nota-se na figura 14 que o modelo não está bem ajustado aos dados. Uma das causas é a distribuição não-normal dos dados (ver figura 9). Já que esta premissa não é satisfeita, é muito difícil obter um bom ajuste.

Depois da verificação do gráfico do ajuste do variograma, ainda na janela do ajuste, temos os “parâmetros estruturais”, onde vamos preencher com os valores que apareceram no relatório de dados. Para isso basta que façamos o seguinte:

(Parâmetros estruturais)  
(Nº de estruturas: 1)  
(Efeito Pepita: 0)  
(Tipo: Gaussiano)  
(Angulo de anisotropia: 0)

(Os valores da *contribuição* e dos *alcances* vão ser selecionados do relatório de dados e devem ser escolhidos os valores da linha que apresentar o menor valor do *Akaike* no caso aqui é -55.638)

(Contribuição: 954125184.0)  
(Alcance máximo e mínimo: 30900.916)

(os alcances devem ser iguais, pois o variograma é do tipo isotrópico)

(Executar)  
(Fechar)

Podemos agora fazer a geração do variograma para as outras amostras seguindo os mesmos passos que acabamos de realizar. Iremos demonstrar os passos para a amostra Mor\_0198 de uma forma menos detalhada por se tratar dos mesmos procedimentos.

(Análise – Geoestatística)  
(Geração de semivariograma)  
(PI ativo: *AmostraMor\_0198\_LAB*)  
(Análise: *Unidirecional*)  
(Amostragem: *Irregular*)  
(Opções: *Semivariograma*)  
Parâmetros de Lag  
(No Lag: 6)  
(Incremento: o padrão)  
(Tolerância: o padrão)  
(Dir: 1)  
(Executar)

Podemos observar o variograma gerado, bem como seus resultados numéricos na figura 15 .

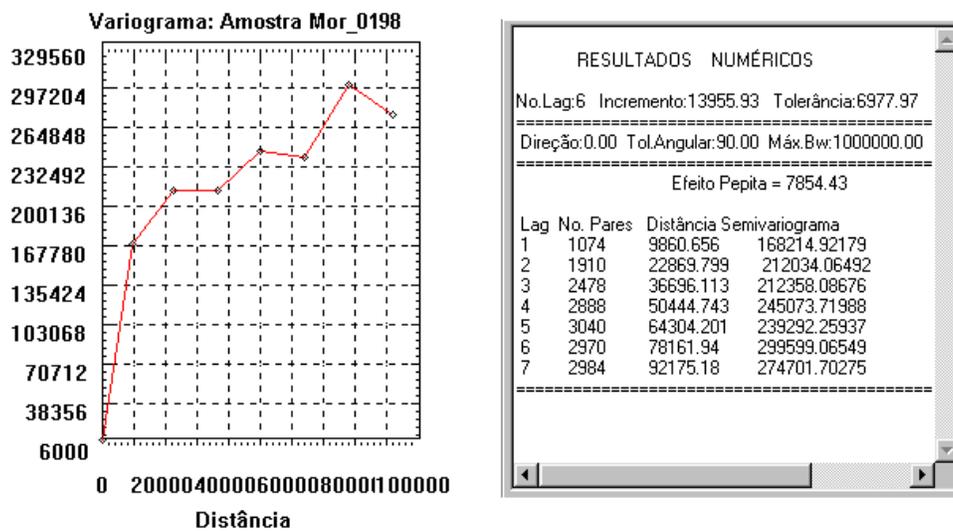


Figura 15 – Variograma gerado para a amostra Mor\_0198

Para o ajuste do modelo usamos:

(Ajuste: Automático)

(Número de estruturas: 1)

(Modelo: Esférico)

(podemos observar o gráfico do modelo de ajuste na figura 16)

