

MINISTÉRIO DA SAÚDE



Série: Capacitação e atualização em geoprocessamento em saúde

Introdução à Estatística Espacial para a Saúde Pública

Brasília • DF
2007

MINISTÉRIO DA SAÚDE
FUNDAÇÃO OSWALDO CRUZ

VOLUME

3

Série: Capacitação e atualização em geoprocessamento em saúde

Introdução à Estatística Espacial para a Saúde Pública

Série B. Textos Básicos de Saúde

Brasília - DF
2007

© 2007 Ministério da Saúde.

Todos os direitos reservados. É permitida a reprodução parcial ou total desta obra, desde que citada a fonte e que não seja para venda ou qualquer fim comercial.

A responsabilidade pelos direitos autorais de textos e imagens desta obra é da área técnica.

A coleção institucional do Ministério da Saúde pode ser acessada, na íntegra, na Biblioteca Virtual em Saúde do Ministério da Saúde: <http://www.saude.gov.br/bvs>

Série B. Textos Básicos de Saúde

Série Capacitação e Atualização em Geoprocessamento em Saúde; 3

Tiragem: 1.ª edição – 2007 – 1.200 exemplares

Elaboração, distribuição e informações:

MINISTÉRIO DA SAÚDE

Secretaria de Vigilância em Saúde

Departamento de Análise de Situação em Saúde
Esplanada dos Ministérios, bloco G
Edifício Sede, 1.º andar, sala 134
CEP: 70058-900, Brasília – DF
E-mail: svs@saude.gov.br
Home page: <http://www.saude.gov.br/svs>

Fundação Oswaldo Cruz

Instituto de Comunicação e Informação
Científica e Tecnológica em Saúde
Avenida Brasil 4365, Manguinhos,
Rio de Janeiro, RJ, CEP 21045-900
Home page: www.fiocruz.br

Equipe Técnica da Elaboração:

Ana Lúcia S. S. Andrade – IPTSP/UFGO
Antônio Miguel V. Monteiro – DPI/INPE
Christovam Barcellos – CICT/FIOCRUZ
Eugênio P. Lisboa – CGVS/SMSPOA
Lisiane M.W. Acosta – CGVS/SMSPOA
Maria Cristina de Mattos Almeida – SMSA/PBH
Maria Regina Varnieri Brito – CNVANT/SMS/POA
Marília Sá Carvalho – ENSP/FIOCRUZ
Marina Atanaka dos Santos – ISC/UMG
Oswaldo Gonçalves Cruz – PROCC/FIOCRUZ
Reinaldo Souza dos Santos – ENSP/FIOCRUZ
Rui Flores – CGVS/SMSPOA
Simonne Almeida e Silva – IPTSP/UFGO
Simone M. Santos – CICT/FIOCRUZ
Virginia Ragoni M. Correia – DPI/INPE
Wayner V. Souza – CpqAM/FIOCRUZ

Coordenação Executiva do projeto

Walter Massa Ramalho
Christovam Barcellos
Mônica M. F. Magalhães
Simone M. Santos

Revisão Técnico-pedagógica

Grácia Maria de Miranda Gondim
Simone M. Santos

Revisão de Português:

Itamar José de Oliveira

*Ilustração da capa, Programação Visual e
Direção de Arte:*

Vera Lucia Fernandes de Pinho

Fotos da capa:

Peter Illiciev / Banco de Imagens / Fiocruz

Tratamento de Imagens e Fotografias:

Os autores

Impresso no Brasil / Printed in Brazil

Ficha Catalográfica

Brasil. Ministério da Saúde. Secretaria de Vigilância em Saúde. Fundação Oswaldo Cruz.

Introdução à Estatística Espacial para a Saúde Pública / Ministério da Saúde, Fundação Oswaldo Cruz; Simone M. Santos, Wayner V. Souza, organizadores. - Brasília : Ministério da Saúde, 2007.

120 p. : il. – (Série B. Textos Básicos de Saúde) (Série Capacitação e Atualização em Geoprocessamento em Saúde; 3)

ISBN 978-85-334-1427-3

1. Estatística espacial. 2. Vigilância em Saúde. 3. Análise espacial. 4. Saúde pública. I. Simone M. Santos (Org.) II. Wayner V. Souza (Org.) III. Título. IV. Série.

NLM W 26.55.14

Catálogo na fonte – Coordenação-Geral de Documentação e Informação – Editora MS – OS 2007/1136

Títulos para indexação:

Em inglês: Spatial Statistics Introduction for Public Health

Em espanhol: Introducción a la Estadística Espacial en la Salud Pública

AUTORES

Ana Lúcia S. S. Andrade

Médica, doutora em Saúde Pública, professora titular do Departamento de Saúde Coletiva do Instituto de Patologia Tropical e Saúde Pública da Universidade Federal de Goiás.

Antônio Miguel Vieira Monteiro

Engenheiro, doutor em Engenharia da Computação, tecnólogo sênior da Divisão de Processamento de Imagens e Coordenador do Programa Espaço e Sociedade do Instituto Nacional de Pesquisas Espaciais.

Christovam Barcellos

Geógrafo, Sanitarista, doutor em Geociências, pesquisador titular do Departamento de Informações em Saúde do Centro de Informação Científica e Tecnológica da Fundação Oswaldo Cruz.

Eugênio P. Lisboa

Médico, Secretaria Municipal de Saúde de Porto Alegre.

Lisiane M.W.Acosta

Enfermeira, especialista em Enfermagem em Saúde Pública, mestranda em Epidemiologia, profissional da Equipe de Vigilância das Doenças Transmissíveis da CGVS da Secretaria Municipal de Saúde de Porto Alegre.

Maria Cristina de Mattos Almeida

Médica Veterinária, mestre em Saúde Pública, epidemiologista da Gerência de Epidemiologia e Informação da Secretaria Municipal de Saúde de Belo Horizonte.

Maria Regina Varnieri Brito

Médica, especialista em Saúde Pública e em Demografia, mestre em Saúde Pública (Epidemiologia). Médica da Equipe de Eventos Vitais e Coordenadora do Núcleo de Vigilância de Agravos Não Transmissíveis da Secretaria Municipal de Saúde de Porto Alegre.

Marília Sá Carvalho

Médica, mestre em Saúde Pública, doutora em Engenharia Biomédica, pós-doutorado em Estatística, pesquisadora titular do Departamento de Epidemiologia e Métodos Quantitativos da Escola Nacional de Saúde Pública Sérgio Arouca da Fundação Oswaldo Cruz.

Marina Atanaka dos Santos

Enfermeira, doutora em Saúde Pública, docente do Departamento de Saúde Coletiva do Instituto de Saúde Coletiva da Universidade Federal de Mato Grosso.

Oswaldo G. Cruz

Biólogo, mestre em Saúde Pública, doutor em Engenharia Biomédica, pesquisador do Programa de Computação Científica da Fundação Oswaldo Cruz.

Reinaldo Souza dos Santos

Biólogo, doutor em Saúde pública, pesquisador associado do Departamento de Endemias Samuel Pessoa da Escola Nacional de Saúde Pública Sérgio Arouca da Fundação Oswaldo Cruz.

Rui Flores

Médico, Sanitarista, Secretaria Municipal de Saúde de Porto Alegre.

Simone Maria dos Santos (organizadora)

Médica, especialista em Saúde Coletiva, mestra e doutoranda em Saúde Pública (Epidemiologia), pesquisadora colaboradora do Departamento de Informações em Saúde do Centro de Informação Científica e Tecnológica da Fundação Oswaldo Cruz.

Simonne Almeida e Silva

Médica, especialista em Medicina Preventiva e Social mestra e doutora em Medicina Tropical (Epidemiologia). Professora adjunta do Departamento de Saúde Coletiva do Instituto de Patologia Tropical e Saúde Pública da Universidade Federal de Goiás.

Wayner V. Souza (organizador)

Estatístico, mestre e doutor em Saúde Pública, tecnólogo sênior e docente do Departamento de Saúde Coletiva do Centro de Pesquisas Aggeu Magalhães da Fundação Oswaldo Cruz



Prefácio



O rápido desenvolvimento de tecnologias para análise de dados no espaço geográfico tem oferecido possibilidades inovadoras ao estudo da situação de saúde e de suas tendências, propiciando melhor compreensão dos fatores socioeconômicos e ambientais, entre outros, que determinam as condições de vida e o estado de saúde da população.

A apropriação desse instrumental pelo setor saúde é sumamente importante no Brasil, onde acentuadas desigualdades se evidenciam não apenas entre as grandes regiões, os estados e os municípios, mas também no interior destas áreas, sobretudo nos espaços urbanos. Sobrepondo-se ao mosaico de situações diferenciadas que caracteriza o território brasileiro, as transformações demográficas observadas nas últimas décadas potencializam a dinâmica de distribuição de doenças e agravos, bem como de demanda e acesso aos serviços.

Tudo isso implica enormes desafios para a organização e a atuação do setor saúde, impondo crescente necessidade de dados sistematizados e de informações para análise crítica, apresentadas oportunamente e de forma adequada à apreensão e uso pelas instâncias decisórias. A visualização espacial dos problemas de saúde e da forma como eles evoluem, correlacionadas aos seus determinantes e às medidas de intervenção que são propostas e colocadas em prática, constitui instrumento poderoso e complementar a outros modelos de análise com este fim.

No contexto brasileiro do setor saúde, tal propósito envolve considerável esforço de apropriação das tecnologias disponíveis e de desenvolvimento de estratégias que viabilizem o seu uso, em sintonia com processos de modernização gerencial da área de informação e informática em saúde na rede de serviços. O seu máximo aproveitamento requer mecanismos efetivos de articulação institucional e intersetorial em todas as esferas do Sistema

Prefácio



Único de Saúde (SUS), que promovam a harmonização de iniciativas e a mobilização dos recursos humanos existentes, em torno de objetivos comuns.

Nesse sentido, os núcleos especializados do Ministério da Saúde têm prestado, de maneira notável, a contribuição ao seu alcance, exercendo com denodo e eficiência a função de desenvolver instrumentos apropriados e de promover iniciativas de capacitação e atualização de profissionais do SUS nas técnicas de geoprocessamento em saúde. Este terceiro livro da série representa, portanto, um novo marco importante no processo de aperfeiçoar a gestão da saúde no Brasil.

João Baptista Risi Junior

*Consultor da Organização Pan-Americana de Saúde
Área técnica de Informação para políticas de saúde*

Apresentação



Para além dos mapas coropléticos (temáticos) é o que pretende o livro **Introdução à Estatística Espacial para a Saúde Pública**, mostrando ao leitor os primeiros passos dos pressupostos estatísticos nas abordagens espaciais. As técnicas estatísticas apresentadas ampliam as análises de modo a permitir a realização de inferências e de testes de hipóteses geradas pelos profissionais de saúde pública para responder questões sobre a distribuição de eventos no seu dia-a-dia.

Esse esforço de tradução da estatística para uma linguagem acessível, pela redução das notações e terminologias estatísticas e identificação de conceitos fundamentais dos pressupostos de uso destas técnicas, é uma meta direcionada aos profissionais do SUS. A crescente demanda por maior capacidade analítica dos seus técnicos, e o desafio de transformação da explosão de dados disponíveis em informações úteis pode ser acompanhada pelo refinamento das análises. Essa tradução de linguagem, que não é trivial, foi executada com sucesso graças à experiência do corpo autoral desta publicação. Participam como autores deste volume um grupo de especialistas (docentes, pesquisadores e técnicos de serviços de saúde) envolvido com o desenvolvimento da área no Brasil, e que há vários anos vem utilizando e propondo técnicas de análise espacial aplicadas à saúde.

Contribuindo nesse esforço, grupos de pesquisadores vêm produzindo Sistemas de Informação Geográfica com pacotes aplicados à análise espacial, que apresentam características animadoras como: a facilidade da interface; a geração de bancos de dados geográficos; a possibilidade de incorporar pequenos aplicativos denominados de “*plug-in*”; e, especialmente o fato de possuir bibliotecas abertas para desenvolvedores externos. O aplicativo TerraView, desenvolvido pelo Instituto Nacional de Pesquisas Espaciais, instituição parceira e atuante na área, é o ambiente computacional de referência, também neste volume. Como alguns procedimentos ainda não estão disponíveis no TerraView, no momento desta edição, lançou-se mão de outros *softwares* compatíveis e gratuitos, como o SatScan e o CrimeStat, para execução de algumas técnicas.

Apresentação



O primeiro capítulo possibilita um passeio introdutório por conceitos da Análise de Dados Espaciais a partir de conceitos da estatística geral para a estatística espacial, de forma suave e encadeada, buscando interagir com o leitor por meio de exemplos cotidianos da realidade brasileira.

Os dois capítulos que seguem tratam de facetas específicas da natureza cartográfica da representação das informações disponíveis – os pontos e os polígonos (áreas). Os autores abordam as limitações e as possibilidades de análise estatística espacial adequada aos indicadores, de acordo com sua topologia.

Complementando o conteúdo teórico, três exercícios dirigidos são apresentados para a aplicação prática do aprendizado. Os problemas abordados têm base na experiência das atividades de Vigilância em Saúde. Os bancos de dados necessários para a execução das análises propostas estão disponíveis para *download* no endereço do projeto descrito abaixo.

Ao tratar desta publicação, tentamos contemplar o público com uma apresentação fundamentada em técnicas estatísticas, focando nas estimações e especialmente nas interpretações. Por meio da adequação da linguagem, esperamos ter alcançado os objetivos expostos, cativando o leitor e estimulando suas incursões nesse universo de técnicas estatísticas. Gostaria ainda de expressar a grande felicidade na participação em um projeto de desenvolvimento que toma corpo como uma rede que reúne saberes e especialmente práticas que permitem viabilizar a aplicação dos desenvolvimentos da geotecnologia à saúde no Brasil. Assim, como este projeto é de construção permanente, o TerraView está sempre disponibilizando ferramentas novas (www.dpi.inpe.br), e o projeto de Capacitação e Atualização em Geoprocessamento em Saúde (www.capacita.geosaude.cict.gov.br) oferecerá produtos voltados para aplicações no setor saúde. Participe, entre em contato, siga os exercícios e boa sorte!

Walter M. Ramalho

Coordenação Executiva do projeto de Capacitação e
Atualização em Geoprocessamento para a Saúde Pública



Sumário

Capítulo 1: Conceitos Básicos em Análise de Dados Espaciais em Saúde ... 13

1.1 - Estudos Ecológicos	15
1.2 - Estatística Espacial	17
• Principais Aplicações	20
• Tipos de Dados	21
A- Pontos	21
B- Áreas	22
• Conceitos Estatísticos Fundamentais	23
A- Dependência espacial	23
B- Estacionariedade	25

Capítulo 2: Análise de Dados em Forma de Pontos 29

2.1 - Princípios de Análise Espacial de Eventos com Representação por Pontos	31
2.2 - Análise Espacial de Eventos	33
• Visualização de Dados Espaciais	33
• Análise Exploratória de Dados Espaciais	33
• Modelagem de Dados Espaciais	33
2.3 - Análise de Padrões de Dados com Representação por Pontos	34
2.4 – Análises Baseadas na Distância entre Pontos	38
• Análise de Aglomerados Hierárquicos	39
A- Distância limiar	40
B- Número mínimo de pontos	41
C- Tamanho da elipse	41
2.5 – Análise da Densidade de Pontos	44
2.6 - Avaliação de Aglomerados considerando a Distribuição da População	48
• Detecção de Aglomerados pela Técnica Estatística Scan	51
• Estimativa Kernel considerando a População	54
2.7 - Interpolação de Dados Pontuais	56

Capítulo 3: Análise Espacial de Dados de Áreas 61

3.1 - Métodos para Espacialização e Tratamento Estatístico de Indicadores	63
• Instabilidade em Pequenas Áreas	63
• Matriz de Vizinhança	65
3.2 - Métodos de Suavização Espacial	66
• Média Móvel Espacial	66
• Estimador Bayesiano Empírico	66
• Autocorrelação Espacial	69
3.3 - Análise da Estrutura Espacial	71
• Análise da Estrutura Espacial - Estatística de Moran	71
• Análise da Estrutura Espacial - Estatística-Função LISA	72
• Análise da Estrutura Espacial - Métodos Inteiramente Bayesianos	73
• Análise da Estrutura Espacial - Método de Markov Chain-Monte Carlo	75
3.4 - Técnicas de Análise de Dados Pontuais Aplicadas a Áreas	78

Anexos: Exercícios 83

Exercício 1: Vigilância e atenção à saúde de gestantes HIV positivas: construção de indicadores com operações entre camadas	85
Exercício 2: Distribuição dos acidentes de trânsito em Belo Horizonte, MG.	97
Exercício 3: Análise espacial da distribuição de casos de malária no Estado de Mato Grosso em 2000 e 2003	105

Referências Bibliográficas 117**Links úteis para download de aplicativos recomendados 119**



Capítulo

Conceitos Básicos em Análise de Dados Espaciais em Saúde

Marilia Sá Carvalho

Oswaldo G. Cruz

Wayner V. Souza

Antônio Miguel V. Monteiro

Capítulo 1:

Conceitos Básicos em Análise de Dados Espaciais em Saúde

1.1 - Estudos Ecológicos

pág. 15

1.2 - Estatística Espacial

pág. 17

Principais Aplicações

pág. 20

Tipos de Dados

pág. 21

Conceitos Estatísticos
Fundamentais

pág. 23

A- Pontos

B- Áreas

A- Dependência espacial

B- Estacionariedade

1.1 - Estudos Ecológicos

O estudo dos padrões de distribuição geográfica das doenças e suas relações com fatores socioambientais constitui-se no objeto do que hoje chamamos de Epidemiologia Geográfica, que tem se constituído em campo de aplicação e desenvolvimento de novos métodos de análise. Esse desenvolvimento tem sido viabilizado pela crescente disponibilidade e desenvolvimento de técnicas e recursos de computação eletrônica, que também tornou possível o desenvolvimento de Sistemas de Informações Geográficas com base na cartografia digital, que vêm sendo utilizados em análise de dados espaciais em saúde.

As técnicas de análise espacial se adequam às necessidades dos estudos ecológicos, que utilizam áreas geográficas como unidade usual de observação. No caso particular da análise espacial da ocorrência de doenças, onde se dispõe de dados referidos a uma área geográfica, deve-se dar atenção especial ao pressuposto básico de que taxas e indicadores epidemiológicos de uma doença, calculados para um determinado período, são uma única realização de um processo que é probabilístico e que se procura identificar.

A opção por estudos epidemiológicos do tipo ecológico pode ser logicamente adequada para o entendimento da variação do risco de adoecer entre diferentes grupos populacionais. A correta compreensão das dimensões envolvidas nos estudos de nível ecológico e de nível individual, além das relações entre eles, possibilita explorar o potencial da abordagem ecológica em saúde pública, mediante o emprego de ferramentas de análise apropriadas que permitam controlar possíveis fatores de confusão e evitar vieses de análise, principalmente os clássicos, como os de produzir inferências cruzadas, ou seja, de um nível para outro.

Inferências enviesadas serão produzidas não só quando se extrapolam resultados do nível ecológico para o individual (falácia ecológica), mas também quando ocorre o oposto, do nível individual para o ecológico (falácia atomística), desconsiderando-se os efeitos de grupo. Apesar de não se dever considerar essas duas abordagens como antagônicas, a análise da variabilidade do risco no nível ecológico é fundamental para a compreensão dos determinantes sociais e ambientais do processo saúde-doença, onde, particularmente, a condição socioeconômica dos grupos populacionais desempenha papel preponderante na explicação das condições de saúde desses grupos. Nesse sentido, uma importante abordagem do ponto de vista ecológico é considerar o espaço como fator multidimensional de estratificação de populações, o que

permite visualizar espacialmente, de forma articulada, as distribuições da doença e do risco, entendido como situação coletiva e definido sob diferentes abordagens. Além disso, pode propiciar ao Sistema de Vigilância em Saúde Pública compreender a dinâmica e identificar diferenças e “causas” das diferenças na ocorrência das doenças nos diversos subconjuntos da população.

QUADRO 1.1 – Classificação dos Estudos Epidemiológicos

TIPO	POSIÇÃO DO INVESTIGADOR	REFERÊNCIA TEMPORAL	DENOMINAÇÃO
Agregado	Observacional	Transversal	Estudos Ecológicos
		Longitudinal	Estudos de tendência ou Séries Temporais
	Intervenção	Longitudinal	Ensaio comunitários
Individual	Observacional	Transversal	Inquéritos
		Longitudinal	Estudos Prospectivos (Coorte)
			Estudos Retrospectivos (Caso-Control)
	Intervenção	Longitudinal	Ensaio Clínicos

Essa abordagem possibilita não só a vigilância dos indivíduos de risco, mas também uma visão antecipada do risco coletivo, coerentemente com o entendimento de que as conseqüências dos processos endêmicos e das intervenções de saúde pública nos vários grupos sociais devem ser analisadas a partir de uma perspectiva ecológica. Tal perspectiva deve considerar a dinâmica evolutiva dos processos sociais e buscar identificar causas da incidência das doenças em grupos populacionais e não as causas da doença no indivíduo. Cada vez mais, trabalhos vêm sendo desenvolvidos utilizando uma abordagem espacial e produzindo análises integradas da associação dos eventos de saúde com indicadores e/ou variáveis socioambientais, com vistas a suprir as necessidades de um Sistema de Vigilância em Saúde.

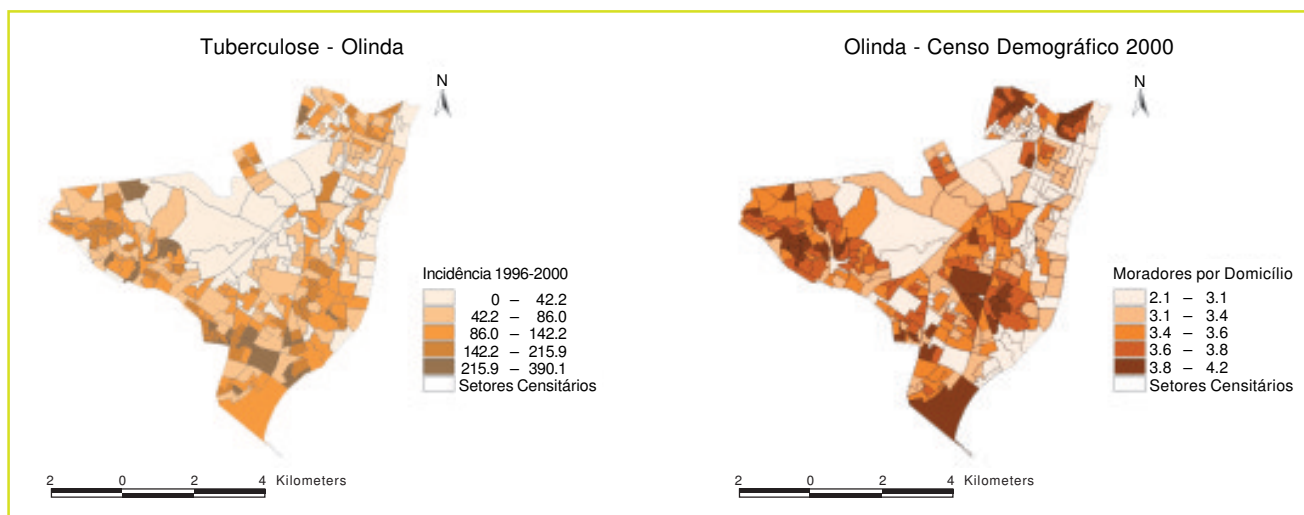
As fontes de informação para essas variáveis e as abordagens para construção desses indicadores, principalmente aqueles capazes de medir índices de carência social, variam e vêm sendo objeto de estudos e debates que podem contribuir para superar desigualdades e iniquidades em saúde.

Em muitos casos essas informações são provenientes dos censos demográficos, disponibilizadas no nível dos setores censitários, possibilitando análises para este nível ou níveis mais agregados, como bairros, distritos e municípios. No caso dos indicadores ambientais especificamente, as informações são obtidas junto a institutos municipais ou estaduais de meio

ambiente ou planejamento urbano, oriundas de levantamentos de campo, levantamentos aéreos ou de imagens de satélite.

A título de exemplo, apresentamos na figura 1.1, um mapa temático representando as taxas médias de incidência de tuberculose, segundo setores censitários do Município de Olinda-PE, para o período 1996-2000 e outro representando a média de moradores por domicílio, para estes mesmos setores censitários de Olinda, segundo o Censo Demográfico de 2000.

FIGURA 1.1 – Olinda: taxa média de incidência de tuberculose no período 1996-2000, e média de moradores por domicílio em 2000, segundo setores censitários.



Fontes: SINAN e Censo Demográfico 2000 - IBGE.

Pode-se observar, mesmo que apenas por inspeção visual, uma coincidência entre vários setores com altas taxas de incidência e altas concentrações de moradores por domicílio. Essa associação espacial entre o indicador escolhido e a incidência de tuberculose não é casual, dado que altas médias de moradores por domicílio apontam para carência social, além de se constituir em fator facilitador da transmissão da doença.

1.2 – Estatística Espacial

Denomina-se estatística espacial o ramo da estatística que permite analisar a localização espacial de eventos. Ou seja, além de identificar, localizar e visualizar a ocorrência de fenômenos que se materializam no espaço, tarefas possibilitadas pelo uso dos SIG, utilizando-se a estatística espacial é possível modelar a ocorrência destes fenômenos, incorporando, por exemplo, os fatores determinantes, a estrutura de distribuição espacial ou a identificação de padrões.

SNOW

Londres, 1854. Região sul, subdistrito de St. Saviour em Southwark. Localidade, paróquia de Christchurch. Entre os dias 9 de julho e 5 de agosto. Livro de Registros Gerais.

1. n. 34, Charlotte Street, 29 de julho, um coronheiro, 29 anos, “Cólica Asiática, 18 horas”. Atendimento de água feito por Lambeth Water Company.
2. n. 45, Gravel Lane, 1º de agosto, a viúva de um fazendeiro, 48 anos, “Cólica, 12 horas”. Atendimento de água feito por Southwark and Vauxhall Company.
3. n. 1, Alpha Place, 1º de agosto, o escrevente de um advogado, 57 anos, “Cólica, 24 horas”. Atendimento de água feito por Southwark and Vauxhall.
4. ...

Ver o debate do Lancet, June 23, 1855, reproduzido em:

<http://www.ph.ucla.edu/epi/snow/reactionandcommitteeaction.html>.

Se tomarmos o exemplo descrito no Box acima, vê-se que o médico inglês John Snow, reconhecendo que os casos de cólera ocorriam em certas localizações, fez uso direto dos mapas de Londres (Figura 1.2). Ele localizou os casos encontrados no registro de óbitos, utilizando o endereço das residências e os poços de provisão de água existentes naquelas ruas. Se fosse hoje, John Snow só publicaria os resultados de seu estudo e fecharia a bomba de água de Broad Street (Figura 1.3), se pudesse comprovar estatisticamente que esta bomba era responsável pela propagação da epidemia de cólera. (ver o site <http://www.ph.ucla.edu/epi/snow.html>)

FIGURA 1.2 – Mapa do Cólera – Soho, Londres.



FIGURA 1.3 – Bomba de água de Broad Street, Londres.



Que tipo de estatística deveria ser usada?

As técnicas estatísticas mais usadas (por exemplo, o teste chi-quadrado), são baseadas em pressupostos que dificultam, ou mesmo impedem, seu uso quando a pergunta relaciona-se a problemas onde a localização espacial é relevante.

Essas são as questões de fundo que iremos abordar neste livro.

Usamos a estatística nos estudos em saúde para modelar a realidade. Modelos são simplificações da realidade usadas para entender um sistema, estudar seu funcionamento, buscar causas de fenômenos, avaliar intervenções, prever desfechos. Segundo Box (1979) “todos os modelos estão errados, alguns modelos são úteis”, ou seja, não existe modelo certo, mas modelos que servem a algum propósito.

Quando então são úteis os modelos estatísticos na análise espacial? Quando queremos estudar a saúde e os fatores relacionados à saúde no contexto ambiental. A idéia é entender, estimar e modelar como esse contexto afeta a saúde das pessoas, seja o contexto derivado de fatores socioeconômicos, da oferta de serviços de saúde, do ambiente físico ou cultural. Sem medir esses contextos, nem padrão de mortalidade e morbidade, nem o espalhamento epidêmico, ou o risco de uma fonte de contaminação ambiental podem ser explicados e ter seu efeito estimado.

A denominação estatística espacial surge em oposição às técnicas estatísticas comuns nos estudos em saúde – testes do tipo chi-quadrado, regressão múltipla, análise de aglomerados, por exemplo – quando se necessita focalizar explicitamente a localização espacial, quando é necessário considerar a possível importância de seu arranjo espacial na análise e interpretação de resultados. Isso porque um dos pressupostos mais gerais na estatística, que permite estimar diversas estatísticas é o da independência entre as observações: o que se mede em um indivíduo não está associado ao medido em outro indivíduo, exceto pelos possíveis fatores comuns que desejamos exatamente descobrir.

O teste **Chi-Quadrado** (ler qui-quadrado) nos diz em que medida os valores observados de duas variáveis categóricas se desviam do valor esperado, caso as duas variáveis não estivessem correlacionadas. Quanto maior o chi-quadrado, mais significativa é a relação entre as duas.

A **regressão linear** é um método para se estimar o valor esperado de uma variável *dependente* dado o valor de um variável *independente*. É chamada “linear” porque se considera que a relação da resposta às variáveis é uma função linear de alguns parâmetros. Quando se ajusta para mais de uma variável independente é chamada regressão múltipla.

Na **análise de aglomerados**, freqüentemente chamada de análise de *cluster*, o objetivo é separar um conjunto de objetos em grupos conforme um padrão de similaridade, medida a partir de um conjunto de variáveis. Esse padrão de similaridade é estabelecido em um espaço n-dimensional dado pelas variáveis medidas para cada observação.

Entretanto, quando olhamos para o espaço, esse pressuposto é pouco realístico, pois “todas as coisas são parecidas, mas coisas mais próximas se parecem mais que coisas mais distantes” (Tobler, 1979). Esse simples fato, que todos conhecemos, tem implicações diretas para abordagem quantitativa baseada na teoria e técnicas estatísticas que aplicamos usualmente sobre nossas observações. Por uma razão muito simples, e que às vezes esquecemos: as técnicas de análise estatística não-espaciais que utilizamos para nossos dados têm como pressuposto fundamental que as observações em análise representam pedaços de evidência independentes sobre as associações que estamos descrevendo e modelando.

Ou, generalizando, a maior parte das ocorrências, sejam estas naturais ou sociais, se relacionam, seja em sinergia ou antagonismo, e esta relação se enfraquece com a distância. E é esta a especificidade deste ramo da estatística: técnicas que permitem modelar os fenômenos cuja distribuição é afetada pela sua localização geográfica e pela sua relação com seus vizinhos.

• Principais Aplicações •

As principais áreas de aplicação da estatística espacial são o mapeamento de doenças, os estudos ecológicos, a identificação de aglomerados espaciais (*cluster*) e o monitoramento de problemas ambientais.

O mapeamento de doenças consiste na descrição do processo de distribuição espacial, visando a avaliar a variação geográfica na sua ocorrência para identificar diferenciais de risco, orientar a alocação de recursos e levantar hipóteses etiológicas. Os métodos têm como objetivo produzir um mapa “limpo”, sem o “ruído” gerado pela flutuação aleatória dos pequenos números, e controlando as diferenças na estrutura demográfica.

Os estudos ecológicos visam a estudar a relação entre incidência de doenças e potenciais fatores etiológicos, que expliquem as diferenças na incidência de determinado evento de saúde. Do ponto de vista estatístico, esses estudos consistem, essencialmente, em modelos de regressão, onde se busca explicar a variação na incidência da doença por meio de outras variáveis. O modelo estatístico se complica pela necessidade de controlar, simultaneamente, o processo espacial, variáveis explicativas e variáveis de confusão.

Um “**cluster**” **espacial** é: qualquer agregado de eventos que não seja meramente casual, cuja identificação é foco de pesquisas na área de estatística espacial. Estes aglomerados podem ser causados por diferentes fatores, tais como agentes infecciosos, contaminação ambiental localizada, efeitos colaterais de tratamentos, cada problema destes com peculiaridades e técnicas particulares.

A avaliação e monitoramento ambiental visam a estimar a distribuição espacial de fatores ambientais relevantes para a saúde, acompanhando potenciais fontes ambientais de problemas de saúde, tais como poluentes químicos, insolação (Raios UV), vegetação, clima, entre outros. Nesse caso, em geral, os modelos estatísticos têm por objetivos fazer a predição espacial ou espaço-temporal de processos com forte correlação espacial e temporal. Esse tipo de técnica não será objeto deste livro.

• Tipos de Dados •

Da mesma forma que o SIG tem diferentes funções dependendo do tipo de dado, algumas técnicas estatísticas são mais apropriadas para um ou outro tipo de dado. Basicamente, são dois tipos que utilizamos: (A) dados que são atributos de **pontos**, ou (B) dados atributos de **áreas**.

A- Pontos

O tipo mais simples é o dado de ponto. Nesse caso temos a localização pontual (coordenadas) de algum evento, o mais comum é o da residência de um caso de doença. É importante observar que o que é aleatório, ou seja, o que será analisado estatisticamente, é a localização. A pergunta associada a esse tipo de dado é sempre o ONDE, mesmo quando outras co-variáveis do indivíduo são medidas.

O melhor exemplo novamente é a investigação da epidemia de cólera em Londres, 1854. John Snow localizou as residências de cada caso de cólera, a idade, profissão e outras variáveis de interesse, e, mais importante, a fonte da água utilizada e a localização dos locais de captação de cada companhia. Suponhamos então que fossemos nós a fazer a pesquisa.

Duas perguntas são importantes. A primeira busca analisar a distribuição espacial dos casos de cólera, verificando se estão distribuídos homoganeamente na cidade, ou ao contrário, concentram-se em alguns locais em detrimento de outros. Vale lembrar que na época a teoria miasmática era dominante, com suporte do Dr. William Farr, que publicou estudo atribuindo aos locais mais baixos, onde os miasmas se concentram, as maiores taxas de mortalidade por cólera.

A segunda pergunta busca estimar o efeito de cada companhia de água sobre a mortalidade. Esse foi um dos aspectos analisados que de fato definiu os rumos da pesquisa. Se fosse hoje, poderíamos analisar simultaneamente a localização geográfica e a companhia de água, o que não era possível na época. John Snow considerou que os casos eram amostras independentes, o que não é tão verdadeiro. Cólera também se transmite de pessoa a pessoa, diretamente ou através de objetos contaminados, embora a magnitude da contaminação pela água seja muito maior.

As técnicas de análise de dados de pontos serão abordadas no Capítulo 2.

B- Áreas

Outro tipo de dado é o proveniente de áreas geográficas com limites definidos, usualmente divisões político-administrativas. São as contagens de casos ou óbitos de alguma doença ou os indicadores provenientes do censo demográfico.

É importante observar a escala em que esses dados são coletados e analisados. Em geral, quanto maior a área geográfica, mais misturada é a população onde ocorre o processo em estudo. Municípios, muito freqüentemente, são compostos por diferentes populações, com diferentes características e riscos à saúde, por exemplo, englobando populações rurais e urbanas, diferentes níveis sociais, acesso a serviços de saúde ou problemas ambientais. No outro extremo, os setores censitários têm população mais semelhante do ponto de vista socioeconômico e demográfico. Ou seja, a capacidade de diferenciar geograficamente regiões diminui com o tamanho da área e respectiva população residente.

Entretanto, quando consideramos pequenas áreas, a disponibilidade do dado começa a ser mais complicada. Informações socioeconômicas por setor censitário são obtidas facilmente, mas contagens de óbitos já são de obtenção mais difícil. Além disso, quanto menor a área, mais raros são os eventos observáveis, o que origina excessiva **flutuação aleatória dos indicadores** e todos os problemas daí decorrentes.

Flutuação aleatória dos indicadores

Como o próprio nome indica, significa que o indicador está variando aleatoriamente, sem qualquer padrão explicável. Ou seja: se em uma área rural nasceram 5 crianças em um dado ano e uma faleceu no primeiro ano de vida, a mortalidade infantil seria 1/5, ou seja, 200 por 1000 nascidos vivos. Entretanto, um epidemiologista sensato, antes de divulgar número tão grave, lembraria que entre 5 recém-natos, um óbito pode ser simplesmente ocasionado por algum acaso, não sendo justo julgar as condições de saúde do município somente por este dado. Porque população no denominador do indicador é muito pequena, o estimador fica instável, flutua aleatoriamente.

Dados de áreas são os mais analisados na área da saúde, embora tenham sua importância freqüentemente subestimada por receio da denominada falácia ecológica, que pode ser definida como uma tentativa de estimar associações entre indivíduos a partir de dados agregados. Por exemplo, ao analisar o problema dos atropelamentos, observamos que as taxas são maiores nos municípios mais ricos. Nunca poderíamos concluir que os indivíduos mais ricos é que são atropelados, o que não seria verdade, mas sim, que nos municípios mais ricos, há mais carros, e estes atropelam os pedestres, que são em geral crianças dos estratos mais pobres da população.

Entretanto, também modelos em nível individual podem ser mal especificados. Cabe ressaltar que existem dois tipos de perguntas etiológicas. A primeira busca as causas dos casos, e a segunda as causas da incidência, sendo rara a doença cuja taxa de incidência não tenha variado largamente, seja ao longo do tempo ou entre populações. As técnicas de análise de dados de áreas serão apresentadas no Capítulo 3.

Há uma terceira situação que são os pontos de localização de amostras, muito utilizados na análise **Geoestatística**. Nesse caso, o dado é uma medida obtida em um ponto do espaço amostrado, portanto, apesar de ser atribuída

ao ponto, a informação do dado é referente à área amostrada (geralmente um espaço contínuo, sem delimitação explícita). Esse ponto, além de uma coordenada geográfica, possui um atributo na forma de um valor medido. Por exemplo, um ponto de coleta meteorológica onde temos uma amostra de variáveis que ocorrem de uma forma contínua, amostradas num determinado ponto. Esse tipo de dado é o mais freqüentemente utilizado em variáveis ambientais. As técnicas mais usuais na análise desse tipo de dados são voltadas para a predição espacial, ou seja, para estimar valores da variável em locais onde não houve medida a partir das mensurações feitas. Uma das técnicas mais utilizadas para a predição é denominada krigagem. Neste livro não iremos abordar esse tipo de técnica. Para maiores detalhes ver Câmara *et al.*, 2004.

• Conceitos Estatísticos Fundamentais •

Ainda que cada um desses objetivos deva ser tratado com técnicas estatísticas específicas, alguns conceitos estatísticos fundamentais definem os problemas metodológicos tratados pela estatística espacial. O entendimento da noção de: (A) **dependência espacial** e (B) **estacionariedade**, orientam a identificação de pressupostos a serem assumidos por diferentes técnicas estatísticas.

A- Dependência espacial

O primeiro destes conceitos é a noção de **dependência espacial** e sua quantificação através de funções como a de autocorrelação espacial.

A **Autocorrelação espacial** é uma medida geral da associação espacial existente no conjunto dos dados. Seu valor varia de -1 a 1 . Valores próximos de zero, indicam a inexistência de autocorrelação espacial significativa entre os valores dos objetos e seus vizinhos. Valores positivos para o índice, indicam autocorrelação espacial positiva, ou seja, o valor do atributo de um objeto tende a ser semelhante aos valores dos seus vizinhos. Valores negativos para o índice, por sua vez, indicam autocorrelação negativa.

Correlograma é uma representação gráfica da autocorrelação, em função da distância, ou seja, indicando em que medida cada dado em uma determinada distância se relaciona com os demais dados.

