

**MINISTÉRIO DA SAÚDE
FUNDAÇÃO OSWALDO CRUZ
ESCOLA NACIONAL DE SAÚDE PÚBLICA**

**AVALIAÇÃO DO MÉTODO
GEOESTATÍSTICO NO ESTUDO DA
DISTRIBUIÇÃO ESPACIAL DA HEPATITE A**

Autor: Roberto de Andrade Medronho

Orientador: Cláudio José Struchiner

Co-orientadores: Cláudio Bettini e Jorge Xavier da Silva

**Rio de Janeiro, RJ - Brasil
Abril de 1999**

AVALIAÇÃO DO MÉTODO GEOESTATÍSTICO NO ESTUDO DA DISTRIBUIÇÃO ESPACIAL DA HEPATITE A

Roberto de Andrade Medronho

**TESE SUBMETIDA AO CORPO DOCENTE DA ESCOLA NACIONAL DE
SAÚDE PÚBLICA DA FUNDAÇÃO OSWALDO CRUZ COMO PARTE DOS
REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR
EM SAÚDE PÚBLICA**

Aprovada por:

Prof. Cláudio José Struchiner, Ph.D.

Prof. Flávio Fonseca Nobre, Ph.D.

Prof. Cláudia Medina Coeli, D.Sc.

Prof. Evandro Freire Coutinho, D.Sc.

Prof. Christóvam Barcellos, D.Sc.

Rio de Janeiro, RJ - Brasil

Abril de 1999

Medronho, Roberto de Andrade

Avaliação do método geoestatístico no estudo da distribuição espacial da hepatite A. Rio de Janeiro, FIOCRUZ, 1999

XV, 146 p.

Tese (Doutorado em Saúde Pública) FIOCRUZ. Escola Nacional de Saúde Pública

1. Hepatite A. 2. Distribuição espacial 3. Mapas. 4. Análise estatística. 5. Modelagem.

I. Fundação Oswaldo Cruz. II. Título.

À querida Claudia
Aos doces Adriana, Rodrigo e Renata
Com vocês aprendo todos dias os mais belos ensinamentos

Agradecimentos

Ao Cláudio José Struchiner, que com sua paciência e proficiência contribuiu de forma inestimável para esta tese chegar a bom termo; meu profundo respeito e admiração.

Ao Cláudio Bettini, pelos competentes e valiosos ensinamentos em geoestatística que viabilizaram esta tese; foram horas e horas muito prazerosas de paciente orientação. Sua humildade, sinceridade e dedicação me comoveram; ganhei um amigo.

Ao Jorge Xavier da Silva, pelos ensinamentos em geoprocessamento e entusiasmo contagiante.

À Marília Bernardes Marques, pela primorosa ajuda nos primeiros passos desta longa estrada do doutorado.

À Claudia Caminha Escosteguy, companheira de todas as horas, pelas severas e oportunas críticas e revisão minuciosa do texto.

À Diana Maul de Carvalho, pelo grande apoio, estímulo e críticas à tese.

Ao Basílio de Bragança Pereira, pelo estímulo, dicas atuais e críticas à tese.

Ao Ronir Raggio Luiz, pela solidariedade, companheirismo e críticas à tese.

Aos companheiros orientandos do Cláudio Struchiner – Tania Zdenka Guillén de Torres, Haroldo José de Matos, Maria Tereza Serrano Barbosa e Mônica Edelenyr, pelas críticas sempre construtivas, inteligentes e o apoio irrestrito.

À Liz Maria de Almeida, pelo grande esforço para viabilizar o PAISQUA e apoio à tese.

À toda equipe do PAISQUA, pelo intenso trabalho desenvolvido, possibilitando a utilização de parte dos dados do projeto nesta tese.

À CAPES, pelo apoio financeiro através da bolsa de doutorado.

À FAPERJ, pelo apoio financeiro ao PAISQUA.

Aos colegas do Núcleo de Estudos de Saúde Coletiva (NESC/UFRJ), pelo apoio, carinho e compreensão em minha ausência.

Aos colegas do Serviço de Epidemiologia do Hospital dos Servidores do Estado (HSE/MS), pelo apoio, carinho e compreensão em minha ausência.

SUMÁRIO

LISTA DE TABELAS	v
LISTA DE FIGURAS	vii
LISTA DE ANEXOS	xi
RESUMO	xiv
ABSTRACT	xv
I. INTRODUÇÃO	1
II. OBJETIVOS	7
III. METODOLOGIA	8
1. Geoestatística	8
1.1. Introdução	8
1.2. Teoria da variável regionalizada	8
1.3. Estacionariedade	9
1.3.1. Estacionariedade estrita	10
1.3.2. Estacionariedade intrínseca	10
1.3.3. Estacionariedade de segunda ordem	11
1.3.4. Quase- Estacionariedade	11
1.4. Variograma	12
1.5. Verificação de eixos de continuidade espacial	15
1.6. Modelagem do variograma	17
1.7. Método de estimativa espacial – <i>krigagem</i>	19
1.7.1. <i>Krigagem</i> simples	20
1.7.2. <i>Krigagem</i> ordinária	20
1.7.3. <i>Krigagem</i> indicadora	21
2. Região de estudo	24
3. Inquérito domiciliar e soroepidemiológico	25
4. Processamento e análise dos dados	27
IV. RESULTADOS	31
1. Análise exploratória convencional	31
1.1. Idade	31

SUMÁRIO (continuação)

1.2. Variáveis sócio-econômicas, domiciliares e peridomiciliares	31
1.3. Soroprevalência de anti-HAV	35
2. Análise exploratória espacial	37
2.1. Descrição da região	37
2.2. Variografia	43
2.3. Modelagem	53
2.4. <i>Krigagem</i>	64
2.5. Validação cruzada	80
V. DISCUSSÃO	92
VI. CONCLUSÕES	97
VII. CONSIDERAÇÕES FINAIS	98
REFERÊNCIAS BIBLIOGRÁFICAS	99
ANEXOS	106

Lista de Tabelas

Tabela 1:	Função de distribuição de probabilidade da variável aleatória local $Z(s)$	27
Tabela 2:	Distribuição dos indivíduos examinados segundo idade, por setor censitário, Parque Fluminense, Duque de Caxias, Rio de Janeiro, 1996	31
Tabela 3:	Parâmetros da distribuição de variáveis domiciliares relevantes por setor censitário, Parque Fluminense, Duque de Caxias, Rio de Janeiro, 1996	33
Tabela 4:	Variáveis domiciliares com diferenças significativas entre os setores censitários, Parque Fluminense, Duque de Caxias, Rio de Janeiro, 1996	34
Tabela 5:	Distribuição da soroprevalência de anti-HAV segundo idade, setores censitários 111 e 112, Parque Fluminense, Duque de Caxias, Rio de Janeiro, 1996	35
Tabela 6:	Parâmetros da distribuição de variáveis domiciliares relevantes, setores censitários 111 e 112, Parque Fluminense, Duque de Caxias, Rio de Janeiro, 1996	36
Tabela 7:	Variáveis domiciliares com diferenças significativas entre os soropositivos e soronegativos, setores censitários 111 e 112, Parque Fluminense, Duque de Caxias, Rio de Janeiro, 1996	37
Tabela 8:	Valores estimados da probabilidade de soropositividade e respectiva variância em cada setor censitário	54
Tabela 9:	Parâmetros de modelagem com correção de anisotropia geométrica, setor censitário 111	56
Tabela 10:	Parâmetros de modelagem com correção de anisotropia geométrica, setor censitário 112	61
Tabela 11:	Valores observados vs. valores estimados pelo modelo omnidirecional no <i>cutoff</i> 0,25 - Setor censitário 111 e 112	81

Lista de Tabelas (continuação)

Tabela 12: Valores observados vs. valores estimados pelo modelo omnidirecional no <i>cutoff</i> 0,20 - Setor censitário 111	85
Tabela 13: Valores observados vs. valores estimados pelo modelo anisotrópico no <i>cutoff</i> 0,20 - Setor censitário 111	85
Tabela 14: Valores observados vs. valores estimados pelo modelo com <i>hole effect</i> no <i>cutoff</i> 0,20 - Setor censitário 111	85
Tabela 15: Valores observados vs. valores estimados pelo modelo omnidirecional no <i>cutoff</i> 0,30 – Setor censitário 112	90
Tabela 16: Valores observados vs. valores estimados pelo modelo anisotrópico no <i>cutoff</i> 0,30 - Setor censitário 112	90
Tabela 17: Valores observados vs. valores estimados pelo modelo com <i>hole effect</i> no <i>cutoff</i> 0,30 - Setor censitário 112	90

Lista de Figuras

Figura 1:	Múltiplas realizações de uma variável regionalizada e a hipótese de estacionariedade	10
Figura 2:	Exemplo de um variograma com seus componentes	13
Figura 3:	Parâmetros requeridos para a construção do variograma experimental	15
Figura 4:	Exemplo de dois parâmetros necessários para a definição de anisotropia geométrica de um variograma em 2D	16
Figura 5:	Distribuição condicional local no ponto s	23
Figura 6:	<i>Box plot</i> da renda familiar em salários mínimos e número de pontos de água por domicílio	33
Figura 7:	Mapa de faixas altimétricas, Parque Fluminense, Duque de Caxias ...	38
Figura 8:	Proporção de chefes com renda menor ou igual a 2 salários mínimos, Setores Censitários, Parque Fluminense, Duque de Caxias	38
Figura 9:	Taxas de prevalência padronizadas por idade em menores de 10 anos por setores censitários, Parque Fluminense, Duque de Caxias, RJ	39
Figura 10:	Aerofoto dos setores censitários 111 e 112, Parque Fluminense, Duque de Caxias, Rio de Janeiro, jan/96	40
Figura 11:	Mapa com a localização das crianças examinadas, setores censitários 111 e 112, Parque Fluminense, Duque de Caxias, Rio de Janeiro	42
Figura 12:	Mapa de variograma dos setores censitários 111 e 112	43
Figura 13:	Mapa de variograma do setor censitário 111	44
Figura 14:	Mapa de variograma do setor censitário 112	45
Figura 15:	Mapa de contorno dos variogramas, setores censitários 111 e 112	46
Figura 16:	Mapa de variograma em 3D, setores censitários 111 e 112	47
Figura 17:	Mapa de contorno dos variogramas, setor censitário 111	48
Figura 18:	Mapa de variograma em 3D, setor censitário 111	49
Figura 19:	Mapa de contorno dos variogramas, setor censitário 112	50
Figura 20:	Mapa de variograma em 3D, setor censitário 112	51
Figura 21:	Diagrama de rosa do setor censitário 111	52
Figura 22:	Diagrama de rosa do setor censitário 112	52

Lista de Figuras (continuação)

Figura 23: Modelagem do semivariograma omnidirecional dos setores censitários 111 e 112	55
Figura 24: Modelagem do semivariograma omnidirecional do setor censitário 111	56
Figura 25: Modelagem do semivariograma do setor censitário 111 no azimute de 137°	57
Figura 26: Modelagem do semivariograma do setor censitário 111 no azimute de 47°	58
Figura 27: Modelagem do semivariograma do setor 111 na direção de 47° com ajuste do <i>hole effect</i>	59
Figura 28: Modelagem do semivariograma omnidirecional do setor censitário 112	60
Figura 29: Modelagem do semivariograma do setor censitário 112 no azimute de 53°	62
Figura 30: Modelagem do semivariograma do setor censitário 112 no azimute de 143°	62
Figura 31: Modelagem do semivariograma do setor 112 na direção de 143° com ajuste do <i>hole effect</i>	64
Figura 32: Mapa com os valores estimados pela <i>krigagem</i> , modelo omnidirecional, setores censitários 111 e 112	65
Figura 33: Mapa com as variâncias dos valores estimados pela <i>krigagem</i> , modelo omnidirecional, setores censitários 111 e 112	67
Figura 34: Mapa com os valores estimados pela <i>krigagem</i> , modelo omnidirecional, setor censitário 111	68
Figura 35: Mapa com as variâncias dos valores estimados pela <i>krigagem</i> , modelo omnidirecional, setor censitário 111	69
Figura 36: Mapa com os valores estimados pela <i>krigagem</i> , modelo anisotrópico, setor censitário 111	70

Lista de Figuras (continuação)

Figura 37: Mapa com as variâncias dos valores estimados pela <i>krigagem</i> , modelo anisotrópico, setor censitário 111	71
Figura 38: Mapa com os valores estimados pela <i>krigagem</i> , modelo com <i>hole effect</i> , setor censitário 111	72
Figura 39: Mapa com as variâncias dos valores estimados pela <i>krigagem</i> , modelo com <i>hole effect</i> , setor censitário 111	73
Figura 40: Mapa com os valores estimados pela <i>krigagem</i> , modelo omnidirecional, setor censitário 112	74
Figura 41: Mapa com as variâncias dos valores estimados pela <i>krigagem</i> , modelo omnidirecional, setor censitário 112	75
Figura 42: Mapa com os valores estimados pela <i>krigagem</i> , modelo anisotrópico, setor censitário 112	76
Figura 43: Mapa com as variâncias dos valores estimados pela <i>krigagem</i> , modelo anisotrópico, setor censitário 112	77
Figura 44: Mapa com os valores estimados pela <i>krigagem</i> , modelo com <i>hole effect</i> , setor censitário 112	78
Figura 45: Mapa com as variâncias dos valores estimados pela <i>krigagem</i> , modelo com <i>hole effect</i> , setor censitário 112	79
Figura 46: Histograma dos valores estimados pela validação cruzada para o modelo omnidirecional, setores censitários 111 e 112	80
Figura 47: Curva ROC dos valores estimados pela validação cruzada para o modelo omnidirecional, setores censitários 111 e 112	81
Figura 48: Histograma dos valores estimados pela validação cruzada para o modelo omnidirecional, setor censitário 111	82
Figura 49: Histograma dos valores estimados pela validação cruzada para o modelo anisotrópico, setor censitário 111	83
Figura 50: Histograma dos valores estimados pela validação cruzada para o modelo com ajuste do <i>hole effect</i> , setor censitário 111	83

Lista de Figuras (continuação)

Figura 51: Curva ROC para o modelo omnidirecional do setor censitário 111	84
Figura 52: Curva ROC para o modelo com anisotropia geométrica do setor censitário 111	84
Figura 53: Curva ROC para o modelo com <i>hole effect</i> do setor censitário 111	85
Figura 54: Curva ROC dos valores estimados pela validação cruzada para os modelos omnidirecional, anisotrópico e com ajuste do <i>hole effect</i> , setor censitário 111	86
Figura 55: Histograma dos valores estimados pela validação cruzada para o modelo omnidirecional, setor censitário 112	87
Figura 56: Histograma dos valores estimados pela validação cruzada para o modelo anisotrópico, setor censitário 112	87
Figura 57: Histograma dos valores estimados pela validação cruzada para o modelo com ajuste do <i>hole effect</i> , setor censitário 112	88
Figura 58: Curva ROC para o modelo omnidirecional do setor censitário 112	88
Figura 59: Curva ROC para o modelo com anisotropia geométrica do setor censitário 112	89
Figura 60: Curva ROC para o modelo com <i>hole effect</i> do setor censitário 112	89
Figura 61: Curva ROC dos valores estimados pela validação cruzada para os modelos omnidirecional, anisotrópico e com ajuste do <i>hole effect</i> , setor censitário 112	91

Lista de Anexos

A.	Inquérito epidemiológico – questionário	107
B.1.	Arquivo de parâmetros do <i>varmap</i> (programa de construção de mapa de variograma) para os setores censitários 111 e 112 em conjunto	118
B.2.	Arquivo de parâmetros do <i>varmap</i> (programa de construção de mapa de variograma) para o setor censitário 111	119
B.3.	Arquivo de parâmetros do <i>varmap</i> (programa de construção de mapa de variograma) para o setor censitário 112	120
B.4.	Arquivo de parâmetros do <i>gamv</i> (programa de construção de variogramas) para o setor censitário 111 e 112 em conjunto, omnidirecional	121
B.5.	Arquivo de parâmetros do <i>gamv</i> (programa de construção de variogramas) para o setor censitário 111, omnidirecional	122
B.6.	Arquivo de parâmetros do <i>gamv</i> (programa de construção de variogramas) para o setor censitário 111, eixos de maior e menor continuidade espacial	122
B.7.	Arquivo de parâmetros do <i>gamv</i> (programa de construção de variogramas) para o setor censitário 112, omnidirecional	123
B.8.	Arquivo de parâmetros do <i>gamv</i> (programa de construção de variogramas) para o setor censitário 112, eixos de maior e menor continuidade espacial	123
B.9.	Arquivo de parâmetros do <i>kt3d</i> (programa de <i>krigagem</i>) para os setores censitários 111 e 112 em conjunto, modelo omnidirecional	124
B.10.	Arquivo de parâmetros do <i>kt3d</i> (programa de <i>krigagem</i>) para o setor censitário 111, modelo omnidirecional	125
B.11.	Arquivo de parâmetros do <i>kt3d</i> (programa de <i>krigagem</i>) para o setor censitário 111, modelo anisotrópico	126
B.12.	Arquivo de parâmetros do <i>kt3d</i> (programa de <i>krigagem</i>) para o setor censitário 111, modelo com <i>hole effect</i>	127

Lista de Anexos (continuação)

B.13.	Arquivo de parâmetros do <i>kt3d</i> (programa de <i>krigagem</i>) para o setor censitário 112, modelo omnidirecional	128
B.14.	Arquivo de parâmetros do <i>kt3d</i> (programa de <i>krigagem</i>) para o setor censitário 112, modelo anisotrópico	129
B.15.	Arquivo de parâmetros do <i>kt3d</i> (programa de <i>krigagem</i>) para o setor censitário 112, modelo com <i>hole effect</i>	130
B.16.	Arquivo de parâmetros do <i>kt3d</i> (opção de validação cruzada) para os setores censitários 111 e 112 em conjunto, modelo omnidirecional	131
B.17.	Arquivo de parâmetros do <i>kt3d</i> (opção de validação cruzada) para o setor censitário 111, modelo omnidirecional	132
B.18.	Arquivo de parâmetros do <i>kt3d</i> (opção de validação cruzada) para o setor censitário 111, modelo anisotrópico	133
B.19.	Arquivo de parâmetros do <i>kt3d</i> (opção de validação cruzada) para o setor censitário 111, modelo com <i>hole effect</i>	134
B.20.	Arquivo de parâmetros do <i>kt3d</i> (opção de validação cruzada) para o setor censitário 112, modelo omnidirecional	135
B.21.	Arquivo de parâmetros do <i>kt3d</i> (opção de validação cruzada) para o setor censitário 112, modelo anisotrópico	136
B.22.	Arquivo de parâmetros do <i>kt3d</i> (opção de validação cruzada) para o setor censitário 112, modelo com <i>hole effect</i>	137
C.1.	Programa para suavização para os valores dos variogramas resultantes do <i>varmap</i> (exemplifica-se apenas o programa referente ao setor censitário 111)	138
C.2.	Programa para traçar a elipse com os eixos de maior e menor continuidade espacial definidos pela variografia em todas as direções (exemplo para o setor censitário 111)	140