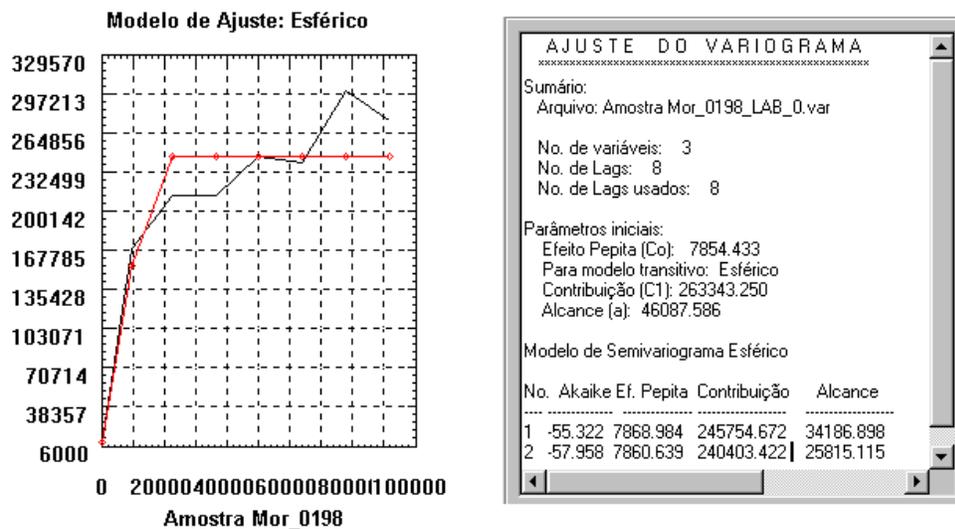


Figura 16 – Ajuste do variograma gerado para a amostra Mor\_0198

Novamente, nota-se na figura 16 a baixa qualidade do ajuste do modelo, visto a distribuição não-normal dos dados. Os parâmetros estruturais foram preenchidos com os seguintes valores extraídos do relatório dos dados:

(Parâmetros estruturais)  
(Nº de estruturas: 1)  
(Efeito Pepita: 0)  
(Tipo: Esférico)  
(Angulo de anisotropia: 0)  
(Contribuição: 240403422.0)  
(Alcance máximo e mínimo: 25815.115)  
(Executar)  
(Fechar)

#### 4.6 – Aplicação da Krigeagem

Depois de obtido o variograma, o seu ajuste e os parâmetros podemos realizar a técnica da Krigeagem para a amostra *Amb\_1299*. Para a realização desta técnica podemos fazer da seguinte maneira:

(Análise – Geoestatística)  
(Krigeagem)  
(PI ativo: *AmostraAmb\_1299*)  
(Verificar modelo) (podemos com isso confirmar se os dados que inserimos anteriormente nos parâmetros estruturais estão corretos)  
(Executar)  
(Tipo: *ordinária*)  
(Parâmetros de interpolação: mínimo: 4 – Máximo 16)  
(Saída)  
(Categoria: *PontosA*)  
(Plano de Informação: *Kri\_G\_Iso\_Amb1299*)  
(Executar)

Aplicamos a Krigeagem da mesma forma na amostra *Mor\_0198\_LAB* e obtivemos o plano de informação *Kri\_E\_Iso\_Mor0198*.

(Análise – Geoestatística)  
(Krigeagem)  
(PI ativo: *AmostraMor\_0198*)  
(Verificar modelo)  
(Executar)  
(Tipo: *ordinária*)  
(Parâmetros de interpolação: mínimo: 4 – Máximo 16)  
(Saída)  
(Categoria: *PontosA*)  
(Plano de Informação: *Kri\_E\_Iso\_Mor0198*)

(Executar)

Podemos observar nas figuras 17, 18, 19 e 20 que o resultado da aplicação desta técnica resultou em mapas escuros. Por se tratarem apenas de experiências de aprendizado sobre modelos que não podem ser aplicados realmente, não iremos fazer considerações sobre os mesmos. Se esses mapas fossem obtidos de forma correta, poderíamos perceber uma variabilidade espacial acentuada próxima a cidades como João Pessoa e Campina Grande frente aos demais municípios que as cercam. Além dessas duas existem outras cidades que apresentam variabilidades espaciais em relação à municípios vizinhos, no centro e na porção oeste do estado. Um cruzamento com um mapa de rodovias do estado poderia trazer informações interessantes.