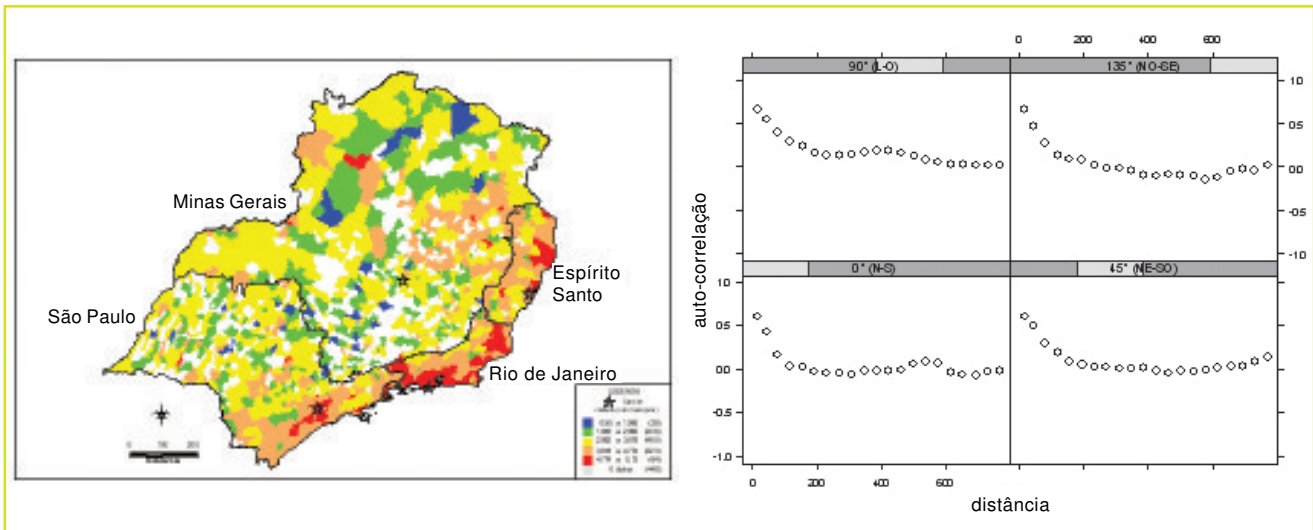
Os métodos básicos aqui são voltados para obter evidências da estrutura de dependência espacial, ou seja, autocorrelação espacial que é uma medida da estrutura de Dependência Espacial. Caso haja evidência de estrutura espacial, o postulado de independência das amostras, base da maior parte dos procedimentos de inferência estatística, é inválido, sendo necessário levar em conta, explicitamente, o espaço na formulação e estimativas dos modelos estatísticos.

Dados espaciais não formam um conjunto de amostras independentes.

Uma importante diferença em relação a dados sem essa peculiaridade é que cada observação não traz uma informação independente, e o conjunto de todas as observações é utilizado integralmente para descrever o padrão do fenômeno estudado.

O correlograma é uma representação gráfica das autocorrelações de um conjunto de dados em uma determinada distância.

FIGURA 1.4 – Autocorrelação Espacial.

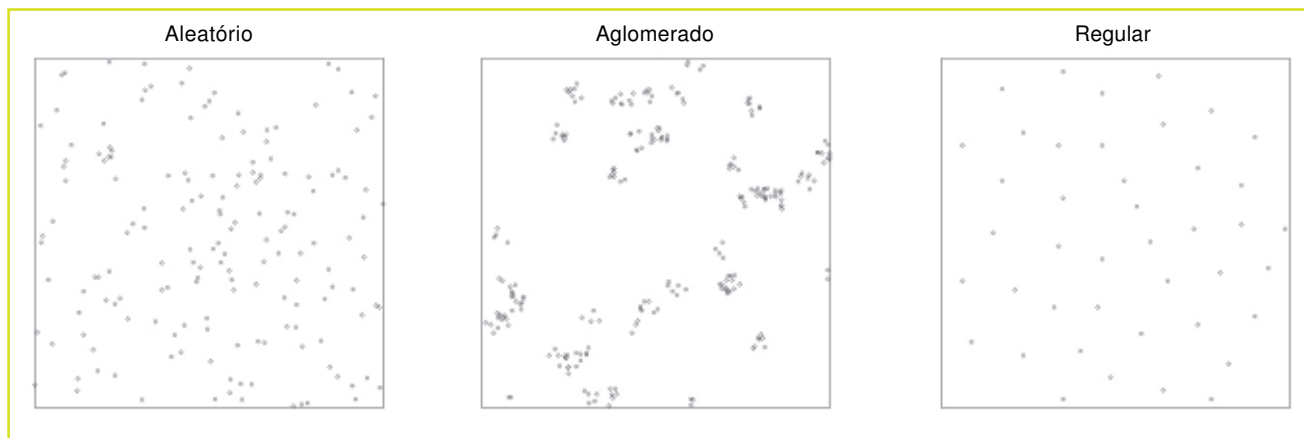


Pode-se notar na figura 1.4 que a autocorrelação nos quatro estados da região sudeste apresentam diferentes padrões. No RJ a autocorrelação decai rapidamente, na medida que a distância aumenta, em SP a autocorrelação decai mais lentamente, no ES a autocorrelação decai voltando a subir, fato que pode ser interpretado pela existência de *cluster* locais, ou ainda o pequeno número de municípios existente nas maiores distâncias do correlograma. O estado de MG apresenta um padrão de correlação próximo a zero, ou seja, aparentemente existe uma completa aleatoriedade espacial.

Dois questões estatísticas devem levar em consideração a presença de dependência espacial: a **identificação de padrões espaciais** e a **análise do efeito** de algum fator de risco sobre um desfecho controlando para a localização geográfica. Para a **identificação de padrões espaciais** (Figura 1.5), precisamos estimar a presença, forma e intensidade da dependência espacial. Diversos métodos foram desenvolvidos para essa finalidade, entre os quais apresentaremos alguns testes para verificar afastamentos da completa **aleatoriedade espacial**, outros que permitem identificar fontes suspeitas de risco à saúde, e outros ainda que localizam as regiões com concentração de casos acima de um valor esperado.

Espacialmente aleatória é qualquer ocorrência cuja distribuição espacial não apresente qualquer padrão espacial detectável. Observe os gráficos abaixo. No primeiro, os pontos estão distribuídos aleatoriamente. No segundo estão aglomerados (*clusterizados*), e no terceiro distribuídos de forma regular, ou seja, não aleatória.

FIGURA 1.5 – Padrões de distribuição espacial de pontos.



Para a **análise de efeito**, o que se está discutindo é qual seria o efeito da presença de dependência espacial ao fazer, por exemplo, um modelo de regressão linear múltipla. Suponhamos, por exemplo, um estudo onde se deseja estimar o efeito da densidade demográfica sobre a taxa de homicídios em cada município da Região Sudeste. Um modelo simples de regressão permitiria estimar essa relação, se considerássemos que a ocorrência da violência em cada município é independente dos demais. Só que não é. Ignorar essa dependência levará a superestimar o efeito da variável independente, pois cada município analisado contém menos informação do que em situação de independência. Na verdade, não considerar essa questão pode afetar seriamente os resultados obtidos de qualquer modelo estatístico.

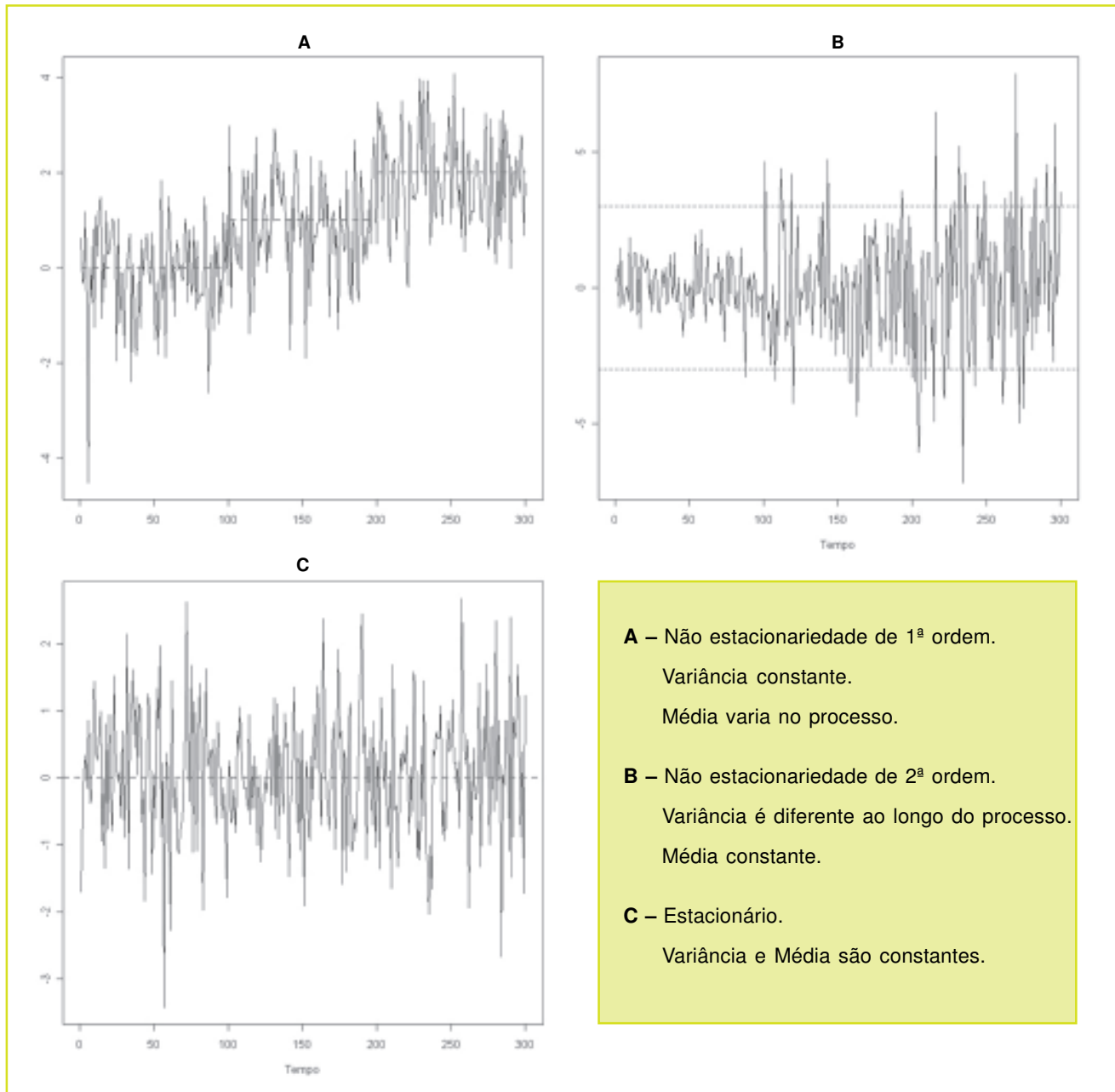
B- Estacionariedade

O outro conceito fundamental é denominado **estacionariedade**. Se a média do processo cuja ocorrência espacializada estudamos não é aproximadamente constante em toda a região, dizemos que o processo não apresenta estacionariedade de primeira ordem. Esse conceito fica mais claro quando pensamos em uma série temporal: quando a média aumenta (ou diminui) ao longo do tempo, dizemos que há uma tendência a aumentar (ou diminuir). No espaço é a mesma idéia. Apenas, em vez de termos uma linha no tempo, temos uma superfície no espaço.

Observe que pode ser difícil avaliar a estacionariedade, pois a flutuação aleatória usual dificulta a visualização (Figura 1.6). Porém identificar esta

questão é importante, não só porque afeta os estimadores estatísticos, mas principalmente porque esta variação da média no espaço pode ser importante para a saúde pública.

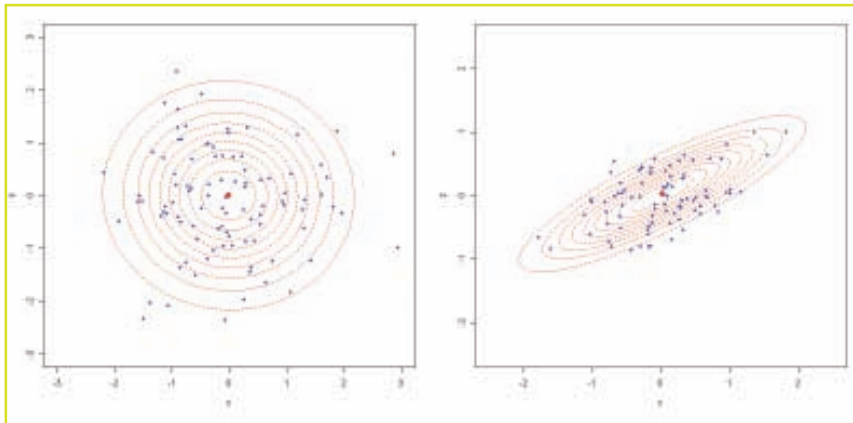
FIGURA 1.6 – Exemplos de análise da estacionariedade numa Série Temporal.



Vimos a **não estacionariedade de primeira ordem**, ou seja, a média variando no espaço. Porém, da mesma forma que ao longo do tempo a variância pode ser não estacionária, ou seja, não temos estacionariedade de segunda ordem. Nesse caso, a variabilidade em determinadas regiões difere. Usualmente, temos estacionariedade de primeira e segunda ordem, pois quando a média não varia, também a variância na sub-região se mantém.

Até agora conceituamos estacionariedade de forma muito similar às séries temporais. Porém há um aspecto que é exclusivo da estatística espacial: a **isotropia** (Figura 1.7). Um processo é isotrópico quando seu comportamento é igual em todas as direções. Ou seja, quando a dependência espacial é igual na direção Norte-Sul ou Leste-Oeste. Um exemplo de processo anisotrópico é a densidade populacional do Brasil. A diminuição de densidade na direção Leste-Oeste, ou em direção ao interior do país, é mais intensa do que quando se caminha na direção do Sul ao Norte. A força da dependência espacial decai mais rapidamente nesse eixo do que acompanhando o litoral.

FIGURA 1.7 – Isotropia e Anisotropia.



Os conceitos apresentados neste capítulo descreveram um conjunto básico de propriedades inerentes aos dados espaciais que de acordo com suas características fazem com que devam ser analisados de forma diferenciada. Nos próximos capítulos serão apresentadas algumas das principais técnicas que podem ser aplicadas na análise de dados espaciais em saúde.



Capítulo

Análise de Dados em Forma de Pontos

Christovam Barcellos

Simonne Almeida e Silva

Ana Lúcia S. S. de Andrade

Capítulo 2:

Análise de Dados em Forma de Pontos

2.1 - Princípios de Análise Espacial de Eventos com Representação por Pontos *pág. 31*

2.2 - Análise Espacial de Eventos *pág. 33*

Visualização de Dados Espaciais *pág. 33*

Análise Exploratória de Dados Espaciais *pág. 33*

Modelagem de Dados Espaciais *pág. 33*

2.3 - Análise de Padrões de Dados com Representação por Pontos *pág. 34*

2.4 - Análises Baseadas na Distância entre Pontos *pág. 38*

Análise de Aglomerados Hierárquicos *pág. 39*

A- Distância limiar

B- Número mínimo de pontos

C- Tamanho da elipse

2.5 - Análise da Densidade de Pontos *pág. 44*

2.6 - Avaliação de Aglomerados considerando a Distribuição da População *pág. 48*

Detecção de Aglomerados pela Técnica Estatística Scan *pág. 51*

Estimativa Kernel considerando a População *pág. 54*

2.7 - Interpolação de Dados Pontuais *pág. 56*

2.1 – Princípios de Análise Espacial de Eventos com Representação por Pontos

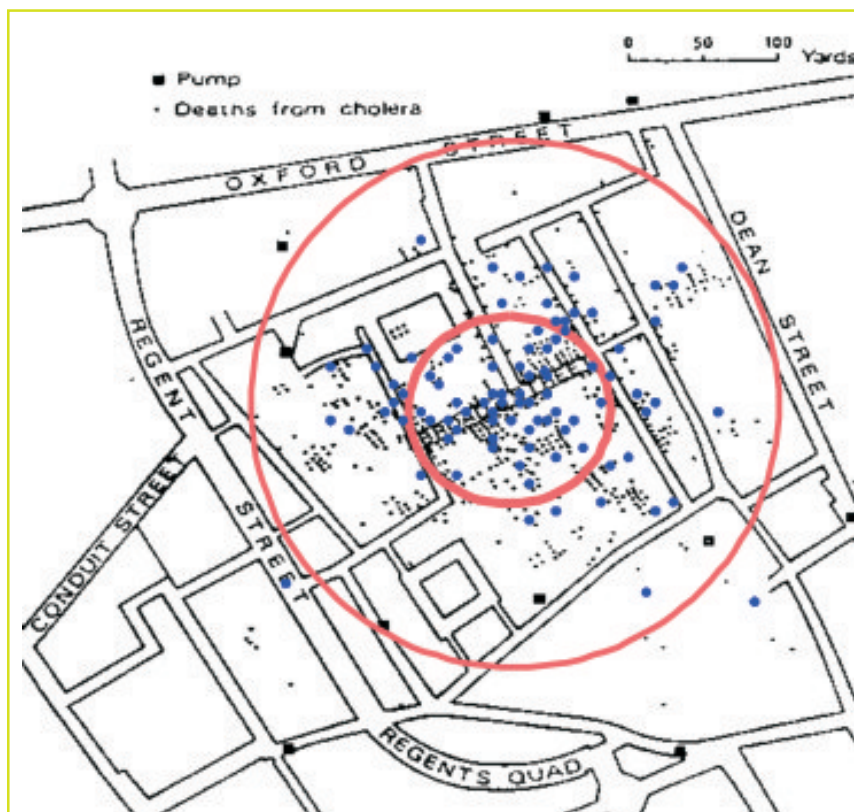
Uma das formas mais tradicionais de se elaborar mapas de interesse para a Saúde Pública é a representação de eventos de saúde em forma de pontos. O uso de pontos para localização de eventos, no momento da construção da base de dados geográfica e durante o procedimento de análise, minimiza problemas associados com as barreiras impostas pela delimitação de áreas administrativas artificiais.

A aplicação de análise de padrão de pontos, assim como dos métodos baseados em SIG, oferecem uma grande flexibilidade para delinear a distribuição espacial dos eventos comparado com as abordagens tradicionais da epidemiologia, que requerem delimitações de áreas usualmente arbitrárias. A localização pontual de eventos é fundamental para possibilitar a avaliação de seu padrão espacial sem a necessidade de unidade de agregação de área predeterminada.

O exemplo mais tradicional desse tipo de abordagem é o mapa da cólera em Londres feito por John Snow. Recordando as hipóteses dessa investigação, era preciso mostrar que as mortes por cólera se concentravam em torno do poço de água da Broad Street (Figura 2.1). O mapa é bastante evidente e foi fácil mostrar essa associação entre água e cólera. Mas em diversos outros casos do nosso dia-a-dia de Vigilância em Saúde, o padrão espacial dos pontos não é tão claro e precisamos usar técnicas de análise espacial para demonstrar e testar os padrões.

Durante a epidemia de cólera na Inglaterra de 1850 a 1860, John Snow produziu diversos mapas para avaliar sua hipótese sobre a transmissão da cólera (Snow, 1999). Um dos mapas mais conhecidos mostra poços de abastecimento de água e as residências de pessoas que morreram por cólera num bairro de Londres. O mapa mostra uma concentração de pontos que representam mortes por cólera, demonstrando um “padrão de distribuição espacial”. À medida que se afasta do centro do mapa, diminui a concentração desses pontos. No centro do mapa existe uma maior densidade de pontos que na periferia. Esse padrão pode indicar a presença de uma fonte comum de exposição.

FIGURA 2.1 – Localização de poços de água e mortes por cólera em Londres 1854 (Baseado em Snow, 1999).



Os eventos em saúde são ocorrências que têm um interesse especial para a Vigilância em Saúde e que, em geral, têm uma dimensão espacial, isto é, podem ser representados em mapas como pontos. Os acidentes de trânsito, casos de sarampo, internações por algum agravo que pode ser usado como evento sentinela, todos estes eventos podem ser representados nos mapas como pontos. Se esses eventos estiverem concentrados no espaço é bem provável que exista uma determinação espacial deste evento e que encontrando os locais com maior concentração, temos uma pista para identificar fatores ambientais, sociais ou ligados à própria assistência à saúde que podem ser objeto de ação da Vigilância em Saúde.

Devemos lembrar que esses eventos nem sempre precisam ser negativos. Poderíamos dispor em um mapa os pontos que representam o endereço de crianças que foram vacinadas. Onde os pontos estiverem mais concentrados pode indicar áreas onde a cobertura da vacinação está melhor. Se tivermos uma lista de endereços de Organizações Não-governamentais (ONG) que atuam na promoção de saúde e colocarmos estes dados no mapa, podemos obter pontos indicando onde estas ONG desenvolvem atividades. Os locais com maior concentração de pontos podem representar áreas melhor trabalhadas.

No caso do John Snow e de outros em que os pontos representam eventos, podemos utilizar técnicas de **análise exploratória de dados espaciais** (ESDA) para avaliar a distribuição dos pontos no espaço. A análise espacial de eventos deve ser realizada levando-se em conta os três grupos de métodos: (a) **visualização**; (b) **exploração**; e, (c) **modelagem** (Bailey e Gatrell, 1995).

2.2 – Análise Espacial de Eventos

- **Visualização de Dados Espaciais** •

O principal objetivo da visualização de dados espaciais é verificar a demonstração gráfica de dados para avaliar padrões específicos de distribuição. O mapa é o modo classicamente usado para visualizar esses dados. Com o advento de ferramentas de SIG, essa abordagem se transformou em algo relativamente rápido, simples e interativo. Dados pontuais, por exemplo, podem ser apresentados em mapas de pontos, da mesma forma como mapas temáticos (coropléticos) podem ser utilizados para a visualização de dados de área. O mapa de John Snow é um bom exemplo de um instrumento de visualização de dados espaciais. Naquele caso não foi utilizada nenhuma técnica de estatística.

- **Análise Exploratória de Dados Espaciais** •

Por sua vez a aplicação de métodos exploratórios de dados espaciais implica a busca de uma descrição dos eventos apresentados, tentando gerar hipóteses ou mesmo modelos apropriados. Pode-se nesse caso usar ferramentas do SIG para medir distâncias e calcular densidades para permitir descrever os padrões espaciais de forma mais clara. O resultado da análise exploratória pode ser um mapa ou mesmo alguns gráficos mais convencionais. Na maioria das vezes, os métodos usados enfatizam a visualização gráfica possibilitando a identificação de padrões, inter-relações, ou mesmo observações atípicas.

- **Modelagem de Dados Espaciais** •

A modelagem de dados espaciais, por outro lado, utiliza métodos que especificam modelos estatísticos e tentam estimar parâmetros. Preocupa-se com os fenômenos estocásticos, ou seja, aqueles que são sujeitos à incerteza, ou influenciados pelas leis de probabilidade. Desse modo, são usados

métodos de análise que consideram a probabilidade de ocorrência de um evento em relação a um conjunto de outros eventos localizados no espaço, ou seja, uma nuvem de pontos.

A diferenciação entre visualizar, explorar e modelar dados é um ponto importante, mas suas delimitações não estão bem definidas, uma vez que existe uma inter-relação entre estes métodos, os quais estão ligados num processo interativo. Essa diferenciação depende basicamente do grau de manipulação dos dados que o método aplicado envolve. Essas abordagens são normalmente acompanhadas de interpretação intuitiva, e têm como objetivo determinar uma impressão inicial da área de estudo e de qualquer padrão óbvio presente. No entanto, como ressaltam Bailey e Gatrell (1995), deve-se estar atento para o fato de que a interpretação intuitiva, do que possa ser um padrão aleatório, pode ser ambígua. A tendência atual é mesclar métodos de visualização e exploratórios devido à semelhança entre eles. Na prática a mescla dos dois procedimentos tem sido muito utilizada, e a denominação “**visualização-exploratória**” tem sido empregada.

Nesse contexto, utiliza-se uma integração dos procedimentos de visualização, ou seja, das ferramentas usadas para mapear dados espacialmente georreferenciados com os procedimentos exploratórios, que analisam os dados na tentativa de detectar padrões, identificar aglomerados, etc. Dependendo do fenômeno representado, seja ele, padrão de pontos, superfície contínua ou dados de área, deve-se utilizar métodos de análise apropriados para cada situação.

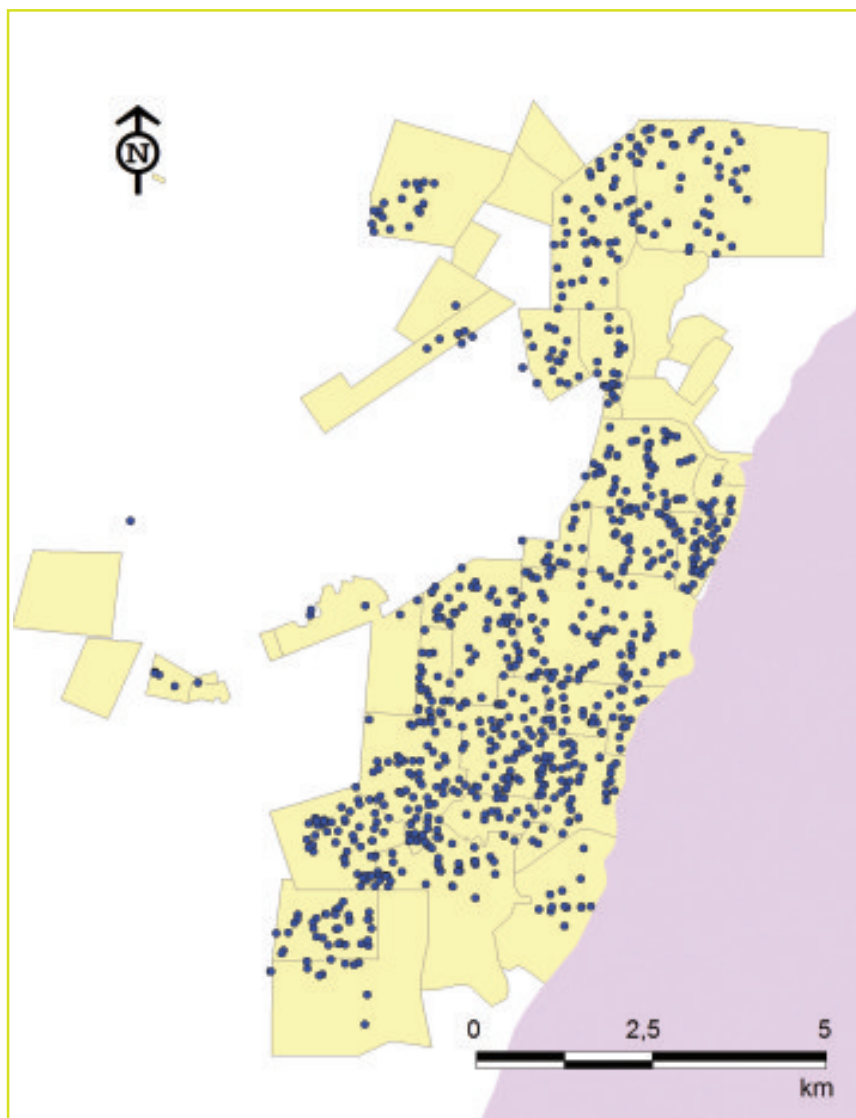
2.3 – Análise de Padrões de Dados com Representação por Pontos

De acordo com Bailey e Gatrell (1995), o **padrão de pontos** é considerado o modo mais simples de representar dados espaciais. É definido como uma base de dados contendo uma série de localização de pontos, numa determinada região de estudo, onde ocorreu o evento de interesse. Em sua apresentação mais simples, esses dados espaciais contêm apenas as coordenadas dos eventos.

Esse é o caso dos mapas que mostram pontos como eventos de saúde, como o de John Snow e de outros como a investigação de hepatite A no Município de Macapá (AP), conforme observado na figura 2.2. Esses dados foram obtidos com base na notificação de casos, usando-se o Sistema

Nacional de Agravos de Notificação (SINAN). A localização de casos foi feita através do registro de coordenadas geográficas no local de residência, utilizando-se um GPS (Neri, 2004).

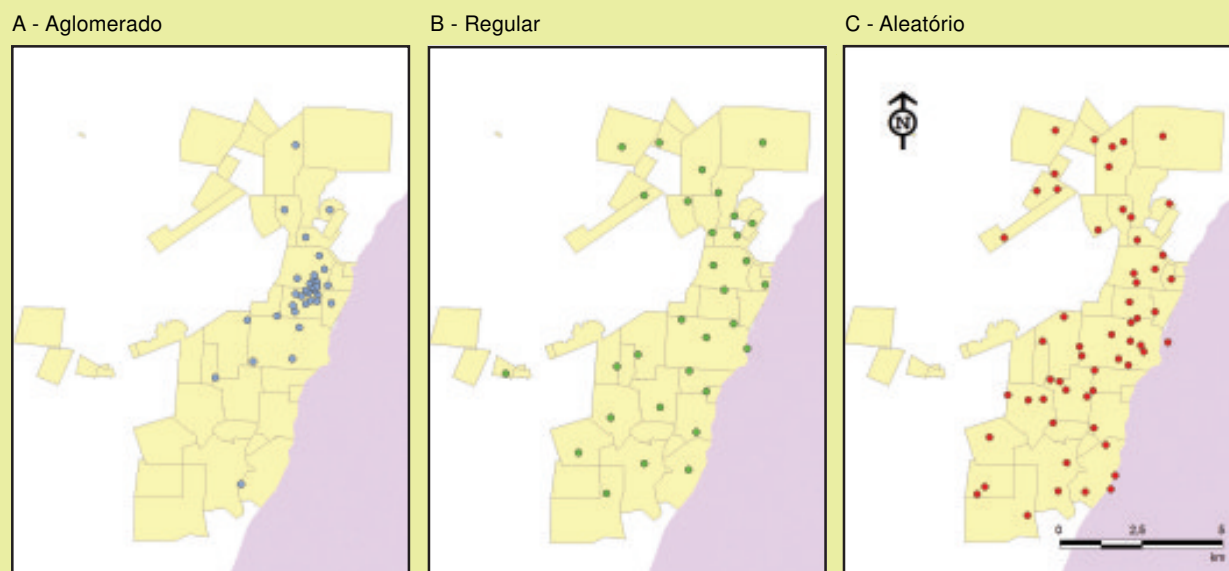
FIGURA 2.2 – Distribuição de casos de hepatite A no Município de Macapá, de 1999 a 2003 (Neri, 2004).



Esse mapa mostra a localização de 933 pontos que representam o local de residência de casos de hepatite A, em Macapá, nos anos de 2000 e 2001. Assim como no caso da cólera, a hepatite A é de transmissão fecal-oral e as áreas com maior incidência da doença podem ter problemas com a qualidade da água consumida pela população ou com o destino de esgotos domésticos. Interessa para a Vigilância em Saúde, portanto, apontar locais em que existe maior intensidade de transmissão para agir sobre estas áreas. Não é uma tarefa fácil nesse caso apontar áreas de maior incidência.

A pergunta que se deve responder neste caso é se existe um padrão espacial de distribuição de pontos. Pensando em um teste estatístico, a pergunta seria reformulada para “**o padrão da distribuição de pontos é diferente do aleatório?**”. Devemos então diferenciar os tipos de distribuição que são esperados quando se trabalha com eventos de saúde: aglomerado, regular e aleatório (Figura 2.3).

FIGURA 2.3 – Tipologia de distribuição padrão de pontos.



Se os riscos estão concentrados no espaço, espera-se um padrão de pontos que mostre uma aglomeração (Figura 2.3 A). Esse **padrão** também é chamado de **agrupado** ou **clusterizado**, usando-se uma expressão do inglês. Esse padrão é diferente do regular (Figura 2.3 B) e do aleatório (Figura 2.3 C).

No **padrão regular** (Figura 2.3 B) existe uma distância média entre os pontos que tende a ser constante. Os pontos estão espalhados em todo o espaço, mas mantendo uma regularidade entre eles. É como se a distribuição de pontos tivesse sido planejada para que um ponto não estivesse muito longe do outro. Na prática de saúde pública este padrão é pouco comum. Um dos casos em que esse padrão pode ocorrer é a localização de unidades básicas de saúde dentro de uma cidade. Se um dos critérios para a localização desses serviços for, por exemplo, que os usuários devem encontrar uma unidade básica a menos de 1 km de sua casa, há uma tendência de se espalhar homogeneamente essas unidades no espaço. O mesmo pode acontecer com amostras ambientais ou armadilhas para o monitoramento de infestação por mosquitos (ovitrapas), que são distribuídas propositalmente no espaço de forma regular.

Já o **padrão aleatório** (Figura 2.3 C) não tem nenhum modelo de distribuição. Ele ocorre por um completo acaso. Nessa última situação, existe uma mescla de distribuição aglomerada e uniforme (Ward e Carpenter, 2000).

Imagine uma doença que não é fortemente determinada por nenhum fator social ou ambiental, como a diabetes. Como os fatores de risco da doença (por exemplo, a predisposição genética e hábitos alimentares) estão em geral espalhados no espaço, a distribuição espacial dos casos da doença vai tender a um padrão aleatório. Mesmo nesses casos, como a diabetes, é uma tarefa da Vigilância em Saúde procurar associações e padrões que possam servir para identificar populações mais vulneráveis. Se existe uma fonte comum de exposição a riscos, o padrão de distribuição de pontos tenderá a ser aglomerado.

As técnicas de análise espacial permitem identificar áreas com concentração aparentemente alta de eventos e possibilitam avaliar se o aparente aglomerado de casos ocorreu ou não ao acaso. De acordo com Besag e Newell (1991) e Lawson e Kulldorf (1999), os testes estatísticos utilizados em análise espacial podem ser agrupados em duas grandes categorias:

- a) **testes globais** e
- b) **testes focais**.

Os **testes globais** compreendem a maioria dos testes para identificação de aglomerados espaciais de doenças, em que todos os casos são processados. Esses procedimentos testam a presença de aglomerados sem ter a habilidade de identificar a localização dos mesmos (trabalham com o conjunto total dos pontos mapeados). Dessa forma, testam a hipótese de que não existe aglomerado na região contra a hipótese alternativa de que existe conglomerado sem especificar sua localização.

Os **testes focais**, ou locais, para a identificação de aglomerados são usados quando se tem uma fonte primária pré-definida, ou seja, verificam a existência de aglomerados em regiões definidas *a priori*, antes da observação do evento. Testam a presença de aglomerados em áreas específicas, como por exemplo, em torno de um foco suspeito de risco aumentado para uma doença. São especialmente úteis quando existe forte suspeita de um agente etiológico.

Os **testes de detecção de aglomerados** podem ser considerados uma terceira categoria na qual é possível identificar o local da ocorrência do aglomerado e, além disso, testar sua significância estatística. Testes globais e testes de detecção de aglomerados complementam-se e têm uma boa aplicação no estudo espacial de aglomerados em diversos cenários epidemiológicos.

Existem diversas técnicas disponíveis de análise exploratória espacial para detectar e mapear “áreas quentes” ou aglomerados. Dentre elas, destacam-se a **técnica de Kernel** e a **estatística espacial scan**, que constituem ferramentas amplamente utilizadas em anos recentes, na área de saúde e ambiente.

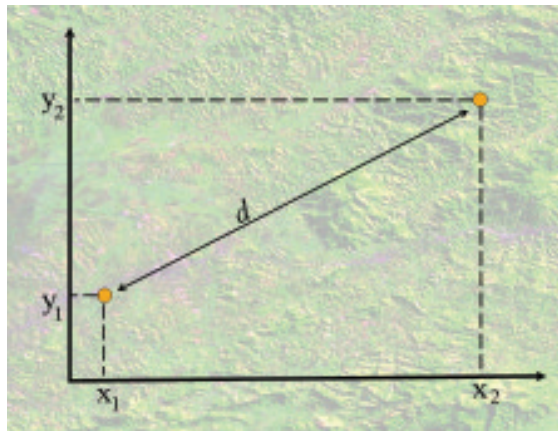
2.4 – Análises Baseadas na Distância entre Pontos

Uma maneira de avaliar os padrões espaciais de pontos e detectar a existência de aglomerados é medir a média da distância entre os pontos. No padrão aglomerado a distância média entre pontos é pequena, já nos padrões aleatório e regular a distância entre pontos é grande. As distâncias entre pontos são medidas facilmente em um SIG, já que o sistema guarda as coordenadas de cada ponto.

A chamada **distância euclidiana** é calculada pela equação:

$$d = \sqrt{\{(x_1 - x_2)^2 + (y_1 - y_2)^2\}}$$

Onde o ponto 1 tem coordenadas x_1 e y_1 , e o ponto 2 tem coordenadas x_2 e y_2 .



A tabela abaixo mostra a estatística de distância entre pontos dos casos da figura 2.3.

	DISTÂNCIA MÉDIA (em metros)	DESVIO PADRÃO DA DISTÂNCIA (em metros)
Aglomerado	358	688
Regular	1094	321
Aleatório	890	643

Pode-se observar que o padrão aglomerado tem a menor distância média entre pontos, o que indica uma concentração de pontos no espaço estudado, mas um grande desvio padrão. O padrão regular, como o próprio nome indica, tem uma grande distância média entre os pontos, mas um pequeno desvio, indicando a homogeneidade da distribuição. Já o padrão aleatório tem uma grande distância média entre pontos e grande desvio padrão.

Essas estatísticas de distância entre pontos podem servir para uma série de técnicas usadas para a detecção de aglomerados de pontos. A análise de aglomerados (*cluster*) é um termo genérico usado em um grande número de técnicas que tentam determinar se uma base de dados é composta por grupos distintos (Everitt, 1974). De modo geral a análise de aglomerados é um método de classificação que utiliza a característica de certos objetos com o intuito de agrupá-los.

• Análise de Aglomerados Hierárquicos •

Nas técnicas de aglomerados hierárquicos, por exemplo, a detecção inicia-se com os eventos separados e procede-se a união das observações mais semelhantes em grupos, com base em critérios pré-definidos.

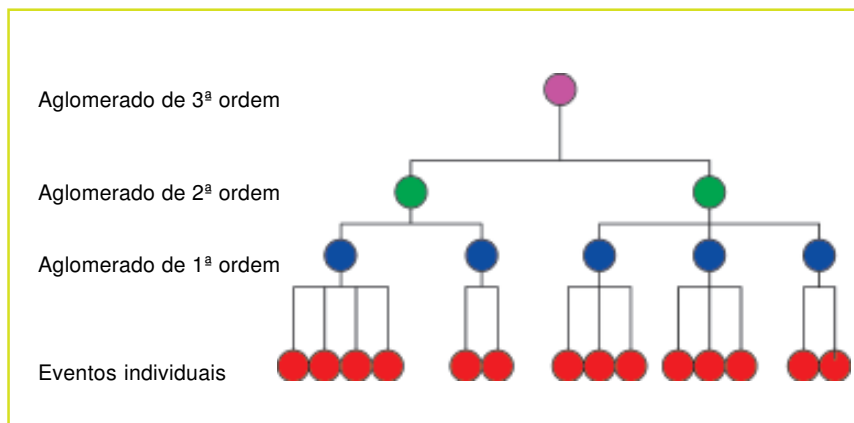
No aglomerado hierárquico (*hierarchical cluster*), a distância entre vizinhos mais próximos freqüentemente é usada como medida de dissimilaridade. Existem diversos tipos de medidas de distâncias métricas disponíveis, mas a distância euclidiana é a mais comumente usada por ser bastante fácil e intuitiva. Na análise hierárquica de aglomerados, a menor distância entre dois objetos é primeiramente identificada e estes objetos são agrupados em um novo aglomerado. As distâncias entre esses aglomerados são calculadas, e estes por sua vez, podem ser agrupados num próximo nível. Segundo Bailey e Gatrell (1995), a análise de hierarquia de aglomerados é um método exploratório de estatística multivariada usado para identificar agrupamento de objetos que pode identificar a localização espacial destes eventos.

Dentre os métodos hierárquicos mais utilizados está o de vizinhos mais próximos. A rotina de **aglomerados hierárquicos de vizinhos mais próximos** (Nearest Neighbor Hierarchical Clustering – NNHC), implementada pelo programa Crimestat 2.0 (Levine, 2002), é uma rotina de agrupamento de distância que une pontos com base em sua proximidade espacial. Ou seja, agrupa casos incidentes que estão espacialmente próximos.

De forma geral, a rotina identifica aglomerados **vizinhos de primeira ordem** os quais representam os pontos que estão mais próximos do que a distância limiar proposta, e que tenham pelo menos o número mínimo de pontos determinados. Na seqüência da análise, os aglomerados de primeira ordem são tratados como pontos individuais que podem ser agrupados em **aglomerados de segunda ordem** e assim sucessivamente até que se atinja um único aglomerado ou que as condições estabelecidas não sejam mais atendidas. Esses aglomerados são identificados por meio de elipses que delimitam um grupo de pontos contidos nos aglomerados. O processo de unificação dos aglomerados identificados pela técnica de aglomerados hierárquicos de vizinhos mais próximos pode ser graficamente representado, de forma espacial, através de mapas, ou ainda de forma não espacial por

meio de árvore hierárquica, também chamada de dendograma, como na figura a seguir.