Lista de Anexos (continuação)

C.3.	Programa para proceder a modelagem dos diversos variogramas omnidirecionais e para a maior e menor continuidade espacial de cada setor (exemplo da modelagem do eixo de 137° para o setor censitário 111)	142
C.4.	Programa para proceder a modelagem do <i>hole effect</i> (exemplo da modelagem do eixo de 57° - menor continuidade - para o setor censitário 111)	144
C.5.	Programa para construção da curva ROC e identificação do <i>cutoff</i> com equilíbrio entre as melhores sensibilidade e especificidade conjuntas (exemplo para o modelo omnidirecional do setor censitário 111)	146

RESUMO

Esta tese avalia o uso das técnicas geoestatísticas para a análise, modelagem e estimativa espacial no campo da Epidemiologia. Para tal, foram analisadas 410 crianças menores de 10 anos residentes nos setores censitários 111 e 112 do segundo distrito do município de Duque de Caxias, Rio de Janeiro, Brasil, que participaram de um inquérito soropidemiológico para hepatite A. A soroprevalência global de anticorpos contra o vírus da hepatite A foi de 24,6%, sendo maior no setor 112 que no setor 111 (29,5% vs. 18,1% - $p < 0,008$). A variografia mostrou de forma consistente eixos de maior continuidade espacial (anisotropia geométrica) da soroprevalência para cada um dos setores, que coincidiram com características geo-ambientais e urbanas propícias à propagação da hepatite A, sugerindo formas distintas de propagação nos setores. Além disso, detectou também o fenômeno de *hole effect*. Foram construídos três modelos para cada setor (omnidirecional, com correção da anisotropia geométrica e com correção do *hole effect*). O procedimento de estimativa espacial usado foi o da *krigagem* indicadora, resultando em mapas distintos segundo o tipo de modelo aplicado, sendo que as correções da anisotropia geométrica e do *hole effect* produziram mapas mais consistentes com a distribuição espacial do fenômeno. Os diferentes modelos foram avaliados através do procedimento de validação cruzada. O resultado obtido pela validação cruzada para cada modelo foi discretizado em diversos pontos de corte, construindo-se curvas ROC (*Receiver Operator Characteristic*) para cada modelo. Estas curvas não mostraram diferença significativa entre os modelos. Identificou-se também o ponto de corte que continha um equilíbrio entre as melhores sensibilidade e especificidade conjuntas de cada modelo, mensurando-se a proporção de casos classificados corretamente neste ponto, para o qual observou-se que o ajuste dos modelos não foi satisfatório. Este estudo discute algumas vantagens que podem ser obtidas através da utilização de técnicas de análise geoestatística à área de Epidemiologia, onde sua incorporação é ainda incipiente, assim como possíveis limitações para seu uso.

ABSTRACT

This study evaluates the use of geostatistics technics for the spatial analysis, modelling and estimation in Epidemiology. It was studied a group of 410 smaller than 10 years old children who participated in a hepatitis A seroepidemiological survey; they lived in the census tract 111 and 112 of the second district of the municipal district of Duque de Caxias, Rio de Janeiro, Brazil. The global serum prevalence of antibodies against hepatitis A virus was 24,6%, being larger in census tract 112 than in census tract 111 (29,5% vs. 18,1% - $p < 0,008$). The variography consistently showed axes of larger spatial continuity (geometric anisotropy) of the serum prevalence for each one of the sections; the axes coincided with favorable geo-environmental and urban characteristics to the propagation of the hepatitis A and they suggested different forms propagation in both census tracts. The variography also detected the hole effect phenomenon. Three models were built for each section (omnidirectional, geometric anisotropy correction and hole effect correction). The indicator kriging was used to proceed the spatial estimation, resulting in different maps according to the specific applied model; the geometric anisotropy and the hole effect corrections produced maps which were more consistent with the spatial distribution of the phenomenon. The different models were appraised through the cross validation procedure. The result obtained by the cross validation for each model was divided in several cutoffs. ROC (Receiver Operator Characteristic) curves were built for each model, showing no significant difference among the models. It was identified the cutoff for the best together sensibility and especificity, and the proportion of correctly classified cases was measured in this cutoff. It was observed that the models fitness was not satisfactory. This study discusses some advantages that can be obtained with the use of geostatistics analysis in Epidemiology, where its incorporation is still incipient, as well as, possible limitations for its use.

I. INTRODUÇÃO:

A hepatite A é uma doença de transmissão fecal-oral. Água, alimentos e moluscos constituem potenciais fontes de contaminação. Entre os fatores de risco estão incluídos os comunicantes de doentes com hepatite A, trabalhadores de centros de cuidados diários, viagem internacional, exposição recente a água ou alimentos contaminados durante um surto, atividade homossexual e uso de droga injetável.

O agente etiológico da hepatite A é um vírus do gênero *Heparnavirus*, da família dos *Picornavirus*; tem forma esférica com diâmetro médio de 27 nm (Bier, 1994).

A hepatite A tem um período de incubação de cerca de 4 semanas, variando entre 15 e 45 dias. Embora a replicação viral ocorra somente no fígado, o vírus pode ser detectado no sangue, bile e fezes durante o período de incubação tardio e na fase pré-ictérica da doença aguda. Anticorpos contra o vírus da hepatite A (anti-HAV) do tipo IgM são detectados entre a terceira e a décima terceira semana, enquanto que os anti-HAV do tipo IgG são detectados a partir da terceira semana, permanecendo no soro indefinidamente. A imunidade provocada pela doença é permanente.

O espectro clínico da infecção pelo vírus da hepatite A (HAV) é amplo, podendo variar desde uma forma assintomática ou subclínica (cerca de 85% dos casos) até uma forma grave e incapacitante. Não existe tratamento específico contra o HAV, nem evidência científica de benefícios com o uso de dieta. A taxa de letalidade é inferior a 0,1% (OPAS, 1987).

Existe grande dificuldade no estudo da incidência da doença devido à elevada subnotificação e ao amplo espectro clínico; assim, sua epidemiologia é melhor definida através da identificação sorológica de anti-HAV (Melnick, 1995).

A partir da década de 70, após o desenvolvimento de técnicas laboratoriais que permitiram a identificação do anticorpo contra o vírus da hepatite A, os estudos de soroprevalência tornaram-se freqüentes. Desde então, trabalhos publicados em diversas regiões revelaram características dessa infecção, tais como grande flutuação da prevalência de anticorpos anti-HAV nas populações estudadas, aumento da soropositividade com a idade, forte correlação com o nível sócio-econômico e as condições de higiene, e inexistência de portadores crônicos (Almeida, 1997).

As populações mais atingidas nos países em desenvolvimento são as crianças, enquanto que nos países desenvolvidos, os adultos jovens. Níveis de higiene e saneamento afetam essa distribuição. O sexo não parece exercer influência no risco de infecção por HAV, exceto em condições excepcionais tais como grupamentos masculinos que trabalham com esgoto, ou que mantêm práticas homossexuais (Leentvaar-Kuijpers *et al*, 1995).

Em um estudo de 540 amostras de soro de crianças e adultos residentes na cidade de São Paulo, Pannuti *et al* (1985) mostraram que as prevalências de anti-HAV variaram de acordo com o nível sócio-econômico, sobretudo na infância. Na classe estratificada como baixa, encontrou-se o extremo de 100% de soropositividade a partir dos 8 anos de idade; já na classe média, valores próximos a 100% só foram atingidos após os 21 anos.

Queiroz *et al* (1995a) encontraram uma soroprevalência variando entre 80,0% e 92,2% nos jovens de rua de Goiânia, em inquérito soro-epidemiológico realizado entre 1991/1992, em 397 indivíduos de 7 a 21 anos trabalhando na ruas ou vivendo em instituições.

Queiroz *et al* (1995b) encontraram em 310 crianças em centros de cuidados diários, com idade entre 3 meses a 9 anos, uma soropositividade para anticorpos IgG/IgM em 69,7% das crianças, sendo que 60% destas entre 1 e 3 anos. Entre os 10 centros analisados, a prevalência de anti-HAV do tipo IgM foi de 3,2%. Os autores verificaram que o risco de infecção aumentava com o período de atendimento nos centros de cuidados diários, ou seja, o risco para crianças atendidas por um ano ou mais era 4,7 vezes maior, quando comparado com crianças com 1 mês de atendimento (IC 95% = 2,3 - 9,9).

Gaspar *et al* (1996) realizaram um inquérito de soroprevalência em 699 amostras de creches e escolas primárias e secundárias em uma área periférica do município do Rio de Janeiro. A soroprevalência do anti-HAV do tipo IgG aumentou com a idade, variando de 0% em crianças até 4 anos, até 61,5% em maiores de 14 anos. Segundo os próprios autores, este padrão era significativamente diferente do observado em um outro estudo realizado em 1980, em que a soropositividade em menores de 5 anos já atingia 100%.

Segundo Koff (1995), cerca de 40% dos indivíduos nos EUA apresentam soropositividade para anticorpos contra o vírus da hepatite A, sendo que as taxas aumentam com a idade, talvez refletindo um efeito coorte devido a pessoas infectadas em períodos passados quando a infecção era mais comum. Entretanto, a incidência de hepatite A tem diminuído nos países desenvolvidos em consequência da melhoria das instalações sanitárias. Amela *et al* (1995) demonstraram uma redução na transmissão da infecção pelo HAV associada ao processo de urbanização ocorrido em Madri e um conseqüente aumento de suscetíveis, com a possibilidade de ocorrência de epidemias explosivas. Neste estudo observou-se, também, um aumento na idade média de apresentação da doença, aumentando assim, a possibilidade de infecção sintomática.

Bolumar *et al* (1995), estudando a soroprevalência de HAV em Valência (Espanha), encontraram um aumento com a idade (*odds ratio* - OR > 50 anos = 69,8; IC 95% = 26,5 - 183,4) e uma diminuição com o nível educacional mais elevado (OR formação universitária = 0,2; IC 95% = 0,1 - 0,5).

Maguire *et al* (1995) realizaram um estudo tipo caso-controle para determinar fatores de risco associados à infecção pelo vírus da hepatite A na Inglaterra. Os autores encontraram um risco mais elevado nos comunicantes domiciliares (OR = 19,8; IC 95% = 4,87 - 80,6), em compartilhar domicílio com criança entre 3 e 10 anos (OR = 1,57; IC 95% = 1,1 - 2,22), e para os que realizaram viagem ao exterior (OR = 19,8; IC 95% = 4,87 - 80,6).

Em um estudo para avaliar a incidência e fatores de risco para a hepatite A na Itália, Mele *et al* (1997) concluíram que a doença está associada principalmente a consumo de alimentos e recomendavam a vacinação de viajantes para áreas endêmicas.

O principal veículo de transmissão da hepatite A nos países em desenvolvimento é a água contaminada. Estudos epidemiológicos demonstram uma forte associação entre nível sócio-econômico baixo e condições de higiene e saneamento precárias com soroprevalência elevada de anticorpos anti-HAV. Por outro lado, nos países desenvolvidos, a melhoria nas instalações sanitárias foi acompanhada de uma redução na incidência de hepatite A. Assim, a análise da soroprevalência de anticorpos anti-HAV pode ser usada como um indicador das condições sanitárias da região em estudo. Neste sentido, o *Projeto de Avaliação dos Impactos do Programa de Despoluição da*

Baía de Guanabara Sobre as Condições de Saúde e Qualidade de Vida (PAISQUA) realizou um inquérito soro-epidemiológico para anticorpos contra o vírus da hepatite A (NESC, 1995), na localidade denominada pelo *Programa de Despoluição da Baía de Guanabara* (PDBG) de Setor Parque Fluminense, que abrange parte do segundo distrito do município de Duque de Caxias, Rio de Janeiro. Este projeto foi desenvolvido pelo Núcleo de Estudos de Saúde Coletiva da Universidade Federal do Rio de Janeiro (NESC/UFRJ) e financiado pela Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ).

No presente estudo foram analisados os resultados referentes ao estudo piloto realizado pelo PAISQUA nos setores censitários 111 e 112, utilizando-se técnicas geoestatísticas para a análise da distribuição, modelagem e estimativa espacial da soroprevalência da hepatite A. Estas técnicas, apesar de terem sido introduzidas a partir da década de 60 por Matheron (1971), somente mais recentemente vêm sendo aplicadas na área da saúde.

Ressalta-se que os estudos epidemiológicos lidam com as categorias relacionadas a *tempo, lugar e pessoa* desde há muito tempo (MacMahon & Pugh, 1970; Lilienfeld & Lilienfeld, 1980). Ainda no século XIX, o clássico estudo de John Snow (1990) utilizou técnicas de mapeamento para analisar a epidemia de cólera ocorrida em 1854 na região de Soho, Londres. Ele demonstrou uma associação espacial entre mortes por cólera e suprimento de água por diferentes bombas públicas de abastecimento, identificando assim a origem da epidemia, mesmo sem conhecer seu agente etiológico. Entretanto, a despeito do desenvolvimento das técnicas quantitativas para o estudo do *tempo* (séries temporais) no campo da Epidemiologia, o mesmo não ocorreu em relação à categoria *espaço*. Com o avanço recente das técnicas de análise espacial, tais como geoprocessamento e geoestatística, aplicadas a diversos campos do conhecimento, é importante analisar a contribuição que estas técnicas possam oferecer à Epidemiologia.

Análise espacial em saúde refere-se ao estudo quantitativo da distribuição das doenças ou serviços de saúde, no qual o objeto de estudo é definido geograficamente (Gesler, 1986). Diversas técnicas de análise espacial vêm sendo utilizadas no campo da saúde (Marshall, 1991). O campo específico da análise de dados espaciais envolve dados que representam o desfecho de um processo operando no espaço; métodos estatísticos são utilizados para descrever e eventualmente explicar tais dados,

freqüentemente buscando relações com outros dados espacialmente definidos (Gatrell & Bailey, 1996; Bailey & Gatrell, 1995).

Segundo Gatrell & Bailey (1996), os métodos para análise espacial podem ser divididos em: visualização, análise exploratória de dados e modelagem. Neste sentido, pode-se distinguir claramente que os procedimentos utilizados para executar a análise espacial não se resumem simplesmente ao mapeamento dos eventos, exigindo também a necessidade de se adicionar um novo significado aos dados. Diversos procedimentos são utilizados para a análise espacial, notadamente, as técnicas de geoprocessamento, de análise geoestatística, de dados em reticulado (*lattice*) e de padrões de ponto - *point patterns* (Cliff & Ord, 1981; Cressie, 1993; Kaluzny *et al*, 1996).

O uso das técnicas de geoprocessamento vem aumentando cada vez mais no planejamento, monitoramento e avaliação das ações de saúde, além de constituir em uma importante ferramenta para análise das relações entre o ambiente e as questões relacionadas à saúde (Barcellos & Bastos, 1996). Especificamente, no campo da Epidemiologia, o geoprocessamento vem sendo utilizado na análise da dinâmica espacial das doenças em suas relações com o ambiente, na avaliação da situação de saúde de uma região, na identificação de regiões e grupos de alto risco (Castillo-Salgado, 1996), sendo também um importante instrumento no apoio às atividades de vigilância epidemiológica e planejamento de ações de prevenção e controle de doenças (Medronho, 1995), especialmente no contexto atual, onde ocorre o aparecimento de novas doenças assim como a emergência de diversas outras já existentes (Clarke *et al*, 1996).

Os recentes avanços na área de computação relacionados aos equipamentos e *softwares* de Sistema de Informação Geográfica (GIS, de *Geographic Information System*) tornaram possível interagir diretamente com grandes bases de dados espaciais e obter, quase que instantaneamente, resultados para uma grande variedade de operações de GIS. Os métodos sofisticados de captura, armazenamento e demonstração de dados em GIS criaram uma demanda para novas ferramentas para executar a análise espacial em geral e a análise estatística espacial em particular (Anselin, 1996).

Diversos estudos epidemiológicos utilizaram GIS para analisar possíveis associações entre a distribuição geográfica de doenças e o ambiente. Exemplos recentes

podem ser citados nas áreas de: vigilância epidemiológica (Nobre *et al*, 1996; Glass *et al*, 1995; Richards, 1993), doenças de veiculação hídrica (Clarke *et al*, 1991; Xavier-da-Silva, *et al*, 1997), mapeamento de risco de dengue (Medronho, 1995), controle da malária (Soares-Filho & Sawyer, 1996; Brêtas & Bessa, 1996; Beck, *et al*, 1994), saúde ambiental (Cuthe *et al*, 1992), modelagem de exposição a campos magnéticos (Wartenberg *et al*, 1992), análise de políticas de controle doenças (Tempalsky, 1994).

Entretanto, a incorporação das técnicas de geoestatística na área da saúde tem-se dado de forma ainda pontual e esparsa, não contemplando os múltiplos aspectos da análise espacial necessários à compreensão do papel do espaço no processo saúde-doença. Assim, faz-se necessário avaliar de forma sistemática tais técnicas, no sentido de aprimorar o entendimento da dependência espacial nos estudos epidemiológicos, sendo esta a principal contribuição do presente estudo.

II. OBJETIVOS

Objetivo geral

Avaliar o uso das técnicas geoestatísticas no estudo da distribuição espacial da soropositividade para anticorpos contra o vírus da hepatite A em indivíduos menores de 10 anos nos setores censitários 111 e 112 do segundo distrito (Campos Elyseos) do município de Duque de Caxias, Rio de Janeiro, Brasil

Objetivos específicos

1. Proceder a análise exploratória espacial da ocorrência de casos soropositivos e soronegativos para anticorpos contra o antígeno da hepatite A.
2. Proceder a modelagem espacial através da variografia.
3. Estimar a probabilidade de ocorrência dos casos em locais não amostrados através do método de *krigagem* indicadora.