**Krigeagem \ Modelo Esférico \ Variável Mor\_0198**

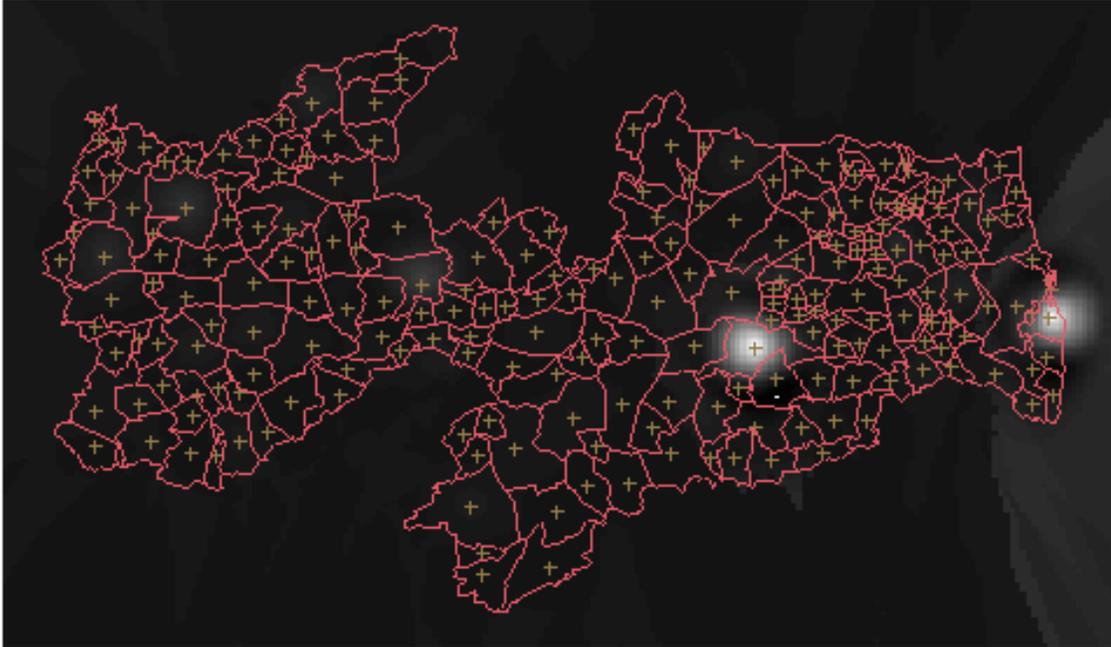


Figura 17 – Mapa resultante da Krigeagem da amostra Mor\_0198

**Variância da Krigeagem \ Amostra Mor\_0198**

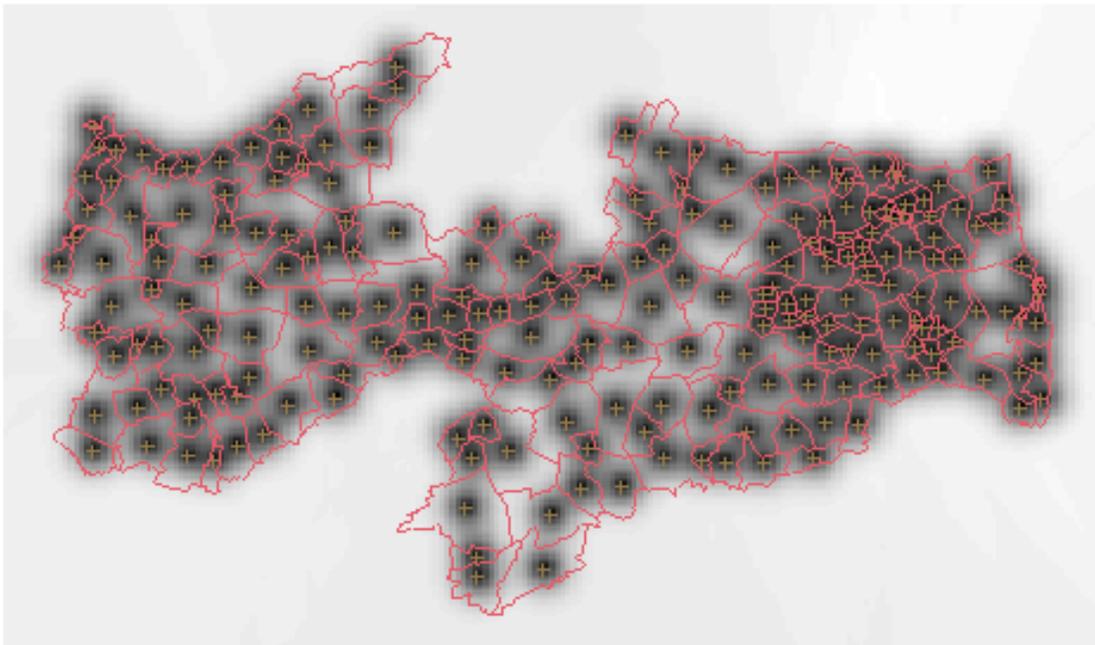


Figura 18 – Mapa da variância da Krigeagem da amostra Mor\_0198

**Krigeagem \ Gaussiana \ Amostra Amb\_1299**



Figura 19 – Mapa resultante da Krigeagem da amostra Amb\_1299

**Variância da Krigeagem \ Amostra Amb\_1299**

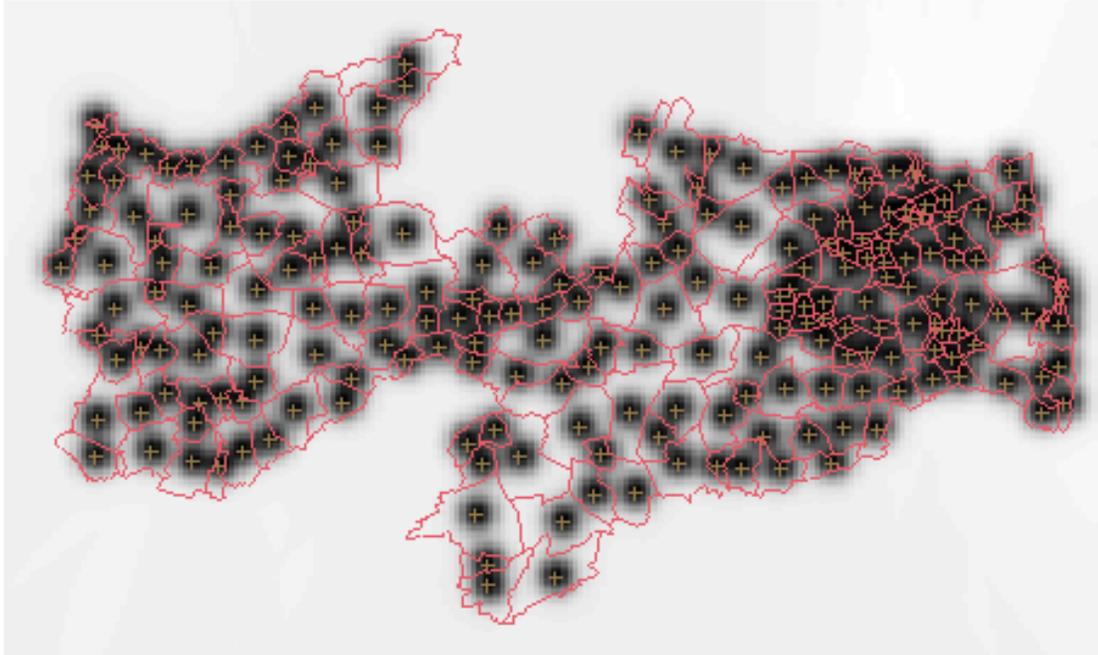


Figura 20 – Mapa da variância da Krigeagem da amostra Amb\_1299

#### 4.7 – Considerações finais

Realizamos a contento os estudos introdutórios que nos serviram de embasamento para as procedimentos de análises espaciais que foram nossa meta. Mesmo com a estatística descritiva dos dados nos mostrando um comportamento de não normalidade, decidimos gerar os variogramas de algumas amostras e aplicar em seguida a técnica da Krigagem. Deixando claro, no entanto, que estas técnicas de interpolação foram realizadas apenas com a finalidade de aprendizado e que os resultados não são confiáveis. Serviram, principalmente, para aperfeiçoar os procedimentos para uma aplicação futura, onde existam dados com uma distribuição considerada normal.

## Bibliografia

CÂMARA, G. e MEDEIROS, J.S.; *Geoprocessamento para Projetos ambientais*. VIII Simpósio Brasileiro de Sensoriamento Remoto, 1996.

CÂMARA, G. et al (ed); *Geoprocessamento – teoria e aplicação*, 1999. Obtido em <<http://www.dpi.inpe.br/gilberto/livro/>> , em 26/01/2000.