FIGURA 2.4 – Dendograma de representação gráfica dos vários passos do processo hierárquico de aglomerado. A utilização dessa forma de apresentação evidencia a ligação entre os eventos e os aglomerados identificados em seus diversos níveis.



Com base no grau de similaridade destes eventos ou nas suas características compartilhadas eles vão sendo agrupados, inicialmente em um primeiro aglomerado, identificado como de primeira ordem. A partir desse ponto cada aglomerado passa a ser identificado como se fosse um evento isolado, e de acordo com as premissas preestabelecidas vão sendo agrupados até alcançar o último aglomerado possível ou até que as condições estabelecidas não sejam mais atendidas.

Ao se conduzir uma análise hierárquica de vizinhos próximos devem ser definidos os seguintes parâmetros:

- a **distância limiar**, que é a probabilidade de selecionar um par ao acaso;
- o **número mínimo de pontos** necessários para constituir um aglomerado e
- o **tamanho da elipse** que vai ser utilizada para identificar o local de ocorrência desses aglomerados.

A– Distância limiar

Se apenas pares de eventos estivessem sendo agrupados, então a distância limiar seria o parâmetro de maior importância a ser considerado. Utilizando-se o valor padrão definido pelo programa ($p < 0,5$), aproximadamente metade dos pares seria selecionada ao acaso se os dados obedecessem a uma distribuição randômica. Contudo, uma vez que, além da distância limiar, utiliza-se o número mínimo de eventos na construção da rotina de análise, a chance de encontrar aglomerados com este número mínimo de pontos é menor. A escolha dos valores dessa distância limiar vai depender do objetivo do estudo. Para intervenções centradas numa localização particular, o uso de uma distância limiar pequena pode ser apropriado; por outro lado, a identificação de áreas mais amplas, em algumas situações, pode ser melhor que a seleção de microambientes.

B- Número mínimo de pontos

O número de eventos mínimos para definir um aglomerado é de suma importância na identificação dos grupos, uma vez que o programa desconsidera a localização do ponto inicial se o aglomerado não congrega este número mínimo de pontos. Em geral o critério do número mínimo de pontos é mais crítico do que a distância limiar, embora a última também possa influenciar os resultados. Levine (2002) sugere que a investigação pode ser realizada com diferentes critérios até que se encontre o critério que melhor se adapte à situação estudada.

C- Tamanho da elipse

Por sua vez, a definição do tamanho da elipse a ser utilizada na visualização dos aglomerados identificados também é influenciada pelo objetivo do estudo. A rotina presente no aplicativo disponibiliza a escolha de valores que vão de um a dois desvios padrões. A utilização de elipses construídas com base em um desvio padrão pode não ser o mais adequado para a visualização do evento em escala regional. Mais uma vez, o investigador deve balancear a necessidade de identificar, de forma acurada, a localização do aglomerado *versus* a necessidade de facilitar a adequada visualização da área estudada.

O aglomerado hierárquico de vizinhos mais próximos, além de ser um método exploratório de análise espacial, pode ser combinado com uma **simulação de Monte Carlo**. Esse método simula uma série de dados de acordo com um modelo prévio de modo a testar a existência de aglomerados que sejam estatisticamente significativos. Nessa situação, pode ser então classificado como um método de modelagem de padrão espacial de pontos, uma vez que utiliza técnicas para testar hipóteses comparando a distribuição observada a um modelo particular, o de aleatoriedade espacial completa (AEC). O modelo padrão para a AEC é de que os eventos seguem um processo homogêneo de Poisson na área estudada, ou seja, os eventos são independentes e uniformemente distribuídos na região. Isso significa dizer que: (a) qualquer evento tem a mesma probabilidade de ocorrer em qualquer posição da região sob estudo, e (b) a posição de qualquer evento é independente da posição de outro evento, ou seja, os eventos não interagem uns com os outros.

A AEC é o primeiro passo na testagem da hipótese de que as observações se distribuem de forma regular, aglomerada ou aleatória. De forma conceitual, a distância limiar define a probabilidade de dois pontos serem agrupados ao acaso. O teste de significância estatística avalia o intervalo de confiança em torno da distância dos vizinhos de primeira ordem. Se a probabilidade é 5%, então aproximadamente 5% de todos os pares de casos seriam identificados numa distribuição aleatória.

No entanto, a rotina hierárquica de vizinhos próximos leva em conta outros parâmetros, e a probabilidade de distribuição levando em conta estes parâmetros não é conhecida. Nesse sentido, é necessária uma simulação de Monte Carlo considerando apenas os aglomerados hierárquicos de primeira ordem.

Vantagens e desvantagens da técnica de aglomerados hierárquicos de vizinhos mais próximos

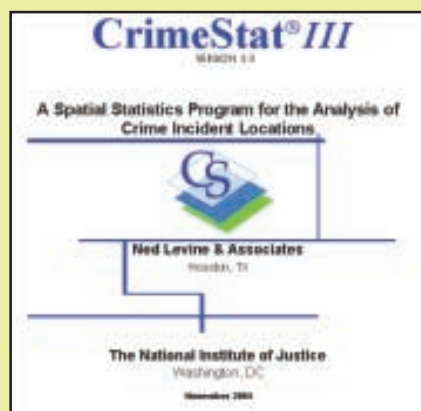
A técnica de aglomerados hierárquicos de vizinhos mais próximos determina se existem aglomerados (ex: “áreas quentes”) na distribuição dos eventos utilizando-se um critério quantitativo. A identificação de aglomerados é feita pelo desenho de elipses que marcam locais com maior concentração de pontos que o esperado segundo um modelo de aleatoriedade.

Dentre as **vantagens** da técnica ressalta-se:

- a) habilidade para identificar áreas onde exista concentração de eventos;
- b) pode ser aplicada para qualquer base de dados, independente da sua abrangência (ex: município ou distrito urbano); e
- c) a ligação entre os diversos aglomerados encontrados pode ser visualizada por meio de grupos segundo níveis de hierárquicos, ou seja, considerando diferentes escalas de aglomerados de eventos que podem ser identificadas por esta técnica.

A principal **desvantagem** é que essa técnica usa dados absolutos e pode produzir uma falsa impressão da existência de grupamentos, uma vez que não considera dados populacionais. Esses aglomerados podem ser conseqüências da própria tendência de concentração de população em determinados locais com mais recursos. Além disso, o tamanho da área estudada é dependente do tamanho da amostra, uma vez que o intervalo de confiança em torno da distância média aleatória é utilizado como um dos critérios. Uma outra desvantagem é a existência de um certo grau de arbitrariedade, uma vez que o número mínimo de eventos a compor um aglomerado é definido pelo investigador, da mesma forma como a definição da probabilidade a ser empregada.

Nas pesquisas e práticas de Vigilância em Saúde a utilização dessa técnica tem sido usada crescentemente. A maioria dos exemplos disponíveis de sua aplicação encontra-se na identificação de áreas de maior ocorrência de crimes (Levine, 2002).



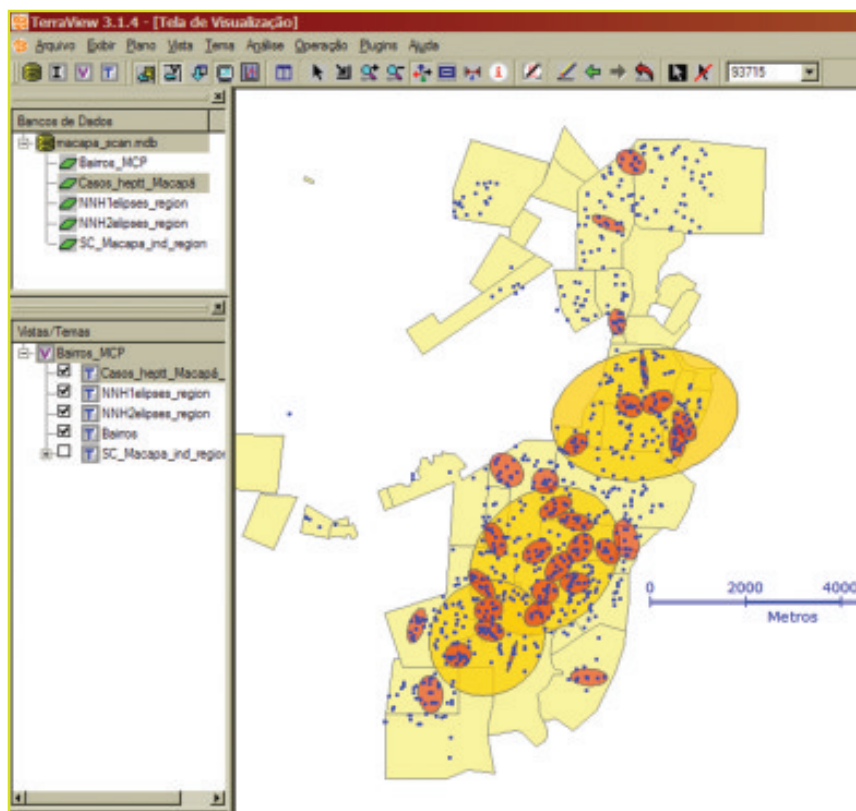
Fonte: <http://www.icpsr.umich.edu/CRIMESTAT/>

O **Crimestat** é um aplicativo gratuito desenvolvido pelo professor Ned Levine para o Instituto de Justiça dos Estados Unidos, para avaliar a incidência de crimes em áreas urbanas daquele país. Esse programa pode também ser usado para identificar padrões de distribuição de eventos de saúde. O programa possui diversas rotinas de análise espacial de pontos como a função K, estimadores kernel, análise espaço-temporal por teste de Knox, e outros.

Como entrada de dados, o programa aceita arquivos em *dbf* que tenham as coordenadas (em qualquer sistema de projeção) e exporta os resultados como um texto ou como objetos cartográficos em formato *shp*. Por exemplo, as elipses que marcam os aglomerados identificados pelo aplicativo são criadas em formato *shp* e facilmente importadas para um programa de SIG. Como esses formatos são os mesmos empregados pelo Terraview, o Crimestat pode ser usado como aplicativo auxiliar para análise e teste de hipóteses que não podem se feitas em outros programas.

Retomando-se o exemplo da distribuição de casos de hepatite A em Macapá, temos a identificação de alguns aglomerados de casos: um grande aglomerado de segunda ordem na área sul da cidade e dez aglomerados de primeira ordem espalhados pela cidade, principalmente nos bairros do sul (Figura 2.5). Esses aglomerados são marcados por uma elipse que é produzida pelo próprio aplicativo Crimestat em formato *shp* (*shape*). Este formato pode ser facilmente importado pelo programa TerraView.

FIGURA 2.5 – Casos de hepatite A e aglomerados de primeira ordem (em forma de pequenas elipses vermelhas) e de segunda ordem (em forma de uma grande elipse laranja).



2.5 – Análise da Densidade de Pontos

Dentre os procedimentos para estimar densidade de eventos, a estimativa Kernel é a mais conhecida e a mais utilizada, desempenhando um papel importante no contexto epidemiológico (Gatrell *et al.*, 1996) para identificar a concentração de casos.

A **estimativa Kernel** é uma técnica de interpolação exploratória que gera uma superfície de densidade para a identificação visual de “**áreas quentes**”. Entende-se a ocorrência de uma área quente como uma concentração de eventos que indica de alguma forma a aglomeração em uma distribuição espacial.

É uma técnica estatística, de interpolação, não paramétrica, em que uma distribuição de pontos ou eventos é transformada numa “superfície contínua de risco” para a sua ocorrência. Esse procedimento permite filtrar a variabilidade de um conjunto de dados, sem, no entanto, alterar de forma essencial as suas características locais (Bailey e Gatrell, 1995). O estimador Kernel não é um método de detecção de aglomerados por si, mas, um método para explorar e mostrar o padrão de pontos de dados em saúde, e é útil a partir do momento em que gera uma superfície contínua a partir de dados pontuais (Cromley e McLafferty, 2002).

O estimador Kernel produz uma superfície contínua, com densidades calculadas em todas as localizações, desta forma, o investigador não precisa definir de forma arbitrária onde “cortar” a “área quente”, isto é, uma densidade a partir da qual se considera que os pontos formam um aglomerado (Levine, 2002). Contudo, é importante ressaltar que a interpretação dos resultados obtidos mediante a análise Kernel é subjetiva e depende do conhecimento prévio da área de estudo. A técnica apresenta, como uma das maiores vantagens, a rápida visualização de áreas que merecem atenção, além de não ser afetada por divisões político-administrativas. Portanto o estimador de intensidade é uma boa alternativa para se avaliar o comportamento dos padrões de pontos em uma determinada área de estudo, sendo considerado muito útil para fornecer uma visão geral da distribuição de primeira ordem dos eventos (Carvalho e Câmara, 2002).

Originalmente a estimativa Kernel foi desenvolvida para se obter uma estimativa suavizada de uma probabilidade univariada ou multivariada com base na amostra de uma observação, ou seja, um histograma suavizado. A estimativa de intensidade de uma distribuição de pontos é muito semelhante à estimativa de probabilidade de densidade bivariada (*bivariate probability density*).

Para a aplicação da estimativa Kernel, é necessária a definição de dois parâmetros básicos, que podem ser visualizados na figura 2.6:

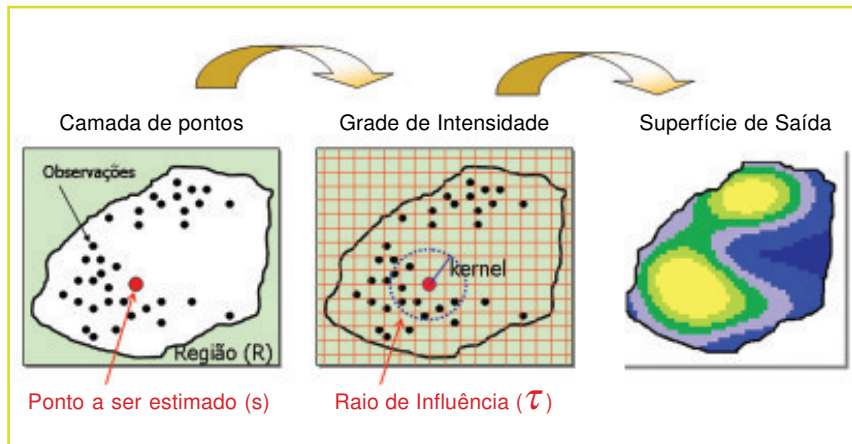
a) **raio de influência** (τ) que define a vizinhança do ponto a ser interpolado e controla o alisamento da superfície gerada. É o raio de um disco, centrado em s , que é uma localização na região R , no qual pontos s_i vão contribuir para a estimativa da função de intensidade;

b) uma **função de estimação k (Kernel)** com propriedades de suavização do fenômeno. Os Kernel normal ou quártico são os mais comumente utilizados. Se s representa uma localização em R e s_1, \dots, s_n são as localizações das n observações, então a intensidade $\lambda(s)$, em s é estimada por:

$$\lambda_{\tau}(s) = \sum_{i=1}^n \frac{1}{\tau^2} K\left(\frac{(s - s_i)}{\tau}\right)$$

Onde k é uma função de densidade bivariada escolhida, ou seja, a função de estimação ou de alisamento, conhecida como Kernel. O parâmetro τ é conhecido como raio de influência ou largura da banda e determina o grau de suavização da superfície de saída. A função bidimensional é ajustada sobre os eventos considerados compondo uma superfície cujo valor será proporcional à intensidade dos eventos por unidade de área.

FIGURA 2.6 – Passos para o cálculo de densidade de pontos segundo a técnica de Kernel (Câmara *et al.*, 2002).



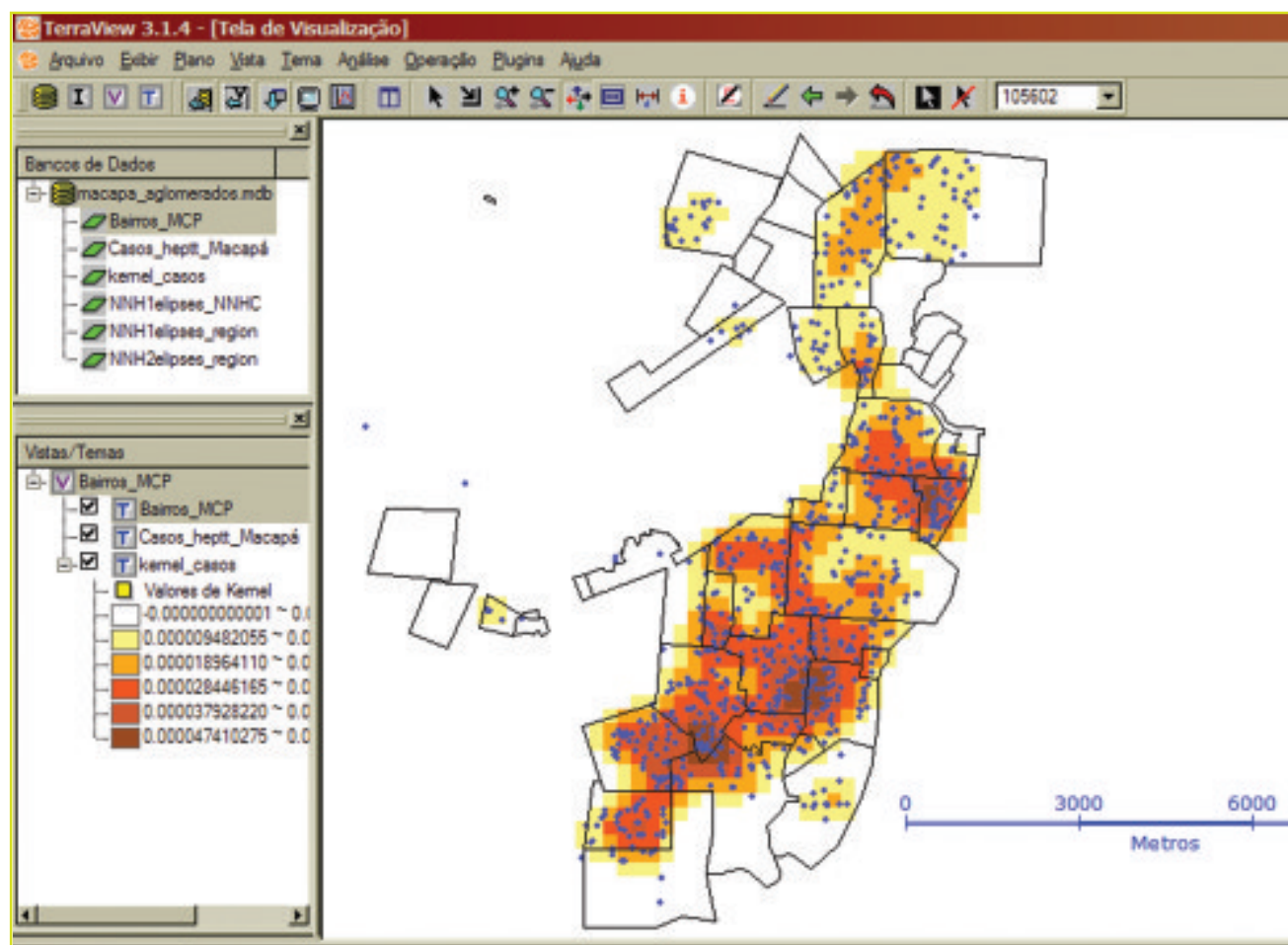
A escolha da função Kernel (se quadrática, normal, triangular ou uniforme) a ser utilizada não é um ponto crítico, contudo, a escolha do raio de influência é crucial, produzindo significantes alterações da estimativa final.

Essa função (Figura 2.6) realiza a contagem dos eventos dentro de uma região de influência, ponderando-os pela distância, de cada um, a uma localização de interesse (Câmara *et al.*, 2002). Do ponto de vista visual, quando se aplica o estimador de intensidade sobre uma grade de localizações na área R , pode-se pensar numa função tridimensional que “visita” cada ponto s

na grade. A distância para cada evento observado s_i , que cai na área de influência, é medida e contribui para a estimativa de intensidade em s de acordo com sua distância de s .

O resultado da estimativa Kernel para o caso da hepatite A em Macapá é mostrado na figura 2.7, a seguir.

FIGURA 2.7 – Casos de hepatite A e superfície de densidade, segundo a técnica de Kernel em Macapá (Dados de Neri, 2004).



Pode-se observar que existem casos distribuídos em toda a cidade, mas estes casos estão mais concentrados na área central-sul (bairros de Buritizal e Congos) e na central-norte (Perpétuo Socorro). Essa já é uma indicação de riscos que pode ser usada para se atuar sobre determinantes da hepatite A. Como estamos trabalhando em um ambiente de Sistema de Informações Geográficas (SIG), que tem diversas funcionalidades, podemos usar este sistema para identificar o nome dos bairros com maior intensidade de transmissão e as características sociais e demográficas destes bairros.

Grade regular de representação do Kernel

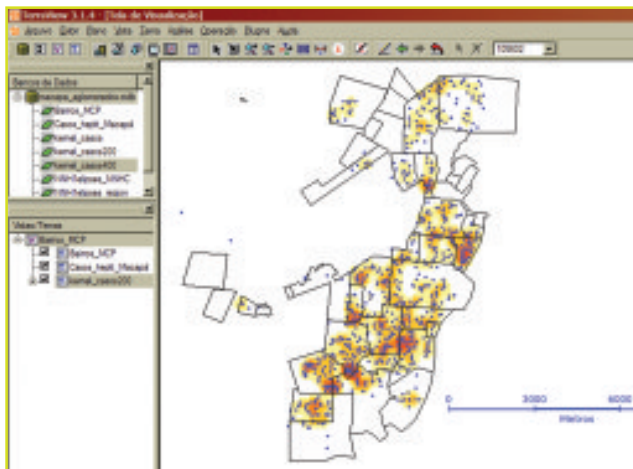
O arquivo gerado pela técnica de kernel é uma grade com **n** colunas por **m** linhas, criando **n x m** células. Quanto maior o número de células, maior a resolução da figura gerada. A definição desse parâmetro afeta a quantidade de memória que vai ser usada para armazenar os resultados. O uso de uma grande quantidade de colunas pode gerar uma sobrecarga de processamento no programa e computador.

Essa grade é uma maneira de representar a superfície contínua com as estimativas de densidade de pontos. Cada uma destas células tem como atributo o valor estimado da densidade de pontos. Observe que o mapa temático gerado tem uma legenda que mostra variações a partir de um número muito pequeno. Os valores representam a quantidade de pontos encontrados em uma unidade de área. No caso do mapa de Macapá, onde se trabalhou com o sistema de coordenadas UTM em unidades de metros, os valores estão em eventos por metro quadrado.

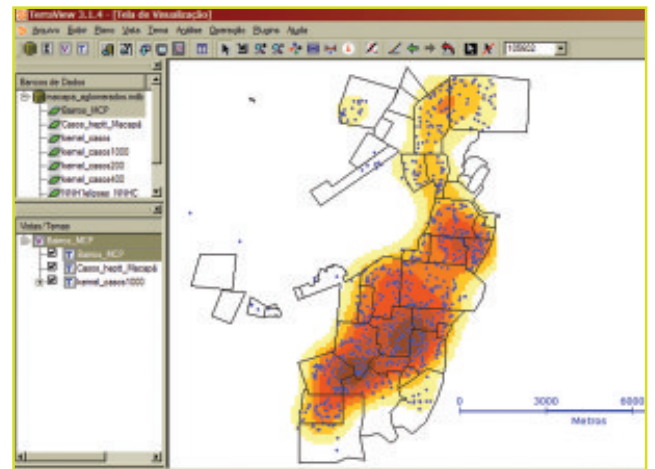
Na figura 2.8 simularam-se dois padrões de suavização variando os valores do raio de influência. Raio de influência mais amplo, como o utilizado na figura 2.8 B, gera uma suavização maior, com superfície mais homogênea, enquanto a suavização produzida por valores pequenos da largura da banda gera um número maior de “áreas quentes” na região de estudo (Figura 2.8 A).

FIGURA 2.8 – Simulação de resultados de Kernel:

A- alisamento produzido por valor pequeno de raio de influência ($\mathcal{T} = 200$ m);



B- alisamento produzido pelo emprego de valor maior do raio de influência ($\mathcal{T} = 1000$ m).



A escolha do padrão de suavização é de importância fundamental na estimativa da densidade dos eventos e depende do objetivo do estudo e do tipo de evento estudado. Para identificar áreas específicas e de menor abrangência para atuação, valores menores de raio de influência podem ser mais indicados (ex: doenças transmitidas por insetos), contudo esta abordagem pode gerar áreas múltiplas e pulverizadas de atuação. Se o objetivo é identificar áreas mais abrangentes para otimizar intervenções, a largura da banda mais

ampla torna-se a melhor opção. Se o objetivo do estudo é explorar os dados para a formulação de hipóteses, diferentes raios de influência devem ser empregados (Silverman, 1986). Na prática, para o cálculo do estimador Kernel, o investigador pode experimentar diferentes valores de τ gerando assim variações de intensidade até encontrar o padrão que melhor se adapte à região estudada refletindo a densidade local dos eventos (Bailey e Gatrell, 1995; Cromely e McLafferty, 2002).

2.6 – Avaliação de Aglomerados considerando a Distribuição da População

Os métodos descritos anteriormente, baseados na distância entre pontos e na densidade de pontos no espaço, podem ser utilizados para avaliar uma aglomeração destes pontos, indicando áreas onde estes pontos se concentram. Nesses casos, o que se busca é responder à pergunta: **os pontos têm uma distribuição espacialmente condicionada?** Ou, a distribuição dos pontos é diferente do que se esperaria de um padrão aleatório?

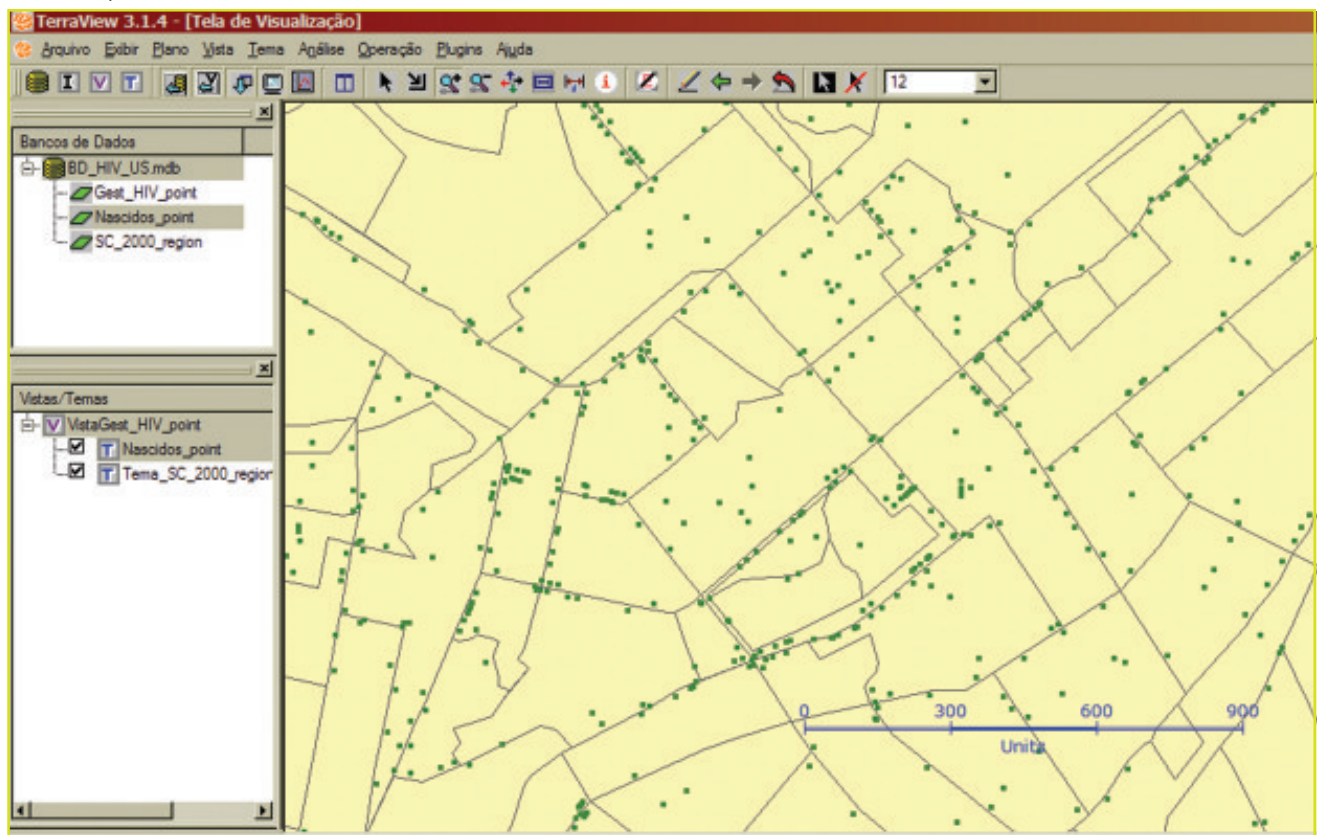
O grande problema das técnicas de detecção de aglomerados é que os pontos de eventos de saúde tendem a ter uma distribuição espacial semelhante à distribuição da população. Todos os eventos de saúde – o nascimento, a infecção, a vacinação, o adoecimento, a internação, a morte – se manifestam em pessoas. Essas pessoas não estão distribuídas aleatoriamente no espaço. Essa distribuição é determinada por fatores históricos e socioeconômicos. Por isso, ao se trabalhar com registros de saúde para avaliar riscos, deve-se estimar a probabilidade de um evento ocorrer, ponderada pela distribuição da população.

A pergunta nesse caso passa a ser: **a distribuição espacial dos pontos é diferente da distribuição da população?** E a melhor forma de responder à pergunta é construir taxas, onde o numerador é o número de eventos de saúde e o denominador a população de referência.

A forma mais usual de se considerar a distribuição da população na avaliação de riscos é a agregação de dados demográficos e de saúde em unidades espaciais discretas, isto é, em áreas pré-concebidas, para posteriormente calcular indicadores epidemiológicos. Essa estratégia apresenta sérias limitações, como a desconsideração das interações entre as unidades espaciais e a instabilidade dos indicadores criados em pequenas áreas (King, 1979).

Se dispusermos de dados em forma de pontos, podemos construir diversas totalizações de conjuntos de eventos dentro de vários tipos de áreas, a partir da agregação destes dados. A figura 2.9 mostra um conjunto de dados originais que podem ser usados para avaliar uma situação em que se quer identificar aglomerados de pontos com taxas de incidência altas. Os eventos de saúde são representados por pontos e os dados de população são atributos dos setores censitários, que são pequenos polígonos.

FIGURA 2.9 – Eventos de saúde (pontos) e setores censitários (polígonos) usados para o cálculo de taxas por meio de técnicas de análise espacial.



O caso mais simples de construção de taxas a partir de pontos é quando se tem uma fonte suspeita de riscos. Essa era exatamente a situação do estudo de John Snow que tentava mostrar que existia um risco maior de morrer por cólera nos moradores em torno do poço de água da *Broad Street*. Também é a situação de muitas investigações que vêm sendo feitas em vigilância ambiental em saúde, em que existe uma suspeita de aumento de casos de doenças, leucemia, por exemplo, em torno de usinas nucleares, depósitos de lixo tóxico ou incineradores.

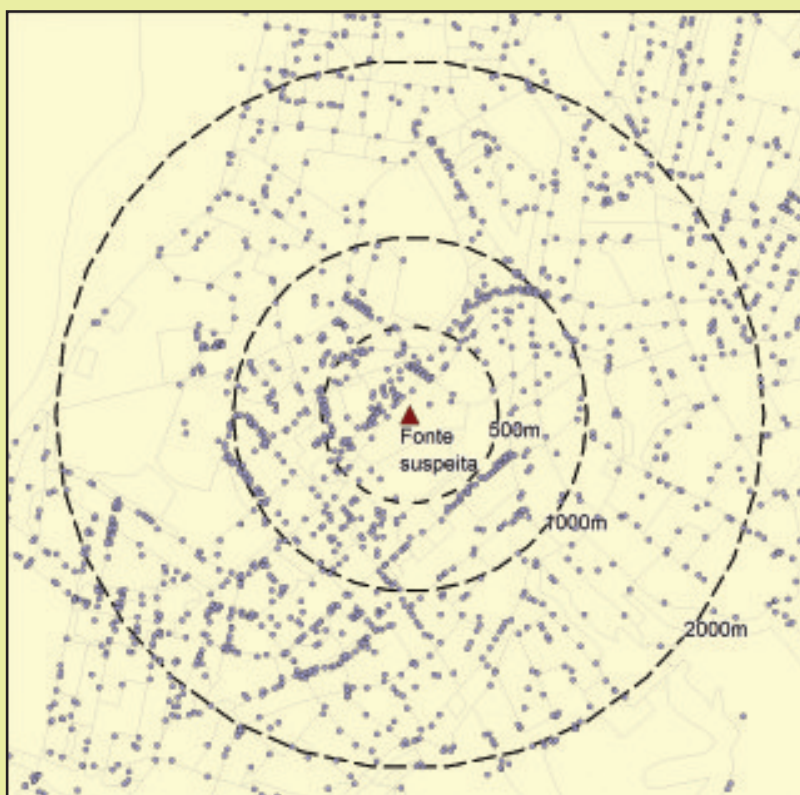
Uma solução analítica para esse problema é apontar a fonte suspeita no mapa, criar uma **área de influência** (*buffer*) em torno desta fonte suspeita, calcular a taxa de incidência do agravo à saúde dentro da área de influência e fora dela. A maneira que se faz isso dentro de um ambiente SIG é bastante

simples. Uma vez delimitada a área de influência, contamos quantos eventos existem dentro da área (pontos dentro do polígono) e qual a população total dentro da área (soma das populações de polígonos pequenos dentro de polígonos maiores).

Área de influência (*buffer*)

Existem várias formas de se construir áreas de influência. A mais simples é um círculo em torno de um ponto. Mas poderíamos também criar vários círculos concêntricos em torno desse ponto, com raios de 500, 1.000, 2.000 e 5.000 metros, por exemplo (Figura 2.10). As áreas de influência podem ser também criadas em torno de linhas e de polígonos.

FIGURA 2.10 – Áreas de influência em torno de uma fonte suspeita de risco.



Ao construir taxas de incidência dentro de cada uma dessas áreas de influência, podemos avaliar se os riscos decaem com o aumento da distância da fonte suspeita. Nem sempre a fonte suspeita é um ponto no mapa. Em algumas situações muito comuns, essa fonte pode ser uma linha. Uma estrada pode ser uma fonte de riscos de contaminação do ar para a população em torno. Também pode representar um risco maior de morte por acidentes de trânsito para as pessoas que moram perto.

No segundo livro desta série mostrou-se como construir áreas de influência usando o Terraview. Dentro dessa área de influência podem-se calcular taxas contando-se o número de eventos e a população total.

Existem técnicas estatísticas que permitem avaliar se as taxas de incidência dentro e fora da área de influência são significativamente diferentes, como se verá a seguir. Mas em diversos outros casos, do cotidiano de Vigilância em Saúde, não existe uma fonte suspeita. Quando se avalia a dengue, por exemplo, não se sabe *a priori* onde estão essas fontes de risco, que podem ser focos do vetor. Esse também é o caso da distribuição da hepatite A. Gostaríamos que as técnicas de análise espacial apontassem onde estão essas fontes de risco, que ainda não conhecemos.

Existe uma técnica, recentemente desenvolvida, que avalia as taxas de incidência por meio de um círculo que caminha no espaço, mudando o raio e a posição do seu centro. Essa técnica faz uma varredura (*scan* em inglês) do espaço estudado, procurando localizar um círculo no local onde haja uma taxa de incidência muito grande, dentro, e muito baixa, fora deste círculo.

A maioria dos métodos estatísticos, para análise de aglomerados de dados pontuais, é descritiva: conseguem detectar a localização do aglomerado sem fazer inferências estatísticas, ou por outro lado, conseguem fazer inferências sem ter a habilidade de detectar a localização do aglomerado. Essa é uma diferença básica da estatística espacial *scan* a qual possui a habilidade de realizar ambos os procedimentos, de tal forma que, uma vez que a hipótese nula é rejeitada consegue-se localizar a área específica do mapa que causou a rejeição (Kulldorff, 1997).

• Detecção de Aglomerados pela Técnica Estatística Scan •

O método leva em conta a distribuição populacional irregular que influencia a distribuição destes pontos sob a hipótese nula. A **estatística espacial *scan*** é definida por uma janela geográfica circular que se move através da área de interesse, ou seja, é uma estatística de varredura onde, o círculo é capaz de incluir diferentes conjuntos de áreas vizinhas. Considera-se uma área como incluída no círculo se o seu centróide está contido na janela. O raio de cobertura da janela pode variar de zero a um valor máximo, de forma que a janela nunca inclua mais que 50% da população total. Com isso, a janela é flexível em tamanho e localização. Para cada localização e tamanho da janela de varredura é verificado se existe uma taxa elevada dentro do círculo quando comparada com a região externa a este círculo.

A técnica cria um conjunto com infinito número de círculos distintos. Cada um deles podendo conter uma série diferente de regiões vizinhas com a possibilidade de conter um aglomerado de eventos. Cada círculo geográfico criado inclui um conjunto de localizações ou áreas administrativas, e é tratado como um aglomerado em potencial.

Para cada janela, o método testa a hipótese nula frente à hipótese alternativa que existe um risco elevado de ocorrência de eventos dentro desta janela em

comparação com o exterior. É comparado o número de casos dentro da janela com o número de casos esperados se os casos estivessem aleatoriamente distribuídos no espaço.

A função de verossimilhança é maximizada para todas as janelas e, aquela com a máxima verossimilhança, é considerada como o aglomerado mais provável. Ou seja, o aglomerado que tem a menor probabilidade de ter ocorrido ao acaso. A razão de verossimilhança para essa janela constitui a estatística máxima da razão de verossimilhança. A sua distribuição sob a hipótese nula é obtida por meio da repetição desta análise num grande número aleatório de replicações do teste geradas sob a hipótese nula.

Além do aglomerado mais verossímil, o método também identifica aglomerados secundários com altos valores de verossimilhança. Alguns desses aglomerados podem estar relacionados ao aglomerado mais verossímil tendo áreas de sobreposição. Aglomerados secundários que estão localizados em outras regiões da área de estudo são de interesse especial e são definidos como aglomerados que não se sobrepõem ao aglomerado de maior verossimilhança.

Dois tipos diferentes de modelos probabilísticos podem ser considerados quando se aplica a estatística espacial scan: (A) **Bernoulli**, e (B) **Poisson**.

A – No **modelo Bernoulli**, cada unidade de medida corresponde a uma entidade ou indivíduo que pode ter qualquer um dos dois *status*: “casos” ou “não casos”. Isso pode representar pessoas com ou sem a doença, ou ainda, pessoas com diferentes estágios de doenças, podendo significar casos e controles de uma população em geral. Os indivíduos em quaisquer dessas situações são identificados como pontos e sua localização constitui o processo pontual, ou ainda, os eventos podem ser agregados para áreas pré-determinadas (bairros, municípios, distritos, etc.).

B – No **modelo Poisson**, o número de eventos em cada área é considerado como distribuído de acordo com a população sob risco conhecida. Sob a hipótese nula o número esperado de casos em cada área é proporcional ao tamanho de sua população. O modelo Poisson requer a contagem de casos e população para cada região de localização dos eventos (ex: bairros, municípios, distritos, etc.), assim como o conhecimento das coordenadas geográficas para cada uma destas localizações.

Para qualquer um dos modelos descritos, Bernoulli ou Poisson, a estatística *scan* ajusta para a densidade populacional irregular presente em praticamente todas as populações, e a análise é condicionada ao número total de casos observados.

A escolha de qual dos modelos deve ser utilizado, Bernoulli ou Poisson, depende do tipo de dado disponível. Para dados binários, o modelo Bernoulli é o indicado, por outro lado, para dados em forma de contagens de eventos o modelo Poisson é o indicado (Kulldorff, 2003).