III. METODOLOGIA:

Este capítulo dividiu-se em quatro seções: a primeira referente à revisão das técnicas geoestatísticas, com ênfase naquelas que foram utilizadas no estudo; a segunda, referente à região de estudo, e as duas últimas referentes às técnicas utilizadas para coleta e análise dos dados.

1. Geoestatística:

1.1. Introdução:

Geoestatística pode ser definida como o “estudo de fenômenos que variam no espaço” (Olea, 1991), no tempo ou no espaço e tempo (Deutsch & Journel, 1998). A geoestatística tem como objetivo principal a análise e modelagem da variabilidade espacial de um fenômeno.

A análise estatística exploratória (Tukey, 1977) habitualmente utilizada, através do cálculo da média, desvio padrão, medidas de locação (mediana, quartis, valores máximo e mínimo), coeficientes de assimetria, coeficientes de variação, gráficos do tipo histograma e *box plot*, entre outros, não é suficiente para caracterizar a continuidade espacial, já que essas técnicas pressupõem que as observações sejam independentes. Assim, diferentes conjuntos de dados contendo valores iguais para as estatísticas habituais, podem, eventualmente, possuir comportamentos distintos no espaço (Medronho *et al*, 1997).

1.2. Teoria da variável regionalizada:

O estudo da variabilidade espacial de diversos fenômenos que ocorrem na superfície terrestre constitui-se, há muito tempo, em uma importante atividade para o homem (Burrough, 1990). No início do século, estes estudos baseavam-se na utilização de grandes quantidades de dados amostrais para descrever a distribuição espacial de um determinado fenômeno em estudo. Posteriormente, Danie G. Krige, trabalhando com dados de concentração de ouro, em 1951, na África do Sul, concluiu que somente a informação dada pela variância não era suficiente para explicar o fenômeno em estudo,

sendo necessário considerar a distância entre as observações (Camargo, 1997). Baseado nestas observações, Matheron (1971) formalizou a *teoria da variável regionalizada*. Esta teoria define como uma variável regionalizada, qualquer variável distribuída no espaço, sendo o conjunto de dados espaciais interpretados como a realização de um processo aleatório

$$\{Z(s), s \in D\}$$

onde D é um subconjunto de \mathfrak{R}^d (habitualmente, $D \subset \mathfrak{R}^2$), ou seja, o índice espacial s varia continuamente na região D (Cressie, 1993).

A variação espacial da variável regionalizada pode ser modelada usando um componente aleatório (flutuação aleatória local) e um determinístico (comportamento estrutural geral). A premissa básica é a de que valores de uma variável em posições próximas são correlacionados e que esta correlação diminui à medida que a distância entre estes valores aumenta (Valencia, 1999).

1.3. Estacionariedade:

Tendo em vista que a variável regionalizada $z(\cdot)$ é considerada uma realização particular do processo aleatório espacial $\{Z(s), s \in D\}$, ao se observar $z(\cdot)$ em n posições $\{s_1, \dots, s_n\} \subset \mathfrak{R}^d$, obtêm-se os dados espaciais $\{z(s_1), \dots, z(s_n)\}$. Estes dados são considerados como uma observação incompleta de uma realização particular do processo aleatório $Z(\cdot)$. Assim, o desenvolvimento da teoria da variável regionalizada é baseado no conjunto das variáveis aleatórias $\{Z(s_1), \dots, Z(s_n)\}$, sendo cada $Z(s_i)$, $i=1, \dots, n$, uma variável aleatória local (Valencia, 1999). Entretanto, uma importante restrição na maioria dos problemas práticos encontrados em geoestatística diz respeito ao fato de que os dados correspondem a uma única realização do processo aleatório, não sendo possível obter replicações. Assim, os dados são considerados como uma amostra de uma realização de um processo aleatório. Para tal, faz-se necessário fazer suposições de homogeneidade do fenômeno sob estudo em determinada região de referência D . Sob esta hipótese de homogeneidade, pode-se, por exemplo, considerar dois valores $z(s)$ e $z(s+\mathbf{h})$ para dois diferentes pontos s e $s+\mathbf{h}$ como sendo duas realizações diferentes do mesmo processo aleatório. A figura 1, a seguir, exemplifica o problema:

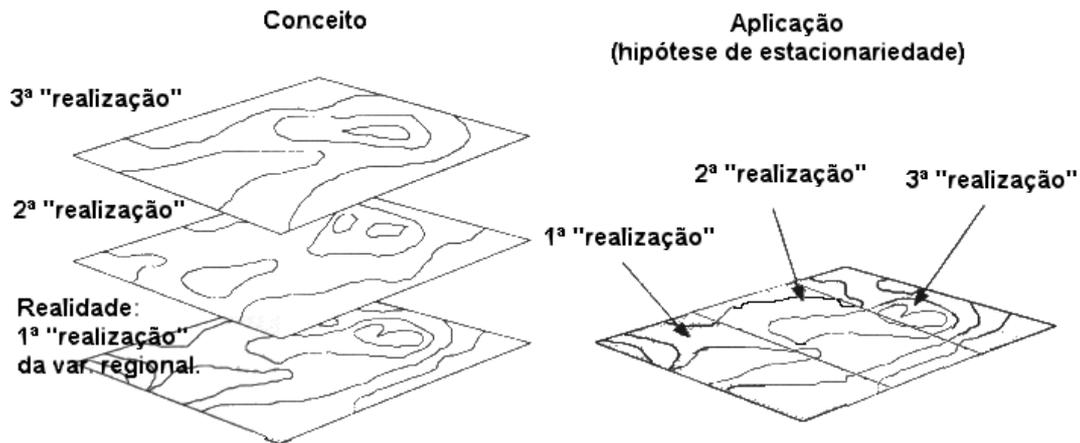


Figura 1 : Múltiplas realizações de uma variável regionalizada e a hipótese de estacionariedade (adaptado de Pannatier, 1996)

Embora a estacionariedade seja uma propriedade do processo aleatório e não da distribuição dos dados, na prática é necessário admitir uma hipótese de estacionariedade para que se possa proceder a inferência estatística, já que tanto a análise da continuidade espacial como da estimativa espacial dependem que os dados sejam provenientes de um processo aleatório estacionário (Valencia, 1999). Journel & Huijbregts (1978) definem vários tipos de processo aleatório estacionário:

1.3.1. Estacionariedade estrita:

Um processo aleatório é dito estacionário, em sentido estrito, quando sua lei de distribuição de probabilidades é invariante por translação. Entretanto, os princípios da geoestatística são baseados nas hipóteses restritivas de estacionariedade, que podem ser intrínseca e de segunda ordem. Neste caso:

$$E\{Z(s)\} = m(s) = E\{Z(s+\mathbf{h})\} = m = \text{constante}$$

1.3.2. Estacionariedade intrínseca:

O processo $Z(s)$ é intrinsecamente estacionário, quando:

- A expectância do processo é constante:

$$E\{Z(s)\} = m, \quad \forall s;$$

- A variância de $\{Z(s) - Z(s+\mathbf{h})\}$ depende somente do deslocamento \mathbf{h} :

$$\text{Var}\{Z(s) - Z(s+\mathbf{h})\} = E\{[Z(s) - Z(s+\mathbf{h})]^2\} = 2\gamma(\mathbf{h}), \quad \forall s;$$

A estacionariedade intrínseca é denominada também de hipótese intrínseca.

1.3.3. Estacionariedade de segunda ordem:

Um processo aleatório é dito estacionário de segunda ordem, quando:

- A expectância do processo é constante:

$$E\{Z(s)\} = m, \quad \forall s;$$

- Para cada par de variável aleatória $\{Z(s), Z(s+\mathbf{h})\}$, a covariância depende somente do vetor deslocamento \mathbf{h} :

$$C(\mathbf{h}) = E\{Z(s) \cdot Z(s+\mathbf{h})\} - m^2, \quad \forall s,$$

onde \mathbf{h} representa o vetor de coordenadas $(\mathbf{h}_u, \mathbf{h}_v, \mathbf{h}_w)$ no espaço tridimensional.

A hipótese de estacionariedade de segunda ordem é uma condição mais restritiva que a intrínseca. Além disso, a estacionariedade de segunda ordem implica na estacionariedade intrínseca, mas a recíproca não é verdadeira.

1.3.4. Quase-Estacionariedade:

As hipóteses de estacionariedade descritas anteriormente estabelecem que a função $m(\cdot)$ não depende da posição espacial, sendo, portanto, constante em toda região D . Entretanto, usualmente, os dados espaciais de uma variável regionalizada indicam o contrário, ou seja, que $m(\cdot)$ depende da posição espacial. Nestes casos, pode-se definir regiões vizinhas em D , onde as hipóteses de estacionariedade sejam válidas.

Ressalta-se que as hipóteses de estacionariedade são relativas à escala em que o fenômeno é observado e à quantidade de dados (Valencia, 1999). Na prática, em geoestatística, a estacionariedade não pode ser provada a partir dos dados, conseqüentemente, não se pode utilizar um teste estatístico para estacionariedade. Dessa forma, é necessário realizar diversas verificações empíricas nos dados para poder aceitar se uma das hipóteses de estacionariedade é razoável.

1.4. Variograma:

O variograma pode ser definido como a variância do incremento $[Z(s) - Z(s+\mathbf{h})]$ (Deutsch & Journel, 1998). Para um processo aleatório estacionário:

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

Obs: o termo variograma, representa por $2\gamma(\mathbf{h})$, e o termo semivariograma, representado por $\gamma(\mathbf{h})$ serão utilizados indistintamente neste texto.

O variograma é função somente do incremento \mathbf{h} , sendo um modelo da variabilidade da variável regionalizada à medida que a distância entre as posições espaciais aumenta. Constitui-se na principal medida utilizada em geoestatística para descrever a variabilidade espacial e compõe-se dos seguintes elementos:

- **alcance** (*range*) - \mathbf{a} : é o valor da distância máxima do variograma a partir da qual as amostras se tornam independentes (Almeida & Bettini, 1994);
- **patamar** (*sill*) - C_1 : valor do variograma correspondente ao seu alcance. A partir deste ponto, considera-se que não existe mais dependência espacial, já que a variância da diferença entre os pares de pontos - $\text{Var}\{Z(s) - Z(s+\mathbf{h})\}$ - torna-se invariante com a distância. O patamar reflete a variância da variável para distâncias superiores ao alcance (Almeida & Bettini, 1994);
- **efeito pepita** (*nugget effect*) - C_0 : teoricamente, $\gamma(0) = 0$. Entretanto, muitas vezes na prática, à medida que \mathbf{h} tende para 0, $\gamma(\mathbf{h})$ se aproxima de um valor positivo denominado de efeito pepita. Este valor revela a descontinuidade do variograma para distâncias menores do que a menor distância entre as observações. Acredita-se que este fenômeno seja devido à variação em micro-escala que causaria uma descontinuidade na origem ou a possíveis erros de medida para observações muito próximas (Cressie, 1993). Entretanto, não é possível quantificar se a maior contribuição provém dos erros de medida ou da variabilidade de pequena escala não captada pela amostragem (Camargo, 1997);
- **contribuição** - C_1 : é a diferença entre o patamar (C) e o efeito pepita (C_0).

A figura 2 ilustra os componentes de um variograma:

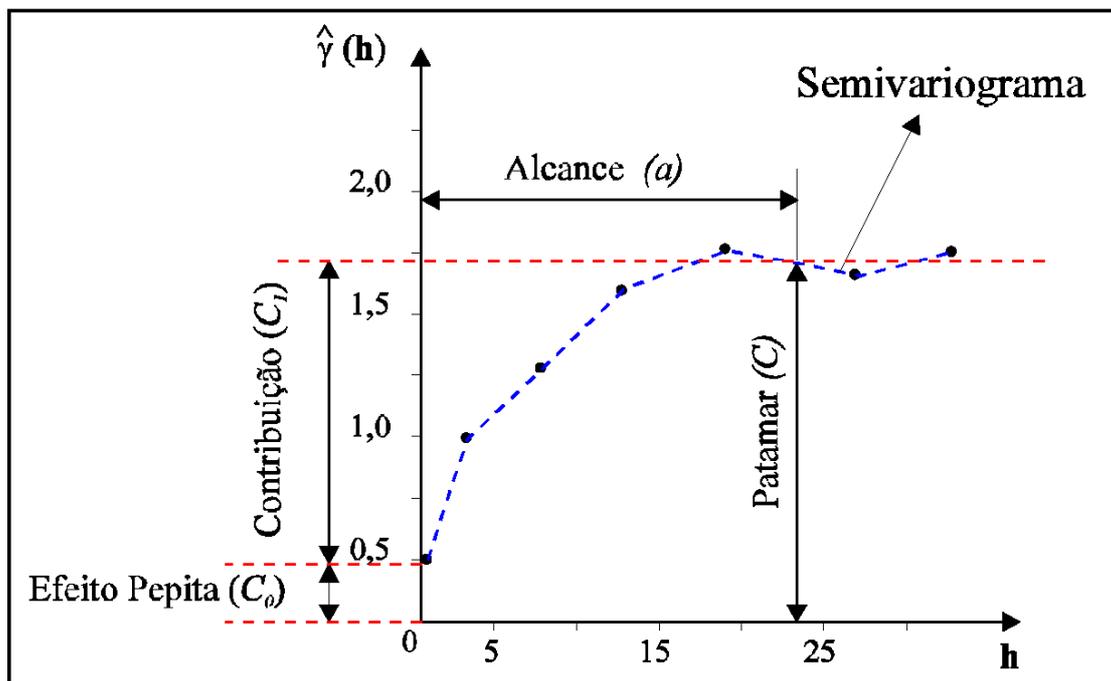


Figura 2: Exemplo de um variograma com seus componentes (fonte: Camargo, 1997)

Como foi descrito no item anterior, a estacionariedade da covariância implica na estacionariedade da variância, assim, as duas relações abaixo tornam-se evidentes:

$$\text{Var}\{Z(s)\} = E\{[Z(s) - m]^2\} = C(0), \quad \forall s$$

$$\gamma(\mathbf{h}) = \frac{1}{2} E\{[Z(s+\mathbf{h}) - Z(s)]^2\} = C(0) - C(\mathbf{h}), \quad \forall s.$$

Esta última relação indica que, sob a hipótese de estacionariedade de segunda ordem, a covariância e o variograma são duas ferramentas equivalentes para caracterizar a autocorrelação espacial entre duas variáveis $Z(s+\mathbf{h})$ e $Z(s)$ separadas por uma distância \mathbf{h} . Dessa forma, pode-se definir uma terceira ferramenta, o **correlograma** (Journel & Huijbregts, 1978):

$$\rho(h) = \frac{C(h)}{C(0)} = 1 - \frac{\gamma(h)}{C(0)}$$

Desta relação, verifica-se que na ausência de autocorrelação espacial, $\rho(\mathbf{h})$ tende a zero e $\gamma(\mathbf{h})$ se aproxima de $C(0)$ - a variância do processo. Quando $\rho(\mathbf{h})$ tende a um, indicando forte autocorrelação espacial, $\gamma(\mathbf{h})$ tende a zero (Szwarcwald & Leal, 1997).

Sob a hipótese intrínseca, é possível estimar o variograma ou semivariograma - $\gamma(\mathbf{h})$ – para os dados disponíveis. Dessa forma, é possível mensurar a heterogeneidade média entre dados separados por um vetor h . Um estimador do semivariograma pode ser calculado da seguinte forma (Goovaerts, 1997):

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(s_i) - z(s_i + h)]^2, \mathbf{h} \in \mathfrak{R}^d$$

onde $N(\mathbf{h})$ é o número de pares experimentais $[z(s_i), z(s_i + \mathbf{h})]$ de dados separados pelo vetor \mathbf{h} .

Para a estimativa dos valores do variograma dos dados amostrais necessita-se definir alguns parâmetros. Inicialmente, deve-se determinar o tamanho da área que será necessário para a construção do variograma experimental. Journel & Huijbergts (1978) propuseram que a construção deve ser feita considerando distâncias iguais ou menores do que a metade da distância máxima da região de estudo. Posteriormente, deve-se proceder a divisão dessa distância em tamanhos iguais, denominados *lag spacing* (Isaaks & Srivastava, 1989). A definição do tamanho de cada *lag* deve levar em conta o número de pares que ele possa conter, já que se este número for muito pequeno, poderão ser encontrados valores de variogramas muito irregulares, e se for muito grande, os valores poderão ser muito suavizados. A seguir, em função dos dados na maioria das vezes não estarem dispostos em um *grid* regular, é necessário a definição de uma tolerância para esse *lag*; habitualmente, utiliza-se a metade do tamanho do *lag* como tolerância.

Outro parâmetro fundamental na construção do variograma experimental é a definição da tolerância angular. Os variogramas devem ser construídos para diferentes direções na etapa exploratória; entretanto, para que se garanta encontrar pares de pontos em determinada direção, é conveniente dar uma determinada tolerância angular. Uma maneira prática de construir variogramas em diferentes direções é definir inicialmente as direções de 0°, 45°, 90° e 135°, dando uma tolerância angular de 22,5°, ou seja, a metade do incremento angular.

Para que o ângulo de busca dos pares não se abra indefinidamente, faz-se necessário restringí-lo através da definição de uma largura de banda ou *bandwidth* (Deutsch & Journel, 1998).