CARVALHO, M. S.; *Aplicação de métodos de análise espacial na caracterização de áreas de risco à saúde*. Rio de Janeiro: Tese de Doutorado em Engenharia Biomédica – COOPE/UFRJ, 1997.

CARVALHO, M. S. e CRUZ, O.G.; *Mortalidade por causas externas – Análise exploratória espacial, Região Sudeste/Brasil*. in anais XI Encontro Nacional de Estudos Populacionais, Caxambu, 19 a 23 de outubro de 1998, CD ROM.

CLARKE, A. B. e DISNEY, R. L.; *Probabilidade e Processos Estocásticos*. Rio de Janeiro. Livros Técnicos e Científicos Editora, 1979.

GERARDI, Lúcia Helena de Oliveira. *Quantificação em Geografia*, São Paulo: DIFEL, 1981

IBGE. *Censo Demográfico Brasileiro de 1991*.<http://www.ibge.gov.br/>, 18 de março de 1999.

INSTITUTO NACIONAL de PESQUISAS ESPACIAIS (INPE); *Manual do SPRING – v 2.0.4*. [online]. <http://www.dpi.inpe.br/~spring/>, 14 de março de 1998.

SMITH, T. ; PEUQUET, D.; MENON, S.; AGARWALL, P. KBGIS-II, A Knowledge-based Geographical Information Systems. *International Journal of Geographical Information Systems*. Vol. 1, nº 2, 1987. Pp. 149-172.

TELES, M. M. F.; *Estudos Sobre os Sistemas de Informação Geográfica*. Relatório Final de Iniciação científica – PIBIC/CNPq-UFPB, 1999.

WEISS. N. A. e YOSELOFF, M. L.; *Matemática Finita*. Rio de Janeiro. Ed. Guanabara Dois, 1978.

## *ANEXOS*

## Cronograma do projeto

Atividades	1	2	3	4	5	6	7	8	9	10	11	12
1- Princípio de análise espacial em Sistemas de informação Geográfica	1	2	3	4								
2- Implantação das novas variáveis sobre o banco de dados já existente					5	6	7	8				
3- Criação de mapas e resultados pertinentes às variáveis presentes no banco de dados									9	10	11	12