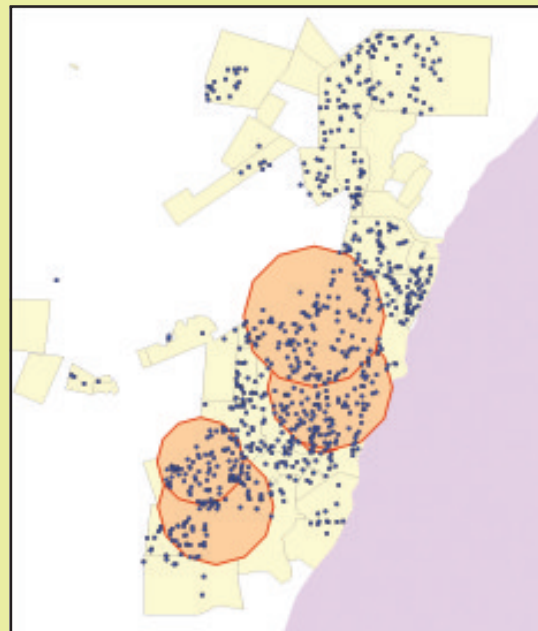
Vantagens e desvantagens da estatística *Scan*

Comparada com outros métodos estatísticos espaciais, a estatística *scan* apresenta **vantagens** que a torna indicada na localização de aglomerados: (a) pode ser ajustada para a densidade populacional; (b) pode ser ajustada por variáveis de confusão; (c) sua aplicação minimiza viés de pré-seleção, uma vez que procura por aglomerados sem especificar seu tamanho e localização; (d) o teste estatístico de razão de verossimilhança leva em conta testagens múltiplas informando um único valor de p ao testar a hipótese nula (Kuldorff *et al.*, 1997).

Por outro lado, uma das principais **desvantagens** do método consiste no fato de que os aglomerados são sempre definidos como círculos. Nessa situação, se uma área de baixa ocorrência de casos está rodeada por áreas com maior número de casos, ela será sempre incluída no aglomerado, apesar de poder apresentar características diferentes (Sankoh *et al.*, 2001). Ou seja, existe a tendência de identificar um aglomerado maior do que ele é, uma vez que regiões onde nenhum caso foi registrado podem ser incorporadas na identificação do aglomerado.

O resultado da técnica de varredura *scan* para o caso da hepatite A em Macapá é mostrado na figura 2.11.

FIGURA 2.11 – Resultado do teste *Scan* dos casos de hepatite A de Macapá. Eventos de saúde (pontos), bairros (polígonos) e círculos mostrando áreas com maior incidência de hepatite A.



Para proceder a análise com a técnica estatística *scan*, utiliza-se o aplicativo SaTScan. Adicionalmente, o aplicativo permite a padronização de taxas por meio do ajuste pela estrutura etária da população. Além disso, a estatística *scan* incorpora a possibilidade de aplicação de métodos multivariados de análise que levam em conta a convergência de variáveis potencialmente de confusão.

O **Satscan**, é um aplicativo gratuito desenvolvido pelos professores Martin Kulldorff, do Instituto de Câncer dos Estados Unidos e Dr. Farzad Mostashari, do Department of Health and Mental Hygiene da Cidade de Nova Iorque.

A entrada de dados no programa é feita com arquivos em *dbf*, *csv* ou *txt*, que tenham as coordenadas (em qualquer sistema de projeção) e exporta os resultados como um arquivo de texto (*txt*) que contém as coordenadas do centro do aglomerado e seu raio nas mesmas unidades em que foram entrados os dados (metros, quilômetros ou graus). Esses dados podem ser convertidos em um objeto cartográfico que pode ser disposto no Terraview. Desse modo, o Satscan pode ser usado como aplicativo auxiliar para análise e teste de hipóteses que não podem se feitas em outros programas.



Fonte: <http://www.satscan.org/>

• Estimativa Kernel considerando a População •

Podemos usar a técnica de Kernel para avaliar a distribuição espacial dos eventos, considerando também a distribuição da população. É uma maneira gráfica de se obter uma estimativa de **taxa de incidência**. A análise mais simplificada consiste na comparação visual de mapas suavizados para cada um dos processos (Santos *et al.*, 2001). Contudo, outras abordagens devem ser consideradas.

O estimador Kernel único apresenta como desvantagem o fato de não levar em conta a distribuição populacional onde os eventos estão inseridos. Se a população se distribuisse uniformemente na região, então um Kernel único de eventos seria suficiente. Contudo, uma vez que isso raramente ocorre, métodos que levem em conta essa população devem, também, ser considerados. Nesse cenário epidemiológico assume-se que a intensidade de eventos vai variar com a densidade populacional (Bailey e Gatrell, 1995).

Para se construir um mapa de densidade demográfica num SIG como o Terraview, deve-se dispor de um conjunto de dados de população como atributos de uma unidade espacial. Os municípios do Brasil são unidades espaciais que possuem dados de população, não só o total, mas também fracionada por faixa etária, sexo, renda e nível de instrução. Dentro de áreas urbanas a melhor fonte de dados de população é o setor censitário.

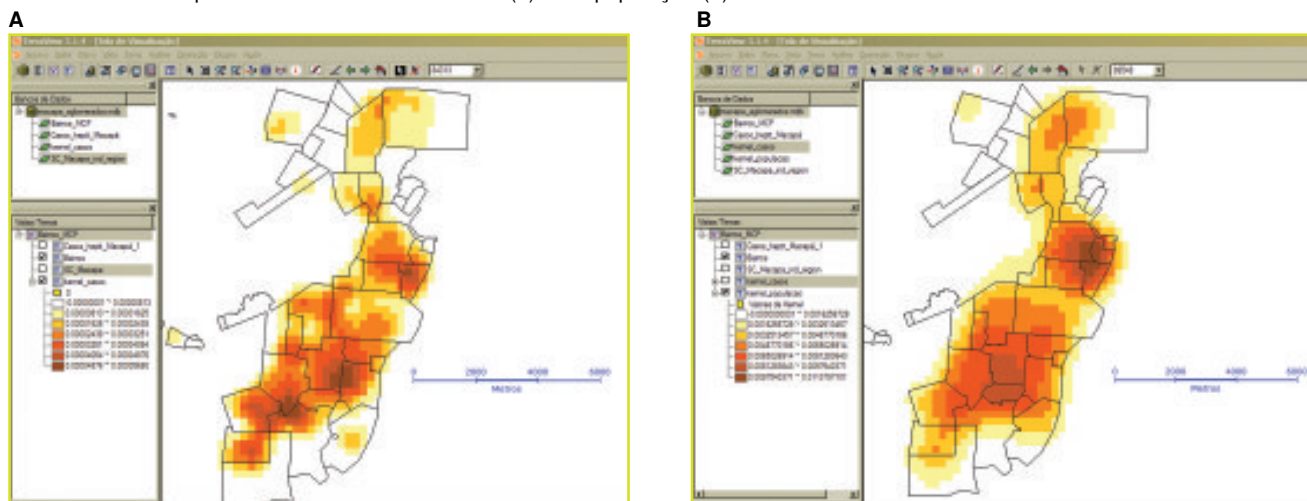
O estimador Kernel, como referido, avalia os eventos tendo o ponto como unidade de análise. Levando-se em conta que o dado populacional é agrupado e a unidade mínima oficial de agrupamento é o setor censitário, ao atribuir-se ao centróide dessa região o dado populacional conhecido, tem-se a possibilidade de estimar a intensidade desse processo (população) usando-se uma função Kernel. Com essa abordagem, Kelsall e Diggle (1995) consideram o problema da estimativa da variação espacial do risco relativo equivalente à estimativa da razão de duas densidades bivariadas. A razão entre as duas funções será a intensidade do risco (Carvalho e Câmara, 2002) e o mapa produzido a partir dessa razão será um mapa de risco relativo (Sabel *et al.*, 2000).

O método é baseado no cálculo da densidade de casos (número de casos por área), produzindo uma superfície de probabilidades onde áreas com casos mais próximos apresentam maior risco, cujo denominador utilizado para a taxa é, em uma outra camada, a densidade de pessoas (habitantes por área, ou simplesmente a densidade demográfica) também como uma superfície contínua.

Se conseguirmos, agora, dividir o kernel do número de eventos pelo kernel de população teremos a distribuição das taxas de incidência. A unidade do kernel de eventos é número de casos por metro quadrado. A unidade do kernel de população é número de habitantes por metro quadrado. A divisão desses dois valores resultaria em número de casos por habitante, que é a unidade de taxa de incidência.

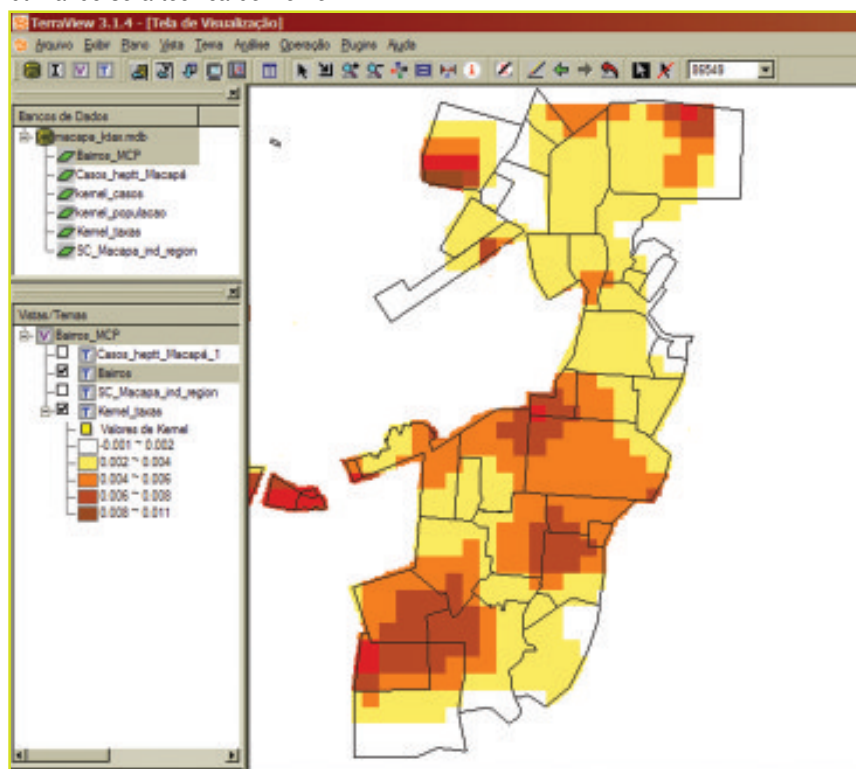
A figura 2.12 mostra a estimativa de distribuição de casos de hepatite A e de população em Macapá usando a técnica de Kernel. A unidade espacial utilizada é o setor censitário, e a variável usada para estimativa é a população total.

FIGURA 2.12 – Superfície de densidade de casos (a) e de população (b) calculados usando-se a técnica de Kernel.



No Terraview e no Crimestat é possível se construir um Kernel dual, ou uma razão de Kernel, que seja calculado utilizando-se a razão entre duas variáveis. A primeira variável será o número de eventos, que está relacionada à camada de casos (pontos), e a segunda variável será a população, que é um atributo da camada setor censitário. O resultado para a análise de casos de hepatite A em Macapá pode ser visto na figura 2.13.

FIGURA 2.13 – Superfície da taxa de incidência de hepatite A em Macapá calculada utilizando-se a técnica de Kernel.



Pode-se observar uma grande variabilidade dos valores obtidos para as taxas. Algumas das áreas com maiores valores situam-se nas periferias da cidade, inclusive fora dos seus limites. Áreas em vermelho, denotando altas taxas estão localizadas fora dos limites dos bairros (polígonos no mapa). Isso acontece devido à extrapolação da estimativa de kernel, chamado efeito de borda. Somente os valores de taxas contidos nos limites da região de estudo podem ser considerados, isto é, as taxas só podem ser calculadas dentro de áreas onde existem dados de população. Nesse caso, podem ser apontados aglomerados na área sul (bairro de Buritizal) e nordeste (bairro do Perpétuo Socorro) da Cidade de Macapá.

2.7 – Interpolação de Dados Pontuais

Esta abordagem também se aplica na comparação entre dois processos espaciais (ex: casos/controles; homens/mulheres) (Kelsall e Diggle, 1995). Essa estratégia é particularmente interessante quando não dispomos de dados de população, ou quando estamos trabalhando com inquéritos ou amostras. Nesse caso, podemos comparar a distribuição espacial de casos com o total de amostras. Imagine um inquérito sobre tuberculose em que algumas pessoas com a doença (sorteadas aleatoriamente no espaço) são perguntadas sobre o abandono do tratamento. A resposta a essa pergunta (sim ou não, portanto uma variável dicotômica) deve ser convertida em um valor numérico (1 ou 0). A densidade de respostas sim em relação ao total de respostas fornece uma estimativa de proporção de abandono do tratamento representada no espaço.

Essa técnica foi usada em Goiânia para avaliar a distribuição espacial da soroprevalência de dengue. Cada pessoa sorteada no inquérito foi localizada no trecho de rua, conforme o endereço anotado no momento da visita. A soropositividade de dengue foi obtida por exame em laboratório a partir de amostras de sangue. A variável soropositividade foi dicotomizada (transformada em 0 ou 1, pontos vermelho e verde no mapa) e foi calculado o Kernel usando-se como atributo o valor da soropositividade (Figura 2.14).

No mapa ao lado observa-se a área periférica da cidade, principalmente no sudeste com altas taxas de prevalência de dengue.

Em alguns casos, principalmente nas avaliações de qualidade ambiental, as variáveis não são dicotômicas. Por exemplo, os resultados da análise de amostras de água mostram uma variação gradual dos valores de cloro, flúor, turbidez e outros parâmetros usados para o controle e vigilância. Essas variáveis são chamadas contínuas e podem ser usadas em uma técnica de Kernel. Nesse caso, a resposta gráfica da técnica é uma grade com valores interpolados da variável selecionada.

O exemplo a seguir (Figura 2.15) mostra o resultado de técnica de Kernel aplicada à qualidade da água no Rio de Janeiro. Cada ponto no mapa mostra um local de coleta de amostra para avaliação da qualidade. A variável usada para avaliar a qualidade da água nesse caso foi a concentração de flúor (em mg/l).

Essa é uma técnica muito utilizada para se obter interpolações de dados no espaço. As interpolações são importantes para se avaliar a distribuição espacial de um processo contínuo no espaço, baseadas em amostras pontuais deste processo. Isso se dá principalmente para os processos ambientais, em que se espera que variem suavemente no espaço e que tenham uma forte dependência espacial. No caso do controle da qualidade da água, é óbvio

FIGURA 2.14 – Estimativa da distribuição espacial de prevalência de dengue em Goiânia (Marteli *et al.*, 2004).

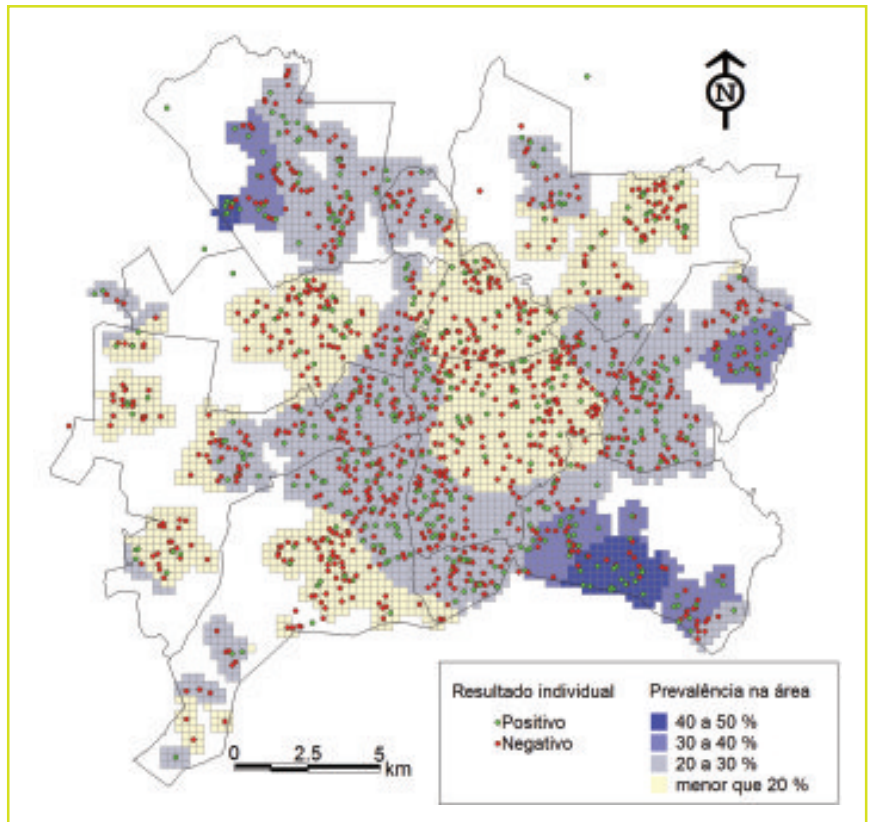
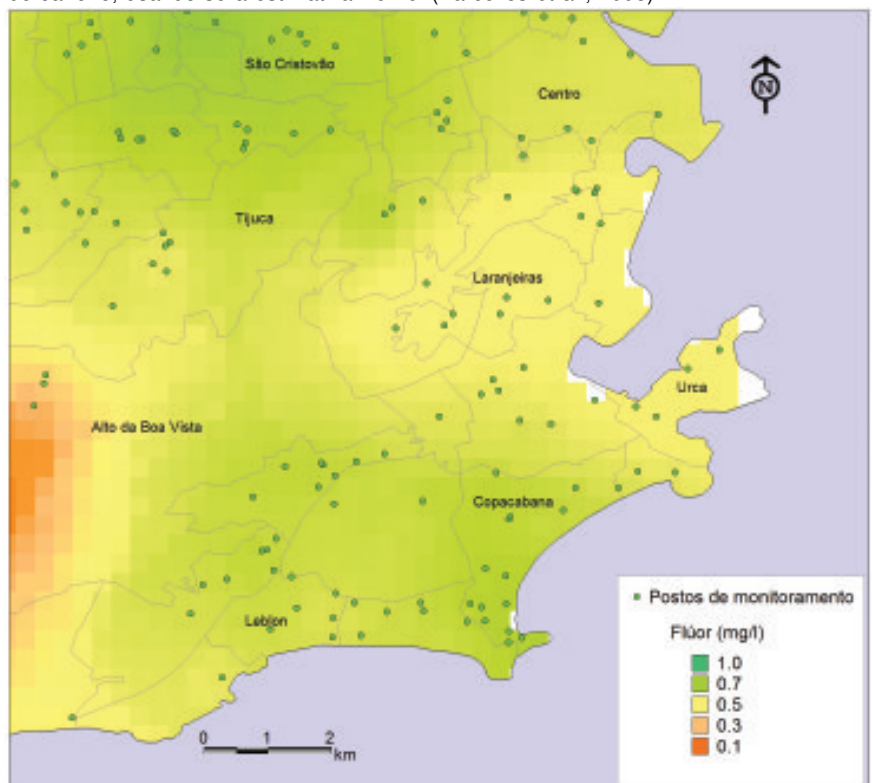


FIGURA 2.15 – Interpolação de valores de flúor na rede de distribuição de água no Rio de Janeiro, usando-se a estimativa Kernel (Barcellos *et al.*, 1998).



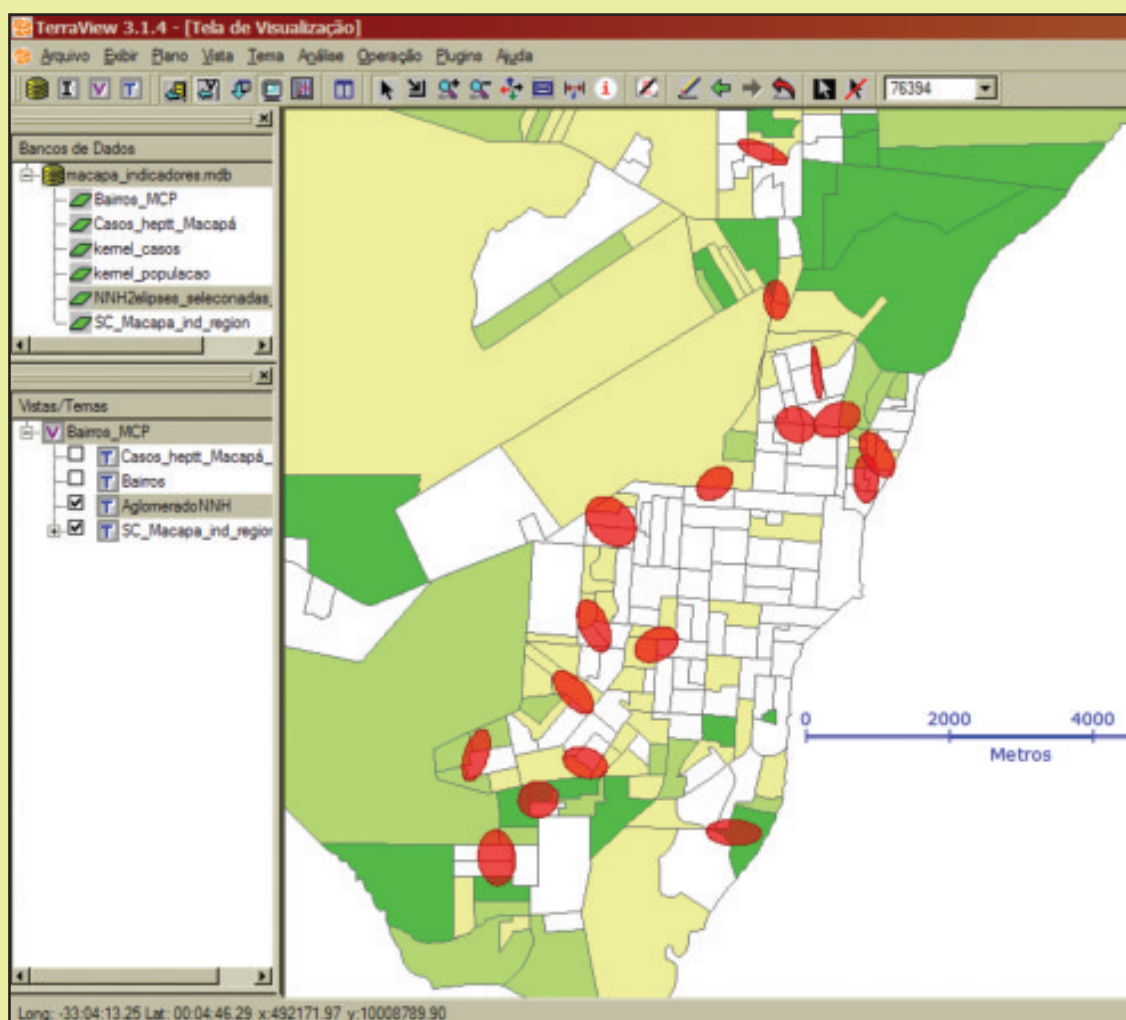
que não se dispõe de recursos para se amostrar todas as residências existentes em uma cidade. Os métodos de interpolação permitem que se estime essa qualidade em áreas não amostradas.

Outros recursos de um SIG

O trabalho de localizar e testar os possíveis aglomerados de pontos pode ser feito dentro de um SIG ou usando-se um outro programa que importe dados deste SIG e exporte resultados para este SIG. O programa Terraview tem sido aperfeiçoado para incorporar diversas ferramentas de análise espacial que podem ser usadas com esse fim. Outras ferramentas estão disponíveis em programas gratuitos, como o R, Crimestat e Satscan.

Depois de identificar esses aglomerados ainda há muito a fazer. A Vigilância em Saúde deve procurar entender como e por que estas áreas, destacadas pelos métodos estatísticos, diferem das outras. Um trabalho importante nesta etapa é retornar os resultados estatísticos para o SIG para verificar o que caracteriza essas áreas. Isso pode ser feito através da confecção de mapas temáticos sobre as condições socioeconômicas, por setor censitário.

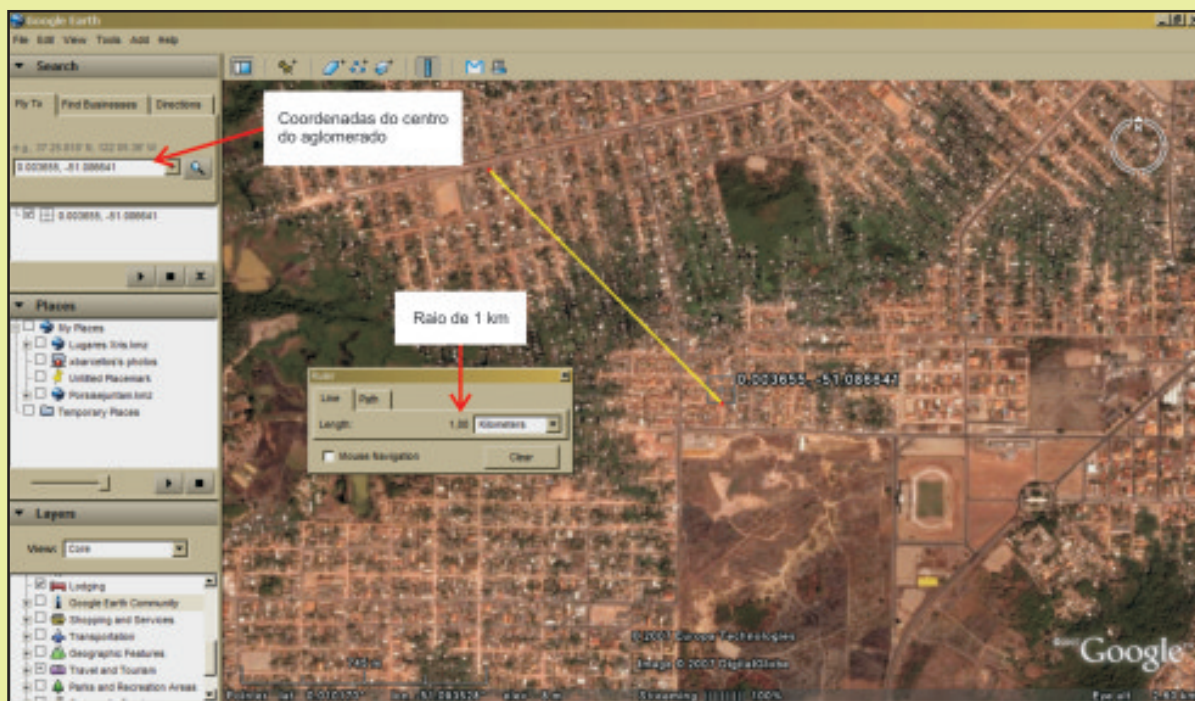
A figura abaixo mostra a proporção de domicílios sem banheiro nos setores censitários (em tons de verde). Sobre esta camada são mostrados os aglomerados de casos de hepatite A segundo o método do vizinho mais próximo (em pequenas elipses vermelhas).



Fonte: Neri, 2004.

Também podem ser adicionadas novas camadas de informação que auxiliem a interpretação dos resultados estatísticos, como a localização de unidades básicas de saúde, de córregos e rios, etc.

Outro recurso importante é avaliar a área identificada pelos métodos estatísticos por meio de outras fontes de informação. A figura abaixo mostra a mesma área ao sul da Cidade de Macapá segundo o programa *Google Earth*.



Fonte: <http://earth.google.com/intl/pt/>

Nesse programa, pode-se entrar com as coordenadas de um local em que se deseja visualizar a imagem de satélite. Com instrumentos de medição de distância se pode calcular um raio de 1 km em torno deste ponto, identificado como o centro do aglomerado de casos. Usando-se esse recurso pode-se visualizar que grande parte das casas dentro desse raio está em áreas de favelas de palafitas, com más condições de saneamento.



Capítulo

Análise Espacial de Dados de Áreas

Wayner V. Souza

Marilia Sá Carvalho

Oswaldo G. Cruz

Virginia Ragoni

Capítulo 3:

Análise Espacial de Dados de Áreas

3.1 - Métodos para Espacialização e Tratamento Estatístico de Indicadores

pág. 63

Instabilidade em Pequenas Áreas

pág. 63

Matriz de Vizinhança

pág. 65

3.2 - Métodos de Suavização Espacial

pág. 66

Média Móvel Espacial

pág. 66

Estimador Bayesiano Empírico

pág. 66

Autocorrelação Espacial

pág. 69

3.3 - Análise da Estrutura Espacial

pág. 71

Análise da Estrutura Espacial - Estatística de Moran

pág. 71

Análise da Estrutura Espacial - Estatística-Função LISA

pág. 72

Análise da Estrutura Espacial - Métodos Inteiramente Bayesianos

pág. 73

Análise da Estrutura Espacial - Método de Markov Chain-Monte Carlo

pág. 75

3.4 - Técnicas de Análise de Dados Pontuais Aplicadas a Áreas

pág. 78

3.1 – Métodos para Espacialização e Tratamento Estatístico de Indicadores

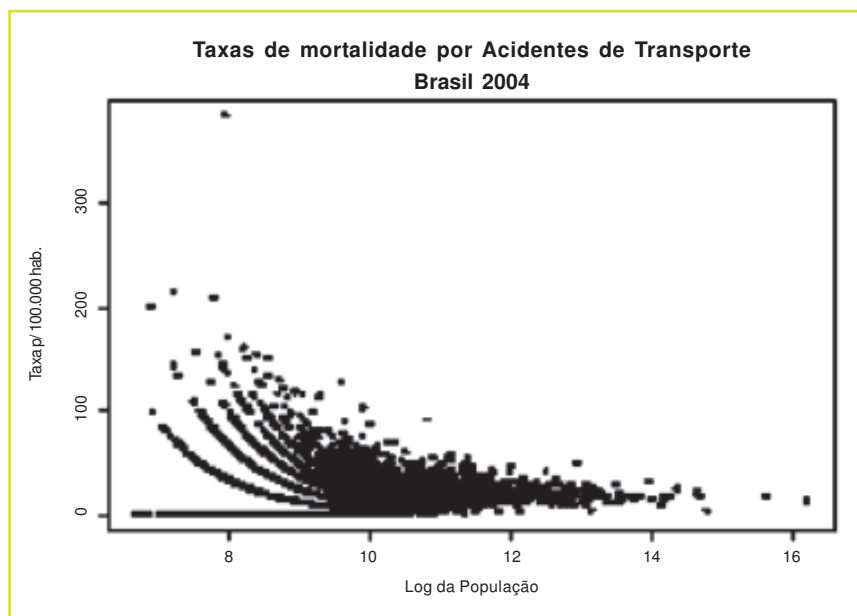
• Instabilidade em Pequenas Áreas •

As taxas brutas, como de incidência e de mortalidade, são os estimadores de risco mais simples e mais usados para o mapeamento da ocorrência de doenças e agravos, e são calculados como a razão entre o número de casos ou óbitos ocorridos na área e o número de pessoas expostas nesta área. Um grande problema associado ao uso de taxas brutas, porém, é a alta instabilidade que elas possuem para expressar o risco de um evento quando ele é raro e a população da região de ocorrência é pequena.

As grandes flutuações que ocorrem com essas taxas podem nada ter a ver com o fenômeno e sim com a ocorrência de um ou dois casos do evento a mais ou a menos numa determinada localidade, causando variações abruptas nas taxas brutas se a sua população for pequena, este efeito não ocorre em grandes populações.

A figura 3.1 mostra a dispersão das taxas de Acidentes de Transporte (CID-BR) *versus* o logaritmo natural da População para todos os municípios do Brasil no ano de 2004. Observe a forma de funil e a grande variabilidade encontrada nos municípios com populações de cerca de 3.000 habitantes ($\log N = 8$).

FIGURA 3.1 – Taxas de mortalidade por acidentes de transporte nos municípios brasileiros em 2004, segundo tamanho populacional.

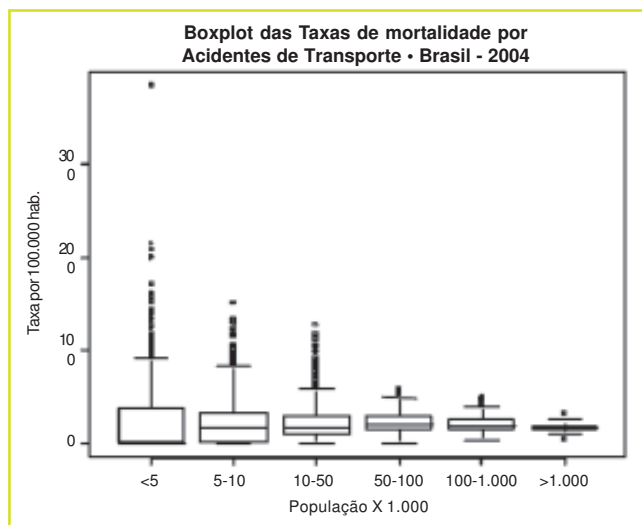


Fonte: DATASUS.

Pode-se observar melhor esse fenômeno construindo um *boxplot* com os mesmos dados, em intervalos da população como pode ser visto na figura 3.2. Cerca de 25% dos municípios brasileiros possuem menos de 5.000 habitantes. Assim sendo, podemos, com freqüência, ser levados a falsas interpretações e conclusões ao analisarmos mapas de taxas brutas.

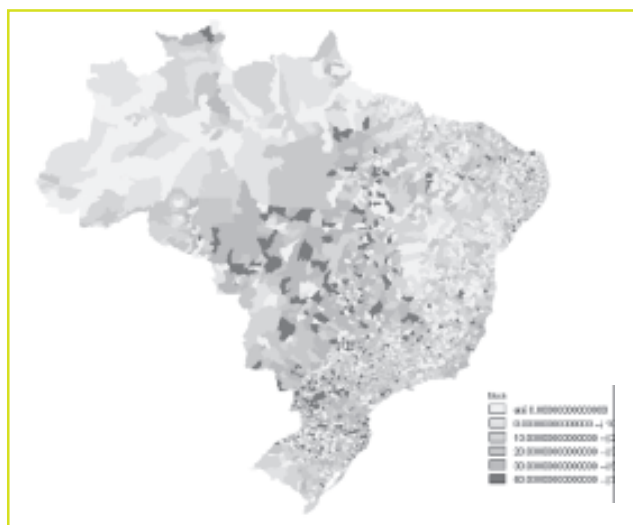
A figura 3.3 mostra o mapa da taxa de acidentes de transporte por 100.000 habitantes para o ano de 2004 nos municípios do Brasil. Pode-se observar, que por simples inspeção visual, não é possível identificar um padrão de ocorrência do evento, principalmente por estarmos representando taxas sujeitas a variações conforme demonstrado nas figuras 3.2 e 3.3.

FIGURA 3.2 – Taxas de mortalidade por acidentes de transporte nos municípios brasileiros em 2004, segundo faixas de tamanho populacional.



Fonte: DATASUS.

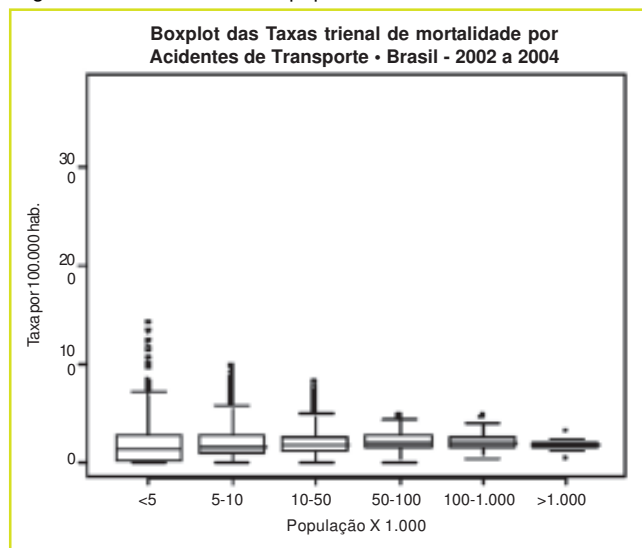
FIGURA 3.3 – Mapa temático das taxas de mortalidade por acidentes de transporte nos municípios brasileiros em 2004.



Log da População

Fonte: DATASUS.

FIGURA 3.4 – Taxas de mortalidade média por acidentes de transporte nos municípios brasileiros em 2002, 2003 e 2004, segundo faixas de tamanho populacional.



Fonte: DATASUS.

Figura 3.5 - Mapa temático das taxas médias de mortalidade por acidentes de transporte nos municípios brasileiros em 2002, 2003 e 2004.



Fonte: DATASUS.

O Estimador por Médias Móveis Temporais é uma das maneiras de se contornar o problema. Uma das mais freqüentes é a utilização da **média** trienal, ou seja, usa-se a média de casos no período de 3 anos para cada município, obtendo-se assim uma estabilização das taxas. A figura 3.4 mostra a média trienal da taxa dos Acidentes de Transporte para os anos de 2002, 2003 e 2004, representados na mesma escala que a figura 3.2. Note que o **efeito funil** ainda existe, mas que houve uma grande redução da variabilidade das taxas nos municípios de populações menores.

Olhando no mapa da figura 3.5, que utiliza a mesma escala da figura 3.3, podemos notar a suavização introduzida.

Existem outras maneiras de se lidar com esse problema, as quais serão discutidas a seguir. A base para esses métodos consiste na criação de uma matriz de vizinhança.

• Matriz de Vizinhança •

Um procedimento necessário para a análise de dados de área é a construção de uma matriz de vizinhança, também conhecida como matriz de distância, matriz de conectividade ou matriz de proximidade. Essa matriz de conectividade indica a relação espacial de cada área com as demais. Essa matriz pode ser composta, por exemplo, apenas pela lista de vizinhos de cada município; ou a distância entre municípios ligados por estradas; ou ainda a conectividade ponderada pelo comprimento da fronteira comum.

A figura 3.6 apresenta os municípios do Estado de Roraima, a partir dos quais foi construída uma matriz baseada na vizinhança, por meio da atribuição de valores: 1 para os municípios que possuem fronteiras em comum, e 0 para os municípios que não compartilham fronteiras.

A matriz de vizinhança referente ao mapa da figura 3.6 pode ser vista na figura 3.7.

Observe, por exemplo, que a capital Boa Vista tem 7 vizinhos e que o município de Uiramutã no extremo norte tem apenas 2 vizinhos.

FIGURA 3.6 – Mapa do Estado de Roraima com divisão por municípios.



Fonte: DATASUS.

FIGURA 3.7 – Matriz de Vizinhança para os municípios do Estado de Roraima.

	Amajari	Alto Alegre	Boa Vista	Bonfim	Cantá	Caracará	Caroebe	Iracema	Mucajá	Normandia	Pacaraima	Rorainópolis	São João da Baliza	São Luiz	Uiramutã
Amajari		1	1	0	0	0	0	0	0	0	1	0	0	0	0
Alto Alegre	1		1	0	0	0	0	1	1	0	0	0	0	0	0
Boa Vista	1	1		1	1	0	0	0	1	1	1	0	0	0	0
Bonfim	0	0	1		1	1	0	0	0	1	0	0	0	0	0
Cantá	0	0	1	1		1	0	1	1	0	0	0	0	0	0
Caracará	0	0	0	1	1		1	1	0	0	0	1	1	1	0
Caroebe	0	0	0	0	0	1		0	0	0	0	0	1	0	0
Iracema	0	1	0	0	1	1	0		1	0	0	0	0	0	0
Mucajá	0	1	1	0	1	0	0	1		0	0	0	0	0	0
Normandia	0	0	1	1	0	0	0	0	0		1	0	0	0	1
Pacaraima	1	0	1	0	0	0	0	0	0	1		0	0	0	1
Rorainópolis	0	0	0	0	0	1	0	0	0	0	0		1	1	0
São João da Baliza	0	0	0	0	0	1	1	0	0	0	0	1		1	0
São Luiz	0	0	0	0	0	1	0	0	0	0	0	1	1		0
Uiramutã	0	0	0	0	0	0	0	0	0	1	1	0	0	0	

3.2. Métodos de Suavização Espacial

• Média Móvel Espacial •

Quando se estabiliza a taxa de ocorrência de um evento utilizando o método de **médias móveis**, o valor atribuído a cada área é igual à soma dos eventos nos vizinhos dividida pela população total (somadas as populações de todas as áreas). Nesse caso, as áreas com maior população têm mais peso na estimativa desse indicador. Entretanto, a área do município central e a de seus vizinhos recebem o mesmo peso sem ser possível atribuir pesos distintos entre os municípios.