A figura 3 exemplifica os parâmetros necessários à construção de um variograma experimental:

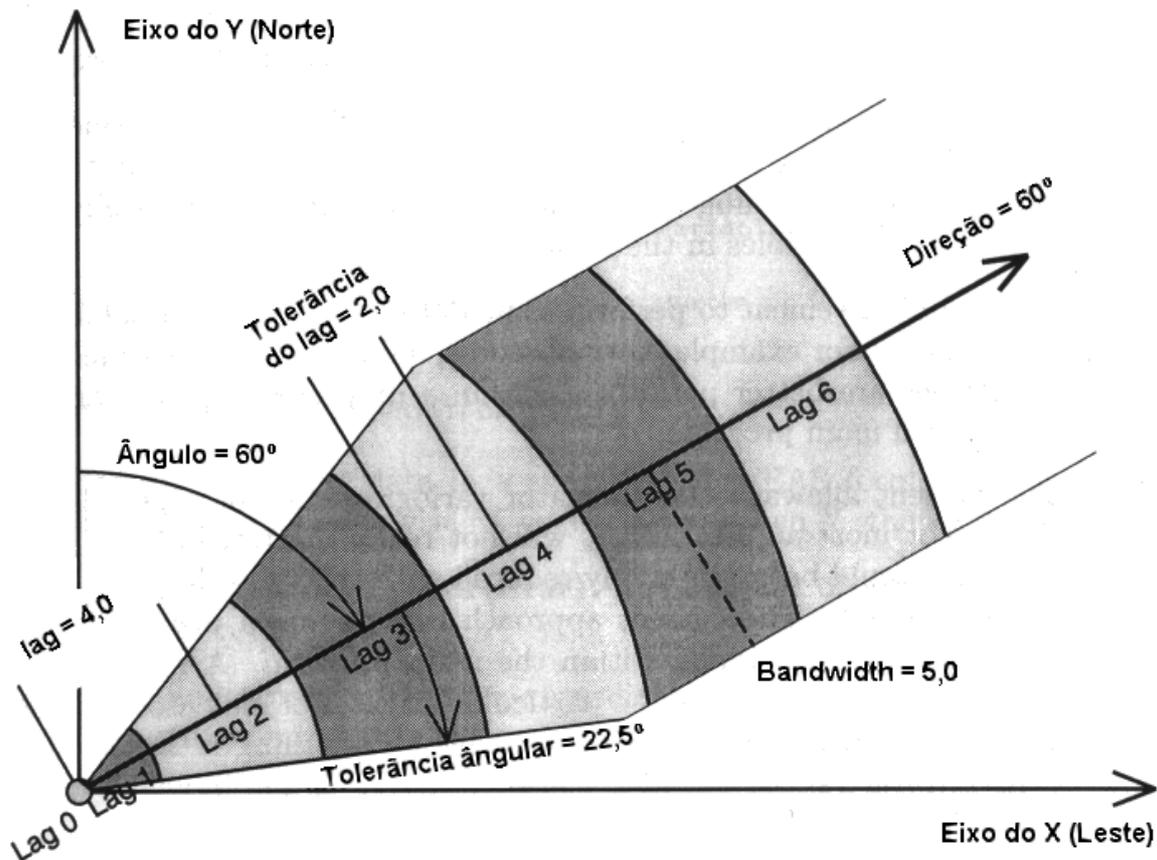


Figura 3: parâmetros requeridos para a construção do variograma experimental (adaptado de Deutsch & Journel, 1998)

Ressalta-se, entretanto, que o claro entendimento do arranjo espacial dos dados, como a detecção de aglomerados (*clusters*), tendências, etc. é essencial para definir melhor os parâmetros do variograma, tais como, tamanho do *lag*, direções e tolerância angular (Deutsch & Journel, 1998).

1.5. Verificação de eixos de continuidade espacial:

Uma etapa fundamental que precede o processo de modelagem do variograma consiste em verificar se a autocorrelação espacial do fenômeno em estudo modifica-se em diferentes direções. Se a autocorrelação espacial é apenas função do deslocamento h , o processo é dito *isotrópico*. Se o processo além da magnitude do deslocamento h ,

depende também da direção desse deslocamento, é denominado de *anisotrópico*. Existem dois tipos básicos de anisotropia (Kaluzny *et al*, 1996):

- **geométrica**: ocorre quando o alcance do variograma muda nas diferentes direções, enquanto que o patamar permanece constante (figura 4);
- **zonal**: ocorre quando o patamar do variograma muda com a direção.

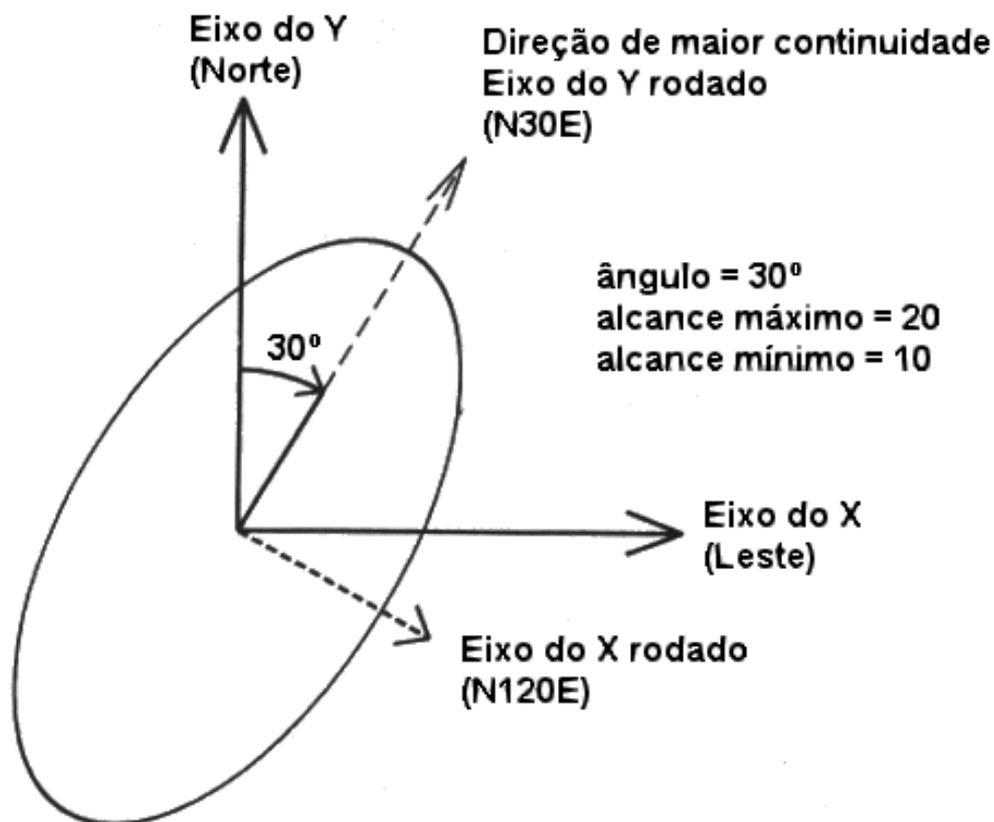


Figura 4: exemplo de dois parâmetros necessários para a definição de anisotropia geométrica de um variograma em 2D (adaptado de Deutsch & Journel, 1998)

Na figura 4, a direção de maior continuidade é especificada por um ângulo de rotação medido em sentido horário em relação ao azimute norte (eixo do Y), conseqüentemente, o maior alcance fica nesta direção e o menor alcance na direção perpendicular. Esta convenção é adotada pela biblioteca de *softwares* GSLIB (Deutsch & Journel, 1998), que foi utilizada nesta tese.

Uma maneira rápida para evidenciar anisotropia é a construção do mapa de variograma (Deutsch & Journel, 1998). Dessa forma, pode-se ter uma visão global dos valores do variograma em todas as direções. O mapa de variograma é um gráfico em 2D

de $\gamma(\mathbf{h}_1, \mathbf{h}_2)$ dos variograma amostral para todos os vetores separação $\mathbf{h} = (\mathbf{h}_1, \mathbf{h}_2)$ disponíveis experimentalmente. O valor de $\gamma(0) = 0$ é plotado no centro da figura. Assim, o mapa de variograma é construído a partir do cálculo do variograma de todos os pares de pontos que distam uma determinada distância \mathbf{h} em uma dada direção. Este valor é plotado no respectivo *pixel*. Este procedimento é feito para toda a área de estudo.

A existência de anisotropia pode ser verificada também através da construção de variogramas amostrais em todas as direções possíveis e a posterior construção do diagrama de rosas (Isaaks & Srivastava, 1989). Se todos os variogramas apresentarem o mesmo alcance, o diagrama assumirá o formato de um círculo, indicando que o fenômeno é isotrópico, caso contrário, será anisotrópico.

Deve-se ressaltar que pode-se mascarar facilmente a continuidade espacial através da definição inadequada do tamanho do *lag*, ângulos de direção ou de um manejo inadequado de valores *outliers*, embora seja raro gerar continuidade espacial onde ela não existe (Deusth & Journel, 1998).

Identificada a anisotropia, deve-se corrigi-la. Esta correção consiste na transformação do vetor de coordenadas original em um novo vetor no qual o valor do modelo de variograma anisotrópico identifica o valor de um modelo isotrópico no novo sistema de coordenadas (Goovaerts, 1997).

1.6. Modelagem do variograma:

Após a identificação das direções de máxima e mínima continuidade espacial, passa-se à etapa de modelagem do variograma experimental, que consiste no ajuste deste a um dos modelos teóricos conhecidos, através de algoritmos de regressão não linear (Szwarcwald & Leal, 1997).

Existem vários modelos de variogramas, os mais comuns são (Deusth & Journel, 1998):

a) Modelo do efeito pepita (*nugget effect*): utilizado para modelar a descontinuidade na origem. Essa descontinuidade é modelada através de um modelo de transição

definido positivo, descontínuo, sendo 0 quando $|h|$ é igual a 0 e o valor da descontinuidade de outra forma.

$$\gamma(h) = c = \begin{cases} 0 & \text{se } h = 0, \\ c, & \text{caso contrário} \end{cases}$$

- b) Modelo esférico:** caracteriza-se por um comportamento linear perto da origem. Atinge o patamar para uma distância igual ao alcance a . A contribuição positiva da variância ou patamar é de valor c . Modela fenômenos contínuos para distâncias relativamente pequenas.

$$\gamma(h) = c \cdot \text{Sph}\left(\frac{h}{a}\right) = \begin{cases} c \cdot \left[1,5 \frac{h}{a} - 0,5 \left(\frac{h}{a}\right)^3 \right], & \text{se } h \leq a \\ c, & \text{se } h \geq a \end{cases}$$

- c) Modelo exponencial:** possui um comportamento linear próximo à origem. Atinge o patamar assintoticamente para grandes distâncias.

$$\gamma(h) = c \cdot \text{Exp}\left(\frac{h}{a}\right) = c \cdot \left[1 - \exp\left(-\frac{3h}{a}\right) \right]$$

- d) Modelo gaussiano:** caracteriza-se por ter um comportamento parabólico próximo à origem. Atinge o patamar assintoticamente. Utilizado para modelar fenômenos muito contínuos.

$$\gamma(h) = c \cdot \left[1 - \exp\left(-\frac{(3h)^2}{a^2}\right) \right]$$

- e) Modelo de potência:** definido por uma potência $0 < \omega < 2$ e uma curva positiva c .

$$\gamma(h) = c \cdot h^\omega$$

- f) Modelo do efeito buraco (*hole effect*):** um variograma apresenta efeito buraco quando oscila com amplitude decrescente ao redor do patamar. Este modelo reflete a periodicidade dos dados, e é definido por um comprimento a para o primeiro pico

(tamanho da característica cíclica básica) e o valor positivo da contribuição da variância c . Utilizado para modelar fenômenos cíclicos.

$$\gamma(h) = c \cdot \left[1,0 - \cos\left(\frac{h}{a} \cdot \pi\right) \right]$$

1.7. Método de estimativa espacial – *krigagem*:

O método de estimativa ou predição espacial visa estimar o valor de um fenômeno em posições não amostradas, empregando informação proveniente de dados amostrados da variável em posições vizinhas. O mais utilizado é o método de *krigagem* (ou *kriging*) que estima de forma não enviesada valores do processo $Z(s)$, em qualquer ponto da região D , através de combinações lineares do conjunto de observações (z_1, \dots, z_n), minimizando assim os erros de predição (Szwarcwald & Leal, 1997). A existência de um modelo de dependência espacial permite manejar o problema de estimativa de valores para localizações não amostradas (Goovaerts, 1997). Entretanto, embora tenha sido este o objetivo inicial da *krigagem*, tem aumentado seu uso para a construção de modelos probabilísticos de incerteza sobre estes valores desconhecidos (Deustch & Journel, 1998).

Em Epidemiologia, a geoestatística tem sido utilizada no estudo da distribuição espacial de diversas doenças. O método de *krigagem* já foi utilizado, entre outras aplicações, para analisar a distribuição geográfica de uma epidemia de resfriado comum na França (Carrat & Valleron, 1992), identificar áreas de risco para câncer infantil na região centro-oeste da Inglaterra (Olivier *et al*, 1992), para modelagem espacial na estimativa dos dados de tuberculose no Brasil (Braga, 1997) e para a caracterização de áreas de risco à saúde (Carvalho, 1997).

A seguir são descritas algumas das mais comuns versões da *krigagem*.

1.7.1. *Krigagem* simples:

Todas as versões de *krigagem* são elaborações do algoritmo de regressão linear generalizado básico e seu estimador correspondente:

$$[Z_{SK}^*(s) - m(s)] = \sum_{\alpha=1}^n \lambda_{\alpha}(s) [Z(s_{\alpha}) - m(s_{\alpha})]$$

onde:

$Z(s)$ é o modelo do processo aleatório para a localização s ;

s_{α} são as n localizações dos dados;

$m(s) = E\{Z(s)\}$ é o valor esperado dependente da localização da variável aleatória $Z(s)$;

Z_{SK}^* é o estimador da regressão linear, também chamado estimador de *krigagem* simples (SK).

Os pesos $\lambda_{\alpha}(s)$ de SK são dados por um sistema de equações normais escritas em sua forma estacionária mais geral como:

$$\sum_{\beta=1}^n \lambda_{\beta}(s) C(s_{\beta}, s_{\alpha}) = C(s_{\beta}, s_{\alpha}), \alpha = 1, \dots, n$$

O algoritmo da SK requer o conhecimento prévio das $(n+1)$ médias $m(s)$, $m(s_{\alpha})$, $\alpha=1, \dots, n$, e a $(n+1)$ por $(n+1)$ matriz de covariância $[C(s_{\alpha}, s_{\beta}), \alpha, \beta = 0, 1, \dots, n]$ com $s_0 = s$.

1.7.2. *Krigagem* ordinária:

A *krigagem* ordinária (OK) filtra a média do estimador de SK, pois requer que a soma dos pesos da *krigagem* seja igual a 1 (Deutsch & Journel, 1998). Isto resulta no seguinte estimador de *krigagem* ordinária:

$$Z_{OK}^*(s) = \sum_{\alpha=1}^n \lambda_{\alpha}^{(OK)}(s) Z(s_{\alpha})$$

e o sistema OK estacionário:

$$\begin{cases} \sum_{\beta=1}^n \lambda_{\beta}^{(OK)}(s) C(s_{\beta} - s_{\alpha}) + \mu(s) = C(s - s_{\alpha}), \alpha=1, \dots, n \\ \sum_{\beta=1}^n \lambda_{\beta}^{(OK)}(s) = 1 \end{cases}$$

onde,

$\lambda_{\beta}^{(OK)}(s)$ são os pesos da OK; e

$\mu(s)$ é o parâmetro *Lagrange* associado com a limitante $\sum_{\beta=1}^n \lambda_{\beta}^{(OK)}(s) = 1$

A *krigagem* ordinária é um dos métodos de estimativa espacial mais utilizado em geoestatística (Journel & Huijbregts, 1978; Cressie, 1993). O preditor da *krigagem* ordinária caracteriza-se pelo fato de (Valencia, 1999):

- ser uma função linear dos dados e ótimo no sentido de minimizar o erro quadrático esperado da predição, denominado também de variância de *krigagem*;
- sua derivação não depender de uma lei de probabilidades (em particular, de uma lei gaussiana); e
- sua derivação não precisar do conhecimento da média do processo aleatório $Z(\cdot)$.

1.7.3. *Krigagem* indicadora:

Em uma região de interesse $D \subset \mathfrak{R}^d$, a variável regionalizada contínua z , torna-se conhecida através de amostragem em um conjunto de pontos $\{s_{\alpha} \in D, \alpha = 1, \dots, n\} = (n)$. Nos pontos s fora deste conjunto, o valor desconhecido de $z(s)$ é representado pela variável aleatória local $Z(s)$, cuja função de distribuição acumulada (FDA) é condicional aos dados (n) . Esta função acumulada local pode ser representada por uma coleção adequada de K variáveis binárias, todas com distribuição de Bernoulli, com parâmetro $p_k(s)$ idêntico à probabilidade acumulada em cada um dos K quantis de $Z(s)$, denominados de pontos de corte (*cutoffs*), tais que:

$$z_k = F_{Z(s)}^{-1}(p_k) \Leftrightarrow P\{Z[s | (n)] \leq z_k\} = p_k(s) = F_{Z(s)}[z_k | (n)]$$

$$k = 1, 2, \dots, K$$

A qualidade desta representação está ligada à escolha conveniente do conjunto $\{z_k, k = 1, \dots, K\}$, que pode ser formado, por exemplo, pelos decis da amostra.

No caso de uma variável contínua, o objetivo da *krigagem* indicadora (IK) é estimar $F_{Z(s)} z_k$, $k = 1, \dots, K$ para todos os pontos s de um subconjunto de D , por exemplo, no caso de D pertencer à \mathfrak{R}^2 , os nós de uma malha ou grade (*grid*) regular de

pontos. A partir desse conjunto de K pontos, é possível reconstituir aproximadamente a FDA condicional. O procedimento consiste nos seguintes passos:

(i) Transformar $Z(s)$ em K indicadores, em todos os pontos amostrais s_α , $\alpha = 1, \dots, n$:

$$I(s_\alpha, z_k) = \begin{cases} 1 & \text{se } Z(s_\alpha) \leq z_k \\ 0, & \text{caso contrário} \end{cases}$$

Desse modo, $I(\cdot)$ tem distribuição de Bernoulli, com parâmetro $p_k(s)$, tal que:

$$P[I(\cdot) = i] = \begin{cases} p_k(s), & \text{se } i = 1 \\ 1 - p_k(s), & \text{se } i = 0 \end{cases}$$

Logo,

$$E[I(\cdot)] = p_k(s) = F_{Z(s)}[z_k | (n)]$$

$$\text{Var}[I(\cdot)] = p_k(s) \cdot [1 - p_k(s)]$$

(ii) Calcular o semivariograma para cada um dos indicadores $I(\cdot, z_k)$, $k = 1, \dots, K$

$$\gamma_{I(\cdot, z_k)}(h) = \frac{n_{01} + n_{10}}{N}$$

onde:

n_{01} e n_{10} = nº de pares com valores distintos do indicador

N = nº total de pares de pontos

(iii) Ajustar um modelo ao semivariograma.