• Estimador Bayesiano Empírico •

Outra forma de suavizar a flutuação associada às pequenas áreas é usando-se o método denominado **estimador bayesiano empírico**. Nesse caso, supomos que a taxa real θ_i , que de fato expressa a força de ocorrência do evento na área i , é uma variável aleatória com média μ_i e variância θ_i^2 .

O melhor estimador dessa taxa é uma combinação linear entre a taxa observada (eventos/população) na área i e um valor médio μ_i ponderados por um fator w_i da seguinte forma:

$$\hat{\theta}_i = w_i t_i + (1 - w_i) \mu_i$$

Essa média usada na ponderação pode ser a taxa **média da região toda de estudo**, quando chamamos o método de **bayesiano empírico global**, ou pode ser a **média dos vizinhos**, método ao qual chamaremos de **bayesiano empírico local**.

O fator de ponderação w_i é tanto menor quanto menor a população em estudo da i ésima área, e reflete o grau de confiança a respeito de cada taxa. Contrariamente, se confiamos bastante no valor estimado da taxa (populações maiores), w_i se aproxima de 1 (um), e, conseqüentemente, o peso $(1 - w_i)$ da média dos vizinhos se aproxima de 0 (zero).

Para o caso de populações reduzidas, a confiança na taxa observada diminui e a estimativa da taxa se aproxima de nosso modelo *a priori* (ou seja, se aproxima de μ_i).

Regiões com populações muito baixas terão uma correção maior, e regiões populosas terão pouca alteração em suas taxas. Logo $\hat{\theta}_i$ será estimado, quando n for pequeno, com maior peso da média da vizinhança, no caso do estimador local ou da taxa geral da região no caso do estimador global.

Chamamos o método de **bayesiano empírico** porque a média e a variância serão estimadas a partir dos dados, considerando que cada valor observado é apenas uma realização de um mesmo processo de média μ_i e variância θ_i^2 .

Estimar a média é fácil:

$$\hat{\mu} = \frac{\sum y_i}{\sum n_i}$$

, onde y_i é a contagem do evento em cada área e n_i a população. No caso da suavização bayesiana local, esta é a taxa do evento entre os vizinhos. No caso global é a taxa da região toda. A variância, da mesma forma, é estimada com base na variância das taxas de cada área em relação à média μ_i .

Voltando então ao fator de ponderação w_i , este é a variância da taxa entre os vizinhos dividida pela soma desta variância com uma fração da média μ_i/n_i :

$$w_i = \frac{\theta_i^2}{\theta_i^2 + \frac{\mu_i}{n_i}}$$

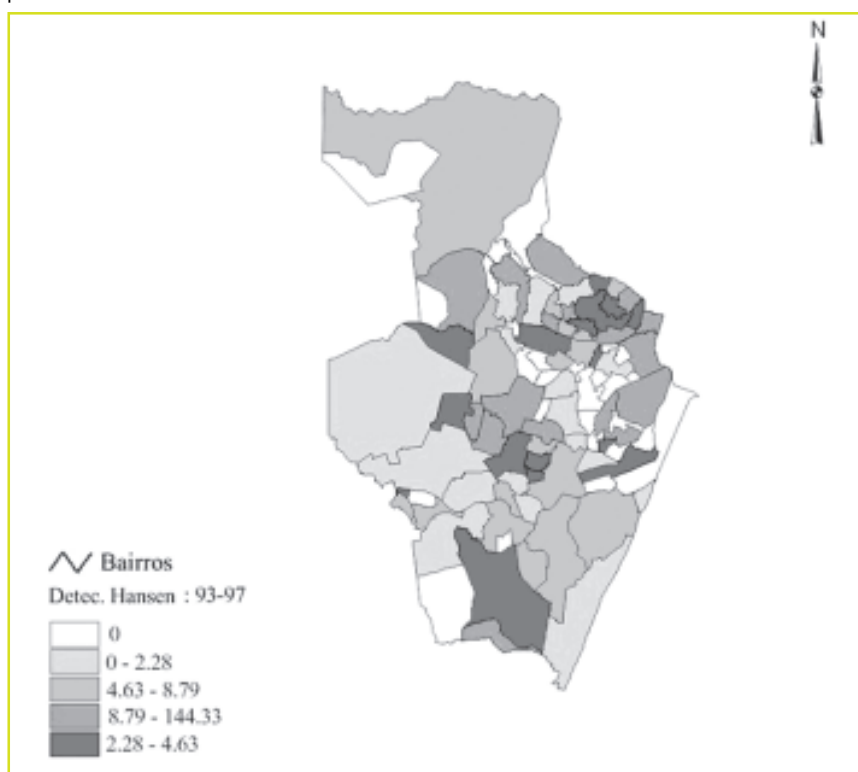
Observe que quando a população n_i é muito grande, o elemento somado ao denominador se aproxima de zero, e o fator de ponderação fica igual a um. Ou seja, os vizinhos não influenciam o valor estimado da taxa. O que faz todo o sentido.

É interessante observar na figura 3.9, comparando com a figura 3.8, que não há mais bairros com taxa igual a zero. Isso significa que os locais com 0 (zero) casos, muito provavelmente, não tinham população suficiente para que os casos aparecessem. Por outro lado, a maior taxa agora é 13,59 em vez de 144,33 por 100.000 habitantes dadas as suavizações dos valores originais, evidenciada no mapa. Além disso, diversos bairros mudaram de classe. O mais importante é que os dados agora refletem melhor o risco real de detecção de casos de hanseníase, e não a presença de flutuação aleatória.

Cabe lembrar que áreas que têm apenas um vizinho não se beneficiam muito dessa técnica, pois somente este vizinho contribuirá para estimar o indicador.

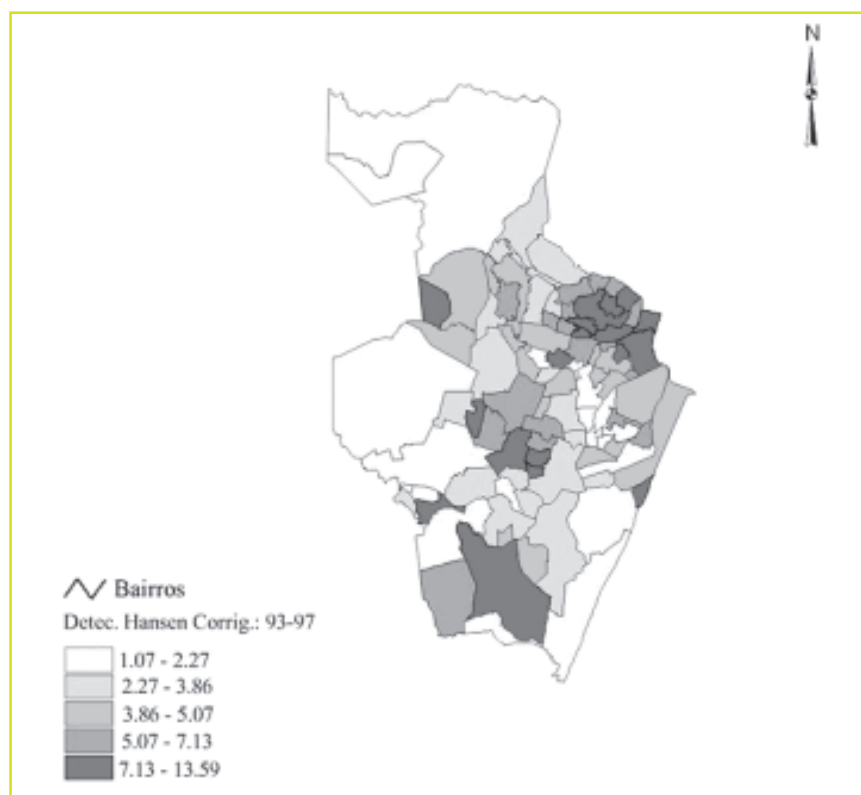
Alguns cuidados são essenciais na hora de calcular o estimador bayesiano, particularmente o tipo (critério) de **matriz de vizinhança** adotada. Pode-se adotar critérios de vizinhança por adjacência (como no exemplo acima), por distância (entre centróides, por exemplo), por quantidade (5 polígonos mais próximos, por exemplo), etc.

FIGURA 3.8 – Recife: Taxas de detecção média de hanseníase em menores de 15 anos, para o período 1993-1997, por bairro.



Fonte: SES-PE, 1998.

FIGURA 3.9 – Recife: Taxas de detecção média obtidas por método bayesiano empírico local, em menores de 15 anos, para o período 1993-1997, por bairro.



Fonte: SES-PE, 1998.

• Autocorrelação Espacial •

A estrutura de dependência entre os valores observados nas várias áreas do fenômeno em estudo é analisada pela função de **autocorrelação espacial**. Autocorrelação, como o próprio nome indica, mede a correlação da própria variável, e, sendo espacial, no espaço. A correlação de uma variável com ela mesma, medida no mesmo local, será sempre 1 (UM). Entretanto, a correlação de uma variável com ela mesma, porém medida nas áreas vizinhas, terá um valor que varia entre -1 e 1 (como qualquer medida de correlação). Quanto mais próximo de 1 (UM), maior a semelhança entre vizinhos. O valor 0 (zero) indica inexistência de correlação, e valores negativos indicam dessemelhança.

Vejamos como isso funciona:

– Em primeiro lugar, como a função de autocorrelação é **espacial**, precisamos utilizar a matriz de vizinhança, como no exemplo ao lado (Tabela 3.1), onde consideramos **vizinhos** os municípios que partilham de **fronteira comum**. Nesse caso, a matriz é simétrica: valor 1 (UM) quando são vizinhos e 0 (ZERO) caso contrário.

– O valor do indicador em estudo, nesse exemplo, a taxa de mortalidade infantil (TMI), está apresentado na última coluna da tabela. A média e a variância apresentadas não são parâmetros globais, simplesmente são parâmetros calculados para a distribuição das taxas dos municípios.

TABELA 3.1 – Matriz de vizinhança espacial por adjacência e Taxas de Mortalidade Infantil (/1000 Nascidos Vivos) - Exemplo

MUNICÍPIO	Água Santa	Bebedouro	Cacimba	Nascente	Poço	TMI
Água Santa	0	1	0	1	1	22,3
Bebedouro	1	0	0	1	0	26,5
Cacimba	0	0	0	0	1	18,4
Nascente	1	1	0	0	0	23,0
Poço	1	0	1	0	0	20,7
Média das Taxas (\bar{y})						22,18
Variância das Taxas						7,17

A função de correlação usual “compara” valores de duas variáveis. A **função de autocorrelação** é a correlação do valor do indicador comparado aos valores do mesmo indicador nos municípios vizinhos. Assim, matematicamente, a função é expressa como:

$$1 = \frac{1}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \times \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{VAR(y)}$$

sendo w_{ij} o indicador de vizinhança (no nosso exemplo 0 ou 1), y_i o valor do indicador na área i e y_j o valor em j , \bar{y} a média e $VAR(y)$ a variância das taxas.

– Traduzindo os somatórios e produtos em palavras, temos no numerador que, sempre que i e j forem vizinhas (1 na matriz acima), somamos o produto do desvio das áreas i e j em relação ao valor médio \bar{y} . No denominador temos simplesmente a variância do indicador.

Na tabela 3.2 apresentamos o cálculo da função de autocorrelação para nosso exemplo. Na última coluna, o valor registrado é cada parcela que entra no somatório. Observe que quando os municípios não são vizinhos, a contribuição deles é **nula** para o cálculo do indicador.

TABELA 3.2 – Auxílio ao cálculo da autocorrelação espacial.

DUPLAS DE MUNICÍPIOS	Vizinhança (W_{ij})	Taxa em i (y_i)	Taxa em j (y_j)	$w_{ij} x(y_i - \bar{y})x(y_j - \bar{y})$
Água Santa x Bebedouro	1	22,3	26,5	0,518
Água Santa x Cacimba	0	22,3	18,4	0,000
Água Santa x Nascente	1	22,3	23,0	0,098
Água Santa x Poço	1	22,3	20,7	-0,178
Bebedouro x Cacimba	0	26,5	18,4	0,000
Bebedouro x Nascente	1	26,5	23,0	3,542
Bebedouro x Poço	0	26,5	20,7	0,000
Cacimba x Nascente	0	18,4	23,0	0,000
Cacimba x Poço	1	18,4	20,7	5,594
Nascente x Poço	0	23,0	20,7	0,000
SOMA	5	—	—	9,576

Ao compararmos essa tabela com a matriz de vizinhança pode-se verificar que as somas obtidas referem-se apenas a uma parte dela (inferior ou superior), com relação à diagonal principal. Como a matriz é simétrica, as somas que procuramos serão exatamente o dobro das aqui obtidas. Portanto:

$$\sum_{i=1}^n \sum_{j=1}^n w_{ij} = 2 \times 5 = 10$$

$$\sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y}) = 2 \times 9,576 = 19,152$$

$$I = \frac{1}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \times \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{VAR(y)} = \frac{1}{10} \times \frac{19,152}{7,17} = 0,267$$

– Pode-se também adotar outra forma de **vizinhança**, por exemplo, por **distância**. Na coluna de vizinhança, w_{ij} passa a ser um valor contínuo, resultante do produto da distância por 1 ou 0, conforme sejam ou não vizinhas as áreas consideradas.

– Poderíamos considerar além de **vizinhos de primeira ordem**, ou vizinhos diretos, os **vizinhos de segunda ordem**. O vizinho k de j será vizinho de segunda ordem de i , se i e j forem vizinhos.

Em nosso exemplo, o valor da função de autocorrelação, quando consideramos apenas vizinhos de primeira ordem, ou vizinhos diretos, é 0,267. O valor da função de autocorrelação, para vizinhos diretos, é denominado **estatística I de Moran**, e utilizado como teste para verificar a presença de *cluster* espacial, necessitando para tanto calcular a significância do valor encontrado.

3.3. Análise da Estrutura Espacial

• Análise da Estrutura Espacial - Estatística de Moran •

O conceito de **estacionariedade** aplicado a áreas geográficas significa que a média do processo medido em estudo não é constante em todas as subáreas.

Um bom exemplo de dado não estacionário é a incidência câncer de pele do tipo basocelular, que aumenta com: (a) a insolação, e conseqüentemente diminui conforme aumenta a latitude; (b) com a cor da pele da população, sendo mais freqüente em população de pele mais branca; e (c) com a presença de buracos da camada de ozônio, que diminuem a proteção contra os raios ultravioleta. Ou seja, fatores que fazem com que haja uma clara tendência espacial.

O **teste de Moran** para detecção de padrão de distribuição de eventos em *cluster* não deve ser usado quando o dado não é estacionário. Uma das formas de avaliar se o dado é ou não estacionário, além de olhar atentamente o mapa em busca de padrões, é utilizar a função de autocorrelação espacial considerando vizinhos de diversas ordens. Quando o dado não é estacionário a função decresce muito lentamente.

Uma observação importante dessa função é que, se houver uma variação na média conforme a subárea geográfica, por exemplo, aumentando com a latitude, os valores da estatística serão muito grandes. Isso porque ao comparar valores do indicador nos municípios com a média global (o numerador), se dois municípios estão do mesmo lado da área geográfica, os desvios em relação ao global serão na mesma direção – positiva ou negativa

– e quanto mais afastados da média, mais contribuirão para a estatística.

Um processo estocástico (aleatório) é estacionário quando sua distribuição de probabilidade é a mesma em qualquer ponto do tempo ou local. Por exemplo, se medirmos a taxa de mortalidade por câncer de pulmão, nos municípios do Brasil poderemos afirmar que a média e a variância em cada estado é aproximadamente constante.

• Análise da Estrutura Espacial - Estatística-Função LISA •

Outra abordagem para a **análise da estrutura espacial é a função LISA**, ou função de autocorrelação local. Da mesma forma que podemos utilizar o método bayesiano empírico global e local, pode-se estimar a autocorrelação de primeira ordem, global – Moran, e local – LISA. Nesse caso, em vez de utilizar o indicador global, utiliza-se o valor do indicador na subárea composta por vizinhos.

Essa estatística é interessante, pois permite comparar o valor de cada município com seus vizinhos. Em geral se utiliza o indicador normalizado, ou seja, a diferença entre a média global e o valor em cada município, dividida pelo desvio padrão, de forma que a unidade do indicador passa a ser unidades de desvio padrão de afastamento da média. Essa técnica permite visualizar graficamente, através do **diagrama de espalhamento de Moran**, o grau de similaridade entre vizinhos. Colocando-se no eixo X o valor do indicador do município i normalizado e no eixo Y o valor da estatística de Moran somente entre os vizinhos de i , cada quadrante do gráfico (Figura 3.10) pode ser interpretado como:

- **Q1** (valores positivos, médias positivas) e **Q2** (valores negativos, médias negativas): indicam pontos de associação espacial positiva, no sentido que uma localização possui vizinhos com valores semelhantes;
- **Q3** (valores positivos, médias negativas) e **Q4** (valores negativos, médias positivas): indicam pontos de associação espacial negativa, no sentido que uma localização possui vizinhos com valores distintos.

De forma similar aos indicadores globais, a significância do índice local de Moran (I_i) deve ser avaliada, utilizando-se hipótese de normalidade ou simulação de distribuição por permutação aleatória nos valores dos atributos (Anselin, 1995).

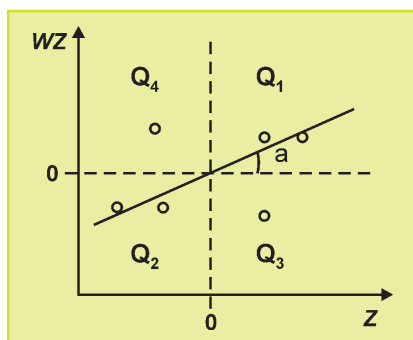
Uma vez determinada a significância estatística de Moran (I_i) é muito útil gerar um mapa indicando as regiões que apresentam correlação local significativamente diferente do resto dos dados.

Esse mapa é denominado por Anselin (1995) de “LISA MAP”.

Na geração do LISA MAP, os índices locais I_i são classificados como:

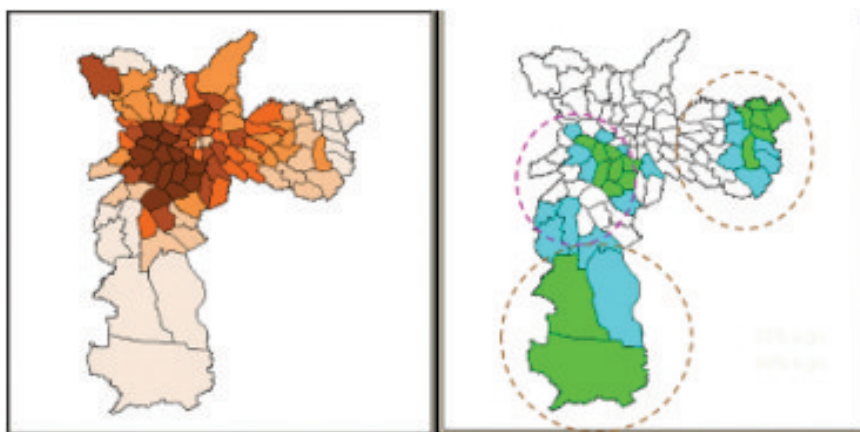
- não significantes;
- com significância de 95% (1,96 sd), 99% (2,54 sd) e 99,9% (3,2 sd).

FIGURA 3.10 – Diagrama de Espalhamento de Moran.



Na figura 3.11, o primeiro mapa mostra indicadores de qualidade de vida no município de São Paulo. No segundo mapa, as áreas em verde têm alta correlação entre vizinhos, isto é os indicadores se parecem. Nas áreas em azul claro, há uma pequena correlação entre indicadores de áreas vizinhas e nas áreas em branco, as áreas se parecem menos. Isso quer dizer que no município de São Paulo existem três grandes áreas homogêneas: uma no sul e outra na zona leste da cidade, com baixos índices de qualidade de vida, e uma área central com melhores índices de qualidade de vida.

FIGURA 3.11 – Indicadores Locais de Autocorrelação Espacial.



• Análise da Estrutura Espacial - Métodos Inteiramente Bayesianos •

A análise da distribuição espacial de eventos de saúde visa a identificar seu padrão espacial, por meio de mapas de risco, relativos às unidades territoriais de análise, de forma a apontar as áreas de maior gravidade do problema e facilitar o planejamento de intervenções de Saúde Pública. Para tal, pode-se utilizar o **método bayesiano completo** para estimação da função de densidade de probabilidade do risco relativo analisado espacialmente, levando em consideração co-fatores socioeconômicos e ambientais, além de indicadores de atenção à saúde e/ou de efetividade das ações de controle que interferem no processo, como forma de resolver as questões relativas à incorporação do espaço como categoria de análise, aqui levantadas.

Assim, a abordagem bayesiana propõe que o número de casos ocorridos em cada uma das n áreas geográficas, $\mathbf{y} = (y_1, y_2, \dots, y_n)$, são realizações de uma variável aleatória $\mathbf{y} = (y_1, y_2, \dots, y_n)$ com **distribuição de Poisson** e com média $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_n)$.

– Sob a hipótese de não haver diferenças entre as taxas de incidência, no espaço, pode-se estimar o vetor $\mathbf{e} = (e_1, e_2, \dots, e_n)$, de casos esperados para cada área, pela simples aplicação da taxa média de incidência do município, unidade da federação (UF), ou país à população de cada uma destas áreas, no período de estudo.

Sendo assim, podemos definir o **risco relativo para a ocorrência do evento** em cada área i (λ_i), como:

$$\lambda_i = \frac{\mu_i}{e_i}$$

O **modelo bayesiano** estima esses riscos, considerando a possível influência de co-fatores socioeconômicos e ambientais, de atenção à saúde e/ou de efetividade das ações de controle.

– Dado isso e a expressão $\lambda_i = \frac{\mu_i}{e_i}$, podemos escrever que $\mu_i = e_i \lambda_i$, ou, aplicando-se uma transformação logarítmica, $\log(\mu_i) = \log(e_i) + \log(\lambda_i)$, onde $\log(\lambda_i)$ é o logaritmo neperiano do risco relativo, para cada área e que pode ser modelado como:

$$\log(\lambda_i) = \alpha_0 + \alpha_1 x_{i1} + \alpha_2 x_{i2} + \dots + \alpha_k x_{ik} + \theta_i + v_i$$

onde α_0 é a constante do modelo e os α_j ($j = 1, 2, \dots, k$) são os parâmetros que refletirão a influência de cada co-fator sobre o risco relativo; x_{ij} ($j = 1, 2, \dots, k$) é o vetor de co-fatores referente a cada área i ; θ_i representa a componente aleatória espacialmente estruturada do risco relativo e v_i o componente aleatório não-espacial.

Sinteticamente, podemos dizer que o risco relativo é modelado por meio de um modelo log-linear, onde são consideradas duas componentes aleatórias: uma espacial e outra não-espacial.

– Portanto temos:

$$\log(\mu_i) = \log(e_i) + \alpha_0 + \sum_{j=1}^k \alpha_j x_{ij} + \theta_i + v_i$$

onde o componente v_i é um **efeito aleatório não-espacial** que permitirá, analisando-se sua distribuição, interpretar se as variáveis explicativas incluídas no modelo são suficientes na explicação do fenômeno, tal qual nos modelos de regressão clássicos.

– O componente θ_i é um efeito aleatório que representa um residual de risco relativo espacialmente estruturado, que pode ter sua estrutura definida usando-se um modelo auto-regressivo condicional normal (CAR-Normal) dado por:

$$p(\theta_i / \theta_{j \neq i}) \sim N \left(\frac{\sum_{j \neq i} w_{ij} \theta_j}{\sum_{j \neq i} w_{ij}}, \frac{\sigma^2}{\sum_{j \neq i} w_{ij}} \right)$$

onde w_{ij} sendo igual a 1 se os setores i e j forem adjacentes e 0 em caso contrário, o que significa dizer que foram considerados vizinhos os setores censitários adjacentes.

O modelo utilizado para definição da estrutura dos θ_i é um **modelo auto-regressivo condicional normal**, dado que o efeito aleatório espacial (θ_i) de

uma determinada área está condicionado tão somente aos efeitos aleatórios de sua vizinhança ($\theta_j ; j \neq i$), através de uma relação linear. Por conseguinte, θ_j tem distribuição normal com média igual à média de sua vizinhança e desvio padrão inversamente proporcional ao tamanho desta vizinhança.

Para implementação do modelo utiliza-se o **método de simulação MCMC (Markov Chain-Monte Carlo)** usando-se o programa WinBugs (<http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml>), que gera amostras da distribuição de $p(\mu / y)$, que é a distribuição conjunta *a posteriori* de α , θ e v , dadas as observações y .

• Análise da Estrutura Espacial - Método de Markov Chain-Monte Carlo •

Define-se um processo **Markoviano** como um **processo estocástico** que represente observações de um sistema, obedecendo a certas leis de probabilidade com o seguinte princípio: a probabilidade de o sistema estar, num determinado estado, no momento t_2 , pode ser deduzida a partir do conhecimento do seu estado num momento anterior t_1 , e não depende da história do sistema antes do momento t_1 .

Os processos de Markov podem ser classificados não só pelo seu parâmetro, mas também pelo conjunto de valores possíveis do processo estocástico, que pode ser contínuo ou discreto. Os processos de Markov, em que o conjunto de valores possíveis é **discreto**, são chamados de **cadeias de Markov**.

– Sendo assim, dado um conjunto discreto de variáveis aleatórias $X = (X_0, X_1, \dots, X_{m-1}, X_m, \dots, X_n)$, pode-se dizer que:

$$P(X_m = x_m / X_0 = x_0, X_1 = x_1, \dots, X_{m-1} = x_{m-1}) = P(X_m = x_m / X_{m-1} = x_{m-1})$$

A obtenção dos vários resultados desse processo estocástico (valores da variável aleatória X) se dá, então, por simulação, usando-se o método de Monte Carlo.

Nesse modelo, o que se necessita obter são os valores médios dos riscos relativos (λ_j) para cada setor censitário, que por sua vez é modelado em função de co-fatores que são ponderados pelos parâmetros α_j . Então temos de obter os valores médios de cada α_j , por simulação, assumindo que os α_j têm uma distribuição *a priori* pouco informativa, no caso uma distribuição Normal, com média igual a zero e variância muito grande, dado que o objetivo da modelagem dos λ_j é rejeitar a hipótese nula, $H_0 : \alpha_j = 0$.

O método de Monte Carlo consiste, então, na geração, através de números aleatórios da distribuição de probabilidade *a priori* adotada, de N (número de simulações) valores para cada α_j , ou seja, $\alpha_{j1}, \alpha_{j2}, \dots, \alpha_{jN}$. Assim, a variável $\alpha_{j+} = \alpha_{j1} + \alpha_{j2} + \dots + \alpha_{jN}$ tem, pelo **teorema do limite central**, distribuição aproximadamente Normal com média igual a $N \times m$ e variância igual a $N \times s^2$, onde m e s são, respectivamente, a média e a variância dos α_{jk} .

– Sendo assim, a variável $\bar{\alpha}_j = \frac{\sum_{k=1}^N \alpha_{jk}}{N}$ terá distribuição Normal com média igual a m e desvio padrão igual a $\frac{s}{\sqrt{N}}$.

Teorema do Limite Central

Sejam $X_1, X_2, X_3, \dots, X_n$ variáveis aleatórias independentes e identicamente distribuídas com média μ finita e variância $0 < \sigma^2 < \infty$ e seja $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$, então $\bar{X} \approx N\left(\mu; \frac{\sigma}{\sqrt{n}}\right)$.

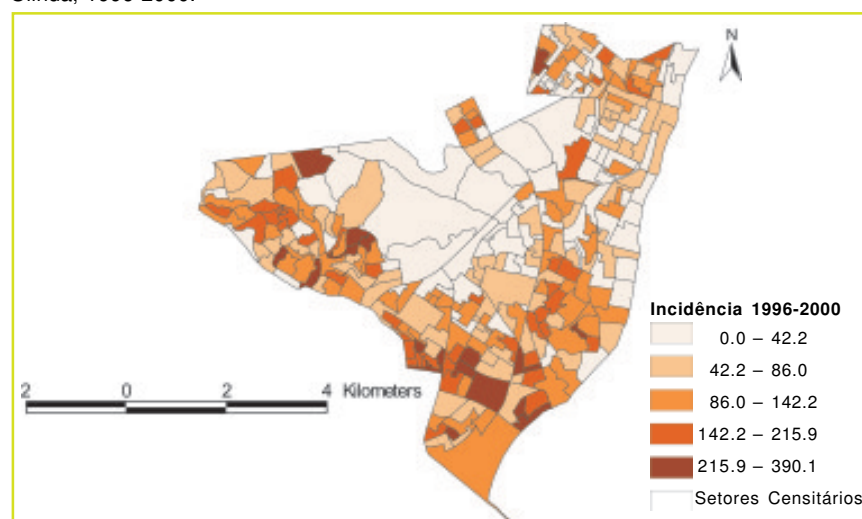
Procedimento idêntico deve ser adotado, com o mesmo método de simulação, para obtenção dos parâmetros θ e v incluídos na modelagem dos λ_i .

A título de exemplo, tomemos o caso da tuberculose em Olinda no período 1996-2000. Os dados referentes ao período são apresentados na tabela 3.3.

TABELA 3.3 – Casos de tuberculose notificados e georreferenciados em Olinda, 1996-2000.

ANO	Total de casos notificados	Referenciados por setor censitário	Proporção georreferenciada (%)	População ajustada para o meio do período	Incidência por 100.000 hab.
1996	433	366	84,5	349.243	124,0
1997	411	339	82,5	353.783	116,2
1998	419	360	85,9	358.381	116,9
1999	381	320	84,0	363.039	104,9
2000	340	293	86,2	367.758	92,5
TOTAL	1984	1678	84,6	358.381	110,7

FIGURA 3.12 – Taxas de Incidência Média de Tuberculose por 100.000 habitantes em Olinda, 1996-2000.



No período de 1996 a 2000, foram registrados 1.984 novos casos de tuberculose em residentes no Município de Olinda, resultando em uma taxa de incidência média superior a 1,0 caso novo para cada 1.000 habitantes, o que representa quase o dobro da média nacional.

O mapa de padrão de incidência segundo setores censitários é apresentado na figura 3.12, ao lado.

Variáveis explicativas e modelo

O modelo empregado foi:

$$\log(\mu_i) = \log(e_i) + \alpha_0 + \alpha_1 \log(x_1) + \alpha_2 \log(1+x_2) + \alpha_3 \log(1+x_3) + \alpha_4 \log(1+x_4) + \alpha_5 \log(1+x_5) + \theta_i + v_i$$

onde:

x_1 : média de habitantes por domicílio no setor;

x_2 : percentual de chefes de família sem instrução ou com menos de um ano de instrução, por setor;

x_3 : percentual de chefes de família sem rendimento ou com renda inferior a um salário mínimo mensal, por setor;

x_4 : número de casos de retratamento no setor;

x_5 : número de domicílios com mais de um caso no período, por setor.

Assim, o risco relativo foi estimado no modelo como:

$$\lambda_i = e^{(\alpha_0 + \sum_{j=1}^k \alpha_j x_j + \theta_i + v_i \lambda_i)}$$

Dessa forma, pode-se calcular a medida de efeito de cada co-variável, na magnitude do risco relativo, como sendo igual à exponencial do respectivo α_j .

A etapa final relativa à modelagem consiste na **análise de convergência dos modelos**, que é a verificação da estabilidade das estimativas produzidas por métodos de simulação, como o **MCMC**, e pode ser efetuada com emprego da rotina **BOA** ("bayesian output analysis", disponível em www.mrc-bsu.cam.ac.uk/bugs) executada pelo *software S-PLUS*.

Existe um grande elenco de métodos de análise de convergência, disponível na literatura especializada. Aqui foi utilizado o de **Heidelberger & Welch** (Brooks, 1998), que consiste em **testar a hipótese nula** de que os valores obtidos por simulação, para cada variável, **formam um processo estacionário**. A rejeição da hipótese nula indica a necessidade de produzir-se um número maior de simulações para o modelo.

Resultados

Pode-se observar os riscos relativos estimados pelo modelo na figura 3.13. Esses riscos são, em realidade, uma representação probabilística do risco de ocorrência da tuberculose, ou seja, uma identificação do processo estocástico gerador do risco, livre da interferência de flutuações aleatórias.

FIGURA 3.13 – Riscos Relativos para a ocorrência da tuberculose em Olinda, segundo setores censitários, com emprego de Método Bayesiano Completo.

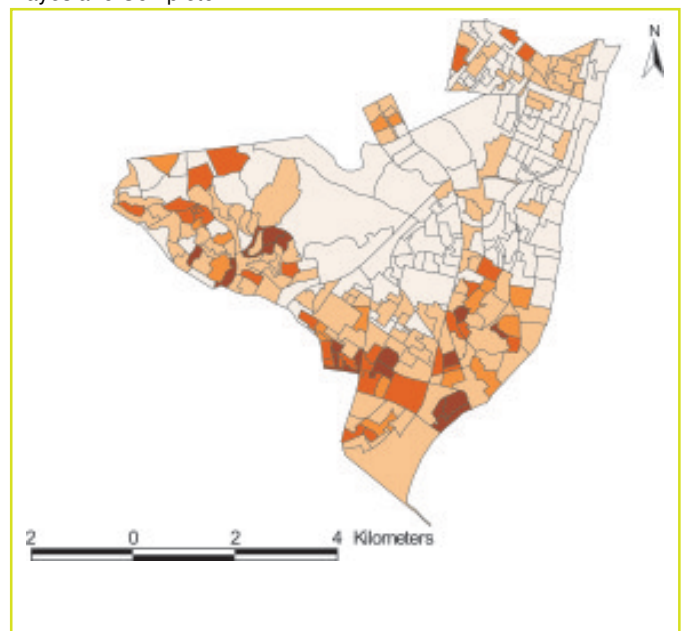


FIGURA 3.14 – Olinda: Setores Censitários com risco relativo para a ocorrência da tuberculose igual ou maior que 150%.



Esses resultados permitem analisar e identificar formas realistas de enfrentamento do problema diante da escassez de recursos, por vezes presente, conforme exemplificado na figura 3.14. Pode-se, por exemplo, no âmbito do SIG, selecionar áreas com risco relativo maior do que 1,5 (150%) como forma de identificar áreas prioritárias para intensificação de ações de vigilância e controle.

As áreas assinaladas (em vermelho) correspondem a 47 setores censitários que totalizaram 629 casos de tuberculose, no período 1996-2000 (37,5% do total), tendo uma população de 59.587 habitantes (16,6% do total), o que implica uma incidência média anual de 211,1 por 100.000 habitantes.

3.4. Técnicas de Análise de Dados Pontuais Aplicadas a Áreas

Na análise espacial de taxas epidemiológicas e eventos de saúde referidos a áreas geográficas, pode-se estar interessado na produção de mapas de superfícies contínuas como forma de produzir uma visualização menos fragmentada da distribuição do evento, identificando as áreas “quentes”. Para tal, pode-se adotar técnicas até aqui descritas como apropriadas à análise de dados pontuais (como o estimador de intensidade Kernel, por exemplo), obviamente cercando-se de alguns cuidados no que diz respeito à interpretação dos resultados.

A título de exemplo, usaremos os dados de uma coorte de casos de tuberculose, notificados entre 2003 e 2005 na Cidade do Recife, gentilmente cedidos pelo Prof. Ricardo Ximenes, da Universidade Federal de Pernambuco, que coordena o Projeto de Pesquisa que coletou os dados.

Nesse exemplo, foram usadas as informações sobre setor censitário e bairro de residência de 1.092 desses casos, permitindo o cálculo das densidades de incidência para cada um desses níveis de agregação.

Na figura 3.15 apresentamos o mapa da Cidade do Recife, com suas divisões administrativas segundo bairros e distritos sanitários, e na figura 3.16, o mapa clássico de padrão das taxas de incidência calculadas para os bairros.

FIGURA 3.15 – Recife – Divisão de bairros (polígonos delimitados em preto) e distritos sanitários (polígonos delimitados em azul).

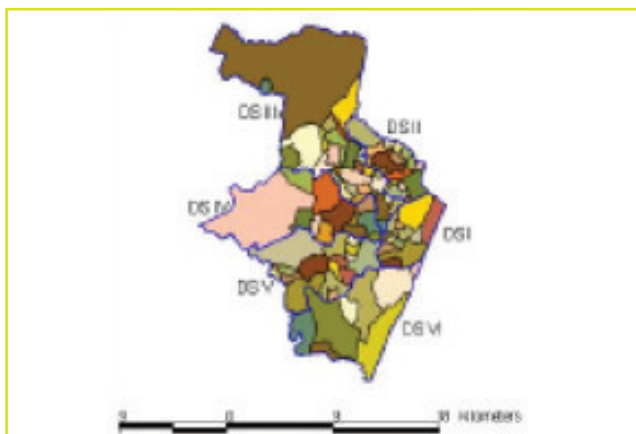
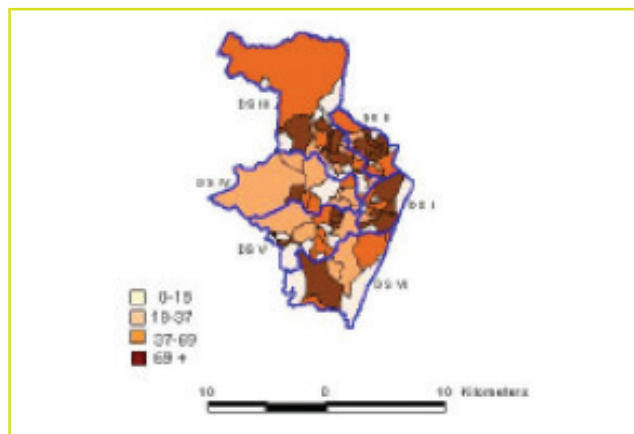


FIGURA 3.16 – Mapa de padrão de incidência de tuberculose segundo bairros – Coorte 2003-2005.



A adoção de técnicas de análise de dados pontuais referida pode ser feita identificando-se as coordenadas geográficas do centróide de cada polígono (bairro ou setor censitário), atribuindo-se a esta localização os casos ocorridos na área, assim como sua população. Note-se que ao adotarmos tal procedimento, estaremos incluindo uma fonte de erro, por atribuirmos todas as ocorrências de uma determinada área a um único ponto, que é seu centróide. Esse erro seria possivelmente menor se atribuíssemos essas ocorrências ao centróide populacional do polígono, o que não é o mais comum, devido a dificuldades de se obter mapas de densidade populacional. O mais usual é atribuir as ocorrências ao centróide geográfico do polígono, o que remete a uma discussão de escala, visando a controlar tal erro.

Temos então que analisar a questão da escala que desejamos trabalhar (bairro ou setores), além de outra que se refere ao tipo de intensidade que queremos estimar (casos ou incidência).

As figuras 3.17 e 3.18 mostram mapas de Kernel apresentando as intensidades de incidência e de casos, respectivamente, adotando-se o bairro como unidade de análise.

FIGURA 3.17 – Mapa de Kernel de **incidência** de tuberculose segundo bairros.

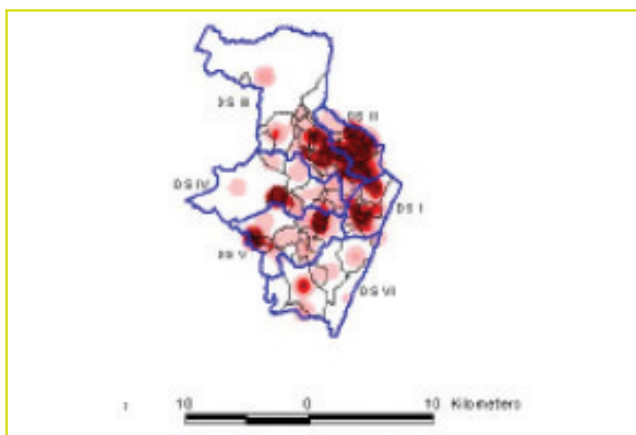
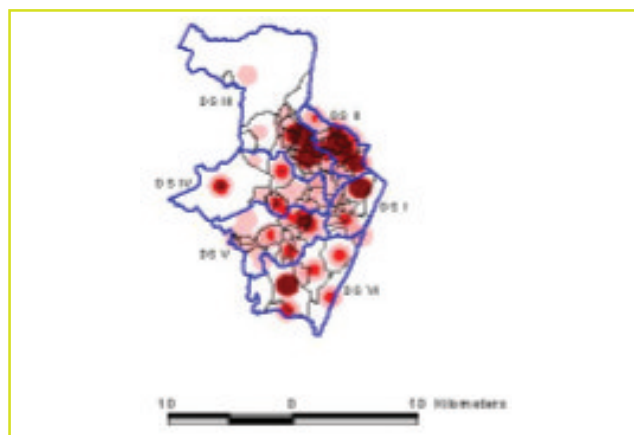


FIGURA 3.18 – Mapa de Kernel de **casos** de tuberculose segundo bairros.



Analisando-se esses mapas podemos identificar que, exceto a região noroeste da cidade (que sempre se apresenta como área crítica para a ocorrência da tuberculose), existem diferenças importantes entre o mapa de intensidade de incidência e de casos. Se quisermos analisar a gravidade do problema, em termos de risco de adoecer, temos que utilizar o mapa que fornece a intensidade da incidência (Figura 3.16), por considerar os tamanhos das populações sob risco, que no caso dos bairros do Recife apresentam variações importantes. Por outro lado, se estamos adicionalmente interessados em planejar a infra-estrutura de vigilância e assistência, também deve ser observado o mapa que fornece informação sobre intensidade de casos (Figura 3.18).

Nas figuras 3.19 e 3.20, apresentamos as mesmas abordagens, agora considerando o setor censitário como unidade de análise. Nesse caso as diferenças entre os dois mapas já não são tão marcantes, dado que tal unidade geográfica é constituída pelo IBGE com tamanhos populacionais com pouca variação, na maioria das vezes.

FIGURA 3.19 – Mapa de Kernel de incidência de tuberculose segundo setores censitários (polígonos não representados).

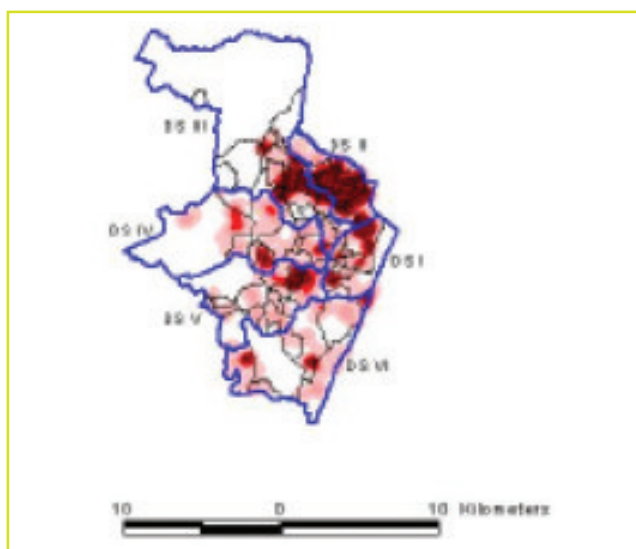
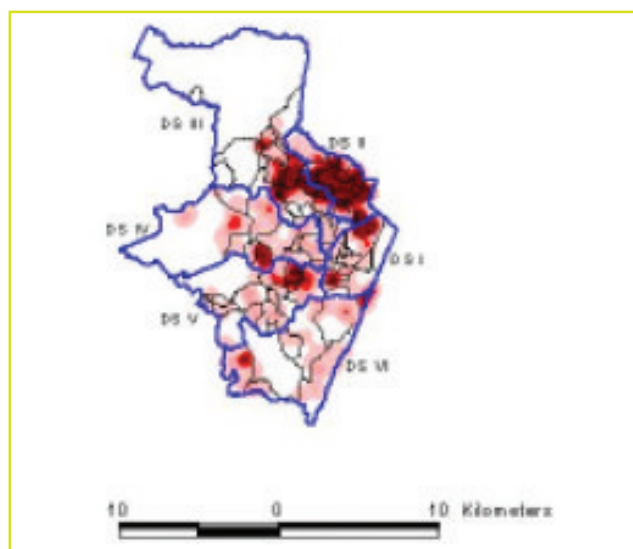


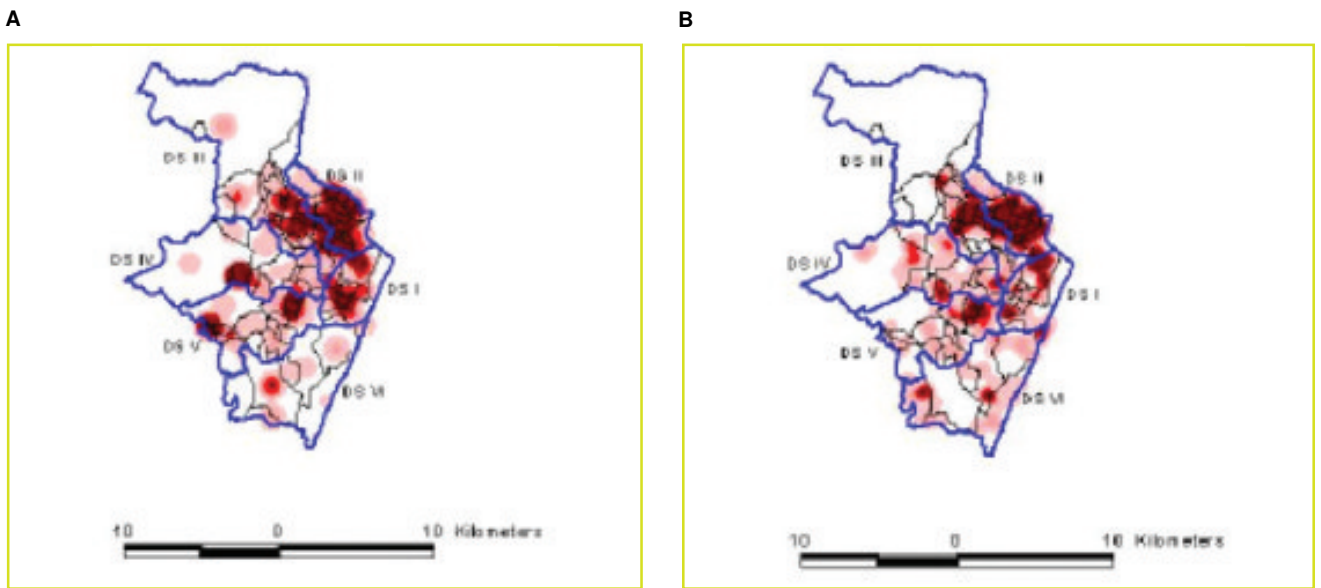
FIGURA 3.20 – Mapa de Kernel de casos de tuberculose segundo setores censitários (polígonos não representados).



Por fim, nas figuras 3.21 A e 3.21 B, comparamos os mapas de intensidade de incidência segundo bairros e segundo setores censitários, respectivamente. Pode-se notar, novamente, a menos da região noroeste da cidade, diferenças importantes entre os dois mapas, decorrentes da questão da heterogeneidade da distribuição espacial do evento, dentro de uma unidade geográfica de análise de maior tamanho.

Quando adotamos o setor censitário como unidade de análise, estamos tentando melhor apreender essa heterogeneidade, identificando os subgrupos populacionais submetidos a um maior risco de adoecer, embora pagando o preço da instabilidade de taxas calculadas para pequenas áreas. Essa é uma questão central quando estamos diante da decisão sobre que escala adotar.

FIGURA 3.21 – Mapa de Kernel de incidência por bairros (A) versus incidência por setor (B).



Deve-se ainda salientar mais uma vez que a abordagem aqui apresentada pode resultar em análises enviesadas, dado que atribuímos ao centróide geográfico do polígono todas as ocorrências de interesse. O centro geográfico da área nem sempre coincide com o centróide populacional, ponto este ao qual estaria mais correto atribuírmos as ocorrências. Em face dessa questão, a adoção do setor censitário como unidade de análise tem a vantagem de introduzir vieses menores, em grande parte de nossas áreas urbanas, visto que as diferenças de localização entre esses dois centróides são minimizadas devido ao menor tamanho geográfico dos setores.



Anexos

EXERCÍCIO 1: Vigilância e atenção à saúde de gestantes HIV positivas: construção de indicadores com operações entre camadas.

Autores: Lisiane M. W. Acosta, Maria Regina V. Brito, Eugênio P. Lisboa, Rui Flores e Christovam Barcellos

EXERCÍCIO 2: Distribuição dos acidentes de trânsito em Belo Horizonte, MG.

Autora: Maria Cristina Mattos

EXERCÍCIO 3: Análise espacial da distribuição de casos de malária no Estado de Mato Grosso em 2000 e 2003.

Autores: Marina Atanaka dos Santos e Reinaldo Souza-Santos

Anexos:

EXERCÍCIOS

Exercício 1: Vigilância e atenção à saúde de gestantes HIV positivas: construção de indicadores com operações entre camadas.

pág. 85

Exercício 2: Distribuição dos acidentes de trânsito em Belo Horizonte, MG.

pág. 97

Exercício 3: Análise espacial da distribuição de casos de malária no Estado de Mato Grosso em 2000 e 2003.

pág. 105

Exercício 1 : Vigilância e atenção à saúde de gestantes HIV positivas: construção de indicadores com operações entre camadas

I - Introdução

A possibilidade concreta dos serviços de saúde atuarem na prevenção da transmissão vertical do vírus HIV é um dos mais importantes avanços no controle da epidemia da Aids. O grande marco para essas ações foi o estudo desenvolvido pelo AIDS Clinical Trial Group (ACTG-076) que, em 1994, comprovou as estratégias que decrescem a transmissão vertical chegando a níveis de 1 a 2 % enquanto que, sem qualquer intervenção, seriam de 25 a 30 %. No Brasil, esses conhecimentos já foram incorporados às políticas públicas. O Programa Nacional de DST/AIDS do Ministério da Saúde (MS) priorizou o controle da transmissão vertical incidindo na rede de saúde por meio de capacitações, fornecimento de insumos e estabelecendo o agravo “gestante HIV e criança exposta” como de notificação compulsória em todo o território nacional no ano de 2000.

Esse esforço de minimizar a transmissão vertical do HIV pressupõe a articulação entre ações de atenção básica à saúde e a vigilância da saúde com base no princípio de territorialização, em áreas de responsabilidade dos serviços de saúde. A territorialização dos serviços de saúde é uma premissa básica do Sistema Único de Saúde (SUS). Os territórios das áreas de atuação das unidades básicas de saúde (UBS), como agregados espaciais, podem ser avaliados através de indicadores sociodemográficos, epidemiológicos, de recursos, de cobertura, e outros que contribuem para análise de situação de saúde e permitem o monitoramento e avaliação de ações de saúde.

O Brasil possui vários sistemas de informação e se destaca internacionalmente pela transparência dos mesmos, em especial pelos dados do DATASUS, que disponibiliza várias informações em saúde. Entre todos os sistemas de informação, serão utilizados para este exercício os que auxiliam para análise espacial da incidência de nascidos vivos expostos ao vírus HIV pela gestação/parto e associação com indicadores de assistência à saúde e características populacionais.

Utilizando-se os dados do censo por setores censitários do município, que compõem os territórios das unidades de saúde, fornecidos pelo Instituto Brasileiro de Geografia e Estatística (IBGE), tem-se os dados populacionais.

O Sistema Nacional de Nascidos Vivos (SINASC) tem condições de informar o número de nascidos vivos, além de outras variáveis como: escolaridade da mãe, tipo de parto, realização ou não de pré-natal, peso no nascimento e outros.

O Sistema de Informação dos Agravos de Notificação (SINAN) possui os dados de morbidade de agravos de notificação que podem ser georreferenciados pelos endereços de residência para obterem-se os casos por unidades de saúde. No caso de vigilância epidemiológica de gestantes HIV positivas e crianças expostas, tem-se escolaridade materna, período de detecção do HIV na mãe, realização de pré-natal, número de consultas realizadas, uso de anti-retrovirais na gestação, parto da criança, aleitamento materno, acompanhamento da criança e se a mesma se tornou infectada pelo HIV ou não.

II - Exercício

Considerando que o objetivo da vigilância epidemiológica não é somente a coleta de dados, mas também a sua análise e a avaliação das políticas assistenciais e criação de novas estratégias de ação, vamos neste exercício determinar:

- A incidência de gestantes HIV positivas e crianças expostas das áreas de atuação dos serviços de atenção básica à saúde.
- Alguns indicadores sociodemográficos e de atenção à saúde nessas áreas.
- A correlação entre incidência de HIV com indicadores sociodemográficos e de atenção à saúde.

Serão utilizados neste exercício os seguintes dados:


NOME DOS ARQUIVOS	DADO GRÁFICO	ATRIBUTOS	FONTE DE INFORMAÇÃO
Area_US_region	Área de atuação das unidades básicas de saúde (UBS) (polígonos)	Código da área	Secretaria Municipal de Saúde
Gest_HIV_point	Local de residência de gestantes HIV positivas (ponto)	Número de registro	SINAN (SUS)
Nascidos_point	Local de residência dos nascidos (ponto)	Número da DN, local de nascimento e número de consultas pré-natal	SINASC (SUS)
SC_2000_region	Setor censitário de 2000 (polígonos)	Código do setor, tipo de setor, população total e número de domicílios	Censo Demográfico (IBGE)

Observação: Para que as pessoas e as áreas não sejam identificadas, foram excluídos dos bancos de dados todos os campos de identificação. Também foram alterados o sistema de projeção e a localização exata dos pontos que pudessem ser usados para localizar pessoas e serviços. Manter o sigilo das informações deve ser uma preocupação de todos os que trabalham com bancos de dados de saúde.

A visualização e análise do banco de dados deve ser feita utilizando-se o aplicativo geográfico TerraView. Portanto, é necessário instalar a versão mais recente desse aplicativo. A aquisição do arquivo de instalação, do manual e do tutorial do TerraView pode ser feita na página (www.dpi.inpe.br/terraview) da Divisão de Processamento de Imagens do Instituto Nacional de Pesquisas Espaciais – DPI/INPE.

O exercício tem os seguintes passos:

1. Inicialmente, é necessário **executar** (abrir) o aplicativo TerraView. Esse procedimento pode ser feito clicando em **Iniciar > Programas > TerraView3x > TerraView3x** ou através do ícone de atalho **TerraView 3x**, na área de trabalho.

2. Em seguida, a tela inicial do TerraView será apresentada. Crie um **novo banco de dados**, clicando no ícone . Opte por criar (em vez de conectar), escolha um **diretório** (sugerimos c:\curso\exerc1) e um nome para o **banco de dados** (por exemplo, HIV_US).

3. Abrindo as camadas de dados: No **Menu Arquivo > Importar Dados:**

a) em **Arquivo**, localize e selecione os arquivos de mapas (**Area_US_region; Gest_HIV_point, Nascidos_point e SC_2000_region**);

b) para selecionar a **chave de ligação** do mapa com os dados, clique **Selecionar Coluna** e selecione a coluna adequada;

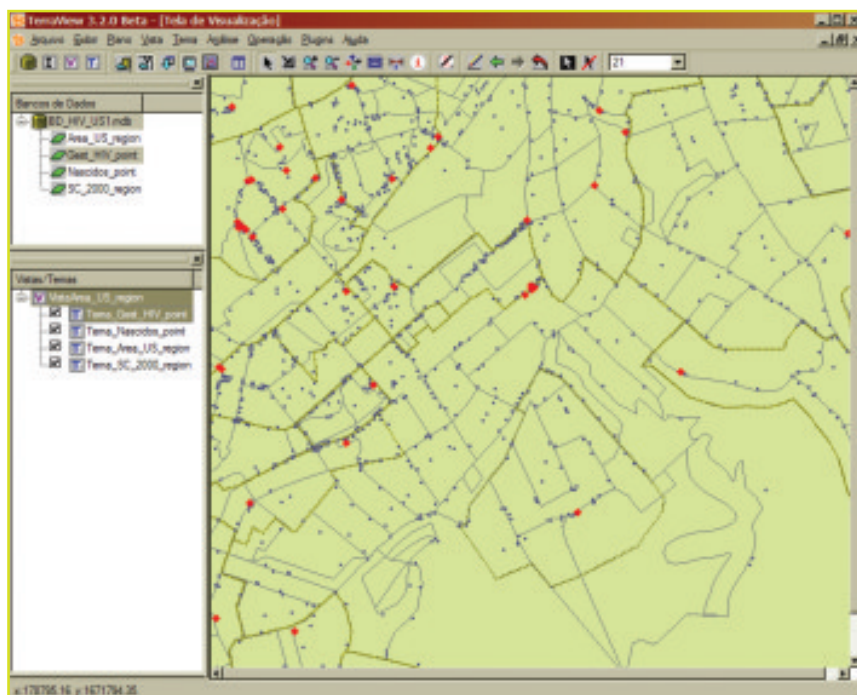
c) clique **OK** para concluir a importação. Repita essa operação para cada arquivo de mapa.

4. Cada camada (ou tema) será disposta em uma vista. Coloque cada camada (ou tema) na mesma vista clicando e arrastando os temas para cima. Escolha

os melhores **padrões de visualização** para cada camada (ou tema). Isso deve ser feito clicando com o botão direito do *mouse* em **Visual-Default**. Como sugestão, deixe a camada de áreas de atuação (**Area_US_region**) em cima de setores censitários (**SC_2000_region**) e ligeiramente transparente. Coloque as camadas de pontos (**Gest_HIV_point** e **Nascidos_point**) em cima das de polígonos com padrões diferentes para que se visualize os dois tipos de eventos (notificação de gestante HIV positiva e nascido vivo).

5. Na **barra de ferramentas**, pressione o botão **Desenhar**. O resultado é mostrado na figura 1.1 ao lado.

FIGURA 1.1 – Visualização dos dados de interesse para análise da transmissão vertical de HIV (área de atuação das unidades básicas de saúde, local de residência de gestantes HIV positivas, local de residência dos nascidos vivos, setor censitário de 2000).



Observe a superposição entre estas camadas. Existem grandes áreas com pequenas quantidades de pontos de nascimentos, provavelmente indicando que é uma área de baixa densidade demográfica. Os limites dos setores censitários não coincidem com os limites de áreas de atuação. Que outros padrões você consegue identificar através desse mapa?

Para análise, foram selecionados os seguintes indicadores:

- **Proporção de setores de favelas:** razão entre o número de setores censitários caracterizados como “sub-normais” durante o censo de 2000 e o número total de setores censitários das áreas das UBS.

- **Prevalência de HIV em Nascidos Vivos:** razão entre o número de nascidos vivos expostos ao HIV pela gestação e parto (dados do SINAN) e o número de nascidos vivos (dados do SINASC) nas áreas das UBS.

- **Taxa bruta de natalidade:** razão entre o número de nascidos vivos (SINASC) e a população total nas áreas das UBS (Censo demográfico).

- **Proporção de nascidos vivos sem pré-natal:** razão entre o número de nascidos vivos sem nenhuma consulta de pré-natal e o total de nascidos vivos nas áreas das UBS (SINASC).

O grande desafio que temos para calcular esses indicadores é que eles se referem a unidades espaciais dispares, isto é, não coincidentes. Devemos transferir os dados de uma camada para outra, no nosso caso os polígonos de áreas de abrangência.

Para obter esses indicadores, será necessário calcular:

- O número de setores censitários do tipo “sub-normal” (tipo 1 segundo os dados do censo) existentes na área de abrangência da unidade de saúde;

- O número total de setores censitários existentes na área de abrangência da unidade de saúde;

- A população total dos setores censitários contidos na área de abrangência da unidade de saúde;

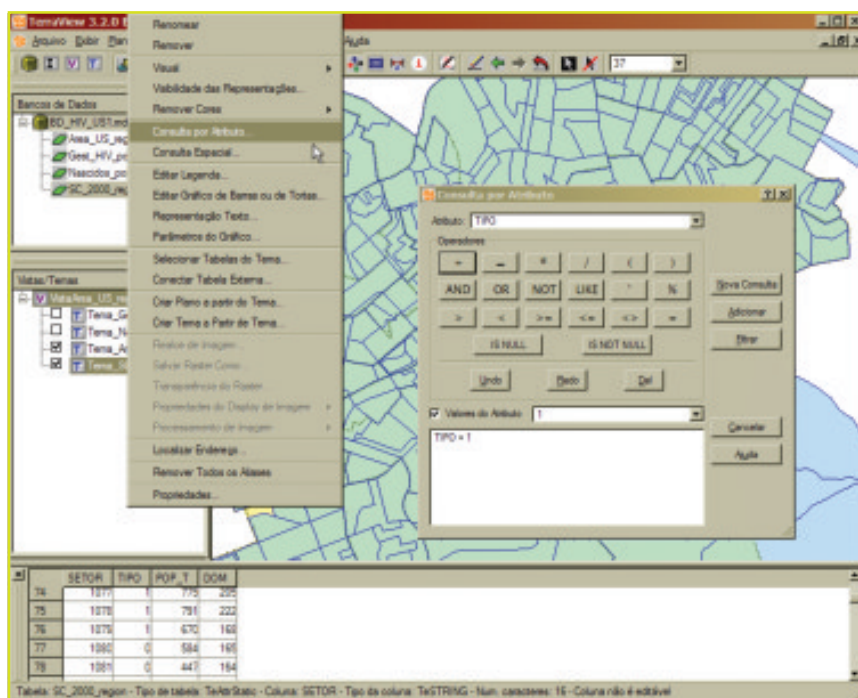
- O número total de nascidos vivos ocorridos na área de abrangência da unidade de saúde;

- O número total de gestantes HIV positivas existentes na área de abrangência da unidade de saúde;

- O número de nascidos vivos sem exames pré-natal ocorridos na área de abrangência da unidade de saúde.

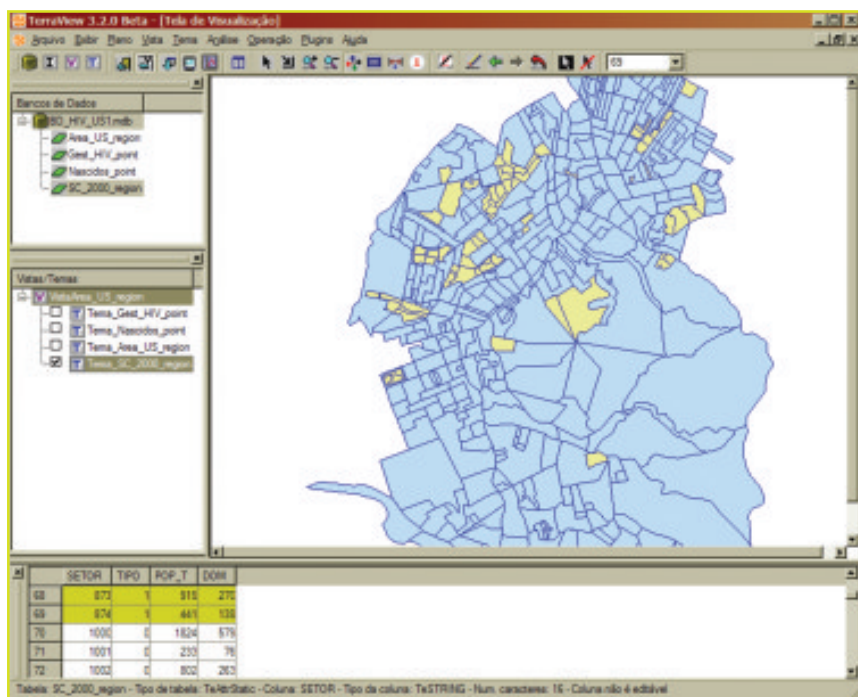
6. Para identificar os setores censitários do tipo “sub-normal” devemos realizar uma **Consulta por Atributos**. Clique com o botão direito do *mouse* no tema **SC_2000_region**.

FIGURA 1.2 – Seleção de setores censitários tipo 1 (sub-normal). Consulta por atributos.



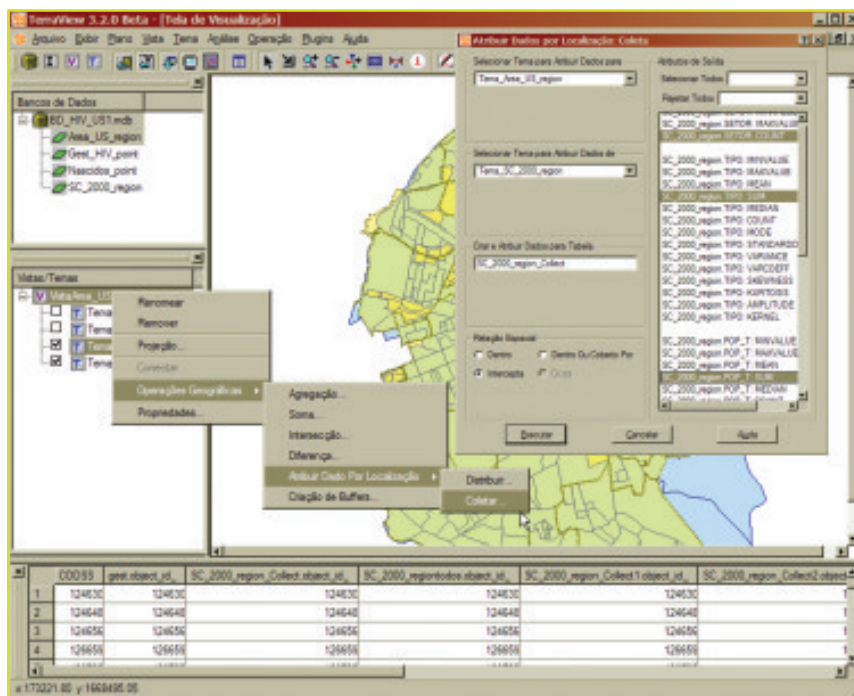
7. Agora selecione o atributo **Tipo** e coloque como valor do atributo **1**, ou escreva a expressão **Tipo = 1**. Clique em **Nova consulta**. Ficarão marcados os setores do tipo “sub-normal” tanto no mapa como na tabela. Observe que estão marcados 69 dos 431 setores.

FIGURA 1.3 – Resultado da seleção de setores censitários tipo 1.



8. O próximo passo é **adicionar essa informação às áreas de abrangência de unidades de saúde**. Para isso faremos uma **operação geográfica** chamada de **atribuir dado por localização** (coletar dados). Clique com o botão direito do *mouse* sobre a **vista**, que contém os temas **Area_US_region** e **SC_2000_region**.

FIGURA 1.4 – Contagem de setores censitários, total e tipo 1 e soma das populações. Atribuir dados por localização.



9. Selecione como **tema para atribuir dados para Area_US_region** e **tema para atribuir dados de SC_2000_region**, isto é, os setores censitários. Como **atributo de saída** precisamos somente das variáveis **SC_2000_region.SETOR: COUNT**, isto é, o número total de setores censitários (de todos os tipos), a **variável SC_2000_region.TIPO: SUM**, isto é, a soma dos setores censitários do tipo 1 (os setores censitários do tipo 0 não serão contados) e a variável **SC_2000_region.POP_T: SUM**, isto é, a soma da população total dos setores censitários. Selecionamos a **relação espacial intercepta**, quer dizer, usando todos os setores que estão totalmente ou parcialmente contidos dentro dos limites das áreas de abrangência. O programa adiciona três novas colunas com o valor da quantidade de setores censitários total e do tipo 1, e a população total nos polígonos de área de abrangência da unidade de saúde.

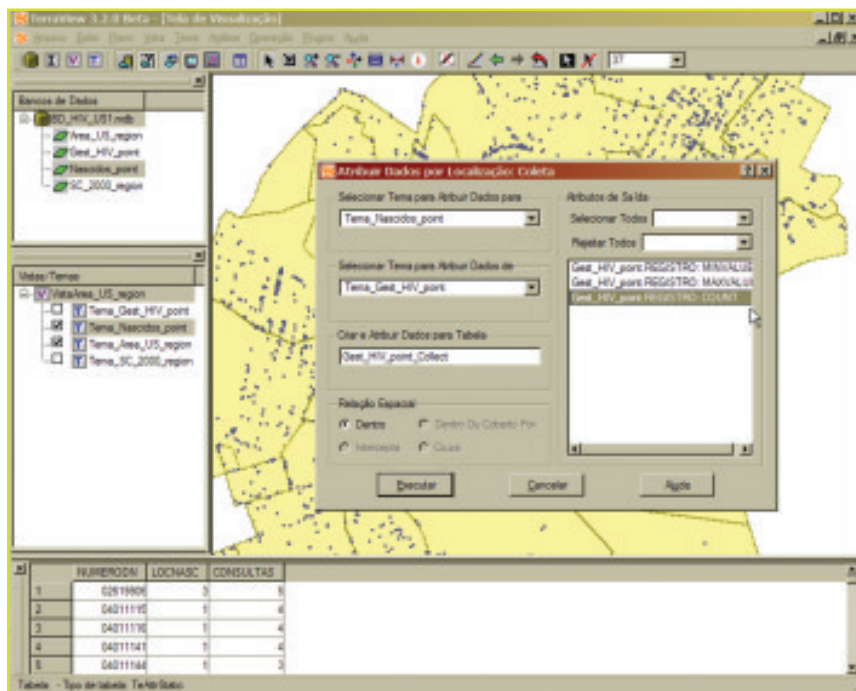
10. Para facilitar o trabalho de **operações entre variáveis**, podemos **mudar o nome das novas variáveis**. Isso pode ser feito clicando-se com o botão direito do *mouse* sobre a **coluna**. Clique na opção **modificar nome da coluna**.

Mude os nomes das variáveis para:

NOME GERADO PELO PROGRAMA	NOVO NOME
SC_2000_region.SETOR: COUNT	N_setores
SC_2000_region.TIPO: SUM	Setores_favela
SC_2000_region.POP_T: SUM	Pop2000

11. Agora podemos repetir esse procedimento para **coletar dados** do tema **Nascidos_Point**. O próximo passo é adicionar essa informação às áreas de abrangência de unidades de saúde. Para isso repetimos a **operação geográfica de atribuir dado por localização** (coletar dados). Selecione como **tema para atribuir dados para Area_US_region** e **tema para atribuir dados de Nascidos_Point**, isto é, os nascidos vivos. Como **atributo de saída** selecione a variável **Nascidos_Point.NUMERODN: COUNT**, isto é, o número total de nascidos. Selecionamos a **relação espacial dentro**, a única possibilidade de relação, já que se trata de uma coleção de pontos dentro de um conjunto de polígonos.

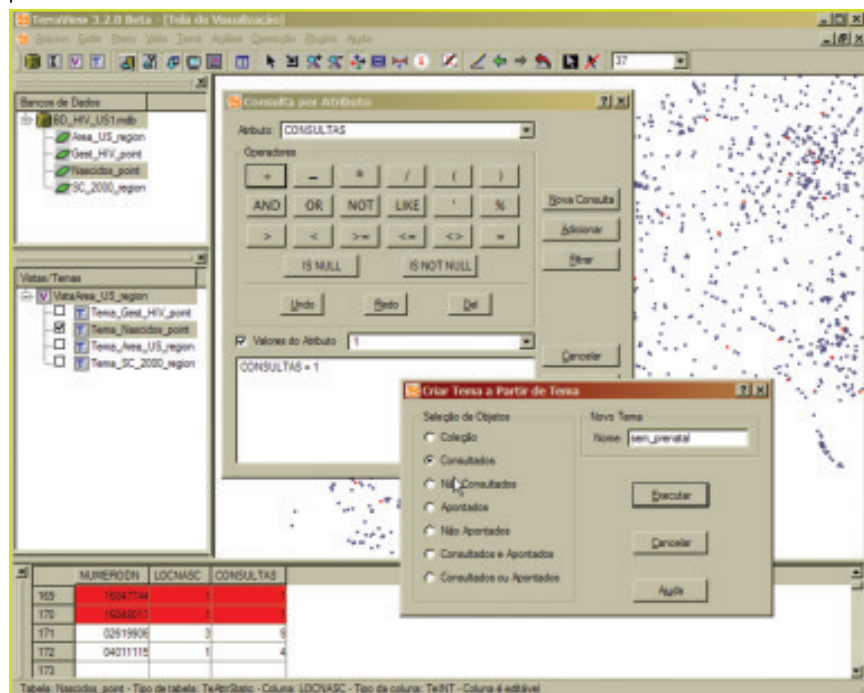
FIGURA 1.5 – Contagem de nascidos vivos em áreas de abrangência. Atribuir dados por localização.



O programa adiciona uma **nova coluna** com o valor da quantidade de nascidos vivos nos polígonos de área de abrangência da unidade de saúde. **Mude o nome** dessa nova variável para **N_Nascidos**.

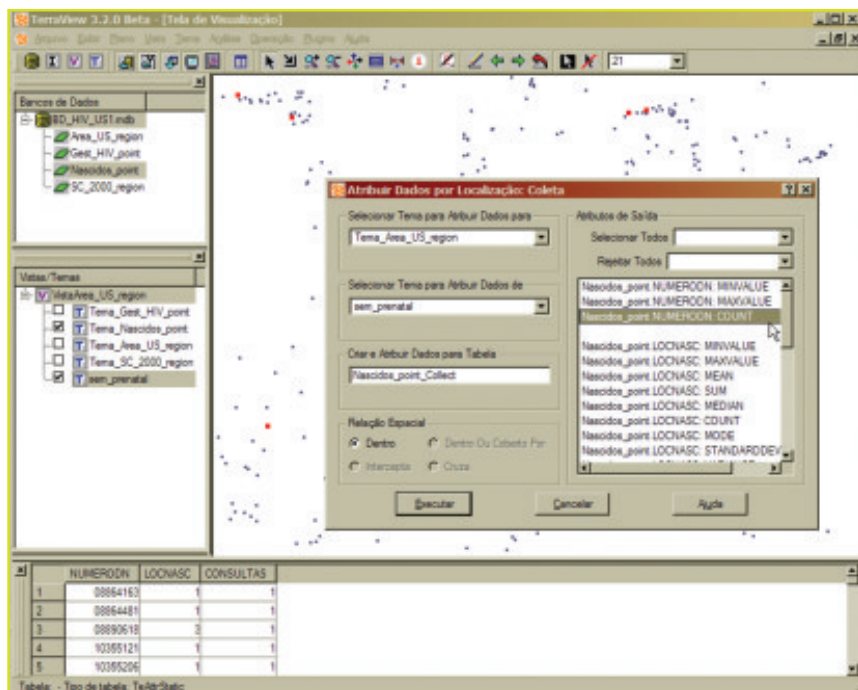
12. Para se obter o número de nascidos vivos sem exames pré-natal devemos selecionar o conjunto de nascidos com atributo consultas categoria 1. Para isso, devemos realizar uma **Consulta por Atributos** e clicar com o botão direito do *mouse* no **tema Nascidos_Point**. Agora selecione o atributo **Consultas** e coloque como valor do atributo **1**, ou escreva a expressão **Consultas = 1**. Clique em **Filtrar**. Ficarão marcados os nascidos sem pré-natal tanto no mapa quanto na tabela. Dos 4.064 nascidos, 170 foram marcados. Esta é uma seleção do conjunto de nascidos que deve ser salva como um novo tema (ou camada). Para isso, clique no **tema Nascidos_Point** com o botão direito do *mouse* e selecione **criar tema a partir de tema**. Aparecerá a janela seguinte em que você deve dar um nome para esse **novo tema** (**sem_prenatal**, por exemplo) e **selecionar somente os objetos consultados**.

FIGURA 1.6 – Criação de um novo tema por consulta. Seleção de nascidos vivos sem pré-natal.



Novamente faremos a **operação geográfica de atribuir dado por localização** (coletar dados). Selecione como **tema para atribuir dados para Area_US_region** e **tema para atribuir dados de Sem_prenatal**, isto é, os nascidos vivos sem consultas pré-natal. Como atributo de saída selecione a variável **Nascidos_Point.NUMERODN: COUNT**.

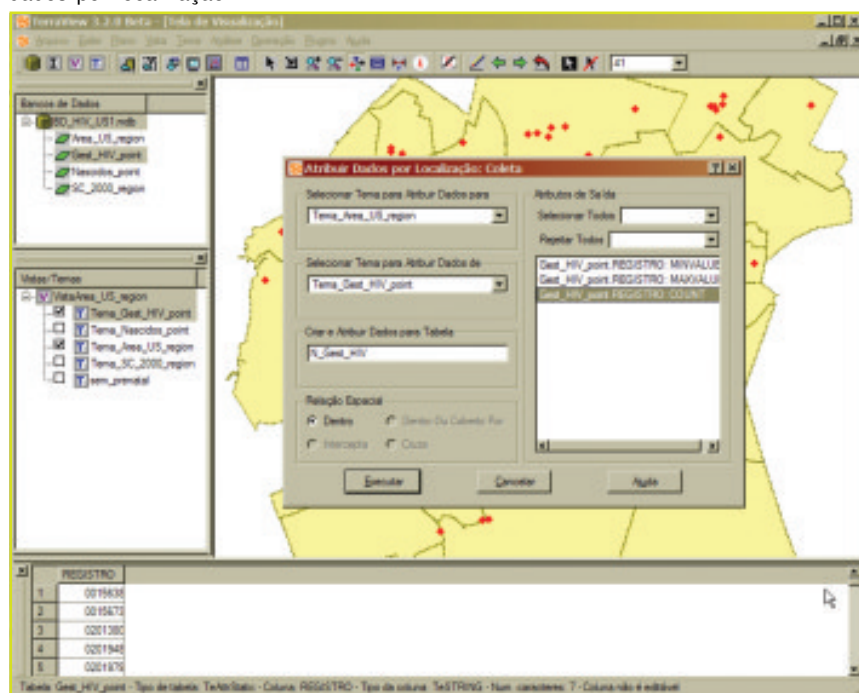
FIGURA 1.7 – Contagem de nascidos vivos sem pré-natal nas áreas de abrangência.



O programa adiciona mais uma coluna com o valor da quantidade de nascidos que não tiveram exames de pré-natal presentes nos polígonos de área de abrangência da unidade de saúde. Mude o nome dessa nova variável para ***N_Nascidos_sempnata***.

13. A última variável necessária para construir os indicadores é o número de gestantes HIV positivas existentes na área de abrangência da unidade de saúde. Para isso, fazemos a **operação geográfica de atribuir dado por localização** (coletar dados). Selecione como **tema para atribuir dados para Area_US_region** e **tema para atribuir dados de Gest_HIV_point**. Como atributo de saída selecione a variável **Gest_HIV_point.REGISTRO:COUNT**. O programa adiciona mais uma coluna com o valor da quantidade de gestantes HIV positivas nos polígonos de área de abrangência da unidade de saúde. Mude o nome dessa nova variável para ***N_gest_HIVpos***.

FIGURA 1.8 – Contagem de gestantes HIV-positivas nas áreas de abrangência. Atribuir dados por localização.



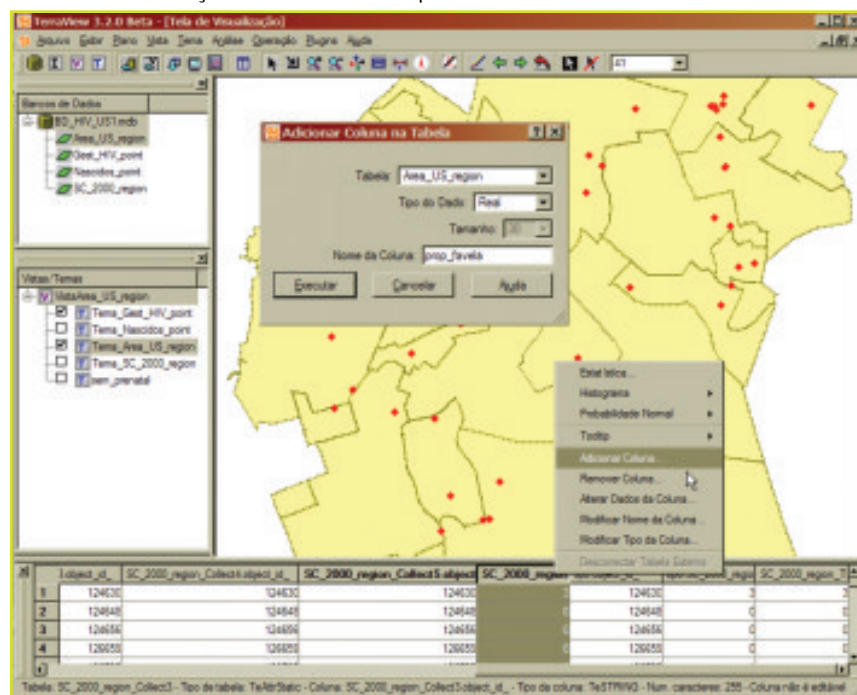
Agora temos todos os dados que precisamos para calcular os indicadores.