(iv) Estimar $F_{Z(s)}[z_k | (n)] = P[Z(s) \leq z_k | (n)] = p_k(s)$ para cada nó do *grid*, através da *krigagem* de cada um dos indicadores $I(s_\alpha, z_k)$, $k = 1, \dots, K$.

$$\hat{F}(s, z_k) = \hat{p}_k(s) = E[I(\cdot)] \approx I_{IK}^*[s, z_k | (n)]$$

onde: $I_{IK}^*[s, z_k | (n)]$ é o valor *krigado* do k -ésimo indicador.

Convém lembrar que $E[I(\cdot)] = p_k(s)$, mas $p_k(s) = P[Z(s) \leq z_k | (n)] = F_{Z(s)}[z_k | (n)]$. Logo, o valor *krigado* do indicador $I(\cdot, z_k)$ no ponto s , relativo ao ponto de corte z_k , é uma estimativa de $F_{Z(s)}[z_k | (n)]$.

Desse modo, em cada ponto s da malha, tem-se um conjunto de K pares $\{z_k, \hat{F}_{Z(s)}[z_k | (n)]\}$, representando a distribuição condicional local, como pode ser mostrado na figura 5:

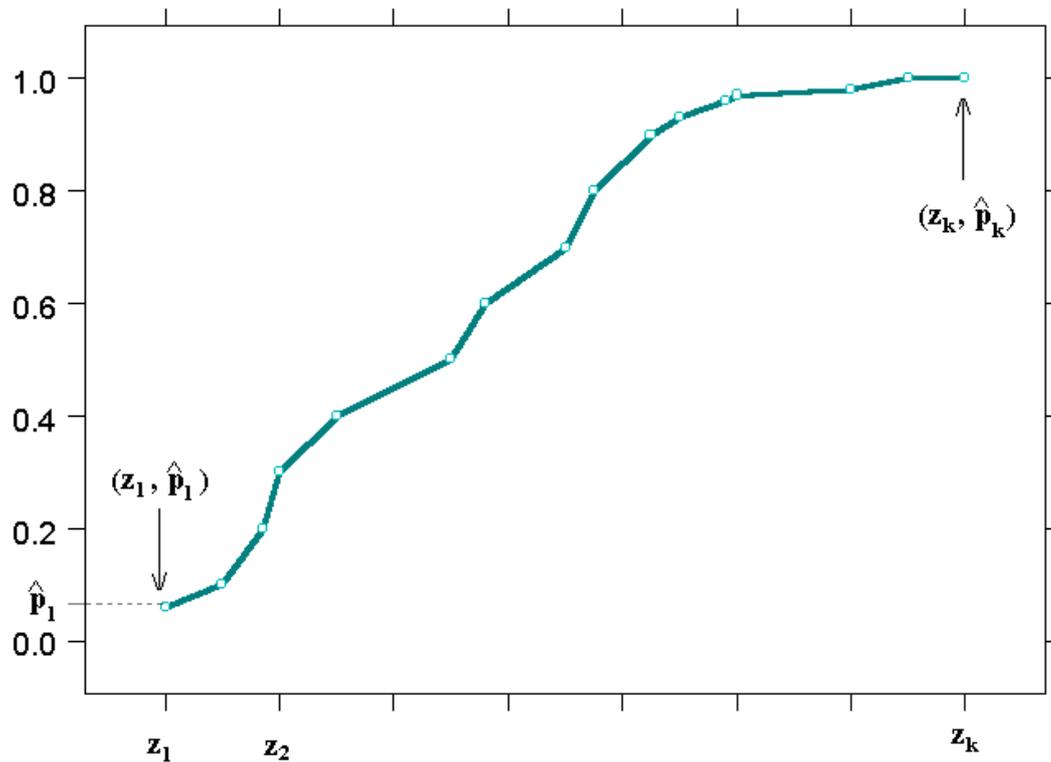


Figura 5: Distribuição condicional local no ponto s

Tratando-se de uma variável aleatória $I(s)$ binária, como é o caso da soropositividade para anti-HAV, o objetivo da *krigagem* indicadora consiste em estimar a função massa de probabilidade, ou sua FDA, para uma única variável aleatória de Bernoulli em cada nó da malha. Nesse caso, o valor krigado em cada nó s representa diretamente a probabilidade de o indicador ser igual a 1, isto é:

$$I_{IK}^*(s) \approx E[I(s)] \approx \hat{p}(s) = \hat{P}[I(s) = 1]$$

Em outros termos, o valor *krigado* em cada nó s representa a probabilidade de se encontrar naquele ponto um indivíduo soropositivo.

Finalmente, ressalta-se que uma das mais importantes características do método de *krigagem* em geral é não estar restrito aos limites das unidades geográficas; desta forma, evitam-se transições súbitas entre áreas vizinhas. O método de *krigagem* tem

como vantagem em relação a outros métodos de estimativa o fato de se basear na variabilidade espacial do dado real mapeado, provendo assim, a variância dos valores estimados. Nesse caso, mapas de erro podem ser usados para decidir onde introduzir novos valores amostrais, ou seja, para aqueles locais onde julga-se haver um erro padrão muito alto (Carrat & Valleron, 1992).

Segundo Camargo (1997), as principais vantagens do método geoestatístico em relação aos métodos usuais de estimativa espacial são:

- Os pesos determinados pela *krigagem* são baseados na análise da correlação espacial descrita pelo variograma ($\lambda_i = f[\gamma(h)]$), e não é apenas uma função da distância ($\lambda_i = f(d_i)$).
- A área de influência do processo de estimativa espacial na *krigagem* é determinada pelo alcance e não por um raio de busca arbitrário.
- A possibilidade de se modelar a anisotropia do fenômeno.
- Existência de tratamento de *clusters*, atribuindo-se pesos adequados para agrupamentos de amostras.

2. Região de estudo:

A região de estudo envolveu os setores censitários 111 e 112 da localidade denominada pelo Programa de Despoluição da Baía de Guanabara (PDBG) de “Setor Parque Fluminense”, que abrange uma parte do segundo distrito do município de Duque de Caxias.

O município de Duque de Caxias está localizado na região metropolitana do estado do Rio de Janeiro, às margens da Baía de Guanabara, fazendo divisa com os municípios do Rio de Janeiro, Belford Roxo, São João de Meriti, Miguel Pereira, Nova Iguaçu, Magé, Vassouras e Petrópolis. Segundo a contagem da população realizada pelo IBGE em 1996, o município possuía uma população de 715.089 habitantes, sendo 139.334 (19,5%) menores de 10 anos de idade (IBGE, 1997).

O “Setor Parque Fluminense” é composto de 43 setores censitários, e reunia em 1991, uma população de 61.410 pessoas (IBGE, 1991) vivendo em condições sanitárias bastante precárias. É delimitado ao norte pelo Canal do Iguaçu, ao sul pelo Canal

Sarapuí, a oeste pelo limite do município e a leste pela Baía de Guanabara. A região não conta com rede de esgoto; a cobertura da coleta de lixo é baixa e de frequência irregular. O abastecimento de água não atinge todos os domicílios e também é irregular, fazendo com que parte da população se utilize de poços rasos como forma alternativa de abastecimento de água (Almeida, 1997).

No “Setor Parque Fluminense” foi realizado um inquérito domiciliar e soroepidemiológico para anticorpos contra o vírus da hepatite A pelo Projeto de Avaliação dos Impactos do PDBG Sobre as Condições de Saúde e Qualidade de Vida (PAISQUA), desenvolvido pelo Núcleo de Estudos de Saúde Coletiva da Universidade Federal do Rio de Janeiro (NESC/UFRJ) e financiado pelo Fundo de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ).

Para estimar a soroprevalência de anti-HAV para toda a região, o PAISQUA realizou inicialmente um estudo piloto nos setores censitários 111 e 112. A seleção destes setores censitários levou em conta o fato de apresentarem melhores condições sanitárias do que os demais setores da região (evitando assim uma possível subestimativa do tamanho da amostra necessário a cada setor censitário). Outros fatores considerados para a escolha desses setores incluíram a facilidade de acesso pela equipe do projeto e a proximidade de um posto de saúde, que poderia servir de base logística para o estudo (Almeida, 1997).

3. Inquérito domiciliar e soroepidemiológico:

Em 1996, o PAISQUA realizou um inquérito soroepidemiológico para anticorpos contra o antígeno da hepatite A nos setores censitários 111 e 112 do segundo distrito do município de Duque de Caxias. Coletaram-se também, por meio de entrevistas domiciliares, variáveis relacionadas ao indivíduo, seu ambiente domiciliar e peridomiciliar. A coleta destes dados foi feita através de um formulário pré-codificado (anexo A). O presente trabalho utilizou o banco de dados resultante deste inquérito. Para minimizar o efeito coorte e os problemas decorrentes de movimentos migratórios, foram estudadas apenas as crianças com idade igual ou maior a um ano e menor que dez anos. Como a hepatite A é uma doença benigna, muito poucos indivíduos morrerão em decorrência dela e, conseqüentemente, ocorrerá pouco *viés de seleção* (Rothman &

Greenland, 1998). Além disso, o estudo de soroprevalência do anti-HAV, cujas técnicas de detecção estão bem estabelecidas e têm custo razoável, permite uma adequada classificação de casos (Almeida, 1997). Os indivíduos amostrados para este estudo foram considerados como membros de uma população dinâmica para um determinado ponto no tempo (Miettinen, 1985).

A coleta sangüínea foi feita por punção da veia cubital. Nos casos em que, por qualquer motivo, a punção venosa não pode ser realizada (recusa ou dificuldade de venopunção), foi feita punção digital com lanceta automática de ponteira descartável. O sangue total era centrifugado e decantado. O soro resultante foi submetido à análise pelo método MEIA (Abbott) para anti-HAV total no Laboratório de Virologia do Serviço de Patologia Clínica do Hospital Universitário Clementino Fraga Filho da Universidade Federal do Rio de Janeiro. Antes da coleta sangüínea, o responsável pela criança era informado sobre os objetivos do trabalho e assinava um termo de consentimento (Almeida, 1997).

Em etapa anterior à realização deste inquérito, o PAISQUA procedeu um cadastramento de domicílios e um censo da população residente nos domicílios particulares dos setores censitários 111 e 112. Foram recenseados 3051 indivíduos com um ano ou mais de idade, sendo 534 (17,5%) menores de 10 anos, dos quais 419 foram incluídos no inquérito, ocorrendo assim a perda de 115 indivíduos (21,5%). Segundo Almeida (1997), não se verificou nenhum padrão nas perdas nas diferentes faixas etárias nos setores censitários. Diversos foram os motivos das perdas relacionadas à recusa por parte do responsável pela criança, entre os quais: o fato da criança ter feito “exame de sangue” recente e “estar em dia” (mais freqüente); o medo de “tomar injeção”; a preocupação da coleta sangüínea provocar anemia nas crianças; o receio do material estar contaminado com o vírus da imunodeficiência humana (HIV) ou o fato da família possuir seguro de saúde tendo assim maior acesso a exames quando necessário.

Para que se pudesse proceder a análise estatística espacial dos dados, foi realizado novo trabalho de campo para identificar a localização do domicílio onde cada criança que participou do inquérito residia, a partir da identificação das coordenadas UTM (*Universal Transverse Mercator*). Foram localizadas 410 crianças, com perda de 9 indivíduos (2,1% em relação ao total de crianças examinadas no inquérito).

4. Processamento e análise dos dados:

A análise exploratória convencional dos dados foi feita utilizando-se os *softwares* Epi Info 6.0a e Stata 5.0. Inicialmente, procedeu-se a análise univariada das diversas variáveis que constituíam o banco de dados. A seguir, foi realizada a análise bivariada no sentido de encontrar diferenças entre a soropositividade e as demais variáveis. Esse mesmo procedimento foi feito para a comparação dos dois setores censitários selecionados neste estudo. A comparação entre proporções foi feita pelo teste do χ^2 e a análise comparativa da distribuição das variáveis foi feita utilizando-se o teste de Kruskal-Wallis, sendo considerado estatisticamente significativo um valor de $p < 0,05$ bi-caudal.

No processo de análise espacial e modelagem da ocorrência de anti-HAV, atribuiu-se o valor 1 (um) aos casos soropositivos e o valor 0 (zero) aos casos soronegativos para anti-HAV. Não foi possível obter para este trabalho os valores mensurados para cada indivíduo com os respectivos *cutoffs*. Assim, somente foi possível proceder a análise da distribuição de valores 0 e 1. Considerou-se, então, para cada ponto de observação, uma variável aleatória de Bernoulli com valores $\{0,1\}$, significando ausência e presença de soropositividade para o anti-HAV, sendo o parâmetro $p = p(s)$ a probabilidade do indivíduo ser soropositivo. Desse modo, pode-se descrever a variável aleatória local $Z(s)$ através da função de probabilidade mostrada na tabela 1, a seguir:

Tabela 1
Função de distribuição de probabilidade da variável aleatória local $Z(s)$

$Z(s)$	0	1	Total
$P[Z(s) = z] = p(s)$	$1 - p$	p	1

Assim,

$$E[Z(s)] = p(s) = p$$

$$Var[Z(s)] = p - p^2 = p(1 - p)$$

$$F[Z(s)] = \begin{cases} 0, & \text{se } x < 0 \\ 1 - p, & \text{se } 0 \leq x < 1 \\ 1, & \text{se } x \geq 1 \end{cases}$$

Baseado nos pressupostos acima, o patamar do variograma deve ser igual a $p(1 - p)$, tendo em vista ser este o valor máximo esperado para a variância do conjunto de observações.

A análise exploratória espacial foi realizada utilizando-se o SAGA/UFRJ, a biblioteca de *softwares* disponíveis na GSLIB 2.0, ArcView 3.0a, Surfer 6.0, o módulo S+SpatialStats do S-Plus 4.5 e o SAS Release 6.12.

Inicialmente, corrigiram-se as coordenadas UTM somente para facilitar os cálculos; assim, subtraiu-se do eixo de coordenadas x o valor de 670.000, e do eixo de coordenadas y o valor de 7.480.000. A seguir, procedeu-se a construção de mapas de variogramas e de variogramas direcionais, no sentido de identificar-se a existência de anisotropia.

Os variogramas experimentais foram construídos, primeiramente, para os dois setores em conjunto. Utilizou-se como *lag* de separação, a divisão em dez partes iguais da metade da maior distância entre dois pontos. Posteriormente, diversos variogramas experimentais foram construídos para os setores censitários 111 e 112, separadamente. O *lag* inicial para a construção dos variogramas foi reduzido para 100m após ter sido constatado que a continuidade espacial não ultrapassava esta distância. Os parâmetros utilizados para a construção dos diferentes variogramas experimentais foram:

- número de *lags*: 10
- distância de separação de cada *lag*: 10m
- tolerância de cada *lag*: 7.5m
- tolerância angular: 30°
- *bandwidth*: 50m

Os parâmetros encontrados na variografia referentes ao efeito pepita (*nugget effect*), contribuição (diferença entre o efeito pepita e o patamar), alcance e efeito buraco (*hole effect*) foram utilizados no processo de modelagem espacial. Para tal, utilizou-se o modelo esférico por ter apresentado melhor ajuste. Ressalta-se, também, que este modelo é o mais usado em geoestatística.

O procedimento de estimativa espacial foi feito através do método de *krigagem* indicadora ordinária, considerando os parâmetros estabelecidos pela modelagem espacial. A GSLIB 2.0 disponibiliza um programa denominado *ik3d* que executa a *krigagem* indicadora simples ou ordinária de variáveis categóricas ou indicadoras cumulativas definidas a partir de variáveis contínuas. Entretanto, a *krigagem* indicadora ordinária de variáveis categóricas também pode ser feita aplicando-se diretamente o programa *kt3d* para o dado categórico, tendo sido este o procedimento efetuado no presente trabalho, pois os dados já estavam no formato 0 e 1.

Utilizaram-se três tipos de estimativa espacial: a primeira usou um modelo omnidirecional baseado nos dados da sorologia anti-HAV disponíveis para os dois setores censitários em conjunto; a segunda levou em consideração a anisotropia existente em cada um dos setores e assim, eles foram analisados separadamente. Finalmente, a este último procedimento foi incorporado o ajuste do efeito buraco para cada setor censitário. Ressalta-se que para a modelagem do efeito buraco foi utilizado um programa desenvolvido no SAS. Além disto, foi necessário compilar o programa *kt3d* para que se pudesse incluir os parâmetros necessários à modelagem do efeito buraco quando da execução da *krigagem* indicadora.

No sentido de se conhecer os modelos que ofereceram melhor ajuste, foi realizado o procedimento de validação cruzada para cada um deles,. Este procedimento foi executado através do programa *kt3d* existente na GSLIB 2.0. Assim, foi possível comparar os três processos de modelagem (omnidirecional, anisotrópico e efeito buraco) para cada setor.

A validação cruzada permite validar o modelo geoestatístico através da re-estimativa, sob determinadas condições, dos valores conhecidos. Estas condições referem-se ao modelo de variograma, tipo de *krigagem* e estratégia de busca. Na validação cruzada, o dado real é retirado um de cada vez e re-estimado a partir dos dados vizinhos, sendo recolocado no banco de dados logo após ter sido re-estimado (Deutsch & Journel, 1998).

Como o resultado obtido pela validação cruzada pertencia ao intervalo [0,1], discretizou-se este resultado em vários pontos de cortes; atribuindo-se 0 aos valores abaixo de um determinado ponto de corte, e 1 aos valores acima deste ponto. Dessa

forma foi possível obter a sensibilidade e especificidade para cada ponto de corte em cada modelo. A partir da plotagem da sensibilidade com valor complementar da especificidade ($1 - \text{especificidade}$) foi possível construir uma curva ROC – *Receiver Operator Characteristic* - (Fletcher *et al*, 1996; Sackett *et al*, 1991) para cada modelo, no sentido de se comparar os diversos ajustes para cada setor censitário específico.

Identificou-se, também, o ponto de corte que continha um equilíbrio entre as melhores sensibilidade e especificidade conjunta de cada modelo, mensurando-se a proporção de casos que foram classificados corretamente neste ponto. Utilizou-se o teste qui-quadrado de McNemar (Fleiss, 1981) para avaliar a significância ($p < 0,05$) da proporção de pares discordantes entre os valores estimados pelos diferentes modelos e os observados para este ponto de corte.