Recordando:

INDICADOR	COMPONENTES	CÁLCULO
Proporção de setores de favelas	Número de setores censitários “sub-normais” / número total de setores censitários.	$\text{Setores_favela} * 100 / \text{N_setores}$
Prevalência de HIV em Nascidos Vivos:	Número de gestantes HIV positivas / Número de nascidos vivos.	$\text{N_gest_HIVpos} * 100 / \text{N_Nascidos}$
Taxa bruta de natalidade	Número de nascidos vivos / População total.	$\text{N_Nascidos} * 100 / \text{Pop2000}$
Proporção de nascidos vivos sem pré-natal	Número de nascidos vivos sem pré-natal / Total de nascidos vivos.	$\text{N_Nascidos_semprnatal} * 100 / \text{N_Nascidos}$

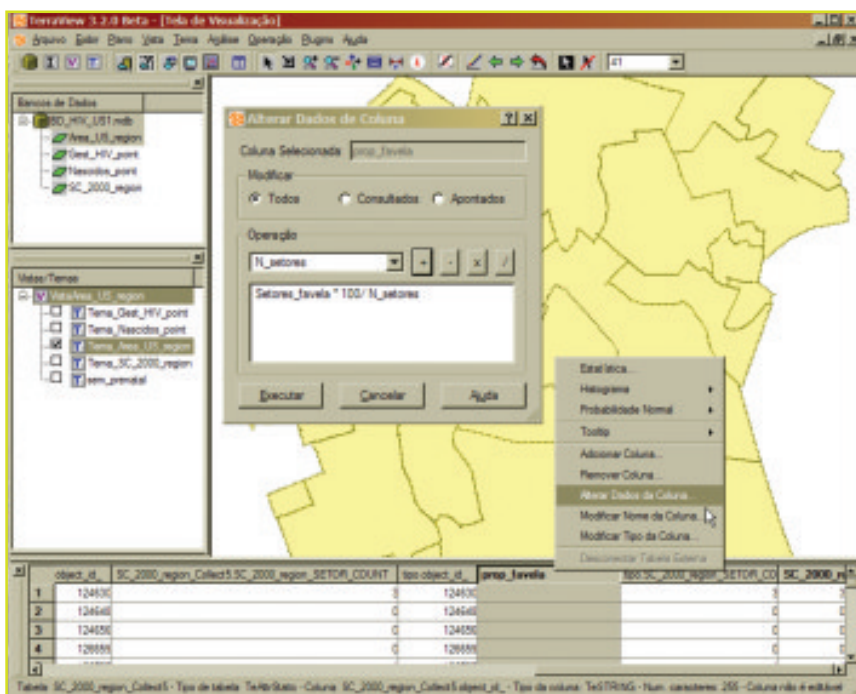
Esses cálculos são realizados criando-se nova coluna na tabela **Área_US_region**. Na tabela, clique com o botão direito do *mouse* em **adicionar coluna** e dê um nome para a coluna do indicador **proporção de favelas** (por exemplo, **prop_favela**).

FIGURA 1.9 – Criação de novas colunas para cálculo de indicadores. Adicionar coluna.



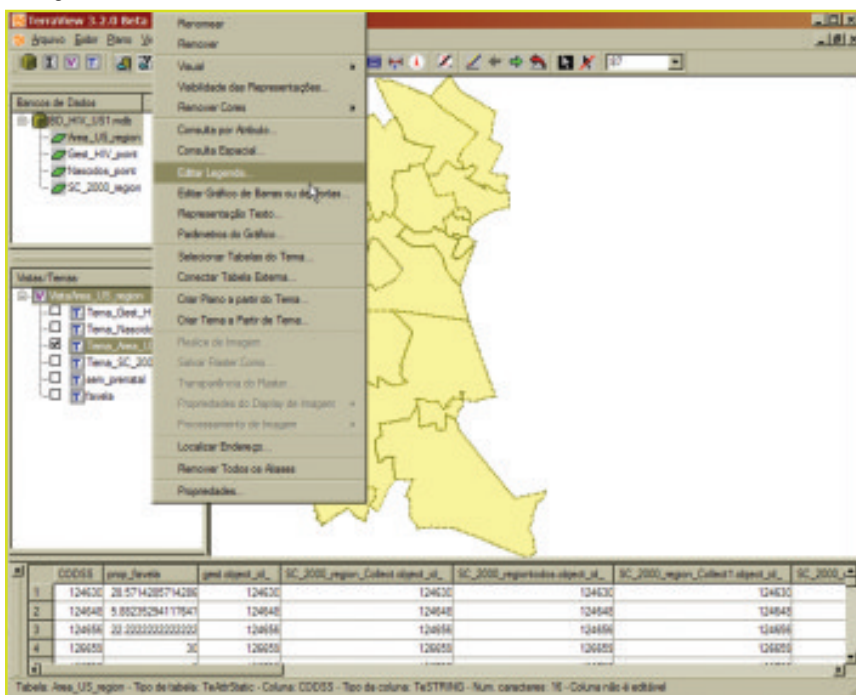
14. Sobre essa coluna nova clique com o botão direito do *mouse* em **alterar dados da coluna** e escreva (com ajuda dos botões disponíveis na janela) a expressão **Setores_favela / N_setores**. Faça a mesma operação para calcular os demais indicadores (adicione colunas e altere os dados dessa coluna).

FIGURA 1.10 – Cálculo de indicadores usando operações entre duas colunas existentes na tabela.



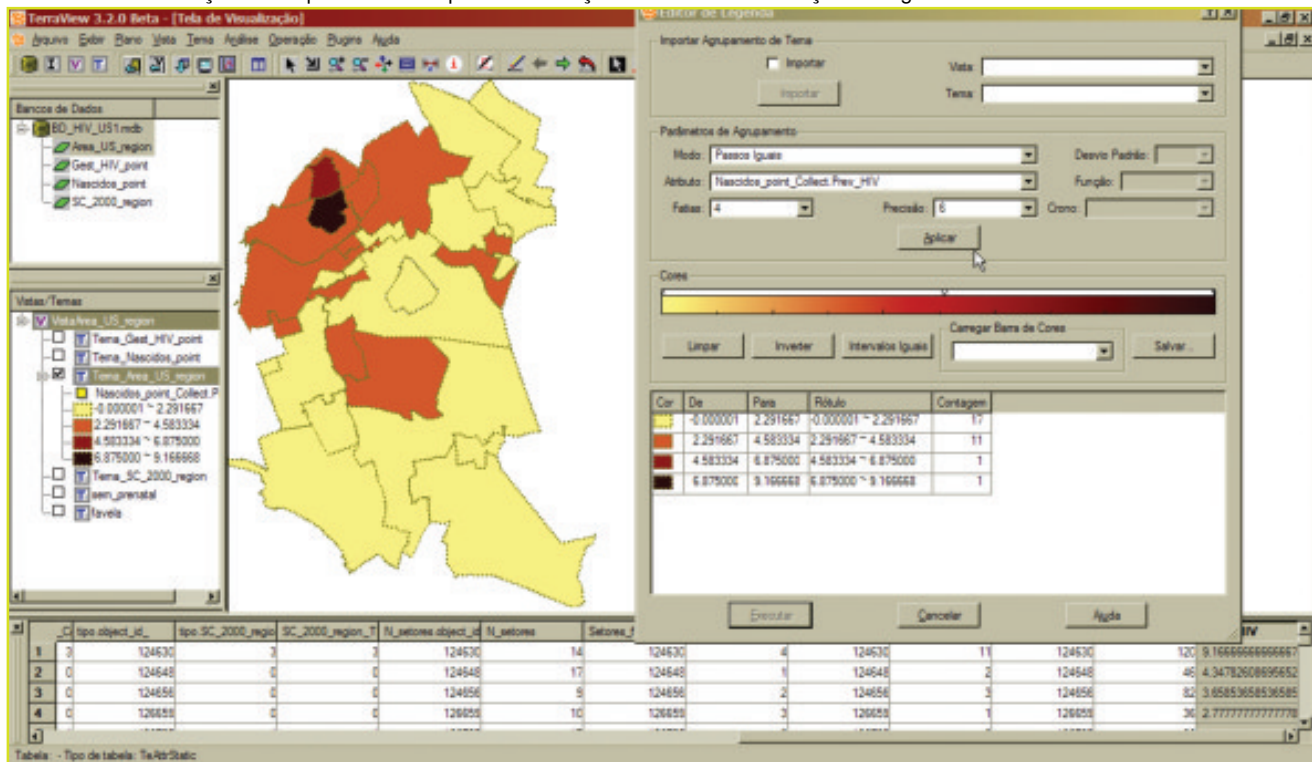
15. A primeira fase de análise desse conjunto de indicadores deve ser a **visualização**, isto é, criar mapas temáticos que permitam analisar a distribuição desses indicadores. Sobre o tema **Area_US_region** clique em **editar legenda**.

FIGURA 1.11 – Criação de mapas temáticos para visualização de indicadores. Edição de legenda.



Obtemos o seguinte mapa:

FIGURA 1.12 – Criação de mapas temáticos para visualização de indicadores. Edição de legenda.



Explore visualmente os indicadores obtidos e identifique áreas que devem ser priorizadas para ações de atenção e prevenção à saúde. Para **análise da associação estatística** entre os indicadores pode ser utilizada a **correlação não-paramétrica** e calculado o **coeficiente de Spearman** para avaliar a associação entre pares de variáveis. Num programa de estatística, pode ser usado um modelo de regressão múltipla que relacione a prevalência de HIV em gestantes como variável dependente, com os demais indicadores como variáveis independentes. Isso pode ajudar a explicar que determinantes estão atuando sobre os riscos de transmissão vertical do HIV.

O aspecto mais importante deste exercício, no entanto, é a possibilidade completamente nova de se criar indicadores através de operações entre camadas em um ambiente de SIG. Esse tipo de cálculo é impossível de ser obtido por meio de bancos de dados convencionais.

Exercício 2 : Distribuição dos acidentes de trânsito em Belo Horizonte, MG.

I - Introdução

Os acidentes de trânsito são atualmente um dos agravos de maior impacto na saúde da população brasileira, tanto pelos óbitos quanto pelas internações e seqüelas deixadas em suas vítimas. Ocupam a segunda posição entre as causas externas de mortalidade no país.

As análises dos acidentes de trânsito devem servir para orientar medidas de prevenção e controle desses agravos. Para o município, a localização pontual de diferentes eventos relacionados a este tema (acidentes, pontos de apoio para as unidades móveis de atendimento pré-hospitalar, serviços de urgência etc.) constitui-se num valioso material para avaliação e planejamento de ações preventivas tanto para o serviço de saúde quanto para os órgãos de trânsito. Pode orientar medidas de controle do tráfego, indicar falha na sinalização de vias ou apontar melhores locais para os pontos de apoio das ambulâncias.

Uma fonte de informação para este propósito é o registro dos atendimentos aos acidentados pelo Serviço de Atendimento Móvel de Urgência (SAMU), que tem o objetivo de reduzir o número de óbitos, o tempo de internação em hospitais e as seqüelas decorrentes da falta de socorro precoce.

II – Exercício

Neste exercício, vamos analisar o padrão de distribuição dos acidentes de trânsito em Belo Horizonte, buscando identificar áreas críticas, ou seja, áreas de maior intensidade deste tipo de acidentes.

Conjunto de dados:

• **Atendimentos a acidentados no trânsito (atendimentos. *)**

Estes dados representam parte dos atendimentos pré-hospitalares aos acidentados no trânsito, realizado pelo Serviço de Atendimento Móvel de Urgência (SAMU) da Secretaria Municipal de Saúde de Belo Horizonte, ocorridos no período de janeiro a abril de 2006. Correspondem aos registros de atendimentos que levaram à liberação de uma unidade de atendimento móvel.

São registros de pacientes atendidos, deste modo, vários atendimentos podem corresponder a um mesmo acidente.

Em Belo Horizonte, os atendimentos do SAMU são georreferenciados ao endereço do local do atendimento. Quando os chamados são referentes a ocorrências em via pública, como é caso dos acidentes de trânsito, estes são referenciados ao endereço em frente ou mais próximo ao local do evento. Os atendimentos para os quais não foi possível fazer o georreferenciamento foram suprimidos deste exemplo.

Tipo de dado: pontos

• **Acidentes de trânsito (acidentes.*)**

Correspondem aos acidentes referentes aos atendimentos acima. É uma consolidação daqueles dados, contendo o número de vítimas atendidas em cada ocorrência. Seguem os mesmos critérios quanto ao seu georreferenciamento.

Tipo de dado: pontos

• **Pontos de apoio (ambulancias.*)**

Correspondem aos pontos de apoio das ambulâncias do SAMU, representadas pelas 15 Unidades de Suporte Básico (USB) e 03 Unidades de Suporte Avançado (USA).

Tipo de dado: pontos

• **Portas de entrada para atendimento às urgências (unid_urgencia.*)**

Representam as 11 principais portas de entrada da urgência do SUS-BH, tanto para traumas quanto para urgências clínicas.

Tipo de dado: pontos

• **Limite municipal (belohorizonte.*)**

Representa o limite do Município de Belo Horizonte.

Tipo de dado: polígono

• **Principais vias de trânsito (ppvias.*)**

Correspondem às vias que promovem a ligação entre diferentes bairros ou setores da cidade, onde a intensidade de veículos e a velocidade no trânsito são maiores.

Tipo de dado: linha.

■ **Sobre os dados:**

1) Estes dados se destinam a ilustrar os exercícios do Programa de Capacitação e Atualização em Abordagens do Espaço em Análises de Saúde Pública. São dados parciais, sujeitos a alteração.

2) Foram utilizados dados do Sistema de Informações Geográficas, gerenciado pela Empresa de Informática e Informação de Belo Horizonte (Prodabel). O georreferenciamento dos dados pontuais do foi feito por meio da camada de endereços.

3) Todos os temas camadas estão suprimidos em sete milhões de unidades no eixo das coordenadas Y.

O Banco de Dados no TerraView (*TransitoBH.mdb*) foi formado pelas camadas importadas a partir dos arquivos *shape*.

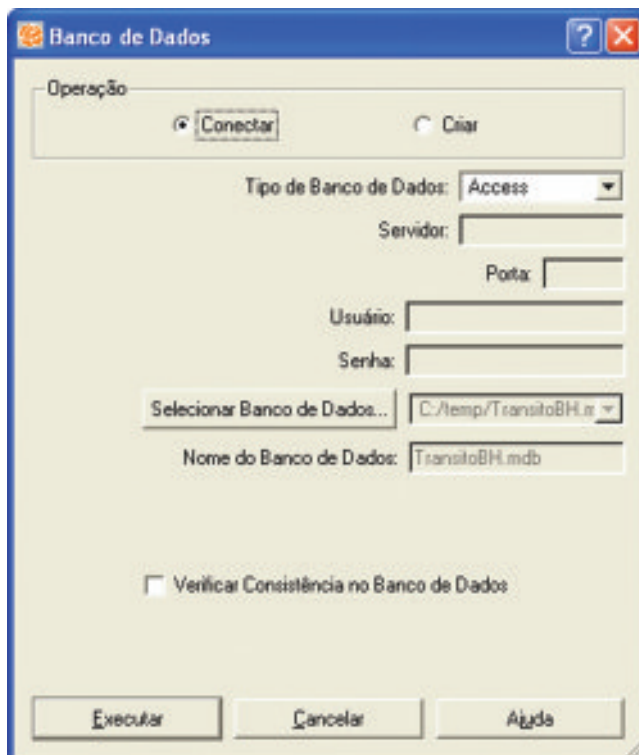
Problema:

Identificar locais de maior densidade de acidentes de trânsito no município a partir do registro de atendimentos do SAMU.

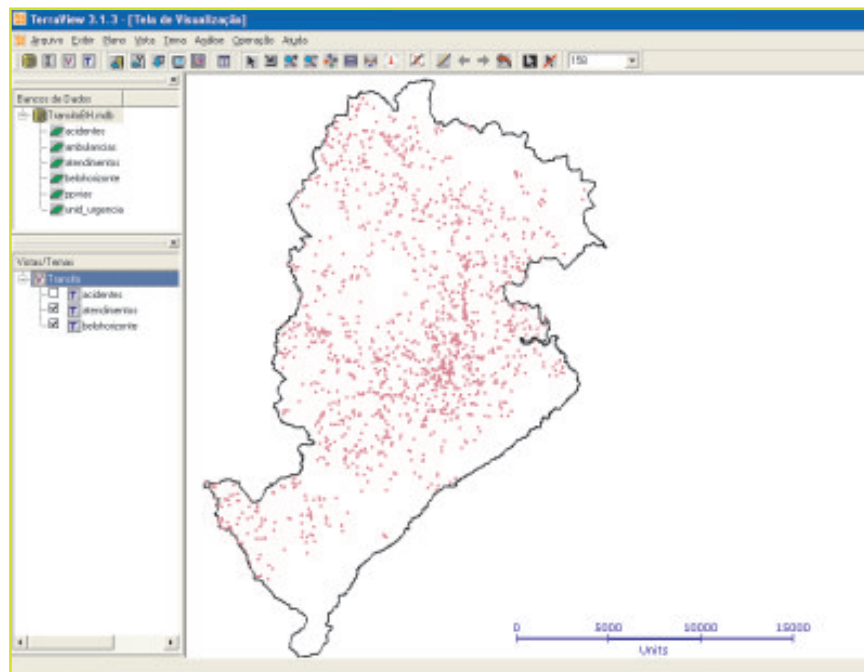
A identificação de locais onde haja uma aglomeração de acidentes pode ajudar na identificação de locais críticos, propiciando o direcionamento das medidas de controle de trânsito (sinalização, radares etc.) e orientando a locação de equipamentos móveis para atendimento aos acidentados.

Em primeiro lugar, deveremos conhecer os dados, avaliar sua completitude, cobertura e qualidade tanto dos dados tabulares quanto dos dados geográficos. Dependendo da maneira de obtenção das coordenadas dos dados pontuais, podem ocorrer falhas durante o processamento. Dados obtidos por GPS podem conter perdas devido a falhas na recepção de sinal dos satélites, não captando as coordenadas. Quando os dados são georreferenciados a uma base geográfica específica, como é o caso de nosso exemplo para os atendimentos pelo SAMU, alguns endereços podem não ser encontrados, caso a base esteja incompleta. Essas perdas deverão ser criticadas, pois podem estar concentradas em alguma região ou em um determinado tipo de acidente.

1. Abra o banco de dados *TransitoBH.mdb*, conheça os planos de informação. Verifique seus atributos.

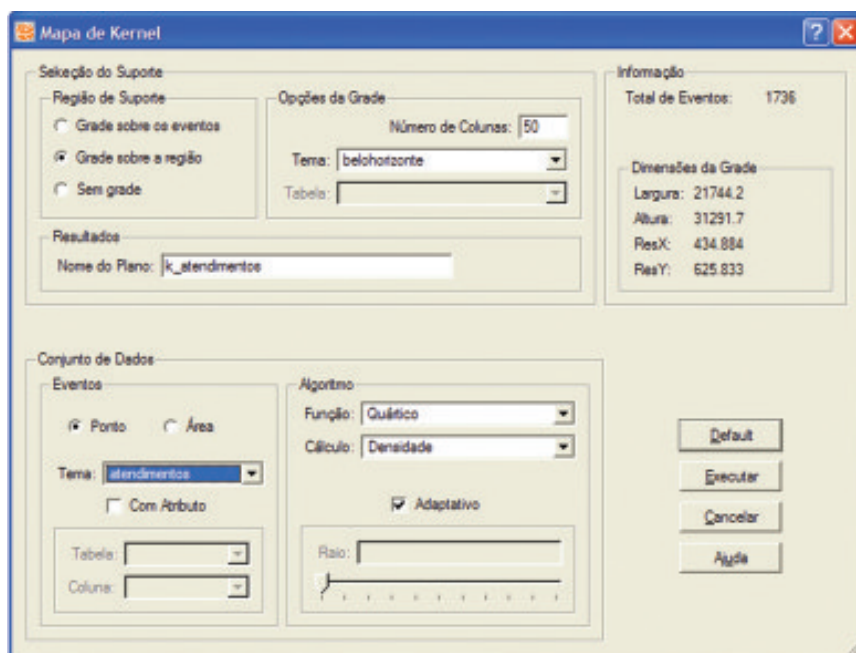


Os acidentes podem ser mapeados e mapas temáticos podem ser elaborados segundo característica da vítima ou do acidente. Nota-se a concentração de acidentes em algumas áreas, especialmente na parte central do município. Podemos verificar alguns pontos enfileirados, que estão localizados ao longo de algumas vias. Entretanto, deve-se ter cautela na interpretação desses mapas de pontos. Alguns acidentes que ocorram num mesmo local, por exemplo, em um cruzamento perigoso, podem registrar as mesmas coordenadas. Daí, um local com 10 eventos pode ser representado da mesma forma que um outro local onde haja apenas uma ocorrência. A simples visualização de pontos é subjetiva e pode não detectar padrões da distribuição dos eventos.



Vamos agora construir um mapa de Kernel para estimar a intensidade dos acidentes de trânsito.

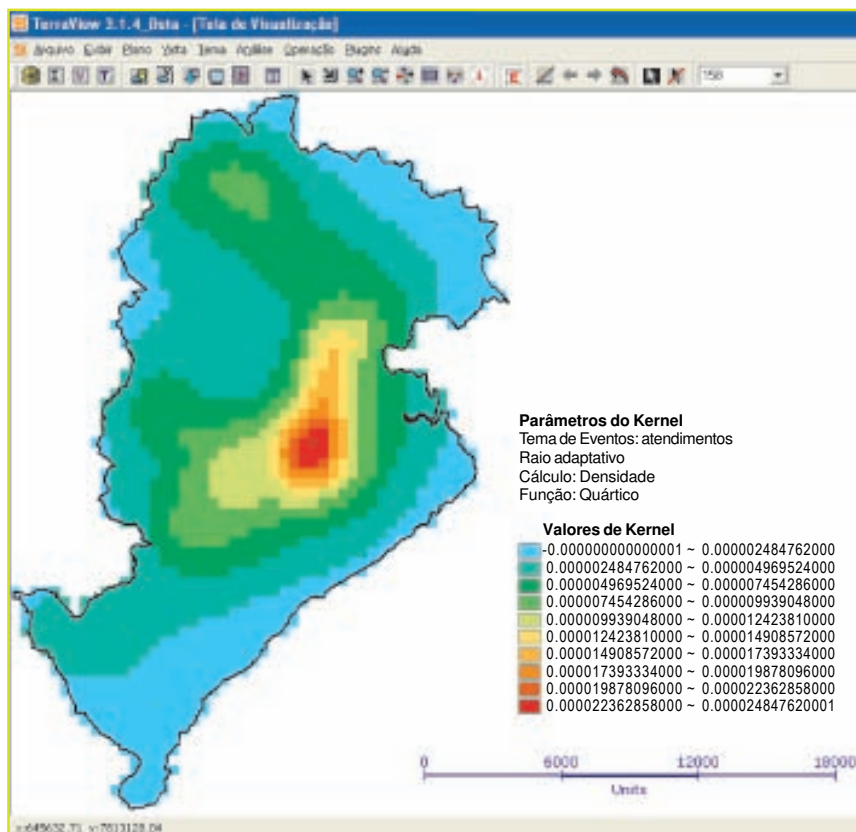
2. Ative o **tema** *atendimentos* na **vista** *Transito*.
3. Vá ao **menu** **Análise > Mapa de Kernel** para escolher os parâmetros a serem utilizados.
 - a) **Região de Suporte**: escolha *Grade sobre a região*
 - b) **Opção de Grade**: escolha *50 colunas* e **Tema** *belohorizonte*
 - c) Em **Resultados**: nomeie o plano (por exemplo *k_atendimentos*)
 - d) Em "**Conjunto de Dados**" marque o tipo de evento (ponto) e o tema a ser analisado (*atendimentos*). Não selecione a opção "Com Atributo".
 - e) Em **Algoritmo** selecione a **Função** "*quártica*" e **Cálculo** da densidade. Selecione a opção "*Adaptativo*".
4. **Execute**.

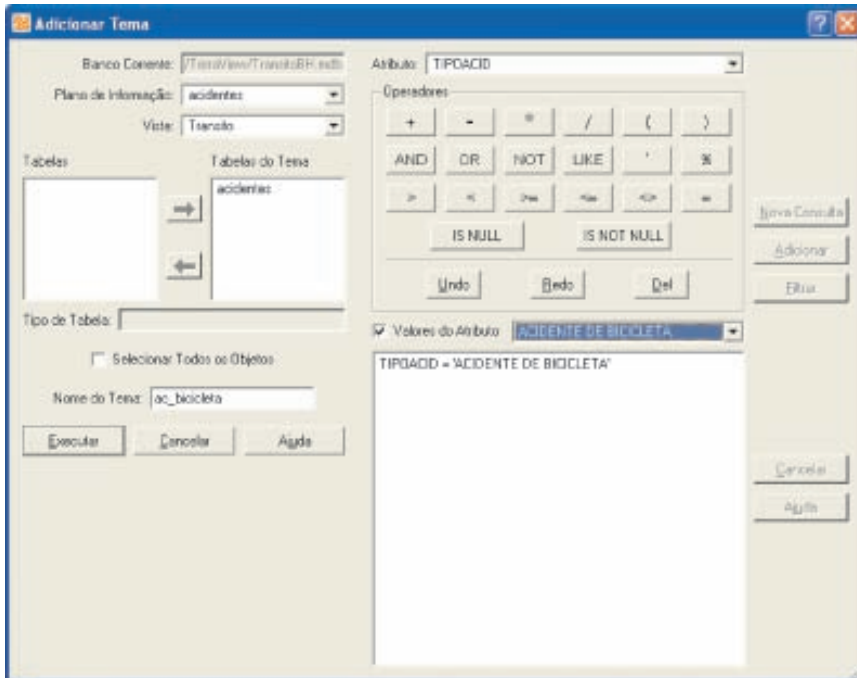


Será gerado um novo plano k_atendimentos. Observe as áreas de maior concentração.

A região central da cidade apresenta a área de maior intensidade quando consideramos os diversos tipos de acidentes (Figura 2.1).

FIGURA 2.1 – Superfície de densidade dos diversos tipos de acidentes.





Agora, vamos trabalhar apenas os acidentes com bicicleta.

Para isso vamos adicionar um novo tema a partir da seleção:

TIPOACID = 'ACIDENTE DE BICICLETA'

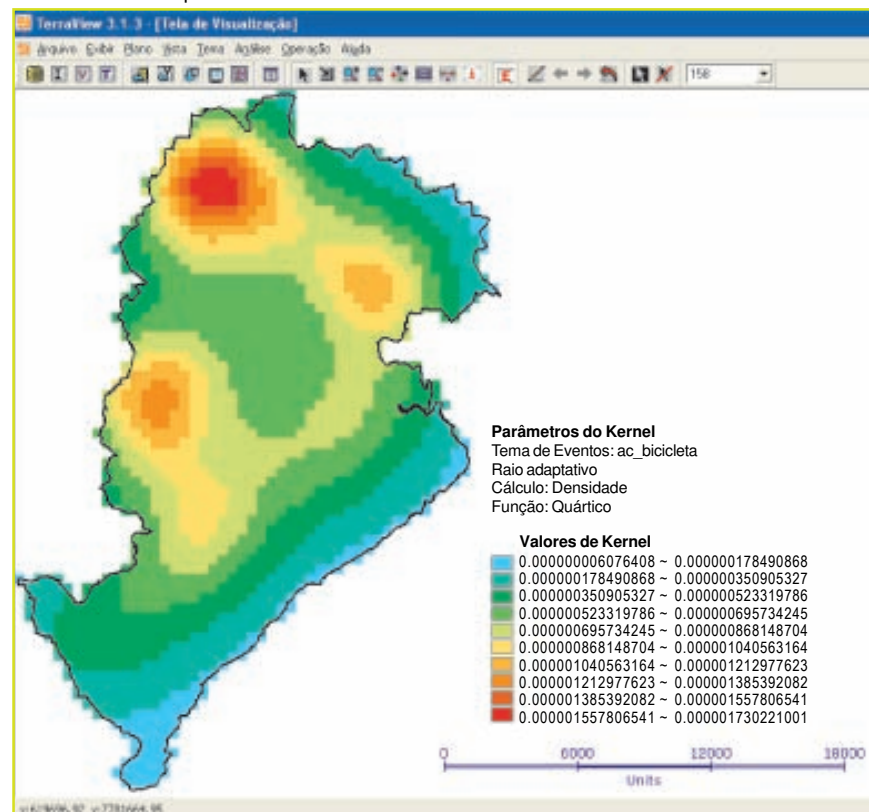
5. **Adicione** um novo tema *acidentes*, entretanto **desmarque** a opção **Selecionar Todos os Objetos**. Isso lhe permitirá fazer a seleção acima.

6. Nomeie este novo tema como *ac_bicicleta*, adicionando-o à **vista Transito**.

Agora você pode fazer um mapa de Kernel, apenas para esse tipo de acidente.

7. Repita os procedimentos do item 3 para os acidentes de bicicleta. Observe as áreas de maior concentração desse tipo de acidente (Figura 2.2).

FIGURA 2.2 – Superfície de densidade de acidentes de bicicleta.



A escolha do método e dos parâmetros da análise deve ser feita de acordo com os objetivos propostos. A detecção mais detalhada dos acidentes pode ser de grande valor para os órgãos gestores do trânsito e para os serviços de saúde do município.

Experimente novas análises variando o número de colunas (células). Por exemplo: 100 colunas. Quanto maior o número de colunas, maior a resolução do mapa gerado.

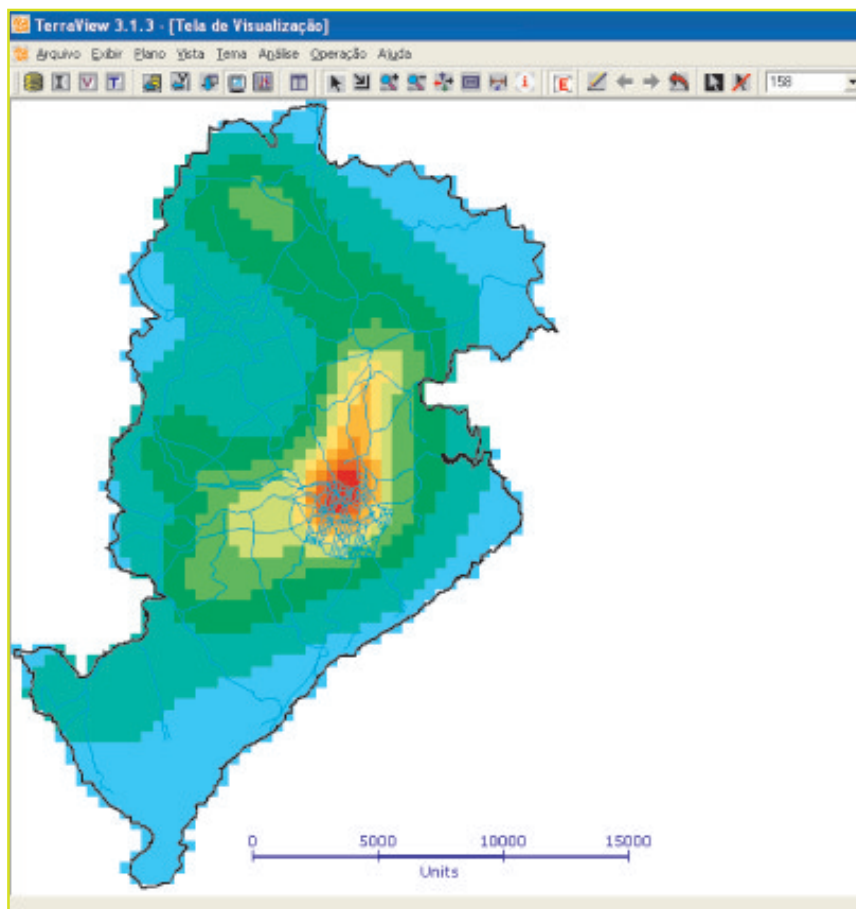
Varie também o raio. Para isso, desmarque a opção “Adaptativo” nos parâmetros do mapa de Kernel e atribua diferentes valores para o raio. Quanto maior o raio, mais suavizada será a superfície estimada. Experimente utilizar um raio de 1.000 metros.

Pode-se ainda ponderar a estimativa utilizando atributos, como o número de vítimas, um escore de gravidade, ou o tempo decorrido até a chegada do socorro.

A sobreposição de outros temas ajuda a caracterizar os locais de maior intensidade, como por exemplo: as vias de tráfego intenso, a localização das ambulâncias do SAMU e as portas de entrada para trauma.

Experimente sobrepor os pontos de apoio das ambulâncias sobre um mapa de Kernel dos acidentes (Figura 2.3).

FIGURA 2.3 – Superfície de densidade dos diversos tipos de acidentes e pontos de apoio das ambulâncias.



■ **Considerações:**

Nos exemplos aqui demonstrados, realizamos alguns procedimentos exploratórios, buscando evidenciar áreas críticas quanto à ocorrência de acidentes de trânsito, que podem subsidiar propostas para a melhoria da segurança no trânsito.

Não devemos, entretanto, desconsiderar que o SAMU atende outros tipos de traumas e urgências clínicas, e é todo este conjunto que determina o planejamento deste serviço.

Exercício 3 : Análise espacial da distribuição de casos de malária no Estado de Mato Grosso em 2000 e 2003.

I - Introdução

No contexto malárico e das propostas de organização dos serviços de saúde descentralizados, especialmente no Setor de Vigilância em Saúde, a identificação das áreas geográficas e grupos populacionais que estão expostos ao maior risco de malária constitui-se em etapa importante para a elaboração de programas de controle, monitoramento e de avaliação em níveis local, regional e nacional. Os métodos de análise espaciais são particularmente úteis para a criação ou delimitação de áreas homogêneas, definição de critérios de monitoramento e avaliação para uma determinada intervenção, estabelecimento de prioridades para planejamento e alocação de recursos (Cruz, 1996). Considerando o processo de descentralização, a unidade de análise importante é a escala do municipal. No entanto, a ocorrência da doença muitas vezes não se restringe aos limites definidos político-administrativamente. Dessa forma, questões como a existência de padrão de distribuição da malária considerando o município e os municípios vizinhos, e a relação entre a ocorrência da malária no município e aqueles circunvizinhos passam a assumir relevância na identificação de áreas homogêneas para definição de estratégias de intervenção.

II – Exercício

Neste exercício faremos a análise espacial dos casos de malária notificados pelos municípios do Estado de Mato Grosso, durante o ano de 2000 e 2003. O objetivo do mesmo é apresentar técnicas e respectivos procedimentos para o desenvolvimento da análise dos dados espaciais e visualização de áreas em mapas. Com isso, espera-se apresentar uma metodologia que auxilie no encontro de áreas homogêneas para auxiliar os gestores na definição de áreas prioritárias para aplicação de recursos e planejamento de intervenções, de acordo com as características específicas de cada área.

Para a realização deste exercício, os seguintes dados foram utilizados, direta ou indiretamente:

- **Malha de municípios de Mato Grosso** – limites (polígonos) dos municípios, de ano 1997 e 2001.

Formato: vetorial (extensões: *.shp*, *.dbf* e *.shx*).

Fonte: Instituto Brasileiro de Geografia e Estatística – IBGE. Outro formato (*.bna*) é disponibilizado pelo DATASUS. Esclarecemos que as duas malhas são necessárias devido à criação de novos municípios, especialmente em 2000. Dessa maneira, para a análise da IPA (Incidência Parasitária Anual) de 2000 será utilizada a malha digital de 1997, e para a IPA de 2003, a malha de 2001.

- **População de 2000** – população residente nos municípios do estado.

Formato: tabular (extensão: *.xls*).

Fonte: Instituto Brasileiro de Geografia e Estatística – IBGE. Censo Demográfico 2000.

- **População em 2003** – estimativa da população residente nos municípios de Mato Grosso, em 2003.

Formato: tabular (extensão: *.xls*).

Fonte: Instituto Brasileiro de Geografia e Estatística – IBGE. Estimativa Populacional Municipal de 2002.

- **Incidência Parasitária Anual (IPA)** – calculada com base nas lâminas positivas notificadas pelos municípios para os anos de 2000 e 2003, dividida pela respectiva população e multiplicado por mil.

Formato: tabular (extensão *.xls*).

Fonte: lâminas positivas por municípios de 2000, dados obtidos na Secretaria do Estado de Mato Grosso, armazenados no Sistema de Informações de Malária (SISMAL); os de 2003 foram obtidos no Sistema de Informações de Vigilância Epidemiológica – Malária (SIVEP-Malária), da Secretaria de Vigilância em Saúde do Ministério da Saúde. Neste exercício, as IPA de 2000 e 2003 já estão calculadas, dispostas em planilha (extensão: *...xls* e *... csv*).

Banco de dados:

Os dados adquiridos foram importados e organizados em um banco do tipo Access criado pelo aplicativo TerraView. Para facilitar a condução do exercício, disponibilizamos o banco de dados (*IPA_MT00-03*) e as informações necessárias ao desenvolvimento das atividades, no sítio www.geocapacita.cict.fiocruz.br.

A visualização e análise do banco de dados deve ser feita por meio do aplicativo geográfico TerraView, sendo, portanto, necessário instalar a versão mais recente deste aplicativo. A aquisição do arquivo de instalação, do manual e do tutorial do TerraView pode ser feita (www.dpi.inpe.br/terraview) da Divisão de Processamento de Imagens do Instituto Nacional de Pesquisas Espaciais – DPI/INPE.

A - Abrindo o banco e visualizando os dados:


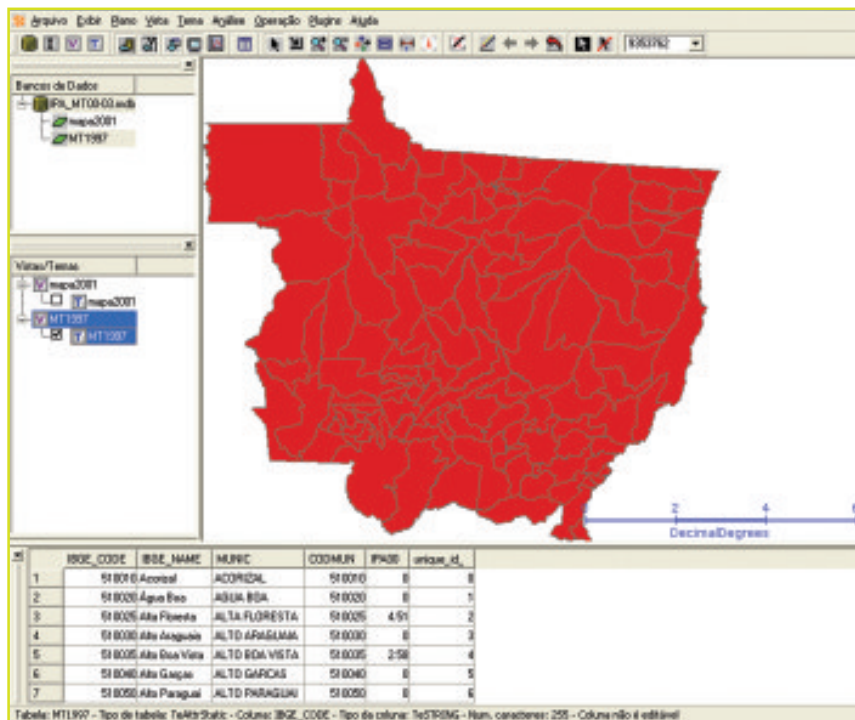
1. Inicialmente, **executar** (abrir) o aplicativo TerraView;
2. **Conectar** o banco de dados  na barra de ferramentas;
3. Selecionar o banco de dados **IPA_MT00-03.mdb**;
4. Visualiza-se o mapa e a tabela de atributos conforme apresentado na figura 3.1;
5. **Importar** dados **mapa2001.bna**;
6. **Importar** dados **MT1997.bna**;
7. Na árvore **Vista/tema** selecionar tema **mapa2001.bna**, importar a tabela **IPAMT03** (arquivo .csv; é importante lembrar que o tipo de variável deve ser alterado e confirmado quando necessário);
8. **Criar** tabela externa **IPAMT03**;
9. **Conectar** tabela externa **IPAMT03**;
10. Na árvore **Vista/tema** selecionar tema **MT1997.bna**, **importar** a tabela **IPAMT00** (arquivo .csv);
11. **Criar** tabela externa **IPAMT00**;
12. **Conectar** tabela externa **IPAMT00**.