ANEXOS

15. Posição dos moradores do domicílio (grau de parentesco com o chefe da família):

N.º		ade (Cidade)	N.º da Família	Sexo 1.M asc 2.Fe m			Amb. Int..			Fezes
01										
02										
03										
04										
05										
06										
07										
08										
09										
10										

P o s i ç ã o n a f a m í l i a (G r a u)

1. Chefe	5. Pai ou Mãe	9. Genro ou Nora	13. Agregado
2. Cônjuge	6. Sogra(o)	10. Irmão(ã)	14. Pensionista
3. Filho(a)	7. Avô(ó) ou Bisavô(ó)	11. Cunhado(a)	15. Empregado(a) Doméstico(a)
4. Enteadado(a)	8. Neto(a) ou Bisneto	12. Outros Parentes	16. Parente Do Empregado

*Completar os itens com 88, nos casos de recusa em dar a informação;
99 nos casos em que a informação solicitada for ignorada*

16. HÁ QUANTO TEMPO A FAMÍLIA MORA NESTE ENDEREÇO?

16. [][]

17. REGISTRAR O CÓDIGO CORRESPONDENTE AO TEMPO DE MORADIA NO ENDEREÇO:

17. [][]

**para anos.01
para meses.02
para dias.03**

18. QUANTOS CÔMODOS EXISTEM NESTE DOMICÍLIO?

18. [][]

Considerar como cômodos, os espaços fechados, delimitados por paredes, mesmo em espaço externo, desde que usado como dormitório, cozinha, sala ou banheiro.

19. QUANTOS CÔMODOS DO DOMICÍLIO SÃO UTILIZADOS COMO DORMITÓRIO?

19. [][]

20. NESTA CASA EXISTE UM CÔMODO EXCLUSIVO PARA PREPARAR ALIMENTOS?

20. [] []

sim.01**não.02**

21. DE QUE TIPO DE MATERIAL SÃO FEITAS AS PAREDES DO DOMICÍLIO?

21. [] []

alvenaria.01**madeira aparelhada.02****pau-a-pique não revestida.03****material aproveitado.04****palha.05****outros.06**

22. QUAL O TIPO DE COBERTURA OU TELHADO DESTE DOMICÍLIO?

22. [] []

laje.01**telha de barro.02****telha de cimento - amianto.03****zinco.04****madeira aparelhada.05****palha.06****outros.07**

23. DE ONDE VEM A ÁGUA UTILIZADA NO DOMICÍLIO?

23. [] []

rede geral.01**poço.02****rede geral e poço.03****outros.04**

24. EXISTE ENCANAMENTO DENTRO DA CASA?

24. [] []

sim.01**não.02**

Preencher os itens 25 a 30 com 00 se o domicílio não tiver encanamento dentro de casa.

25. HÁ QUANTO TEMPO EXISTE ENCANAMENTO DENTRO DE CASA? 25. [][]

26. REGISTRAR O CÓDIGO CORRESPONDENTE AO TEMPO QUE TEM ENCANAMENTO DENTRO DE CASA. 26. [][]

para anos.01
para meses.02
para dias.03

27. HÁ QUANTO TEMPO O DOMICÍLIO RECEBE ÁGUA DA REDE DA CEDAE ? 27. [][]

*Completar os itens com 88, nos casos de recusa em dar a informação;
99 nos casos em que a informação solicitada for ignorada*

28. REGISTRAR O CÓDIGO CORRESPONDENTE AO TEMPO EM QUE A CASA RECEBE ÁGUA DA REDE DA CEDAE. 28. [][]

para anos.01
para meses.02
para dias.03

29. QUANTOS PONTOS DE ÁGUA EXISTEM NO DOMICÍLIO? 29. [][]

30. QUANTAS VEZES POR SEMANA A ÁGUA DA RUA ENTRA NESTA CASA?
Completar com o n.º de dias, de 01 a 07, conforme a periodicidade semanal informada; para períodos inferiores a 1 vez por semana completar com n.º 08

30.01 - NO VERÃO 30.01. [][]

30.02 - NO INVERNO 30.02. [][]

31. VOCÊS SABEM EM QUE DIA(S) DA SEMANA ENTRA ÁGUA DA RUA NESTA CASA? 31. [][]

sim.01
não.02

32. CASO O DOMICÍLIO TENHA BOMBA, ESTA É UTILIZADA PARA 32. [][]

puxar água do poço.01
puxar água da rede.02
puxar água do poço e da rede.03

Caso o domicílio não use água de poço, completar os itens de 33 a 40 com 00.

33. CASO VOCÊS UTILIZEM ÁGUA DO POÇO, ONDE O MESMO ESTÁ LOCALIZADO?

33. [][]

no próprio domicílio.01
em outro domicílio.02

Caso a resposta seja 02, registrar o endereço onde poço esta localizado:

34. COMO FOI FEITO O POÇO DE ONDE VOCÊS RETIRAM A ÁGUA ?

34. [][]

foi escavado.01
foi perfurado.02
ambos.03

Completar os itens 35 e 36 com 00, caso o poço não tenha tampa

35. DE QUE MATERIAL É FEITA A TAMPA DO POÇO ?

35. [][]

cimento ou metal.01
outros.02

O entrevistador deve avaliar a tampa do poço e registrar suas condições no item 36; considerar como adequadas apenas as tampas de cimento ou metal, em bom estado de conservação e capazes de vedar a boca do poço

36. A TAMPA DO POÇO É:

36. [][]

adequada.01
inadequada.02

37. A PROTEÇÃO OU A BOCA DO POÇO FICA ACIMA DO NÍVEL DO CHÃO ?

37. [][]

sim.01
não.02

38. HÁ QUANTO TEMPO VOCÊS USAM ÁGUA DE POÇO ?

38. [][]

39. REGISTRAR O CÓDIGO DO TEMPO EM QUE USAM ÁGUA DE POÇO.

39. [][]

para anos.01
para meses.02
para dias.03

40. DE ONDE VEM A ÁGUA QUE VOCÊS USAM PARA BEBER ?

- da rede.01
do poço.02
da rede e do poço.03
outros.04

40. [][]

*Completar os itens 41 a 43, com 00 caso não haja armazenamento de água;
no item 42 usar 00 caso a água seja armazenada em recipientes sem tampa.*

41. ONDE VOCÊS ARMAZENAM A ÁGUA PARA USO DOMÉSTICO ?

- caixa d'água.01
cisterna.02
caixa d'água e cisterna.03
outros.04

41. [][]

*Completar os itens com 88, nos casos de recusa em dar a informação;
99 nos casos em que a informação solicitada for ignorada*

42. [][]

- cimento.01
outros.02

43. DE QUANTO EM QUANTO TEMPO É FEITA A LIMPEZA :

43.01 DA CAIXA D'ÁGUA
43.02 DA CISTERNA
43.03 DE OUTROS

- pelo menos de 3 em 3 meses.01
pelo menos de 6 em 6 meses.02
pelo menos 1 vez por ano.03
menos de 1 vez por ano.04

43.01. [][]

43.02. [][]

43.03. [][]

Completar com 00 se nunca foi feita limpeza dos recipientes em uso

44. VOCÊS TÊM FILTRO(OU PURIFICADOR) DE ÁGUA EM CASA ?

- sim.01
não.02

44. [][]

45. VOCÊS COSTUMAM FERVER A ÁGUA ANTES DE BEBER ?

- sim.01

45. [][]

*Completar os itens com 88, nos casos de recusa em dar a informação;
99 nos casos em que a informação solicitada for ignorada*

44. VOCÊS TÊM FILTRO(OU PURIFICADOR) DE ÁGUA EM CASA ? 44. [][]
sim.01
não.02

45. VOCÊS COSTUMAM FERVER A ÁGUA ANTES DE BEBER ? 45. [][]
sim.01
não.02

46. QUAL O TIPO DE PRODUTO QUE VOCÊS USAM NA ÁGUA ? 46. [][]
cloro.01
outros.02
nenhum.00

47. QUANTOS BANHEIROS EXISTEM NO DOMICÍLIO? 47. [][]

*Considerar também os banheiros externos;
 Completar com 00 até o item 49, no caso a resposta seja nenhum banheiro,.*

48. O BANHEIRO DESTE DOMICÍLIO TEM: 48. [][]
vaso sanitário, pia, chuveiro (ou banheira).01
vaso sanitário e pia.02
vaso sanitário e chuveiro (ou banheira).03
vaso sanitário.04
pia e/ou chuveiro.05
não tem vaso sanitário no banheiro.00

49. AS INSTALAÇÕES SANITÁRIAS DO SEU BANHEIRO SÃO USADAS EXCLUSIVAMENTE PELOS MORADORES DESTE DOMICÍLIO ? 49. [][]
sim.01
não.02
não tem vaso.00

50. ONDE FICA O VASO SANITÁRIO QUE VOCÊS UTILIZAM ? 50. [][]
Usar 00 para vaso no domicílio
em outro domicílio.01
é vaso sanitário comunitário.02
não usam vaso sanitário.03

51. PARA ONDE VAI O ESGOTO DESTA DOMICÍLIO ? 51. [][]
- rede geral.01**
rede pluvial ou outra rede.02
direto para a rua.03
vala negra.04
rio, lagoa ou mar.05
sumidouro.06
outros.07
não tem instalações sanitárias.00
52. HÁ QUANTO TEMPO VOCÊS TEM ESSA INSTALAÇÃO DE ESGOTO? 52. [][]
53. REGISTRE O TEMPO DESSA INSTALAÇÃO ? 53. [][]
- para anos.01**
para meses.02
para dias.03
- Caso não exista coleta de lixo completar os itens 54 e 55 com 00*
54. COMO É FEITA A COLETA DE LIXO EM SEU DOMICÍLIO ? 54. [][]
- na porta da casa.01**
na caçamba de lixo.02
55. QUANTAS VEZES POR SEMANA O LIXO É RECOLHIDO NA PORTA DA SUA CASA ? 55. [][]
- Completar com o n.º de dias, de 01 a 07, conforme a periodicidade semanal informada; para períodos inferiores a 1 vez por semana completar com n.º 08*
56. CASO NÃO HAJA COLETA DE LIXO, QUAL O DESTINO DO LIXO DE SEU DOMICÍLIO: 56. [][]
- é queimado.01**
é enterrado.02
é jogado em terreno baldio.03
é jogado em rio, lago ou mar.04
outros.05
existe coleta.00
57. HÁ QUANTO TEMPO O LIXO É RECOLHIDO DESTA MANEIRA ? 57. [][]
58. REGISTRE O TEMPO EM QUE O RECOLHIMENTO É FEITO DESSA FORMA . 58. [][]
- para anos.01**
para meses.02
para dias.03
59. NASCEU ALGUMA CRIANÇA NESTE DOMICÍLIO DE JANEIRO DE 1996 59. [][]

ATÉ AGORA ?

sim.01
não.02

Caso a resposta seja afirmativa, anotar o(s) nome(s) da(s) mãe(s):

60. FALECEU ALGUMA PESSOA NESTE DOMICÍLIO DE JANEIRO DE 1996 ATÉ AGORA ?

60. [][]

sim.01
não.02

No caso de resposta afirmativa, anotar o nome(S) do(S) falecido(S):

Completar os itens com 88, nos casos de recusa em dar a informação;
99 nos casos em que a informação solicitada for ignorada

negativa

61. QUANTOS MORADORES DESTA DOMICÍLIO TIVERAM PROBLEMAS DE SAÚDE NOS ÚLTIMOS QUINZE DIAS ?

61. [][]

Aplicar um formulário azul para cada pessoa que adoeceu

62. QUANTOS MORADORES FORAM INTERNADOS NO ÚLTIMO ANO POR MOTIVO DE DOENÇA ?

62. [][]

Aplicar um formulário amarelo para cada pessoa que foi internada

63. QUANTAS CRIANÇAS MENORES DE 5 ANOS TIVERAM DIARRÉIA NOS ÚLTIMOS 15 DIAS ?

63. [][]

Verificar se houve preenchimento de formulário azul ou amarelo, caso alguma criança tenha apresentado diarreia (pelo menos 3 evacuações com fezes amolecidas no dia)

64. QUANTAS PESSOAS, NA SUA CASA, JÁ TIVERAM HEPATITE ?

64. [][]

65. QUANTAS PESSOAS, NA SUA CASA, JÁ TIVERAM LEPTOSPIROSE ?

65. [][]

66. QUANTAS PESSOAS, NA SUA CASA, JÁ TIVERAM DENGUE ?

66. [][]

67. QUAL É O SEXO DO INFORMANTE ?

67. [][]

Feminino.01
Masculino.02

68. QUAL É A IDADE DO INFORMANTE?

68. [][]

Anotar a série e o grau e depois codificar os itens 69 e 70 com o número de séries ou anos de escolaridade concluídos e 00 caso a resposta seja nunca estudou.

69. QUAL O ÚLTIMO GRAU OU SÉRIE QUE O INFORMANTE ESTUDOU?
(_____ série(ou período) do _____ grau)

69. [][]

70. QUAL A SÉRIE E GRAU QUE A DONA DA CASA (OU OUTRO RESPONSÁVEL PELOS CUIDA-DOS DA CASA) ESTÁ CURSANDO (OU ÚLTIMA SÉRIE QUE CURSOU)?

70. [][]

(_____ série do _____ grau)

71. QUAL A RENDA MENSAL DESTA FAMÍLIA?

Valor total da renda R\$ _____

Para Codificar o item, usar o número de salários mínimos, considerando uma casa decimal

sem renda no momento.00

71. [][], []

72. TÉRMINO DA ENTREVISTA: ____:____(HORAS/MINUTOS)

73. SITUAÇÃO DA ENTREVISTA:

completa.01

não realizada- domicílio fechado no dia da visita.02

domicílio desocupado.03

não realizada - informante incapaz.04

parcialmente realizada.05

recusa.06

73. [][]

NOMES E RUBRICAS:

74. Entrevistador: _____

75. Supervisor: _____

76. Coordenador: _____

74. [][]

75. [][]

76. [][]

OBSERVAÇÕES :

(registrar nesse campo as datas de agendamentos de entrevistas, caso esta não tenha se realizado na primeira visita e outras informações que julgar importante)

Data : ___ / ___ / ___

Completar o quadro com o número dos exames realizados e a ordem do morador submetido aos respectivos exames, segundo item 15 (composição da família)

REGISTROS DE EXAMES			
NO	Primeiro nome	No. do EXAME DE SANGUE	No. do EXAME de FEZES

FOI FEITA COLETA DE ÁGUA NO DOMICÍLIO?

Sim 1.

não 2.

ANEXO B

ARQUIVOS DE PARÂMETROS DA GSLIB 2.0

B.1. Arquivo de parâmetros do *varmap* (programa de construção de mapa de variograma) para os setores censitários 111 e 112 em conjunto:

Parameters for VARMAP

```

START OF PARAMETERS:
arcgeo12.dat          -file with data
1 3                  - number of variables: column numbers
-1.0e21 1.0e21      - trimming limits
0                    -1=regular grid, 0=scattered values
 50 50 1            -if =1: nx, ny, nz
1.0 1.0 1.0         - xsiz, ysiz, zsiz
1 2 0               -if =0: columns for x,y, z coordinates
vmap12.out           -file for variogram output
 20 20 0            -nxlag, nylag, nzlag
 5.0 5.0 1.0        -dxlag, dylag, dzlag
5                    -minimum number of pairs
0                    -standardize sill? (0=no, 1=yes)
1                    -number of variograms
1 1 1               -tail, head, variogram type

```

```

1 = traditional semivariogram
2 = traditional cross semivariogram
3 = covariance
4 = correlogram
5 = general relative semivariogram
6 = pairwise relative semivariogram
7 = semivariogram of logarithms
8 = semimadogram
9 = indicator semivariogram - continuous variable
10= indicator semivariogram - categorical variable

```

B.2. Arquivo de parâmetros do *varmap* (programa de construção de mapa de variograma) para o setor censitário 111:

Parameters for VARMAP

START OF PARAMETERS:

```
arcgeol.dat          -file with data
1 3                 - number of variables: column numbers
0.0 1.0e21         - trimming limits
0                  -1=regular grid, 0=scattered values
50 50 1            -if =1: nx, ny, nz
1.0 1.0 1.0        - xsiz, ysiz, zsiz
1 2 0              -if =0: columns for x,y, z coordinates
vmap1.out           -file for variogram output
20 20 0            -nxlag, nylag, nzlag
5.0 5.0 1.0        -dxlag, dylag, dzlag
10                 -minimum number of pairs
0                  -standardize sill? (0=no, 1=yes)
1                  -number of variograms
1 1 1              -tail, head, variogram type
```

```
1 = traditional semivariogram
2 = traditional cross semivariogram
3 = covariance
4 = correlogram
5 = general relative semivariogram
6 = pairwise relative semivariogram
7 = semivariogram of logarithms
8 = semimadogram
9 = indicator semivariogram - continuous variable
10= indicator semivariogram - categorical variable
```

B.3. Arquivo de parâmetros do *varmap* (programa de construção de mapa de variograma) para o setor censitário 112:

Parameters for VARMAP

```

START OF PARAMETERS:
arcgeo2.dat          -file with data
1 3                 - number of variables: column numbers
-1.0e21 1.0e21      - trimming limits
0                   -1=regular grid, 0=scattered values
 50 50 1           -if =1: nx, ny, nz
1.0 1.0 1.0        - xsiz, ysiz, zsiz
1 2 0              -if =0: columns for x,y, z coordinates
vmap2.out           -file for variogram output
 20 20 0           -nxlag, nylag, nzlag
 5.0 5.0 1.0       -dxlag, dylag, dzlag
5                   -minimum number of pairs
0                   -standardize sill? (0=no, 1=yes)
1                   -number of variograms
1 1 1              -tail, head, variogram type

```

```

1 = traditional semivariogram
2 = traditional cross semivariogram
3 = covariance
4 = correlogram
5 = general relative semivariogram
6 = pairwise relative semivariogram
7 = semivariogram of logarithms
8 = semimadogram
9 = indicator semivariogram - continuous variable
10= indicator semivariogram - categorical variable

```

B.4. Arquivo de parâmetros do *gamv* (programa de construção de variogramas) para o setor censitário 111 e 112 em conjunto, omnidirecional:

Parameters for GAMV

```

START OF PARAMETERS:
arcgeo12.dat          \file with data
1  2  0              \  columns for X, Y, Z coordinates
1  3  0              \  number of variables, column numbers
-1.0e21      1.0e21  \  trimming limits
gamomn12.out         \file for variogram output
12                  \number of lags
10.0                \lag separation distance
7.5                 \lag tolerance
1                   \number of directions
0.0  90.0  50.0    0.0  90.0  50.0 \azm, atol, bandh, dip, dtol, bandv
0                   \standardize sills? (0=no, 1=yes)
1                   \number of variograms
1  1  1             \tail var., head var., variogram type

```

B.5. Arquivo de parâmetros do *gamv* (programa de construção de variogramas) para o setor censitário 111, omnidirecional:

```

Parameters for GAMV
*****

START OF PARAMETERS:
arcgeol.dat          \file with data
1  2  0              \  columns for X, Y, Z coordinates
1  3                  \  number of variables, column numbers
-0.1    1.1          \  trimming limits
gmomni1.out          \file for variogram output
10                 \number of lags
10.0                \lag separation distance
7.5                 \lag tolerance
1                   \number of directions
0.0  90.0 50.0  0.0  90.0  50.0 \azm,atol,bandh,dip,dtol,bandv
0                   \standardize sills? (0=no, 1=yes)
1                   \number of variograms
1  1  1              \tail var., head var., variogram type

```

B.6. Arquivo de parâmetros do *gamv* (programa de construção de variogramas) para o setor censitário 111, eixos de maior e menor continuidade espacial:

```

Parameters for GAMV
*****

START OF PARAMETERS:
arcgeol.dat          \file with data
1  2  0              \  columns for X, Y, Z coordinates
1  3                  \  number of variables, column numbers
-0.1    1.1          \  trimming limits
gmanisol.out         \file for variogram output
10                 \number of lags
10.0                \lag separation distance
7.5                 \lag tolerance
2                   \number of directions
137.0  30.0 50.0 0.0  90.0  50.0 \azm,atol,bandh,dip,dtol,bandv
 47.0  30.0 50.0 0.0  90.0  50.0 \azm,atol,bandh,dip,dtol,bandv
0                   \standardize sills? (0=no, 1=yes)
1                   \number of variograms
1  1  1              \tail var., head var., variogram type

```

B.7. Arquivo de parâmetros do *gamv* (programa de construção de variogramas) para o setor censitário 112, omnidirecional:

```

Parameters for GAMV
*****

START OF PARAMETERS:
arcgeo2.dat          \file with data
1 2 0                \ columns for X, Y, Z coordinates
1 3                  \ number of variables, column numbers
-0.1 1.1             \ trimming limits
gmomni2.out          \file for variogram output
10                   \number of lags
10.0                 \lag separation distance
7.5                  \lag tolerance
1                    \number of directions
0.0 90.0 50.0 0.0 90.0 50.0 \azm,atol,bandh,dip,dtol,bandv
0                    \standardize sills? (0=no, 1=yes)
1                    \number of variograms
1 1 1                \tail var., head var., variogram type

```

B.8. Arquivo de parâmetros do *gamv* (programa de construção de variogramas) para o setor censitário 112, eixos de maior e menor continuidade espacial:

```

Parameters for GAMV
*****

START OF PARAMETERS:
arcgeo2.dat          \file with data
1 2 0                \ columns for X, Y, Z coordinates
1 3                  \ number of variables, column numbers
-0.1 1.1             \ trimming limits
gmaniso2.out         \file for variogram output
10                   \number of lags
10.0                 \lag separation distance
7.5                  \lag tolerance
2                    \number of directions
53.0 30.0 50.0 0.0 90.0 50.0 \azm,atol,bandh,dip,dtol,bandv
143.0 30.0 50.0 0.0 90.0 50.0 \azm,atol,bandh,dip,dtol,bandv
0                    \standardize sills? (0=no, 1=yes)
1                    \number of variograms
1 1 1                \tail var., head var., variogram type

```

B.9. Arquivo de parâmetros do *kt3d* (programa de *krigagem*) para os setores censitários 111 e 112 em conjunto, modelo omnidirecional:

```

Parameters for KT3D
*****

START OF PARAMETERS:
arcgeo12.dat          \file with data
1  2  0  3  0        \  columns for X, Y, Z, var, sec var
-1.0e21  1.0e21      \  trimming limits
0                    \option: 0=grid, 1=cross, 2=jackknife
xvk.dat              \file with jackknife data
1  2  0  3  0        \  columns for X,Y,Z,vr and sec var
3                    \debugging level: 0,1,2,3
kt3d.dbg             \file for debugging output
ktomni12.out         \file for kriged output
160  2402.5  5.0     \nx,xmn,xsiz
200  5402.5  5.0     \ny,ymn,ysiz
1  0.5  1.0         \nz,zmn,zsiz
1  1  1             \x,y and z block discretization
5  50               \min, max data for kriging
0                   \max per octant (0-> not used)
50.0  50.0  50.0    \maximum search radii
0.0  0.0  0.0       \angles for search ellipsoid
1  2.302            \0=SK,1=OK,2=non-st SK,3=exdrift
0 0 0 0 0 0 0 0 0  \drift: x,y,z,xx,yy,zz,xy,xz,zy
0                   \0, variable; 1, estimate trend
extdrift.dat        \gridded file with drift/mean
4                   \  column number in gridded file
1  0.12             \nst, nugget effect
1  0.06  0.0  0.0  0.0 \it,cc,ang1,ang2,ang3
    10.0 10.0 10.0   \a_hmax, a_hmin, a_vert

```

B.10. Arquivo de parâmetros do *kt3d* (programa de *krigagem*) para o setor censitário 111, modelo omnidirecional:

```

Parameters for KT3D
*****

START OF PARAMETERS:
arcgeol.dat          \file with data
1  2  0  3  0       \  columns for X, Y, Z, var, sec var
-1.0e21  1.0e21     \  trimming limits
0                   \option: 0=grid, 1=cross, 2=jackknife
xvk.dat              \file with jackknife data
1  2  0  3  0       \  columns for X,Y,Z,vr and sec var
3                   \debugging level: 0,1,2,3
kt3d.dbg             \file for debugging output
ktomni1.out          \file for kriged output
80  2752.5  5.0      \nx,xmn,xsiz
140  5402.5  5.0     \ny,ymn,ysiz
1  0.5  1.0         \nz,zmn,zsiz
1  1  1             \x,y and z block discretization
5  50               \min, max data for kriging
0                   \max per octant (0-> not used)
50.0  50.0  50.0    \maximum search radii
0.0  0.0  0.0       \angles for search ellipsoid
1  2.302            \0=SK,1=OK,2=non-st SK,3=exdrift
0 0 0 0 0 0 0 0 0  \drift: x,y,z,xx,yy,zz,xy,xz,zy
0                   \0, variable; 1, estimate trend
extdrift.dat        \gridded file with drift/mean
4                   \  column number in gridded file
1  0.13             \nst, nugget effect
1  0.03  0.0  0.0  0.0 \it,cc,ang1,ang2,ang3
    5.0  5.0  10.0    \a_hmax, a_hmin, a_vert

```

B.11. Arquivo de parâmetros do *kt3d* (programa de *krigagem*) para o setor censitário 111, modelo anisotrópico:

```

Parameters for KT3D
*****

START OF PARAMETERS:
arcgeol.dat          \file with data
1  2  0  3  0       \  columns for X, Y, Z, var, sec var
-1.0e21  1.0e21     \  trimming limits
0                   \option: 0=grid, 1=cross, 2=jackknife
xvk.dat             \file with jackknife data
1  2  0  3  0       \  columns for X,Y,Z,vr and sec var
3                   \debugging level: 0,1,2,3
kt3d.dbg            \file for debugging output
ktanisol.out        \file for kriged output
80  2752.5  5.0     \nx,xmn,xsiz
140  5402.5  5.0    \ny,ymn,ysiz
1  0.5  1.0         \nz,zmn,zsiz
1  1  1             \x,y and z block discretization
5  50              \min, max data for kriging
0                  \max per octant (0-> not used)
50.0  50.0  50.0   \maximum search radii
0.0  0.0  0.0      \angles for search ellipsoid
1  2.302           \0=SK,1=OK,2=non-st SK,3=exdrift
0 0 0 0 0 0 0 0 0  \drift: x,y,z,xx,yy,zz,xy,xz,zy
0                  \0, variable; 1, estimate trend
extdrift.dat        \gridded file with drift/mean
4                  \  column number in gridded file
1  0.14            \nst, nugget effect
1  0.03 137.0  0.0  0.0 \it,cc,ang1,ang2,ang3
50.0  20.0 10.0   \a_hmax, a_hmin, a_vert

```

B.12. Arquivo de parâmetros do *kt3d* (programa de *krigagem*) para o setor censitário 111, modelo com *hole effect*:

```

Parameters for KT3D
*****

START OF PARAMETERS:
arcgeol.dat          \file with data
1  2  0  3  0       \  columns for X, Y, Z, var, sec var
-0.001  1.001      \  trimming limits
0                   \option: 0=grid, 1=cross, 2=jackknife
xvk.dat             \file with jackknife data
1  2  0  3  0       \  columns for X,Y,Z,vr and sec var
3                   \debugging level: 0,1,2,3
kt6OK.dbg          \file for debugging output
kt6d11.out         \file for kriged output
80  2752.5  5.0     \nx,xmn,xsiz
140 5402.5  5.0     \ny,ymn,ysiz
1   0.5  1.0       \nz,zmn,zsiz
1   1  1           \x,y and z block discretization
5  50             \min, max data for kriging
0                 \max per octant (0-> not used)
50.0 50.0 50.0    \maximum search radii
  0.0 0.0 0.0     \angles for search ellipsoid
1   2.302         \0=SK,1=OK,2=non-st SK,3=exdrift
0 0 0 0 0 0 0 0 0 \drift: x,y,z,xx,yy,zz,xy,xz,zy
0                 \0, variable; 1, estimate trend
extdrift.dat       \gridded file with drift/mean
4                 \  column number in gridded file
2   0.14          \nst, nugget effect
1   0.03 47.0  0.0  0.0 \it,cc,ang1,ang2,ang3
  99999. 50.0 10.0     \a_hmax, a_hmin, a_vert
5   0.03 137.0 0.0  0.0 \it,cc,ang1,ang2,ang3
  99999.  20.0 10.0    \a_hmax, a_hmin, a_vert

```

B.13. Arquivo de parâmetros do *kt3d* (programa de *krigagem*) para o setor censitário 112, modelo omnidirecional:

Parameters for KT3D

```

START OF PARAMETERS:
arcgeo2.dat          \file with data
1  2  0  3  0      \  columns for X, Y, Z, var, sec var
-1.0e21  1.0e21    \  trimming limits
0                  \option: 0=grid, 1=cross, 2=jackknife
xvk.dat             \file with jackknife data
1  2  0  3  0      \  columns for X,Y,Z,vr and sec var
3                 \debugging level: 0,1,2,3
kt3d.dbg           \file for debugging output
ktomni2.out        \file for kriged output
140  2302.5  5.0   \nx,xmn,xsiz
140  5702.5  5.0   \ny,ymn,ysiz
1  0.5  1.0       \nz,zmn,zsiz
1  1  1           \x,y and z block discretization
5  50            \min, max data for kriging
0               \max per octant (0-> not used)
50.0  50.0  50.0 \maximum search radii
0.0  0.0  0.0    \angles for search ellipsoid
1  2.302         \0=SK,1=OK,2=non-st SK,3=exdrift
0 0 0 0 0 0 0 0  \drift: x,y,z,xx,yy,zz,xy,xz,zy
0               \0, variable; 1, estimate trend
extdrift.dat       \gridded file with drift/mean
4                \  column number in gridded file
1  0.09          \nst, nugget effect
1  0.12  0.0  0.0  0.0 \it,cc,ang1,ang2,ang3
    15.0 15.0 10.0    \a_hmax, a_hmin, a_vert

```

B.14. Arquivo de parâmetros do *kt3d* (programa de *krigagem*) para o setor censitário 112, modelo anisotrópico:

```

Parameters for KT3D
*****

START OF PARAMETERS:
arcgeo2.dat          \file with data
1  2  0  3  0       \  columns for X, Y, Z, var, sec var
-1.0e21  1.0e21     \  trimming limits
0                   \option: 0=grid, 1=cross, 2=jackknife
xvk.dat             \file with jackknife data
1  2  0  3  0       \  columns for X,Y,Z,vr and sec var
3                   \debugging level: 0,1,2,3
kt3d.dbg            \file for debugging output
ktanis2a.out        \file for kriged output
140  2302.5  5.0    \nx,xmn,xsiz
140  5702.5  5.0    \ny,ymn,ysiz
1  0.5  1.0        \nz,zmn,zsiz
1  1  1            \x,y and z block discretization
5  50             \min, max data for kriging
0                 \max per octant (0-> not used)
50.0  50.0  50.0   \maximum search radii
0.0  0.0  0.0     \angles for search ellipsoid
1  2.302          \0=SK,1=OK,2=non-st SK,3=exdrift
0 0 0 0 0 0 0 0 0 \drift: x,y,z,xx,yy,zz,xy,xz,zy
0                 \0, variable; 1, estimate trend
extdrift.dat       \gridded file with drift/mean
4                 \  column number in gridded file
1  0.15           \nst, nugget effect
1  0.06  53.0  0.0  0.0 \it,cc,ang1,ang2,ang3
    20.0  15.0  10.0   \a_hmax, a_hmin, a_vert

```

B.15. Arquivo de parâmetros do *kt3d* (programa de *krigagem*) para o setor censitário 112, modelo com *hole effect*:

```

Parameters for KT3D
*****

START OF PARAMETERS:
arcgeo2.dat          \file with data
1  2  0  3  0      \  columns for X, Y, Z, var, sec var
-0.001  1.001      \  trimming limits
0                  \option: 0=grid, 1=cross, 2=jackknife
xvk.dat             \file with jackknife data
1  2  0  3  0      \  columns for X,Y,Z,vr and sec var
3                 \debugging level: 0,1,2,3
kt6OK.dbg           \file for debugging output
kt6d12.out          \file for kriged output
140 2302.5 5.0      \nx,xmn,xsiz
140 5702.5 5.0      \ny,ymn,ysiz
1  0.5  1.0        \nz,zmn,zsiz
1  1  1            \x,y and z block discretization
5  50             \min, max data for kriging
0                 \max per octant (0-> not used)
50.0 50.0 50.0     \maximum search radii
  0.0  0.0  0.0    \angles for search ellipsoid
1  2.302          \0=SK,1=OK,2=non-st SK,3=exdrift
0 0 0 0 0 0 0 0 0 \drift: x,y,z,xx,yy,zz,xy,xz,zy
0                 \0, variable; 1, estimate trend
extdrift.dat       \gridded file with drift/mean
4                 \  column number in gridded file
2  0.15           \nst, nugget effect
1  0.06 143.0 0.0 0.0 \it,cc,ang1,ang2,ang3
  99999. 20.0 10.0    \a_hmax, a_hmin, a_vert
5  0.06 53.0 0.0 0.0 \it,cc,ang1,ang2,ang3
  99999. 15.0 5.0    \a_hmax, a_hmin, a_vert

```

B.16. Arquivo de parâmetros do *kt3d* (opção de validação cruzada) para os setores censitários 111 e 112 em conjunto, modelo omnidirecional:

```

Parameters for KT3D
*****

START OF PARAMETERS:
arcgeo12.dat          \file with data
1  2  0  3  0        \  columns for X, Y, Z, var, sec var
-1.0e21  1.0e21      \  trimming limits
1                    \option: 0=grid, 1=cross, 2=jackknife
xvk.dat              \file with jackknife data
1  2  0  3  0        \  columns for X,Y,Z,vr and sec var
3                    \debugging level: 0,1,2,3
kt3d.dbg             \file for debugging output
ktomn12x.out         \file for kriged output
160  2417.6  5.0     \nx,xmn,xsiz
180  5458.5  5.0     \ny,ymn,ysiz
1  0.5  1.0         \nz,zmn,zsiz
1  1  1             \x,y and z block discretization
5  50              \min, max data for kriging
0                  \max per octant (0-> not used)
50.0  50.0  50.0    \maximum search radii
0.0  0.0  0.0       \angles for search ellipsoid
1  2.302           \0=SK,1=OK,2=non-st SK,3=exdrift
0 0 0 0 0 0 0 0 0  \drift: x,y,z,xx,yy,zz,xy,xz,zy
0                  \0, variable; 1, estimate trend
extdrift.dat        \gridded file with drift/mean
4                  \  column number in gridded file
1  0.12            \nst, nugget effect
1  0.06  0.0  0.0  0.0  \it,cc,ang1,ang2,ang3
    10.0 10.0  10.0     \a_hmax, a_hmin, a_vert

```

B.17. Arquivo de parâmetros do *kt3d* (opção de validação cruzada) para o setor censitário 111, modelo omnidirecional:

```

Parameters for KT3D
*****

START OF PARAMETERS:
arcgeol.dat          \file with data
1  2  0  3  0       \  columns for X, Y, Z, var, sec var
-1.0e21  1.0e21     \  trimming limits
1                   \option: 0=grid, 1=cross, 2=jackknife
xvk.dat              \file with jackknife data
1  2  0  3  0       \  columns for X,Y,Z,vr and sec var
3                   \debugging level: 0,1,2,3
kt3d.dbg             \file for debugging output
ktomnilx.out         \file for kriged output
160  2417.6  5.0     \nx,xmn,xsiz
180  5458.5  5.0     \ny,ymn,ysiz
1  0.5  1.0         \nz,zmn,zsiz
1  1  1             \x,y and z block discretization
5  50              \min, max data for kriging
0                  \max per octant (0-> not used)
50.0  50.0  50.0    \maximum search radii
0.0  0.0  0.0       \angles for search ellipsoid
1  2.302           \0=SK,1=OK,2=non-st SK,3=exdrift
0 0 0 0 0 0 0 0 0  \drift: x,y,z,xx,yy,zz,xy,xz,zy
0                  \0, variable; 1, estimate trend
extdrift.dat        \gridded file with drift/mean
4                  \  column number in gridded file
1  0.13            \nst, nugget effect
1  0.03  0.0  0.0  0.0  \it,cc,ang1,ang2,ang3
    5.0  5.0  10.0     \a_hmax, a_hmin, a_vert

```

B.18. Arquivo de parâmetros do *kt3d* (opção de validação cruzada) para o setor censitário 111, modelo anisotrópico:

```

Parameters for KT3D
*****

START OF PARAMETERS:
arcgeol.dat          \file with data
1  2  0  3  0       \  columns for X, Y, Z, var, sec var
-1.0e21  1.0e21     \  trimming limits
1                   \option: 0=grid, 1=cross, 2=jackknife
xvk.dat             \file with jackknife data
1  2  0  3  0       \  columns for X,Y,Z,vr and sec var
3                   \debugging level: 0,1,2,3
kt3d.dbg           \file for debugging output
ktanislx.out       \file for kriged output
160  2417.6  5.0   \nx,xmn,xsiz
180  5458.5  5.0   \ny,ymn,ysiz
1  0.5  1.0       \nz,zmn,zsiz
1  1  1           \x,y and z block discretization
5  50            \min, max data for kriging
0                \max per octant (0-> not used)
50.0  50.0  50.0  \maximum search radii
0.0  0.0  0.0     \angles for search ellipsoid
1  2.302         \0=SK,1=OK,2=non-st SK,3=exdrift
0 0 0 0 0 0 0 0  \drift: x,y,z,xx,yy,zz,xy,xz,zy
0                \0, variable; 1, estimate trend
extdrift.dat       \gridded file with drift/mean
4                 \  column number in gridded file
1  0.14           \nst, nugget effect
1  0.03 137.0  0.0  0.0 \it,cc,ang1,ang2,ang3
50.0  20.0  10.0  \a_hmax, a_hmin, a_vert

```

B.19. Arquivo de parâmetros do *kt3d* (opção de validação cruzada) para o setor censitário 111, modelo com *hole effect*:

```

Parameters for KT3D
*****

START OF PARAMETERS:
arcgeol.dat          \file with data
1  2  0  3  0       \  columns for X, Y, Z, var, sec var
-0.001  1.001       \  trimming limits
1                   \option: 0=grid, 1=cross, 2=jackknife
xvk.dat             \file with jackknife data
1  2  0  3  0       \  columns for X,Y,Z,vr and sec var
3                   \debugging level: 0,1,2,3
kt6OK.dbg           \file for debugging output
kt6d11x.out         \file for kriged output
80  2752.5  5.0     \nx,xmn,xsiz
140 5402.5  5.0     \ny,ymn,ysiz
1  0.5  1.0        \nz,zmn,zsiz
1  1  1            \x,y and z block discretization
5  50              \min, max data for kriging
0                  \max per octant (0-> not used)
50.0  50.0  50.0   \maximum search radii
  0.0  0.0  0.0    \angles for search ellipsoid
1  2.302           \0=SK,1=OK,2=non-st SK,3=exdrift
0 0 0 0 0 0 0 0 0  \drift: x,y,z,xx,yy,zz,xy,xz,zy
0                  \0, variable; 1, estimate trend
extdrift.dat        \gridded file with drift/mean
4                  \  column number in gridded file
2  0.14            \nst, nugget effect
1  0.03  47.0  0.0  0.0 \it,cc,ang1,ang2,ang3
  99999.  50.0  10.0    \a_hmax, a_hmin, a_vert
5  0.03  137.0  0.0  0.0 \it,cc,ang1,ang2,ang3
  99999.  20.0  10.0    \a_hmax, a_hmin, a_vert

```

B.20. Arquivo de parâmetros do *kt3d* (opção de validação cruzada) para o setor censitário 112, modelo omnidirecional:

```

Parameters for KT3D
*****

START OF PARAMETERS:
arcgeo2.dat          \file with data
1  2  0  3  0      \  columns for X, Y, Z, var, sec var
-1.0e21  1.0e21    \  trimming limits
1                  \option: 0=grid, 1=cross, 2=jackknife
xvk.dat             \file with jackknife data
1  2  0  3  0      \  columns for X,Y,Z,vr and sec var
3                 \debugging level: 0,1,2,3
kt3d.dbg           \file for debugging output
ktomni2x.out       \file for kriged output
160  2417.6  5.0   \nx,xmn,xsiz
180  5458.5  5.0   \ny,ymn,ysiz
1  0.5  1.0       \nz,zmn,zsiz
1  1  1           \x,y and z block discretization
5  50            \min, max data for kriging
0               \max per octant (0-> not used)
50.0  50.0  50.0  \maximum search radii
0.0  0.0  0.0     \angles for search ellipsoid
1  2.302         \0=SK,1=OK,2=non-st SK,3=exdrift
0 0 0 0 0 0 0 0  \drift: x,y,z,xx,yy,zz,xy,xz,zy
0               \0, variable; 1, estimate trend
extdrift.dat       \gridded file with drift/mean
4                \  column number in gridded file
1  0.09          \nst, nugget effect
1  0.12  0.0  0.0  0.0  \it,cc,ang1,ang2,ang3
    15.0 15.0 10.0     \a_hmax, a_hmin, a_vert

```

B.21. Arquivo de parâmetros do *kt3d* (opção de validação cruzada) para o setor censitário 112, modelo anisotrópico:

```

Parameters for KT3D
*****

START OF PARAMETERS:
arcgeo2.dat          \file with data
1  2  0  3  0      \  columns for X, Y, Z, var, sec var
-1.0e21  1.0e21    \  trimming limits
1                  \option: 0=grid, 1=cross, 2=jackknife
xvk.dat             \file with jackknife data
1  2  0  3  0      \  columns for X,Y,Z,vr and sec var
3                 \debugging level: 0,1,2,3
kt3d.dbg           \file for debugging output
ktanis2x.out       \file for kriged output
160  2417.6  5.0   \nx,xmn,xsiz
180  5458.5  5.0   \ny,ymn,ysiz
1  0.5  1.0       \nz,zmn,zsiz
1  1  1           \x,y and z block discretization
5  50            \min, max data for kriging
0                \max per octant (0-> not used)
50.0  50.0  50.0  \maximum search radii
0.0  0.0  0.0     \angles for search ellipsoid
1  2.302         \0=SK,1=OK,2=non-st SK,3=exdrift
0 0 0 0 0 0 0 0  \drift: x,y,z,xx,yy,zz,xy,xz,zy
0                \0, variable; 1, estimate trend
extdrift.dat       \gridded file with drift/mean
4                 \  column number in gridded file
1  0.15           \nst, nugget effect
1  0.06  53.0  0.0  0.0  \it,cc,ang1,ang2,ang3
    20.0  15.0  10.0     \a_hmax, a_hmin, a_vert

```

B.22. Arquivo de parâmetros do *kt3d* (opção de validação cruzada) para o setor censitário 112, modelo com *hole effect*:

```

Parameters for KT3D
*****

START OF PARAMETERS:
arcgeo2.dat          \file with data
1  2  0  3  0       \  columns for X, Y, Z, var, sec var
-0.001  1.001      \  trimming limits
1                   \option: 0=grid, 1=cross, 2=jackknife
xvk.dat             \file with jackknife data
1  2  0  3  0       \  columns for X,Y,Z,vr and sec var
3                   \debugging level: 0,1,2,3
kt6OK.dbg           \file for debugging output
kt6d12x.out         \file for kriged output
140 2302.5  5.0     \nx,xmn,xsiz
140 5702.5  5.0     \ny,ymn,ysiz
1  0.5  1.0        \nz,zmn,zsiz
1  1  1            \x,y and z block discretization
5  50             \min, max data for kriging
0                 \max per octant (0-> not used)
50.0 50.0 50.0    \maximum search radii
  0.0  0.0  0.0   \angles for search ellipsoid
1  2.302          \0=SK,1=OK,2=non-st SK,3=exdrift
0 0 0 0 0 0 0 0 0 \drift: x,y,z,xx,yy,zz,xy,xz,zy
0                 \0, variable; 1, estimate trend
extdrift.dat       \gridded file with drift/mean
4                 \  column number in gridded file
2  0.15           \nst, nugget effect
1  0.06 143.0 0.0 0.0 \it,cc,ang1,ang2,ang3
  99999. 20.0 10.0   \a_hmax, a_hmin, a_vert
5  0.06 53.0 0.0 0.0 \it,cc,ang1,ang2,ang3
  99999. 15.0 5.0   \a_hmax, a_hmin, a_vert

```

ANEXO C

PROGRAMAS DESENVOLVIDOS NO SAS

C.1. Programa para suavização para os valores dos variogramas resultantes do *varmap* (exemplifica-se apenas o programa referente ao setor censitário 111):

```

/* programa varmap.sas */
/* programa para ler o grid do varmap produzido pela gslib, */
/* na forma de tripa ( 41 * 41 linhas), */
/* gerar coordenadas (x,y) e exibir mapa de contorno */
/* arquivo de entrada = c:\medronho\tese\vmmap111.dat */

data a;
  infile 'vmmap111.dat';
  input varm v2-v6;
  iy = 1 + int( (_N_ - 0.5) / 41 ) ;
  ix = _N_ - (iy - 1) * 41 ;
  y = iy*5 - 105.0;
  x = ix*5 - 105.0;
run;

data varmap;
  set a;
  keep ix iy x y varm;
  if varm ge 0.;
run;

proc means;
  run;

goptions device=win;

proc g3grid data=varmap out= suave;
  grid y*x=varm / axis1= -100.0 to 100.0 by 2
                axis2= -100.0 to 100.0 by 2
                near = 50;
run;

title h=0.9 'MAPA DE VARIOGRAMA DO SETOR 111';

axis1 length= 60 pct minor=none value=(h=0.8 c=blue f=swissl)
      c= red;
axis2 length= 25 pct minor=none value=(h=0.8 c=blue f=swissl)
      c= red;
pattern1 c = blue v = s;
pattern2 c = green v = s;
pattern3 c = yellow v = s;
pattern4 c = orange v = s;
pattern5 c = red v = s;

proc gcontour DATA=suave;
  plot y*x = varm/
      pattern join
      caxis=blue coutline=green ctext=black
      vaxis = axis1
      haxis = axis2
      levels = 0.0 to 0.4 by 0.1 ;

```

run;

C.2. Programa para traçar a elipse com os eixos de maior e menor continuidade espacial definidos pela variografia em todas as direções (exemplo para o setor censitário 111):

```

/*-----*/
/*  PROGRAMA ELIP111.SAS                               */
/*    traçado da elipse para o setor 111              */
/*     $(x/a)**2 + (y/b)**2 = 1$                        */
/*     $y = \pm b * \text{sqrt}(1 - x^2 / a^2)$               */
/*  CONVENCAO: a = alcance na direcao x = ~leste     */
/*              b = alcance na direcao y = ~norte     */
/*  considerar  angulos de rotacao alfa em [-45,+45]  */
/*              direcao NW ==> angulo negativo        */
/*              direcao NE ==> angulo positivo        */
/*  considerar  excentricidade >1 para b>a           */
/*              excentricidade <1 para b<a           */
/*  ATENCAO:    angulos sempre em RADIANOS           */
/*  Exemplo:    cos(alfa * 3.14159 / 180.)            */
/*-----*/
/*  No caso do setor 111:                             */
/*  Azimute do eixo maior = - 43 ou 137 > 45         */
/*  usar alfa = -11.                                  */
/*  alcance na direcao proxima do norte: b = 50.0   */
/*  alcance na direcao proxima do leste: a = 20.0   */
/*  centro da figura no ponto (40,40)                */
/*-----*/

data elipse;
do x = -40 to 40 by 0.01;
y1= 50.0 * sqrt( 1.0 - (x /20.0 )**2);
y2= -50.0 * sqrt( 1.0 - (x /20.0 )**2);
xa1 = 40. + cos(-43.*3.14159/180.0)*x + sin(-43.*3.14159/180.0)*y1;
xa2 = 40. + cos(-43.*3.14159/180.0)*x + sin(-43.*3.14159/180.0)*y2;
ya1 = 40. - sin(-43.*3.14159/180.0)*x + cos(-43.*3.14159/180.0)*y1;
ya2 = 40. - sin(-43.*3.14159/180.0)*x + cos(-43.*3.14159/180.0)*y2;
output; end; run;

title1 h=0.85 'DIAGRAMA DE ROSAS DO SETOR CENSITARIO 111';
title2 h=0.75 'RAZAO DE ANIS = 50.0/20.0 ANG DE MAIOR CONT = 137°';
symbol1 v=point i=spline c=blue;
symbol2 v=point i=spline c=blue;

axis1 length=31.9 pct order= 0 to 80 by 20 minor=none
      label=(h=0.75 'LESTE') value=(h=0.7) ;
axis2 length=60 pct order= 0 to 80 by 20 minor=none
      label=(h=0.75 a=90 'NORTE') value=(h=0.7) ;

goptions device=win;
proc gplot;
plot ya1 *xa1 ya2*xa2 / frame overlay
      haxis=axis1
      vaxis=axis2;
run;

```

C.3. Programa para proceder a modelagem dos diversos variogramas omnidirecionais e para a maior e menor continuidade espacial de cada setor (exemplo da modelagem do eixo de 137° para o setor censitário 111):

```

/* programa modelo.sas
/* função: modelagem de variograma omnidirecional e anisotrópico
/* notação: i = identificador; h = lag; gama = variograma
/* ----- */

data a;
input i h gama v4-v6;
drop i v5-v6;
v4 = .;
cards;

    2      3.791      .14286      35      .22857      .34286
    3      9.342      .14894      47      .12766      .34043
    4     22.485      .15789      95      .24211      .22105
    5     30.257      .16667     180      .23889      .23889
    6     40.497      .17593     216      .22685      .22685
    7     49.969      .17241     203      .17241      .21182
    8     60.270      .16919     198      .14141      .23737
    9     70.500      .18000     250      .14000      .28400
   10     80.119      .20513     273      .16850      .27839
   11     89.710      .18939     264      .18561      .26894
   12     98.282      .16418     201      .17413      .28358
;
run;

data b;
do h= 0.0 to 100. by 0.5;
gama = .;
if h < 50.0 then v4 = 0.14 + 0.03* (1.5* h/50. - 0.5* (h/50.)**3 );
else v4 = 0.17 ;
output; end;
run;

proc append base=a new=b;
run;

data c;
set a;
if (v4=0 and h < 50.0) then v4 = 0.14 + 0.03* (1.5* h/50. - 0.5*
(h/50.)**3 );
else if ( v4=0 and h > 50.) then v4 = 0.17 ;
run;

proc sort data=c;
by h; run;

options device=win;

title1 h=1.0 'SETOR CENSITARIO 111';
title2 h=0.9 'Semivariograma indicador no azimute 137';
title3 h=0.8 'a = 50; c0 = 0.14; cc = 0.03 (maior continuidade)';

symbol1 v=circle c=red i=none h=1.2;
symbol2 v=point c=blue i=spline w= 4 ;

```

```
axis1 length= 50 pct order=0 to 100 by 20 minor=none;  
axis2 length= 55 pct order=0 to 0.25 by 0.05 minor=none;  
proc gplot;  
  plot ( gama v4) *h / overlay frame  
        haxis=axis1 vaxis=axis2;  
run;
```

C.4. Programa para proceder a modelagem do *hole effect* (exemplo da modelagem do eixo de 57° - menor continuidade - para o setor censitário 111):

```

/* ----- */
/* Programa hole.sas                               */
/* Função: ajustar hole effect aos dados do setor 111 */
/* Autor : Claudio Bettini                         */
/* Data: 02-02-99                                  */
/* Adaptado por Roberto A. Medronho em: 02-03-99 */
/* Notação: h = lag; gama = variograma; nlag = nº pontos em cada lag */
/* ----- */

```

```

data sc111;
  input h gama nlag;
CARDS;
  3.68   .150  20.
  11.45  .233  43.
  22.04  .207  94.
  30.38  .198  189.
  40.42  .144  187.
  50.44  .154  240.
  59.76  .167  221.
  70.47  .137  223.
  79.71  .132  254.
  90.24  .140  215.
  98.60  .135  200.
;
run;

proc nlin method=dud;

  parms a=20.0  c=0.03  c0=0.14;

  file print;

  model
    gama = c0 + c * (1. - exp(-3.*h/(10.*a)))*cos( (h/a)*3.14159 );
  bounds 0 <= c0 < 0.25;
  bounds 0 < c < 0.25;
  bounds 0 < a < 50;

  output out=bomb predicted=yp r=yr;
run;

```

```

/* ----- */
/*                               Gráficos                               */
/* ----- */

```

```

goptions device = win;
  title 'EFEITO BURACO NO SETOR 111';
symbol1 v=circle i=none c=red h=1.2;
symbol2 v=point i=spline c=blue w=4;

axis1 length=55 pct minor=none order=0 to 0.3 by 0.1;
axis2 length=50 pct minor=none order=0 to 100 by 20;

proc gplot;
  plot gama*h  yp*h / overlay frame
        vaxis=axis1 haxis=axis2;
  label gama= 'SEMIVARIOGRAMA'
         h = 'INTERDISTANCIA';

```

```
run;
```

C.5. Programa para construção da curva ROC e identificação do *cutoff* com equilíbrio entre as melhores sensibilidade e especificidade conjuntas (exemplo para o modelo omnidirecional do setor censitário 111):

```

/* ----- */
/* programa roclin.sas */
/* função: construir curva ROC e identificar cutoff de equilíbrio */
/* notação: sens = sensibilidade; espe = especificidade; */
/* sens1 = 1 - sens; esp1 = 1 - esp */
/* ----- */

data a;
  input sens espe sens1 esp1;
  x = esp1;
  y = 1. -x;
  cards;
1.000 0.224 0.000 0.776
0.939 0.327 0.061 0.673
0.788 0.395 0.212 0.605
0.667 0.558 0.333 0.442
0.424 0.701 0.576 0.299
0.333 0.830 0.667 0.170
0.242 0.925 0.758 0.075
0.242 0.973 0.758 0.027
0.000 0.973 1.000 0.027
0.000 0.986 1.000 0.014
0.000 0.993 1.000 0.007
;
run;

goptions device =win;

title h=1.0 'CURVA ROC DO MOD OMNI 111';

symbol1 v=circle c=blue i=join h=1.8 lline w=4;
symbol2 v=point c=red i=join w=2;

proc gplot;
  plot (sens y)* x /overlay frame
          vaxis = 0 to 1 by 0.2
          haxis = 0 to 1 by 0.2;
  label sens = 'SENSIBILIDADE'
         x = '1 - ESPECIFICIDADE';
run;

```