FIGURA 3.1 – Visualização dos dados do banco **IPA_MT00-03.mdb** no TerraView.



A malha dos municípios do Estado de Mato Grosso será apresentada na janela de visualização de mapas e os dados tabulares, referentes ao tema *mapa MT2001*, serão mostrados na tabela de atributos. A descrição dos dados dessa tabela é a seguinte:

- IBGE_CODE – código (IBGE) de identificação dos municípios;
- IBGE_NAME – nome dos municípios;


- MUNIC - nome dos municípios da tabela externa conectada;
- CODMUN - código (IBGE) de identificação dos municípios da tabela externa;
- IPA03 – IPA (Incidência Parasitária Anual) de 2003, na tabela externa;
- No tema MT1997, encontra-se a IPA, na tabela externa, do ano 2000 (IPA00).

ANALISANDO OS DADOS

A definição de vizinhança aqui adotada levou em conta os municípios que **compartilham a mesma fronteira** sendo, neste caso, considerados como vizinhos de “primeira ordem”. A primeira questão a ser verificada é se a IPA observada em um município possui correlação com os valores das IPAs apresentadas pelos municípios vizinhos, considerando o estado. Para verificar essa correlação optou-se pelo cálculo dos Índices de Moran Global e Local.

B – Calculando índices de Moran Global e Local.

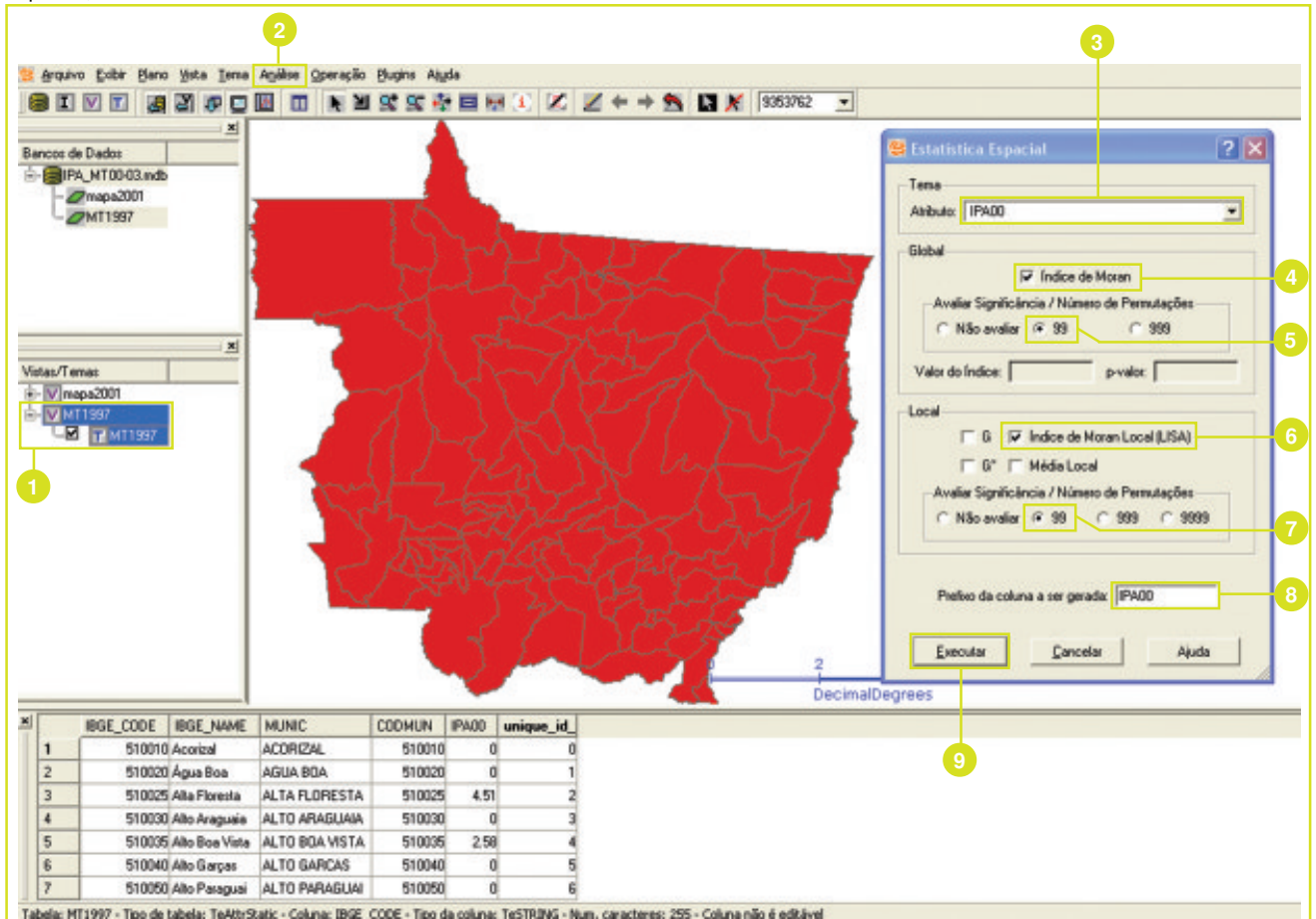
Para se conhecer os índices de Moran Global e Local das IPAs de 2000 e 2003, efetue os seguintes passos, apresentados na figura 3.2:

1. Na árvore de **vistas/temas**, selecione e habilite o tema **MT1997**;
2. Desloque o cursor de apontamento  para **Análise** na tela de visualização aberta e clique. Selecione a opção **Estatística espacial**;
3. Na janela da **Estatística Espacial**, selecionar do tema o **atributo IPA00**;
4. Selecionar o **Índice de Moran Global** (caso tenha interesse em detectar indícios de correlação espacial da IPA dos municípios em relação aos seus vizinhos, considerando o estado como um todo);
5. Em **Índice de Moran Global**, selecionar o **nível de significância**, neste caso optou-se por *99*;
6. Em **Índice de Moran Local**, selecionar o **Índice de Moran Local (LISA)**;
7. Selecionar o **nível de significância**, *99* para o Índice de Moran Local;
8. Nomear o prefixo da coluna a ser gerada, como por exemplo IPA00;
9. **Executar**.

Nesta operação o TerraView calcula a matriz de vizinhança (proximidade) automaticamente, levando em consideração os polígonos (municípios) contíguos.

Repetir a seqüência anterior selecionando o **mapa2001.bna** e o atributo **IPA03**. Nomear a coluna gerada como **IPA03**.

FIGURA 3.2 – Etapas da execução de cálculos dos Índices de Moran, na análise espacial, no TerraView.



O Índice de Moran Global da IPA obtido foi igual a 0,223582 (p-valor = 0,01) em 2000 e 0,184715 (p-valor = 0,02) em 2003. Neste exercício, o valor de I encontrado foi maior que 0 (zero), o que indica evidência de autocorrelação espacial positiva, isto é, as áreas tendem a ser similares entre si quanto à variável IPA. No entanto, por fornecer um único valor, como medida de associação espacial para todo o conjunto de dados, é necessário decompor este índice global de associação espacial e examinar padrões numa escala de maior detalhe para evidenciar agrupamentos (leia-se municípios) de incidência de malária semelhantes (*clusters*) e/ou discrepantes (*outliers*).

Inicialmente, vamos realizar análise visual no mapa temático para verificar a existência de padrão espacial construído a partir de estratificação de risco de transmissão. O Programa Nacional de Controle da Malária – PNCM caracteriza as áreas do país de acordo com os seguintes estratos:

- **Áreas de baixo risco de transmissão** – regiões, estados ou municípios com taxa de incidência de lâmina positiva menor que 10 por 1.000 habitantes, no período de um ano;
- **Áreas de médio risco de transmissão** – regiões, estados ou municípios com taxa de incidência de lâmina positiva maior ou igual a 10 e menor que 50 lâminas positivas por 1.000 habitantes, no período de um ano;

• **Áreas de alto risco de transmissão** – regiões, estados ou municípios com taxa de incidência maior ou igual a 50 lâminas positivas por 1.000 habitantes, no período de um ano.

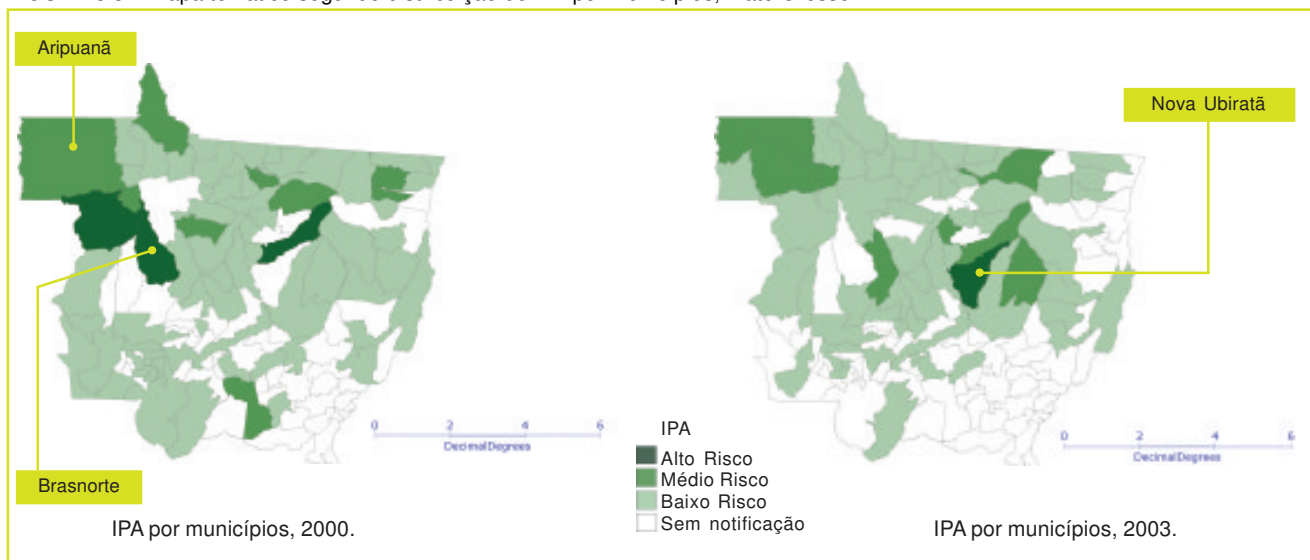
C – Estratificando para mapa temático.

Para a visualização do mapa temático recorrendo a estratificação anterior, utilizaremos **Editar Legenda**, conforme os passos a seguir, apresentados na figura 3.3:

1. Na árvore de **vistas/temas**, selecione e habilite o tema **MT1997**;
2. Pressione o botão direito do *mouse* e selecione **Editar legenda**;
3. Em **Parâmetro de Agrupamento**:
 - a) Selecionar o modo **Passos Iguais**;
 - b) Selecionar o atributo **IPAMT00.IPA00**;
 - c) Definir por **4 fatias** e **precisão 2**;
4. Escolher **4 cores**;
5. **Aplicar**.
6. Modificar os agrupamentos para:
 - a) De 0.00 para 0.01;
 - b) De 0.01 para 10.00;
 - c) De 10.00 para 50.00;
 - d) De 50.00 para 160.00;
7. **Executar**;
8. Pode-se, clicando com o *mouse* esquerdo, **renomear a legenda e/ou alterar o visual**. Pode-se renomear a legenda conforme exemplo abaixo:
 - a) Sem notificação;
 - b) Baixo Risco;
 - c) Médio Risco;
 - d) Alto Risco;
9. Salvar tela de visualização em arquivo *.bmp*.

Na definição do primeiro intervalo, correspondente aos municípios sem notificação, utilizamos os seguintes valores: De = 0.00 e Para = 0.01. Fizemos isso porque o TerraView não considera o valor apresentado na janela “para”, mas sim o imediatamente anterior a este. Por isso, se for digitado 0.00 o programa não selecionará nenhum registro com valor 0.00. **Faça um teste!**

FIGURA 3.3 – Mapa temático segundo distribuição de IPA por municípios, Mato Grosso.



Observa-se na figura 3.3, que a ocorrência da malária está dispersa, porém os municípios com maior número de notificações estão localizados na região norte do estado. Na região sul, há municípios que não apresentam casos nos dois anos estudados. O que chama atenção é o Município de Nova Ubiratã, que em 2000 não notificou nenhum caso e em 2003 passa a figurar como município de alto risco de transmissão. Desempenho contrário é observado em Brasnorte, que passa de alto risco para nenhuma notificação.

A distribuição de casos por estratos de risco nos indica em 2000 a existência de 9 municípios classificados como de médio risco e 3 municípios de alto risco. Essa configuração é modificada em 2003, com a existência de 8 municípios de médio risco e 1 de alto risco. O Município de Aripuanã foi desmembrado em Municípios de Colniza, Rondolândia e Aripuanã. Esta área aparentemente manteve o quadro de médio risco, exceto Rondolândia. Essa descrição ilustra a mobilidade existente na demarcação dos contornos territoriais definidos político-administrativamente. E, no nosso caso, a adoção de municípios como unidade de análise requer precaução principalmente quando se busca acompanhar/monitorar o comportamento e a dinâmica da endêmica ao longo do tempo.

O mapa temático já nos indica que a distribuição espacial da malária no estado não ocorre de forma homogênea. A heterogeneidade observada visualmente reafirma a necessidade de aprofundar a análise em maior escala. Assim, após a localização e classificação dos municípios do estado por estratos de risco, questiona-se:

- Será que os casos de malária, apresentados pelos municípios, têm relação com os casos de malária dos vizinhos? Ou vice-versa?

Para responder a essa questão escolhemos o **Índice de Moran Local**, mais especificamente o **Box Map**, como recurso para a identificação/localização de áreas homogêneas constituídas por municípios com associação espacial. Esse é construído com base nos valores normalizados (valores dos atributos – IPA – subtraídos de sua média e divididos pelo desvio-padrão), no qual o atributo foi classificado conforme sua posição em relação aos quadrantes do **Diagrama de Espalhamento de Moran**, recebendo uma cor correspondente quando gerado o mapa (Câmara, Carvalho, Cruz & Correa, 2004). Dessa forma, classificamos os municípios de acordo com sua localização nos quadrantes do Diagrama de Espalhamento de Moran, em áreas de agrupamentos de municípios homogêneos, consideradas como: quadrante 1 (+ / +) e quadrante 2 (- / -) para áreas de associação espacial positiva, e quadrante 3 (+ / -) ou quadrante 4 (- / +) para áreas de associação espacial negativa.

O quadrante 1 (+ / +) indica que há presença de município com valor positivo da IPA normalizada e possui a média dos valores normalizados dos municípios vizinhos também positiva, isto é, ambos os valores são positivos; o quadrante

2 (– / –) localiza os municípios com valor negativo da IPA normalizada e possui a média dos valores normalizados dos municípios vizinhos também negativos, isto é, ambos os valores são negativos. Dessa forma, os quadrantes 1 e 2 (Q1 e Q2) indicam áreas de associação espacial positivas, levando-se em conta que o município possui vizinhos com valores semelhantes (Câmara, Carvalho, Cruz & Correa, 2004).

O quadrante 3 (+ / –) indica município com valor positivo da IPA normalizada e possui a média dos valores normalizados dos municípios vizinhos negativos e; quadrante 4 (– / +) distribui os municípios com valor negativo da IPA normalizada e possui a média dos valores normalizados dos municípios vizinhos com valor positivo. Tanto o quadrante 3 (Q3) como o 4 (Q4) não acompanham a tendência global e indicam áreas de associação espacial negativas, considerando que numa localização há vizinhos com valores distintos, considerada como região de transição entre os dois regimes espaciais, no caso os municípios dos quadrantes 1 e 2 (Câmara, Carvalho, Cruz & Correa, 2004).

Adotou-se, com base no mapa gerado pelo **Box Map**, quatro áreas distintas de prioridade para o controle da malária:

- área de maior prioridade para controle: municípios agregados no quadrante Q1 (+ / +);
- área de menor prioridade para controle: municípios agregados no quadrante Q2 (– / –);
- área de prioridade intermediária para controle: municípios agregados no quadrante Q3 (+ / –);
- área de prioridade intermediária para controle: municípios agregados no quadrante Q4 (– / +).

Antes de gerar um novo mapa é importante que seja criado um novo tema com as mesmas informações. Dessa forma não perdemos o mapa temático gerado na etapa anterior.

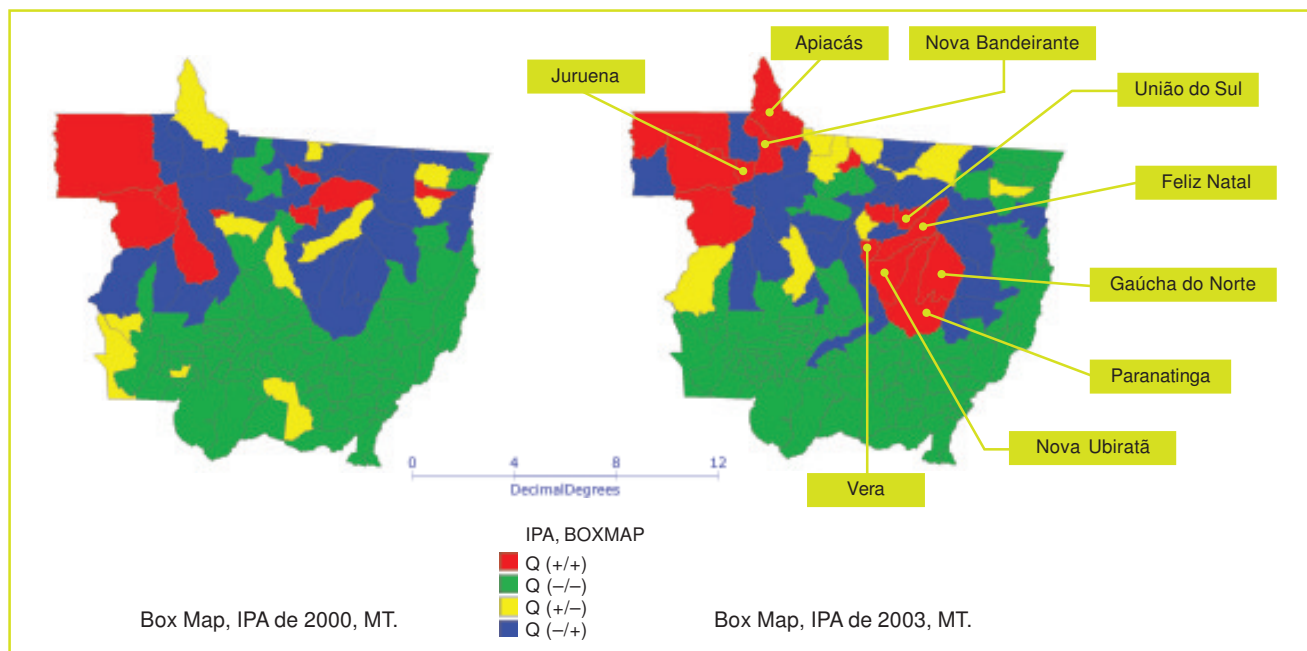
D – Visualizando o Box Map.

Para a visualização do **Box Map** execute os passos a seguir, o resultado é apresentado na figura 3.4:

1. Marque o tema **MT1997** e clique com o botão esquerdo do mouse para acionar o *menu*;
2. Na árvore de vistas/temas, selecione e habilite o tema **MT1997**;
3. Clique em **Criar tema a partir de tema** (tenha certeza de que estejam selecionados todos os objetos) e nomear como **Box00**;
4. Para a visualização do mapa Box Map, pressione o botão direito do *mouse* e selecione **Editar legenda**;
5. Em **Parâmetro de Agrupamento**:
 - a) Selecionar o modo **Valor Único**;

- b) Selecionar o atributo **MT1997.IPA00BoxMap**;
 - c) Definir por **precisão 1**;
 - d) Escolher **4 cores**;
- Aplicar.**
6. Modificar os valores e os rótulo para:
 - a) Valor = 1, rótulo = Q(+ / +);
 - b) Valor = 2, rótulo = Q(- / -);
 - c) Valor = 3, rótulo = Q(+ / -);
 - d) Valor = 4, rótulo = Q(- / +);
 7. **Executar**;
 8. Pode-se, clicando com o *mouse* esquerdo, **renomear a legenda e/ou alterar o visual**. Pode-se renomear a legenda conforme exemplo abaixo:
 - a) Área de maior prioridade;
 - b) Área de menor prioridade;
 - c) Área de prioridade intermediária Q3;
 - d) Área de prioridade intermediária Q4;
 9. Salvar tela de visualização em arquivo *.bmp*.

FIGURA 3.4 – Box Map da ocorrência da IPA nos municípios de Mato Grosso.



Na figura 3.4 observa-se que 10 municípios em 2000 foram classificados como de alta prioridade para controle e 14 em 2003. Pelo Índice de Moran Local, os municípios de Apiacás e Feliz Natal passaram de Área de prioridade intermediária Q3 para Área de maior prioridade Q1, isto é, apresentavam em 2000, IPA acima da média que seus vizinhos, e em 2003 ambos (os municípios e seus vizinhos) apresentaram os valores positivos. Destacam-se também municípios de prioridade intermediária Q4 (IPA normalizada menor que seus vizinhos) em 2000, modificarem o *status* para municípios de maior prioridade.

Os municípios de prioridade intermediária (seja ela Q3 ou Q4), estão em sua maioria entre os de menor e maior prioridade e são conhecidos como áreas de transição. Esta visualização destaca áreas de transição, de maior ou menor prioridade, e possibilita monitorar a dinâmica do evento e definição de estratégias de intervenção adequadas a especificidades locais, considerando a vizinhança.

O **Moran Map** foi construído para visualizar no interior de cada área de prioridade identificada pelo **Box Map**, áreas ou municípios onde a dependência espacial é mais pronunciada, destacando áreas de autocorrelação espacial estatisticamente significativa. O Moran Map é a representação do Índice de Moran Local, onde os municípios são classificados em 4 quadrantes (Q + / +, Q - / -, Q + / - e Q - / +) e municípios não-significantes estatisticamente (Câmara, Carvalho, Cruz & Correa, 2004). Os municípios de Q + / + são considerados como áreas críticas de ocorrência de malária.

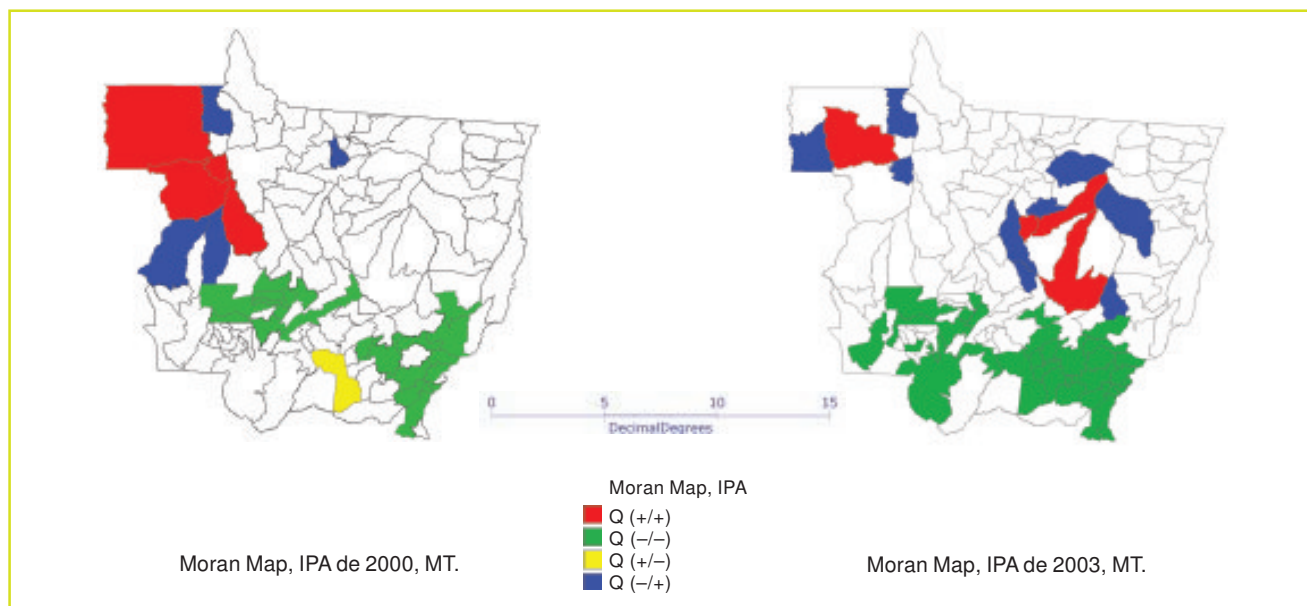
E – Visualizando o Moran Map.

Para a visualização do mapa **Moran Map**, utilizaremos **Editar Legenda**, execute os passos a seguir, o resultado é apresentado na figura 3.5:

1. Na árvore de vistas/temas, selecione e habilite o tema **MT1997**;
2. Pressione o botão direito do *mouse* e selecione **Editar legenda**;
3. Em **Parâmetro de Agrupamento**:
 - a) Selecionar o modo **Valor Único**;
 - b) Selecionar o atributo **MT1997.IPA00MoranMap**;
 - c) Definir por **precisão 1**;
 - d) Escolher **5 cores**;
4. **Aplicar**.
5. Modificar os valores e os rótulos para:
 - a) Valor = 0, rótulo = não significativa
 - b) Valor = 1, rótulo = Q(+ / +);
 - c) Valor = 2, rótulo = Q(- / -);
 - d) Valor = 3, rótulo = Q(+ / -);
 - e) Valor = 4, rótulo = Q(- / +);
6. **Executar**;
7. Pode-se, clicando com o *mouse* esquerdo, **renomear a legenda e/ou alterar o visual**. Pode-se renomear a legenda conforme exemplo abaixo:
 - a) Não Significante
 - b) Área de maior prioridade;
 - c) Área de menor prioridade;
 - d) Área de prioridade intermediária Q3;
 - e) Área de prioridade intermediária Q4;
8. Salvar tela de visualização em arquivo *.bmp*.

Na visualização do padrão local gerado pelo Moran Map notamos agrupamentos de autocorrelação espacial local estaticamente significativa, apontando-os como o local onde a dependência espacial é ainda mais pronunciada, seja de área de maior prioridade, menor prioridade ou prioridade intermediária. A importância da sua utilidade para monitoramento ou estratificação de áreas de risco reside na identificação de municípios críticos, no caso de áreas de maior risco.

FIGURA 3.5 – Moran Map da ocorrência da IPA nos municípios de Mato Grosso.



O recurso para estratificação de prioridade utilizando o Box Map propicia estabelecer metas que considerem a incidência do município principal e de seus vizinhos. Essas metas necessitam ser diferenciadas de acordo com as áreas ou agrupamentos de prioridade. Dessa forma, a curto prazo, (metas anuais) poderia-se estabelecer as seguintes metas:

- os municípios localizados na área intermediária de prioridade Q3 determinariam como meta a redução da IPA para valores próximos ou menores que a média de seus vizinhos;

- os municípios localizados na área intermediária de prioridade Q4 teriam a meta de reduzir a IPA para valores próximos à média dos municípios de menor prioridade e, empenhar esforços para não sofrer a influência de seus vizinhos e aumentar a IPA;

- os municípios localizados na área de maior prioridade estabeleceriam como objetivo reduzir a IPA, procurando se aproximar aos valores apresentados pelos municípios vizinhos localizados na área intermediária de prioridade Q4, ou reduzir a IPA para valores próximos à média dos municípios de menor prioridade;

- os municípios localizados na área de menor prioridade estabeleceriam como meta a eliminação de casos autóctones e reforçariam o sistema de vigilância e controle da malária na detecção precoce de casos importados.

Referências Bibliográficas

- Bailey TC, Gatrell AC, 1995. Interactive spatial data analysis. 1 ed. Longman Group Limited, Essex, 413 pp.
- Barcellos C, Barbosa KC, Pina MF, Magalhães MMAF, Paola JCMD, Santos SM, 1998. Interrelacionamento de dados ambientais e de saúde: Análise de risco à saúde aplicada ao abastecimento de água no Rio de Janeiro utilizando Sistemas de Informações Geográficas, *Cadernos de Saúde Pública*, 14(3) 597-605
- Barcellos C, Acosta LMW, Lisboa EP, Brito MRV, Flores R, 2006. Estimativa da prevalência de HIV em gestantes em micro-áreas de Porto Alegre através de técnicas de análise espacial. *Revista de Saúde Pública*, 40(5): 928 - 930.
- Besag J, Newell J, 1991. The detection of clusters in rare diseases. *J. R. Stat. Soc. A*, 154: 143-155.
- Box, GEP. 1979. Robustness is the strategy of scientific model building. In *Robustness in Statistics*, (Launer, RL & Wilkinson, GN, eds), pp. 201-236, Academic Press, New York, NY.
- Câmara G, Monteiro AMV, Druck S, Carvalho MS, 2002. Análise espacial e geoprocessamento. In S Druck, MS Carvalho, G Câmara and AMV Monteiro, Análise espacial de dados geográficos. Disponível em <http://www.dpi.inpe.br/gilberto/livro/analise/>. Acessado em 17 de setembro de 2003.
- Câmara G, Carvalho MS, Cruz OG, Correa V, 2004. Análise espacial de áreas. In. Embrapa Cerrados. Análise espacial de dados geográficos. Planaltina, DF: Embrapa; 2004. p.157-209.
- Carvalho MS, Câmara G, 2002. Análise de eventos pontuais. In S Druck, MS Carvalho, G Câmara and AMV Monteiro, Análise espacial de dados geográficos. Disponível em <http://www.dpi.inpe.br/gilberto/livro/analise/>. Acessado em 17 de setembro de 2003.
- Cavalcante MS, 2004. Transmissão Vertical do HIV em Fortaleza: Revelando a Situação Epidemiológica em uma Capital do Nordeste. RBGO: Revista da Federação Brasileira das Sociedades de Ginecologia e Obstetrícia/FEBRASCO., -V.6, n.2, São Paulo.
- Cromley EK, McLafferty SL, 2002. GIS and Public Health. 1 ed. The Guilford Press, New York, 340 pp.
- Everitt B, 1974. Cluster Analysis. Heinemann Educational Books, London.
- Gatrell AC, Bailey TC, 1996. Interactive spatial data analysis in medical geography. *Soc Sci Med*, 42: 843-855.
- Kelsall JE, Diggle PJ, 1995. Non-parametric estimation of spatial variation in relative risk. *Stat Med*, 14: 2335-2342.

- King PE, 1979. Problems of spatial analysis in geographical epidemiology. *Social Science & Medicine*, 13: 249-252.
- Kulldorff M, 1997. A spatial scan statistic. *Commun. Statist. - Theory Meth*, 26: 1481-1496.
- Kulldorff M, 2003. SaTScan™ user Guide for version 3.1. Disponível em <http://www.satscan.org>. Acessado em 21 de junho de 2003.
- Lawson AB, Kulldorff M, 1999. A review of cluster detection methods. In AB Lawson, A Biggeri, D Böhning, E Lesaffre, JF Viel and R Bertollini, Disease mapping and risk assessment for public health, John Wiley and Sons, Chichester. p. 99-110.
- Levine N, 2002. CrimeStat: A spatial statistics program for the analysis of crime incident locations. Ned Levine & Associates, Houston, TX, and the National Institute of Justice. Washington, DC.
- MS (Ministério da Saúde), 2002. Boletim Epidemiológico da AIDS/Ministério da Saúde. Coordenação Nacional de DST e Aids. Ano XVI, nº 01, 14ª a 52ª semanas epidemiológicas abril a dezembro 2002. Brasília. <http://www.opas.org.br/sistema/arquivos/matriz.pdf> Acessado em 27/out/2004.
- MS (Ministério da Saúde), 2003. Vigilância em Saúde. Dados e Indicadores Seleccionados 2003. Brasília.
- Neri SHA, 2004. A utilização das ferramentas de geoprocessamento para identificação de comunidades expostas a hepatite A nas áreas de ressacas dos municípios de Macapá e Santana (AP). Dissertação de mestrado. COPPE/UFRJ. Rio de Janeiro
- Rede Interagencial de Informação para a Saúde (RIPSA), 2002. Indicadores básicos de saúde no Brasil: conceitos e aplicações. Brasília: OPAS, 2002.
- Sabel CE, Gatrell AC, Loytonen M, Maasilta P, Jokelainen M, 2000. Modelling exposure opportunities: estimating relative risk for motor neurone disease in Finland. *Soc Sci Med*, 50: 1121-1137.
- Sankoh OA, Ye Y, Sauerborn R, Muller O, Becher H, 2001. Clustering of childhood mortality in rural Burkina Faso. *Int J Epidemiol*, 30: 485-492.
- Santana R, Dhalia C, Barreira D, Barbosa A, PAZ, Cardoso A, 2002. Implantação da vigilância epidemiológica de gestantes HIV+ e crianças expostas. *Revista Brasileira de Epidemiologia*, Suplemento Especial; Comunicações Coordenadas p.73, Curitiba.
- Santos SM, Barcellos C, Carvalho MS, Flores R, 2001. Detecção de aglomerados espaciais de óbitos por causas violentas em Porto Alegre, Rio Grande do Sul, Brasil, 1996. *Cad Saúde Pública*, 17: 1141-1151.
- Silverman BW, 1986. Density estimation for statistics and data analysis. Chapman and Hall, London.
- SMS (Secretaria Municipal de Saúde de Porto Alegre). PRÁ-SABER: Informações de Interesse à Saúde. Volume 4 nº 3, dezembro de 1999, Porto Alegre.
- Snow J, 1999. Sobre a Maneira de Transmissão do Cólera. Rio da Janeiro/São Paulo: ABRASCO/HUCITEC.
- Tobler, WR, 1979. Cellular geography. *Philosophy in geography*, pp. 379-386.
- Ward MP, Carpenter TE, 2000. Techniques for analysis of disease clustering in space and in time in veterinary epidemiology. *Prev Vet Med* 45: 257-284.

Links úteis para *download* de aplicativos recomendados

Crimestat - <http://www.icpsr.umich.edu/CRIMESTAT/>

Google Earth - <http://earth.google.com/intl/pt/>

SatScan - <http://www.satscan.org/>

TerraView - <http://www.dpi.inpe.br/terraview/index.php>

WinBugs - <http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml>

Outros títulos da série: CAPACITAÇÃO E ATUALIZAÇÃO EM GEOPROCESSAMENTO EM SAÚDE

VOLUME 1 – Abordagens Espaciais na Saúde Pública

Organizadores: Simone M. Santos e Christovam Barcellos

VOLUME 2 – Sistemas de Informações Geográficas e Análise Espacial na Saúde Pública

Organizadores: Simone M. Santos e Reinaldo Souza-Santos

VOLUME 3 – Introdução à Estatística Espacial para a Saúde Pública

Organizadores: Simone M. Santos e Wayner V. Souza

MAIS INFORMAÇÕES: Secretaria de Vigilância em Saúde

Ministério da Saúde

Esplanada dos Ministérios

edifício sede, Bloco E, 1º andar

Brasília - DF CEP 70.058-900

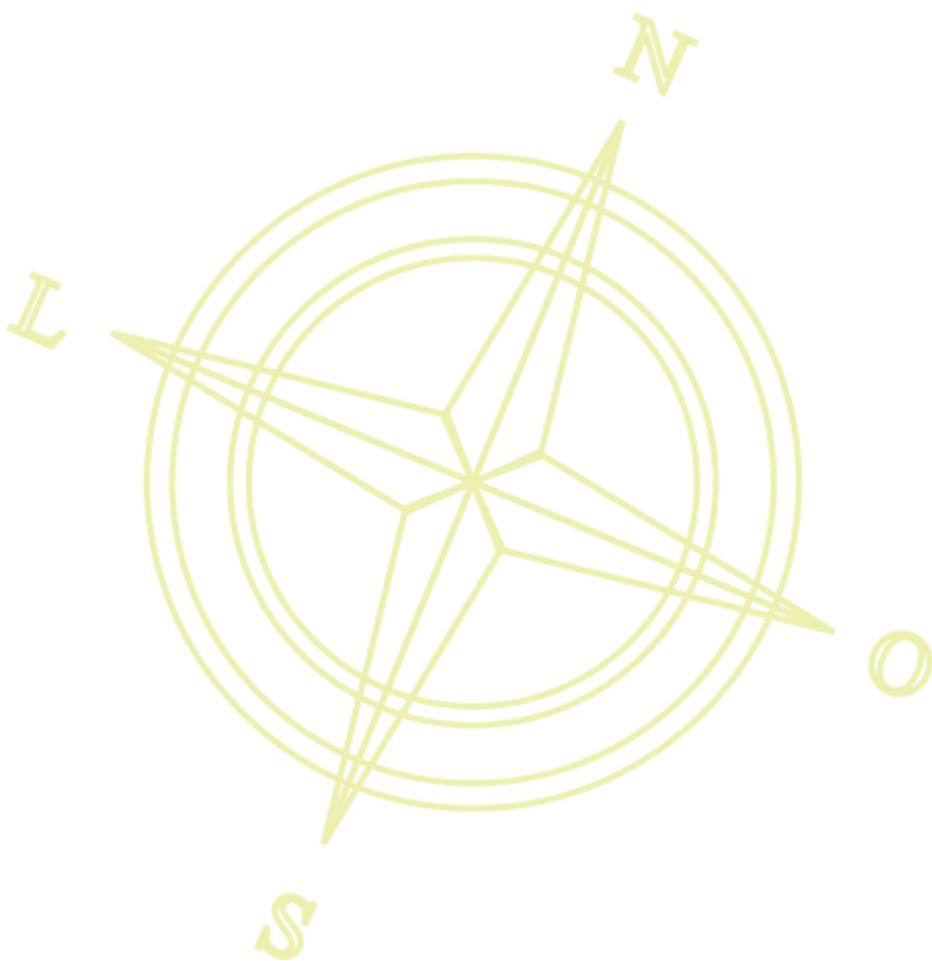
e-mail: svs@saude.gov.br

Tel: (61) 3315.3777

Visite nosso site:

www.capacita.geosaude.cict.fiocruz.br





Esta obra foi produzida no Rio de Janeiro pelo Departamento de Programação Visual – ICICT – Fiocruz, em outubro de 2007. A composição empregou as tipologias Arial, Arial Narrow, Arial Black e Zapf Humanist 601. O papel utilizado para miolo é o Couchè fosco L2 - 95 g/m². E, para a capa, Supremo Alto Alvura 275 g/m².

Série: Capacitação e atualização em
geoprocessamento em saúde

Volume 3

Introdução à Estatística Espacial para a Saúde Pública

O desenvolvimento de novas técnicas de análise espacial, da computação, da cartografia digital e dos Sistemas de Informações Geográficas, possibilitou a expansão da Epidemiologia Geográfica. A facilitação do uso de técnicas de análise que permitem inferências válidas sobre o papel do nível ecológico, impulsiona a retomada desse tipo de estudo. Soma-se a esse movimento, a formação de recursos humanos em epidemiologia, ampliando o crescimento qualitativo e quantitativo de profissionais que atuam na área.

A epidemiologia brasileira recebe este volume, organizado por Simone Santos e Wayner Souza, em momento de intenso desenvolvimento do tema na Pós-graduação em Saúde Coletiva. A abrangência dos conteúdos e a clareza do texto tomam mais este volume um companheiro obrigatório de quem se aventura a investigar questões de saúde na sociedade brasileira.

*Comissão de Epidemiologia da ABRASCO
Associação Brasileira de Pós-graduação em Saúde Coletiva*



Secretaria de
Vigilância em Saúde

Ministério
da Saúde



ISBN 978-85-334-1427-3



9 788533 414273